Fakultät für Luftfahrt, Raumfahrt und Geodäsie
Signalverarbeitung in der Erdbeobachtung

From Remote Sensing Data to Urban Patterns: A Topology Guided Data Fusion Paradigm

Jingliang Hu, M.Sc.

Vollständiger Abdruck der von der Fakultät für Luftfahrt, Raumfahrt und Geodäsie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation

Vorsitzender: Prof. Dr.phil.nat. Urs Hugentobler

Prüfer der Dissertation:

1. Prof. Dr.-Ing. habil. Xiaoxiang Zhu
2. Prof. Dr.-Ing. habil. Richard Bamler
3. Prof. Dr.-Ing. Peter Reinartz
   Universität Osnabrück

Die Dissertation wurde am 22.01.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Luftfahrt, Raumfahrt und Geodäsie am 15.07.2020 angenommen.

Abstract

The United Nations (UN) reported in *'The World's Cities in 2018'* that 55.3% of the population on Earth lived in urban settlement in 2018 and an estimation of this number is 60% for the year of 2030. Urban settlement is so relevant to human development that the UN listed urban sustainable development as one its 17 sustainable development goals (SDGs) in 2015. To accomplish the urban sustainable development, it is of great importance to retrieve strategic urban geographic information to support urban development.

Remote sensing technologies appear as solutions of retrieving urban geographic information on a large scale with timely updates. Current state-of-the-art remote sensing products that map urban settlements on a global scale are Global Urban Footprint (GUF), Global Human Settlement Layer (GHSL), and World Settlement Footprint 2015 (WSF). The availability of these layers have invaluable contributions on providing crucial urban information. However, these products only have a binary urban/non-urban scheme. The simple scheme is universally applicable because of a high generalization, but detailed information inside urban areas is missing. The urban layer of the next generation shall have a classification scheme that provides detailed information inside urban areas and is universally applicable at the same time. Regarding data sources, the GUF is derived from synthetic aperture radar (SAR) data, the GHSL is derived from optical data, and the recent WSF applies both data. Future products will take advantages of both data sources as WSF. Among different types of optical data, hyperspectral data carry a rich spectral information and is an ideal optical data for classification tasks. And polarimetric SAR (PolSAR) data is often the choice among SAR data for classification tasks. Therefore, this dissertation aims at urban classification tasks of a finer scheme than urban/non-urban categorization and utilize both optical and SAR data, mainly hyperspectral and PolSAR data.

Regarding the topic of fusing optical and SAR data for the purpose of urban classification, some of the major challenges are posed by, (1) the severe difference between optical and SAR data; (2) the high dimension of remote sensing data; and (3) the lack of data annotations. Manifold-based data fusion techniques is a promising solution to the topic with respect to the listed three challenges. First, the technique is capable of bridging gaps crossing data domains which is helpful for dealing the difference between optical and SAR data. Second, the intrinsic low-dimensional property of the manifold theory is ideal to handle high dimensional remote sensing data. At last, the semi-supervised manifold alignment associates the labeled and unlabeled data samples via manifold structures so that the semantic information carried by annotations is implicitly propagated to unlabeled data. It eventually amplifies the impact of the data annotations. Therefore, the methodological goal of this dissertation is to explore manifold-based techniques for data fusion and classification of urban areas.

Four methodological objectives were achieved in the dissertation. They are listed as follows.

A data fusion algorithm, named MIMA, was developed for the purpose of fusion and classification of optical and SAR data in urban areas. MIMA is a semi-supervised manifold data fusion algorithm that finds a latent space where the data sources are fused. During the fusion procedure, it simultaneously considers annotated data and explores data structures using topological data analysis. According to comprehensive evaluations under multiple circumstances, MIMA outperforms the state-of-the-art manifold-based data fusion algorithms up to 2.5% in terms of classification accuracy. For the three data sets used for evaluations

in this dissertation, this improved percentage corresponds to approximately 350, 600, and 8000 data samples, respectively. Importantly, MIMA shows better performance on these cases where the amount annotated data occupy a small proportion of data for inference.

A comparative review of manifold-based algorithms for fusing hyperspectral and PolSAR data on urban classification tasks was conducted. Two important questions were discussed in this study: (1) where should the data fusion procedure be carried out, in data domain or manifold domain? (2) which supervision strategy is more efficient, supervision, semi-supervision, or un-supervision? Extensive experiments showed that the data fusion should be carried out in manifold domain, and the semi-supervision is the most efficient supervision strategy.

A classification pipeline encapsulating MIMA was developed to deal with a classification task of 42 major cities around the world. MIMA in this pipeline is a modified version which is dedicated to fuse and classify Sentinel-1 and Sentinel-2 data. To tackle the practical issue of hyperparameter tunning, this dissertation comes up with an ensemble solution. This classification pipeline outperforms other state-of-the-art large scale pipelines according to the evaluations over cities of Munich, Mumbai, and New York.

An urban pattern analysis was carried out in this dissertation. Based on the maps produced by the classification pipeline and a global population data, the urban morphological patterns of 42 major cities around the world are studied. Eight different urban patterns were identified in terms of their morphological formations and population distributions. The eight urban patterns are super compact, compact, compact-green, open, open-green, open-informal, informal, and industrial.

Keywords Heterogeneous data fusion, Human settlement, Hyperspectral data, Large scale urban classification, Local climate zone, Manifold learning, MIMA, Multispectral data, Optical data, Polarimetric SAR (PolSAR), Semi-supervised manifold learning, Sentinel-1, Sentinel-2, Synthetic aperture radar (SAR), Topological data analysis (TDA)

Zusammenfassung

Laut dem Bericht The World's Cities in 2018 der Vereinten Nationen (UN) lebten 2018 55,3% der Weltbevölkerung in urbanen Gebieten. Sie erwarten, dass diese Zahl bis 2030 auf 60% ansteigen wird. Die Ansiedlung in urbanen Gebieten ist für die Entwicklung der Menschheit so wichtig, dass die UN die nachhaltige, urbane Entwicklung als eines ihrer 17 nachhaltigen Entwicklungsziele (engl. Sustainable Developement Goals, SDGs) erklärt hat. Um dieses Ziel zu erreichen, sind strategische, geografische Informationen notwendig, die die Planung für die Zukunft unterstützen.

Mit Hilfe von Erdbeobachtungstechnologien lassen sich aktuelle, weitreichende Geoinformationen von urbanen Gebieten erstellen. Stand heute gibt es drei globale Produkte aus der Erdbeobachtung, die die urbanen Gebiete weltweit kartografieren: der Global Urban Footprint (GUF), der Global Human Settlement Layer (GHSL) und der World Settlement Footprint 2015 (WSF). Die Verfügbarkeit dieser Produkte leistet einen unschätzbaren Beitrag zur Bereitstellung wichtiger, urbaner Informationen. All diesen Produkten ist gemein, dass ihnen ein binäres Schema (urban, nicht urban) zu Grunde liegt. Dieses einfache Schema lässt sich universell einsetzen, da es sehr generell ist, aber andererseits fehlen detaillierte Informationen über die inneren Strukturen der urbanen Gebiete. Die nächste Generation der urbanen Karten sollte ein Klassifikationsschema aufweisen, das detaillierte Informationen über die innere Struktur widerspiegelt und gleichzeitig universell anwendbar ist. In Bezug auf die Datenquellen basiert der GUF auf Daten von synthetischem Aperturradar (SAR), während der GHSL aus optischen Daten berechnet wurde und der aktuelle WSF auf Daten aus beiden vorher genannten Quellen zurückgreift. Zukünftige Produkte werden wie der WSF die Vorteile beider Datenquellen vereinen. Im Hinblick auf optische Datenquellen bieten Hyperspektraldaten eine große Bandbreite an Informationen und eignen sich daher ideal für Klassifikationsfragestellungen. Bei Radardaten ist polarimetrisches SAR (PolSAR) die häufigste Wahl für Klassifikationsaufgaben. Vor diesem Hintergrund zielt diese Dissertation auf urbane Klassifikationsaufgaben ab, denen ein detaillierteres Klassifikationsschema als ein binäres Schema zu Grunde liegt. Dabei werden sowohl optische, insbesondere Hyperspektraldaten, als auch PolSAR Daten verwendet.

Bezüglich der Fusion von optischen und SAR-Daten zum Zweck der urbane Klassifikation sind einige der größten Herausforderungen, (1) die drastischen Unterschiede zwischen optischen und SAR-Daten; (2) die hohe Dimension von Fernerkundungsdaten; und (3) der Mangel an gekennzeichneten Daten. Manifoldbasierte Verfahren für die Datenfusion sind eine vielversprechende Lösung hinsichtlich der drei aufgeführten Herausforderungen. Erstens, ist das Verfahren in der Lage die Lücke zwischen Datendomänen zu überbrücken, was hilfreich ist die Unterschiede zwischen optischen und SAR-Daten zu adressieren. Zweitens, die intrinsische Eigenschaft der Manifoldtheorie niedrige Dimensionalität zu bevorzugen ist ideal um hochdimensionale Fernerkundungsdaten zu verarbeiten. Zuletzt, die teilüberwachte Manifoldangleichung assoziiert gekennzeichnete und ungekennzeichnete Datensample mittels Manifoldstrukturen, wodurch die semantische Information der Kennzeichnungen implizit zu ungekennzeichneten Daten propagiert wird. Dies verstärkt letztendlich den Einfluss der Datenkennzeichnung. Daher ist das methodologische Ziel dieser Dissertation manifoldbasierte Techniken der Datenfusion und zur Klassifikation urbaner Gebiete zu untersuchen.

Im Rahmen der Dissertation wurden die folgenden vier methodologischen Ziele erreicht.

Ein Datenfusionsalgorithmus namens MIMA wurde entwickelt, mit dem Zweck der Fusion und Klassifikation von optischen und SAR-Daten in urbanen Gebieten. MIMA ist ein semiüberwachter Datenfusionsalgorithmus der einen latenten Raum findet in dem all Datenquellen fusioniert sind. Während des Fusionierungsprozesses berücksichtigt er gekennze-

ichnete Daten und erkundet Datenstrukturen mittels topologischer Datenanalyse. Eine umfassende Analyse unter verschiedenen Bedingungen zeigt, dass MIMA eine 2.5% höhere Klassifikationsgenauigkeit erreicht als state-of-the-art manifoldbasierte Datenfusionsalgorithmen. Für die drei Datensätze, die im Rahmen der Evaluierung in dieser Dissertation verwendet wurden, entspricht dieser verbesserte Prozentsatz jeweils ungefähr 350, 600, beziehungsweise 8000 Datensamples. Insbesondere zeigt MIMA eine verbesserte Leistung in der Inferenz für jene Fälle, wo gekennzeichnete Daten nur einen geringen Anteil ausmachen.

Ein Vergleich von manifoldbasierten Algorithmen für die Fusion von hyperspektralen und PolSAR-Daten mit der Aufgabe der urbanen Klassifikation wurde durchgeführt. Zwei wichtige Fragen wurden in dieser Studie diskutiert: (1) wo soll die Datenfusion durchgeführt werden, in der Daten- oder der Manifolddomäne? (2) welche überwachungsstrategie ist effizienter, überwacht, teilüberwacht, oder unüberwacht? Umfassende Experimente zeigten, dass die Datenfusion in der Manifolddomäne ausgeführt werden sollte, und das Teilüberwachung die effizienteste überwachungsstrategie ist.

Eine Klassifikationspipeline die MIMA beinhaltet wurde entwickelt um 42 Großstädte, verteilt über die gesamte Welt, zu klassifizieren. In dieser Pipeline wird eine modifizierte Version von MIMA eingesetzt, deren Aufgabe es ist Sentinel-1 und Sentinel-2 Daten zu fusionieren und zu klassifizieren. Um das konkrete Problem des Hyperparametertunings zu lösen, schlägt diese Dissertation einen Ensembleansatz vor. Diese Klassifikationspipeline übertrifft andere state-of-the-art großflächige Pipelines laut der Evaluierung für die Städte München, Mumbai und New York.

Eine urbane Musteranalyse wurde in dieser Dissertation durchgeführt. Basierend auf den Karten, die von der Klassifikationspipeline produziert wurden und Daten der Weltbevölkerung, wurden die urbanen morphologischen Muster von 42 Großstädten auf der gesamten Welt untersucht. Acht unterschiedliche Muster wurden identifiziert bezüglich ihrer morphologischen Anordnungen und Bevölkerungsverteilung. Diese acht urbanen Muster sind superkompakt, kompakt, kompakt-grün, offen-grün, offen-informell, informell, und industriell.

List of Abbreviations

Abbreviation	Description
AA	average accuracy
CEOS	Committee on Earth Observation Satellites
CCF	canonical correlation forest
COSPACE	common subspace learning
DEM	digital elevation model
EMR	electromagnetic radiation
EMW	electromagnetic wave
ER+	Estrogen Receptor-positive
ESA	European Space Agency
FPC	foliage projective cover
GEO	the Group on Earth Observation
GGF	generalized graph fusion
GHSL	global human settlement layer
GIS	geographic information science
GLCM	gray-level co-occurrence matrix
GUF	global urban footprint
HSI	hyperspectral image
JRC	joint research center
KEMA	Kernel manifold alignment
KNN	k-nearest neighbor
KSVM	Gaussian kernel support vector machine
LCLU	land cover land use
LCZ	local climate zone
LeMA	learnable manifold alignment
LiDAR	light detection and ranging
LPP	locality preserving projection
LSVM	linear support vector machine
MA	manifold alignment
MDGs	Millennium Development Goals
MIMA	MAPPER-induced manifold alignment

Abbreviation	Description
MP	morphological profile
NDSI	Normalised Difference Snow Index
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
NBI	number of intervals in MAPPER
OA	overall accuracy
OLR	overlap rate in MAPPER
OSM	open street map
PA	producer accuracy
PC	principal component
PCA	principal component analysis
PolSAR	polarmetric synthetic aperture radar
RF	random forest
SAR	synthetic aperture radar
SCI	spinal cord injury
SDGs	Sustainable Development Goals
SSE	error sum of square
SSMA	semi-supervised manifold alignment
SVM	support vector machine
T2D	type 2 diabetes
TBI	traumatic brain injury
TDA	topological data analysis
TEEC	topological enhanced ensemble classification
UA	user accuracy
UN	Unitied Nation
nm	nanometer
dB	decibel
γ	normalized coherence

List of Symbols

Symbol	Description
i	order indicator
j	order indicator
k	data source indicator
k'	data source indicator
$\mathbf{K}$	number of data sources
κ	number of neighbors
N	number of data instances
n	number of dimensions
l	number of dimensions
$\mathbf{x}$	a data instance in vector
$\mathbf{x}_i^k$	the i^{th} data instance of k^{th} data source
$\mathbf{x}_j^k$	the j^{th} data instance of k^{th} data source
$\mathbf{X}$	a data set $\mathbf{X} = \{\mathbf{x} \in \mathbf{R}^n \mid \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$
$\mathbf{X}_s^k$	subset of $\mathbf{X}^k$
$\mathbf{Z}$	joint matrix of multiple data
$\mathbf{B}^k$	annotated data
$\mathbf{y}$	a low dimensional representation of $\mathbf{x}$
$\mathbf{y}_i^k$	a low dimensional representation of $\mathbf{x}_i^k$
$\mathbf{y}_j^k$	a low dimensional representation of $\mathbf{x}_j^k$
$\mathbf{Y}$	a low dimensional representation of $\mathbf{X}$
$\mathcal{M}$	manifold
$\mathbf{A}$	adjacent matrix
$\mathbf{A}_{ij}$	element of i^{th} row and j^{th} column of $\mathbf{A}$
$\mathbf{A}(i,j)$	element of i^{th} row and j^{th} column of $\mathbf{A}$
$\mathbf{A}_t$	adjacent matrix
$\mathbf{A}_m$	adjacent matrix derived by MAPPER
$\mathbf{A}_s$	similarity matrix
$\mathbf{A}_d$	dissimlarity matrix

Symbol	Description
$\mathbf{A}_s^{k,k'}$	similarity matrix of k^{th} and k'^{th} data
$\mathbf{A}_d^{k,k'}$	dissimlarity matrix of k^{th} and k'^{th} data
$\tilde{\mathbf{A}}_t$	joint adjacent matrix of multiple data
$\tilde{\mathbf{A}}_m$	joint adjacent matrix derived by MAPPER
$\tilde{\mathbf{A}}_s$	joint similarity matrix of multiple data
$\tilde{\mathbf{A}}_d$	joint dissimlarity matrix of multiple data
$\mathbf{W}$	weight matrix
$\mathbf{W}_{ij}$	element of i^{th} row and j^{th} column of $\mathbf{W}$
$\mathbf{D}$	degree matrix
$\mathbf{L}$	Laplacian matrix
$\mathcal{L}$	loss function
$\mathbf{R}$	real number
$\mathbf{R}^n$	n dimensional real space
$\mathbf{R}^l$	l dimensional real space
$\mathbf{f}$	a projection $\mathbf{y} = \mathbf{f}\mathbf{x}$
$\mathbf{f}^k$	a projection $\mathbf{y}^k = \mathbf{f}^k\mathbf{x}^k$
$\mathbf{S}$	scattering matrix
H	horizontal polarization
V	vertical polarization
HV	signal received in H, and emitted in V
S_{HH}	complex signal of HH SAR data
S_{HV}	complex signal of HV SAR data
S_{VH}	complex signal of VH SAR data
S_{VV}	complex signal of VV SAR data
μ	weighting parameter in SSMA
$\mathcal{A}$	loss term of similarity in SSMA
$\mathcal{B}$	loss term of dissimilarity in SSMA
$\mathcal{C}$	loss term of topology in SSMA

Symbol	Description
f	filter function in MAPPER
f^{-1}	inverse function of f
p	a continuous and surjective projection
Z	a projected continues space of data $\mathbf{X}$
z	a data point of $\mathcal{Z}$
U	a covering of Z
α_i	the i^{th} vertex in a k-simplex
$\mathbf{C}_i$	a vector presenting features of i^{th} city
$\tilde{\mathbf{C}}$	cluster center of a urban pattern
$\|\cdot\|^2$	L_2 norm
$\cdot^*$	complex conjugate
$<\cdot>$	expectation
ϵ	a neighborhood radius
σ	kernel width
λ	eigenvalue
$\sum$	sum

Contents

1 Introduction

1.1 Motivations and Objectives

Sustainable development is a strategic guidance for the development of human kind. The United Nations (UN) has been working on this subject for about three decades. It was started with Agenda 21 in June 1992. In 2000, the UN announced Millennium Development Goals (MDGs) that is a schedule of sustainable development ought to be achieved in 2015. In the year of 2015, the UN started a new round of 15-years schedule, Sustainable Development Goals (SDGs). It is a major challenge to implement the SDGs globally, timely, and evaluate them objectively. Remote sensing techniques step on stage at this point and exhibit an unbiased, affordable, and timely global solution of providing relevant geographic information.

As reported in UN's report of world's city (UN, 2018), 55.3% of the world's population lived in urban area in 2018, which is estimated to be 60% in 2030. Therefore, urban sustainable development has an impact on daily lives of more than half of the Earth inhabitants and their descendants. The tremendous influence makes urban sustainable development itself a major topic in UN's SDGs. This topic has also drawn a lot of attention of the remote sensing community.

To provide relevant geographic information for urban sustainable development, the remote sensing community started to provide global urban maps in early 1990s (Klotz et al., 2016). The state-of-the-art global urban maps are Global Urban Footprint (GUF) (Esch et al., 2013) and Global Human Settlement Layer (GHSL) (Pesaresi et al., 2013). GUF is a 12 meter resolution binary map of built-ups which is produced from synthetic aperture radar (SAR) data. GHSL is a ten meter resolution world map of the density of built-ups that was derived from optical remote sensing data (Melchiorri et al., 2019). Both GUF and GHSL provided invaluable urban map that benefits urban anaylsis on a global scale. For example, Taubenböck et al. (2019) analyzed global urban development by using GUF data of about 1700 cities. Melchiorri et al. (2018) analyized global urbanization using GHSL for last decades. Motivated by the significant role of urban areas in future and demands of strategic urban geoinformation, this dissertation mainly focuses on studies of urban area.

GUF and GHSL have demonstrated the great potential on providing strategic information, but both products categorize the Earth surface as built-up and non-built-up. The categorization is not fine enough to provide information within a city. Therefore, the next generation global urban layer should have a finer categorization of urban regions. This dissertation works towards this direction and mainly focuses on land cover land use classification schemes within cities. Those schemes describe morphological structures and functionality types of urban blocks.

From the perspective of data sources, GUF utilized SAR data, and GHSL was produced based on optical data. Studies (Zhang, 2010; Joshi et al., 2016; Schmitt and Zhu, 2016; Hu et al., 2019) have highlighted that the complementarity of heterogeneous data sources motivates a fusion of them for producing future remote sensing products. As a type of optical data, hyperspectral data has been proven as an ideal optical data source for classification tasks (Gu et al., 2017; Ghamisi et al., 2017). Thanks to recently missions, such as DESIS and EnMAP, accessing hyperspectral data is becoming convenient. For the choice of SAR data, polarimetric SAR data was often selected for classification tasks in literature (Lee and Pottier, 2009; Wurm et al., 2017). This dissertation mostly focuses on the fusion of hyperspectral data and PolSAR data for urban classification tasks.

To summarize a literature review on fusing optical and SAR data for classification tasks, which is discussed in Chapter 3, the current state-of-the-art practice extracts features from both data sources, fuses features, feeds them to classifiers, and delivers classification results. In terms of methodology, statistical models often very powerful, but might fail to describe complex remote sensing data; kernel models and ensemble models are current state-of-the-art models; deep learning models have not been fully explored due to limited amount of annotated data. To the author's knowledge, manifold models had never been explored for fusing optical and SAR data before this dissertation, although it has been proven excellent for data fusion in general (Wang and Mahadevan, 2011; Tuia et al., 2014). Moreover, semi-supervised manifold models do not require a large amount of annotated data as deep learning models, and amplifies the impact of annotated data. Such properties suit the background of this dissertation well.

Motivated by above-mentioned facts, this dissertation aims to explore manifold model for the fusion of optical and SAR data and apply the fusion on urban region classification tasks. Four important objectives of this dissertation are listed as follows:

- Development of fusion algorithm: this dissertation shall develop a data fusion algorithm that is able to accomplish fusion and classification of multi/hyper-spectral and PolSAR data for urban area. Due to limited access to these data and specially labeled data, an algorithm of unsupervision or weak supervision is preferred.
- Evaluation of the developed algorithm: this dissertation shall carry out a comprehensive evaluation. Besides showing a state-of-the-art performance of the developed algorithm, the evaluation shall be also enable a profound discussion regarding a wide cover of related fusion algorithms.
- Deployment of the developed algorithm: this dissertation shall develop a functional classification system that is based on developed fusion algorithm. This classification system should be deployed on a large scale classification task of urban areas.
- Demonstration of a large scale application: this dissertation shall demonstrate that urban classification maps produced by the deployed system are able to supply strategic geoinformation to interested stakeholders.

1.2 Dissertation Outline

This is a publication based dissertation whose work is summarized in five peer-reviewed journal papers (Four as first author, and one as second author). Three of them are published and two are under reviewing process. The papers are attached in the appendix of this dissertation. Chapter 2 summarizes basic knowledge for works of this dissertation. Chapter 3 provides an overview of the state of the art. Chapter 4 summarizes works have been done within this dissertation. And chapter 5 concludes this PhD project and provides opinions on future works.

2 Basics

Targeting at readers who are already in the research field of remote sensing, necessary basics are included in this chapter.

2.1 Remote Sensing Data

Remote sensing is a science and technology that measures properties of an object of interest without a physical contact (Khorram, 2012; Pohl and Van Genderen, 2017). It generally refers to the technology of observing the Earth by sensors mounted on airborne or spaceborne vehicles. These sensors record electromagnetic signals that are reflected or emitted by the object of interest on the Earth. With the recorded data, scientists are able to retrieve geophysical information without an on-site investigation.

2.1.1 Optical data

Around eight minutes and twenty seconds, electromagnetic radiations (EMR) emitted by the Sun reach the Earth. The EMR travels through the Earth atmosphere, interacts with the Earth surface, and travels back to outer space through the atmosphere again. During the travelling of the EMR, it is either absorbed or scattered. On its way back to outer space, an optical sensor measures the radiance that reaches its lens and records the radiance as a digital number (Alparone et al., 2015). The digital number is converted to top-of-atmosphere radiance, ground-level radiance, and surface reflectance by radiometric calibration, atmospheric correction, and surface correction in a sequential processing (Bioucas-Dias et al., 2013). The surface reflectance is the ratio of the reflected to the incident EMR. Its value relates to physical properties of the illuminated object. Optical sensors work on the electromagnetic waves whose wavelength range from 400 nm to 3000 nm, including visible light, near infrared, and short-wavelength infrared. Within this wavelength range, radiance recorded by optical sensors is mainly EMR reflected by the Earth surface instead of a thermal emission (Alparone et al., 2015).

Two important concepts in optical data are spatial resolution and spectral resolution. The spatial resolution is expressed as a distance. Objects in image can be distinguished only if they are located further than the distance (Alparone et al., 2015). The spectral resolution is the width of the wavelength range within which the photons are collected to form a single spectral band, such as red, or green. According to (Alparone et al., 2015; Grohnfeldt, 2017), an optical data could be categorized as panchromatic, multispectral, superspectral, hyperspectral, and ultraspectral imagery, in terms of different spectral resolutions.

- ◇ panchromatic, a unique broadband typically ranges from 400 nm to 1000 nm.
- ◇ multispectral, the number of bands is smaller than 10, including visible bands.
- ◇ superspectral, the number of bands is larger than 10.
- ◇ hyperspectral, the spectral resolution is higher than 10 nm.
- ◇ ultraspectral, the spectral resolution is at the magnitude of 1 nm.

Due to unique physical property and status, a material reflects EMR differently on each of these spectral bands, which results a spectral signature that can help identifying this material. Therefore, the higher the spectral resolution, the more complete the recorded spectral signature. However, due to limited photon, it is a technically challenge to achieve both high spatial and spectral resolutions.

2.1.2 SAR data

SAR sensor emits a frequency modulated pulse, a 'chirp', towards a target, and records the back scattered signals. After a series of digital processing, SAR data is presented as a complex-valued range-azimuth image (Cumming and Wong, 2005). Each complex number describes the amplitude and phase of a received echo that is scattered by the illuminated targets. The amplitude and phase of a SAR image reflect the geometric shape and surface roughness of the illuminated targets.

Polarimetric SAR (PolSAR) is a SAR technique that utilizes the polarization state of emitted signals. A PolSAR sensor is able to emit horizontally or vertically polarized EMW, and measures either the horizontal or the vertical components of the received signal. A 2 by 2 scattering matrix (2.1) describes PolSAR data, where H and V indicate horizontal and vertical polarization, respectively; and in S_{ij}, i indicates the polarization of the received signal and j indicates it of the emitted signal.

$$\mathbf{S} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \tag{2.1}$$

Being able to emit horizontally and vertically polarized signals enables observing the targets' reaction to signals of different polarization. Measuring the horizontal and the vertical components of reflected signals fully recovers the polarization states of a reflected signals. Therefore, comparing to single polarized SAR data, PolSAR data enables analyzing the response of observed objects in terms of polarization. More specifically, via polarimetric decomposition (Cloude and Pottier, 1997; Yamaguchi et al., 2006; Lee and Pottier, 2009), one can look into the scattering mechanisms, such as surface scattering, volume scattering, double bounce, and helix scattering. By comparing to a single polarized SAR data, PolSAR data is more capable of extracting geo/bio-physical parameters for forest, agricultural, and urban applications (Neumann et al., 2009; Hajnsek et al., 2009; Moreira et al., 2013; Xiang et al., 2015).

2.1.3 Imaging geometry of optical and SAR data

The different imaging principles of SAR and optical sensors result in the most important difference between an optical image and a SAR image, namely the geometry. Taking the lens of an optical sensor as an originate, objects locating within minimum distinguishable azimuth and elevation angles will be imaged into one resolutions cell. It is convenient to interpret an optical image since people observe this kind of images in daily life. On the other hand, a resolution cell of a SAR image records signals that reflected by objects whose distances to the sensor are the same. As shown in the right plot of Fig. 2.1, due to the side-looking imaging geometry of a SAR sensor, objects that have the same distances to the sensor would seldom locate close by. They often have a difference on elevation. It makes the interpretation of a SAR image challenging. As the example in Fig. 2.1, the section D in the optical image represents the ground; however, the corresponding section in the SAR image includes signals reflected by the ground, the facade of the lower building, and the roof of the lower building. All these contributors of the section D of the SAR image have the same distances to the sensor, but are not located close by.

Due to different imaging principles, contents of an optical image and a SAR image are very different from each other. A visual example is shown in Fig. 2.1. Thanks to SAR imaging principles, the difference would be larger if the resolution getting higher or variations of the height in the observed region getting larger.

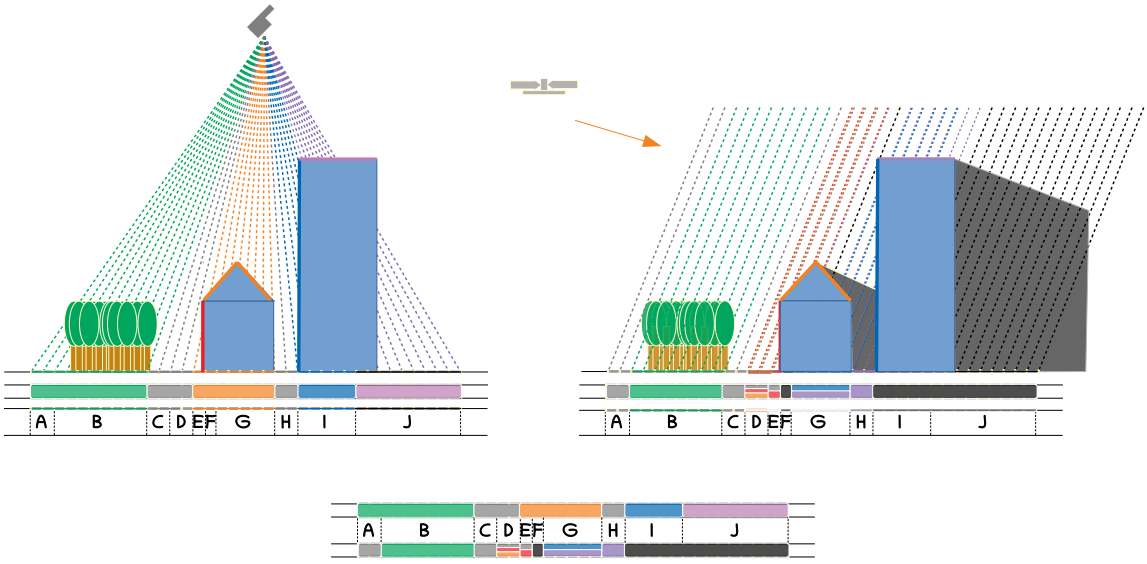


Fig. 2.1. A sketch of the imaging geometry of an optical image (left) and a SAR image (right) [Inspired by (Palubinskas, 2012)]. A resolution cell in the sketch is bounded by a pair of dashed line. The sketch supposes the optical image and the SAR image have the same size of resolution cell. The bottom sketch compares imaged contents of the optical image and the SAR image in every section: A: trees in the optical image, ground in the SAR image; B: trees in both image; C: ground in both image; D: ground in the optical image, ground, facade of the lower building, and roof of the lower building in the SAR image; E: roof of the lower building in the optical image, ground and facade of the lower building in the SAR image; F: roof of the lower building in the optical image, shadowing in the SAR image; G: roof of the lower building in the optical image, facade and roof of the higher building in the SAR image; H: ground in the optical image, roof of the higher building in the SAR image; I: facade of the higher building in the optical image, shadowing in the SAR image; J: roof of the higher building in the optical image, shadowing in the SAR image.

Fig. 2.2. A visualization of an optical image and a SAR image of the same scene, the city of Augsburg, Germany. The optical image is a Sentinel-2 image with red-green-blue channels. The SAR image is a Sentinel-1 intensity image in dB.

2.1.4 Complementarity of optical and SAR data

Two aspects, illumination source of the EMW and wavelength range of the EMW, are the reasons for the complementarity of optical and SAR data.

The illumination source of an optical data is the Sun (passive remote sensing). Therefore, weather condition has a significant impact on collecting optical data. As a result, it poses limitations for certain regions, such as tropic areas frequently covered by cloud, or certain applications, such as rapid disaster response. On the other hand, a SAR sensor has its own illumination source (active remote sensing) so that it can collect data day and night. Thanks

to the different illumination sources, a part of the complementarity is that SAR data could be collected at some time or regions that an optical data can not be acquired.

Regarding the wavelength range of applied EMW, an optical sensor operates on the wavelength ranging from 400 nm to 3000 nm. However, SAR sensors normally operate on the wavelength range of centimeters to meters. Another part of the complementarity is that EMW of different wavelength ranges reveal different properties of a target. Moreover, since the microwave is capable of penetrating a cloud, SAR images can still be acquired over cloudy areas.

2.2 Manifold Learning for Data Fusion

For the term of manifold learning in this dissertation, it means a group of machine learning techniques that assume an underlying manifold laying in a sampled data, attempt to retrieve the manifold, and utilize the manifold for a certain goal.

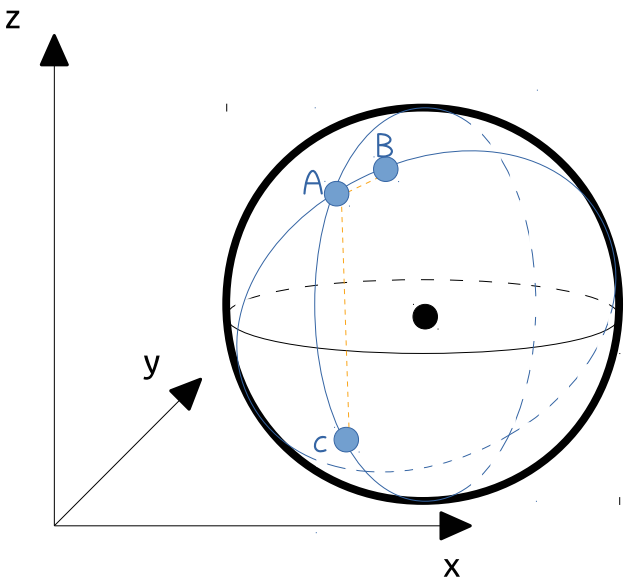


Fig. 2.3. Points A, B, and C locate on the surface of the sphere which is a 2 dimensional manifold embedded in a 3 dimensional space. The black point is the center of the sphere, and the black ellipse is the equator of the sphere. Solid curves lay on the surface facing the reader, and dashed curves lay on the surface of the other side. The orange dashed edges are inside the sphere. If two data points are local, such as A and B, a Euclidean distance (orange dashed edge) can be used to approximate the distance on manifold (solid blue curve connecting A and B). If two data points are not local, such as A and C, a Euclidean distance is not a valid approximation for the distance on manifold (solid blue curve connecting A and C).

2.2.1 Topological interpretation of manifold

A manifold is a smooth hyperplane. It is embedded in a Euclidean space which has a higher dimension than the manifold. As a simple example in Fig. 2.3, the surface of a sphere is a 2-dimensional manifold that is embedded in a 3-dimensional Euclidean space. Apart from the intuitive example for explanation, manifolds are often too complex to be described by an analytic formulation. It is not mathematically convenient to represent a manifold.

From topological point of view, a local region on a manifold and a hyperplane having the same dimension as the manifold are homeomorphic to each other (Edelsbrunner, 2014, p.58), (Ghrist, 2014, p.10). An intuitive interpretation is that a local region on a manifold

is a hyperplane that has a same dimension as the manifold (Edelsbrunner and Harer, 2010, p.32). With the sphere example, it means that an ant standing on a point of a sphere would see its surrounding as a flat ground. Therefore, it brings a mathematical convenience that the local property of a manifold can be approximated by local measures in a Euclidean space. This local property provides an access of approximating a manifold, as explained in Fig. 2.3.

2.2.2 Manifold dimension reduction

Given a data set $\mathbf{X} = \{\mathbf{x} \in \mathbf{R}^n | \ \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ that is measured in an n-dimensional space, dimension reduction is a task to find a representation $\mathbf{Y} = \{\mathbf{y} \in \mathbf{R}^l | \ \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N\}$ of $\mathbf{X}$ where $l << n$ so that the amount of data is reduced while the essential information should be preserved. The most classical technique is principal component analysis (PCA). PCA linearly transfers the data $\mathbf{X}$ into an orthogonal Euclidean space where axis sequentially account largest variance of the data. With a predefined percentage of variance to be preserved, PCA accomplishes dimension reduction by ignoring the dimensions with lowest variances.

Manifold learning technique, however, has a more intuitive mathematical theory than PCA for dimension reduction. Because of the mathematical definition that a manifold $\mathcal{M}$ is a lower dimensional space that embedded in a higher dimension space, finding the underlying manifold is naturally a procedure of dimension reduction. Because real world data is often measured by an over-parameterized system, the assumption of existing an underlying manifold often holds (Bengio et al., 2013). The manifold-based dimension reduction has drawn a lot of attention and been intensively studied.

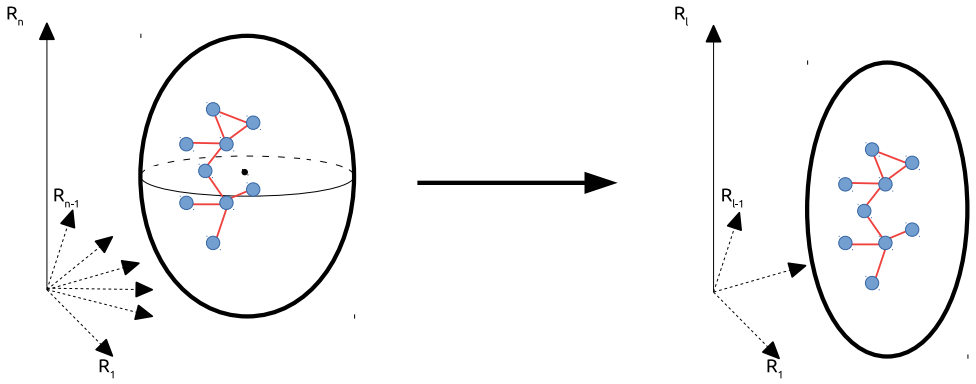


Fig. 2.4. On the left plot, blue points are data points lay on an l dimensional manifold (bold ellipse) that is embedded in an n dimensional space. The red lines represent local adjacent properties of data points. Manifold dimension reduction projects those data points into an l dimensional space ($l << n$), while the local adjacent properties are all preserved. Different models focus on different local properties to be preserved.

Highlighted as the red edges in Fig. 2.4, the key point in this technique is how to approximate the underlying manifold via the local structures of the original space. Among existing studies, Roweis and Saul (2000) represented a data point as a linear combination of its neighbors, and find a lower dimensional manifold where the data point is the same linear combination of the same neighbors. The local linear relationship stayed unchanged while the number of dimensions decreased. Tenenbaum et al. (2000) estimated manifold distances of every data pair. If the two data points are neighbors, their manifold distance is estimated by the Euclidean distance. If the two data points are not neighbors, their manifold distance is estimated by adding the Euclidean distances of edges that form a shortest path between them. Therefore, a global structure of the manifold can be measured by local Euclidean distances, and the global structure is preserved while reducing the dimension. He and Niyogi (2004) designed a dimension reduction technique locality preserving projection (LPP) that only preserves local connectivity and ignored non-local structure of the original data. Many

other similar studies accomplish manifold-based dimension reduction with different strategies of modelling local structures (Belkin and Niyogi, 2002; He et al., 2005; Lin and Zha, 2008).

The LPP introduced by He and Niyogi (2004) is a representative manifold algorithm for dimension reduction. The technical detail is briefly introduced in this section. LLP estimates local connectivity of data points in its manifold $\mathcal{M} \in \mathbf{R}^l$ by calculating local connectivity in its original space $\mathbf{X} \in \mathbf{R}^n$. This connectivity property is preserved while decreasing the number of dimensions. He and Niyogi (2004) defined the connectivity of two data points $\mathbf{x}_i$ and $\mathbf{x}_j$ by either $\|\mathbf{x}_i - \mathbf{x}_j\|^2 < \epsilon$ or $\mathbf{x}_i$ is among the κ nearest neighbors of $\mathbf{x}_j$. Both ϵ and k are predefined. The local connectivity property of $\mathbf{X}$ can be represented by an N by N adjacent matrix $\mathbf{A}$, where $\mathbf{A}_{ij} = 1$ if $\mathbf{x}_i$ and $\mathbf{x}_j$ are connected, and $\mathbf{A}_{ij} = 0$ elsewhere. One can further model the strength of the connection between $\mathbf{x}_i$ and $\mathbf{x}_j$ with a weight as in matrix $\mathbf{W}$

$$\mathbf{W}_{ij} = \mathbf{A}_{ij} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}} \tag{2.2}$$

Optionally, the weight matrix $\mathbf{W}$ can be identical to the adjacent matrix $\mathbf{A}$. Therefore, with the matrix $\mathbf{A}$ or $\mathbf{W}$, the local connectivity of the underlying manifold can be modeled.

To achieve data representation on the manifold, a projection $\mathbf{f}$ should be found to link $\mathbf{X}$ and $\mathbf{Y}$ by $\mathbf{y}_i = \mathbf{f}\mathbf{x}_i$. To preserve the local connectivity in the underlying manifold, a loss function is designed as Eq. 2.3.

$$\mathcal{L} = \sum_{ij} (\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{W}_{ij} = \sum_{ij} (\mathbf{f}\mathbf{x}_i - \mathbf{f}\mathbf{x}_j)^T \mathbf{W}_{ij} \tag{2.3}$$

Minimizing the loss function encourages data points that are neighbors in the original space staying as neighbors in the pursued manifold. Having been proven in (He and Niyogi, 2004), the solution of minimizing $\mathcal{L}$ is given by eigenvectors corresponding to smallest eigenvalues achieved from the generalized eigenvalue problem (2.4).

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{f} = \lambda \mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{f}, \tag{2.4}$$

where $\mathbf{D}$ is the degree matrix that $\mathbf{D}_{ii} = \sum_{j=1}^{N} \mathbf{W}_{ij}$ and $D_{ij} = 0 (i \neq j)$, and $\mathbf{L}$ is the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$. Therefore, the optimized projection $\mathbf{f}$ is capable of projecting data $\mathbf{X}$ into its lower dimensional manifold $\mathcal{M}$ so that the dimension reduction is accomplished.

Manifold dimension reduction can also be applied for remote sensing data fusion. Liao et al. (2014) stacked hyperspectral data and LiDAR data resulting a high dimension data cube, applied a manifold-based dimension reduction technique similar to LPP, and accomplished dimension reduction and data fusion at the same time.

2.2.3 Manifold alignment

Manifold alignment attempts to match two manifolds so that the information or knowledge crossing domains can be retrieved or combined. It has been successfully applied to accomplishing challenging tasks, such as querying images that are semantically identical but very different in contents (images of different illumination) (Pei et al., 2012, 2013), finding similar documents of different languages (Wang and Mahadevan, 2008), and so on.

To achieve the alignment of manifolds, the correspondences cross domains have to been known. The colors in Fig. 2.5 represent the correspondences. With or without prior information of the correspondence separates the manifold alignment techniques into **supervised manifold alignment** or **unsupervised manifold alignment**.

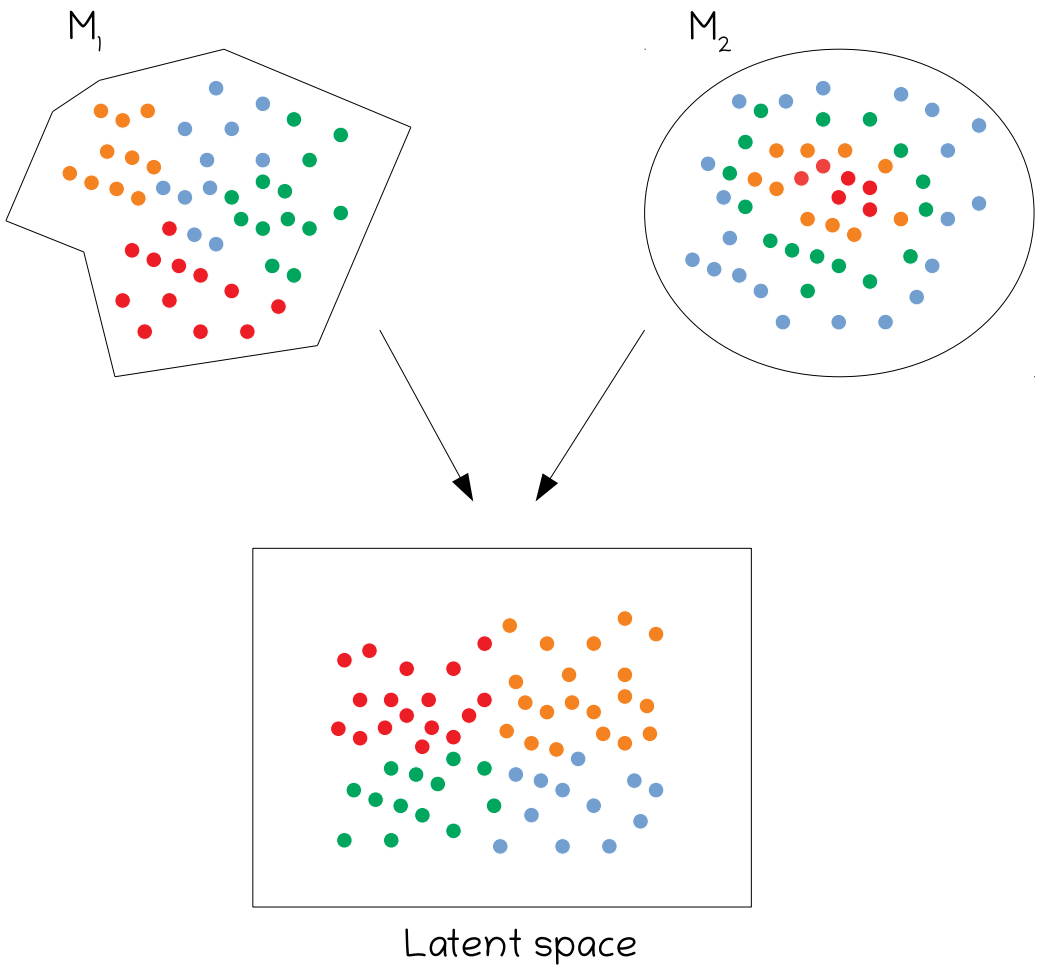


Fig. 2.5. This figure demonstrates the principle of the manifold alignment with an example of two data sets. $\mathcal{M}_1$ and $\mathcal{M}_2$ represent the manifolds of two data sets. The colors stand for the correspondences of data points cross manifolds. Both data sets are projected into a latent space where data correspondences are preserved. Therefore, the two manifolds are aligned.

Intuitively, for **unsupervised manifold alignment**, the main task is to find the correspondence. Most existing studies focus on developing similarity measures of data samples crossing domains. Wang and Mahadevan (2009b) introduced a distance that measures the similarity of two local structures. Therefore, they were able to find local correspondences between domains and accomplish manifold alignment. Pei et al. (2012) introduced a similarity measure of two manifold which is able to find the correspondence even with the presence of small local distortions. Escolano et al. (2011) applied a symmetrized normalized-entropy-square variation as a dissimilarity measure of different manifolds with which they accomplished an alignment. Pei et al. (2013) introduced a manifold-to-manifold distance so that a query pattern could be compared to reference patterns. Wang and Mahadevan (2013) measure two samples of different domains by utilizing a sample of correspondence as a middle point of the edge linking the two samples. With the correspondence-based measure, it accomplished the alignment between manifolds.

Due to extra effort required to provide a prior correspondence cross domains, there is a only a handful of studies of **supervised manifold alignment**. Wang and Mahadevan (2008) accomplished manifold alignment by first applying manifold-based dimension reduction on multiple data sources, and then aligned the multiple lower dimensional data via rotating and shifting the subsets which have correspondence. Tan et al. (2014) applied manifold alignment to match the networks on different social media platforms, such as Facebook and Twitter. The correspondence between networks is built by matching usernames.

Besides image query and cross-lingual retrieval, various applications have been accomplished by manifold alignment techniques. Majeed et al. (2015) utilized a manifold alignment technique for indoor localization. Schneider et al. (2014) accomplished appearance estimation with manifold alignment. In remote sensing, Yang and Crawford (2015) applied manifold alignment for multitemporal hyperspectral image classification. Liao et al. (2016) realized visualizing hyperspectral data as a natural image via manifold alignment.

According to existing literature, manifold alignment is an intuitive solution for data fusion problems. One can assume that there exist an underlying manifold in each of the data sources to be fused. Aligning these manifolds is a natural procedure of data fusion. Additionally, the low dimension essence of the manifold is able to tackle *'curse of dimension'* (Donoho et al., 2000) which is often problematic for real world data.

2.3 Topological Data Analysis (TDA)

Topological data analysis (TDA) is a mathematical tool of applied topology. It attempts to derive relevant information that is associated to the shape of the data.

One TDA working pipeline aims to detect prominent structures hidden in the data shape by a concept of persistent homology. Its general procedure follows three steps (Ghrist, 2008): (1) organizing a data set as a point cloud with a pairwise similarity measurement; (2) representing the topological space of the point cloud with simplicial complex; and (3) deriving prominent topological structures via the persistence. For details, please refer to (Edelsbrunner et al., 2000; Zomorodian and Carlsson, 2005; Edelsbrunner and Harer, 2010; Edelsbrunner, 2014). This pipeline ranks the prominence of topological structures of data by using the persistence. The most prominent structures in data often carry relevant information. Ghrist (2008) applied this pipeline on a natural image data set and found three closed circular structures within in the data. They found out the circles indicate patterns of image intensity change, line orientations, and strip structures. Erden and Cetin (2017) utilized the persistent homology to detect periodic signals in a respiratory rate measurement.

Another TDA pipeline is MAPPER. It is designed to derive a guided topological structure of a data set (Singh et al., 2007). The technical details of MAPPER is introduced in section 4.2.2. This section discusses its applications in other research fields. MAPPER has been applied to reveal valuable knowledge in multiple studies by interpreting the shape of data sets. Nicolau et al. (2011) applied MAPPER on analyzing breast cancer transcriptional data and found a sub-group of Estrogen Receptor-positive (ER+) breast cancers. 100% survival and no metastasis have been found among patients suffering this sub-type of cancer. Nielson et al. (2015) analyzed the data of preclinical traumatic brain injury (TBI) and spinal cord injury (SCI) by using MAPPER. They found a previously unknown pattern of co-occurring TBI and SCI, as well as an unknown harmful effect of a drug treatment. Li et al. (2015) analyzed medical records of patients suffering from type 2 diabetes (T2D) using MAPPER, and identified a sub-group of T2D that is unknown to doctors. Among these studies, MAPPER helped discovering invaluable knowledge which helps future practice.

Topological data analysis is a branch of data analysis focusing on studying relevant knowledge from the topological structure of data. It is an art of simplification for ignoring massive information in data, but focusing on the prominent topological structure and the guided topological representation.

3 State-of-the-art in Optical and SAR Data Fusion

This chapter provides an overview of current research status for the topic of fusing optical and SAR data from five aspects: (1) fusion terminology, (2) fusion visions, (3) fusion levels, (4) fusion techniques, and (5) fusion applications. And at the end of this chapter, the author provides a summary.

3.1 Fusion Terminology

The author found that terms, such as multi-sources, multi-modal, multi-sensory, and so on, often appear in the titles of data fusion articles. These terms suits various fusion topics, such as the fusion of panchromatic data and multispectral data, the fusion of hyperspectral data and LiDAR data, or the fusion of optical data and SAR data. However, since each of these fusion topics poses different challenge and fuses different data contents, arbitrarily using those terms with general correct sense would not help constructing a systematic understanding of data fusion in remote sensing. With a hierarchical structure of remote sensing data sources, this dissertation attempts to use these terms in terms of data contents to be fused and sensor configurations.

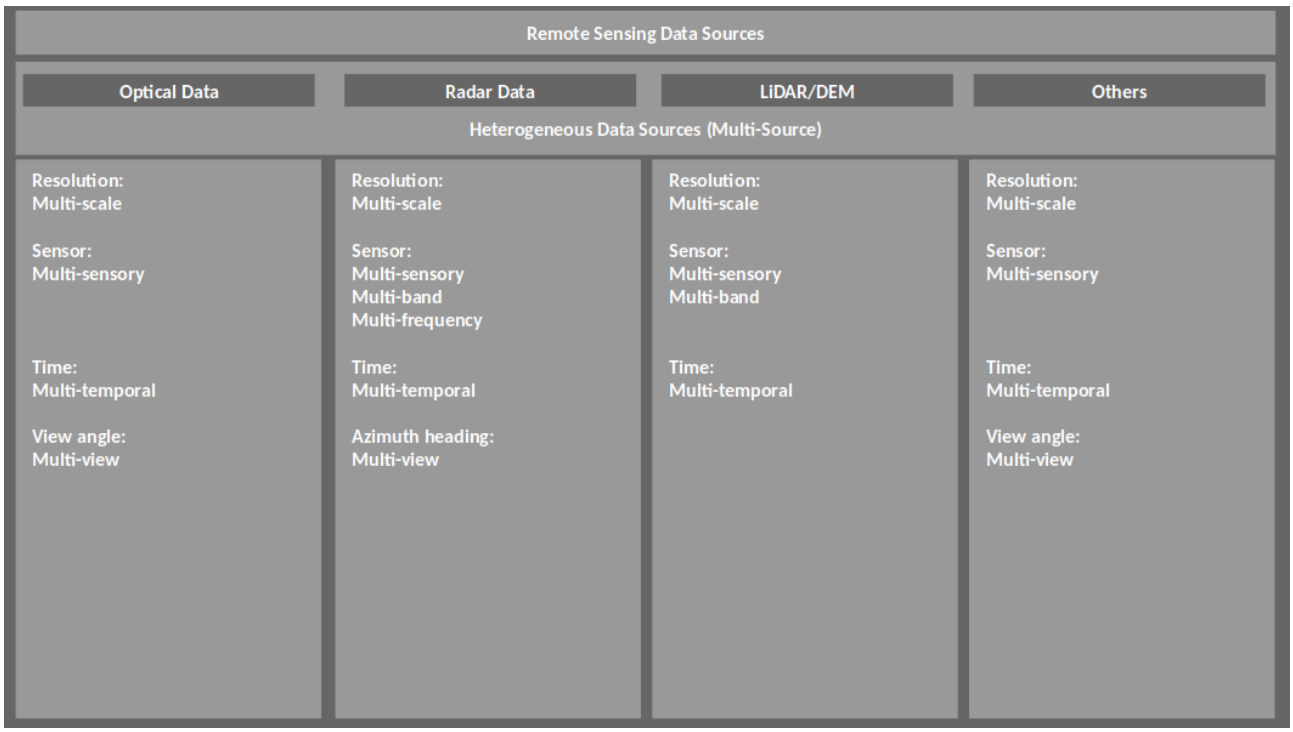


Fig. 3.1. A hierarchical structure of remote sensing data sources.

As shown in Fig. 3.1, this dissertation first divides remote sensing data sources into optical data, SAR data, LiDAR data, and others. This level is categorized by principles of data acquisition and data contents. For example, optical data provide spectrum information; SAR data provide information of shape, roughness, and dialectic properties; and LiDAR data provide height information. This dissertation treat them as heterogeneous data sources. For SAR data, an extra fold of the heterogeneity is that its imaging geometry result in a huge differences of data representation comparing to optical data and LiDAR data. In this dissertation, the fusion of heterogeneous data sources is named as multi-source data fusion or heterogeneous data fusion. Within each of these data sources, they differ from each other

by resolutions, sensors, acquisition times, and view angles. Respectively, this dissertation would name them as data fusion of multi-scale, multi-sensory, multi-temporal, and multi-view.

From this perspective, this dissertation focuses on fusing optical data and SAR data, which falls in the range of a multi-source data fusion problem.

3.2 Fusion Visions

This section starts reviewing the state-of-the-art by summarizing strategic or high level opinions on the research topic of fusing optical and SAR data that are provided in related review articles. Those options cover various aspects including motivations, current status, methodology, and future directions.

Regarding the motivation of fusing optical and SAR data, multiple studies (Gómez-Chova et al., 2015; Mura et al., 2015; Schmitt and Zhu, 2016) highlighted the complementarity of optical and radar data sets. Furthermore, fusing the two data sources is capable of solving ambiguities and conflicts in applications. And spatial and spectral information from both data sources could cooperate with each other.

To summarize existing developments, Zhang (2010) had reviewed the status of multi-source remote sensing data fusion in the year of 2010, and claimed that studies had mainly focused on new feature extraction, knowledge representation and classifier combination. A recent review article (Joshi et al., 2016) summarized about 100 articles and pointed out some issues. For example, studies typically focused on small regions, universal strategy of adequate evaluation was missing, and concrete justification of methodologies was often lacking.

Some other researchers paid special attentions to the processing of data sources. Palubinskas (2012) focused on understanding of the data sources for fusion and listed a number of related details that might have potential influence, such as speckle reduction of SAR data, imaging geometry of both data sets, high dimensionality, and so on. Similarly, Lahat et al. (2015) highlighted that successful data fusion is based on concrete understanding of individual data source. Oloyede and Hancke (2016) had summarized fusion strategies from biometric field. Apart from the different input data, general methodology is very similar to the ones in remote sensing. It reminds us that more attention might be required on how characteristics of remote sensing data should be handled while applying general fusion models.

On the side of methodology, Gómez-Chova et al. (2015) pointed out that manifold alignment could be applied on data fusion. Salentinig and Gamba (2015) preferred maximum likelihood classifier and neural network. And Schmitt and Zhu (2016) emphasised the estimation theory and machine learning. Baltrušaitis et al. (2019) reviewed the data fusion by deep learning from a general perspective, which has a great value of advisory for remote sensing applications on methodology. Both Zhu et al. (2017) and Reichstein et al. (2019) emphasised that deep learning is a promising methodology for remote sensing applications and argued that it is important for remote sensing experts to integrate physical knowledge of this field into the design of deep learning structure.

Looking into the future as a remote sensing expert, Zhu et al. (2019) recently pointed out the data fusion should not be limited in traditional remote sensing applications, such as land cover land use classification, but aim at future application, such as urban 3D structure reconstruction.

To summary, fusing optical and SAR data is motivated by complementarity of being heterogeneous data; current issues are lacking of generalization and understanding of data

sources; manifold alignment, machine learning, and deep learning are promising methodologies; and deep learning coping with field knowledge will be developed for complex applications in the future.

3.3 Fusion Level

Fusion level considers where the data fusion technically takes place. Its categories might vary slightly from one researcher to another. Generally, it is a consensus that fusion level can be divided into: *data level fusion*, *feature level fusion*, and *decision level fusion*, as shown in Fig. 3.2. This categorization is not only found in remote sensing field (Zhang, 2010; Salentinig and Gamba, 2015; Schmitt and Zhu, 2016), but also in biometric (Oloyede and Hancke, 2016) and machine learning (Lahat et al., 2015; Baltrušaitis et al., 2019).

3.3.1 Data level fusion

Data level fusion joins physical measures of different sensors.

Li et al. (2017b) fused normalized difference vegetation index (NDVI) of LandSat and intensity (dB) of ENVISAT ASAR data. Haas and Ban (2017) fused measures of spectral bands from Sentinel-2 data and intensity of Sentinel-1 data. Carreiras et al. (2017) fused surface reflectance of LandSat-5 and gamma nought (γ^o) of ALOS PALSAR. Baumann et al. (2018) fused measures of spectral bands from LandSat data and σ_0 value of Sentinel-1 data. And Lin et al. (2019) fused measures of WorldView-2 data, GeoEye-1 data, TerraSAR-X data, and LiDAR data.

Generally, data level fusion only requires necessary data pre-processing such as image coregistration. It makes the fusion practically straightforward and simple. However, it posts bigger challenges to the fusion models, especially for fusing optical and SAR data. For example, the fusion model has to deal with the geometric differences, the noise differences, and differences of data unit (reflectance and SAR intensity). As shown in Table. 1, about 10% articles that are reviewed in this dissertation carried out their fusion at data level.

3.3.2 Feature level fusion

Feature level fusion joins representations of data sources.

A variety of feature extractors were developed (Haralick et al., 1973; Cloude and Pottier, 1997; Benediktsson et al., 2001). They are designed to extract certain physical information from the data sources. For example, texture information is often extracted by the gray-level co-occurrence matrix GLCM (Bruzzone et al., 1999; Zhu et al., 2012; Bechtel et al., 2016; Erinjery et al., 2018; Zhou et al., 2018b; Tabib Mahmoudi et al., 2019) and wavelet transformations (Wang et al., 2016; Pereira et al., 2018). Spatial information was extracted by the morphological profile (Aravena Pelizari et al., 2018). Geometric information was extracted by using edge detection (Tupin and Roux, 2003) and objects based parameters, such as area and perimeters (Kussul et al., 2016; Aravena Pelizari et al., 2018). Polarimetric information was extracted by decomposition (Baghdadi et al., 2016; Yang et al., 2017; He et al., 2017; Demuzere et al., 2019; Zhang et al., 2019a). SAR coherence feature was applied in (Wegner et al., 2010; Zhou et al., 2018b). More similar representations are physical content index, such as NDVI, Normalised Difference Snow Index (NDSI), and Normalised Difference Water Index (NDWI) (Ward et al., 2014; Robson et al., 2015; Baghdadi et al., 2016; Zhang et al., 2019b). These hand-craft feature extractors are chosen to directly provide relevant physical information for targeted applications.

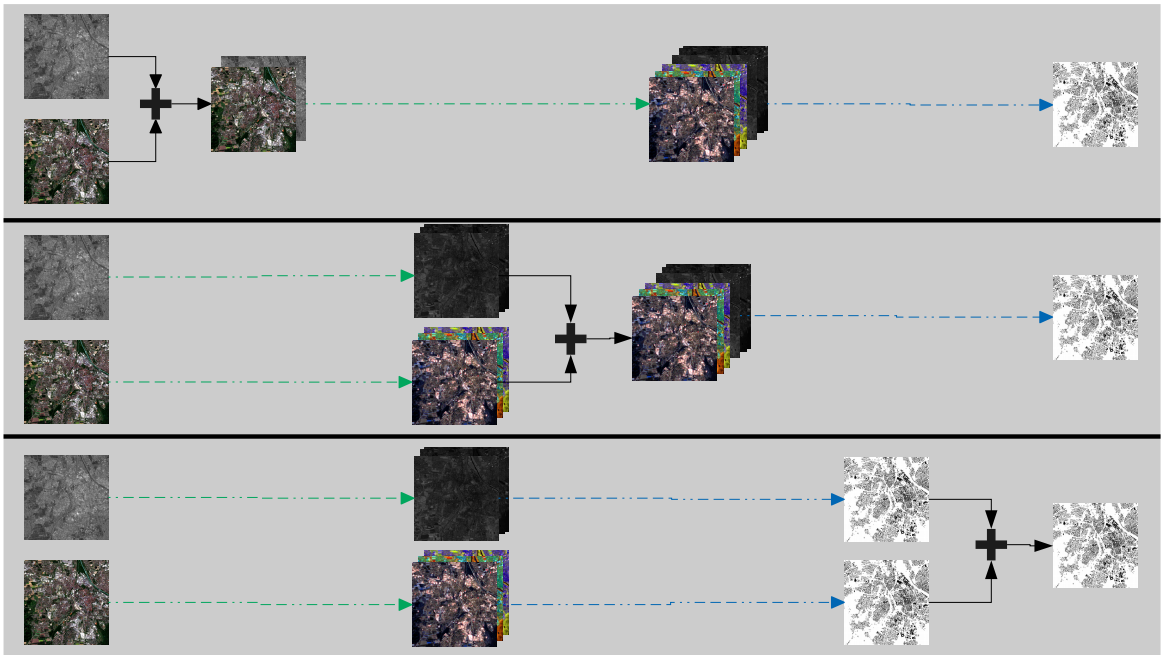


Fig. 3.2. The data fusion workflow has generally two procedures: feature extraction from physical data sets (green arrow) and decision from extracted features (blue arrow). The plus sign and black arrow indicate where the fusion takes place in the workflow. From top to bottom: data level fusion, feature level fusion, and decision level fusion.

Besides these feature extractors which are driven by deriving representations of physical information, neural network extracts and fuses representations that are directly driven by the loss functions (Serpico and Roli, 1995; Bruzzone et al., 1999; Vaglio Laurin et al., 2013; Baghdadi et al., 2016; Kussul et al., 2016; Zhang et al., 2019a).

Apparently, feature extraction is a crucial part in a data fusion workflow. For the conventional hand-designed feature extraction, it is a challenge to find out a set of informative features from endless combinations of numerous feature operators. The recent advanced deep learning technique is able to extract powerful features in an end-to-end fashion, yet heavily relies on large amount of training data. Comparing to data level fusion, feature level fusion is less demanding on the fusion models since the fused features are more informative in terms of targeted applications. Two thirds of the reviewed articles carried out the fusion in the feature level, which makes it the most frequently practiced fusion strategy among all.

3.3.3 Decision level fusion

Decision level fusion combines decisions of different data sources.

Since optical and SAR data measure different physical quantities, physical models were naturally employed in the decision fusion strategy (Ban et al., 2010; Taubenböck et al., 2012; Gamba and Aldrighi, 2012; Qin et al., 2015; Chen et al., 2017; Tong et al., 2018). Apart from fusing optical and SAR data using physical model, statistical fusion is also a commonly practiced technique. Such statistical fusion is usually done by modeling the relation between the individual decisions of the two data sources and the final decision. Often applied models are Bayesian model (Solberg et al., 1994; Lehmann et al., 2012; Reiche et al., 2018), evidence model (Hegarat-Mascle et al., 1997), joint distribution (Qin et al., 2017; Yousif and Ban, 2017; Guo et al., 2019), or sequential model (Revill et al., 2013). About a quarter of reviewed literature carried out the fusion at this level.

The decision fusion often supported by a strong expert knowledge. It appears as a empirical decision in physical models and as predefined distributions in statistical models.

3.4 Fusion Techniques

To summarize technical developments of the fusion of optical and SAR data, Table. 1 demonstrates an statistical investigation result of recent research articles that are published in several journals, such as Remote Sensing of Environment, IEEE Transactions on Geoscience and Remote Sensing, and so on. Mainly six types of fusion models are found in literature. This part briefly summarizes these fusion models.

		Fusion Level			
		Data Level	*Feature Level*	*Decision Level*	Total number
Fusion Technique	Physical model	2	9	11	22
	Feature model	-	5	-	5
	Statistical model	4	9	11	24
	Kernel model	1	10	1	12
	Ensemble model	1	14	1	16
	Neural network model	-	7	-	7
Summary		8	54	24	86

Table 1. This table summarizes 86 reviewed research papers for the topic of fusing optical data and SAR data from **the fusion level** and **the fusion technique**.

3.4.1 Physical model

The physical fusion model summarizes fusion techniques whose fusion procedure is powered by either a data characteristic, a knowledge based rule, a model of physical procedures, or a model of natural phenomenon.

Some studies are motivated to accomplish the fusion of optical and SAR data because of the different data characteristics. Lin et al. (2019) classified non-shadowed areas with optical data, SAR data, and LiDAR data, and classified shadowed areas mainly with SAR data and LiDAR data. This fusion procedure explicitly considered the fact that SAR and LiDAR data are independent from Sun illumination. Spröhnle et al. (2017) utilized the fact that SAR data has a strong reflected signal for man-made structures so that it detected edges from SAR data which were used as seeds for detecting dwelling objects in optical data. It helped to differ close-standing dwellings that can not be separated in optical data. Similarly, Iannelli and Gamba (2019) extracted seeds from SAR data which were applied in optical data to detect urban extent. Ban et al. (2010) applied a decision rule to fuse crop classification results by considering class-wise classification performance of optical data and SAR data, respectively.

Powered by knowledge and experiences, remote sensing experts are often able to link physical meaning of data to a phenomenon under study. After detailed analysis of optical data and SAR data in terms of glaciers, Robson et al. (2015) developed a classification scheme that explicitly utilizes physical information carried by NDVI, NDWI, NDSI, DEM, and PolSAR coherence for detecting glacier. Tabib Mahmoudi et al. (2019) considered physical content of SAR and optical data, such as texture contrast, vegetation coverage, and spectral content, and developed an object classification scheme. Tupin and Roux (2003) detected linear

structures in optical and SAR data, and fused those features to outline buildings, with a geometric regularization.

Some objects or phenomenon under study can be mathematically modeled. For example, Zheng et al. (2017) estimated snow thickness on sea ice by using optical data, SAR data, and a snow melting model. Tao et al. (2019) estimated the soil moisture by developing a backscattering model that use both SAR and optical data.

Some studies fused data sources in a way which simply suits facts. Chen et al. (2017) derived sea water body from SAR data, and mangrove forest from optical data. It fused both information to derive a better mangrove forest due to a fact that mangrove grows with sea water. Qin et al. (2015) applied MODIS derived NDVI as a mask to a PALSAR derived forest map so that non-vegetation land cover that is similar to the forest in PALSAR data can be eliminated from the forest map. Dong et al. (2013) combined a phenology feature of rubber tree that derived from optical data with a forest map derived from SAR data by intersection. Gessner et al. (2015) derived vegetation land cover with MODIS data, urban cover with TanDEM-X and TerraSAR-X data, and water body with ASAR data, finally fused all information to achieve a land cover classification. To detect an area impacted by flood, an intuitive way is to detect the water body before and after a flooding event and to find the difference. Tong et al. (2018) derived water maps of before and after flooding event by using optical data and SAR data, respectively. Flooded area is achieved by differing water maps. Taubenböck et al. (2012) extracted urban extent using optical data and SAR data of different time, and detected change of urban extent along times. Similarly, Lehmann et al. (2015) derived a temporal monitoring of forest areas by using optical data and SAR data of sequential times.

Fusion studies with a physical model are often designed for a very specialized application, such as flood detection, snow thickness detection, dwelling detection, and soil moisture estimation. It requires not only the knowledge of remote sensing techniques, but also a rich expertise of the associated application. And those models often employ simple techniques, such as thresholding and logic operations, which are empirical operations resulted from strong expertise. On the other hand, those fusion models have a poor power of generalization. They can barely be applied to other applications. Even with the same application of a different region, these models often have to be re-studied to deliver an acceptable result.

3.4.2 Feature model

Feature fusion models extract or select a new set of features from data sources to be fused in terms of designed objectives. Typical techniques are principal component analysis (Pearson, 1901), canonical correlation analysis (Härdle and Simar, 2007), locality preserving projections (He and Niyogi, 2004), and so on.

Pereira et al. (2018) fused PALSAR-1 data and LandSat data via multiple feature analysis models (wavelet-based models, principal components analysis, Brovey transformation, and Ehlers fusion), and classified these feature for agricultural land classification. Yang et al. (2017) extracted about 150 features out of PolSAR and multispectral data which have concrete physical meaning in a sense of SAR polarimetry or spectrum, optimized the feature selection procedure with a goal of minimizing intra-class variance and maximizing inter-class variance, and finally classified the selected features with a kernel based classifier, multiclass relevance vector machine. Wang et al. (2016) fused data features with discrete wavelet transforms, extracted objects afterwards, and classified the objects with random forest and support vector machine. Yang et al. (2016) fused SAR and optical data by representing them in an orthogonal system via Gram-Schmidt algorithm, and classified those extracted features with a support vector machine. Liu et al. (2018) fused optical and SAR data of different time into a latent space to accomplish change detection.

Fusion studies with a feature model often attempt to derive informative features that joins complementarity of data sources under the guidance of a loss function. The loss function is of key importance and often directly related to a targeted application. For example, a loss function of feature extraction for a classification task often aims at maximize inter-class variance and minimize intra-class variance. For this kind of fusion model, some details could have a great influence, such as normalization, number of dimension for each data source, and so on.

3.4.3 Statistical model

Statistical models fuse optical and SAR data with a statistical assumption which could be a distribution of data noise, distributions of data in classes, or statistical connection along time. Prendes et al. (2015) statistically modeled a homogeneous area as a joint distribution of optical and SAR data, and sovled it with a weighted maximum likelihood estimator. Wan et al. (2019) applied a joint conditional posterior probability model to fuse optical data and SAR data. Guo et al. (2019) fused decisions of the two data sources via a statistical model. Reiche et al. (2018) accomplished a temporal forest monitoring task with a Bayesian model using previous results as prior and future results as evidence for each time point. Zhang et al. (2019b) trained a kernel ridge regression (KRR) model to find the relation between forests that were derived from SAR and optical data respectively, and predicted a SAR-derived forest for 2011 to 2014 when SAR data was not accessible. Næsset et al. (2016) fused multi-sources remote sensing data by applying a multivariate regression model for estimating above ground biomass and utilized a logistic regression model to predict a forest area. Wang et al. (2017) carried out a statistical analysis on multiple remote sening data sources of training areas, and decided some thresholds that were statistically significant for the detection of red cedar. Lucas et al. (2014) statistically analyzed a landSat-derived foliage projective cover (FPC) and PALSAR data corresponding to training areas, and located statistical significant thresholds for mapping forest. Similar ideas can be found in (Qin et al., 2017)

Fusion studies with a statistical model often have a strong assumption, such as the noise distribution and class data distributions. It often works well if the assumption is valid. However, it might be difficult to find a concrete statistical assumption for certain remote sensing applications.

3.4.4 Kernel model

Kernel models project data from its feature space where a linear algorithm is not able to solve a targeted problem to another high-dimensional space where the problem is linearly solvable (Cortes and Vapnik, 1995). The practical power of a kernel model is that, instead of projecting all data into an unknown high-dimensional space, the only thing that is required in the model is how to calculate a dot product of the unknown space (Camps-Valls et al., 2008). The solution is a kernel. A kernel takes a data pair from the original feature space and outputs a dot product of another space for the data pair. Regarding data fusion, data fused at the feature level often results in a complex feature space where linear models fail. Projecting those features into a high-dimensional space where linear models work is an efficient procedure of data fusion. A typical kernel method is support vector machine (SVM). It finds a decision boundary which has maximum distances to the training samples.

As SVM has been proven efficient in remote sensing data fusion (Camps-Valls et al., 2008; Mountrakis et al., 2011; Huang and Zhang, 2013), it is frequently applied in the fusion of optical data and SAR data. It was reported in urban applications (Haas and Ban, 2017; Ban and Jacob, 2013; Xu et al., 2017), forest monitoring (Wang et al., 2016; Rajah et al., 2018,

2019), agricultural classification (Waske and Benediktsson, 2007), snow detection (He et al., 2017), and land cover land use classification (Sukawattanavijit et al., 2017; Xiao et al., 2016). In stead of SVM, Sun et al. (2019) developed a multiple kernel learning for urban impervious surface mapping. And Yang et al. (2017) applied a multiclass relevance vector machine on rice phenology estimation. The kernel based fusion model was compared to other fusion models. Zhang et al. (2015) compared different fusion models, including a neural network, a kernel model (SVM), statistical models (maximum likelihood and minimum distance), for an urban classification task, and concluded that SVM had the best performance.

The kernel fusion model was widely utilized in the fusion of optical data and SAR data, mostly on classification task with SVM. It often performs quite well. However, practically, its shortage lays at high computational cost caused by a grid searching of hyper-parameters. It became very noticeable when dealing with large data sets. And pre-setting of a hyper-parameter grid is also crucial for the final performance.

3.4.5 Ensemble model

Ensemble models joint a number of models to increase the performance of final prediction (Rokach, 2010). A typical algorithm of ensemble model is the random forest (Breiman, 2001). A random forest trains a group of tree structures which utilize randomly different subsets of features and randomly different subsets of training samples. Fusing decisions of those tree structures often deliver an improved result comparing to the ones provided by individual tree structure. The underlying power of an ensemble model roots from the statistical principle of *Law of Large Numbers* which leads a model a robust performance and avoiding overfitting. Since the tree structures in random forest utilize randomly different subsets of features, it means random forest implicitly carries out a feature selection. When multiple data sources were put into random forest, the algorithm automatically carries out a data fusion mechanism by trying out different feature combinations. Practically, it often works well. Villa et al. (2015a) compared different fusion models on the fusion of LandSat-8 data and TerraSAR-X data, including a statistical model (classification and regression trees), an ensemble model (random forest), a kernel model (support vector machines), and a neural network model (multilayer perceptron), and concluded the random forest had the best performance. Zhou et al. (2018b) also reported a superior performance of random forest by comparing to SVM, on a task of urban land cover mapping.

Since random forest has a stat-of-the-art performance and low computational cost, it was widely applied on the fusion of optical data and SAR data. It was applied on forest mapping (Carreiras et al., 2017), urban land cover land use classification (Stefanski et al., 2014; Bechtel et al., 2016; Forget et al., 2018; Zhou et al., 2018a), impervious surface estimation (Zhang et al., 2014; Tan et al., 2015; Aravena Pelizari et al., 2018), and agriculture application (van Beijma et al., 2014; Villa et al., 2015b; Erinjery et al., 2018).

3.4.6 Neural Network model

Neural network models have been proven successful machine learning methods that surpass human capability (Krizhevsky et al., 2012a; He et al., 2016). These models are capable of learning very complicated nonlinear relationship, but depends on a large amount of training data set. However, in the fusion of optical data and SAR data, less than 10% of the studies employed a neural network model, according to our survey in Table. 2. Among a small proportional studies, Serpico and Roli (1995) and Bruzzone et al. (1999) started studied of neural network for agriculture classification tasks in 1990s. Baghdadi et al. (2016) and Kussul et al. (2016) have recently applied the model on agriculture applications. It was also employed on urban applications (Pacifici et al., 2008; Villa et al., 2015a; Zhang

et al., 2015, 2019a) and forest (Vaglio Laurin et al., 2013). Zhang et al. (2015) compared a neural network model with other fusion models, and reported that SVM had the best performance. Villa et al. (2015a) also reported a neural network model, multilayer preceptron, is not comparable to random forest regarding classification accuracy. Their conclusions might be a result of the neural network models applied in their studies were not a stat-of-the-art structure. And lacking of training data set could also be a reason why neural network was not performing well. It is very expensive to label remote sensing data for a high demand of expert knowledge, especially for the SAR data. In order to fully apply the power of neural network, one problem needs to be solved is to produce large amount of labeled data set of optical and SAR data pairs.

		Fusion Level			
		Data Level	*Feature Level*	*Decision Level*	Total number
Fusion Application	Agriculture	2	14	6	22
	Forest	3	9	6	18
	Flood detection	-	3	1	4
	Snow detection	-	2	1	3
	Glacier monitoring	-	1	1	2
	Urban	3	25	9	37
Summary		8	54	24	86

Table 2. This table summarizes 86 reviewed research papers for the topic of fusing optical data and SAR data from **the fusion level** and **the application**.

3.5 Fusion Applications

Applications are the fundamental driving power of fusing optical and SAR data. Among summarized research articles, roughly six different applications are involved, which are: agriculture, forest, flood detection, snow detection, glacier monitoring, and urban, as shown in Table. 2. The three most popular applications in the reviewed articles are urban applications, agriculture applications, and forest applications, which are 37 (43.02%), 22 (25.58%), and 18 (20.93%), out of 86 articles, respectively. These percentages somehow reflect the distribution of demands in the remote sensing field.

In urban applications, more specific tasks are urban expansion (Taubenböck et al., 2012; Li et al., 2017b), change detection (Qin et al., 2017; Wan et al., 2019; Demuzere et al., 2019), classification (Ban et al., 2010; Haas and Ban, 2017; Tabib Mahmoudi et al., 2019), urban extent detection (Iannelli and Gamba, 2019; Forget et al., 2018), and impervious surface estimation (Zhang et al., 2019a; Sun et al., 2019). In forest applications, popular tasks are classification (Carreiras et al., 2017; Cremon et al., 2014; Erinjery et al., 2018; Pereira et al., 2018; Vaglio Laurin et al., 2013; Lehmann et al., 2012), monitoring (Reiche et al., 2015; Lehmann et al., 2015; Lucas et al., 2014; Zhang et al., 2019b; Reiche et al., 2018; Qin et al., 2015), detection (Baumann et al., 2018; Wang et al., 2017; Chen et al., 2017; Dong et al., 2013), and biomass estimation (Næsset et al., 2016). In agricultural applications, studies focus on classification (Bendjebbour et al., 2001; Ramsey et al., 2014; Bruzzone et al., 1999; Kussul et al., 2016; Gessner et al., 2015; Waske and Benediktsson, 2007; Hegarat-Mascle

et al., 1997), soil moisture retrieve (Baghdadi et al., 2016; Tao et al., 2019), detection (Rajah et al., 2018; van Beijma et al., 2014; Guo et al., 2019), crop monitoring(Villa et al., 2015b; Yang et al., 2017; Veloso et al., 2017; Revill et al., 2013). Among these tasks, classification and detection are the most frequently studied cases.

3.6 Summary

For the fusion of optical data and SAR data, according to the survey shown in Table. 1 and 2, this dissertation concludes that: (1) most studies fused data sources at the feature level; (2) the urban, agriculture, and forest are three most popular applications among literature; and (3) random forest was the mostly used technique and reported good performance. According to author's latest literature survey, hyperspectral data is baraly studied for the fusion with the SAR data.

4 Summary of the Work

The objectives of this dissertation are addressed in four published peer-reviewed journal papers and one peer-reviewed journal paper under review. This chapter gives a summary of these papers including the following main contributions:

- A comprehensive research on feature engineering of Sentinel-1 dual-Pol SAR data is presented in Appendix A and summarized in Section 4.1 which targeted on an LCZ classification task of 29 major cities around the world.
- A manifold data fusion algorithm (MIMA) is proposed in Appendix B and summarized in Section 4.2 which incorporates a recently developed mathematical tool, topological data analysis, with a semi-supervised manifold alignment technique to accomplish fusion and classification of optical and SAR data.
- A comprehensive comparison of manifold-based data fusion algorithms for the fusion and classification of hyperspectral and PolSAR data is presented in Appendix C and Section 4.3 which mainly discussed the impacts of fusion domains and supervision strategies, and secondarily served as an intensive evaluation of MIMA.
- A classification system was developed based on MIMA with an ensemble strategy tackling the practical issue of hyperparameter tuning in Appendix E. Its classification performance was evaluated over three cities, Munich, Mumbai, and New York. It outperformed three other state-of-the-art large scale classification systems.
- An urban analysis of 42 major cities around the world was carried out in Appendix D and Appendix E. Combined with a global population data set, 42 produced urban classification maps were applied for analyzing the urban morphological structures and its related population distributions. Eight urban patterns were identified from the 42 cites.

4.1 A Demonstration of Data Complementarity

This section discusses a fundamental question: ***'why should one fuse optical and SAR data?'***. This question has been theoretically discussed in Section 2.1.4. In the following, benefits of the fusion is demonstrated by experiments of a land cover land use classification task.

4.1.1 Feature engineering

Feature engineering is a crucial prerequisite to data fusion for three reasons.

Power of discrimination. Feature engineering extracts features from a data source which improves the representation capability of the original data for a certain purpose. It is the discriminating power for classification tasks. This procedure has a vital influence on the final performance of a classification and data fusion task.

Number of dimensions. Feature engineering should provides an appropriate number of features. A small number of features might lead to insufficient discriminating power of features. A large number of features leads to a well known problem of *'curse of dimensionality'* (Donoho et al., 2000).

Credit of fusion. A data fusion strategy should demonstrate its superior performance against single data usage. Therefore, the feature engineering of both cases should be identical for the purpose of comparison. Moreover, it should also grantee that the usage of single data is at least close to its upper limit in terms of the feature design. In this case, it is persua-

sive that the final superiority is contributed by the fusion strategy but not by the unexplored potential of individual data set.

Due to above-mentioned reasons, before answering the question: ***'why should one fuse optical and SAR data?'***, this part attempts to construct a persuasive baseline by carrying out a reasonable feature engineering for each of the optical and SAR data.

4.1.1.1 Feature engineering of Sentinel-1 data

Feature engineering of Sentinel-1 data was comprehensively analyzed in Appendix A in terms of a land cover land use classification scheme named local climate zone (Stewart and Oke, 2012). The study was carried out on 29 major cities which locate on different continents, in different cultural regions, and in different environments.

Four types of features that are often reported in literature are investigated in this work. They are *polarimetric features*, *local statistical features*, *texture features*, and *morphological features*.

Polarimetric features were reported being able to retrieve information of shapes, orientations, dielectric properties and scattering procedures (Lee and Pottier, 2009; Moreira et al., 2013; Schmitt et al., 2015). The ones employed in this study are intensity of VH ($|S_{VH}|^2$), intensity of VV ($|S_{VV}|^2$), relative phase of VH and VV ($actan2(\langle S_{VH}S_{VV}^*\rangle)$), and normalized coherence γ of VH and VV (Eq. 4.1),

$$\gamma = \frac{\langle S_{VH}S_{VV}^*\rangle}{\sqrt{\langle|S_{VH}|^2\rangle\langle|S_{VV}|^2\rangle}}, \tag{4.1}$$

where S_{VV} and S_{VH} are the VV and VH complex elements in Eq. 2.1, and (*) denotes complex conjugate.

local statistical features, such as mean and variance, have been derived from remote sensing images as features describing local neighborhood. It has been proven efficient in terms of local climate zone classification in (Bechtel and Daneke, 2012; Yokoya et al., 2017). Therefore, this work investigated a group of statistical indicators. They are: maximum, minimum, mean, standard deviation, and median of a local neighborhood of 100 meter by 100 meter. All these indicators are derived from each of the four polarimetric features.

Texture features were derived by GLCM. It has been proven efficient in SAR classification in multiple studies (Zhu et al., 2012; Geng et al., 2015; Bechtel et al., 2016; Zhou et al., 2018b; Tabib Mahmoudi et al., 2019). Therefore, GLCM statistical features including contrast, dissimilarity, homogeneity, angular second moment, maximum probability, entropy, mean, variance, and correlation, are employed in this study.

Morphological features were extracted by morphological profile (MP) which was applied in literature of optical classification (Benediktsson et al., 2005; Ghamisi et al., 2014), SAR classification (Wurm et al., 2017), and classification of their fusion (Aravena Pelizari et al., 2018). This study also took it into consideration.

As shown in Fig. 4.1, the combination of statistical features and morphological profiles had the best performance. Therefore, this dissertation concludes from this study that the combination of statistical features and morphological profiles is a reasonable feature combination of Sentinel-1 data which could be applied in further data fusion.

4.1.1.2 Feature engineering of hyperspectral data

Feature engineering of hyperspectral data is a well studied topic in literature. Among them, spectral-spatial feature extraction is the most efficient and frequent practice (Fauvel et al., 2012; Ghamisi et al., 2013; Zhao and Du, 2016; Zhang et al., 2016). In this dissertation, in order to prepare a reasonable feature combination of hyperspectral data for further data

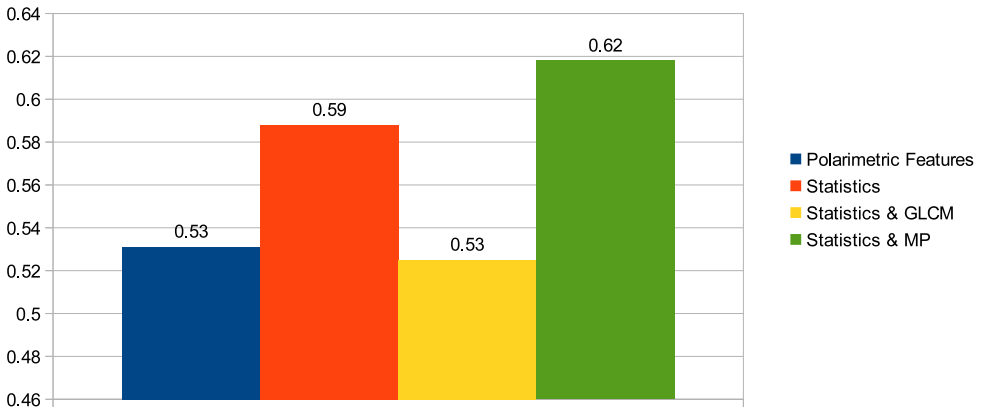


Fig. 4.1. This chart demonstrates LCZ classification overall accuracy that are achieved by classifying indicated feature combinations of Sentinel-1 data. The employed classifier is a canonical correlation forest (CCF). Training samples are collected from 20 cities, and testing samples are collected from the other 9 cities.

fusion, it follows the well-proven spectral-spatial feature extraction strategy. More specifically, the feature engineering of hyperspectral data in this dissertation follows the strategy appeared in (Ghamisi et al., 2013, 2014; Rasti et al., 2017). It carried out a dimension reduction on hyperspectral data to tackle the problem of *'curse of dimensionality'*, meanwhile the dimension-reduced data preserves the spectral information. It then derived spatial features from the dimension-reduced data as a representation of spatial information. For joining both information, they concatenated the dimension-reduced data and the derived spatial features as a data cube. To follow this strategy, this dissertation applies a PCA-based dimension reduction and derives the morphological profiles as the spatial features.

This strategy is tested over the hyperspectral data of a scene for the city Berlin, Germany (Okujeni et al., 2016). As shown in Table. 3, spectral-spatial feature combination had the best performance in terms of overall accuracy and had a reasonable number of dimensions. The experiment result supported the conclusions in multiple studies (Ghamisi et al., 2013, 2014; Rasti et al., 2017). Therefore, this feature combination of the hyperspectral data is applied in this dissertation for further fusions.

Table 3. This table demonstrates performance of feature combinations in terms of classification accuracy for the hyperspectral data of the scene in the city Berlin, Germany (Okujeni et al., 2016). A random forest was employed as the classifier. HSI: original hyperspectral data; HSI-PC: principal components of hyperspectral data; HSI-PC-MP: principal components of hyperspectral data and derived morphological profiles. The number in the bracket parendissertation indicates the number of features.

RF	**HSI [244]**	**HSI-PC [4]**	**HSI-PC-MP [28]**
OA	0.6881	0.6704	0.7021
KAPPA	0.5283	0.5018	0.5437

4.1.2 Data complementarity

Having confirmed the feature engineering strategies, this part attempts to answer the question ***'why should one fuse optical and SAR data?'*** by classification experiments over two real world data sets.

The experiment classifies a land cover land use scheme over two scenes, one in the city Berlin, Germany, and the other in the city Augsburg, Germany. Both data sets include Sentinel-1 data and simulated space-borne hyperspectral data. The classification outcomes

Table 4. This table demonstrates classification accuracy of using PolSAR data, hyperspectral data, and both data, on a scene of city Berlin, Germany. Five classifiers were employed. The joint usage of two data sources outperforms usages of individual data source, in terms of kappa coefficient, average accuracy, and overall accuracy, of every classifier and also the mean overall accuracy.

Data	Classifier	Forest	Residential Area	Industrial Area	Low Plants	Soil	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
PolSAR	*1NN*	40.64	57.67	25.14	32.94	56.88	32.19	30.37	33.85	0.2927	38.71	48.92	56.76
	LSVM	33.02	77.92	13.85	36.46	72.6	40.64	32.23	37.68	0.4012	43.05	60.94	
	KSVM	34.36	69.94	20.38	30.61	68.27	38.62	32.79	42.82	0.3566	42.23	55.76	
	RF	35.61	72.3	25.63	28.66	66.38	43.9	37.87	45.39	0.3789	44.47	57.61	
	CCF	37.96	76.87	24.87	30.69	64.72	38.82	36.88	41.34	0.4035	44.02	60.56	
HSI	*1NN*	68.78	63.87	30.01	57.58	90.73	55.76	32.86	73.89	0.4599	59.18	61.64	70.14
	LSVM	69.2	82.5	18.55	65.7	79.06	53.59	44.77	72.81	0.585	60.77	73.48	
	KSVM	72.58	78.68	35.43	63.74	74.18	56.87	31.58	74.29	0.5625	60.92	71.34	
	RF	66.65	79.64	30.25	57.44	75.33	47.77	35.17	78.1	0.5437	58.79	70.21	
	CCF	71	81.86	31.54	68.95	81.36	53.47	38.35	74.81	0.597	62.67	74.03	
HSI & PolSAR	*1NN*	64.83	69.7	32.89	65.27	83.81	54.77	34.59	63.51	0.4975	58.67	65.44	73.73
	LSVM	66.57	86.24	30.48	75.3	79.61	53.52	40.12	76.11	0.6329	63.49	76.93	
	KSVM	67.27	80.93	41.78	64.02	72.37	57.58	33	74.6	0.5764	61.44	72.36	
	RF	63.46	84.99	37.79	74.38	82.72	56.26	40.61	82.09	0.6266	65.29	76.26	
	CCF	71.51	86.27	34.05	72.03	83.24	56.3	44.33	77.7	0.6445	65.68	77.67	

are demonstrated in Table. 4 and 5. PolSAR data (PolSAR), hyperspectral data (HSI), and a concatenation of both data (HSI & PolSAR), are classified by five different classifiers.

For both scenes, a classification based on two data sources outperforms classification using individual data source with all five classifiers, in terms of kappa coefficient, average accuracy, and overall accuracy. The mean value of overall accuracy over the five classifiers provides the same outcome.

Based on the intensive experiment, it is obvious that the data complementarity is benefiting the classification application. Therefore, this dissertation provides the answer to the question ***'why should one fuse optical and SAR data?'*** by a theoretical analysis in section 2.1.4 and an experiment demonstration in this section. The conclusion and experiments are parts of the work represented in Appendix C.

Table 5. This table demonstrates classification accuracy of using PolSAR data, hyperspectral data, and both data, on a scene of city Augsburg, Germany. Five classifiers were employed. The joint usage of two data sources outperforms usages of an individual data source, in terms of kappa coefficient, average accuracy, and overall accuracy, of every classifier and also the mean overall accuracy crossing classifiers.

Data	Classifier	Forest	Residential Area	Industrial Area	Low Plants	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
PolSAR	*1NN*	64	35.88	38.8	55.02	22.54	38.9	18.66	0.2897	39.11	39.11	48.21
	LSVM	86.93	46.44	39.15	73.17	25.37	44.29	21.8	0.3952	48.16	48.16	
	KSVM	86.51	64.49	31.41	81.98	22.39	41.98	19.12	0.4131	49.7	49.7	
	RF	81.88	63.44	47.76	88.46	28.88	38.71	14.63	0.4396	51.97	51.97	
	CCF	82.29	61.85	47.8	88.37	30.34	38.07	16.1	0.4414	52.12	52.12	
HSI	*1NN*	27.9	52.49	61.1	78.2	60.66	24.9	55.24	0.4341	51.5	51.5	51.33
	LSVM	25.44	50.22	75.93	67.46	38.32	15.15	57.93	0.3841	47.21	47.21	
	KSVM	31.2	65.2	70.71	86.37	55.98	20.8	54.63	0.4748	54.98	54.98	
	RF	25.59	58.29	70.29	84.34	40.41	15.98	52.98	0.4131	49.7	49.7	
	CCF	27.29	64.56	75.71	84.68	48.29	16.54	55.66	0.4546	53.25	53.25	
HSI & PolSAR	*1NN*	34.76	58.17	55.93	84.56	57.73	34.9	54.88	0.4682	54.42	54.42	56.71
	LSVM	31	65.95	73.29	83.85	36.9	25.07	42.85	0.4315	51.28	51.28	
	KSVM	40.59	67.83	67.07	92.59	45.24	27.1	55.78	0.4937	56.6	56.6	
	RF	61.27	73.88	70.1	94.98	47.51	25.63	59.17	0.5542	61.79	61.79	
	CCF	46.07	75.63	78.05	95.51	58.07	18.49	44.22	0.5267	59.44	59.44	

4.2 A TDA Guided Manifold Alignment Data Fusion — the MIMA algorithm

The previous section demonstrated the complementarity of optical and SAR data, namely hyperspectral and PolSAR data. This section summarizes Appendix B which investigated the fusion of optical and SAR data using manifold-based techniques and introduced a new semi-supervised data fusion algorithm, Mapper-Induced Manifold Alignment (MIMA). MIMA utilized recently emerged topological data analysis to model the underlying manifold of data sets and accomplished the fusion by aligning those derived manifolds.

4.2.1 Semi-supervised manifold alignment

Wang and Mahadevan (2011) introduced a semi-supervised manifold alignment (SSMA) for the task of data fusion. It retrieves the underlying manifold for each of the data sources by using local connectivity, and pursues a latent space where the manifold of each data source is preserved while manifolds are aligned by the guidance of annotated data. Tuia et al. (2014) applied this idea on the fusion of multiple optical remote sensing images. Tuia and Camps-Valls (2016) later upgraded SSMA with a kernel technique which is capable of dealing complicated nonlinear alignment. The technique has been proven to be efficient for data fusion. However, according to the author's literature review, it has never been studied for the fusion of optical and SAR data. Therefore, it is worthy to study the topic of fusing optical and SAR data using manifold alignment techniques.

4.2.1.1 SSMA merits for the fusion

Optical and SAR data have a severe difference due to their imaging geometry and physical measures. As mentioned in section 2.2.3, manifold alignment is capable of bridging gaps between data domains, such as cross-lingual gap (Wang and Mahadevan, 2008). Therefore, manifold alignment based technique is promising for the fusion in terms of bridging the difference of optical and SAR data.

Thanks to the low dimensional essence of a manifold, manifold-based techniques are capable of finding a low dimensional representation of a high dimensional data. Moreover, the technique is attractive for solving the fusion of optical and SAR data, because remote sensing data often have a redundantly high dimension, especially for the hyperspectral data.

A manifold essentially represents relations of data samples in a data set, including samples with and without annotation. Therefore, the semantic information carried by the labels is propagated to samples without annotation via the manifold structure. It implicitly amplifies the effect of annotated data. It is well known in remote sensing that labeled data is expensive to access. Therefore, in terms of making the most usage of annotated data, semi-supervised manifold alignment is also a promising option for the methodology.

4.2.1.2 An intuitive interpretation of SSMA

SSMA aims to optimize projections that map multiple data sources into a latent space. While optimizing the projections, SSMA attempts to accomplish three goals: ***(a)*** pulling data of the same label together; ***(b)*** pushing data of different label apart; and ***(c)*** preserving the manifold structure of individual data source. These are three properties pursued for the fused data. Targeting at ***(a)*** and ***(b)*** is equivalent to maximizing inter-class variance and minimizing intra-class variance which benefits classification tasks. Targeting at ***(c)*** preserves the intrinsic manifold of each data source, and implicitly propagates label information to unlabeled data via connections of the labeled and unlabeled data. The following paragraphs introduce SSMA in detail.

4.2.1.3 The technical detail of SSMA

K data sources are $\{\mathbf{X}^k \in \mathbf{R}^{n_k} |\ \mathbf{x}_1^k, \mathbf{x}_2^k, ... \mathbf{x}_i^k, ..., \mathbf{x}_{N_k}^k\}$ where $\{k = 1, 2, ..., k, ..., K\}$, $\mathbf{x}_i^k$ is an n_k by 1 vector representing the i^{th} data point of the k^{th} data source, and N_k is the number of data points in $\mathbf{X}^k$. A subset of the k^{th} data source $\mathbf{X}_s^k \subset \mathbf{X}^k$ is equipped with labels $\mathbf{B}^k$ where $b_j^k \in \mathbf{B}^k$ is the label of j^{th} data point $\mathbf{x}_{s,j}^k$. Manifold alignment extracts an N_k by N_k adjacency matrix $\mathbf{A}^k$ for $\mathbf{X}^k$ where $\mathbf{A}^k(i,j) = 1$ if $\mathbf{x}_i^k$ and $\mathbf{x}_j^k$ are connected, otherwise $\mathbf{A}^k(i,j) = 0$. The adjacency matrix $\mathbf{A}^k$ is interpreted as a representation of topological structure for $\mathbf{X}^k$. A combined adjacency matrix $\tilde{\mathbf{A}}_t$ can be formed as (4.2) to represent topological structures of all data sources.

$$\tilde{\mathbf{A}}_\mathbf{t} = \begin{pmatrix} \mathbf{A}^1 & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & \mathbf{A}^K \end{pmatrix} \tag{4.2}$$

Furthermore, with given labels, one can construct a similarity matrix (4.3) and a dissimilarity matrix (4.4).

$$\tilde{\mathbf{A}}_s = \begin{pmatrix} \mathbf{A}_s^{1,1} & ... & \mathbf{A}_s^{1,K} \\ ... & ... & ... \\ \mathbf{A}_s^{K,1} & ... & \mathbf{A}_s^{K,K} \end{pmatrix} \tag{4.3}$$

where $\mathbf{A}_s^{k,k'}(i,j) = 1$ if $\mathbf{x}_i^k$ and $\mathbf{x}_j^{k'}$ have the same label, otherwise $\mathbf{A}_s^{k,k'}(i,j) = 0$

$$\tilde{\mathbf{A}}_d = \begin{pmatrix} \mathbf{A}_d^{1,1} & ... & \mathbf{A}_d^{1,K} \\ ... & ... & ... \\ \mathbf{A}_d^{K,1} & ... & \mathbf{A}_d^{K,K} \end{pmatrix} \tag{4.4}$$

where $\mathbf{A}_d^{k,k'}(i,j) = 1$ if $\mathbf{x}_i^k$ and $\mathbf{x}_j^{k'}$ do not have the same label, otherwise $\mathbf{A}_d^{k,k'}(i,j) = 0$

Intuitively, Eq. 4.2 represents topological structures of all K data sources in a combined matrix, Eq. 4.3 connects data samples within a data source and across data sources as long as they share a common label, and Eq. 4.4 marks data samples within a data source and across data sources as long as they don't share a common label.

Within the manifold alignment fusion, a projection $\mathbf{f}_k$ maps data $\mathbf{X}^k$ into a joint manifold while attempting to accomplish the three properties. Therefore, three terms are further formulated.

$$\mathcal{A} = \sum_{k=1}^{K} \sum_{k'=1}^{K} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \|\mathbf{f}_k \mathbf{x}_i^k - \mathbf{f}_{k'} \mathbf{x}_j^{k'}\|^2 \mathbf{A}_s^{k,k'}(i,j). \tag{4.5}$$

Minimizing Eq. (4.5) pulls data of the same class together in the latent space, which meets property ***(a)***.

$$\mathcal{B} = \sum_{k=1}^{K} \sum_{k'=1}^{K} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \|\mathbf{f}_k \mathbf{x}_i^k - \mathbf{f}_{k'} \mathbf{x}_j^{k'}\|^2 \mathbf{A}_d^{k,k'}(i,j). \tag{4.6}$$

Maximizing Eq. (4.6) pushes data of different classes apart, which suits property ***(b)***.

$$\mathcal{C} = \sum_{k=1}^{K} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \|\mathbf{f}_k \mathbf{x}_i^k - \mathbf{f}_k \mathbf{x}_j^k\|^2 \mathbf{A}_t^k(i,j). \tag{4.7}$$

Minimizing Eq. (4.7) preserves the topological structure of individual data set, corresponding to property ***(c)***. To simutanously achieve these three goals, a cost function is formulated as Eq. (4.8) to be minimized.

$$\mathcal{L}(f_1,...,f_K) = (\mathcal{A}+\mathcal{C})/\mathcal{B}, \tag{4.8}$$

Proven in (Wang and Mahadevan, 2009a, 2011), the solution $\mathbf{f}_1,...,\mathbf{f}_K$ that minimizing the cost function $\mathcal{L}(\mathbf{f}_1,...,\mathbf{f}_K)$ is the eigenvectors corresponding to smallest non-zero eigenvalues of the generalized eigenvalue decomposition of (4.9).

$$\mathbf{Z}(\mu\mathbf{L}_t + \mathbf{L}_s)\mathbf{Z}^T\mathbf{f} = \lambda\mathbf{Z}\mathbf{L}_d\mathbf{Z}^T\mathbf{f}, \tag{4.9}$$

where $\mathbf{Z} = \begin{pmatrix} \mathbf{X}^1 & ... & 0 \\ ... & ... & ... \\ 0 & ... & \mathbf{X}^K \end{pmatrix}$, $\mathbf{L}$ is the Laplacian matrix $\mathbf{L}_u = \tilde{\mathbf{A}}_u - \mathbf{D}_u$, $u \in \{s,d,t\}$, and $\mathbf{D}$ is the degree matrix $\mathbf{D}_u(i,i) = \sum_{j=1}^{N_1+N_2+...+N_K} \tilde{\mathbf{A}}_u(i,j)$ and $\mathbf{D}_u(i,j) = 0$ when $i \neq j$.

As long as projections $\mathbf{f}_1,...,\mathbf{f}_K$ are achieved, K data sources can be projected onto the fused manifold.

4.2.2 Topological method of data analysis — MAPPER

Singh et al. (2007) introduced MAPPER as a mathematical tool of data analysis. It attempts to reveal relevant information by analyzing the shape of data from a topological perspective. MAPPER mainly consists of three parts. To summarize MAPPER for an acceptable interpretation, this part provides an intuitive explanation, a mathematical explanation, and a high level interpretation, with an intuitive example shown in Fig. 4.2.

4.2.2.1 An intuitive explanation

- ◇ *Filter function*. Filter function emphasises one aspect of a data, such as geometric shape, physical insight, and so on. MAPPER observes topological structure of the data from the perspective of selected filter function. For the data in Fig. 4.2, the filter function is the horizontal distance to the wrist. Fig 4.2 (B) colorized data points using filtered values.
- ◇ *Data separation*. Given an overlap percentage and a number of intervals, MAPPER divides the filtered values into sequential overlapped intervals as in Fig. 4.2 (C). Correspondingly, data points are assigned to sequential overlapped data intervals as shown Fig. 4.2 (D). Data in intervals have the same dimension as the original data.
- ◇ *Clustering and visualization*. Data in each interval are clustered. Clusters might include common data points due to overlapping. MAPPER constructs a graph where a node represents a cluster, and an edge links two clusters sharing data points. The graph serves as a topological structure of the data. For example, the graph in Fig. 4.2 (E) is a topological structure of a human hand data that is derived by MAPPER.

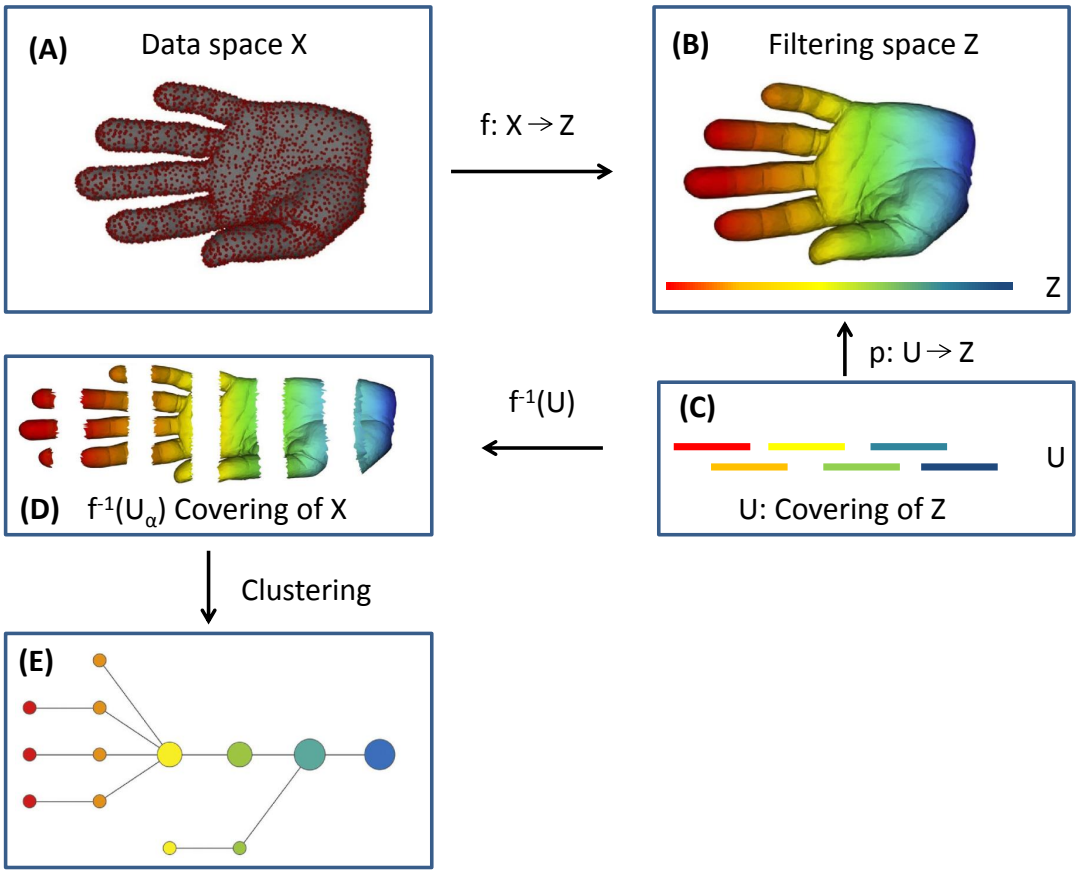


Fig. 4.2. MAPPER Workflow of deriving topological structure of data with an example of a human hand data. (A): Data space X, point cloud data of a human hand; (B): Filtered space Z, points colorized by the filter value; Filter function f: assigning data points with their horizontal distances to the right end; (C): U covering of Z, overlapped intervals of the filtered value; (D): $f^{-1}(U_\alpha)$ covering of X, separating original data into bins according to intervals in (C), data in bins remain their original dimension; (E) $f^{-1}(U_\alpha)$ covering of X, achieved by clustering bins of data. Modified from (Lum et al., 2013).

4.2.2.2 A theoretical explanation

First of all, the concept of covering in topology has to be explained. Munkres (2014) explained it as: let $p : U \to Z$ be continuous and surjective. If every point z of Z has a neighborhood C that is evenly covered by p, then p is called a covering map, and U is defined to be a covering space of Z, then p is a local homeomorphism of U with Z. It means that, in terms of function p, the preimage in U and the image in Z share the same topological properties locally.

- ◇ *Filter function*. A continuous function $f : X \to Z$ projects space X to another space Z, as from Fig. 4.2 (A) to Fig. 4.2 (B).
- ◇ *Data separation*. The space Z in Fig. 4.2 (B) is equipped with a covering space U in Fig. 4.2 (C). Assuming that covering space U is a k-simplex spanned by a set $\{\alpha_1, \alpha_2, ..., \alpha_k\}$ so that $U = \{U_\alpha\}$, since f is continuous, $f^{-1}(U_\alpha)$ forms a covering of space X and could be used to represent topological space X of given data, as Fig. 4.2 (D).
- ◇ *Clustering and visualization*. The set $\{\alpha_1, \alpha_2, ..., \alpha_k\}$, as the vertices of k-simplex, are k connected components in topological space X which are separated clusters. Thus, $f^{-1}(U_\alpha)$ is achieved to represent data space X, as shown in Fig 4.2 (E).

For readers interested in more mathematical details, please refers to (Singh et al., 2007) and

Appendix B.

4.2.2.3 A high level interpretation

Topology is an art of simplification. It ignores complicated content in a data but focuses on one specific aspect. As in conventional manifold learning, modelling topological structure mainly focuses on the aspect of local properties. However, MAPPER puts an emphasis on an additional aspect which is pointed by the selected filter function.

4.2.3 MAPPER-Induced Manifold Alignment — MIMA

In SSMA, local property is derived by a κ nearest neighborhood (Wang and Mahadevan, 2011). The value of κ is reported as a choice of empirical trials (Tuia et al., 2014). Apart from the fact that no optimal strategy is available for choosing a value of κ, the κ itself is a hard threshold that would be applied to the whole data. However, MAPPER derives local structures by clustering. Especially, a modified version of MAPPER in the proposed MIMA algorithm applied spectral clustering (Von Luxburg, 2007) for deriving local structures. The spectral clustering have been proven as an optimized graph-cut that is capable of unbiased grouping (Shi and Malik, 2000) and have an eigen-gap strategy of deciding the number of clusters (Ng et al., 2002). Therefore, the local structure derived by MIMA is *a data-driven and optimized local structure*.

Besides local structures, MAPPER analyzes the data from an perspective defined by a given filter function. It provides a technical interface for cooperating *field knowledge*.

At the last, with the help of sequential overlapping data intervals, MAPPER derives *a regional-to-global topological structure*. For the complex remote sensing data, especially SAR data, MIMA derived structure is more robust to outliers than one derived by κ nearest neighbor.

In MIMA, a MAPPER derived topological structure is embedded into SSMA to accomplish the fusion of optical and SAR data. The topological structure of data source $\mathbf{X}^k$ derived by MAPPER can be represented as a $N_k \times N_k$ matrix $\mathbf{A}_m^k$, where N_k is the number of instances in $\mathbf{X}^k$, and $\mathbf{A}_m^k(i,j) = 1$, when data instances $\mathbf{x}_i^k$ and $\mathbf{x}_j^k$ are in the same cluster or in linked clusters, otherwise, $\mathbf{A}_m^k(i,j) = 0$. In MIMA, the topological matrix $\tilde{\mathbf{A}}_t$ in equation (4.2) is replaced by $\tilde{\mathbf{A}}_m$ (4.10).

$$\tilde{\mathbf{A}}_m = \begin{pmatrix} \mathbf{A}_m^1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \mathbf{A}_m^K \end{pmatrix}. \tag{4.10}$$

The detailed steps of MIMA are summarized in Algorithm 1 in which OLR, NBI, and f represent overlapping rate, number of intervals, and a selected filter function.

4.2.4 Evaluation of MIMA for classification tasks

To evaluate the performance of MIMA, it was tested with two data fusion and classification tasks. Two real world data sets were applied. They are a scene of the west part of the city Berlin for a land cover land use (LCLU) classification (Fig. 4.3), and a scene of the whole city Berlin for a LCZ classification (Fig. 4.4).

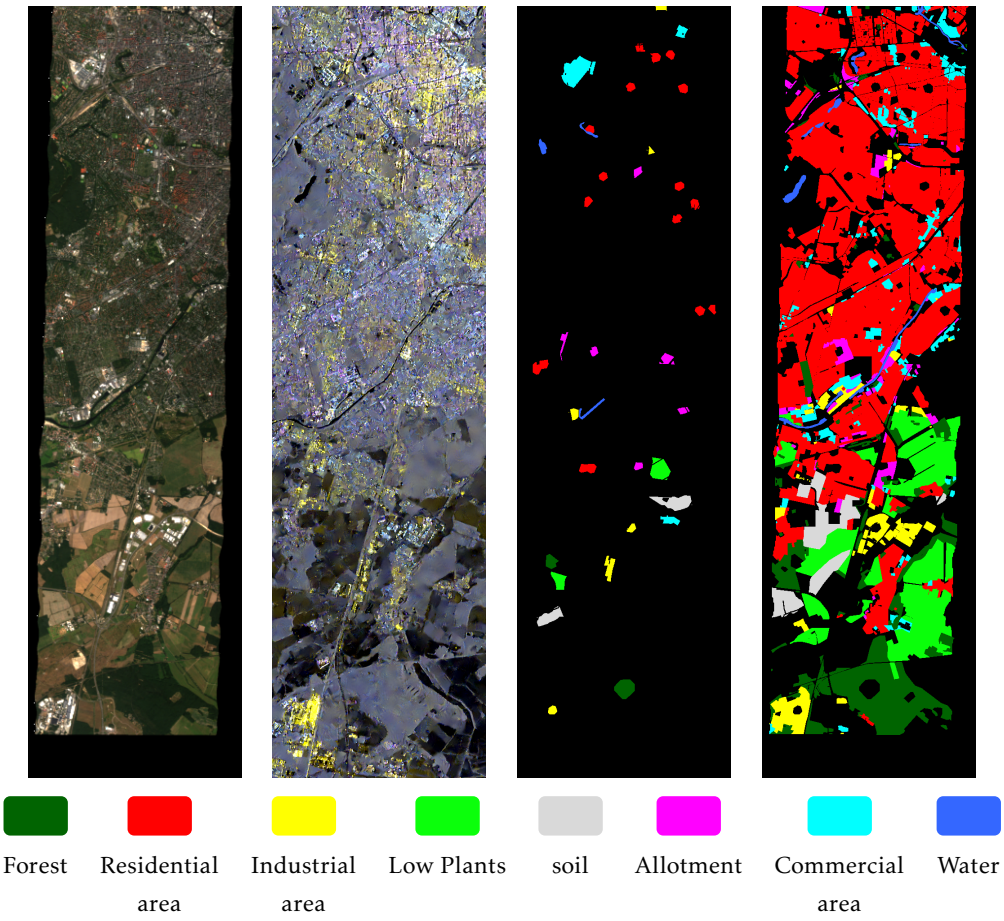


Fig. 4.3. A land cover land use classification data set of a scene in the west of city Berlin, Germany. From left to right: RGB components of simulated EnMAP data (Okujeni et al., 2016); A false RGB image of Sentinel-1 dual-Pol data; LCLU training set; LCLU testing set.

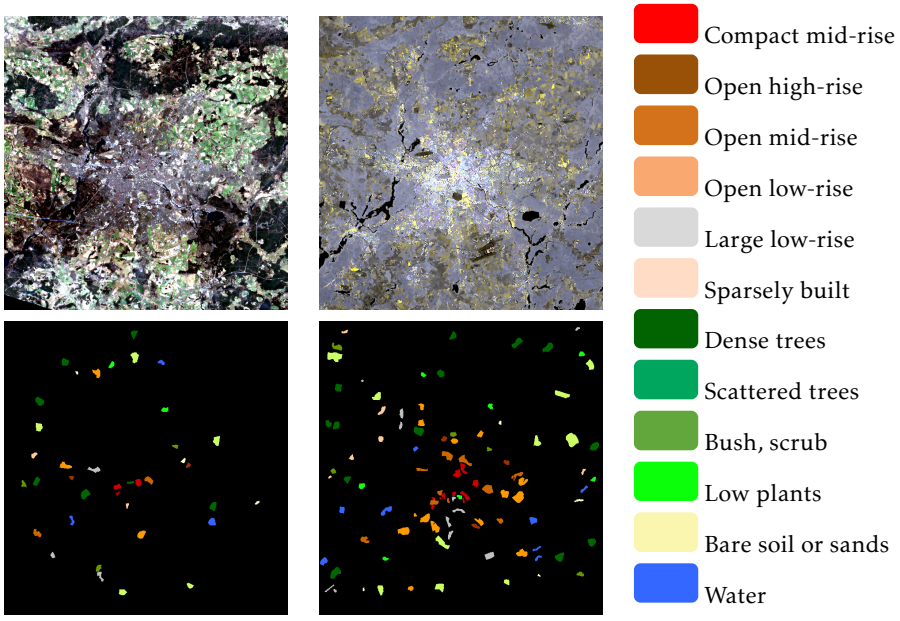


Fig. 4.4. A LCZ classification data set of a scene for the city Berlin, Germany. From left to right, first row: RGB components of LandSat-8 data; RGB component of Sentinel-1 dual-Pol data; second row: LCZ training set; LCZ testing set.

Algorithm 1 MIMA Algorithm

procedure MAPPER($\mathbf{X}^k$; OLR; NBI; f)
  finding the parameter space $\mathbf{X}^k f$
  dividing $\mathbf{X}^k f$ into NBI intervals with OLR% sequential overlapping
  dividing $\mathbf{X}$ into correspoinding intervals
  for all data in NBI intervals **do**
    Spectral clustering (Ng et al., 2002)
  end for
  $$\mathbf{A}_m^k(i,j) = \begin{cases} 1, & \text{if } \mathbf{x}_i^k \text{ and } \mathbf{x}_j^k \text{ in the same cluster;} \\ 1, & \text{if } \mathbf{x}_i^k \text{ and } \mathbf{x}_j^k \text{ in the linked clusters;} \\ 0, & \text{otherwise.} \end{cases}$$
  Return $\mathbf{A}_m^k$
end procedure
procedure MIMA($\{(\mathbf{X}^k, \mathbf{B}^k)|(k = 1, 2, ..., K)\}$; OLR; NBI; f)
  for all $k \in \{1, 2, ..., K\}$ **do**
    $\{\mathbf{X}^k, \mathbf{B}^k\} \rightarrow \tilde{\mathbf{A}}_s$ similarity matrix E.q. (4.3)
    $\{\mathbf{X}^k, \mathbf{B}^k\} \rightarrow \tilde{\mathbf{A}}_d$ dissimilarity matrix E.q. (4.4)
    MAPPER($\mathbf{X}^k$; OLR; NBI; f) $\rightarrow \tilde{\mathbf{A}}_m$ MAPPER topology E.q. (4.10)
  end for
  Constructing loss function E.q. (4.8) with $\tilde{\mathbf{A}}_s, \tilde{\mathbf{A}}_d, and \tilde{\mathbf{A}}_m$
  Optimizing projections $\{\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_K\}$ by solving E.q. (4.9)
  for all $k \in \{1, 2, ..., K\}$ **do**
    $\tilde{\mathbf{X}}^k = \mathbf{X}^k \mathbf{f}_k$
  end for
  Return $\{\tilde{\mathbf{X}}^k|(k = 1, 2, ..., K)\}$
end procedure

4.2.4.1 Algorithms for Comparison

In this evaluation, MIMA is compared to COSPACE (Hong et al., 2019a), LeMA (Hong et al., 2019b), LPP (He and Niyogi, 2004), Semi-supervised version of LPP (LPP-SE), and SSMA (Tuia et al., 2014). The concept of all these algorithms fall into the categoriy of manifold learning, yet they can be categorized into three sub-groups, (1) joint dimension reduction fusion (LPP and LPP-SE), (2) label-driven manifold alignment fusion (COSPACE and LeMA), and (3) partial label-driven manifold alignment fusion (SSMA and MIMA). The joint dimension reduction fusion stacks data as a high dimension data and applies a dimension reduction. The label-driven manifold alignment carries out data fusion based on the guidance of existed label. As a result, the label has a strong impact. The partial label-driven manifold alignment considers data structures and label information at the same time.

A quantitative comparison is demonstrated by classification performance in terms of overall classification accuracy. Three different classifiers are employed in the experiments which are κ nearest neighbor classifier (KNN), a linear support vector machine (LSVM), and a Gaussian kernel support vector machine (KSVM).

4.2.4.2 Results and Discussion

Fig. 4.5 demonstrates classification overall accuracy achieved by applying three classifiers on fused data of different algorithms. First, it is obvious that MIMA performs the best. Second, the partial label-driven manifold alignment outperforms the other two fusion strategies, it means a combination of data structures and label information is beneficial to LCLU classification.

Fig. 4.5. Classification performance in terms of overall accuracy (OA) for the experiments applied on the Berlin LCLU data set. Charts show classification results of three classifiers, from left to right, KNN, LSVM, and KSVM.

Fig. 4.6. Classification performance in terms of overall accuracy (OA) for the experiments applied on the Berlin LCZ data set. Charts show classification results of three classifiers, from left to right, KNN, LSVM, and KSVM.

Fig. 4.6 demonstrates the LCZ classification accuracy. It is obvious that LeMA performs the best and MIMA has the second best performance. LCZ is a challenging classification scheme (Hu et al., 2018; Demuzere et al., 2019) with a comparatively strong semantic content. Therefore, data structures is less helpful than label information. On this regard, LeMA has a strategy of selecting pseudo-label which have a high probability to be correctly labeled. In classification evaluation of LeMA, those correct-prone pseudo-label were also used for training classifiers. In the case of evaluatioin of LCZ task, among original classification setting of 3170 training samples and 18205 testing samples, 1231 samples out of the testing samples were given pseudo-label by LeMA and were utilized for training classifiers. It increased training samples by 38.83% and occupied 6.76% of the testing data for validation. The author believes that this change in classification setting had a big contribution to the fact that LeMA outperforms the other algorithms with a considerable margin in the experiment of LCZ data set.

To sum up, according to the evaluation, it proves that MIMA is practically powerful on efficiently combining data structures and label information for data fusion and classification.

4.3 A Comparative Review of Manifold Data Fusion Algorithms

This section is mainly a summary of Appendix C. It comparatively reviewed manifold-based data fusion techniques in terms of their performance on the fusion and classification of hyperspectral and PolSAR data. It mainly discussed the impacts of two important factors regarding manifold data fusion: fusion domains and supervision strategies. The review also included proposed algorithm MIMA in the comparison. This section can be also treated as a comprehensive evaluation of MIMA.

4.3.1 Land cover land use data sets

Two data sets consisting of hyperspectral data and dual-Pol SAR data are used for the comparison of algorithms. One of them is the Berlin LCLU data set that is shown in Fig. 4.3. The other data set is a scene for the city of Augsburg, Germany, as shown in Fig. 4.7. Both data sets consist of simulated space-borne hyperspectral data, Sentinel-1 dual-Pol data, and ground reference data of land cover land use classes.

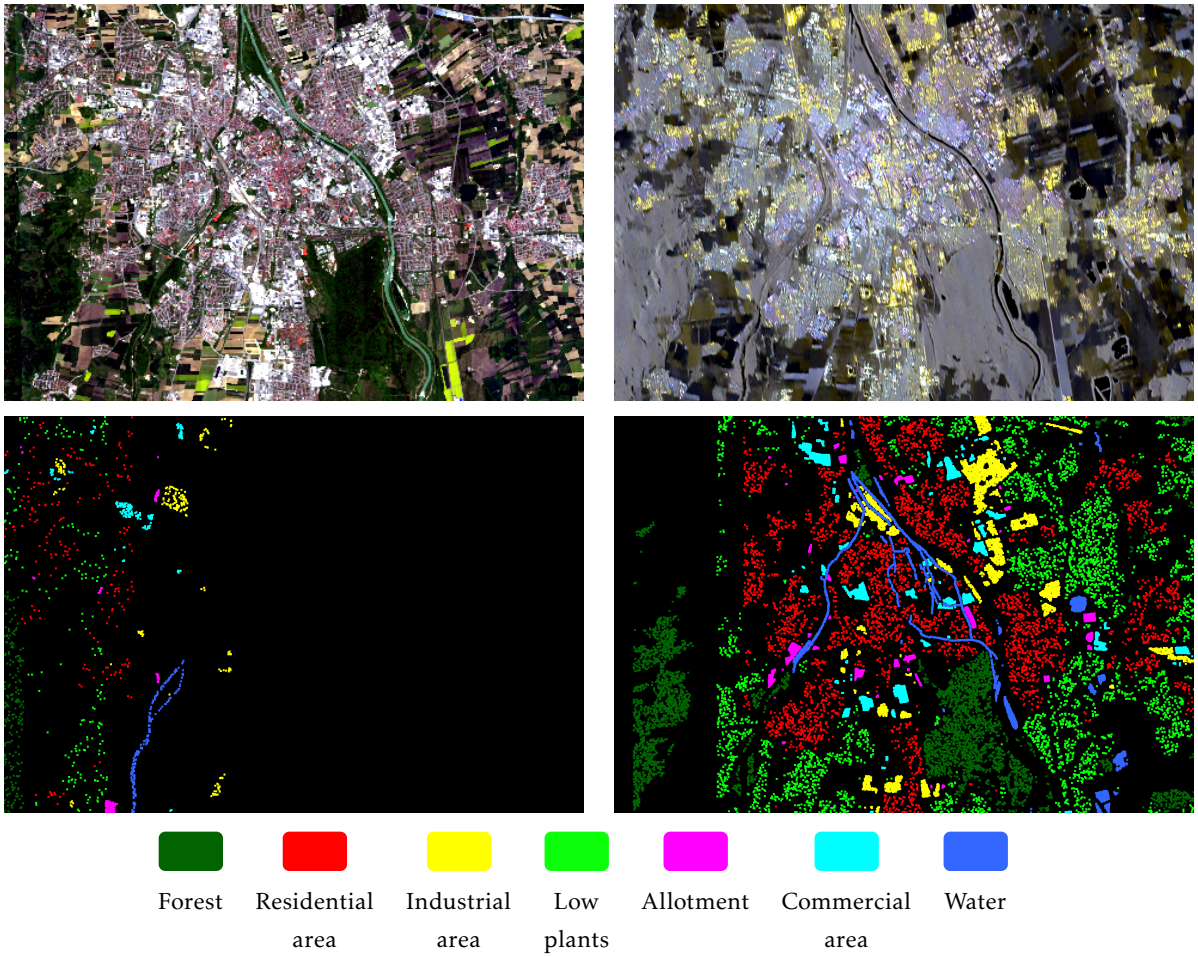


Fig. 4.7. The Augsburg LCLU data set. From left to right, top to bottom: RGB components of the hyperspectral image; Sentinel-1 dual-Pol data; The training data; The testing data.

4.3.2 Fusion domain and supervision strategy

I implemented 14 variants of four data fusion algorithms which embrace two important research questions in manifold-based fusion of hyperspectral and PolSAR data: (1) in which domain should the fusion procedure be carried out — *data domain* or *manifold domain*; and (2) how to make use of existing labeled data when formulating a graph to represent a manifold — *supervision*, *semi-supervision*, or *un-supervision*.

Table 6. Quantitative performance comparison on the Berlin data, in terms of class-specific accuracy, kappa coefficient, average accuracy, overall accuracy, and mean overall accuracy. The mean overall accuracy is calculated from five overall accuracy achieved by using different classifiers. Hyperparameter selection is achieved by grid searching. The kappa coefficient, average accuracy, and the overall accuracy that larger than 0.66, 65%, and 79% are marked in bold. And the three highest mean overall accuracy are also marked in bold. The SU, SE, and UN following abbreviations of algorithms represent supervision, semi-supervision, and un-supervision, respectively. Algorithm MIMA used principal component as filter function, while MIMAD used data density as filter function.

Algorithm	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Soil	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
HSI+POL	*1NN*	64.83	69.7	32.89	65.27	83.81	54.77	34.59	63.51	0.4975	58.67	65.44	73.73
	LSVM	66.57	86.24	30.48	75.3	79.61	53.52	40.12	76.11	0.6329	63.49	76.93	
	KSVM	67.27	80.93	41.78	64.02	72.37	57.58	33	74.6	0.5764	61.44	72.36	
	RF	63.46	84.99	37.79	74.38	82.72	56.26	40.61	82.09	0.6266	**65.29**	76.26	
	CCF	71.51	86.27	34.05	72.03	83.24	56.3	44.33	77.7	0.6445	**65.68**	77.67	
LPP_UN	*1NN*	69.53	69.07	34.56	66.09	80.27	57.51	32.18	64.56	0.5009	59.22	65.65	74.18
	LSVM	70.1	87.05	32.52	70.97	79.26	58.88	36.48	72.61	0.6354	63.48	77.27	
	KSVM	71.19	85.77	41.43	70.95	82.36	53.97	30.77	72.68	0.6297	63.64	76.69	
	RF	56.2	85.87	28.9	69.28	76	49.9	38.64	67.07	0.5874	58.98	74.25	
	CCF	68.41	86.68	34.35	71.96	80.07	54.07	37.54	75.93	0.6325	63.63	77.04	
LPP_SU	*1NN*	63.86	67.04	34.79	71.42	79.06	54.39	28.17	72.32	0.4817	58.88	64.25	71.26
	LSVM	64.41	81.51	34.12	70.1	81.56	56.74	29.1	71.38	0.578	61.11	72.9	
	KSVM	67.06	81.6	43.96	72.17	82.34	57.81	25.04	69.69	0.5908	62.46	73.77	
	RF	64.71	80.89	30.98	65.55	72.26	55.27	32.9	69.36	0.5596	58.99	71.67	
	CCF	64.25	81.99	33.72	74.47	75.59	55.89	33.77	69.76	0.5883	61.18	73.7	
LPP_SE	*1NN*	68.22	72.17	38.92	73.21	73.43	58.09	30.65	74.02	0.5327	61.09	68.26	73.52
	LSVM	64.68	85.37	38.15	74.36	79.63	59.18	29.75	77.41	0.6194	63.57	76.04	
	KSVM	69.02	81.93	41.67	70.74	77	59.76	30.77	76.17	0.6001	63.38	74.15	
	RF	66.96	83.15	29.66	72.12	66.45	56.39	34.17	74	0.5919	60.36	74.03	
	CCF	64.86	85.09	34.63	71.85	66.83	56.05	34.33	75.05	0.6044	61.09	75.12	
GGF_UN	*1NN*	69.28	71.37	36.65	66.54	83.51	56.94	31.34	63.82	0.5186	59.93	67.17	75.31
	LSVM	68.11	88.76	34.14	76.11	79.29	54.93	36.54	75.14	0.655	64.13	78.7	
	KSVM	72.18	84.64	37.08	70.29	81.88	57.25	34.49	74.44	0.6254	64.03	76.15	
	RF	68.97	86.55	29.13	70.39	81.23	49.45	41.85	62.88	0.6242	61.31	76.58	
	CCF	70.53	87.51	31.29	76.34	70.86	51.95	42.06	67.95	0.6448	62.31	77.98	
GGF_SU	*1NN*	65.57	69.99	37.73	68.89	80.13	51.96	28.71	76.62	0.5013	59.95	66.05	71.59
	LSVM	63.6	82.87	36.49	69.8	82.34	56.58	29.62	76.22	0.5906	62.19	73.77	
	KSVM	69.99	80.63	46.43	60.43	77.21	53.92	25.15	78.77	0.5695	61.57	71.98	
	RF	62.01	81.42	32.09	67.3	74.08	53.3	38.83	65.17	0.5678	59.28	72.17	
	CCF	65.54	83.4	31.38	70.58	72.26	51.15	37.02	68.24	0.5906	59.95	74	
GGF_SE	*1NN*	66.96	70.63	36.07	69.65	80.62	55.65	29.49	76.35	0.5119	60.68	66.77	72.40
	LSVM	63.06	83.52	37.69	73.01	81.94	55.48	29.11	79.87	0.6007	62.96	74.54	
	KSVM	70.19	82.26	41.52	67.92	80.35	54.38	31.22	82.51	0.5988	63.79	74.19	
	RF	65.27	80.56	34.49	67.01	75.85	54.57	38.72	66.98	0.5716	60.43	72.21	
	CCF	60.15	83.94	35.07	74.18	74.3	51.22	35.51	68.64	0.5942	60.37	74.29	
SSMA_SE	*1NN*	69.83	73.8	38	75.68	69.64	60.09	29.41	72.27	0.5474	61.09	69.54	**76.40**
	LSVM	65.49	86.97	37.63	79.08	80.06	55.63	34.46	73.37	0.6445	64.09	77.77	
	KSVM	69.38	85.81	37.49	78.3	80.54	55.42	33.29	73.21	0.6405	64.18	77.39	
	RF	64.5	90.08	30.25	77.68	65.58	49.41	36.85	67.95	0.644	60.29	78.45	
	CCF	66.66	89.12	33.05	79.51	68.95	54.91	39.47	71.01	0.6557	62.84	78.89	
SSMA_UN	*1NN*	68.46	69.61	32.87	72.87	78.51	54.88	34.76	67.95	0.5159	59.99	66.68	75.13
	LSVM	66.86	87.58	35.97	77.55	78.59	55.44	36.3	76.15	0.649	64.3	78.1	
	KSVM	70.55	85.61	36.23	74.18	79.83	57.57	35.55	73.14	0.6346	64.08	76.97	
	RF	58.91	87.37	26.35	69.77	80.7	53.14	41.94	60.71	0.6079	59.86	75.74	
	CCF	67	88.05	33.05	74.11	81.85	55	41.91	70.52	0.6467	63.94	78.14	
SSMA_SU	*1NN*	69.88	71.34	34.87	68.69	71.01	57.88	32.38	73.52	0.5199	59.94	67.21	75
	LSVM	67.56	86.73	38.76	79.67	77.21	56.87	32.27	75.45	0.6457	64.31	77.85	
	KSVM	71.6	83.96	35.72	75.92	61.57	59.59	37.1	72.65	0.6204	62.26	75.84	
	RF	60.53	87.82	33.22	77.13	70.16	52.42	38.82	63.66	0.6242	60.47	76.94	
	CCF	64.09	88.37	30.57	76.73	62.56	51.86	36.99	59.9	0.6257	58.89	77.14	

Table 6. *Cont.*

Algorithm	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Soil	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
MIMA	*1NN*	69.91	70.2	33.39	69.63	61.94	53.49	35.07	68.62	0.5055	57.78	66.26	**76.22**
	LSVM	67.76	84.97	36.22	78.36	79.08	57.74	38	70.25	0.6328	64.05	76.85	
	KSVM	71.06	84.24	41.01	76.11	69.87	55.82	32.97	68.97	0.6233	62.51	76.11	
	RF	65.1	90.31	32.54	80	82.77	50.79	35.08	71.01	**0.6642**	63.45	**79.6**	
	CCF	70.86	88.06	36.54	80.42	76.88	57.21	39.61	73.21	**0.667**	**65.35**	**79.36**	
MIMA_UN	*1NN*	72.57	68.39	35.96	70.18	79.27	62.58	30.73	67.41	0.513	60.89	66.25	75.85
	LSVM	68.21	88.59	36.62	74.6	80.79	55.87	29.86	76.08	0.6495	63.83	78.29	
	KSVM	71.78	87.1	36.85	73.13	82.31	58.05	31.79	73.14	0.6449	64.27	77.81	
	RF	67.92	88.44	27.36	77.22	81.32	50.9	35	61.04	0.6417	61.15	78.08	
	CCF	71.06	88.19	29.72	77.55	79.81	55.71	39.99	69.67	0.658	63.96	78.86	
MIMA-D	*1NN*	71.31	72.3	35.31	74.51	76.66	57.37	33.48	71.84	0.5423	61.6	68.92	**76.75**
	LSVM	67.59	86.85	36.8	81.07	78.3	56.4	38.97	75.88	0.6549	**65.23**	78.38	
	KSVM	70.01	85.33	36.79	78.84	78.52	56.83	36.44	76.08	0.6425	64.86	77.37	
	RF	67.02	89.85	33.09	80.46	83.21	50.61	37.95	74.27	**0.6698**	64.56	**79.81**	
	CCF	68.91	89.18	34.79	78.63	75.48	51.74	39.85	69.45	**0.6628**	63.5	**79.28**	
MIMA-D_UN	*1NN*	72.57	68.39	35.96	70.18	79.27	62.58	30.73	67.41	0.513	60.89	66.25	75.52
	LSVM	68.21	88.59	36.62	74.6	80.79	55.87	29.86	76.08	0.6495	63.83	78.29	
	KSVM	71.78	87.1	36.85	73.13	82.31	58.05	31.79	73.14	0.6449	64.27	77.81	
	RF	67.92	88.44	27.36	77.22	81.32	50.9	35	61.04	0.6417	61.15	78.08	
	CCF	71.06	88.19	29.72	77.55	79.81	55.71	39.99	69.67	0.658	63.96	78.86	

Table 7. Quantitative performance comparison on the Augsburg data, in terms of class-specific accuracy, kappa coefficient, average accuracy, overall accuracy, and mean overall accuracy. The mean overall accuracy is calculated from five overall accuracy achieved by using different classifiers. Hyperparameter selection is achieved by grid searching. The kappa coefficient, average accuracy, and the overall accuracy that larger than 0.56, 62.5%, and 62.5% are marked in bold. And the three highest mean overall accuracy are also marked in bold. The SU, SE, and UN following abbreviations of algorithms represent supervision, semi-supervision, and un-supervision, respectively. Algorithm MIMA used principal component as filter function, while MIMAD used data density as filter function.

Algorithm	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
HSI+POL	*1NN*	34.76	58.17	55.93	84.56	57.73	34.9	54.88	0.4682	54.42	54.42	56.71
	LSVM	31	65.95	73.29	83.85	36.9	25.07	42.85	0.4315	51.28	51.28	
	KSVM	40.59	67.83	67.07	92.59	45.24	27.1	55.78	0.4937	56.6	56.6	
	RF	61.27	73.88	70.1	94.98	47.51	25.63	59.17	0.5542	61.79	61.79	
	CCF	46.07	75.63	78.05	95.51	58.07	18.49	44.22	0.5267	59.44	59.44	
LPP_UN	*1NN*	44.9	60.61	53.29	86.56	61.37	34.76	56.32	0.4963	56.83	56.83	57.42
	LSVM	28.17	64.93	76.63	81.54	38.27	17.88	53.93	0.4356	51.62	51.62	
	KSVM	40.98	67.98	73.49	92.32	45.49	22.68	53.66	0.4943	56.66	56.66	
	RF	73.66	66.15	65.8	89.54	51.24	25.78	55.17	0.5456	61.05	61.05	
	CCF	59.63	70.71	72.8	92.2	51.9	22.78	56.51	0.5442	60.93	60.93	
LPP_SU	*1NN*	31.93	55.83	56.95	78.51	49.07	33.98	42.76	0.415	49.86	49.86	52.97
	LSVM	40.85	63.1	63.29	87.46	49.17	32.61	36.05	0.4542	53.22	53.22	
	KSVM	54.24	63.93	66.32	87.2	45.05	28.49	29.41	0.4577	53.52	53.52	
	RF	44.46	60.93	62.78	90.07	44.15	30.95	41.88	0.4587	53.6	53.6	
	CCF	52.07	62.15	64.66	90.17	44.24	28.88	40.51	0.4711	54.67	54.67	
LPP_SE	*1NN*	49.76	59.15	53	85.05	60.98	40.05	55.15	0.5052	57.59	57.59	56.06
	LSVM	43.49	65.51	77.22	85.07	40.76	20.8	41.05	0.4565	53.41	53.41	
	KSVM	37.66	71.27	75.22	93.22	48.44	20.54	45.49	0.4864	55.98	55.98	
	RF	27.17	63.22	72.2	91.78	54.46	26.54	55.66	0.485	55.86	55.86	
	CCF	47.2	66.46	73.22	90.93	56.07	23.27	45.17	0.5039	57.47	57.47	

Table 7. *Cont.*

Algorithm	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
GGF_UN	*1NN*	41.37	57.22	49.68	82.63	61.61	38.2	56.32	0.4784	55.29	55.29	55.81
	LSVM	29.17	63.76	74.83	82.12	36.54	19.71	56.71	0.438	51.83	51.83	
	KSVM	34.51	69.22	73.71	92.34	45.32	23.9	59.61	0.4977	56.94	56.94	
	RF	60.22	65.61	61.29	89.73	46.46	31.78	56.56	0.5194	58.81	58.81	
	CCF	47.9	70.9	72.22	92.44	43.34	23.05	55	0.5081	57.84	57.84	
GGF_SU	*1NN*	31.93	55.83	56.95	78.51	49.07	33.98	42.76	0.415	49.86	49.86	53.36
	LSVM	40.85	63.15	63.29	87.46	49.2	32.61	36.05	0.4543	53.23	53.23	
	KSVM	51.17	64.07	65.46	86.78	44.37	30.9	31.49	0.4571	53.46	53.46	
	RF	44.93	61.05	60.15	89.93	42.88	32.68	45.02	0.4611	53.8	53.8	
	CCF	51.2	62.61	66.93	90.24	46.78	28.76	48.59	0.4918	56.44	56.44	
GGF_SE	*1NN*	44.8	58.17	63.32	84.54	56.05	32.95	46.83	0.4778	55.24	55.24	56.19
	LSVM	53.02	66.54	66.95	84.61	47.27	29.41	31.88	0.4661	54.24	54.24	
	KSVM	67.54	68.24	66.8	87.12	41.32	23.98	24.41	0.4657	54.2	54.2	
	RF	42.88	64.9	68.07	92.56	56.68	26.63	56.54	0.5138	58.32	58.32	
	CCF	47	65.83	67.88	92.51	57.29	27.07	55.15	0.5212	58.96	58.96	
SSMA_SE	*1NN*	30.88	58.68	61.39	82.05	77.27	27.78	54.02	0.4868	56.01	56.01	**57.52**
	LSVM	26.22	66.63	78.2	72.44	42.9	16.1	55.27	0.4296	51.11	51.11	
	KSVM	31.44	69.54	78.8	93	59.05	17.73	53.76	0.5055	57.62	57.62	
	RF	75.34	72.15	64.66	91.61	48.88	30.12	43.24	0.5433	60.86	60.86	
	CCF	65.85	73.24	72.61	93.61	55	23.8	50.05	0.557	62.02	62.02	
SSMA_UN	*1NN*	31.61	56.85	57.29	80.71	73.98	26.61	54.83	0.4698	54.55	54.55	56.54
	LSVM	26.51	67.12	76.8	73.78	41.07	15.71	55.83	0.428	50.98	50.98	
	KSVM	32.56	68.2	74	89.29	58.63	18.88	55.32	0.4948	56.7	56.7	
	RF	75.15	67.93	63.17	87.93	44.29	31	50.83	0.5338	60.04	60.04	
	CCF	75.27	69.07	60.95	89.83	50.07	32.41	45.56	0.5386	60.45	60.45	
SSMA_SU	*1NN*	26.71	52.78	61.15	80.22	69.93	26.07	54.46	0.4522	53.05	53.05	54.53
	LSVM	25.2	57.2	77.56	70.29	36.85	16.68	53.76	0.3959	48.22	48.22	
	KSVM	28.68	60.68	74.83	87.9	56.2	17.46	50.39	0.4602	53.74	53.74	
	RF	49.76	67.1	67.12	91.9	47.27	28.85	54.32	0.5105	58.05	58.05	
	CCF	64.07	69.12	66.78	92.41	52.39	27.63	44.63	0.5284	59.58	59.58	
MIMA	*1NN*	27.68	57.07	62.56	81.39	72.17	26.46	55.51	0.4714	54.69	54.69	**58.01**
	LSVM	23.61	71.93	78.63	79.98	44.29	13.76	54.51	0.4445	52.39	52.39	
	KSVM	34.15	68.12	72.9	92.27	53.51	22.07	59.34	0.5039	57.48	57.48	
	RF	66.22	76.88	65.51	92.8	47.78	26.27	59.02	0.5575	62.07	62.07	
	CCF	76.78	77.49	65.12	92.73	50	28.78	53.15	**0.5734**	**63.44**	**63.44**	
MIMA_UN	*1NN*	34.34	55.24	54.85	80.76	71.41	28.44	53.39	0.4641	54.06	54.06	56.56
	LSVM	28.39	67.78	76.05	74.17	40.73	19.22	53.02	0.4323	51.34	51.34	
	KSVM	31.9	68.12	74.78	90.78	59.44	20.1	57.9	0.505	57.57	57.57	
	RF	58.95	66.54	71.76	89.68	51.68	25.24	42.68	0.5109	58.08	58.08	
	CCF	83.71	67.61	68.07	89.73	57.98	27.68	37.37	0.5536	61.74	61.74	
MIMA-D	*1NN*	28.76	57.63	62.68	80.22	74.83	24.98	55.49	0.4743	54.94	54.94	56.5
	LSVM	25.27	67.44	78.46	73.29	39.85	15.41	52.17	0.4198	50.27	50.27	
	KSVM	33.12	68.95	70.41	92.8	55.8	20.05	60.59	0.5029	57.39	57.39	
	RF	52.54	72.27	73.56	92.24	49.83	24.29	47.71	0.5207	58.92	58.92	
	CCF	55.51	73.66	72.98	92.85	53.61	24.05	54.22	0.5448	60.98	60.98	
MIMA-D_UN	*1NN*	34.93	56.02	56.54	80.46	75.29	26.39	54.41	0.4734	54.86	54.86	**60.29**
	LSVM	87.22	55.24	48	57.41	36.2	41.95	54.41	0.4674	54.35	54.35	
	KSVM	35.27	67.93	77.54	91.83	66.56	17.37	54.73	0.5187	58.75	58.75	
	RF	82.95	65.17	58.12	88.41	54.32	34.05	56.63	**0.5661**	**62.81**	**62.81**	
	CCF	78.54	72.29	65.85	92.63	49.88	26.68	53.56	**0.5657**	**62.78**	**62.78**	

4.3.2.1 Fusion domain

Manifold data fusion algorithms that are chosen in this work are LPP (He and Niyogi, 2004), generalized graph fusion (GGF) (Liao et al., 2014), SSMA (Wang and Mahadevan, 2011; Tuia et al., 2014), and MIMA (Hu et al., 2019). The LPP and GGF concatenate different data as one high dimensional data and execute a dimension reduction. Essentially, they assume an underlying manifold existing in the concatenated data. Therefore, the fusion procedure is carried out in *data domain*. On the other hand, the SSMA and MIMA assume each data set having an underlying manifold. They first extract a manifold from each of the data sets, and fuses the derived manifolds. The fusion procedure is carried out in *manifold domain*.

4.3.2.2 Supervision strategy

When approximating a manifold underlying a data, one can model it by only using labels (*supervision*), only using data structures (*unsupervision*), or jointly using labels and data structures (*semi-supervision*). The implemented 14 variants of the four fusion algorithms enables discussing the impacts of the three supervision strategies.

4.3.3 Result and discussion

In the experiments, all algorithms were tested on the two data sets. Five different classifiers were employed for the LCLU classification. They are nearest neighbor classifier (Friedman et al., 1976), linear SVM (LSVM) (Cristianini et al., 2000; Scholkopf and Smola, 2001), Gaussian kernel SVM (KSVM) (Cristianini et al., 2000; Scholkopf and Smola, 2001), random forest (RF) (Breiman, 2001), and canonical correlation forest (CCF) (Rainforth and Wood, 2015).

Regarding fusion domain, as mean OAs demonstrated in Table. 6 and Table. 7, manifold domain fusion (SSMA and MIMA) generally outperform data domain fusion (LPP and GGF). It refers that manifold domain fusion is more suitable for the fusion of hyperspectral and PolSAR data. The reason might be that, for this data fusion, it is more reasonable to fuse manifolds of data sets than assuming the concatenated data has an underlying manifold. Due to the severe difference of the two data, a joint manifold might be too complicated to approximate. Therefore, fusing hyperspectral and PolSAR data at an early stage is not an appropriated fusion strategy.

Regarding supervision strategy, according to mean OAs in Table. 6 and Table. 7, impact of supervision strategy has an obvious pattern. For the fusion in data domain (LPP and GGF), by sorting in a decreasing order in terms of classification performance, the supervision strategies have an sequence of un-supervision, semi-supervision, and supervision. For manifold domain fusion, by the same sorting, the sequence is semi-supervision, unsupervision, and supervision. The discussion mainly focus on the case of manifold domain fusion since it is more reasonable for the fusion of the dual-Pol SAR data and hyperspectral data. The semi-supervision is better because it utilizes both label information and data structures. Implicitly, the label information is able to propagate to unlabeled data via the local connections in derived data structures.

For conclusion, data domain fusion has a poorer performance than manifold domain fusion, and it is not able to efficiently cooperate label information and data structure. It might be too complicated for a data domain fusion model to deal with gaps among hyperspectral data, PolSAR data, and label information simultaneously. On the other hand, manifold domain fusion is able to efficiently cooperate information from both data sources and label information at the same time.

4.4 An Automatic Classification System based on MIMA

This section summarizes parts of contents in Appendix D and E. An automatic classification system is developed based on a modified version of MIMA. And an ensemble strategy is deployed in the system for solving the practical issue of hyper-parameter tuning. The performance of the system is evaluated over three cities, Munich, Mumbai, and New York. Having confirmed its superior performance, the system was deployed on a task of LCZ classification for 42 urban areas around the world.

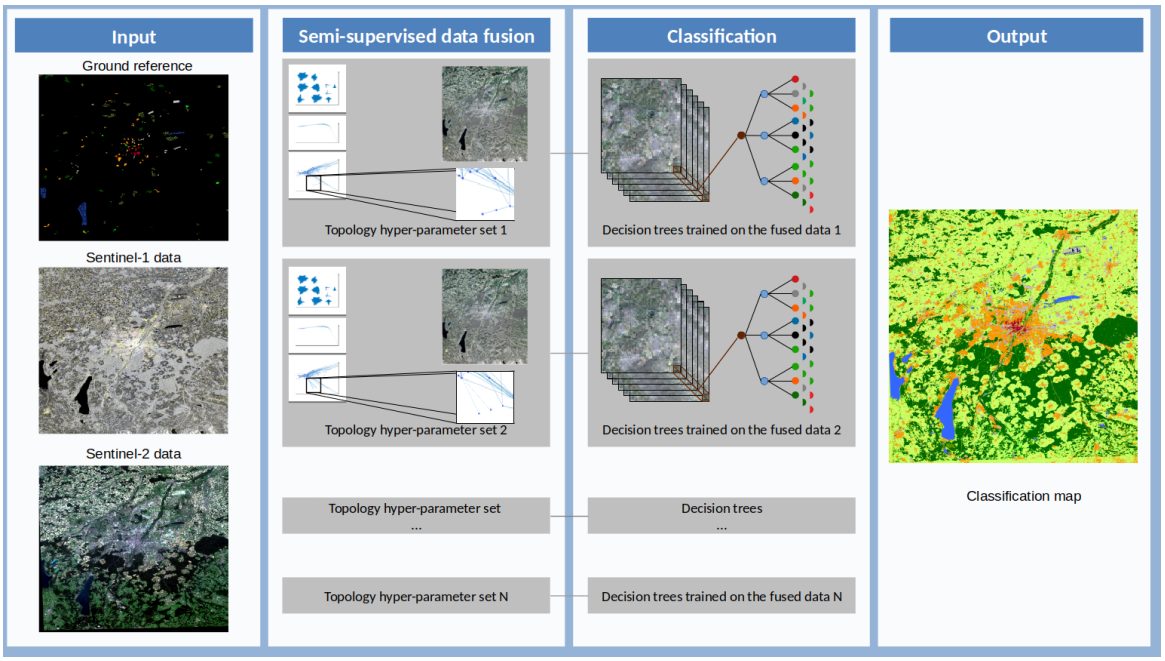


Fig. 4.8. An automatic classification system developed based on MIMA

4.4.1 Classification system

The workflow of the designed classification system is shown in Fig. 4.8. It takes Sentinel-1 data, Sentinel-2 data, and ground reference data as input, fuses data by MIMA, and produces classification maps for urban areas by an ensemble classification scheme.

4.4.1.1 A modified MIMA

To adapt MIMA for fusing Sentinel-1 and Sentinel-2 data and accomplishing LCZ classification, one of the three objectives in MIMA was eliminated, namely "pushing data of different labels apart". Because it is found that MIMA empirically has a better performance on the task without this objective. Optimizing the three objectives is over-modeling for the LCZ classification with Sentinel-1 and Sentinel-2 data. The solution of the modified MIMA boils down from generalized eigenvalue decomposition of Eq. (4.9)) to of Eq. (4.11).

$$\mathbf{Z}(\mathbf{L}_t + \mathbf{L}_s)\mathbf{Z}^T\mathbf{f} = \lambda\mathbf{Z}(\mathbf{D}_t + \mathbf{D}_s)\mathbf{Z}^T\mathbf{f}, \tag{4.11}$$

4.4.1.2 A Topological Enhanced Ensemble Classification System (TEEC)

The random forest is chosen as the classifier of this classification system. The reasons of using it are: (1) it has a state-of-the-art performance; (2) it doesn't require parameter tuning; (3) its high computational efficiency, and (4) it has been proven efficient in both large scale

classification tasks (Hu et al., 2018; Li et al., 2017a) and local climate zone classification tasks (Bechtel et al., 2015; Yokoya et al., 2017; Demuzere et al., 2019).

Random forest aims at improving classification performance by combining multiple decision tree classifiers. It predicts a label for an instance by counting predictions of a number of decision trees. In avoid all decision trees provide identical predictions, random subset of features and random subset of training samples are utilized to train each of the decision trees. Therefore, every single decision tree considers slightly different information for the classification task and predicts different labels for an instance. This configuration creates a diversity that empowers random forest. The diversity makes the random forest statistically follow the *Law of Large Numbers* so that it has robust performance, avoids overfitting that the decision tree suffers, while preserving the computational efficiency of the decision tree (Breiman, 2001; Elghazel et al., 2011; Rainforth and Wood, 2015).

Hyper-parameter tuning is often a troublesome issue for a system of production. As shown in Algorithm 1, there exist hyper-parameters in MIMA which do not show a pattern or a statistical significant optimum in terms of classification performance (Hu et al., 2019). The author found that varying values of the hyper-parameters in MIMA leads to achieve topological structures of slight differences, as shown in the zoomed-in parts of MIMA module Fig. 4.8. Inspired by the *Law of Large Numbers* concept in random forest, this work fuses Sentinel-1 and Sentinel-2 data with varying hyper-parameters and ensembles classification results of each fusion. Therefore, the variations of topological structures introduce a diversity on an additional dimension of data topology.

4.4.2 Evaluation of classification system

To demonstrate the performance of the classification system (TEEC), three other classification systems are chosen for comparison. They are: (1) classifying Sentinel-1 data using random forest (S1); (2) classifying Sentinel-2 data using random forest (S2); (3) classifying the concatenation of Sentinel-1 and Sentinel-2 data using random forest (CON). The three classification systems are chosen for comparison because similar strategies have appeared on large scale classification tasks in literature (Zhu et al., 2012; Ban et al., 2015; Yokoya et al., 2017; Li et al., 2017a; Lisini et al., 2017; Esch et al., 2017; Demuzere et al., 2019). For all of the four classification systems, they are all operated in an automatic way.

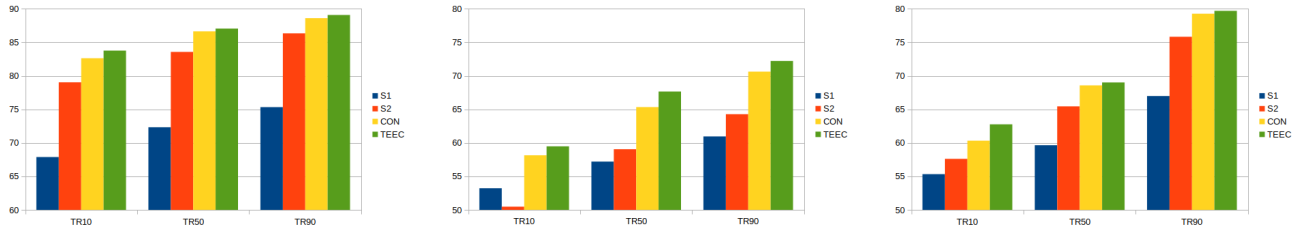


Fig. 4.9. Evaluation of four classification systems, S1, S2, CON, and TEEC. Training samples are block-wise selected. The x-axis depicts the percentages of ground reference that is used as training samples, TR10: training samples occupies 10% data of ground reference data; TR50: 50%; TR90: 90%. The y-axis denotes a mean overall accuracy of ten overall accuracy which are achieved by 10-folds cross-validation evaluation. From the left to right, charts are result of city Munich, Germany; Mumbai, India; and New York, United States.

As shown in Fig. 4.9, the proposed classification system (TEEC) outperforms the other systems in terms of the mean overall accuracy of a 10-fold cross-validation evaluation, under all scenarios of different training sets and for all three cities. More importantly, by comparing accuracy differences of systems, one can find that the proposed system is specially good at those cases where the amount of testing data is much larger than the amount of training data (TR10). This is exactly the real case in remote sensing practice. With the sophisticated evaluation strategy, the author concludes the proposed classification system is a good option among state-of-the-art techniques to accomplish the task of classifying 42 cities.

4.4.3 So2Sat LCZ42 dataset

With current developing of machine learning, accessing label data is of great importance. Although deep learning had been already studied several decades before, it exhibited its power only after the available of computational power and large labelled data set (Deng et al., 2009; Krizhevsky et al., 2012b). It is a concrete and persuasive historical event reminding researchers how important label data set is. At the time when this project started, there was no available labeled data set for the fusion of hyperspectral data and PolSAR data. It was because of limited access to hyperspectral and PolSAR data, and annotating remote sensing data is a very costly task on both expense and labor. With recent satellite missions, such as Sentinel missions of European Copernicus Program, it is more and more convenient to fetch remote sensing data. However, annotating remote sensing data is different from identifying cat and dog. It requires expert knowledge. Some cases are even difficult for experts, such as annotating SAR data.

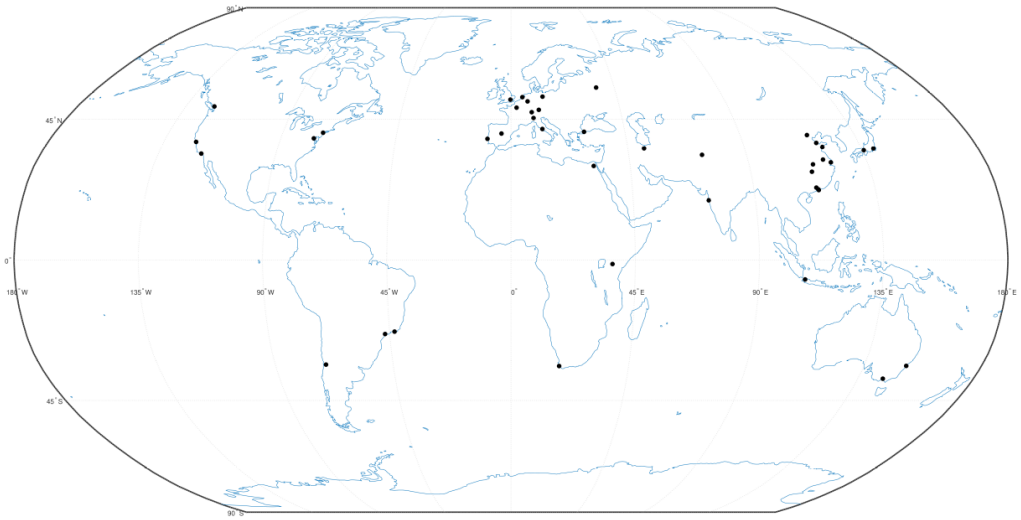


Fig. 4.10. The geographic locations of 42 cities in So2Sat LCZ42 data set.

The author participated in a collaborative work of creating an annotated remote sensing data set, named So2Sat LCZ42. It provides about half a million LCZ labeled data pairs of Sentinel-1 and Sentinel-2 data which are dual-Pol SAR and multispectral data, respectively. It includes data samples that spatially distribute over 42 major cities around the world whose distribution is shown in Fig. 4.10. The data set is described in Appendix D.

The developed classification system was deployed to produce classification maps for the 42 urban areas. The resulting classification maps of urban areas are taken as input data set for a study of analyzing urban patterns. Details about this study is summarized in the next section.

4.5 An Urban Pattern Analysis over 42 Cities

This section summarizes a study of urban pattern analysis over 42 major cities around the world. The analysis is based on LCZ classification maps produced by the classification system described in previous section and a global data of population distribution that is provided by Joint Research Center (JRC) of European Commission. This section summarizes contents in Appendix E.

4.5.1 A global population data

The population data used in this paper is a spatial raster data set depicting the distribution of global population of the year 2015 at a resolution of 250 meters (Schiavina et al., 2019). It was dis-aggregated from census or administrative population data to grid cells, constrained to the distribution of built-up that mapped in the Global Human Settlement Layer (GHSL) global layer (Freire et al., 2016).

4.5.2 Urban clustering

4.5.2.1 Feature extraction

The classification system produced classification maps for these 42 cities. These maps categorized the urban areas as compact built-up, open built-up, light weight built-up, industrial structure, or one of the seven natural types. It is important to note that all seven natural types of LCZ are grouped into one nature class in this urban pattern analysis.

As shown in Fig. 4.11, the analysis first extracts shares of the five urban land consumption types within a city. The water body was excluded here. These percentages describes how a city is formed in terms of morphological characteristics of land. On the side of population, it extracts the percentages of population that distributes on each of the five land consumption types. These population percentages describes how population distribution is related to the urban morphology. Therefore, it results in ten features to represent one city. The complete list of the features is shown in Table. 8.

Regarding the population data, we need to point out that, the state-of-the-art population distribution data is produced by census data and remote sensing human settlement layer. Population census data were evenly distributed over the human settlement layer with constraining to administrative boundaries. It results a strong correlation between the amount of population and the area of built-up types. Nevertheless, there still some interesting phenomenon could be found.

To sum up, for each city, in total ten features are extracted for upcoming clustering.

4.5.2.2 Clustering

To chose the clustering algorithm and the number of clusters, the error sum of squares (SSE) was applied as a quantitative indicator. **SSE** is the sum of squared error between each data record and the mean center of its cluster, $\mathbf{SSE} = \sum_{i=1}^{N}(\mathbf{C}_i - \tilde{\mathbf{C}})^2$, where N is the number of data instances, $\mathbf{C}_i$ is a 10 by 1 vector representing the feature of the i^{th} city, $\tilde{\mathbf{C}}$ is the mean center of the cluster that $\mathbf{C}_i$ belongs (Jain et al., 1999). When the number of clusters is fixed, the smaller value of SSE, the more concentrate of these clusters.

Regarding deciding the number of clusters, there are generally two strategies among literature. First, the number of clusters is an input parameter of a clustering algorithm. Such as K-mean clustering (Arthur and Vassilvitskii, 2007), spectral clustering (Ng et al., 2002; Von Luxburg, 2007), and agglomerating clustering. Second, the number of clusters is decided by a given criteria, such as DBSCAN (Ester et al., 1996), mean shift (Comaniciu and Meer, 2002), modified spectral clustering (Zelnik-Manor and Perona, 2005), and persistence-based clustering (Chazal et al., 2013).

For the sake of simplicity, the author chose the K-mean clustering, spectral clustering, and the agglomerating clustering as candidate algorithms. They are evaluated by the indicator **SSE** with the number of clusters given as 2 to 42. With the same clustering algorithm, the larger the number of clusters, the smaller the SSE. However, a large number of clusters

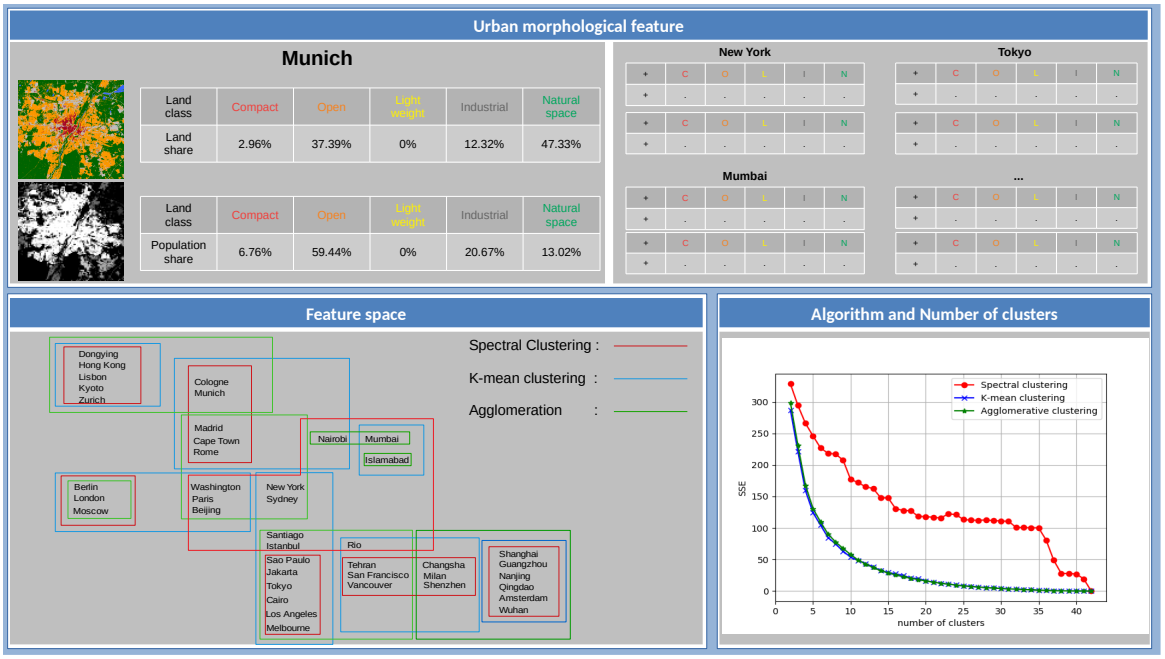


Fig. 4.11. The workflow of the clustering module. Upper block: feature extraction based on LCZ maps and population data; Bottom left: three clustering algorithms with number of clusters as input parameter are chosen as candidate clustering algorithms; Bottom right: the SSE is applied as an indicator for choosing algorithm and deciding number of clusters.

means over-clustering. A simple way of choosing the number of clusters is named as the rule of elbow (Jain et al., 1999). It infers the point where increasing the number of clusters brings un-interestingly small reduction of SSE. Depending on the experiment results shown in Fig. 4.11, the K-mean clustering has a better performance in terms of **SSE** so that it was chosen as the clustering algorithm. The number of clusters equals to 8 according to the rule of elbow.

4.5.3 Urban morphology analysis

Table 8. This table presents feature values of ten mean cluster centers for the eight morphological clusters. The unit of numbers is percentage (%).

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Share of compact area	63	10	5	31	15	50	29	13
Share of open area	8	27	35	11	55	11	10	40
Share of lightweight area	0	0	0	9	0	0	0	5
Share of industrial area	17	31	8	2	8	12	6	12
Share of natural area	12	31	52	47	22	27	55	30
Share of population in compact area	79	18	9	40	24	69	64	25
Share of population in open area	5	35	59	5	59	9	14	38
Share of population in lightweight area	0	0	0	43	0	0	0	13
Share of population in industrial area	13	36	14	3	9	12	8	16
Share of population in natural area	4	12	18	8	8	9	14	7

The first cluster concludes a group of cities which has a super compact urban morphology. This cluster has on average 63% of land having been consumed as compact built-up. And 79% of their population resides within the compact built-up area. The second cluster represents an industrial urban morphology. It includes cities whose 31% of land are occupied by industrial structure. Seven out of the total nine cities in this cluster are from China. By cross check with optical images, the industrial structures are mostly large and low buildings functioning as factories and warehouses. For the city of Amsterdam, tanks and warehouses near the harbor were also identified. The third cluster mainly includes European cites that has an

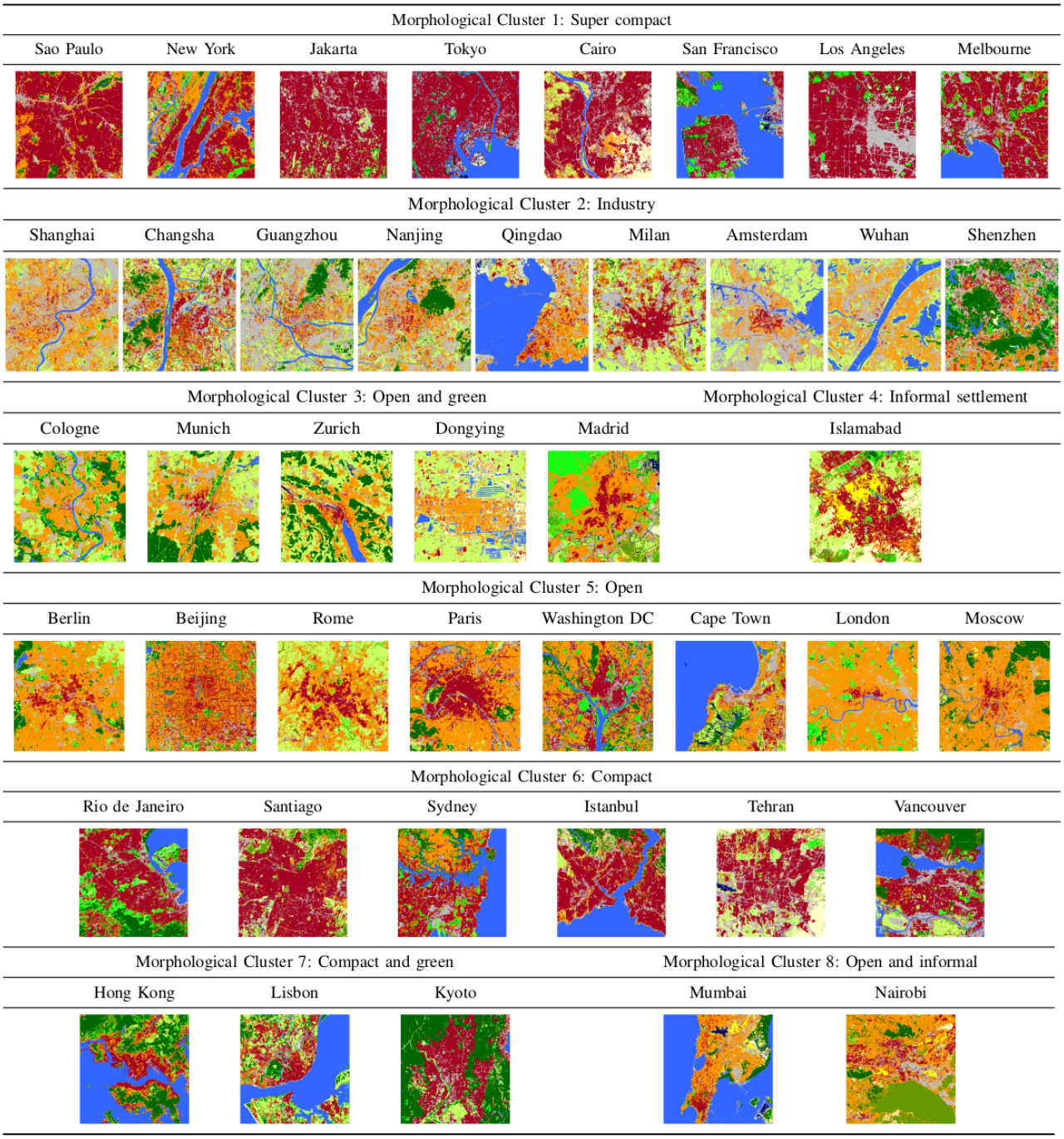


Fig. 4.12. Visualization of LCZ classification maps for 42 cities. These maps are produced by proposed classification system and organized in clusters.

open and green urban morphology. They are characterized by having more than 50% of land as natural cover and nearly 60% population is distributed in open areas. The fourth cluster has the morphological characteristics that a large portion of informal settlement exists. With the 42 cities, it includes only the city of Islamabad which has 9% of land constructed as light weight built-up and has 43% of population residing in light weight built-ups. According to our labeling experiences and visual inspection, the light weight built-up mostly reflects the distribution of informal settlements. 9% is a very high percentage among the selected 42 cities. The 9% land area of light weight build-up in Islamabad is home to 43% of the population in this city, which makes those areas heavily populated. The fifth cluster owns an open urban morphology. This cluster includes eight cities whose 55% of land were constructed into open built-ups. A very interesting point is that all those eight cities in this

cluster are the capital city to eight different countries. The sixth cluster has a compact urban morphology. It gathered cities that have high percentages of compact built-up and population distributed in there, namely 50% and 69% respectively. Comparing to the first cluster which also has a high percentage of compact area, there exists 10% shortage in the sixth cluster. And the sixth cluster has 27% of land being natural cover, which is 15% higher than it in the first cluster. The seventh cluster has a compact and green urban morphology. It has 29% and 55% of their land covered by compact built-up and natural cover, respectively. However, an average 64% of their population resides in the 29% compact built-up areas. The last cluster has mixture urban morphology of light weight built-up and open built-up. In this cluster, the open built-up occupies 40% of the land and the light weight built-ups accounts for about 5%.

To sum up, we have identified eight different urban morphological clusters among the 42 cities around the world. Those urban morphological clusters are: super compact, industry, open-green, informal settlement, open, compact, compact-green, and open-informal.

5 Conclusion and Outlook

5.1 Conclusion

This dissertation explored the fusion of optical and SAR data for urban classification tasks. It developed a data fusion algorithm for the fusion and classification of optical and SAR data, built up a classification system based on the developed fusion algorithm, and demonstrated a large scale application of produced classification maps. From developing algorithm to a large scale application demonstration, the objectives defined in Chapter 1 have been fulfilled. The most important findings of this dissertation are listed as follows:

- Regarding the fusion of optical and SAR data, it benefits the classification task according to all experiments in this dissertation. The optical data has a much better distinguishing power compared to the SAR data. The fusion of them could improve the classification performance of the optical data by a range of 2% to 9% according to experiments in this dissertation.

- Regarding the developed algorithm MIMA, its superior performance among manifold-based algorithms has been proven in comprehensive evaluations in Appendix B, C, and E. The evaluations have included the fusion of LandSat-8 and Sentinel-1 data for the LCZ classification, fusing simulated space-borne hyperspectral data and Sentinel-1 data for land cover land use classification, and fusing Sentinel-1 and Sentinel-2 data for modified LCZ classification. The evaluations have involved about twenty algorithms for comparison and have been carried out over multiple locations, including Berlin, Munich, Augsburg, Mumbai, and New York.

- Regarding data fusion levels, it is more efficient to fuse optical and SAR data at a later stage than fusing them at data level. It has been proven in this dissertation, it is more efficient to fuse optical and SAR data in the manifold domain than fusing them in the data domain. It is a clue that fusing data of severe differences as optical and SAR data should be carried out at a semantic or information level.

- Regarding supervision strategy, it is efficient to utilize semi-supervision which benefits from both data structures and label semantics at the same time. Attention should be paid on how to design the input data so that the data structure is related to label semantics. Un-supervision is very welcomed in remote sensing due to lack of annotated data sets. For the same reason, methodologies heavily rely on annotated data are very expensive to implement, but often are straightforward solutions.

- Regarding hyperparameter tuning in classification, it is a frequent practice to train an algorithm on training data, tune hyperparameters on validation data, and evaluate the trained algorithm on testing data. However, an optimal hyperparameter selection based on validation data can not grantee a good performance on testing data. For a remote sensing application, the amount of labeled data for training and validating algorithms is an extreme tiny proportion compared to data of inference. This dissertation carried out an ensemble strategy to tackle hyperparameter selection in the manifold extraction which is statistically more robust than the frequent practice.

- Regarding urban pattern analysis, this dissertation provides valuable information by identifying groups of cities based on the morphological formation. These patterns

provide fine information to stakeholders. For example, planners and decision makers of a super compact city could statistically analyze the cities of the super compact group. The resulted information are more referable than the information derived from all cities.

- Regarding urban classification, it is an invaluable resource that having a large to global scale urban map whose categorization scheme is finer than a binary mask of built-ups. For instance, GHSL based global population data has registered urban inhabitant over industrial areas. This type of inaccurate information is misleading for further applications, such as public health surveillance, per-person land consumption, and wealthy distribution analysis. It can be avoided by a detailed categorization scheme. The author believes it would also be a great resource for supporting UN SDGs.

5.2 Outlook

According to the studies of remote sensing data fusion and classification in this dissertation, a few research topics are very relevant for future studies.

5.2.1 Unsupervised data fusion and urban mapping

The amount of remote sensing data is increasing exponentially. However, the amount of annotated data is extremely rare compared to the total data amount. It is expensive to annotate remote sensing data due to the requirement of expertise. Moreover, annotated data in remote sensing has a poor capability of generalization due to many reasons. For example, annotations might change due to different data acquisition time and different tasks. On this regard, un-supervised machine learning methods are highly preferable. It will be more feasible to develop un-supervised algorithms for future missions, including urban mapping.

5.2.2 Machine learning based annotation assistant

To access annotated data in remote sensing, crowd sourced and geo-tagged data are becoming more and more popular due to its large volume. But their quality is poor. Geographic information system (GIS) data can also be used as annotated data. The annotated data carry semantic meaning defined by GIS clients which might be too complex to be detected in remote sensing data. Additionally, for applications as vegetation species classification and vegetation health surveillance, the annotated data have to be produced by experts. Instead of pure manual labeling, it would be efficient to develop a machine learning based annotation assistant. At the beginning phase, unsupervised analysis reveals a realistic definition of annotations and recommends initial labels. The expert makes the decisions based on the recommendations. With the initial labels, supervised algorithms can be trained for better recommendations. At last, the assistant should be able to evaluate the quality of annotated data. Importantly, the developed assistant should be applicable to a variety of applications.

5.2.3 Generalization ability of trained model

A trained machine learning model should be capable of performing consistently on data of different regions, different times, and even different sensors. However, the model is often not able to provide satisfying results when these impacting factor change. This is fatal when producing a remote sensing products of a large to global coverage. Therefore, the future study should improve the generalization ability of trained models. The study is related to transfer learning and domain adaptation in machine learning research.

5.2.4 Data fusion and classification system

To build a data fusion and classification system operating constantly, the system design is also a non-trivial engineering study. A sophisticated system in future is required to be able to fuse heterogeneous data sources, fuse multi-scale data, fuse multi-temporal data, handle situations that one or two data sources are missing, manage huge amount of data sets, update data product regularly, and timely distribute products. It should be designed to allow future upgrade and extension. At the same time, it should serve as an infrastructure for research projects.

References

Alparone, L., Aiazzi, B., Baronti, S., Garzelli, A., 2015. Remote Sensing Image Fusion. Crc Press.

Aravena Pelizari, P., Spröhnle, K., GeiÃŸ, C., Schoepfer, E., Plank, S., Taubenböck, H., 2018. Multi-sensor feature fusion for very high spatial resolution built-up area extraction in temporary settlements. Remote Sensing of Environment 209: 793–807.

Arthur, D., Vassilvitskii, S., 2007. k-means++: The advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 1027–1035.

Baghdadi, N. N., Hajj, M. E., Zribi, M., Fayad, I., 2016. Coupling SAR C-Band and Optical Data for Soil Moisture and Leaf Area Index Retrieval Over Irrigated Grasslands. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (3): 1229–1243.

Baltrušaitis, T., Ahuja, C., Morency, L., 2019. Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (2): 423–443.

Ban, Y., Hu, H., Rangel, I. M., 2010. Fusion of Quickbird MS and RADARSAT SAR data for urban land-cover mapping: object-based and knowledge-based approach. International Journal of Remote Sensing 31 (6): 1391–1410.

Ban, Y., Jacob, A., 2013. Object-Based Fusion of Multitemporal Multiangle ENVISAT ASAR and HJ-1b Multispectral Data for Urban Land-Cover Mapping. IEEE Transactions on Geoscience and Remote Sensing 51 (4): 1998–2006.

Ban, Y., Jacob, A., Gamba, P., 2015. Spaceborne sar data for global urban mapping at 30 m resolution using a robust urban extractor. ISPRS Journal of Photogrammetry and Remote Sensing 103: 28–37.

Baumann, M., Levers, C., Macchi, L., Bluhm, H., Waske, B., Gasparri, N. I., Kuemmerle, T., 2018. Mapping continuous fields of tree and shrub cover across the Gran Chaco using Landsat 8 and Sentinel-1 data. Remote Sensing of Environment 216: 201–211.

Bechtel, B., Alexander, P., Böhner, J., Ching, J., Conrad, O., Feddema, J., Mills, G., See, L., Stewart, I., 2015. Mapping local climate zones for a worldwide database of the form and function of cities. ISPRS International Journal of Geo-Information 4 (1): 199–219.

Bechtel, B., Daneke, C., 2012. Classification of local climate zones based on multiple earth observation data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5 (4): 1191–1202.

Bechtel, B., See, L., Mills, G., Foley, M., 2016. Classification of local climate zones using sar and multispectral data in an arid environment. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (7): 3097–3105.

Belkin, M., Niyogi, P., 2002. Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Advances in neural information processing systems, 585–591.

Bendjebbour, A., Delignon, Y., Fouque, L., Samson, V., Pieczynski, W., 2001. Multisensor image segmentation using Dempster-Shafer fusion in Markov fields context. IEEE Transactions on Geoscience and Remote Sensing 39 (8): 1789–1798.

Benediktsson, J. A., Arnason, K., Pesaresi, M., 2001. The use of morphological profiles in classification of data from urban areas. In: Remote Sensing and Data Fusion over Urban Areas, IEEE/ISPRS Joint Workshop 2001, IEEE, 30–34.

Benediktsson, J. A., Palmason, J. A., Sveinsson, J. R., 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. IEEE Transactions on Geoscience and Remote Sensing 43 (3): 480–491.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence 35 (8): 1798–1828.

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., Chanussot, J., 2013. Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and remote sensing magazine 1 (2): 6–36.

Breiman, L., 2001. Random forests. Machine learning 45 (1): 5–32.

Bruzzone, L., Prieto, D. F., Serpico, S. B., 1999. A neural-statistical approach to multitemporal and

multisource remote-sensing image classification. IEEE Transactions on Geoscience and Remote Sensing 37 (3): 1350–1359.

Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Rojo-Alvarez, J. L., Martinez-Ramon, M., 2008. Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection. IEEE Transactions on Geoscience and Remote Sensing 46 (6): 1822–1835.

Carreiras, J. M., Jones, J., Lucas, R. M., Shimabukuro, Y. E., 2017. Mapping major land cover types and retrieving the age of secondary forests in the Brazilian Amazon by combining single-date optical and radar remote sensing data. Remote Sensing of Environment 194: 16–32.

Chazal, F., Guibas, L. J., Oudot, S. Y., Skraba, P., 2013. Persistence-based clustering in riemannian manifolds. Journal of the ACM (JACM) 60 (6): 41.

Chen, B., Xiao, X., Li, X., Pan, L., Doughty, R., Ma, J., Dong, J., Qin, Y., Zhao, B., Wu, Z., Sun, R., Lan, G., Xie, G., Clinton, N., Giri, C., 2017. A mangrove forest map of China in 2015: Analysis of time series Landsat 7/8 and Sentinel-1a imagery in Google Earth Engine cloud computing platform. ISPRS Journal of Photogrammetry and Remote Sensing 131: 104–120.

Cloude, S. R., Pottier, E., 1997. An entropy based classification scheme for land applications of polarimetric sar. IEEE transactions on geoscience and remote sensing 35 (1): 68–78.

Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis & Machine Intelligence (5): 603–619.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning 20 (3): 273–297.

Cremon, d. H., De FÃ¡tima Rossetti, D., Zani, H., 2014. Classification of Vegetation over a Residual Megafan Landform in the Amazonian Lowland Based on Optical and SAR Imagery. Remote Sensing 6 (11): 10931–10946.

Cristianini, N., Shawe-Taylor, J., et al., 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.

Cumming, I. G., Wong, F. H., 2005. Digital processing of synthetic aperture radar data. Artech house 1 (3).

Demuzere, M., Bechtel, B., Mills, G., 2019. Global transferability of local climate zone models. Urban climate 27: 46–63.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, Ieee, 248–255.

Dong, J., Xiao, X., Chen, B., Torbick, N., Jin, C., Zhang, G., Biradar, C., 2013. Mapping deciduous rubber plantations through integration of PALSAR and multi-temporal Landsat imagery. Remote Sensing of Environment 134: 392–402.

Donoho, D. L., et al., 2000. High-dimensional data analysis: The curses and blessings of dimensionality. AMS Math Challenges Lecture 1: 32.

Edelsbrunner, H., 2014. A short course in computational geometry and topology. No. Mathematical methods. Springer.

Edelsbrunner, H., Harer, J., 2010. Computational topology: an introduction. American Mathematical Soc.

Edelsbrunner, H., Letscher, D., Zomorodian, A., 2000. Topological persistence and simplification. In: Proceedings 41st Annual Symposium on Foundations of Computer Science, IEEE, 454–463.

Elghazel, H., Aussem, A., Perraud, F., 2011. Trading-off diversity and accuracy for optimal ensemble tree selection in random forests. In: Ensembles in Machine Learning Applications. Springer, 169–179.

Erden, F., Cetin, A. E., 2017. Period estimation of an almost periodic signal using persistent homology with application to respiratory rate measurement. IEEE Signal Processing Letters 24 (7): 958–962.

Erinjery, J. J., Singh, M., Kent, R., 2018. Mapping and assessment of vegetation types in the tropical rainforests of the Western Ghats using multispectral Sentinel-2 and SAR Sentinel-1 satellite imagery. Remote Sensing of Environment 216: 345–354.

Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., Strano, E., 2017. Breaking new ground in mapping human settlements from space–the global urban foot-

print. ISPRS Journal of Photogrammetry and Remote Sensing 134: 30–42.
Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A., Dech, S., 2013. Urban footprint processorâ€"fully automated processing chain generating settlement masks from global data of the tandem-x mission. IEEE Geoscience and Remote Sensing Letters 10 (6): 1617–1621.
Escolano, F., Hancock, E., Lozano, M., 2011. Graph matching through entropic manifold alignment. In: CVPR 2011, IEEE, 2417–2424.
Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd, Vol. 96, 226–231.
Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J., Tilton, J. C., 2012. Advances in spectral-spatial classification of hyperspectral images. Proceedings of the IEEE 101 (3): 652–675.
Forget, Y., Shimoni, M., Gilbert, M., Linard, C., 2018. Complementarity Between Sentinel-1 and Landsat 8 Imagery for Built-Up Mapping in Sub-Saharan Africa .
Freire, S., MacManus, K., Pesaresi, M., Doxsey-Whitfield, E., Mills, J., 2016. Development of new open and free multi-temporal global population grids at 250 m resolution .
Friedman, J. H., Bentley, J. L., Finkel, R. A., 1976. An algorithm for finding best matches in logarithmic time. ACM Trans. Math. Software 3 (SLAC-PUB-1549-REV. 2): 209–226.
Gamba, P., Aldrighi, M., 2012. SAR Data Classification of Urban Areas by Means of Segmentation Techniques and Ancillary Optical Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5 (4): 1140–1148.
Geng, J., Fan, J., Wang, H., Ma, X., Li, B., Chen, F., 2015. High-resolution sar image classification via deep convolutional autoencoders. IEEE Geoscience and Remote Sensing Letters 12 (11): 2351–2355.
Gessner, U., Machwitz, M., Esch, T., Tillack, A., Naeimi, V., Kuenzer, C., Dech, S., 2015. Multi-sensor mapping of West African land cover using MODIS, ASAR and TanDEM-X/TerraSAR-X data. Remote Sensing of Environment 164: 282–297.
Ghamisi, P., Benediktsson, J. A., Ulfarsson, M. O., 2013. Spectral–spatial classification of hyperspectral images based on hidden markov random fields. IEEE Transactions on Geoscience and Remote Sensing 52 (5): 2565–2574.
Ghamisi, P., Dalla Mura, M., Benediktsson, J. A., 2014. A survey on spectral–spatial classification techniques based on attribute profiles. IEEE Transactions on Geoscience and Remote Sensing 53 (5): 2335–2353.
Ghamisi, P., Plaza, J., Chen, Y., Li, J., Plaza, A. J., 2017. Advanced spectral classifiers for hyperspectral images: A review. IEEE Geoscience and Remote Sensing Magazine 5 (1): 8–32.
Ghrist, R., 2008. Barcodes: the persistent topology of data. Bulletin of the American Mathematical Society 45 (1): 61–75.
Ghrist, R. W., 2014. Elementary applied topology. Vol. 1. Createspace Seattle.
Gómez-Chova, L., Tuia, D., Moser, G., Camps-Valls, G., 2015. Multimodal Classification of Remote Sensing Images: A Review and Future Directions. Proceedings of the IEEE 103 (9): 1560–1584.
Grohnfeldt, C. H., 2017. Multi-sensor data fusion for multi-and hyperspectral resolution enhancement based on sparse representations. Ph.D. thesis, Technische Universität München.
Gu, Y., Chanussot, J., Jia, X., Benediktsson, J. A., 2017. Multiple kernel learning for hyperspectral image classification: A review. IEEE Transactions on Geoscience and Remote Sensing 55 (11): 6547–6565.
Guo, Y., Jia, X., Paull, D., Benediktsson, J. A., 2019. Nomination-favoured opinion pool for optical-SAR-synergistic rice mapping in face of weakened flooding signals. ISPRS Journal of Photogrammetry and Remote Sensing 155: 187–205.
Haas, J., Ban, Y., 2017. Sentinel-1a SAR and sentinel-2a MSI data fusion for urban ecosystem service mapping. Remote Sensing Applications: Society and Environment 8: 41–53.
Hajnsek, I., Jagdhuber, T., Schon, H., Papathanassiou, K. P., 2009. Potential of estimating soil moisture under vegetation cover by means of polsar. IEEE Transactions on Geoscience and Remote Sensing 47 (2): 442–454.

Haralick, R. M., Shanmugam, K., et al., 1973. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics (6): 610–621.

Härdle, W., Simar, L., 2007. Applied multivariate statistical analysis. Vol. 22007. Springer.

He, G., Feng, X., Xiao, P., Xia, Z., Wang, Z., Chen, H., Li, H., Guo, J., 2017. Dry and Wet Snow Cover Mapping in Mountain Areas Using SAR and Optical Remote Sensing Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (6): 2575–2588.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778.

He, X., Cai, D., Yan, S., Zhang, H.-J., 2005. Neighborhood preserving embedding. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, Vol. 2, IEEE, 1208–1213.

He, X., Niyogi, P., 2004. Locality preserving projections. In: Advances in neural information processing systems, 153–160.

Hegarat-Mascle, S. L., Bloch, I., Vidal-Madjar, D., 1997. Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing. IEEE Transactions on Geoscience and Remote Sensing 35 (4): 1018–1031.

Hong, D., Yokoya, N., Chanussot, J., Zhu, X. X., 2019a. Cospace: Common subspace learning from hyperspectral-multispectral correspondences. IEEE Transactions on Geoscience and Remote Sensing .

Hong, D., Yokoya, N., Ge, N., Chanussot, J., Zhu, X. X., 2019b. Learnable manifold alignment (lema): A semi-supervised cross-modality learning framework for land cover and land use classification. ISPRS journal of photogrammetry and remote sensing 147: 193–205.

Hu, J., Ghamisi, P., Zhu, X., 2018. Feature extraction and selection of sentinel-1 dual-pol data for global-scale local climate zone classification. ISPRS International Journal of Geo-Information 7 (9): 379.

Hu, J., Hong, D., Wang, Y., Zhu, X. X., 2019. A comparative review of manifold learning techniques for hyperspectral and polarimetric sar image fusion. Remote Sensing 11 (6): 681.

Hu, J., Hong, D., Zhu, X. X., 2019. Mima: Mapper-induced manifold alignment for semi-supervised fusion of optical image and polarimetric sar data. IEEE Transactions on Geoscience and Remote Sensing : 1–16.

Huang, X., Zhang, L., 2013. An SVM Ensemble Approach Combining Spectral, Structural, and Semantic Features for the Classification of High-Resolution Remotely Sensed Imagery. IEEE Transactions on Geoscience and Remote Sensing 51 (1): 257–272.

Iannelli, G. C., Gamba, P., 2019. Urban Extent Extraction Combining Sentinel Data in the Optical and Microwave Range. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (7): 2209–2216.

Jain, A. K., Murty, M. N., Flynn, P. J., 1999. Data clustering: a review. ACM computing surveys (CSUR) 31 (3): 264–323.

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M. R., Kuemmerle, T., Meyfroidt, P., Mitchard, E. T. A., Reiche, J., Ryan, C. M., Waske, B., 2016. A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. Remote Sensing 8 (1): 70.

Khorram, S. (Hrsg.), 2012. Remote sensing. SpringerBriefs in space development. Springer, Berkeley, CA, oCLC: ocn812190033.

Klotz, M., Kemper, T., Geiß, C., Esch, T., Taubenböck, H., 2016. How good is the map? a multi-scale cross-comparison framework for global settlement layers: Evidence from central europe. Remote Sensing of Environment 178: 191–212.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012a. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, 1097–1105.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012b. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, 1097–1105.

Kussul, N., Lemoine, G., Gallego, F. J., Skakun, S. V., Lavreniuk, M., Shelestov, A. Y., 2016. Parcel-Based Crop Classification in Ukraine Using Landsat-8 Data and Sentinel-1a Data. IEEE Journal of

Selected Topics in Applied Earth Observations and Remote Sensing 9 (6): 2500–2508.

Lahat, D., Adali, T., Jutten, C., 2015. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. Proceedings of the IEEE 103 (9): 1449–1477.

Lee, J.-S., Pottier, E., 2009. Polarimetric radar imaging: from basics to applications. CRC press.

Lehmann, E. A., Caccetta, P., Lowell, K., Mitchell, A., Zhou, Z.-S., Held, A., Milne, T., Tapley, I., 2015. SAR and optical remote sensing: Assessment of complementarity and interoperability in the context of a large-scale operational forest monitoring system. Remote Sensing of Environment 156: 335–348.

Lehmann, E. A., Caccetta, P. A., Zhou, Z., McNeill, S. J., Wu, X., Mitchell, A. L., 2012. Joint Processing of Landsat and ALOS-PALSAR Data for Forest Mapping and Monitoring. IEEE Transactions on Geoscience and Remote Sensing 50 (1): 55–67.

Li, C., Gong, P., Wang, J., Zhu, Z., Biging, G. S., Yuan, C., Hu, T., Zhang, H., Wang, Q., Li, X., et al., 2017a. The first all-season sample set for mapping global land cover with landsat-8 data. Science Bulletin 62 (7): 508–515.

Li, L., Cheng, W.-Y., Glicksberg, B. S., Gottesman, O., Tamler, R., Chen, R., Bottinger, E. P., Dudley, J. T., 2015. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Science translational medicine 7 (311): 311ra174–311ra174.

Li, S., Wang, Y., Chen, P., Xu, X., Cheng, C., Chen, B., 2017b. Spatiotemporal Fuzzy Clustering Strategy for Urban Expansion Monitoring Based on Time Series of Pixel-Level Optical and SAR Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (5): 1769–1779.

Liao, D., Qian, Y., Zhou, J., Tang, Y. Y., 2016. A manifold alignment approach for hyperspectral image visualization with natural color. IEEE Transactions on Geoscience and Remote Sensing 54 (6): 3151–3162.

Liao, W., Pižurica, A., Bellens, R., Gautama, S., Philips, W., 2014. Generalized graph-based fusion of hyperspectral and lidar data using morphological features. IEEE Geoscience and Remote Sensing Letters 12 (3): 552–556.

Lin, T., Zha, H., 2008. Riemannian manifold learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (5): 796–809.

Lin, Y., Zhang, H., Li, G., Wang, T., Wan, L., Lin, H., 2019. Improving Impervious Surface Extraction With Shadow-Based Sparse Representation From Optical, SAR, and LiDAR Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (7): 2417–2428.

Lisini, G., Salentinig, A., Du, P., Gamba, P., 2017. Sar-based urban extents extraction: from envisat to sentinel-1. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (8): 2683–2691.

Liu, Z., Li, G., Mercier, G., He, Y., Pan, Q., 2018. Change Detection in Heterogenous Remote Sensing Images via Homogeneous Pixel Transformation. IEEE Transactions on Image Processing 27 (4): 1822–1834.

Lucas, R. M., Clewley, D., Accad, A., Butler, D., Armston, J., Bowen, M., Bunting, P., Carreiras, J., Dwyer, J., Eyre, T., Kelly, A., McAlpine, C., Pollock, S., Seabrook, L., 2014. Mapping forest growth and degradation stage in the Brigalow Belt Bioregion of Australia through integration of ALOS PALSAR and Landsat-derived foliage projective cover data. Remote Sensing of Environment 155: 42–57.

Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., Carlsson, G., 2013. Extracting insights from the shape of complex data using topology. Scientific reports 3: srep01236.

Majeed, K., Sorour, S., Al-Naffouri, T. Y., Valaee, S., 2015. Indoor localization and radio map estimation using unsupervised manifold alignment with geometry perturbation. IEEE Transactions on Mobile Computing 15 (11): 2794–2808.

Melchiorri, M., Florczyk, A., Freire, S., Schiavina, M., Pesaresi, M., Kemper, T., 2018. Unveiling 25 years of planetary urbanization with remote sensing: Perspectives from the global human settlement layer. Remote Sensing 10 (5): 768.

Melchiorri, M., Pesaresi, M., Florczyk, A. J., Corbane, C., Kemper, T., 2019. Principles and applications of the global human settlement layer as baseline for the land use efficiency indicatorâ€"sdg 11.3. 1. ISPRS International Journal of Geo-Information 8 (2): 96.

Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K. P., 2013. A tutorial on synthetic aperture radar. IEEE Geoscience and remote sensing magazine 1 (1): 6–43.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66 (3): 247–259.

Munkres, J., 2014. Topology. Pearson Education.

Mura, M. D., Prasad, S., Pacifici, F., Gamba, P., Chanussot, J., Benediktsson, J. A., 2015. Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing. Proceedings of the IEEE 103 (9): 1585–1601.

Næsset, E., Ørka, H. O., Solberg, S., Bollandsås, O. M., Hansen, E. H., Mauya, E., Zahabu, E., Malimbwi, R., Chamuya, N., Olsson, H., Gobakken, T., 2016. Mapping and estimating forest area and aboveground biomass in miombo woodlands in Tanzania using data from airborne laser scanning, TanDEM-X, RapidEye, and global forest maps: A comparison of estimated precision. Remote Sensing of Environment 175: 282–300.

Neumann, M., Ferro-Famil, L., Reigber, A., 2009. Estimation of forest structure, ground, and canopy layer characteristics from multibaseline polarimetric interferometric sar data. IEEE Transactions on Geoscience and Remote Sensing 48 (3): 1086–1104.

Ng, A. Y., Jordan, M. I., Weiss, Y., 2002. On spectral clustering: Analysis and an algorithm. In: Advances in neural information processing systems, 849–856.

Nicolau, M., Levine, A. J., Carlsson, G., 2011. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proceedings of the National Academy of Sciences 108 (17): 7265–7270.

Nielson, J. L., Paquette, J., Liu, A. W., Guandique, C. F., Tovar, C. A., Inoue, T., Irvine, K.-A., Gensel, J. C., Kloke, J., Petrossian, T. C., et al., 2015. Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. Nature communications 6: 8581.

Okujeni, A., Van Der Linden, S., Hostert, P., 2016. Berlin-urban-gradient dataset 2009-an enmap preparatory flight campaign (datasets). GFZ Data Services .

Oloyede, M. O., Hancke, G. P., 2016. Unimodal and Multimodal Biometric Sensing Systems: A Review. IEEE Access 4: 7532–7555.

Pacifici, F., Frate, F. D., Emery, W. J., Gamba, P., Chanussot, J., 2008. Urban Mapping Using Coarse SAR and Optical Data: Outcome of the 2007 GRSS Data Fusion Contest. IEEE Geoscience and Remote Sensing Letters 5 (3): 331–335.

Palubinskas, G., 2012. How to fuse optical and radar imagery? In: 2012 IEEE International Geoscience and Remote Sensing Symposium, 2171–2174.

Pearson, K., 1901. Liii. on lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2 (11): 559–572.

Pei, Y., Huang, F., Shi, F., Zha, H., 2012. Unsupervised image matching based on manifold alignment. IEEE transactions on pattern analysis and machine intelligence 34 (8): 1658–1664.

Pei, Y., Kim, T.-K., Zha, H., 2013. Unsupervised random forest manifold alignment for lipreading. In: Proceedings of the IEEE International Conference on Computer Vision, 129–136.

Pereira, L. O., Freitas, C. C., SantÂ´Anna, S. J. S., Reis, M. S., 2018. Evaluation of Optical and Radar Images Integration Methods for LULC Classification in Amazon Region. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (9): 3062–3074.

Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., Lu, L., et al., 2013. A global human settlement layer from optical hr/vhr rs data: concept and first results. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6 (5): 2102–2131.

Pohl, C., Van Genderen, J. L., 2017. Remote sensing image fusion: a practical guide. Taylor & Francis, Boca Raton, FL.

Prendes, J., Chabert, M., Pascal, F., Giros, A., Tourneret, J., 2015. A New Multivariate Statistical

Model for Change Detection in Images Acquired by Homogeneous and Heterogeneous Sensors. IEEE Transactions on Image Processing 24 (3): 799–812.

Qin, Y., Xiao, X., Dong, J., Chen, B., Liu, F., Zhang, G., Zhang, Y., Wang, J., Wu, X., 2017. Quantifying annual changes in built-up area in complex urban-rural landscapes from analyses of PALSAR and Landsat images. ISPRS Journal of Photogrammetry and Remote Sensing 124: 89–105.

Qin, Y., Xiao, X., Dong, J., Zhang, G., Shimada, M., Liu, J., Li, C., Kou, W., Moore, B., 2015. Forest cover maps of China in 2010 from multiple approaches and data sources: PALSAR, Landsat, MODIS, FRA, and NFI. ISPRS Journal of Photogrammetry and Remote Sensing 109: 1–16.

Rainforth, T., Wood, F., 2015. Canonical correlation forests. arXiv preprint arXiv:1507.05444 .

Rajah, P., Odindi, J., Mutanga, O., 2018. Feature level image fusion of optical imagery and Synthetic Aperture Radar (SAR) for invasive alien plant species detection and mapping. Remote Sensing Applications: Society and Environment 10: 198–208.

Rajah, P., Odindi, J., Mutanga, O., Kiala, Z., 2019. The utility of Sentinel-2 Vegetation Indices (VIs) and Sentinel-1 Synthetic Aperture Radar (SAR) for invasive alien species detection and mapping. Nature Conservation 35: 41–61.

Ramsey, E., Rangoonwala, A., Chi, Z., Jones, C. E., Bannister, T., 2014. Marsh Dieback, loss, and recovery mapped with satellite optical, airborne polarimetric radar, and field data. Remote Sensing of Environment 152: 364–374.

Rasti, B., Ghamisi, P., Gloaguen, R., 2017. Hyperspectral and lidar fusion using extinction profiles and total variation component analysis. IEEE Transactions on Geoscience and Remote Sensing 55 (7): 3997–4007.

Reiche, J., Hamunyela, E., Verbesselt, J., Hoekman, D., Herold, M., 2018. Improving near-real time deforestation monitoring in tropical dry forests by combining dense Sentinel-1 time series with Landsat and ALOS-2 PALSAR-2. Remote Sensing of Environment 204: 147–161.

Reiche, J., Verbesselt, J., Hoekman, D., Herold, M., 2015. Fusing Landsat and SAR time series to detect deforestation in the tropics. Remote Sensing of Environment 156: 276–293.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566 (7743): 195–204.

Revill, A., Sus, O., Barrett, B., Williams, M., 2013. Carbon cycling of European croplands: A framework for the assimilation of optical and microwave Earth observation data. Remote Sensing of Environment 137: 84–93.

Robson, B. A., Nuth, C., Dahl, S. O., Hölbling, D., Strozzi, T., Nielsen, P. R., 2015. Automated classification of debris-covered glaciers combining optical, SAR and topographic data in an object-based environment. Remote Sensing of Environment 170: 372–387.

Rokach, L., 2010. Ensemble-based classifiers. Artificial Intelligence Review 33 (1-2): 1–39.

Roweis, S. T., Saul, L. K., 2000. Nonlinear dimensionality reduction by locally linear embedding. science 290 (5500): 2323–2326.

Salentinig, A., Gamba, P., 2015. Combining SAR-Based and Multispectral-Based Extractions to Map Urban Areas at Multiple Spatial Resolutions. IEEE Geoscience and Remote Sensing Magazine 3 (3): 100–112.

Schiavina, M., Freire, S., MacManus, K., 2019. Ghs population grid multitemporal (1975, 1990, 2000, 2015). European Commission, Joint Research Centre (JRC)[Dataset] PID: DOI: 10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218 PID: http://data.europa.eu/89h/0c6b9751-a71f-4062-830b-43c9f432370f .

Schmitt, A., Wendleder, A., Hinz, S., 2015. The kennaugh element framework for multi-scale, multi-polarized, multi-temporal and multi-frequency sar image preparation. ISPRS journal of photogrammetry and remote sensing 102: 122–139.

Schmitt, M., Zhu, X. X., 2016. Data Fusion and Remote Sensing: An ever-growing relationship. IEEE Geoscience and Remote Sensing Magazine 4 (4): 6–23.

Schneider, T., Schauerte, B., Stiefelhagen, R., 2014. Manifold alignment for person independent appearance-based gaze estimation. In: 2014 22nd International Conference on Pattern Recogni-

tion, IEEE, 1167–1172.

Scholkopf, B., Smola, A. J., 2001. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press.

Serpico, S. B., Roli, F., 1995. Classification of multisensor remote-sensing images by structured neural networks. IEEE Transactions on Geoscience and Remote Sensing 33 (3): 562–578.

Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. Departmental Papers (CIS) : 107.

Singh, G., Mémoli, F., Carlsson, G. E., 2007. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In: SPBG, 91–100.

Solberg, A. H. S., Jain, A. K., Taxt, T., 1994. Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images. IEEE Transactions on Geoscience and Remote Sensing 32 (4): 768–778.

Spröhnle, K., Fuchs, E., Pelizari, P. A., 2017. Object-Based Analysis and Fusion of Optical and SAR Satellite Data for Dwelling Detection in Refugee Camps. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (5): 1780–1791.

Stefanski, J., Kuemmerle, T., Chaskovskyy, O., Griffiths, P., Havryluk, V., Knorn, J., Korol, N., Sieber, A., Waske, B., 2014. Mapping Land Management Regimes in Western Ukraine Using Optical and SAR Data. Remote Sensing 6 (6): 5279–5305.

Stewart, I. D., Oke, T. R., 2012. Local climate zones for urban temperature studies. Bulletin of the American Meteorological Society 93 (12): 1879–1900.

Sukawattanavijit, C., Chen, J., Zhang, H., 2017. GA-SVM Algorithm for Improving Land-Cover Classification Using SAR and Optical Remote Sensing Data. IEEE Geoscience and Remote Sensing Letters 14 (3): 284–288.

Sun, G., Kong, Y., Jia, X., Zhang, A., Rong, J., Ma, H., 2019. Synergistic Use of Optical and Dual-Polarized SAR Data With Multiple Kernel Learning for Urban Impervious Surface Mapping. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (1): 223–236.

Tabib Mahmoudi, F., Arabsaeedi, A., Alavipanah, S. K., 2019. Feature-Level Fusion of Landsat 8 Data and SAR Texture Images for Urban Land Cover Classification. Journal of the Indian Society of Remote Sensing 47 (3): 479–485.

Tan, S., Guan, Z., Cai, D., Qin, X., Bu, J., Chen, C., 2014. Mapping users across networks by manifold alignment on hypergraph. In: Twenty-Eighth AAAI Conference on Artificial Intelligence.

Tan, W., Liao, R., Du, Y., Lu, J., Li, J., 2015. Improving urban impervious surface classification by combining Landsat and PolSAR images: A case study in Kitchener-Waterloo, Ontario, Canada. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 1917–1920.

Tao, L., Wang, G., Chen, W., Chen, X., Li, J., Cai, Q., 2019. Soil Moisture Retrieval From SAR and Optical Data Using a Combined Model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (2): 637–647.

Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., Dech, S., 2012. Monitoring urbanization in mega cities from space. Remote Sensing of Environment 117: 162–176.

Taubenböck, H., Weigand, M., Esch, T., Staab, J., Wurm, M., Mast, J., Dech, S., 2019. A new ranking of the world's largest citiesâ€"do administrative units obscure morphological realities? Remote Sensing of Environment 232: 111353.

Tenenbaum, J. B., De Silva, V., Langford, J. C., 2000. A global geometric framework for nonlinear dimensionality reduction. science 290 (5500): 2319–2323.

Tong, X., Luo, X., Liu, S., Xie, H., Chao, W., Liu, S., Liu, S., Makhinov, A., Makhinova, A., Jiang, Y., 2018. An approach for flood monitoring by the combined use of Landsat 8 optical imagery and COSMO-SkyMed radar imagery. ISPRS Journal of Photogrammetry and Remote Sensing 136: 144–153.

Tuia, D., Camps-Valls, G., 2016. Kernel manifold alignment for domain adaptation. PloS one 11 (2): e0148655.

Tuia, D., Volpi, M., Trolliet, M., Camps-Valls, G., 2014. Semisupervised manifold alignment of multimodal remote sensing images. IEEE Transactions on Geoscience and Remote Sensing 52 (12): 7708–7720.

Tupin, F., Roux, M., 2003. Detection of building outlines based on the fusion of SAR and optical features. ISPRS Journal of Photogrammetry and Remote Sensing 58 (1-2): 71–82.

UN, 2018. United nations, the world's cities in 2018. https://www.un.org/en/events/citiesday/assets/pdf/the_worlds_cities_in_2018_data_booklet.pdf, accessed: 2019-06-14.

Vaglio Laurin, G., Liesenberg, V., Chen, Q., Guerriero, L., Del Frate, F., Bartolini, A., Coomes, D., Wilebore, B., Lindsell, J., Valentini, R., 2013. Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa. International Journal of Applied Earth Observation and Geoinformation 21: 7–16.

van Beijma, S., Comber, A., Lamb, A., 2014. Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. Remote Sensing of Environment 149: 118–129.

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F., Ceschia, E., 2017. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. Remote Sensing of Environment 199: 415–426.

Villa, P., Fontanelli, G., Crema, A., 2015a. Integration of multi-seasonal Landsat 8 and TerraSAR-X data for urban mapping: An assessment. In: 2015 Joint Urban Remote Sensing Event (JURSE), 1–4.

Villa, P., Stroppiana, D., Fontanelli, G., Azar, R., Brivio, P. A., 2015b. In-Season Mapping of Crop Type with Optical and X-Band SAR Data: A Classification Tree Approach Using Synoptic Seasonal Features. Remote Sensing 7 (10): 12859–12886.

Von Luxburg, U., 2007. A tutorial on spectral clustering. Statistics and computing 17 (4): 395–416.

Wan, L., Xiang, Y., You, H., 2019. A Post-Classification Comparison Method for SAR and Optical Images Change Detection. IEEE Geoscience and Remote Sensing Letters 16 (7): 1026–1030.

Wang, C., Mahadevan, S., 2008. Manifold alignment using procrustes analysis. In: Proceedings of the 25th international conference on Machine learning, ACM, 1120–1127.

Wang, C., Mahadevan, S., 2009a. A general framework for manifold alignment. In: AAAI fall symposium: manifold learning and its applications, 53–58.

Wang, C., Mahadevan, S., 2009b. Manifold alignment without correspondence. In: Twenty-First International Joint Conference on Artificial Intelligence.

Wang, C., Mahadevan, S., 2011. Heterogeneous domain adaptation using manifold alignment. In: IJCAI proceedings-international joint conference on artificial intelligence, Vol. 22, 1541.

Wang, C., Mahadevan, S., 2013. Manifold alignment preserving global geometry. In: Twenty-Third International Joint Conference on Artificial Intelligence.

Wang, J., Xiao, X., Qin, Y., Dong, J., Geissler, G., Zhang, G., Cejda, N., Alikhani, B., Doughty, R. B., 2017. Mapping the dynamics of eastern redcedar encroachment into grasslands during 1984â€“2010 through PALSAR and time series Landsat images. Remote Sensing of Environment 190: 233–246.

Wang, X., Guo, Y., He, J., Du, L., 2016. Fusion of HJ1b and ALOS PALSAR data for land cover classification using machine learning methods. International Journal of Applied Earth Observation and Geoinformation 52: 192–203.

Ward, D., Petty, A., Setterfield, S., Douglas, M., Ferdinands, K., Hamilton, S., Phinn, S., 2014. Floodplain inundation and vegetation dynamics in the Alligator Rivers region (Kakadu) of northern Australia assessed using optical and radar remote sensing. Remote Sensing of Environment 147: 43–55.

Waske, B., Benediktsson, J. A., 2007. Fusion of Support Vector Machines for Classification of Multisensor Data. IEEE Transactions on Geoscience and Remote Sensing 45 (12): 3858–3866.

Wegner, J. D., Ziehn, J. R., Soergel, U., 2010. Building detection and height estimation from high-resolution insar and optical data. In: 2010 IEEE International Geoscience and Remote Sensing Symposium, 1928–1931.

Wurm, M., Taubenböck, H., Weigand, M., Schmitt, A., 2017. Slum mapping in polarimetric sar data using spatial features. Remote sensing of environment 194: 190–204.

Xiang, D., Ban, Y., Su, Y., 2015. Model-based decomposition with cross scattering for polarimetric sar urban areas. IEEE Geoscience and Remote Sensing Letters 12 (12): 2496–2500.

Xiao, Y., Jiang, Q., Wang, B., Li, Y., Liu, S., Cui, C., 2016. Object-oriented fusion of RADARSAT-2 polarimetric synthetic aperture radar and HJ-1a multispectral data for land-cover classification. Journal of Applied Remote Sensing 10 (2): 026021.

Xu, R., Zhang, H., Lin, H., 2017. Urban Impervious Surfaces Estimation From Optical and SAR Imagery: A Comprehensive Comparison. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (9): 4010–4021.

Yamaguchi, Y., Yajima, Y., Yamada, H., 2006. A four-component decomposition of polsar images based on the coherency matrix. IEEE Geoscience and Remote Sensing Letters 3 (3): 292–296.

Yang, H. L., Crawford, M. M., 2015. Spectral and spatial proximity-based manifold alignment for multitemporal hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 54 (1): 51–64.

Yang, J., Ren, G., Ma, Y., Fan, Y., 2016. Coastal wetland classification based on high resolution SAR and optical image fusion. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 886–889.

Yang, Z., Shao, Y., Li, K., Liu, Q., Liu, L., Brisco, B., 2017. An improved scheme for rice phenology estimation based on time-series multispectral HJ-1a/B and polarimetric RADARSAT-2 data. Remote Sensing of Environment 195: 184–201.

Yokoya, N., Ghamisi, P., Xia, J., 2017. Multimodal, multitemporal, and multisource global data fusion for local climate zones classification based on ensemble learning. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, 1197–1200.

Yousif, O., Ban, Y., 2017. Fusion of SAR and optical data for unsupervised change detection: A case study in Beijing. In: 2017 Joint Urban Remote Sensing Event (JURSE), 1–4.

Zelnik-Manor, L., Perona, P., 2005. Self-tuning spectral clustering. In: Advances in neural information processing systems, 1601–1608.

Zhang, H., Lin, H., Li, Y., 2015. Impacts of Feature Normalization on Optical and SAR Data Fusion for Land Use/Land Cover Classification. IEEE Geoscience and Remote Sensing Letters 12 (5): 1061–1065.

Zhang, H., Wan, L., Wang, T., Lin, Y., Lin, H., Zheng, Z., 2019a. Impervious Surface Estimation From Optical and Polarimetric SAR Data Using Small-Patched Deep Convolutional Networks: A Comparative Study. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (7): 2374–2387.

Zhang, J., 2010. Multi-source remote sensing data fusion: status and trends. International Journal of Image and Data Fusion 1 (1): 5–24.

Zhang, L., Zhang, Q., Du, B., Huang, X., Tang, Y. Y., Tao, D., 2016. Simultaneous spectral-spatial feature selection and extraction for hyperspectral images. IEEE Transactions on Cybernetics 48 (1): 16–28.

Zhang, Y., Ling, F., Foody, G. M., Ge, Y., Boyd, D. S., Li, X., Du, Y., Atkinson, P. M., 2019b. Mapping annual forest cover by fusing PALSAR/PALSAR-2 and MODIS NDVI during 2007â€“2016. Remote Sensing of Environment 224: 74–91.

Zhang, Y., Zhang, H., Lin, H., 2014. Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. Remote Sensing of Environment 141: 155–167.

Zhao, W., Du, S., 2016. Spectralâ€“spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing 54 (8): 4544–4554.

Zheng, J., Geldsetzer, T., Yackel, J., 2017. Snow thickness estimation on first-year sea ice using microwave and optical remote sensing with melt modelling. Remote Sensing of Environment 199: 321–332.

Zhou, T., Li, Z., Pan, J., 2018a. Multi-Feature Classification of Multi-Sensor Satellite Imagery Based on Dual-Polarimetric Sentinel-1a, Landsat-8 OLI, and Hyperion Images for Urban Land-Cover Classification. Sensors 18 (2): 373.

Zhou, T., Zhao, M., Sun, C., Pan, J., 2018b. Exploring the Impact of Seasonality on Urban Land-Cover Mapping Using Multi-Season Sentinel-1a and GF-1 WFV Images in a Subtropical Monsoon-

Climate Region. ISPRS International Journal of Geo-Information 7 (1): 3.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine 5 (4): 8–36.

Zhu, Z., Woodcock, C. E., Rogan, J., Kellndorfer, J., 2012. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. Remote Sensing of Environment 117: 72–82.

Zhu, Z., Zhou, Y., Seto, K. C., Stokes, E. C., Deng, C., Pickett, S. T., Taubenböck, H., 2019. Understanding an urbanizing planet: Strategic directions for remote sensing. Remote Sensing of Environment 228: 164–182.

Zomorodian, A., Carlsson, G., 2005. Computing Persistent Homology. Discrete & Computational Geometry 33 (2): 249–274.

List of Figures

List of Tables

Acknowledgement

The past five years have been and will always be an important journey in my life. It has not only formed the foundation of my future career but also shaped me on every aspect. Many people have supported, educated, and influenced me, for which I cannot express enough gratitude.

First of all, I would like to express my deepest gratitude to my doctoral advisor, Prof. Xiao Xiang Zhu. Even before my Ph.D. project, she guided me into the remote sensing field by supervising my master thesis. She was the one who believed in me, and offered me a chance of doctoral candidacy. In last five years, she was the one who always broadens my horizon, demonstrates me what is diligence via actions, and supports me constantly. It was my fortunes to have her guiding me on the journey.

I would like to sincerely thank Prof. Richard Bamler for being my second supervisor. It was him who always provide critical comments in scientific discussions such as seminars. And those comments always pick out essential contents and inspire further thoughts. The same sincere gratitude goes to Prof. Peter Reinartz who were interested to review my dissertation, and provided constructive comments and questions. I am grateful to Prof. Urs Hugentobler whom I always look up to since my master program. It is my honor to have Prof. Urs Hugentobler being the chairman of the examination committee of my thesis.

It was a tremendous pleasure to work in the group SiPEO, TU München, and the institute IMF, German Aerospace Center. Both institutes provide me invaluable opportunity to approach pioneer scientists in the field of remote sensing. I would like to thank Prof. Andreas Schmitt, Dr. Lichao Mou, Dr. Pedram Ghamisi, Dr. Danfeng Hong, Dr. Yuanyuan Wang, and Dr. Hannes Taubenböck. They all supported me by having intensive scientific discussions and sharing cutting-edge knowledge. Every bit of their selfless support is a building brick of my dissertation. I would also like to thank Prof. Michael Schmitt who supported me immensely on both scientific and bureaucratic matters, Prof. Naoto Yokoya, who shared me his knowledge and opinions, Dr. Martin Werner, who explained me his inspiring interpretation of topology, Dr. Claas Grohnfeldt, who shared me his knowledge of hyperspectral images, Dr. Gerald Baier, who helped me on many technical matters, Dr. Yilei Shi, who discussed feature design of SAR data with me, Dr. Chunping Qiu, who I worked with on So2Sat project, Dr. Jian Kang, who exhibited me a clear path of a research project, Dr. Sina Montazeri, with whom I enjoyed the breaks of office hours, Matthias Häberle, who turned the break time into a nice business meeting, Dr. Nan Ge, who encouraged me to explore topology, Yao Sun, who enlighten me with her focus, Eike Hoffmann, who helped me with ubuntu, Dr. Rong Liu, whom I worked with in hyperspectral working group. Yuansheng Hua, Qingyun Li, Kun Qian, Dr. Cong Luo, and Fei Han, who together made me a nice doctoral hat.

Last but not least, I want to thank my dearest friends, Dr. Jian Xu, Dr. Bingbing Duan, Yuanxin Xia, Chaonan Ji, and Dr. Song Liu. Talking to them released a lot pressure and created tons of joys. I cannot thank enough to my parents. They have been supporting me unconditionally under any circumstances. I want to thank my girlfriend Meijie for standing with me and reminding me the bright side of things whenever I needed.

Appendices

A Hu J, Ghamisi P, and Zhu X X (2018). Feature extraction and selection of sentinel-1 dual-pol data for global-scale local climate zone classification. ISPRS International Journal of Geo-Information, 7(9), 379.

https://www.mdpi.com/2220-9964/7/9/379

International Journal of Geo-Information

MDPI

Article

Feature Extraction and Selection of Sentinel-1 Dual-Pol Data for Global-Scale Local Climate Zone Classification

Jingliang Hu [1,2], Pedram Ghamisi [2,3] and Xiao Xiang Zhu [1,2]*

[1] Signal processing in Earth observation (SiPEO), Technische Universität München (TUM), 80333 Munich, Germany; jingliang.hu@dlr.de
[2] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Weßling, Germany; p.ghamisi@gmail.com
[3] Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology (HIF), Exploration, D-09599 Freiberg, Germany
* Correspondence: xiao.zhu@dlr.de; Tel.: +49-08153-28-3531

Received: 31 July 2018; Accepted: 10 September 2018; Published: date

Abstract: The concept of the local climate zone (LCZ) has been recently proposed as a generic land-cover/land-use classification scheme. It divides urban regions into 17 categories based on compositions of man-made structures and natural landscapes. Although it was originally designed for temperature study, the morphological structure concealed in LCZs also reflects economic status and population distribution. To this end, global LCZ classification is of great value for worldwide studies on economy and population. Conventional classification approaches are usually successful for an individual city using optical remote sensing data. This paper, however, attempts for the first time to produce global LCZ classification maps using polarimetric synthetic aperture radar (PolSAR) data. Specifically, we first produce polarimetric features, local statistical features, texture features, and morphological features and compare them, with respect to their classification performance. Here, an ensemble classifier is investigated, which is trained and tested on already separated transcontinental cities. Considering the challenging global scope this work handles, we conclude the classification accuracy is not yet satisfactory. However, Sentinel-1 dual-Pol SAR data could contribute the classification for several LCZ classes. According to our feature studies, the combination of local statistical features and morphological features yields the best classification results with 61.8% overall accuracy (OA), which is 3% higher than the OA produced by the second best features combination. The 3% is considerably large for a global scale. Based on our feature importance analysis, features related to VH polarized data contributed the most to the eventual classification result.

Keywords: Sentinel-1 dual-Pol data; local climate zone; global scale; feature extraction; GLCM; morphological profile; canonical correlation forest

1. Introduction

The local climate zone (LCZ) classification system is designed as a categorical scheme with 17 classes that describe urban landscapes [1,2]. These classes are defined based on surface structures and surface covers, which are specifically (1) the height of the surface structure, (2) spatial density of the surface structure, and (3) covering material of the surface (i.e., as shown in Figure 1). Eventually, the scheme yields thematic maps of a 100-m ground sampling distance (GSD), with each pixel labeled as one of the 17 classes. The LCZ was designed for the study of urban temperature behavior. It provides a research framework for urban heat island studies and standardizes the worldwide exchange of urban temperature observations [1]. As also shown in Figure 1, the scheme essentially demonstrates

morphological structures of urban local neighborhood. The urban morphological structure is an influential factor on thermal behaviour. It may also reveal the economic status and the population distribution of a particular city. For example, slum districts, which are less economically developed regions with massive population concentration, normally appear as the seventh class in Figure 1. Therefore, thanks to the described morphological structure, the LCZ map acts as a valuable source for a wide variety of studies in urban areas.

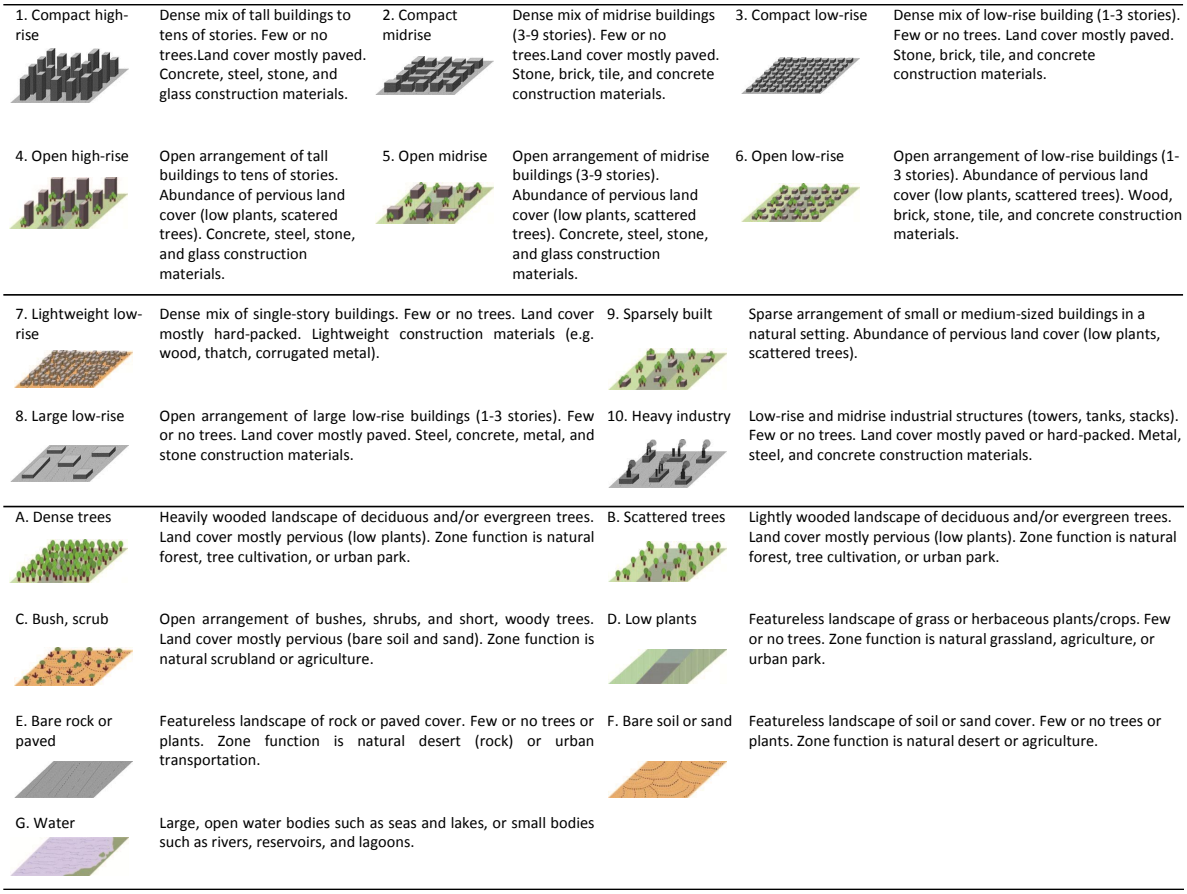


Figure 1. Description of LCZ classes. (adapted from [1])

Following the introduction of LCZs, the World Urban Database and Portal (WUDAPT, http://www.wudapt.org) was initiated [3,4]. WUDAPT has been mainly developed by researchers to obtain high-quality land-cover/land-use information globally, usually via crowdsourcing [5,6], games [7], or other challenges [8]. The WUDAPT project presents a suggested workflow to produce the LCZ map by taking advantage of remote sensing techniques. The production process briefly functions as follows [3]: First, the region of interest (ROI) for a particular city is defined and labels of all classes are manually selected. Second, multispectral images captured by LandSat-8 are prepared for the ROI of the target city. Finally, a supervised classification (i.e., random forest [9]) is applied to the multispectral data to produce the final classification map. Besides the standard production of the WUDAPT project, studies of LCZ classification mainly focus on using optical remote sensing data [10–12].

One of the key factors to define LCZ classes is height. A few studies on LCZ classification were introduced to fuse the digital surface model (DSM) with the optical data in order to use both height and spectral information [13,14]. The Thematic Mapper (TM) data captured by LandSat-5 and Enhanced Thematic Mapper Plus (ETM+) data of LandSat-7 were fused with the normalized digital surface model (NDSM) and airborne Interferometric Synthetic Aperture Radar (InSAR) using feature concatenation and then applied to multiple classifiers for LCZ-related classification in [13]. In [14], LandSat-8 data and the digital surface model (DSM) were also concatenated at the feature

level and then classified by a random forest and support vector machine [15]. It was concluded that spectral features can significantly contribute to the classification task. Geographical information system (GIS)-based approach is also developed to produce the LCZ map in [16–18]. Although Open Street Map (OSM) provides a free-accessible GIS dataset, the completeness of the GIS data needs to be improved, especially for the developing countries. Besides the aforementioned data sources, the Sentinel-1 mission provides a dual-polarimetric synthetic aperture radar (dual-Pol SAR) with free access and global coverage. It has also been studied for LCZ classification in [19]. Researchers have proven that the combination of Sentinel-1 dual-Pol data and LandSat-8 data can improve the performance of LCZ classification. However, they only take used of amplitudes of VV and VH channels and their corresponding texture features derived by gray-level co-occurrence matrix (GLCM). Amplitudes of the two channels only constitute one part of the polarimetric information provided by the dual-Pol data. One important feature, the coherence of the two channels, is missing. Therefore, the first goal of this work is to comprehensively investigate the polarimetric information of Sentinel-1 dual-Pol data for the LCZ classification task.

Global LCZ mapping offers substantial help in exploring local climates on regional and worldwide scales. Several studies have successfully produced LCZ classification maps corresponding to one city by only labeling samples of that city. In this manner, the produced classification maps have achieved high classification accuracy (e.g., overall accuracies (OAs) are beyond 80%) since both training and test samples are located in the same city. However, the collection of accurate training samples are either expensive or time-demanding. Therefore, in the remote sensing community, there is great interest in training models based on the samples available for some cities and applying the trained models to other cities. However, this is a challenging task. For example, one study [20] attempted to select training samples from one city for the classification of another city using RF. The classification accuracies dropped to 18.2%, which indicates that the knowledge transferability between different cities should be carefully considered. In this regard, our second goal is to develop a classifier with adequate generalization capability to be applied to any other cities. The difficulty in this task lies in that the classifier needs to be trained using a limited number of training samples while remaining applicable to handling transcultural, transnational, and cross-environmental data in a worldwide context.

To cope with the aforementioned challenge of generalization, the 2017 Geoscience and Remote Sensing Society (GRSS) data fusion contest of the year 2017 proposed training the classifier on five cities (Berlin, Hong Kong, Paris, Rome, and Sao Paulo) and testing the results on four other cities (Amsterdam, Chicago, Madrid, and Xi'an). Although deep learning-based classification methods have proven to be strong in terms of classification accuracy a generalization capability in the remote sensing community [21–24], the ensemble-based canonical correlation forest (CCF) classification strategy achieved the best performance in the contest, among more than 800 submissions [8,25]. Therefore, this work uses the CCF classifier to pursue a solution for our task. The CCF classifier is an advanced version of a random forest, which is a shallow classifier.

In contrast with the automatic feature selection and extraction of deep learning methods, the feature design is of key importance to shallow classifiers. From the perspective of feature space, especially for dual-Pol SAR data, a limited number of features is not adequate for a complicated classification like the LCZ task. An informative and appropriate number of features should be derived for the subsequent classification task. References [13,25] indicates that local statistical features are informative features regarding LCZ classifications. Texture features derived from GLCM have been proven to be informative for applications of SAR data [26–30]. Mathematical morphological features obtained from a morphological profile have been highly effective in multi/hyperspectral image classification [31–33]. Consequently, besides polarimetric features, we investigate the performances of local statistical features, GLCM features, and morphological profiles for LCZ classification on a global scale.

To sum up, the aims of the study are threefold for local climate zone classification. (1) Comprehensive polarimetric information of the Sentinel-1 dual-Pol data is investigated, which includes intensities of VV and VH channels as well as the coherence and relative phase of the two channels. (2) Classification on a global scale is studied by training and testing the CCF on the separated data of transcontinental cities, which involves terabytes of data volume. (3) Four features (polarimetric feature, local statistical feature, texture feature, and mathematical morphological feature) that were proven to be successful in related tasks are evaluated in our scenario of global-scale LCZ classification.

The rest of the paper is organized as follows. Section 2 demonstrates the principle of selecting a study area and describes the Sentinel-1 dual-Pol data and its data preparation. Section 3 introduces our methodology of global-scale LCZ classification. Section 4 discusses the experiment results regarding feature extraction and selection. Lastly, Section 5 concludes this work.

2. Study Area and Data Set

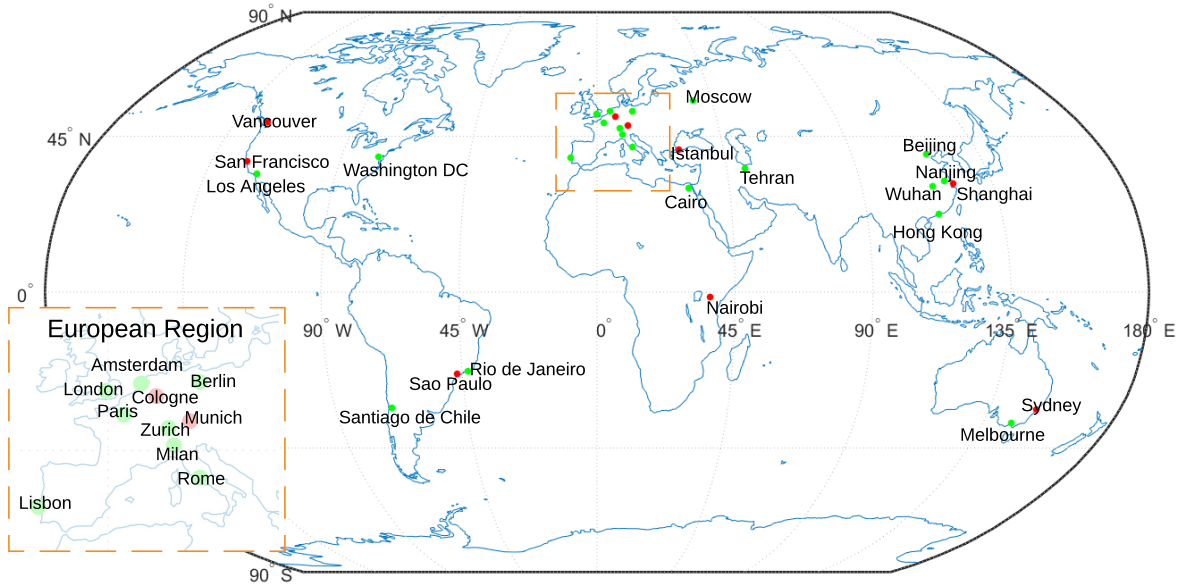


Figure 2. World-wide distribution of selected 29 cities of interest. Red: cities for testing. Green: cities for training.

2.1. Study Areas

Our study aims to produce LCZ maps on a global scale and also focuses on cities of high population density. In total, 29 cities were selected and listed in Table 1. They are located on all continents except Antarctica, shown in Figure 2. This geographical distribution ensures that cities of interest include transcultural, transnational, and cross-environmental regions. While selecting these cities, population was another criteria under consideration. Among all cities, the population of each city was at least one million in 2016 and is expected to grow in the future according to UN statistics [34]. To enable our framework to solve the global challenge, cities of different regions were selected for both training and testing. The selection is shown in Figure 2 and Table 1.

2.2. Ground Truth

To produce LCZ classification maps in a supervised manner, we manually created the ground truth for the selected 29 cities. Generally, the labeling procedure followed the WUDAPT project standard procedure [4]. First, the region of interest (ROI) was decided for each selected city as a 50-by-50-kilometers rectangle centered at the city center. Within the rectangle, the ground truth, polygons of LCZ classes, were manually delineated by observing satellite images on Google Earth

Table 1. List of cities of interest, with information on the regions and populations [34]. List of chosen combinations of cities for training and testing.

Region	City	Training City	Testing City	Polulation at Year		
				2000	2016	2030
Australia	Melbourne	Y	-	3,461,000	4,258,000	5,071,000
	Sydney	-	Y	4,052,000	4,540,000	5,301,000
Eastern Asia	Beijing	Y	-	10,162,000	21,240,000	27,706,000
	Nanjing	Y	-	6,160,000	8,270,000	9,750,000
	Wuhan	Y	-	6,638,000	7,979,000	9,442,000
	Hong Kong	Y	-	6,835,000	7,365,000	7,885,000
	Shanghai	-	Y	13,959,000	24,484,000	30,751,000
Western Asia	Tehran	Y	-	7,128,000	8,516,000	9,990,000
	Istanbul	-	Y	8,744,000	14,365,000	16,694,000
Africa	Cairo	Y	-	13,626,000	19,128,000	24,502,000
	Nairobi	-	Y	2,214,000	4,070,000	7,140,000
Europe	Amsterdam	Y	-	1,005,000	1,099,000	1,213,000
	Berlin	Y	-	3,384,000	3,578,000	3,658,000
	London	Y	-	8,613,000	10,434,000	11,467,000
	Paris	Y	-	9,737,000	10,925,000	11,803,000
	Zurich	Y	-	1,078,000	1,259,000	1,494,000
	Milan	Y	-	2,985,000	3,104,000	3,162,000
	Rome	Y	-	3,385,000	3,738,000	3,842,000
	Lisbon	Y	-	2,672,000	2,902,000	3,192,000
	Moscow	Y	-	10,005,000	12,260,000	12,200,000
	Cologne	-	Y	963,000	1,042,000	1,095,000
	Munich	-	Y	1,202,000	1,454,000	1,548,000
North America	Washington DC	Y	-	3,949,000	5,013,000	5,690,000
	Los Angeles	Y	-	11,798,000	12,317,000	13,257,000
	San Francisco	-	Y	3,230,000	3,299,000	3,615,000
	Vancouver	-	Y	1,959,000	2,523,000	2,930,000
South America	Rio de Janeiro	Y	-	11,307,000	12,981,000	14,174,000
	Santiago de Chile	Y	-	5,658,000	6,544,000	7,122,000
	Sao Paulo	-	Y	17,014,000	21,297,000	23,444,000

(https://www.google.com/earth/). Then, LandSat-8 images were prepared for the ROIs of each city. Afterwards, the software SAGA GIS (www.saga-gis.org/en/index.html), taking the delineated ground truth and LandSat-8 data as inputs, produced an LCZ classification map using the random forest classifier. The produced classification map is aim to be used as an additional validation source for checking the correctness and completeness of the ground truth data. By manually cross checking the classification map, images on Google Earth, and the delineated ground truth, the ground truth is modified if necessary, in the regard of correctness and completeness. Eventually, the delineated polygons of LCZ classes are created as the ground truth data, which are used for training and testing in this work. The delineated ground truth data are shown in Figure 3 with Sentinel-1 data as background.

2.3. Sentinel-1 Dual-Pol Data

The Sentinel-1 mission, as the SAR component of the European Copernicus program, has a constellation of two satellites each mounted with a C-band Synthetic Aperture Radar sensor. It has global coverage with a temporal resolution of six days. Additionally, the data is freely accessible.

The sensor collects data in four modes: (1) Stripmap (SM), (2) Interferometric Wide swath (IW), (3) Extra Wide swath (EW), and (4) Wave (WV). We used a level-1 product, which is focused single look complex data collected from the Interferometric Wide swath mode, because of its large coverage and

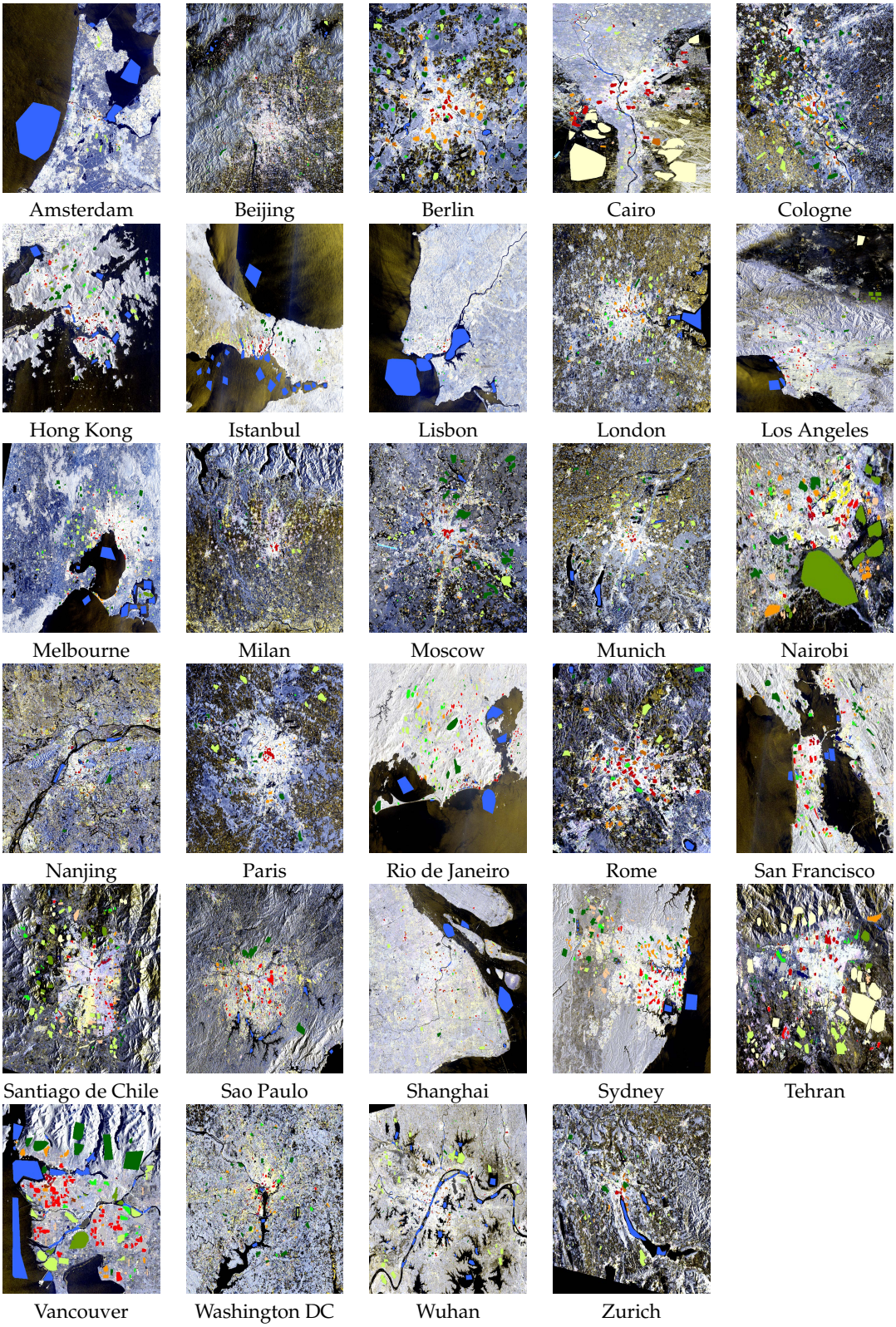


Figure 3. Processed Sentinel-1 Dual-Pol (VV and VH) data of 29 cities are shown in Pauli basis, overlapped with the labeled ground truth.

availability. The level-1 Interferometric Wide swath SLC product consists of one image per sub-swath (three-sub-swaths, IW1, IW2, and IW3), per polarization channel (two polarization channels: VH and VV), resulting in six images in total. The properties of different swaths are shown in Table 2 while Table 3 shows the common properties of all sub-swaths.

Table 2. Properties of different sub-swath of Level-1 Interferometric Wide SLC product.

Beam ID	IW 1	IW 2	IW 2
Spatial resolution rg × az m	2.7 × 22.5	3.1 × 22.7	3.5 × 22.6
Pixel spacing rg × az m	2.3 × 14.1	2.3 × 14.1	2.3 × 14.1
Incidence angle	32.9	38.3	43.1

Table 3. General properties that apply to all sub-swaths.

Product ID	IW SLC
Pixel value	Complex
Coordinate system	Slant range
Bits per pixel	16 I and 16 Q
Polarization	VV and VH
Ground range coverage km	251.8
Equivalent number of looks (ENL)	1
Radiometric resolution	3
Number of looks (range x azimuth)	1 × 1

2.4. Data Preparation

Figure 4 presents the flowchart of data preparation for this work. It generally consists of two main parts—data downloading and data preprocessing—which are indicated as orange and blue blocks, respectively.

Regarding data downloading, the Sentinel-1 data set is accessible to any users via the Copernicus Open Access Hub (https://scihub.copernicus.eu/) (also known as the Sentinels Scientific Data Hub). An open-source toolbox, named SentinelSat (https://github.com/sentinelsat/sentinelsat), provides the utilities of searching, downloading, and retrieving the metadata of Sentinel satellite images. An automatic data downloading tool was developed using SentinelSat, which functions based on an ROI file and a given time period of data collection.

For data processing, an ESA toolbox, the Sentinel Application Platform (SNAP, https://step.esa.int/main/toolboxes/snap/), was designed to work with data provided by Sentinel missions. The Sentinel-1 tool box [35] was integrated as a module to deal with all Sentinel-1 data products. The toolbox provides a powerful kit, named the graph processing tool (GPT), which is able to handle large data processing. Based on the GPT, an automatic Sentinel-1 data preprocessing chain was developed in our work so that data could be prepared for the classification task.

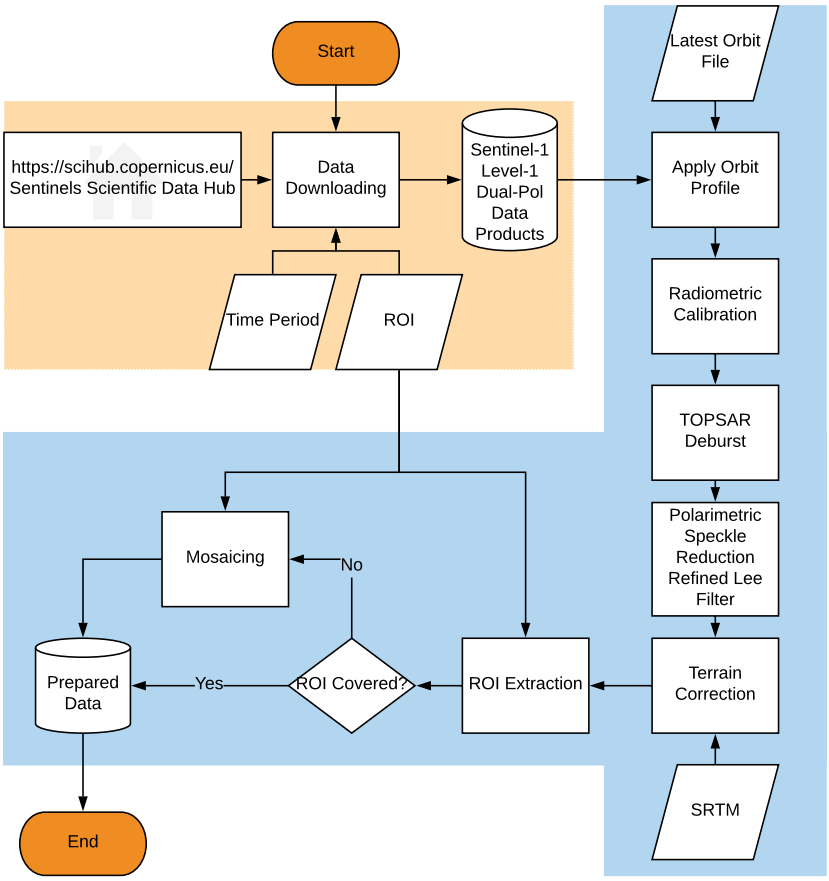


Figure 4. Flowchart of Sentinel-1 data preparation. Module with orange background indicates data downloading part. Module with blue background indicates data preprocessing part.

As shown in the flowchart, a series of data preprocessing modules are applied to Level-1 Sentinel-1 dual-Pol data products by GPT. The functionalities and configurations of these modules are explained in detail as follows:

- Apply Orbit Profile: This module of preprocessing downloads the latest released orbit profile so that a precisely geocoded product can be achieved.
- Radiometric Calibration: Radiometric calibration aims to convert the digital number of the pixel to a radiometrically calibrated backscatter, which is directly related to the radar backscatter of the scene. To extract the relative phase and the correlation between VV and VH, the product of calibration was chosen as a complex valued image.
- TOPSAR Deburst: For each polarization channel, the Sentinel-1 IW product has three swaths. Each swath image consists of a series of bursts. TOPSAR Deburst merges all these bursts and swaths into a single SLC image.
- Polarimetric Speckle Reduction: Speckle reduction was conducted by using the SNAP-integrated refined Lee filter, with a window size of seven by seven [36,37].
- Terrain Correction: Terrain correction eliminates the distortion introduced by the topographical variations. To accomplish the correction, the SRTM was used as the DEM to provide height information. The data was re-sampled to a 10-m GSD by the nearest-neighbor interpolation. The data was geocoded into the WGS84/UTM coordinate system, in which the manually labeled ground truth data was coordinated, so that the ground truth data and Sentinel-1 data could be matched in terms of geo-location.

After the data preprocessing, the analysis-ready dual-Pol data is organized in the common PolSAR covariance matrix. The processed Sentinel-1 dual-Pol data of 29 cities are shown in Figure 3.

3. Methodology

The main building blocks of the proposed global-scale classification approach are feature extraction and classification, which will be detailed in the following subsections.

3.1. Feature Extraction

Following the data preparation described in Section 2, the Sentinel-1 dual-Pol data was processed to the commonly used PolSAR covariance matrix, with a size of two by two. Based on the above mentioned preprocessed dual-Pol data, four different types of features were derived. They were, namely, polarimetric, local statistical, texture, and mathematical morphological feature, which will be described in the following subsections:

3.1.1. Polarimetric

SAR polarimetry allows for the retrieve of shape, orientation, and dielectric property information of scatterers [38,39]. Since there are multiple polarimetric channels, it provides more information than single-pol SAR data. However, the richness of polarimetry is achieved by sacrificing the spatial resolution. To balance the trade-off, instead of a fully polarized SAR, Sentinel-1 mission provides partially polarized SAR data, known as dual-Pol data, with the VV and VH channels. To use the polarimetric information of Sentinel-1 data, we used the intensity of the VH channel ($|S_{VH}|^2$), intensity of VV channel ($|S_{VV}|^2$), normalized coherence of VH and VV $\frac{\langle S_{VH} S_{VV}^* \rangle}{\sqrt{\langle |S_{VH}|^2 \rangle \langle |S_{VV}|^2 \rangle}}$, and relative phase of VH and VV ($actan2(\langle S_{VH} S_{VV}^* \rangle)$) , where S_{VV} and S_{VH} are the complex signals of VV and VH channels, and * denotes complex conjugate. These four features contain essential polarimetric information provided by the dual-Pol data. This polarimetry combination is able to distinguish specular scattering from diffuse scattering [40]. For the purpose of LCZ classification, these features are highly beneficial to differ classes with different surface roughnesses, such as water, plant, building, and soil. We named them as *Pol-Baseline* in our experiments.

3.1.2. Local Statistical

Since the morphological structure of an urban neighborhood is one of the essential factors that local climate zone classes try to describe, it is natural to derive features describing the local neighborhood. It has been shown that simple statistical parameters of a local neighborhood the mean and standard deviation are suitable features for the classification in [13,25]. In our global-scale task, we extracted five statistical parameters: maximum, minimum, mean, standard deviation, and median of local patches. Since the ground sampling distance (GSD) of the LCZ map was suggested to be 100 ms [3,25], the local patch in this work was defined as a size of 10 by 10 pixels, corresponding to the suggested 100 meters GSD. Those parameters were derived from all four polarimetric features (*Pol-Baseline*), resulting in 20 features. We named the local statistical feature the *Stat-Feature*.

3.1.3. Texture

In general, SAR data is well known for containing texture information [27,29,41]. Dual-Pol SAR data is even richer in this regard, simply because it has one more channel. The GLCM is used here to extract texture features. The GLCM describes the distribution of co-occurring values of an image in a given area. It provides a statistical view of texture based on the image histogram [42]. This work extracts the GLCM-based texture information from Sentinel-1 dual-Pol data for LCZ classification. The GLCM statistical features used for describing the distributions include contrast, dissimilarity, homogeneity, angular second moment, maximum probability, entropy, mean, variance, and correlation. For more details of those features, please refer to [42]. Those features can characterize specific characteristics of images, such as homogeneity, contrast, and organized structures presented in an image. For the purpose of LCZ classification, the compactness relates to the spatial distribution of deterministic scatterers in a SAR image, which can be represented by GLCM features. Thus, the

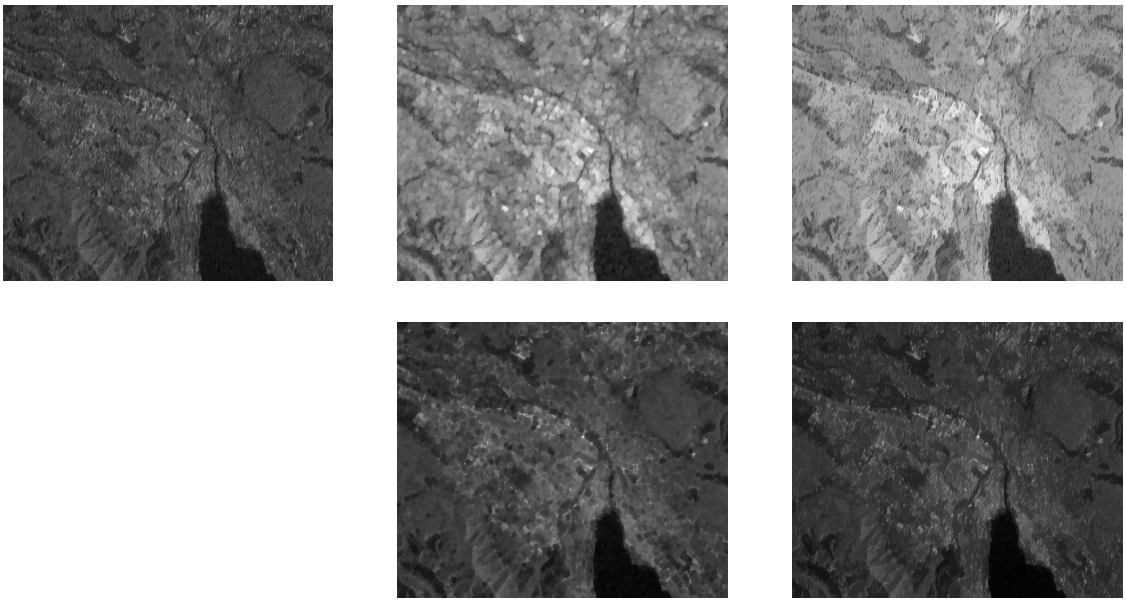
GLCM-based texture features are expected to be beneficial to classify LCZ classes with respect to compactness. To compute the GLCM, it was applied to intensity images of VV and VH channels. For the computational efficiency, the intensity value was quantized into 32 bins. The orientations 0°, 45°, 90° and 135° were chosen. A window size of 11 pixels was chosen. Since the ground sampling distance of the data was 10 ms, the window size suits the 100-m resolution definition of the local climate zone product. The ESA SNAP toolbox was used for GLCM extraction because of its fast implementation. The feature was named the *GLCM-Feature*.

3.1.4. Mathematical Morphological

Spatial-contextual information also plays an important role in LCZ classification. Morphological profiles are regarded as one of the most effective techniques for the extraction of spatial-contextual features and have been intensively used in the remote sensing community for information extraction and scene classification using optical data [33,43–45] and SAR data [27]. Very recently, the advantage of using morphological profiles has become apparent in the application of LCZ classification [8].

The main building blocks of morphological profiles are opening and closing operation [46]. These morphological operators simplify the input gray scale image by removing structures with respect to a predefined structuring element. Structuring elements have a unique structure with a known shape and size (e.g., a disk with a radius of 5 pixels). Therefore, morphological profiles can be produced using a sequence of opening and closing operations, where a structuring element of increasing size applied to a gray scale image to accurately extract spatial features [44]. In [47], an advanced version of opening and closing (opening and closing by reconstruction) was introduced to further improve the ability of the conventional opening and closing operators in terms of information extraction and shape preservation. Opening and closing by reconstructions satisfy the following criterion: If the structuring element cannot fit the structure of the image (objects), then it will be totally removed, otherwise, it will be totally preserved. Reconstruction operators remove objects smaller than structuring element without altering the shape of those objects and reconstruct connected components from the preserved objects. Figure 5 shows an image captured over the city of Zurich along with its corresponding opening, opening by reconstruction, closing and closing by reconstruction.

Figure 5. Morphological opening and closing operations on intensity of VH channel with a radius of 5, for the data of city Zurich. From left to right, top to bottom: VH channel in dB, opening, opening by reconstruction, closing, and closing by reconstruction.



The structuring element has two important parameters: shape and size. Different objects with various surroundings have different degrees of response when considering morphological profiles under different structuring element sizes and shapes. Therefore, morphological profiles can extract the spatial features under different scales and geometric properties. In this paper, the profile was produced by considering the intensity images of VV and VH as inputs. The investigated structuring element in this paper is circular with the diameters set to 1, 3, and 5. We named the morphological feature the *"MP-Feature"*.

3.2. Classifiers

A CCF [48] was chosen to pursue a solution to the task of global local climate zone classification in this work. There are two reasons for this selection: (1) CCF, as a member of random forest methods, is a non-parametric algorithm with a low computation cost, which also provides the function of feature importance analysis; (2) CCF was proven to not only outperform other classifiers in computer vision [48] but also as highly effective in local climate zone classification [25].

To recall a CCF, necessary notations are introduced as follows. Let $X = [x_1, x_2, ..., x_N]^T \in \mathbb{R}^{N \times P}$ denote the training data, with N instances and P features. Let $Y = [y_1, y_2, ..., y_N]^T \in \mathbb{Z}^{N \times C}$ be the label of training data, where C is the number of classes. If y_i, a $C \times 1$-sized vector, indicates that data instance x_i belongs to class two, the second element of vector y_i equals one, and other elements are zero. Accordingly, the training data set is represented as $D = \{X, Y\}$. Similarly, let $\hat{X} = [\hat{x}_1, \hat{x}_2, ..., \hat{x}_N]^T \in \mathbb{R}^{N \times P}$ denote the data for the prediction and $\hat{Y} = [\hat{y}_1, \hat{y}_2, ..., \hat{y}_N]^T \in \mathbb{R}^{N \times C}$ denote the predicted label. Differing from the label in the training data, all elements in predicted label, vector $\hat{y}_i$, ranges in $[0, 1]$, represent the probabilities of every class that $\hat{x}_i$ falls under.

3.2.1. Canonical Correlation Analysis (CCA)

Canonical correlation analysis was designed to analyze the linear relation between sets of variables [49]. Let two data sets be represented as $W \in \mathbb{R}^{k \times a}$ and $V \in \mathbb{R}^{k \times b}$, with k instances and the numbers of features as a and b, respectively. CCA pursues canonical coefficients $P = [p_1, p_2, ..., p_v] \in \mathbb{R}^{a \times v}$ and $Q = [q_1, q_2, ..., q_v] \in \mathbb{R}^{b \times v}$ so that data sets W and V are linearly mapped into a latent space ($WP \in \mathbb{R}^{k \times v}$ and $VQ \in \mathbb{R}^{k \times v}$) where they have maximum correlation. Canonical coefficients are coupled in a pair-wise fashion $\{p_i, q_i\}$ and given by (1).

$$\begin{aligned} & \underset{p_i \in \mathbb{R}^a, q_i \in \mathbb{R}^b}{\text{argmax}} \ corr(Wp_i, Vq_i) & & \\ \textit{subject to} \quad & ||p_i||_2 = 1, ||q_i||_2 = 1 & & \\ & (Wp_i)^T (Wp_j) = 0 & i \neq j & \\ & (Vq_i)^T (Vq_j) = 0 & i \neq j & \\ \textit{where} \quad & i = [1, 2, ..., v] & v = min(rank(W), rank(V)) & \end{aligned} \tag{1}$$

These projected data sets WP and VQ are situated in a v dimensional space. The largest v correlation coefficients associate with the 1^{st} to v^{th} dimension of the space. The solution of the optimization problem (1) is boiled down to a generalized eigenvalue problem [50].

3.2.2. CCFs

As a CCF is an advanced version of random forests, it is introduced by first presenting the random forest and then explaining the improvements made by CCF.

Let $RF = \{t_i\}_{i=1,...,L}$ denote random forests, where a random forest RF consists of L decision trees t_i. An individual decision tree recursively divides a feature space by axis-aligned split until the pure node or stop criteria is achieved. Since a decision tree is a deterministic classifier, training on the same data results in identical trees. To introduce randomness, there are two general strategies among random forest methods. The first one is bagging, where subsets of the whole training data

($X_{sub} \in \mathbb{R}^{N_s \times P}, N_s < N$) are randomly sampled with replacements for training decision trees [51]. The second strategy is to use random subspaces of original data ($X_{sub} \in \mathbb{R}^{N \times P_s}, P_s < P$) to train decision trees [52]. Both strategies decorrelate decision trees and create diversity in predictions. Statistically, diversity encourages a random forest to follow the *Law of Large Numbers* so that it does not suffer from overfitting as individual decision tree [48,53,54]. Most importantly, the diversity is essential to the power of random forest. At the final prediction phase, an averaging of outputs of all trees is applied for the regression task (2).

$$RF(\hat{x}) = \frac{1}{L}\sum_{i=1}^{L} t_i(\hat{x}) \tag{2}$$

For the classification task, besides the majority voting of a unique label, a random forest can also provide probabilities of each class that data instance $\hat{x}$ falls under (3).

$$\begin{aligned} & RF(\hat{x}) = \hat{y} = [pb_1, pb_2, ..., pb_C] \\ \textit{where:} \quad & pb_j = \tfrac{1}{L}\textstyle\sum_{i=1}^{L} t_i = j, j \in [1, 2, ..., C] \end{aligned} \tag{3}$$

Regarding the CCF, there are two key improvements [48]. First, instead of splitting the feature space in axis-aligned manner, it finds the hyperplane in the projected space, where input features $X \in \mathbb{R}^{N \times P}$ and training label $Y \in \mathbb{Z}^{N \times C}$ have maximum correlation. This enables CCF to find natural class boundaries in the feature space instead of restricting it to the axis. The second improvement is called projection bootstrapping. Instead of bagging the training data, it takes in all training data. However, it selects training samples exactly as bagging does to find the CCA projections. Then, all training data are mapped into the canonical correlation space to find hyperplanes. In this fashion, diversity is introduced by bagging-learned CCA projections.

3.2.3. Feature Importance Analysis

To better understand how different features work for our global-scale LCZ classification task, the function of feature importance analysis in a CCF is applied to gain a quantitative insight. Before the analysis, the principle of feature analysis is recalled in this section.

Once a CCF is trained, it predicts the out-of-bag sample or the validation sample. With the ground truth label, one could estimate a prediction error E_{pred}. To achieve the importance analysis of feature P_i, only values of feature P_i are randomly permuted in samples. Therefore, the trained CCF predicts the label of samples with permuted P_i and achieves another prediction error E_{perm}. Hypothetically, E_{pred} should be smaller than E_{perm}. The feature importance indication of P_i is given by $\frac{E_{perm} - E_{pred}}{E_{pred}}$, where the higher the value of this indication, the more important is feature P_i is [53].

4. Experiments and Discussions

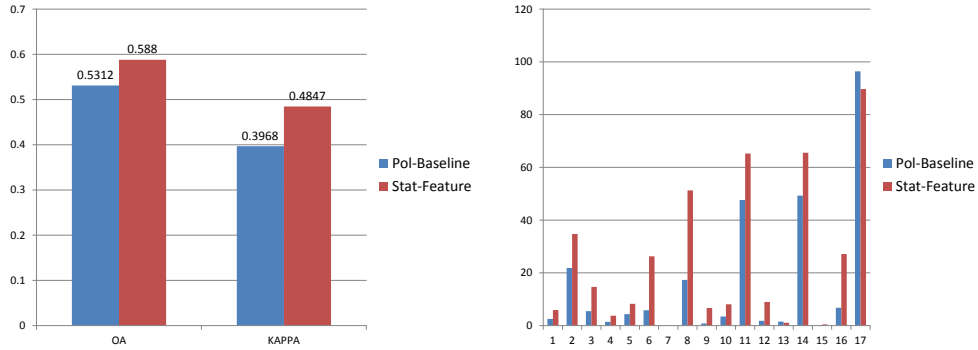


Figure 6. Left figure presents the (OA) and kappa coefficient. Right figure illustrates producer accuracies of all 17 classes. These classes are: 1: Compact high-rise, 2: Compact mid-rise, 3: Compact low-rise, 4: Open high-rise, 5: Open mid-rise, 6: Open low-rise, 7: Light weight low-rise, 8: Large low-rise, 9: Sparsely built, 10: Heavy industry, 11: Dense trees, 12: Scattered trees, 13: Bush, scrub, 14: Low plants, 15: Bare rock or paved, 16: Bare soil or sand, 17: Water.

In this section, the performance of different feature types were analyzed. In this context, firstly, the polarimetric feature and local statistical feature were quantitatively compared, and the one with better performance was later treated as a benchmark feature. Secondly, the chosen benchmark feature was combined with the texture feature and morphological feature, respectively, allowing the performance of the latter two types of features to be compared and analyzed. Afterward the feature importance was analyzed using the CCF. Lastly, the performance of the Sentinel-1 dual-Pol data was discussed regarding the application of local climate zone classification. For all following experiments, the training was based on 20 cities and the testing was conducted on 9 cities, as listed in Table 1. The performance of the classification approach is evaluated based on three quantitative indicators, overall accuracy (OA), kappa coefficient, and average accuracy (AA). OA is simply achieved by dividing the total number of correctly classified samples by the number of overall classified samples, and reported in percentage. Kappa coefficient is a statistical measurement, which estimates inter-rater agreement and is of no unit. AA is the average of all class-specific accuracies.

4.1. Benchmark Feature Selection

The *Pol-Feature* is the identical preprocessed dual-Pol data organized at the pixel level with a ground sampling distance of 10 by 10 ms. The *Stat-Feature* derives straightforward statistical parameters out of a neighborhood of 100 by 100 ms. The extent is identical to the resolution of the targeted local climate zone classification map. Although the *Stat-Feature* involves spatial information, it essentially describes the polarimetric feature as well. Therefore, our first experiment was to analyze performances of both features. The better one, in terms of OA and kappa, was selected as the benchmark feature, which is later combined with the *GLCM-Feature* and the *MP-Feature* for further analysis.

According to Figure 6, the *Stat-Feature* outperformed the *Pol-Feature* by 5.68% and 0.088, in terms of OA and the kappa coefficient, respectively. It also performed better generally on producer accuracies. Considering the challenge of global scale in our work, the difference is quite dramatic, but not surprising. It has been proven in remote sensing that involving spatial information can significantly improve the classification performance [55–57]. Consequently, the *Stat-Feature* was selected as the benchmark feature to work with the other two features.

4.2. Texture Feature

The speckle is omnipresent in SAR images as an intrinsic characteristic [58], which is normally regarded as noise during SAR data interpretation. However, when extracting texture information, the speckle becomes a valuable source containing rich texture information. To quantitatively test to what extent speckle filtering impacts texture, the *GLCM-Feature* was extracted from data sets with (*GLCM-Feature-F*) and without (*GLCM-Feature-UF*) speckle filtering.

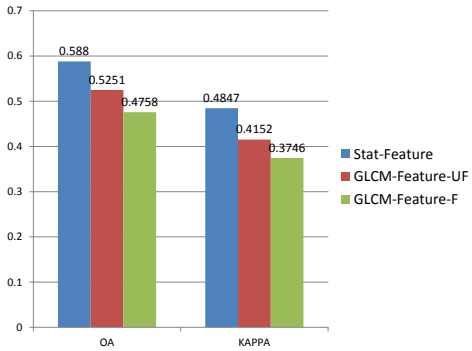


Figure 7. Classification evaluation using OA and kappa coefficient as the evaluation metrics.

To summarize results in Figure 7, firstly, by comparing classification performances of *GLCM-Feature-F* and *GLCM-Feature-UF*, speckle filtering led to a massive loss of texture information. Secondly, the *GLCM-Feature* downgraded the classification performance of the *Stat-Feature*. The reason could be as follows. (1) Originating from the GLCM design, the radiometric resolution is decreased to 32 statistically equalized bins for computational efficiency, thus causing information loss. (2) The equalized bin is statistically decided in the individual data set. For a global scale task, data sets collected from different locations, at different times, with different incident angles would have very diverse intensity responses in imageries. Therefore, the method of GLCM texture extraction is unable to ensure that data sets with the same textures appear the same in the feature space.

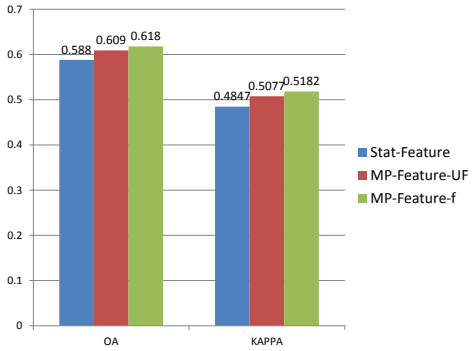


Figure 8. Classification evaluation using OA and Kappa coefficient as the evaluation metrics.

4.3. Morphological Feature

To set up comprehensive experiments, like *GLCM-Feature* extraction, the *MP-Feature* was applied to data sets with *MP-Feature-F* and without *MP-Feature-UF* speckle filtering. Therefore, the performance of the *MP-Feature* could be tested regarding speckle SAR data.

According to results shown in Figure 8, *MP-Feature* improved the classification by 3% in OA. Furthermore there was almost no difference in the performance between *MP-Feature-UF* and *MP-Feature-F*.

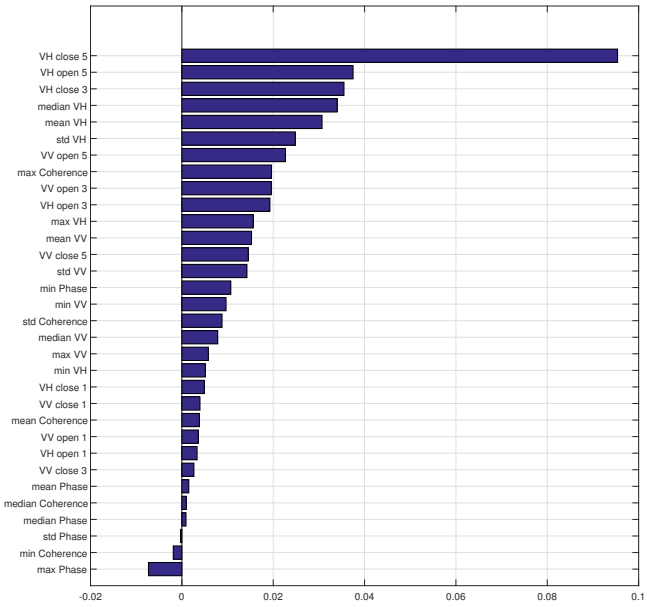


Figure 9. Feature importance obtained by CCF.

4.4. Analysis of Feature Importance

Figure 9 shows the feature importance achieved by canonical correlation forests using all test samples. The empirical importance indication reveals several interesting phenomenons. Firstly, the most important three features are components of the *MP-Feature*. The morphological spatial feature plays a very important role in our global LCZ classification task since LCZ classes describe the morphological property of an urban local neighborhood. Secondly, the VH polarized data contributes the most to the classification because eight out of top ten important features (all top six), are related to VH data. Thirdly, the coherence of VH and VV data also contribute the classification; however, it has often been ignored when using PolSAR data. Lastly, features on the relative phase between VV and VH barely provide any information because they are all among the least important features.

4.5. Class-Wise Analysis

Table 4. Experiments settings regarding feature combinations.

Feature Name	**Feature Combination Code**					
	A	**B**	**C**	**D**	**E**	**F**
Pol-Feature	Y	-	-	-	-	-
Stat-Feature	-	Y	Y	Y	Y	Y
GLCM-Feature-F	-	-	Y	-	-	-
GLCM-Feature-UF	-	-	-	Y	-	-
MP-Feature-F	-	-	-	-	Y	-
MP-Feature-UF	-	-	-	-	-	Y

Table 5. The producer accuracies, OA, and kappa coefficient of the CCF classification approach on different feature combinations (as detailed in Table 4). The number of training and test samples are listed for different classes. The best accuracy for each class is shown in a bold typeface. The metric OA and class-specific accuracy are reported in percentages, and the Kappa coefficient is unitless.

Class	Train	Test	A	B	C	D	E	F
Compact high-rise	4402	2050	2.54	5.9	**14.29**	4.93	6.49	6
Compact mid-rise	21,708	8426	21.84	34.75	**46.24**	31.34	36.11	35.06
Compact low-rise	19,502	21,004	5.5	14.66	12.06	13.97	**14.84**	14.38
Open high-rise	11,683	3185	1.44	3.77	**8.7**	2.35	3.05	2.95
Open mid-rise	17,085	5618	4.34	8.26	**18.08**	10.89	9.08	7.17
Open low-rise	26,126	17,951	5.83	26.27	18.37	19.93	**28.32**	26.64
Light weight low-rise	722	1115	0	0	0	0	0	0
Large low-rise	34,792	17,874	17.33	51.27	49	47.76	**55.68**	54.64
Sparsely built	14,640	6924	0.81	6.69	6.04	2.47	**8.17**	7.32
Heavy industry	9129	5801	3.45	8.08	4.67	4.4	**8.17**	7.74
Dense trees	69,731	43,652	47.64	65.26	53.36	51.39	67.48	**67.51**
Scattered trees	21,926	8938	1.83	8.97	5	5.65	**9.07**	7.56
Bush, scrub	19,396	14,864	1.53	1.08	**3.61**	0.45	1	0.91
Low plants	97,243	35,064	49.31	65.56	56.8	64.29	**69.31**	68.34
Bare rock or paved	6119	3989	0.15	0.45	0.28	**1**	0.45	0.3
Bare soil or sand	78,543	3284	6.76	27.13	**35.99**	5.85	29.75	27.92
Water	309,387	137,753	**96.42**	89.72	68.56	81.7	94.28	93.11
OA			53.12	58.8	47.58	52.51	**61.8**	60.9
KAPPA			0.3968	0.4847	0.3746	0.4152	**0.5182**	0.5077

Table 5 shows class-specific accuracies of different feature combinations. Table 4 gives detailed information about the combinations. In the comparison of feature combinations, class-specific accuracies, OA, and kappa suggest that classification using the *Stat-Feature* and *MP-Feature* derived from filtered data provided the best classification accuracy. By comparing feature combinations **A** and **B**, the *Stat-Feature* improved classification accuracies significantly. Morphological description was also very important for our classification task because adding the *MP-Feature* to the *Stat-Feature* further improved classification accuracies, by comparing combinations **B** and **E**.

Since the task is very challenging in aspects of global scale and complicated LCZ classes, with the CCF strategy, Sentinel-1 dual-Pol data were not able to provide satisfactory classification accuracies.

However, there was valuable information on how Sentinel-1 dual-Pol data could contribute to this task. Despite GLCM is not suitable on extracting textures from multiple cities, the texture features of Sentinel-1 dual-Pol data did excel in distinguishing compactness and height, because the combination **C** provided better accuracies for the first six classes compared to other combinations. By comparing the accuracies of bare soil or sand in combinations **B** and **C**, one can determine that a SAR speckle texture is a good option for classifying this specific class.

4.6. Sentinel-1 Data for LCZ Classification

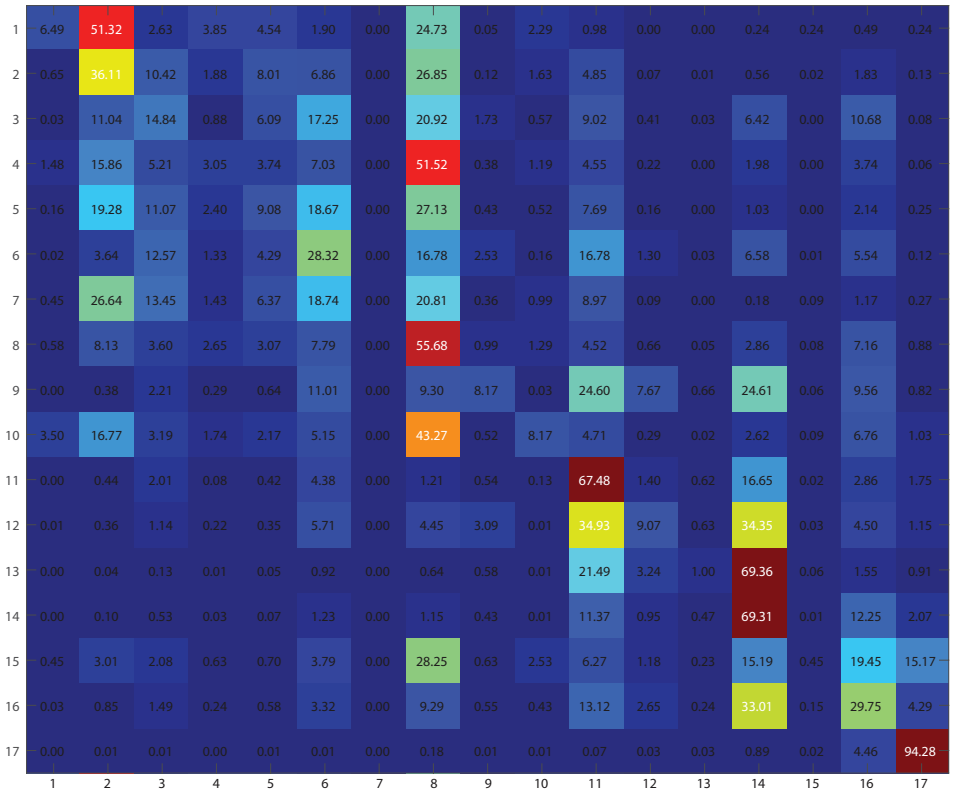


Figure 10. Confusion matrix of the classification framework on feature combination **E**, which is the best feature combination in terms of OA. Numbers are reported in percentages.

According to Table 5 and Figure 10, it is obvious that the classification accuracy is not satisfactory. With reference to the proposed working flow shown in Figure 10, the first ten urban classes are very hard to be classified with the individual use of Sentinel-1 dual-Pol data. The two classes, compact high-rise and the compact mid-rise, are confused with each other. In addition, most urban classes are confused into the class of large low-rise. Eventually only accuracies for the water, low plants, and dense trees are acceptable. However, as it is the first study of LCZ classification using Sentinel-1 dual-Pol data, the present outcome leaves rooms for improvement.

5. Conclusions

This paper describes the first attempt to produce local climate zone classification maps on a global scale using freely accessible Sentinel-1 dual-Pol data sets. The set of features, which were reported as being informative in the literature were assessed for our specific goal. We discovered that the local statistical feature excelled in optical classification and also well with dual-Pol SAR data. However, the GLCM feature, which has often been reported as highly informative in SAR classification, was not suitable for our global scale task due to the lack of a generalization capability since its dependency on a statistical distribution of an individual data set limits its transferability to other data sets. The morphological profile functioned quite well in our task, improving local statistical features, texture features by 3% and 9.29% in terms of OA. Morphological profiles extracted the spatial morphological structure of the data, which suited the essential content of local climate zone classes. According to the feature importance analysis of the CCF classifier, the VH polarized data had the biggest contribution to our classification task. The often-ignored feature, the coherence of VV and VH, contributed more to the classification compared to contributions of the VV polarised data. Moreover, the relative phase

of VV and VH polarized data barely provided informative content for classifications, even dragging down the performance.

By far, classification accuracies of the Sentinel-1 data are not very appealing. However, considering the challenges of a global scale and transferability in our task, Sentinel-1 data still contributes in different manners for certain classes. The ensemble CCF strategy in this paper was chosen for the classification step due to its high generalization ability and superior performance over deep learning-based classifiers [8]. One potential reason for why CCF outperformed a deep learning-based classification method in [8] may have been due to the limited number of training samples. However, one deficiency of the CCF is that it demands hand-designed extracted features to further boost its classification performance. While endless combinations of features that can be fed to a CCF exist, this paper took the very first steps to evaluate the most informative sets of features and critically compare their performance for a global LCZ classification.

In the future, we intend to further examine this challenging task by investigating a deep learning strategy in order to take advantage of its capability of automatic feature extraction and selection. No matter which strategy is under consideration, one more critical detail should also be studied is that, how large is the neighborhood in remote sensing data is optimal for classifying LCZ classes. Regarding the data source, we also intend to investigate using both radar and optical remote sensing data, so that data fusion might contribute to this task. Last but not the least, we are interested to develop a solution to our global scale task, which has a strong capability on transferring generalization.

Author Contributions: J.H. was responsible for the research design, the data preparation, the experiments, the analysis, and the manuscript. P.G. was responsible for the research design, the analysis, and the manuscript. X.Z. was responsible for the conceptual framework, the research design, and the analysis.

Funding: This research is jointly supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. [ERC-2016-StG-714087], Acronym: *So2Sat*), Helmholtz Association under the framework of the Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de), and the Bavarian Academy of Sciences and Humanities in the framework of Junges Kolleg.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Stewart, I.D.; Oke, T.R. Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society* **2012**, *93*, 1879–1900.
2. Stewart, I.D.; Oke, T.R.; Krayenhoff, E.S. Evaluation of the 'local climate zone'scheme using temperature observations and model simulations. *International Journal of Climatology* **2014**, *34*, 1062–1080.
3. Bechtel, B.; Alexander, P.J.; Böhner, J.; Ching, J.; Conrad, O.; Feddema, J.; Mills, G.; See, L.; Stewart, I. Mapping local climate zones for a worldwide database of the form and function of cities. *ISPRS International Journal of Geo-Information* **2015**, *4*, 199–219.
4. WUDAPT. http://www.wudapt.org/. Accessed: 2018-04-10.
5. See, L.; Comber, A.; Salk, C.; Fritz, S.; van der Velde, M.; Perger, C.; Schill, C.; McCallum, I.; Kraxner, F.; Obersteiner, M. Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts. *PLoS One* **2013**, *e69958*.
6. Foody, G.M.; See, L.; van der Velde, M.; Perger, C.; Schill, C.; Boyd, D.S. Assessing the Accuracy of Volunteered Geographic Information arising from Multiple Contributors to an Internet Based Collaborative Project. *Transactions in GIS* **2013**, *17*, 847–860.
7. Bayas, J.C.L.; See, L.; Fritz, S.; Sturn, T.; Perger, C.; Dürauer, M.; Karner, .; Moorthy, I.; Schepaschenko, D.; Domian, D.; McCallum, I. Crowdsourcing In-Situ Data on Land Cover and Land Use Using Gamification and Mobile Technology. *Remote Sensing* **2016**, *8*, 905.
8. Yokoya, N.; Ghamisi, P.; Xia, J.; Sukhanov, S.; Heremans, R.; Tankoyeu, I.; Bechtel, B.; Saux, B.L.; Moser, G.; Tuia, D. Open Data for Global Multimodal Land Use Classification: Outcome of the 2017 IEEE GRSS Data

Fusion Contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2018**, pp. 1–15. doi:10.1109/JSTARS.2018.2799698.

9. Liaw, A.; Wiener, M.; others. Classification and regression by randomForest. *R news* **2002**, *2*, 18–22.
10. Bechtel, B. Multitemporal Landsat data for urban heat island assessment and classification of local climate zones. Urban Remote Sensing Event (JURSE), 2011 Joint. IEEE, 2011, pp. 129–132.
11. Danylo, O.; See, L.; Bechtel, B.; Schepaschenko, D.; Fritz, S. Contributing to WUDAPT: A local climate zone classification of two cities in Ukraine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2016**, *9*, 1841–1853.
12. Xu, Y.; Ren, C.; Cai, M.; Edward, N.Y.Y.; Wu, T. Classification of local climate zones using ASTER and Landsat data for high-density cities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2017**, *10*, 3397–3405.
13. Bechtel, B.; Daneke, C. Classification of local climate zones based on multiple earth observation data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2012**, *5*, 1191–1202.
14. Xu, Z.; Chen, J.; Xia, J.; Du, P.; Zheng, H.; Gan, L. Multisource Earth Observation Data for Land-Cover Classification Using Random Forest. *IEEE Geoscience and Remote Sensing Letters* **2018**.
15. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intelligent Systems and their applications* **1998**, *13*, 18–28.
16. Lelovics, E.; Unger, J.; Gál, T.; Gál, C.V. Design of an urban monitoring network based on Local Climate Zone mapping and temperature pattern modelling. *Climate research* **2014**, *60*, 51–62.
17. Gál, T.; Bechtel, B.; Unger, J. Comparison of two different Local Climate Zone mapping methods **2015**.
18. Geletič, J.; Lehnert, M. GIS-based delineation of local climate zones: The case of medium-sized Central European cities. *Moravian Geographical Reports* **2016**, *24*, 2–12.
19. Bechtel, B.; See, L.; Mills, G.; Foley, M. Classification of local climate zones using SAR and multispectral data in an arid environment. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2016**, *9*, 3097–3105.
20. Kaloustian, N.; Tamminga, M.; Bechtel, B. Local climate zones and annual surface thermal response in a Mediterranean city. 2017 Joint Urban Remote Sensing Event (JURSE), 2017, pp. 1–4. doi:10.1109/JURSE.2017.7924597.
21. Hu, J.; Mou, L.; Schmitt, A.; Zhu, X.X. FusioNet: A two-stream convolutional neural network for urban scene classification using PolSAR and hyperspectral data. Urban Remote Sensing Event (JURSE), 2017 Joint. IEEE, 2017, pp. 1–4.
22. Mou, L.; Ghamisi, P.; Zhu, X.X. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 3639–3655.
23. Mou, L.; Ghamisi, P.; Zhu, X.X. Unsupervised spectral-spatial feature learning via deep residual conv-deconv network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2018**.
24. Kang, J.; Körner, M.; Wang, Y.; Taubenböck, H.; Zhu, X.X. Building instance classification using street view images. *ISPRS Journal of Photogrammetry and Remote Sensing* **2018**.
25. Yokoya, N.; Ghamisi, P.; Xia, J. Multimodal, multitemporal, and multisource global data fusion for local climate zones classification based on ensemble learning. Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International. IEEE, 2017, pp. 1197–1200.
26. Wurm, M.; Taubenböck, H.; Weigand, M.; Schmitt, A. Slum mapping in polarimetric SAR data using spatial features. *Remote sensing of environment* **2017**, *194*, 190–204.
27. Zhu, Z.; Woodcock, C.E.; Rogan, J.; Kellndorfer, J. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. *Remote Sensing of Environment* **2012**, *117*, 72–82.
28. Ban, Y.; Jacob, A.; Gamba, P. Spaceborne SAR data for global urban mapping at 30 m resolution using a robust urban extractor. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *103*, 28–37.
29. Geng, J.; Fan, J.; Wang, H.; Ma, X.; Li, B.; Chen, F. High-resolution SAR image classification via deep convolutional autoencoders. *IEEE Geoscience and Remote Sensing Letters* **2015**, *12*, 2351–2355.
30. Du, P.; Samat, A.; Waske, B.; Liu, S.; Li, Z. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *105*, 38–53.

31. Dalla Mura, M.; Atli Benediktsson, J.; Waske, B.; Bruzzone, L. Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *International Journal of Remote Sensing* **2010**, *31*, 5975–5991.
32. Plaza, A.; Benediktsson, J.A.; Boardman, J.W.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A.; others. Recent advances in techniques for hyperspectral image processing. *Remote sensing of environment* **2009**, *113*, S110–S122.
33. Ghamisi, P.; Dalla Mura, M.; Benediktsson, J.A. A survey on spectral–spatial classification techniques based on attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing* **2015**, *53*, 2335–2353.
34. United Nations, The World's Cities in 2016. http://www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2016_data_booklet.pdf. Accessed: 2018-04-10.
35. Veci, L.; Lu, J.; Foumelis, M.; Engdahl, M. ESA's Multi-mission Sentinel-1 Toolbox. EGU General Assembly Conference Abstracts, 2017, Vol. 19, p. 19398.
36. Lee, J.S.; Grunes, M.R.; De Grandi, G. Polarimetric SAR speckle filtering and its implication for classification. *IEEE Transactions on Geoscience and remote sensing* **1999**, *37*, 2363–2373.
37. Lee, J.; Ainsworth, T.L.; Wang, Y.; Chen, K. Polarimetric SAR Speckle Filtering and the Extended Sigma Filter. *IEEE Transactions on Geoscience and Remote Sensing* **2015**, *53*, 1150–1160. doi:10.1109/TGRS.2014.2335114.
38. Lee, J.S.; Pottier, E. *Polarimetric radar imaging: from basics to applications*; CRC press, 2009.
39. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine* **2013**, *1*, 6–43.
40. Schmitt, A.; Wendleder, A.; Hinz, S. The Kennaugh element framework for multi-scale, multi-polarized, multi-temporal and multi-frequency SAR image preparation. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *102*, 122–139.
41. Yokoya, N. Texture-Guided Multisensor Superresolution for Remotely Sensed Images. *Remote Sensing* **2017**, *9*, 316.
42. Haralick, R.M.; Shanmugam, K.; others. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* **1973**, pp. 610–621.
43. Benediktsson, J.A.; Arnason, K.; Pesaresi, M. The use of morphological profiles in classification of data from urban areas. Remote Sensing and Data Fusion over Urban Areas, IEEE/ISPRS Joint Workshop 2001. IEEE, 2001, pp. 30–34.
44. Benediktsson, J.A.; Palmason, J.A.; Sveinsson, J.R. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing* **2005**, *43*, 480–491.
45. Benediktsson, J.; Ghamisi, P. *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*; 2015.
46. Fauvel, M.; Benediktsson, J.A.; Chanussot, J.; Sveinsson, J.R. Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles. *IEEE Transactions on Geoscience and Remote Sensing* **2008**, *46*, 3804–3814. doi:10.1109/TGRS.2008.922034.
47. Crespo, J.; Serra, J.; Schafer, R.W. Theoretical aspects of morphological filters by reconstruction. *Signal processing* **1995**, *47*, 201–225.
48. Rainforth, T.; Wood, F. Canonical correlation forests. *arXiv preprint arXiv:1507.05444* **2015**.
49. Hotelling, H. Relations between two sets of variates. *Biometrika* **1936**, *28*, 321–377.
50. Xia, J.; Yokoya, N.; Iwasaki, A. Hyperspectral image classification with canonical correlation forests. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 421–431.
51. Breiman, L. Bagging predictors. *Machine learning* **1996**, *24*, 123–140.
52. Ho, T.K. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence* **1998**, *20*, 832–844.
53. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
54. Elghazel, H.; Aussem, A.; Perraud, F. Trading-off diversity and accuracy for optimal ensemble tree selection in random forests. In *Ensembles in Machine Learning Applications*; Springer, 2011; pp. 169–179.
55. Ghamisi, P.; Benediktsson, J.A.; Ulfarsson, M.O. Spectral–spatial classification of hyperspectral images based on hidden Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing* **2014**, *52*, 2565–2574.

56. Ghamisi, P.; Benediktsson, J.A.; Sveinsson, J.R. Automatic Spectral-Spatial Classification Framework Based on Attribute Profiles and Supervised Feature Extraction. *IEEE Trans. Geoscience and Remote Sensing* **2014**, *52*, 5771–5782.
57. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.A.; Chanussot, J.; Tilton, J.C. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE* **2013**, *101*, 652–675.
58. Hu, J.; Guo, R.; Zhu, X.; Baier, G.; Wang, Y. Non-local means filter for polarimetric SAR speckle reduction-experiments using TerraSAR-x data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2015**, *2*, 71.

Appendices

B Hu J, Hong D, and Zhu X X (2019). MIMA: MAPPER-Induced Manifold Alignment for Semi-Supervised Fusion of Optical Image and Polarimetric SAR Data. IEEE Transactions on Geoscience and Remote Sensing. doi: 10.1109/TGRS.2019.2924113

https://ieeexplore.ieee.org/document/8802291

MIMA: MAPPER-Induced Manifold Alignment for Semi-Supervised Fusion of Optical Image and Polarimetric SAR Data

Jingliang Hu, *Student Member, IEEE,* Danfeng Hong, *Student Member, IEEE,* and Xiao Xiang Zhu, *Senior Member, IEEE*

***Abstract*—Multi-modal data fusion has recently been shown promise in classification tasks in remote sensing. Optical data and radar data, two important yet intrinsically different data sources, are attracting more and more attention for potential data fusion. It is already widely known that, a machine learning based methodology often yields excellent performance. However, the methodology relies on a large training set, which is very expensive to achieve in remote sensing. The semi-supervised manifold alignment (SSMA), a multi-modal data fusion algorithm, has been designed to amplify the impact of an existing training set by linking labeled data to unlabeled data via unsupervised techniques. In this paper, we explore the potential of SSMA in fusing optical data and polarimetric SAR data, which are multi-sensory data sources. Furthermore, we propose a MAPPER-induced manifold alignment (MIMA) for semi-supervised fusion of multi-sensory data sources. Our proposed method unites SSMA with MAPPER, which is developed from the emerging topological data analysis (TDA) field. To our best knowledge, this is the first time that SSMA has been applied on fusing optical data and SAR data, and also the first time that TDA has been applied in remote sensing. The conventional SSMA derives a topological structure using k-nearest-neighbor (*kNN*), while MIMA employs MAPPER, which considers the field knowledge and derives a novel topological structure through the spectral clustering in a data-driven fashion. Experiment results on data fusion with respect to land cover land use classification and local climate zone classification suggest superior performance of MIMA.**

***Index Terms*—Hyperspectral image, MAPPER, multi-modal data fusion, multi-sensory data fusion, multispectral image, PolSAR, semi-supervised manifold alignment (SSMA), topological data analysis (TDA).**

I. Introduction

In recent decades, data fusion has attracted a lot of attention in the remote sensing community [1], [2], [3], [4], motivated by the simple fact that multiple data sources reveal complementary physical properties of observed scenes. For example, optical RGB data normally possesses high spatial resolution [5], while multi/hyperspectral data contains spectral information [6], and synthetic aperture radar (SAR) data gives dialectic and geometric properties. Thus, it is valuable to develop algorithms that are able to take advantage of different data sources for applications. In this regard, machine learning techniques are becoming increasingly important due to their excellent performance [7], [8], [9]. As is generally known in machine learning, the training data set is of great importance [10]. Most successful techniques require a large set of training data [11]. However, accessing a large training data set is very expensive, especially in remote sensing, because labeling a training data set in this field requires expertise that is more complicated than identifying dogs and cats. Therefore, a semi-supervised learning technique is a good option for remote sensing tasks, as the unlabeled data set is linked to the training data set by unsupervised approaches in the learning of the technique. It amplifies the effect of the existing training data set. Considering the importance of data fusion and precious training data, this paper studies a semi-supervised learning technique, named manifold alignment, to fuse optical image and polarimetric SAR (PolSAR) data for the purpose of classification.

The authors are with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany, and also with Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), 80333 Munich, Germany (e-mails: jingliang.hu@dlr.de; danfeng.hong@dlr.de; xiaoxiang.zhu@dlr.de).

This work is jointly supported bythe German Research Foundation (DFG) under grant ZH 498/7-2, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. [ERC-2016-StG-714087], Acronym: *So2Sat*), and the Helmholtz Association under the framework of the Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de)

(*Corresponding author: Xiao Xiang Zhu.*)

A. Fusion of optical and SAR data

Due to the rapid development of remote sensing missions such as LandSat-8, Sentinel-2, EnMAP for optical remote sensing and TerreSAR-X, Tandem-X, Sentinel-1 for radar remote sensing, a huge amount of optical data and SAR data have been collected; the data volume can be expected to increase over time. The fusion of the two data sets hold great potential for use in various applications [12]. Besides the data availability, the other reason to fuse them is that dialectic and geometric properties provided by SAR data are complementary to the spectral information of optical data. However, fusing them in practice is not as straightforward as the argument for doing it. The difficulty lies in the intrinsic differences in their imaging geometry. Because of the slanted looking angle of the SAR sensor, SAR images have an oblique appearance, with distortions of foreshortening, shadowing, and layover. This results in image geometry that is severely dissimilar to the nadir looking optical data. The extent of SAR distortions is positively correlated to height. This will pose substantial challenges when fusing these two data sets, especially in urban areas with large height fluctuations. To

date, some studies have explored fusing these two data sources. We categorize those studies into three types based on their purposes: (1) registration oriented, (2) detection oriented, and (3) classification oriented.

1) Registration is actually a prerequisite of any further fusion. However, precise registration of SAR and optical image is very challenging due to geometric differences. A conditional generative adversarial network [13] was trained to generate an artificial SAR image given a real-world optical image, and the optical image and SAR data are then registered by matching the artificial SAR data with the real-world SAR data. This technique was shown to be effective in a suburban area. A 3D registration is introduced in [14] to align optical and SAR data by imitating the physical procedure of optical and SAR imaging based on a digital surface model. A pseudo-siamese convolutional neural network architecture [15] was trained to identify corresponding optical and SAR data in image patches and showed promising preliminary results. By far, although progress has been made in recent years, precise SAR and optical data registration has not achieved a robust solution yet, especially for complex urban areas. Thus, for other purposes of fusing these two data sources, the straightforward approach is registration by geographic coordinates.

2) Detection tasks have been proven successful by using optical and SAR data for the purpose of detecting building outlines [16], crops [17], water [18], [19] and urban areas [20]. Since the detection task focuses on specific targets, studies extract representation of those targets from each of the two data sources so that they can work together to identify targets. For example, for detecting crops, optical data provides spectral signatures and SAR provides scattering mechanisms of interested targets; these characteristics are extracted and used together to identify the target under detection.

3) Classification is more challenging than detection tasks for more than one class of interest is under consideration. This paper focuses on this challenges [21]. Recently, a number of studies [22], [23], [24], [25] have tried to solve classification tasks by using both optical and SAR data. In general, these fusing strategies all extracts features from the individual data set, then concatenate all the features and feed them into various classifiers. The most important part of this procedure is to extract hand-crafted informative features [26] regarding classification. A two-stream convolutional neural network (CNN) [27] derives high level features of individual data sets by utilizing the power of CNN and then concatenates those features for classification. In brief, concatenation is the main strategy for fusing SAR and optical data so far, which is an effective and straightforward approach.

B. Semi-supervised manifold alignment (SSMA)

SSMA pursues a projection for each input data source, and maps corresponding data into a shared latent space [28], [29]. These properties hold within this space: ***(a)*** data of the same classes locate close to each other; ***(b)*** data of different classes locate far from each other; ***(c)*** the topological property of individual data is preserved. These three properties make SSMA to be a promising candidate for our task from a methodological perspective for two reasons. First, the first two properties promote classification-wise advantageous information from any data source to be used. Second, the final property implicitly connects unlabeled data to training data, which amplifies the functionality of the training data. These two factors meet our need for an algorithm that fuses data sets with the maximum usage of the training data.

In the remote sensing community, SSMA has been investigated for various applications. It was applied to fuse an RGB image and hyperspectral image so that visualization of hyperspectral image could be achieved in the latent space, exhibiting more spectral information than conventional visualization methods in [30]. A kernel manifold alignment was introduced in [31] to fuse multiple optical remote sensing data into a latent space by nonlinear projections for a classification task. Manifold alignment was also used in [32] to align spectral signatures from different optical data sets by projecting them into a latent space so that object detection was achieved.

In regard to remote sensing data fusion, different data sources observe the same region of interest. Essentially, the observed target is a single object that appears differently in data sources due to sensor specifications. Thus, this question arises. Although, theoretically, SSMA is a good choice, does one latent space of observed objects, where data sources can be aligned? If it exists, can we find that space by using SSMA? Tuia et al. [33] applied SSMA to find the underlying space of multiple optical data sets under three scenarios: different looking angles, multi-temporal, and different sensors. In this work, we aim to fuse multi-sensory data sets, namely optical image and polarimetric SAR data, by SSMA.

C. Topology and MAPPER

One important feature of SSMA resides its exploration of the topological structure of data. The conventional manifold based method [28], [29], [34], [35], [36] approximates topological properties by using the *kNN*. They essentially assume that the underlying manifold of a data is a Riemannian manifold which can be locally approximated by Euclidean measurement [37], [38], [39]. Recently, topological data analysis (TDA) has emerged as a new mathematical sub-field of big data analysis, by means of studying topological properties in the data [40], [41], [42], [43]. One TDA tool, named MAPPER, resolves a computable approximation of the Reeb graph which represents the topological structure of a data with respect to one interested intrinsic property of the data [44], [45].

A general explanation of topology is that it is an art of simplification. It ignores complex information of the object under studying, and rather focuses on one meaningful aspect of it. On this regard, conventional manifold methods focus on the aspect of the local connection or the local structure. On the side of MAPPER (Reeb graph), it focuses on the topological structure of data related to the interested intrinsic property.

In real applications, the MAPPER has been proven capable of revealing unknown knowledge in medical studies, by interpreting topological structures of data sets. The tool was applied to analyze breast cancer transcriptional data and uncovered a

sub-group of Estrogen Receptor-positive (ER+) breast cancers. Patients suffering this kind of cancer exhibit 100% survival and no metastasis. This finding was previously unknown and is invaluable for future treatment [46]. MAPPER was also applied to analyze the data of preclinical traumatic brain injury (TBI) and spinal cord injury (SCI). It revealed a previously unknown pattern of co-occurring TBI and SCI, as well as a previously unknown harmful effect of an experimental drug treatment [47]. With the help of MAPPER, Li et al. [48] explored complex medical records of type 2 diabetes (T2D) patients and revealed previously unknown sub-groups within T2D. All the above discoveries are invaluable and contribute to greater precision in the practice of medicine.

Besides the inspiration of these successful studies in medicine and the sound theoretical foundation of the MAPPER [44], the other reasons which motivated the authors to utilize the MAPPER to explore the topological structure of the remote sensing data are listed as following:

1) The field knowledge The MAPPER focuses on the topological structure of data related to an intrinsic property. In practice, the intrinsic property is quantitatively derived from the data by an expert-designed filter function. The quantified property operates as a lens through which the MAPPER observes the data and extracts the topological structure of the data. Therefore, the choice of the lens, equally the filter function, introduces a field knowledge into the procedure of extracting topological structure. To our best knowledge, the ability of extracting topological structure from a field-knowledge perspective is unprecedented for the manifold-related technique in remote sensing.

2) The regional-to-global topological structure Instead of focusing on local structures of data points in conventional manifold-based techniques, the MAPPER focuses on an intrinsic property introduced by the filter function. Under the guidance of the filtered values, MAPPER divides a data into several bins, derives topological structure of each bin, and collects those structures together as a global one. This results in a regional-to-global topological structure. For the complex remote sensing data, especially SAR data, the regional derived structure is more robust to outliers than the local derived one.

3) The data-driven and optimized topology The spectral clustering is embedded into MAPPER in this work, leading to a data-driven and optimized topological structure.

- A data-driven topology. The eigen-gap concept in the spectral clustering detects the number of clusters [49]. This ensures the derived topological structure suits the distribution of the data. Rather, conventional techniques derive the topological structure of the whole data set, with the *kNN* of a fixed k [50].
- An optimized topology. The spectral clustering is an optimized graph-cut algorithm, which is capable of unbiased grouping [51], [52]. However, a conventional manifold technique directly relies on the precision of the similarity measurement. Although sophisticated similarity measurements have been developed in remote sensing, the high dimensionality and the complexity of the data still pose challenges on the measurement.

D. Summary

The contributions of this paper are three-fold.

- This work studies the fusion of heterogeneous remote sensing data sources, namely, the optical data and the polarimetric SAR data, with the semi-supervised manifold alignment technology.
- To our best knowledge, this is the first time that the topological data analysis (TDA) technique has been applied in the remote sensing community.
- A novel MAPPER-induced manifold alignment is proposed for semi-supervised data fusion. Its performance on the fusion of polarimetric data and optical data regarding classifications is quantitatively analyzed.

The remainder of this paper is organized as follows. In Section II, MAPPER and SSMA are reviewed, and MIMA is introduced. The experimental setup, results, and comparisons are provided in Section III. Finally, Section IV provides conclusions and remarks on the work.

II. Methodology

In this section, we first introduce the background of the topological data analysis tool called MAPPER. Then, we review the basics of semi-supervised manifold alignment. Finally, the novel MIMA is introduced.

A. MAPPER

In order to introduce MAPPER [44] in a comprehensive and understandable way, we first provide an intuitive example, shown in Fig. 1. The theoretical foundation of MAPPER is then introduced from the perspective of applied topology. Due to heavy reliance on mathematical concepts for this paper, we note that notations in section II-A represent separate meaning from the other notations in the rest of the paper.

1) Intuitive explanation MAPPER is a mathematical tool developed from applied topology to analyze and visualize big data sets [44]. The algorithm essentially consists of three components:

- Filter function selection. MAPPER first requires a filter function which derives a filtering space where the interested intrinsic property is quantified. The chosen filter function should reveal physical meaning or geometric property of the data. It allows a specialist to introduce field knowledge into data analysis. For the example shown in Fig. 1, the filter function is chosen as the distance to the wrist, so that a filtering space in Fig. 1 (B) is derived from the data point cloud.
- Data separation. In the filtering space, the continuous value range is sliced into overlapped intervals with a given overlap percentage and number of intervals, shown in Fig. 1 (C). Guided by the overlapped intervals [54], the original input data can be separated into overlapped data bins accordingly, as shown in Fig. 1 (D). The separated data in bins have the same dimension as the original data.
- Clustering and visualization construction. Clustering is applied on each data bin. Clusters of adjacent data bins might include common data points. MAPPER constructs

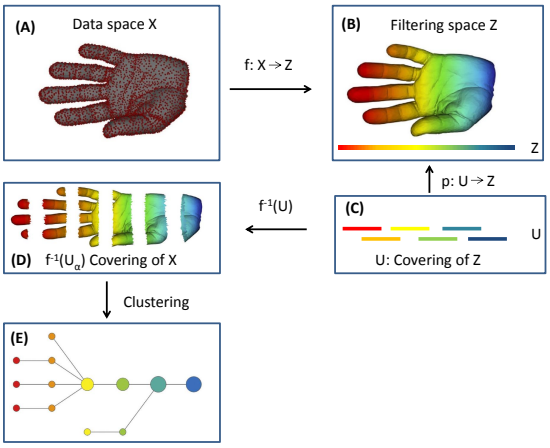


Fig. 1: Example of MAPPER approach to derive the topological structure of the point cloud of a human hand. (A): Data space X, point cloud data of a human hand; (B): Filtered space Z, points colorized by the filter value; Filter function f: assigning data points with their horizontal distances to the right end; (C): U covering of Z, overlapped intervals of the filtered value; (D): $f^{-1}(U_\alpha)$ covering of X, separating original data into bins according to intervals in (C), data in bins remain their original dimension; (E) $f^{-1}(U_\alpha)$ covering of X, achieved by clustering bins of data. Modified from [53].

a graph where a node represents a cluster, and an edge represents a link of two clusters. The link is generated for two clusters if they share common data points. Therefore, the graph serves as a simplified visualization of the topological structure of a data set. For example, the graph in Fig. 1 (E) is derived by MAPPER to represent the topological structure of the point cloud data of a human hand.

It is worth to highlight that the filter function is not seen as a dimension reduction, but quantifies a filtered space which guides the separation of the original data. As mentioned above, topology is an art of simplification. The conventional manifold learning focuses on the local structure of individual data points. On the other hand, MAPPER derives the topological structure of data while focusing on the property quantified by the filter function.

2) Theoretical foundation First, it is necessary to introduce the concept of covering in topology. In [55], it is explained as: let $p : U \rightarrow Z$ be continuous and surjective. If every point z of Z has a neighborhood C that is evenly covered by p, then p is called a covering map, and U is defined to be a covering space of Z, then p is a local homeomorphism of U with Z. It means that, in terms of function p, the preimage in U and the image in Z share the same topological properties locally.

The rest of the theoretical foundation is introduced in blocks corresponding to the three components of the MAPPER.

- Filter function selection. According to MAPPER [44], data is situated in a topological space X as illustrated in Fig. 1 (A). A continuous function $f : X \rightarrow Z$ projects space X to another space Z, as shown in Fig. 1 (B).
- Data separation. The space Z is equipped with a covering space U, as shown in Fig. 1 (C). Assuming that covering space U is a k-simplex spanned by a set $\{\alpha_1, \alpha_2, ..., \alpha_k\}$ so that $U = \{U_\alpha\}$, since f is continuous, $f^{-1}(U_\alpha)$ forms a covering of space X and could be used to represent topological space X of given data , as shown in Fig. 1 (D).
- Clustering and visualization construction. The set $\{\alpha_1, \alpha_2, ..., \alpha_k\}$, as the vertices of k-simplex, are k connected components in topological space X which can be achieved by clustering. Thus, $f^{-1}(U_\alpha)$ is achieved to represent data space X, as shown in Fig 1 (E).

B. Semi-supervised manifold alignment (SSMA)

Let $\mathbf{X}_i = [\mathbf{x}_i^1, ..., \mathbf{x}_i^k, ..., \mathbf{x}_i^{n_i}] \in \mathbb{R}^{m_i \times n_i}$ be a matrix representing the i^{th} data source, with m_i dimensions by n_i instances. The term $\mathbf{x}_i^k$ denotes the k^{th} instance of the i^{th} data source. Let K denote the total number of data sources. SSMA learns a set of K projections $\{f_1, ..., f_K\}$. The i^{th} projection f_i maps the i^{th} data source $\mathbf{X}_i$ into the latent space, where all the K data sources are aligned in terms of the three desired properties discussed in the Introduction. The properties are formulated by three matrices, called the similarity matrix, dissimilarity matrix, and topology matrix. More specifically, the similarity matrix (1) is computed by labeled information to pursue property ***(a)***: the data of same class located close to each other.

$$W_s = \begin{pmatrix} W_s^{1,1} & ... & W_s^{1,K} \\ ... & ... & ... \\ W_s^{K,1} & ... & W_s^{K,K} \end{pmatrix} \tag{1}$$

The dissimilarity matrix is formed as (2) to accomplish property ***(b)***: data of different classes located far from one another.

$$W_d = \begin{pmatrix} W_d^{1,1} & ... & W_d^{1,K} \\ ... & ... & ... \\ W_d^{K,1} & ... & W_d^{K,K} \end{pmatrix} \tag{2}$$

The topology matrix (3) describes the topological structure of the data, which aims at the property ***(c)***: the topological property of individual data is preserved.

$$W_t = \begin{pmatrix} W_t^{1,1} & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & W_t^{K,K} \end{pmatrix} \tag{3}$$

Each of the matrices (1), (2), and (3) is a matrix with the size $(n_1+n_2+...+n_k) \times (n_1+n_2+...+n_k)$. In each matrix, the $W^{i,j}$ is a matrix representing the relationship between the i^{th} and j^{th} data sources on the individual property.

Similarity matrix W_s and dissimilarity matrix W_d are generated based on label information. If x_i^p and x_j^q share a same label, then $W_s^{i,j}(p,q) = 1$, otherwise $W_s^{i,j}(p,q) = 0$. If x_i^p and x_j^q belong to different classes, then $W_d^{i,j}(p,q) = 1$, otherwise $W_s^{i,j}(p,q) = 0$.

Since the topological structure of the individual data set is preserved, the matrix W_t is a block-wise diagonal matrix. The topological structure is conventionally given by the *kNN*,

which means $W_t^{i,i}(p,q) = 1$ if x_i^p and x_i^q are neighbors in a given *kNN* neighborhood. Otherwise, $W_t^{i,i}(p,q) = 0$.

In order to simultaneously model the three properties of the latent space, three terms are formulated for the cost function:

$$A = \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{p=1}^{n_i}\sum_{q=1}^{n_j} \|f_i^T x_i^p - f_j^T x_j^q\|^2 W_s^{i,j}(p,q). \quad (4)$$

Minimizing Eq. (4) has the effect of pulling data of the same class together in the latent space, which meets property ***(a)***.

$$B = \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{p=1}^{n_i}\sum_{q=1}^{n_j} \|f_i^T x_i^p - f_j^T x_j^q\|^2 W_d^{i,j}(p,q). \quad (5)$$

Maximizing Eq. (5) tends to push data of different classes away, which is consistent with property ***(b)***.

$$C = \sum_{i=1}^{K}\sum_{p=1}^{n_i}\sum_{q=1}^{n_i} \|f_i^T x_i^p - f_i^T x_i^q\|^2 W_t^{i,i}(p,q). \quad (6)$$

Minimizing Eq. (6) preserves the topological structure of individual data set, corresponding to property ***(c)***. Eqs. (4 - 6) can be combined into the final cost function, which is formulated as (7):

$$\mathcal{L}(f_1, ..., f_K) = (A + C)/B, \quad (7)$$

and hence an optimization problem (8) can be written as

$$\underset{f_1,...,f_K}{\operatorname{argmin}} \mathcal{L}(f_1, ..., f_K) \quad (8)$$

Proven in [29], the solution $f_1, ..., f_K$ that minimizing the cost function $\mathcal{L}(f_1, ..., f_K)$ is given by the smallest non-zero eigenvectors of the generalized eigenvalue decomposition of (9). And the matrix D and the matrix L in (9) are the degree matrix and the Laplacian matrix, respectively.

$$Z(\mu L_t + L_s)Z^T x = \lambda Z L_d Z^T x, \quad (9)$$

where

$$Z = \begin{pmatrix} X_1 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & ... & 0 & X_K \end{pmatrix},$$

$$L_a = W_a - D_a,\ a \in \{s, d, t\}$$

$$D_a(p,q) = \begin{cases} \sum_{q=1}^{m_1+...+m_k} W_a(p,q) & p = q \\ 0 & p \neq q \end{cases}.$$

C. MAPPER-induced manifold alignment for semi-supervised data fusion (MIMA)

As introduced in the last section, three properties are pursued in SSMA while projections are being learned. Essentially, the first two properties seek to minimize intra-class variance and maximize inter-class variance for the projected data by using label information. This is a goal commonly pursued by many classification strategies, such as linear discriminant analysis [56]. The third property, preserving topological structure, brings two powerful characteristics to SSMA. First, the topological structure is extracted from data, both with and without a label. Thus, SSMA builds up connections among them, which implicitly propagates the label information to unlabeled data. This would amplify the usage of existing labels. Since the label is valuable, the propagation property of the topological term is highly valued. Second, topology emphasizes a notion of nearness, but can distort or even ignore large distances [44]. This is a desirable property for the purpose of classification. For instance, data of one class located in a certain extent of feature space, and locations with large distance to the extent are meaningless for classifying the specific class. This is also proven truth in classification using

Algorithm 1: MAPPER($\mathbf{X}_i$,b,c,$\mathbf{F}$)

Input: $\mathbf{X}_i \in \mathbb{R}^{m_i \times n_i}$: the i^{th} data source with n_i instances and m_i dimensions, **b**: the number of bins, **c**: overlapping percentage of adjacent bins, $\mathbf{F}$: filter function.

Output: $\mathbf{W}_c^{i,i}$: adjacent matrix with the size of $n_i \times n_i$.

calculate the parameter space $\mathbf{X}_i\mathbf{F}$
divide $\mathbf{X}_i\mathbf{F}$ into **b** intervals with **c**% overlap of adjacent intervals
divide data $\mathbf{X}_i$ into **b** data bins corresponding to intervals achieved in 2
for (each data bin):
  Spectral clustering
end for
Construct topological matrix

$$\mathbf{W}_c^{i,i}(p,q) = \begin{cases} 1, & \text{if } p \text{ and } q \text{ in the same cluster;} \\ 1, & \text{if } p \text{ and } q \text{ in the linked clusters;} \\ 0, & \text{otherwise.} \end{cases}$$

Return $\mathbf{W}_c^{i,i}$

Algorithm 2: MIMA ($\{\mathbf{X}_i, \mathbf{Y}_i\}$, **b**, **c**, $\mathbf{F}$)

Input: $\{\mathbf{X}_i, \mathbf{Y}_i\}$, $i \in \{1, ..., K\}$: K data sources and label, **b**: the number of bins, **c**: overlapping percentage of adjacent bins, $\mathbf{F}$: filter function.

Output: $\hat{\mathbf{X}}_i$, $i \in \{1, ..., K\}$: the learned latent features of the K data sources.

Construct $\mathbf{W}_s$ by labeled data $\mathbf{Y}_i$, $i \in \{1, ..., K\}$;
Construct $\mathbf{W}_d$ by labeled data $\mathbf{Y}_i$, $i \in \{1, ..., K\}$;
for (i = 1:K):
  $\mathbf{W}_c^{i,i} = \mathbf{MAPPE}(\mathbf{X}_i, \mathbf{b}, \mathbf{c}, \mathbf{F})$
end for
Construct $\mathbf{W}_c = \begin{pmatrix} W_c^{1,1} & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & W_c^{K,K} \end{pmatrix}$
Compute the projections $\{f_1, ..., f_K\}$ by solving Eq (9)
for (i = 1:K):
  $\hat{\mathbf{X}}_i = \mathbf{X}_i f_i$
end for
Return $\hat{\mathbf{X}}_i$, $i \in \{1, ..., K\}$

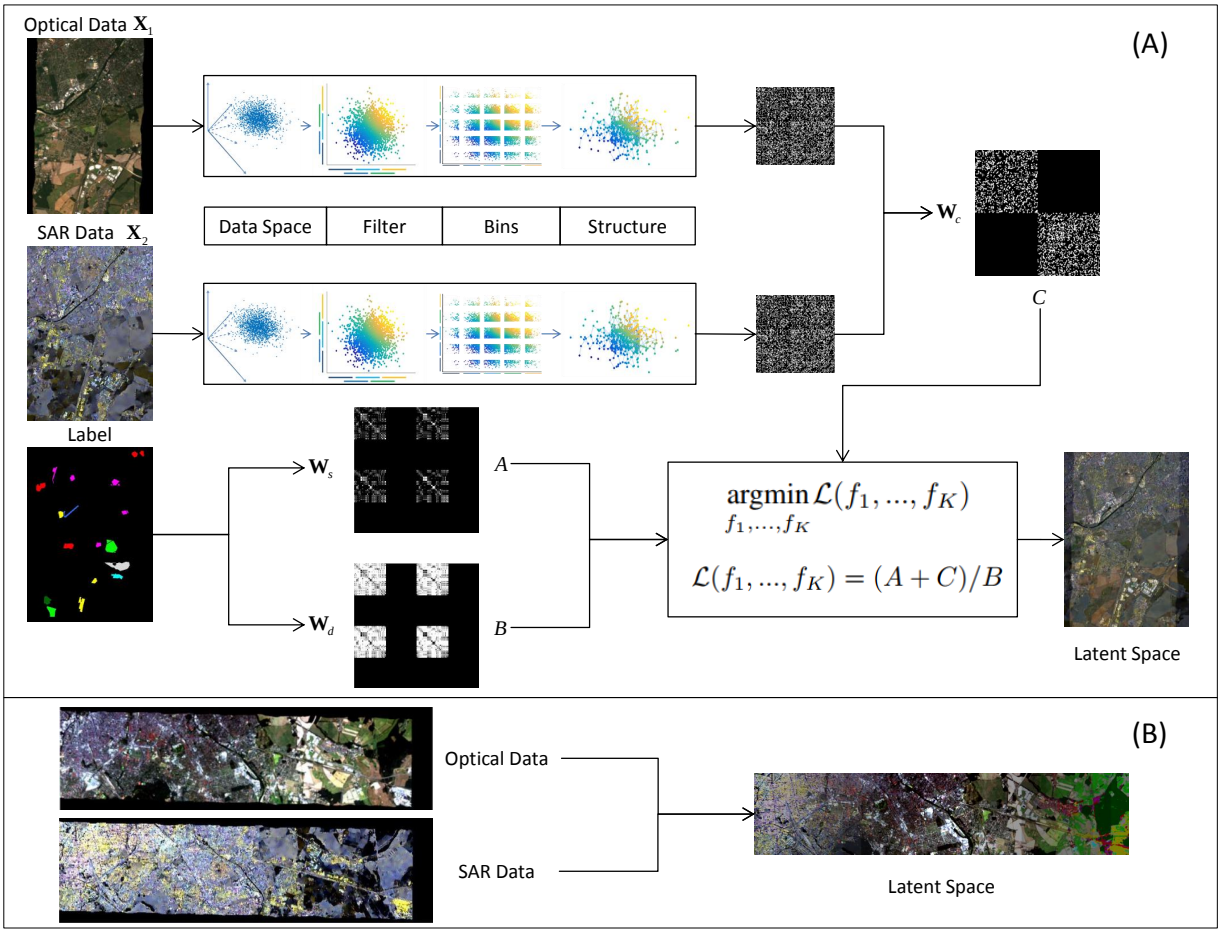


Fig. 2: The flowchart of the algorithm MIMA. (A) Training phase: a topological graph (W_c) is derived from the optical data and the SAR data by MAPPER. A similarity graph (W_s) and a dissimilarity graph (W_d) are formed by using the label information. Therefore, three regularization terms A, B, and C are formulated as Eq. 4, Eq. 5, and Eq. 6, respectively. Lastly, the projection to the latent space is learned by optimizing $\underset{f_1,...,f_K}{\operatorname{argmin}} \mathcal{L}(f_1, ..., f_K)$, where $\mathcal{L}(f_1, ..., f_K) = (A + C)/B$. (B) Testing phase: the out-of-sample optical data and SAR data are projected into the latent space to accomplish fusion.

topology [36], [35].

In order to achieve the topological term, *kNN* commonly serves as the tool to approximate topological structure in conventional methods [29], [57], chosen for its simplicity. In our proposed MIMA, we utilize MAPPER to extract topological structure. There are two reasons to do so. First, when applying MAPPER, field knowledge could be introduced by choosing the filtering function $\mathbf{F}$. In remote sensing classification, field knowledge is essential for the complicity of data. Second, when using *kNN*, nearness is decided solely by the parameter K, which is manually given. Once the K is determined, it is applied to all data without any adaptation. However, nearness is achieved by clustering in MAPPER, which is a more robust approach than deciding nearness by giving a threshold value K. Furthermore, in order to empower MAPPER to decide the nearness in an adaptive manner, the original single-linkage clustering [58], [44] is replaced by the spectral clustering [51] in MIMA. The reason is that the spectral clustering is able to detect the number of clusters by the concept of the eigengap [51]. Thus, when clustering each data bin, the number of clusters is decided based on the data itself, meaning that the nearness is derived in a data-driven manner. For different data bins, the numbers of clusters are different, meaning the nearness is derived for different parts of data in adaptive fashion. Thus, our improved version of MAPPER is capable of deriving topological structure in an automatic and adaptive fashion.

Although the original goal of MAPPER is to provide a simplified visualization of a complicated data set, as shown in Fig 1, one can also derive the comprehensive topological structure of all data points using MAPPER. The topological structure of data source $\mathbf{X}_i$ could be represented as an $n_i \times n_i$ matrix W_c^i, where n_i is the number of instances: $W_c^i(p, q) = 1$, when data instances p and q are in the same cluster or in linked clusters, otherwise, $W_c^i(p, q) = 0$. In MIMA, the topological matrix W_t in equation (3) is replaced by W_c (10).

$$W_c = \begin{pmatrix} W_c^{1,1} & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & W_c^{K,K} \end{pmatrix}. \tag{10}$$

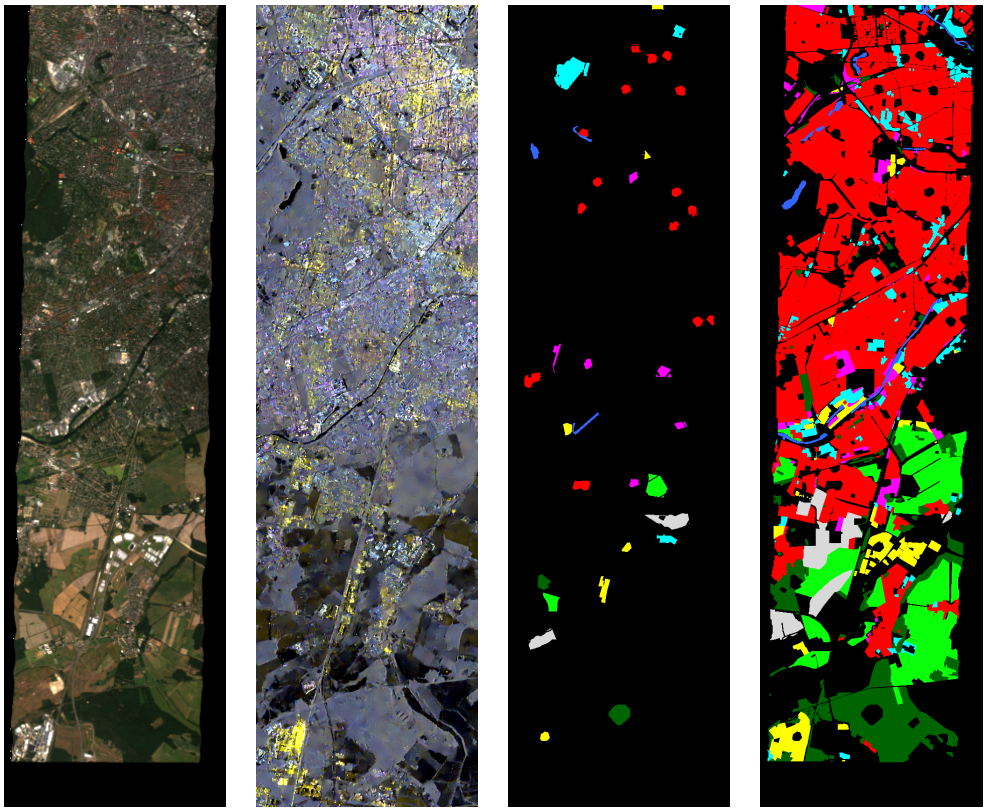


Fig. 3: LCLU data set. From left to right: RGB components of simulated EnMAP data; RGB component of Sentinel-1 dual-Pol data; LCLU training set; LCLU testing set.

The detailed steps of MIMA are summarized in Algorithm 1 and Algorithm 2.

III. Experiments and Discussion

A. Data set and feature design

1) Land cover land use data set (LCLU data set) As shown in Fig. 3, the LCLU data set consists of three data sources: a hyperspectral image, dual-Pol SAR data, and ground truth data. The hyperspectral image is a simulated spaceborne EnMAP scene with a size of 817 by 220, a 30-meter ground sampling distance (GSD), and 244 spectral bands ranging from $400\,\mathrm{nm}$ to $2500\,\mathrm{nm}$ [59]. The dual-Pol SAR data is a VH-VV polarized Sentinel-1 single look complex (SLC) data collected by interferometric wide swath mode.[1] The Sentinel-1 SLC data is preprocessed by the ESA SNAP toolbox.[2] The processed dual-Pol SAR data has a GSD of 13 meters and a size of 1723 by 476. It is organized as the commonly used PolSAR covariance matrix. The ground truth is a land cover land use data set derived from an Open Street Map data.[3]

2) Local climate zone data set (LCZ data set) The local climate zone data set is demonstrated in Fig. 4. It consists of a multispectral image, a dual-Pol SAR data, and a ground truth data. The multispectral image is a scene of LandSat-8 data with a size of 2220 by 2143, a 30-meter GSD, and 11 bands. The dual-Pol SAR data is also a VV-VH polarized Sentinel-1 data processed by the ESA SNAP toolbox. It has a 13.9-meter GSD, a size of 4795 by 4632, and is organized as the commonly used PolSAR covariance matrix. The ground truth is a local climate zone label released by the IEEE GRSS IADF for the data fusion contest in 2017.[4]

3) Label configuration For both the LCLU data set and the LCZ data set, as shown in Fig. 3 and Fig. 4, the training label and the testing label are block-wise separated so that the transferring ability of algorithms is under examination and the risk of implicitly including testing samples into the training procedure is avoided [60]. The label information is detailed in Table I and Table II.

4) Unlabeled data Regarding SSMA and MIMA, the training procedures involves both labeled data and unlabeled data. The unlabeled data was selected by the clustering strategy in [16] so that cluster centers of unlabeled data were selected. In this work, for a more general case, the unlabeled data for training is randomly selected outside the extent of training set. For both the LCLU data set and the LCZ data set, 6000 unlabeled data instances are selected to be involved in training.

5) Feature design of the LCLU data set In order to conduct fair comparisons among algorithms, two principles are pursued on the design of input features of individual data sources. The

[1] https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/interferometric-wide-swath

[2] http://step.esa.int/main/toolboxes/snap/

[3] http://download.geofabrik.de/

[4] http://www.grss-ieee.org/

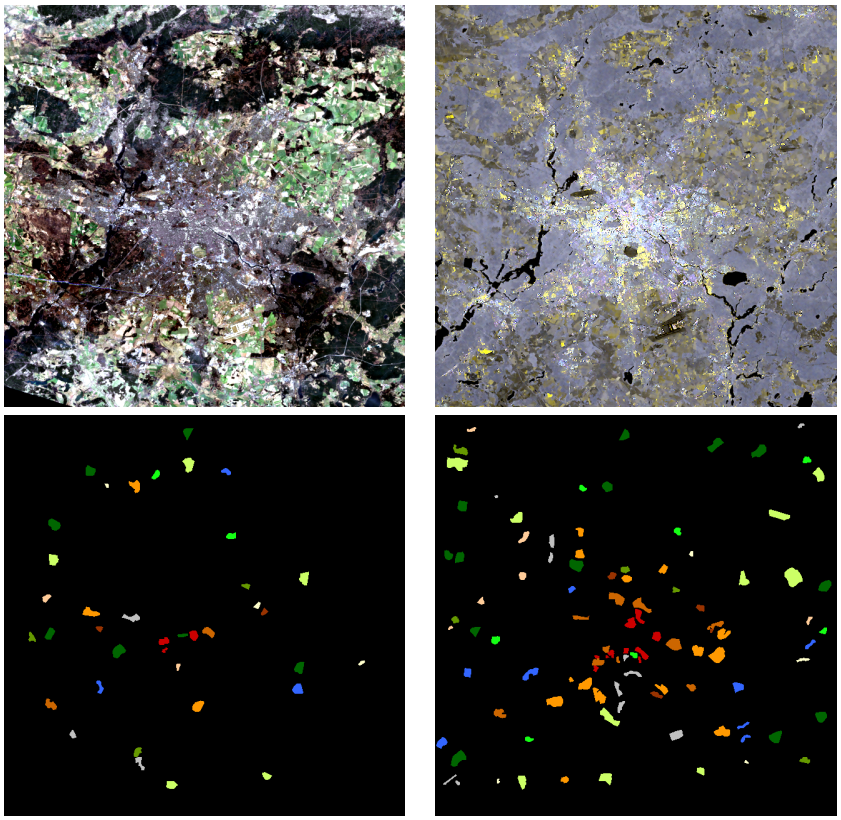


Fig. 4: LCZ data set. From left to right, first row: RGB components of LandSat-8 data; RGB component of Sentinel-1 dual-Pol data; second row: LCZ training set; LCZ testing set.

TABLE I: Summary of training and testing for LCLU data set

Class	Number of Samples	
Name	Train Set	Test Set
Forest	298	52455
Residential area	756	262903
Industrial area	296	17462
Low plants	344	56683
Soil	428	14505
Allotment	281	11322
Commercial area	560	20909
Water	153	5539
Total	3116	441778

TABLE II: Summary of training and testing for LCZ data set

Class	Number of Samples	
Name	Train Set	Test Set
Compact mid-rise	198	1138
Open high-rise	83	412
Open mid-rise	213	2023
Open low-rise	375	3260
Large low-rise	233	1189
Sparsely built	92	577
Dense trees	769	3423
Scattered trees	136	756
Bush, scrub	181	689
Low plants	560	3305
Bare soil or sands	84	192
Water	246	1241
Total	3170	18205

first principle is simply that input features of each data source should be the same for all algorithms. The second principle is that, when an individual data source is used for classification, the input feature should enable reasonably good performance. This is to ensure that later improvements do not originate from the unexplored potential of one data source, but from the fusion or the fusion algorithms. For example, due to the well-known curse-of-dimensionality [61], conducting classification on selected dimensions of hyperspectral images could result in better performance than using the data with all dimensions [62]. If the original full dimensional data were used in our case, it would then be unclear later whether the improvement comes from the fusion or from the dimension reduction.

Regarding the feature design of the simulated EnMAP data, the spectral-spatial feature concept is adopted by extracting morphological profiles from extracted informative subdimensions [63], [64]. Specifically, the first four principal components (PCs) are extracted, which accounts for 99% of the variances of the simulated EnMAP data. The morphological profile is then extracted from these four PCs with radius equal to one, two, and three. In total, 28 features are extracted from the simulated EnMAP data set.

Regarding the feature design of Sentinel-1 dual-Pol data,

TABLE III: The nine algorithms in experimental comparisons. Their fusion strategies are MA-fusion (manifold alignment fusion) and DR-fusion (joint dimension reduction fusion). The learning resource are the Label (annotated data-label records), the Pseudo-label (prediction from a classifier), and the data structure (the distribution of data in feature space). The parameters of these algorithms are, k: the k^{th} neighbor in *kNN* for approximating topological structure; dn: the number of dimensions in the projected space; μ: the importance weighting of topological structure; α and β: learning rates.

Algorithm	Principle		Parameters						
	Fusion Strategy	Learning Resource	k	dn	μ	α	β	b	c
(*A*) POL	PolSAR only	Label	-	-	-	-	-	-	-
(*B*) OPT	Optical only	Label	-	-	-	-	-	-	-
(*C*) OPT-POL	Feature concatenation	Label	-	-	-	-	-	-	-
(*D*) COSPACE	MA-fusion	Label	-	-	-	✓	✓	-	-
(*E*) LeMA	MA-fusion	Label & Pseudo-label	-	-	-	✓	✓	-	-
(*F*) LPP	DR-fusion	Data structure	✓	✓	-	-	-	-	-
(*G*) LPP-SE	DR-fusion	Label & Data structure	✓	✓	-	-	-	-	-
(*H*) SSMA	MA-fusion	Label & Data structure	✓	✓	✓	-	-	-	-
(*I*) MIMA	MA-fusion	Label & Data structure	-	✓	✓	-	-	✓	✓

four polarimetric features are derived: intensity of the VH channel, intensity of the VV channel, the coherence of VV and VH, and the intensity ratio of VV and VH. Since the morphological profile was proven to promote classification of PolSAR [25], [65], [66], it is also used to extract spatial information from dual-Pol data here with radius equal to one, two, and three. In total, it results in 28 features from Sentinel-1 dual-Pol data.

6) Feature design of the LCZ data set The feature design for the LCZ data set also follows principles described in the feature design of LCLU data set.

Regarding the feature design of the LandSat-8 data, in order to achieve feature combination for reasonable good performance, the feature extraction and selection follows the strategy in first prize work from the GRSS IADF data fusion contest in 2017 [8]. Local statistical parameters (mean and standard deviations in a 100×100-meter neighborhood) and morphological profiles are extracted from original LandSat-8 data. For details, please refer to [8]. In total, 34 features are used in our work.

Regarding the feature design of Sentinel-1 dual-Pol data in the LCZ data set, the data source and the preprocessing are the same as those in the LCLU data set. The prepared fundamental features are the four polarimetric features. However, due to the local climate zone describes an urban local neighborhood at a grid with a 100×100-meter unit cell, feature extraction is different in the LCZ data set than in the in LCLU data set. Local statistical parameters, mean and standard deviation of local 100×100-meter cell, are derived from all four polarimetric features, resulting in eight features. Morphological profiles are therefore extracted from these eight features with radius equal to one and three. Thus, 40 features are prepared in total.

B. Experiments setting

In experiments, nine algorithms, which are listed in Table III, are applied to extract features from optical and dual-Pol SAR data. Hereafter, three classifiers are used to test the performance of these algorithms in terms of classification accuracy. The seven algorithms are: (*A*) dual-Pol SAR data (POL), (*B*) optical data (OPT), (*C*) fusing of optical and dual-Pol SAR data by feature concatenation (OPT-POL), (*D*) fusing optical and dual-Pol SAR data by COSPACE [67], (*E*) fusing optical and dual-Pol SAR data by LeMA [68], (*F*) fusing optical and dual-Pol SAR data by unsupervised joint dimension reduction using locality preserving projection [35] (LPP), (*G*) fusing optical and dual-Pol SAR data by semi-supervised joint dimension reduction using locality preserving projection (LPP-SE), (*H*) fusing optical and dual-Pol SAR data by SSMA [33], and (*I*) fusing optical and dual-Pol SAR data by the proposed MIMA.

Among these nine algorithms, parameter tuning is required by (*D*) COSPACE, (*E*) LeMA, (*F*) LPP, (*G*) LPP-SE, (*H*) SSMA, and (*I*) MIMA, as shown in Table III. For the (*D*) COSPACE and (*E*) LeMA, two learning rates, α and β need to be set for the optimization. They are tunned by searching in a grid of $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. Regarding the parameter k in (*F*) LPP, (*G*) LPP-SE and (*H*) SSMA, it has been reported in [33] that the parameter does not have significant influence on the result and is recommended to be nine. For the parameters dn and μ, they will be discussed later in our experiments. In MIMA, two more parameters have to be decided: (1) $\mathbf{b}$: the number of intervals for dividing the data, and (2) $\mathbf{c}$: the overlapping rate. Since these two parameters have limited influences, as discussed in section III-F, specially when compared with the other parameters. Their values are chosen as 5 and 50%, respectively, which is a result of consulting other studies [46], [48] and summarizing practical experiences of the author.

Besides the parameter tunning, one important part of MIMA is to select the filtering function. As discussed before, the filtering function provides a perspective of observing the data and introduces field knowledge. As principal components are widely used in classification of remote sensing data and have been proven to be effective, and this is the first attempt of applying MAPPER in remote sensing, the first and second principal components are chosen to serve as the field knowledge in this work.

As shown in Table III, (*D*) COSPACE, (*E*) LeMA, (*H*) SSMA, and (*I*) MIMA are all fall into the manifold alignment fusion strategy. However, their learning resources are different. COSPACE is designed to learn a joint latent space via the existed labeled data. In addition to the labeled data, LeMA

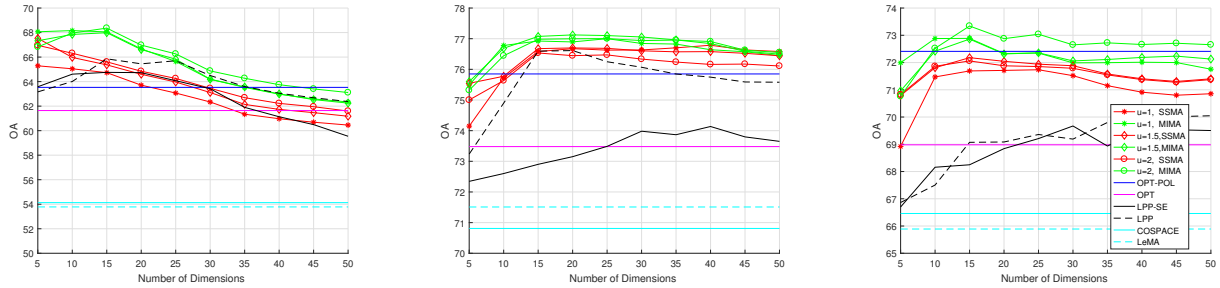


Fig. 5: Classification performance in terms of overall accuracy (OA) for the experiments applied on the LCLU data set. The charts show the results of the three classifiers, from left to right, *ONE-NN*, *LSVM*, and *KSVM*.

TABLE IV: Quantitative performance comparison with the different algorithms on the LCLU data, in terms of class-specific accuracy, kappa coefficient, average accuracy, overall accuracy, and mean overall accuracy. The best performance achieved is shown in bold. Power (a, b) of learning rates $\alpha = 10^a$ and $\beta = 10^b$ are shown for COSPACE and LeMA in terms of the best performance. Number of dimensions (dn) is indicated as (dn) for the best performance of LPP and LPP-SE. Parameter values of μ and number of dimensions (dn) are indicated as (μ, dn) for the best performance of SSMA and MIMA.

Algorithm	Parameter	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Soil	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
POL	-	1NN	35.19	51.03	21.54	25.43	62.1	29.24	29.22	28.36	0.2353	35.26	43.19	
		LSVM	30.32	75.65	7.93	26.18	85.17	39.99	28.92	28.42	0.3614	40.32	57.84	51.02
		KSVM	34.15	66.71	9.17	20.51	67.35	31.2	42.17	22.44	0.2999	36.71	52.04	
OPT	-	1NN	68.78	63.87	30.01	57.58	90.73	55.76	32.86	73.89	0.4599	59.18	61.64	
		LSVM	69.2	82.5	18.55	65.7	79.06	53.59	44.77	72.81	0.585	60.77	73.48	68.03
		KSVM	66.54	77.68	26.18	58.33	65.16	54.94	39.64	72.74	0.5269	57.65	68.98	
OPT-POL	-	1NN	67.02	66.41	32.99	63.54	83.57	55.54	33.53	67.09	0.478	58.71	63.53	
		LSVM	70.13	84.63	30.43	71.58	79.22	56.27	37.67	75.03	0.6183	63.12	75.85	70.60
		KSVM	61.76	82.97	29.7	61.17	76.55	52.98	40.79	70.36	0.5667	59.53	72.41	
COSPACE	$(-1, -2)$	1NN	47.72	58.67	24.27	55.99	61.32	47.67	29.81	60.93	0.3596	48.3	54.13	
	$(-1, -3)$	LSVM	55.04	82.24	27.48	66.28	75.98	57.63	23.66	51.49	0.5379	54.98	70.81	63.80
	$(-1, -4)$	KSVM	34.14	85.08	17.03	52.93	48.37	46.87	22.51	36.52	0.4395	42.93	66.46	
LeMA	$(-1, -1)$	1NN	50.93	57.07	26.88	56.86	63.67	49.68	26.92	61.53	0.3613	49.19	53.78	
	$(-1, -1)$	LSVM	68.52	80.13	25.38	64.57	76.54	62.99	29.41	70.12	0.5601	59.71	71.51	63.73
	$(-2, -2)$	KSVM	36.57	84.21	13.64	48.01	55.22	45.66	27.45	36.02	0.4345	43.35	65.89	
LPP	(15)	1NN	70.81	68.84	34.57	64.82	89.64	51.91	33.22	76.17	0.5057	61.25	65.86	
	(20)	LSVM	69.39	85.56	32.84	71.41	81.94	57.46	38.23	76.67	0.6255	64.19	76.55	70.86
	(50)	KSVM	59.94	79.37	21.41	66.37	63.32	54.66	46.84	61.78	0.5394	56.71	70.18	
LPP-SE	(15)	1NN	67.21	68.2	37.33	65.06	79.01	55.61	29.5	75.57	0.4898	59.69	64.75	
	(40)	LSVM	66.19	83.62	36.88	66.44	81.5	60.61	28.31	75.43	0.5926	62.37	74.12	69.53
	(30)	KSVM	62.07	79.3	34.73	55.15	76.33	55.04	39.41	72.97	0.5326	59.38	69.71	
SSMA	$(1.5, 5)$	1NN	69.26	71.78	33.67	67.44	84.3	55.94	31.93	71.08	0.5222	60.67	67.53	
	$(1, 40)$	LSVM	66.07	85.31	34.67	80.05	78.83	57.13	34.59	70.1	0.632	63.34	76.82	72.25
	$(1.5, 15)$	KSVM	59.65	83.6	30.21	67.64	43.58	55.47	41.93	68.44	0.5642	56.32	72.4	
MIMA	$(2, 15)$	1NN	71.29	72.5	35.6	71.85	66.86	59.78	31.83	71.29	0.5329	60.13	68.36	
	$(1.5, 20)$	LSVM	67.58	86.05	36.32	76.79	78.35	58.75	33.71	76.06	**0.6367**	**64.2**	**77.15**	**73.04**
	$(2, 15)$	KSVM	64.3	84.21	34.46	65.73	54.74	57.45	40.54	69.71	0.5844	58.89	73.6	

also uses the pseudo-labeled data, predictions of a trained classifier on unlabeled data, to include unlabeled data into the procedure of data fusion. For SSMA and MIMA, they utilize the labeled data and extract the data distribution under the guidance of mathematical assumptions, to achieve data fusion. Therefore, when a large amount of labeled data exists or the data distribution is not correlated with labels, LeMA would be more appropriate than SSMA and MIMA, and vice versa.

Since our goal is to assess the performance of fusion, three classical classifiers are chosen: the one-nearest-neighbor classifier (*ONE-NN*), the linear support vector machine (*LSVM*), and the Gaussian kernel support vector machine (*KSVM*). In this work, parameter tuning of *LSVM* and *KSVM* are done in a heuristic procedure [69].

C. Classification on the LCLU data set

This section demonstrates and discusses the experimental results obtained on the LCLU data set.

1) Fusion vs. non-fusion. As shown in Fig. 5 and Table IV, classification on fused hyperspectral imagery and dual-Pol SAR data outperforms classification on the individual data source, in terms of classification accuracies. Among the fusion algorithms, our proposed MIMA provides the best classification performance, which, in terms of overall accuracy, exceeds classifications on dual-Pol SAR data by 25%, 20%, and 21% and exceeds classifications on hyperspectral imagery by 7%, 4%, and 5%, using *ONE-NN*, *LSVM*, and *KSVM*, respectively. This proves that fusion of hyperspectral imagery and dual-Pol SAR data is advantageous to LCLU classification.

2) Fusion categories. Based on properties of fusion algorithms listed in Table III, to simplify the discussion, we

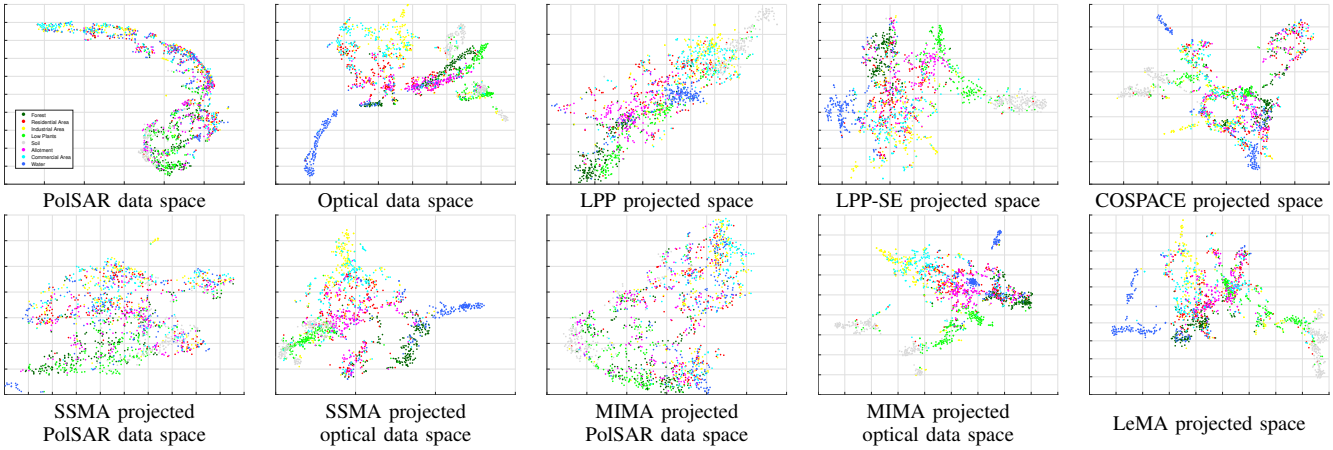


Fig. 6: Visualization of the optical data and the PolSAR data of the LCLU data set, using t-SNE [70] in their original and projected spaces. The x and y axis are the first and second dimensions resluted from the t-SNE. The first row are: the PolSAR data in the original space, the optical data in the original space, LPP jointly projected space, LPP-SE jointly projected space, and COSPACE projected space, respectively. The second row are: the PolSAR data in SSMA projected space, the optical data in SSMA projected space, the PolSAR data in MIMA projected space, the optical data in MIMA projected space, and LeMA projected space, respectively.

would like to divide these seven fusion algorithms into four categories: (1) feature concatenation (OPT-POL); (2) joint dimension reduction fusion (LPP and LPP-SE); (3) label-driven manifold alignment (COSPACE and LeMA); and (4) data-driven manifold alignment fusion (SSMA and MIMA). According to the classification accuracy in Table I, with the feature concatenation (OPT-POL) serves as the benchmark, it is obvious to find that: (a) joint dimension reduction fusion algorithms achieve similar classification accuracy to the feature concatenation fusion; (b) the overall accuracy provided by label-driven manifold alignment are around 7% lower than the accuracy achieved by the feature concatenation; (c) the data-driven manifold alignment fusion outperforms the feature concatenation by 2%. Discussions regarding the three findings are detailed as follows.

It is well known that the dimension reduction technique is capable of boosting the classification accuracy, due to the curse-of-dimensionality [61]. However, according to the finding (a) above, this doesn't suit the feature concatenation fusion in out experiment. Because the curse-of-dimensionality has been tackled in our feature design. The finding (a) also validates that the improvement of our proposed method is not a side effect of dimension reduction.

The label-driven manifold alignment fusion learns projections that map original data sources to a latent space purely based on the label, and applies learned projections on the unlabeled data to accomplish fusion. The finding (b) gives a clear clue that this type of fusion can not provide a proper fusion result for the LCLU data set. This could because the label-driven learned latent space is not applicable to a general case, namely the unlabeled data. Thus, the label-driven manifold alignment fusion might provide a destructive fusion when the label data can not represent the data distribution which is often the case in remote sensing.

The data-driven manifold alignment fusion also learns projections that map original data sources to a latent space. However, the latent space is jointly defined by the label and the data structure explored from the original data sources, including labeled and unlabeled data. The finding (c) suggests that the data-driven manifold alignment fusion is an effective fusion strategy, which improves the overall accuracy about 2% by comparing to the feature concatenation fusion.

3) MIMA vs. SSMA. As shown in Fig. 5 and Table IV, the proposed MIMA has superior performance to SSMA. In Fig. 5, verified with three different classifiers, classifications on MIMA-fused data outperform classifications on SSMA-fused data, when parameter μ and the number of dimensions are the same for both fusion strategies. The classification performance of the best parameter combinations is shown in Table IV. It is clear that the novel MIMA strategy still outperforms SSMA strategy, not only verifying the superior performance of the proposed novel MIMA algorithm, but also proving that a MAPPER-derived topological structure is more effective than a *kNN*-derived structure regarding LCLU classification.

4) Parameter μ. As shown in Fig. 5, with *ONE-NN* and *KSVM* classifiers, a higher value of μ results in better classification performance for both SSMA and MIMA algorithms. Recalling that a higher value of μ assigns stronger weight on topological structure of data in the fusing phase, this is solid evidence that topological structure benefits our classification task. We also find that the way MIMA derives the structure is more beneficial to this LCLU classification than the way SSMA accomplishes it.

5) Fusion visualization. In Fig. 6, we visualize the fused features of different algorithms using the t-SNE algorithm [70]. It is obvious that the joint dimension reduction technique results a set of features which is less discriminative than the original feature. This is also reflected on the classification results, shown in Fig. 5. On the other side, when using

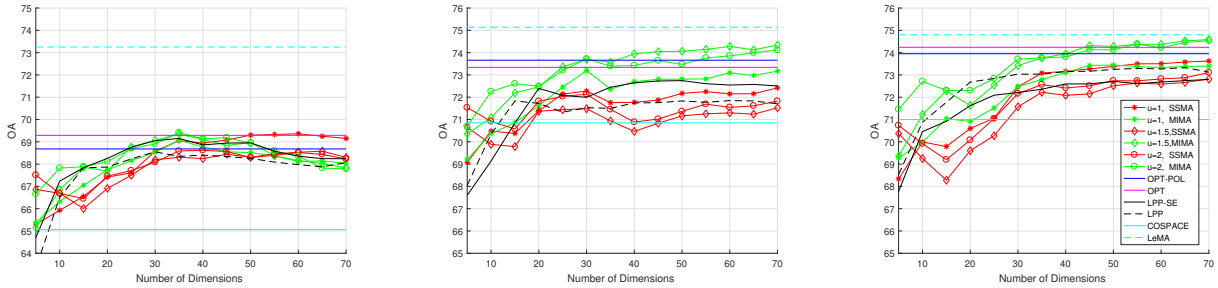


Fig. 7: Classification performance in terms of overall accuracy (OA) for the experiments applied on the LCZ data set. The charts show results for the three classifiers, from left to right, *ONE-NN*, *LSVM*, and *KSVM*.

TABLE V: Quantitative performance comparison with the different algorithms on the LCZ data, in terms of class-specific accuracy, kappa coefficient, average accuracy, overall accuracy, and mean overall accuracy. The best performance achieved is shown in bold. Power (a, b) of learning rates $\alpha = 10^a$ and $\beta = 10^b$ are shown for COSPACE and LeMA in terms of the best performance. Number of dimensions (dn) is indicated as (dn) for the best performance of LPP and LPP-SE. Parameter values of μ and number of dimensions (dn) are indicated as (μ, dn) for the best performance of SSMA and MIMA.

Algorithm	Parameter	Classifiers	Compact mid-rise	Open high-rise	Open mid-rise	Open low-rise	Large low-rise	Sparsely built	Dense trees	Scattered trees	Bush, scrub	Low plants	Bare soil or sands	Water	KAPPA	AA	OA	Mean OA
POL	-	1NN	43.76	29.13	21.4	59.26	33.64	18.54	82.33	10.98	19.59	70.62	28.13	54.23	0.4566	39.3	52.66	
		LSVM	24.78	18.93	23.92	71.1	48.95	7.11	96.23	2.25	14.51	94.19	32.29	88.88	0.5688	43.6	63.03	59.47
		KSVM	40.6	20.87	24.67	75.46	46.76	10.75	95.3	0.13	14.37	92.04	33.85	66.4	0.5637	43.43	62.72	
OPT	-	1NN	72.32	39.56	40.63	67.24	53.99	20.97	98.63	33.99	32.66	79.46	72.4	98.95	0.648	59.23	69.29	
		LSVM	72.93	47.33	52.55	74.66	53.66	20.45	99.5	37.7	28.01	84.87	79.69	99.19	0.6933	62.55	73.34	72.29
		KSVM	70.3	36.17	51.16	73.19	65.85	23.92	99.77	34.79	38.9	90.23	65.63	94.36	0.7024	62.02	74.24	
OPT-POL	-	1NN	61.42	35.44	26.3	73.83	50.55	23.92	98.71	37.3	35.27	82	75.52	98.39	0.6413	58.22	68.68	
		LSVM	49.3	33.01	47.45	80.4	61.14	22.18	99.42	46.16	23.66	89.62	89.06	99.03	0.6957	61.7	73.66	72.10
		KSVM	57.47	32.52	40.04	79.51	61.56	22.36	99.8	37.17	43.11	92.38	77.08	98.07	0.6991	61.75	73.95	
COSPACE	$(-1, -1)$	1NN	70.74	32.77	8.35	67.82	42.39	14.56	98.48	32.54	34.83	82.45	73.96	97.66	0.5961	54.71	65.06	
	$(-1, -3)$	LSVM	69.86	63.83	25.41	78.44	54.58	18.02	99.15	34.39	44.56	83.27	77.6	93.07	0.6653	61.85	70.85	68.97
	$(-1, -1)$	KSVM	68.63	45.39	25.75	75.74	71.83	10.05	99.39	31.48	42.09	87.5	60.94	90.01	0.6651	59.07	71	
LeMA	$(-1, -1)$	1NN	82.07	75.49	37.47	68.31	56.85	37.44	99.62	56.61	39.04	81.66	87.5	99.92	0.6953	68.5	73.25	
	$(-3, -2)$	LSVM	82.07	78.64	46.27	76.69	55	34.14	99.24	51.59	40.93	81.24	75.52	99.6	**0.7166**	68.41	**75.14**	**74.40**
	$(-1, -1)$	KSVM	75.22	74.27	42.26	72.73	67.45	35.88	99.74	50	47.46	86.35	86.98	87.11	0.7119	**68.79**	74.8	
LPP	(30)	1NN	60.98	62.62	33.22	68.65	51.3	26.69	99.77	44.31	33.96	75.22	85.94	98.23	0.6407	61.74	68.55	
	(60)	LSVM	49.12	62.86	45.13	68.77	60.22	25.3	99.8	44.31	50.8	83.06	79.17	99.19	0.6761	63.98	71.76	71.19
	(65)	KSVM	60.37	54.85	43.45	71.01	81.5	24.09	99.8	35.71	45.43	86.02	69.27	92.51	0.6927	63.67	73.27	
LPP-SE	(35)	1NN	58.88	41.99	19.67	76.35	44.41	19.76	99.36	52.25	42.38	83	77.6	99.68	0.6453	59.61	69.15	
	(50)	LSVM	48.59	50	22.24	79.33	64.51	30.68	99.68	53.97	50.22	89.23	83.33	99.44	0.6867	64.27	72.77	71.57
	(70)	KSVM	43.23	44.17	25.46	77.88	81.92	20.45	99.77	48.68	51.38	89.77	70.31	96.21	0.6873	62.44	72.8	
SSMA	$(1, 60)$	1NN	80.49	53.16	17.55	75.25	49.29	23.4	99.3	37.04	45.86	78.52	74.48	99.11	0.6496	61.12	69.36	
	$(1, 70)$	LSVM	72.06	65.53	36.38	71.81	58.37	27.73	99.59	37.17	53.85	82.09	78.13	98.55	0.6838	65.1	72.33	71.70
	$(1, 70)$	KSVM	71.53	61.17	32.43	78.22	80.07	23.74	99.8	33.6	50.8	80.94	56.25	96.54	0.6949	63.76	73.4	
MIMA	$(2, 35)$	1NN	75.04	49.51	23.18	75.03	48.53	25.48	99.15	40.61	38.17	78.82	75.52	98.87	0.6496	60.66	69.42	
	$(1.5, 70)$	LSVM	73.73	53.64	43.85	78.22	60.39	26	99.36	40.34	49.06	83.69	74.48	98.31	0.7055	65.09	74.36	72.78
	$(1.5, 70)$	KSVM	70.83	47.09	44.34	80.09	76.87	22.88	99.65	35.19	45.57	84.96	54.69	90.09	0.7072	62.69	74.57	

manifold alignment techniques, it is clear that the derived feature is more discriminative than the original ones.

D. Classification on the LCZ data set

This section demonstrates and discusses the experimental results obtained on the LCZ data set.

1) LeMA. The most outstanding phenomenon appears in Fig. 7 and Table. V is that LeMA outperforms all the other algorithms by 2 to 6%, which is considered a large margin in this experiment. Since LeMA and COSPACE both accomplish the fusion by using the labeled data and have similar performance in the experiment of LCLU data set, it is very interesting to find out the reason why LeMA not only outperforms COSPACE but also all the other fusion algorithm with a large margin. The difference between COSPACE and LeMA is that, while learning the projections from the labeled data, LeMA, additionally, includes pseudo-label into the learning phase. The pseudo-label are predictions of unlabeled data inferred by a trained classifier. LeMA has a strategy of selecting pseudo-label which have a high probability to be correctly labeled. In our classification evaluation, those correct-prone pseudo-label are also used for training classifiers of fused data. In the case of the experiment on the LCZ data set, by comparing to original 3170 training records and 18205 testing records, there are 1231 additional pseudo-label selected from the test data set which are used for training classifiers. It increases the training data set by 38.83% and occupies 6.76% of the testing data for validation. We believe the change in classification setting is the main reason that LeMA performs the best in the experiment of LCZ data set. On the other hand, in the experiment of the LCLU data set (3116 training records and 441778 testing records), LeMA has 721 additional pseudo-label, which increases the training data set by 23.14% and occupies 0.16% of the testing data for validation.

2) Fusion. According to Table. V, all fusion algorithms, except LeMA, have similar performance to the classification using only the multispectral imagery. Based on the 0.19% difference between OPT-POL and OPT, we might infer that features extracted from dual-Pol SAR data do not benefit the LCZ classification scheme in terms of overall accuracy.

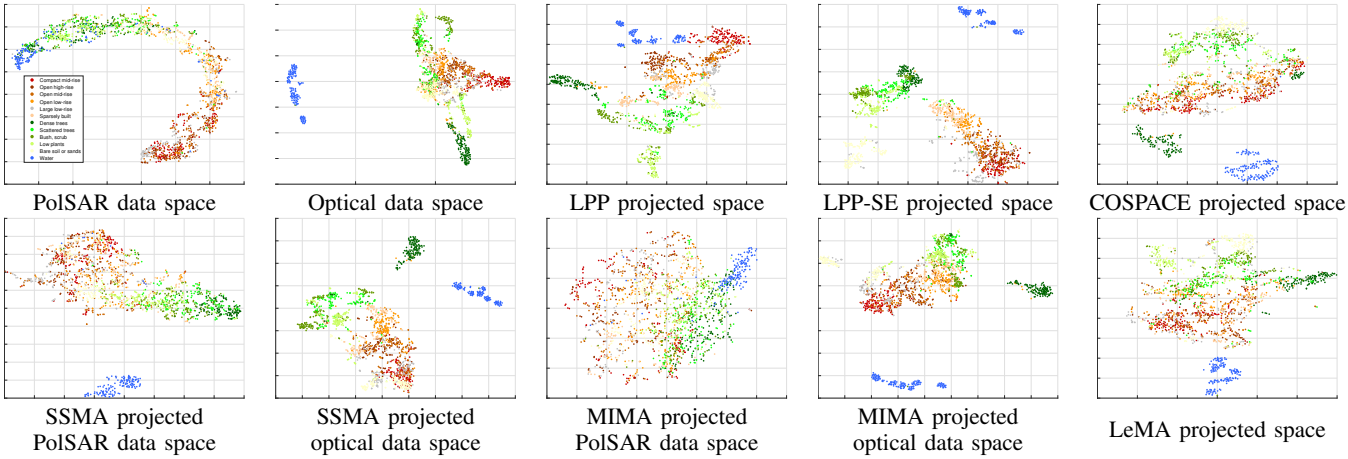


Fig. 8: Visualization of the optical data and the PolSAR data of the LCZ data set, using t-SNE [70] in their original and projected spaces. The x and y axis are the first and second dimensions resluted from the t-SNE. The first row are: the PolSAR data in the original space, the optical data in the original space, LPP jointly projected space, LPP-SE jointly projected space, and COSPACE projected space, respectively. The second row are: the PolSAR data in SSMA projected space, the optical data in SSMA projected space, the PolSAR data in MIMA projected space, the optical data in MIMA projected space, and LeMA projected space, respectively.

However, the dual-Pol SAR data do benefit classifications of certain classes, namely, scattered trees, low plants, bare soil or sand, and water. By comparing COSPACE, LPP, LPP-SE, SSMA, and MIMA to OPT, mean overall accuracy improve -3.32%, -1.1%, -0.72%, -0.59%, and 0.49%, respectively. This infers that few can be done by these fusion algorithms. However, among these fusion algorithms, our propose MIMA outperforms these four former algorithms by 3.81%, 1.59%, 1.21%, and 1.08%.

3) Data-driven manifold alignment. When the fusion is carried out by the OPT-POL, essentially, the information given by the label decides the classification boundary. However, in addition to the label, the data-driven manifold alignment involves the topological structures of the data to find the classification boundary. As shown in Fig. 7, SSMA cannot compete with OPT-POL. This means that data structure is not beneficial with respect to LCZ classification. This is actually reasonable. The LCZ classification scheme describes the contents of an urban local neighborhood relating to the morphological structure, man-made or natural components, and height of structures. However, the topological structure derived from the remote sensing data reveals data structure in terms of its physical meanings, such as covering materials for multispectral images and geometric information for SAR data. Thus, the structure is not directly related to LCZ concepts. On this regard, the data-driven manifold alignment is more appropriate for the LCLU classification, since the information derived in the topological structure is directly related to LCLU classes.

Despite the challenges that LCZ classes pose, when comparing to OPT-POL, the proposed MIMA slightly improves 1.02% overall accuracy with the *LSVM*. Comparing to LeMA, the performance of MIMA differ by -3.83%, -0.79%, and 0.23% in terms of overall accuracy with three different classifier. We consider the performance are comparable. Only the -3.83% indicates a big difference. However, this is because 38.83% and 6.76% differences of training and testing records have a huge impact on the classifier *1NN*. With the two other classifiers, even with fewer training samples, the proposed MIMA is able to provide comparable classification accuracy.

4) Parameter μ. According to Fig. 7 and Table V, trends in terms of μ show that SSMA achieves its best performance when parameter μ equals one and performances are downgraded by increasing μ without a pattern. However, MIMA exhibits a pattern that classification accuracy increases as the value of parameter μ increases. This means that putting higher weights on the topological structure while fusing with MIMA would provide better classification performance in terms of OA.

5) Fusion visualization. We visualize the fused features of different algorithms by t-SNE, as illustrated in Fig. 8. However, it is difficult to carry out a detailed analysis according to the visualization results. In general, the manifold alignment based fusion provides spaces where classes concentrate well. To our knowledge, the optical data are projected into a more discriminative subspace via the proposed MIMA.

TABLE VI: The computational cost of algorithms in comparison. The time listed in this table are means of ten repetitions on each algorithm, carried out on the LCZ data set. The unit is reported in second

Fusion algorithm	(*D*) COSPACE	(*E*) LeMA	(*F*) LPP	(*G*) LPP-SE	(*H*) SSMA	(*I*) MIMA
Time (sec)	324.3175	1080.3080	4.6141	6.0033	45.2369	129.3108

E. Computational cost

In order to show the computational efficiency of algorithms in comparison, experiments of ten repetitions over the LCZ data set had been carried out for every algorithm in comparison. All the experiments are accomplished on a desktop with

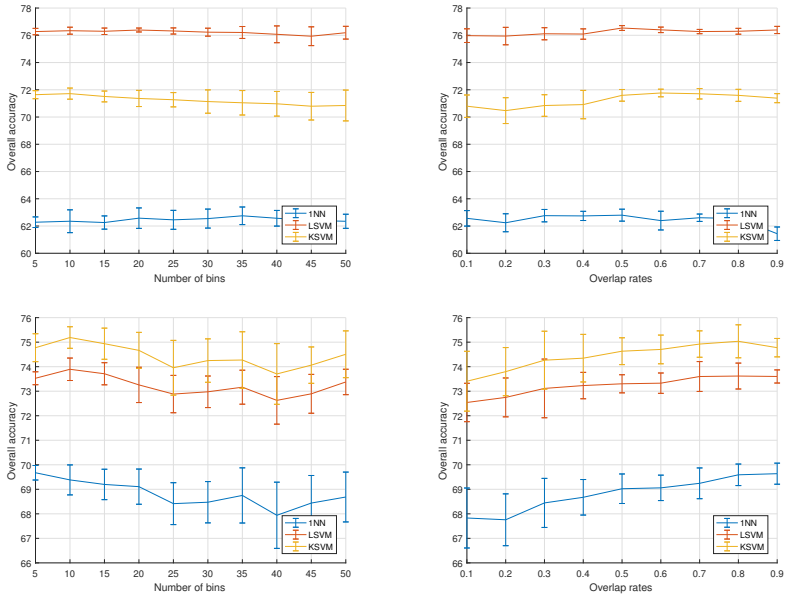


Fig. 9: Plots of the classification overall accuracies achieved by applying the classifiers: the *ONE-NN*, the *LSVM*, and the *KSVM* on MIMA fused feature, while only values of two parameters varies, the number of data bins and the overlap rate. From left to right: (1) plot of overall accuracies achieved on the LCLU data set; the curve and the error bar represent the mean and the standard deviation, which achieved statistically with varying overlap rates; (2) plot of overall accuracies achieved on the LCLU data set; the curve and the error bar represent the mean and the standard deviation, which achieved statistically with varying number of bins; (3) plot of overall accuracies achieved on the LCZ data set; the curve and the error bar represent the mean and the standard deviation, which achieved statistically with varying overlap rates; (4) plot of overall accuracies achieved on the LCZ data set; the curve and the error bar represent the mean and the standard deviation, which achieved statistically with varying number of bins;

a processor of Intel Core i7-4790 CPU (3.60 GHz). The mean time of these repetitions are listed in Table. VI.

Our proposed MIMA do suffer from comparably high computational cost, as shown in Table VI. This is due to the high computational cost of the spectral clustering [51] in MIMA. If the algorithm efficiency is of key importance for a targeted application, more studies could be carried out to find a less demanding clustering algorithm as a substitute. However, considering 9170 optical pixels with 34 dimensions and 9170 SAR pixels with 40 dimension are involved in the training of the algorithm, we think two minutes which required by MIMA is still acceptable.

F. Data bins and overlap rates

As described in the section II-C, there are two parameters brought to MIMA by the MAPPER. They are: the number of data bins and the overlap rate of adjacent data bins. Among all experiments in the previous sections, the number of data bins is chosen as 5 and the overlap rate is selected as 50%, based on the experience of medical studies [46], [48]. However, in this section, the impact of those two parameters are discussed, in terms of the remote sensing data.

Theoretically, the number of data bins has a similar effect to the value k of the *kNN*, which controls the extent of a local neighborhood. Because the local topological structure is derived by the clustering in smaller slices of the data, when a larger number of bins is applied. On the other hand, the overlap rate controls the strength of the connection between adjacent local neighborhoods. Although the theoretical concept is clear, their impacts are really depending on the data set which it works with.

Fig. 9 demonstrate the impact of the number of data bins and the overlap rate by using the LCLU data set and the LCZ data set, in terms of classification accuracy. The number of bins is set to values from 5 to 50 with an interval of 5. The overlap rate is set to values from 0.1 to 0.9 with an interval of 0.1. For the sake of simplicity, the parameter μ is set to 2 and the latent space dimension is set to 50, for the analysis in this section.

According to the upper two plots in Fig. 9, regarding the LCLU classification, the number of bins and the overlap rate do not have a significant influence in terms of the overall accuracy. However, based on the bottom two plots in Fig. 9, regarding the LCZ classification, one can recommend a large overlap rate around 90% and the number of data bins around 10. Thus, the decision of both parameters really depends on the data set and the targeted classification scheme. Last but not the least, it also relates to the choices of the filtering function. One more interesting point is that, by comparing the fluctuation of curves in Fig. 5, Fig. 7, and Fig. 9, we can observe that impacts of μ and dn are much larger than impacts of the number of bins and the overlap rate.

IV. Conclusion

In this paper, we propose a MAPPER-induced manifold alignment for semi-supervised fusion of optical data and polarimetric SAR data, inspired by the semi-supervised technique and the emerging field of topological data analysis. Specifically, we embed a successful topological data analysis tool, MAPPER, into SSMA, to accomplish heterogeneous data fusion. Furthermore, our modified version of MAPPER functions adaptively to data by improving clustering. The performance of MIMA on fusing optical data and polarimetric SAR data is superior to that of SSMA, LPP, COSPACE, LeMA, and the feature concatenation, with respect to LCLU classification and LCZ classification. SSMA-based method is applied to fuse optical data and SAR data for the first time. This is also the first time that topological data analysis is applied in remote sensing field.

In the future, further experiments will be conducted to explore the potential of the proposed MIMA by selectively introducing field knowledge of remote sensing data. In this manner, physical meanings of different remote sensing data can be explicitly introduced into data fusion, instead of treating it as a data-driven machine learning problem. We believe an expert knowledge driven MIMA can further improve the fusion performance.

Acknowledgment

The authors would like to thank the following institutions for providing data sets used in this study: German Research Center for Geosciences, IEEE GRSS IADTC and OpenStreetMap.

References

[1] J. Zhang, "Multi-source remote sensing data fusion: status and trends," *International Journal of Image and Data Fusion*, vol. 1, no. 1, pp. 5–24, 2010.

[2] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1585–1601, 2015.

[3] L. Gómez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal classification of remote sensing images: A review and future directions," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1560–1584, 2015.

[4] X. Liu, C. Deng, J. Chanussot, D. Hong, and B. Zhao, "Stfnet: A two-stream convolutional neural network for spatiotemporal image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, 2019.

[5] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, "ORSIm Detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features," *arXiv preprint arXiv:1901.07925*, 2019.

[6] D. Hong and X. Zhu, "SULoRA: Subspace unmixing with low-rank attribute embedding for hyperspectral data analysis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1351–1363, 2018.

[7] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.

[8] N. Yokoya, P. Ghamisi, and J. Xia, "Multimodal, multitemporal, and multisource global data fusion for local climate zones classification based on ensemble learning," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017, pp. 1197–1200.

[9] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1923–1938, 2019.

[10] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[11] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, 2017.

[12] J. Hu, D. Hong, Y. Wang, and X. X. Zhu, "A comparative review of manifold learning techniques for hyperspectral and polarimetric sar image fusion," *Remote Sensing*, vol. 11, no. 6, p. 681, 2019.

[13] N. Merkle, S. Auer, R. Mller, and P. Reinartz, "Exploring the potential of conditional adversarial networks for optical and sar image matching," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 1–10, 2018.

[14] S. Auer, I. Hornig, M. Schmitt, and P. Reinartz, "Simulation-based interpretation and alignment of high-resolution optical and sar images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 11, pp. 4779–4793, 2017.

[15] L. H. Hughes, M. Schmitt, L. Mou, Y. Wang, and X. X. Zhu, "Identifying corresponding patches in sar and optical images with a pseudo-siamese cnn," *IEEE Geoscience and Remote Sensing Letters*, 2018.

[16] F. Tupin and M. Roux, "Detection of building outlines based on the fusion of sar and optical features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 1-2, pp. 71–82, 2003.

[17] M. Campos-Taberner, F. J. García-Haro, G. Camps-Valls, G. Grau-Muedra, F. Nutini, L. Busetto, D. Katsantonis, D. Stavrakoudis, C. Minakou, L. Gatti *et al.*, "Exploitation of sar and optical sentinel data to detect rice crop and estimate seasonal dynamics of leaf area index," *Remote Sensing*, vol. 9, no. 3, p. 248, 2017.

[18] K. Irwin, D. Beaulne, A. Braun, and G. Fotopoulos, "Fusion of sar, optical imagery and airborne lidar for surface water detection," *Remote Sensing*, vol. 9, no. 9, p. 890, 2017.

[19] C. Lamarche, M. Santoro, S. Bontemps, R. dAndrimont, J. Radoux, L. Giustarini, C. Brockmann, J. Wevers, P. Defourny, and O. Arino, "Compilation and validation of sar and optical data products for a complete and global map of inland/ocean water tailored to the climate modeling community," *Remote Sensing*, vol. 9, no. 1, p. 36, 2017.

[20] L. Fatone, P. Maponi, and F. Zirilli, "Fusion of sar/optical images to detect urban areas," in *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (Cat. No.01EX482)*, 2001, pp. 217–221.

[21] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *arXiv preprint arXiv:1902.10858*, 2019.

[22] D. Dimov, F. Löw, M. Ibrakhimov, G. Stulina, and C. Conrad, "Sar and optical time series for crop classification," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 1–4.

[23] R. Gaetano, D. Cozzolino, L. D'Amiano, L. Verdoliva, and G. Poggi, "Fusion of sar-optical data for land cover monitoring," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017, pp. 5470–5473.

[24] G. V. Laurin, V. Liesenberg, Q. Chen, L. Guerriero, F. Del Frate, A. Bartolini, D. Coomes, B. Wilebore, J. Lindsell, and R. Valentini, "Optical and sar sensor synergies for forest and land cover mapping in a tropical site in west africa," *International Journal of Applied Earth Observation and Geoinformation*, vol. 21, pp. 7–16, 2013.

[25] Z. Zhu, C. E. Woodcock, J. Rogan, and J. Kellndorfer, "Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using landsat and sar data," *Remote Sensing of Environment*, vol. 117, pp. 72–82, 2012.

[26] D. Hong, Z. Pan, and X. Wu, "Improved differential box counting with multi-scale and multi-direction: A new palmprint recognition method," *Optik-International Journal for Light and Electron Optics*, vol. 125, no. 15, pp. 4154–4160, 2014.

[27] J. Hu, L. Mou, A. Schmitt, and X. X. Zhu, "Fusionet: A two-stream convolutional neural network for urban scene classification using polsar and hyperspectral data," in *Urban Remote Sensing Event (JURSE), 2017 Joint*. IEEE, 2017, pp. 1–4.

[28] J. Ham, D. D. Lee, and L. K. Saul, "Semisupervised alignment of manifolds." in *AISTATS*, 2005, pp. 120–127.

[29] C. Wang and S. Mahadevan, "A general framework for manifold alignment." in *AAAI fall symposium: manifold learning and its applications*, 2009, pp. 53–58.

[30] D. Liao, Y. Qian, J. Zhou, and Y. Y. Tang, "A manifold alignment approach for hyperspectral image visualization with natural color," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3151–3162, 2016.

[31] D. Tuia, D. Marcos, and G. Camps-Valls, "Multi-temporal and multi-source remote sensing image classification by nonlinear relative normal-

ization," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 120, pp. 1–12, 2016.
[32] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 107, pp. 50–63, 2015.
[33] D. Tuia, M. Volpi, M. Trolliet, and G. Camps-Valls, "Semisupervised manifold alignment of multimodal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 12, pp. 7708–7720, 2014.
[34] D. Hong, N. Yokoya, and X. X. Zhu, "Learning a robust local manifold representation for hyperspectral dimensionality reduction," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 2960–2975, 2017.
[35] X. He and P. Niyogi, "Locality preserving projections," in *Advances in neural information processing systems*, 2004, pp. 153–160.
[36] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
[37] A. Hatcher, *Algebraic topology*. Tsinghua University Press, 2005.
[38] T. Lin and H. Zha, "Riemannian manifold learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 5, pp. 796–809, 2008.
[39] D. Hong, N. Yokoya, J. Xu, and X. Zhu, "Joint & progressive learning from high-dimensional data for multi-label classification," in *Proc. ECCV*, 2018, pp. 478–493.
[40] H. Chintakunta, M. Robinson, and H. Krim, "Introduction to the special session on topological data analysis, icassp 2016," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 6410–6414.
[41] F. Chazal and B. Michel, "An introduction to topological data analysis: fundamental and practical aspects for data scientists," *arXiv preprint arXiv:1710.04019*, 2017.
[42] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, 2005.
[43] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," in *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*. IEEE, 2000, pp. 454–463.
[44] G. Singh, F. Mémoli, and G. E. Carlsson, "Topological methods for the analysis of high dimensional data sets and 3d object recognition." in *SPBG*, 2007, pp. 91–100.
[45] M. Carriere, B. Michel, and S. Oudot, "Statistical analysis and parameter selection for mapper," *The Journal of Machine Learning Research*, vol. 19, no. 1, pp. 478–516, 2018.
[46] M. Nicolau, A. J. Levine, and G. Carlsson, "Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival," *Proceedings of the National Academy of Sciences*, vol. 108, no. 17, pp. 7265–7270, 2011.
[47] J. L. Nielson, J. Paquette, A. W. Liu, C. F. Guandique, C. A. Tovar, T. Inoue, K.-A. Irvine, J. C. Gensel, J. Kloke, T. C. Petrossian *et al.*, "Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury," *Nature communications*, vol. 6, p. 8581, 2015.
[48] L. Li, W.-Y. Cheng, B. S. Glicksberg, O. Gottesman, R. Tamler, R. Chen, E. P. Bottinger, and J. T. Dudley, "Identification of type 2 diabetes subgroups through topological analysis of patient similarity," *Science translational medicine*, vol. 7, no. 311, pp. 311ra174–311ra174, 2015.
[49] D. Hong, N. Yokoya, and X. Zhu, "The K-LLE algorithm for nonlinear dimensionality reduction of large-scale hyperspectral data," in *Proc. WHISPERS*. IEEE, 2016, pp. 1–5.
[50] D. Hong, W. Liu, J. Su, Z. Pan, and G. Wang, "A novel hierarchical approach for multispectral palmprint recognition," *Neurocomputing*, vol. 151, pp. 511–521, 2015.
[51] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in neural information processing systems*, 2002, pp. 849–856.
[52] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
[53] P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson, "Extracting insights from the shape of complex data using topology," *Scientific reports*, vol. 3, p. srep01236, 2013.
[54] D. Hong, W. Liu, X. Wu, Z. Pan, and J. Su, "Robust palmprint recognition based on the fast variation vese–osher model," *Neurocomputing*, vol. 174, pp. 999–1012, 2016.
[55] J. R. Munkres, *Topology*. Prentice Hall, 2000.
[56] G. McLachlan, *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons, 2004, vol. 544.
[57] D. Hong, N. Yokoya, and X. X. Zhu, "Local manifold learning with robust neighbors selection for hyperspectral dimensionality reduction," in *Proc. IGARSS*. IEEE, 2016, pp. 40–43.
[58] A. K. Jain and R. C. Dubes, "Algorithms for clustering data," 1988.
[59] A. Okujeni, S. Van Der Linden, and P. Hostert, "Berlin-urban-gradient dataset 2009-an enmap preparatory flight campaign (datasets)," *GFZ Data Services*, 2016.
[60] R. Hänsch, A. Ley, and O. Hellwich, "Correct and still wrong: The relationship between sampling strategies and the estimation of the generalization error," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017, pp. 3672–3675.
[61] D. L. Donoho *et al.*, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*, vol. 1, p. 32, 2000.
[62] M. D. Farrell and R. M. Mersereau, "On the impact of pca dimension reduction for hyperspectral detection of difficult targets," *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 2, pp. 192–195, 2005.
[63] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 2005.
[64] B. Rasti, P. Ghamisi, and R. Gloaguen, "Hyperspectral and lidar fusion using extinction profiles and total variation component analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3997–4007, 2017.
[65] M. Wurm, H. Taubenböck, M. Weigand, and A. Schmitt, "Slum mapping in polarimetric sar data using spatial features," *Remote sensing of environment*, vol. 194, pp. 190–204, 2017.
[66] J. Hu, P. Ghamisi, and X. X. Zhu, "Feature extraction and selection of sentinel-1 dual-pol data for global-scale local climate zone classification," *ISPRS International Journal of Geo-Information*, vol. 7, no. 9, p. 379, 2018.
[67] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, "Cospace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, 2019.
[68] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 147, pp. 193–205, 2019.
[69] B. Schölkopf, A. J. Smola *et al.*, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
[70] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

Jingliang Hu (S'19) received the B.Sc. degree in Geographic Information System from Chengdu University of Technology, China, in 2011, the M.Sc. degree in Cartography and Geographic Information System from Wuhan University, China, in 2014, and M.Sc. degree in Space Science and Technology from Technical University of Munich (TUM), Germany, in 2014. He is currently pursuing his Ph.D. degree in Signal Processing in Earth Observation, Technical University of Munich (TUM), Germany. Since July 2015, he is a Research Associate at Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Germany. His research interests include topological data analysis (TDA), data fusion, machine learning, and deep learning.

Danfeng Hong (S'16) received the B.Sc. degree in Computer Science and Technology from the Neusoft College of Information, Northeastern University, China, in 2012, the M.Sc. degree in Computer Vision, Qingdao University, China, in 2015. He is currently pursuing his Ph.D. degree in Signal Processing in Earth Observation, Technical University of Munich (TUM), Munich, Germany.

Since 2015, he is also a Research Associate at Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. In 2018, he was a visiting student in GIPSA-lab, Grenoble INP, CNRS, Univ. Grenoble Alpes, Grenoble, France, under the supervision of Prof. Jocelyn Chanussot.

His research interests include signal / image processing and analysis, pattern recognition, machine / deep learning and their applications in Earth Vision.

Xiao Xiang Zhu (S'10–M'12–SM'14) received the Master (M.Sc.) degree, her doctor of engineering (Dr.-Ing.) degree and her Habilitation in the field of signal processing from Technical University of Munich (TUM), Munich, Germany, in 2008, 2011 and 2013, respectively.

She is currently the Professor for Signal Processing in Earth Observation (www.sipeo.bgu.tum.de) at Technical University of Munich (TUM) and German Aerospace Center (DLR); the head of the department "EO Data Science" at DLR's Earth Observation Center; and the head of the Helmholtz Young Investigator Group "SiPEO" at DLR and TUM. Prof. Zhu was a guest scientist or visiting professor at the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, the University of Tokyo, Tokyo, Japan and University of California, Los Angeles, United States in 2009, 2014, 2015 and 2016, respectively. Her main research interests are remote sensing and Earth observation, signal processing, machine learning and data science, with a special application focus on global urban mapping.

Dr. Zhu is a member of young academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She is an associate Editor of IEEE Transactions on Geoscience and Remote Sensing.

Appendices

C Hu J, Hong D, Wang Y, and Zhu X X (2019). A comparative review of manifold learning techniques for hyperspectral and polarimetric sar image fusion. Remote Sensing, 11(6), 681.

https://www.mdpi.com/2072-4292/11/6/681

remote sensing

MDPI

Article

A Comparative Review of Manifold Learning Techniques for Hyperspectral and Polarimetric SAR Image Fusion

Jingliang Hu [1,2,*], Danfeng Hong [1,2], Yuanyuan Wang [2] and Xiao Xiang Zhu [1,2,*]

[1] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany; danfeng.hong@dlr.de

[2] Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), 80333 Munich, Germany; y.wang@tum.de

* Correspondence: jingliang.hu@dlr.de (J.H.); xiaoxiang.zhu@tum.de (X.X.Z.)

Received: date; Accepted: date; Published: date

Abstract: In remote sensing, hyperspectral and polarimetric synthetic aperture radar (PolSAR) images are the two most versatile data sources for a wide range of applications such as land use land cover classification. However, the fusion of these two data sources receive less attention than many other, because of their scarce data availability, and relatively challenging fusion task caused by their distinct imaging geometries. Among the existing fusion methods, including manifold learning-based, kernel-based, ensemble-based, and matrix factorization, manifold learning is one of most celebrated techniques for the fusion of heterogeneous data. Therefore, this paper aims to promote the research in hyperspectral and PolSAR data fusion, by providing a comprehensive comparison between existing manifold learning-based fusion algorithms. We conducted experiments on 16 state-of-the-art manifold learning algorithms that embrace two important research questions in manifold learning-based fusion of hyperspectral and PolSAR data: (1) in which domain should the data be aligned—the *data domain* or the *manifold domain*; and (2) how to make use of existing *labeled data* when formulating a graph to represent a manifold—supervised, semi-supervised, or unsupervised. The performance of the algorithms were evaluated via multiple accuracy metrics of land use land cover classification over two data sets. Results show that the algorithms based on manifold alignment generally outperform those based on data alignment (data concatenation). Semi-supervised manifold alignment fusion algorithms performs the best among all. Experiments using multiple classifiers show that they outperform the benchmark data alignment-based algorithms by ca. 3% in terms of the overall classification accuracy.

Keywords: data fusion; generalized graph fusion; hyperspectral image; data alignment; locality preserving projections; manifold alignment; manifold learning; MAPPER-induced manifold alignment; polarimetric SAR; manifold alignment; MIMA

1. Introduction

1.1. Related Work

Multi-modal data fusion [1–7] continuously draws attention in the remote sensing community. The fusion of optical and synthetic aperture radar (SAR) data, two important yet intrinsically different data sources, has also began to appear frequently in the context of multi-modal data fusion [8–14]. With the rapid development of Earth observation missions, such as the Sentinel-1 [15], Sentinel-2 [16], and the upcoming EnMAP [17], the availability of both data sources create a huge potential for Earth-oriented information retrieval. Among all optical data [18–20], hyperspectral data are well

known for their distinguishing power that originates from their rich spectral information [21–24]. Similarly, polarimetric SAR (PolSAR) data are a popular choice for classification task in the field of SAR because it can reflect the geometric and the dielectric property of the scatterers [25–29]. It is of great interest to investigate the fusion of hyperspectral and PolSAR images, especially with the application to land use land cover classification (LULC).

Few studies attempted to address the challenge of fusing hyperspectral and PolSAR data. Jouan and Allard [30] proposed a hierarchical fusion strategy for land cover classification using PolSAR and hyperspectral images. In their work, hyperspectral images are firstly used to distinguish vegetation and non-vegetation area. The PolSAR data are used to classify the non-vegetation area to man-made objects, water, or bare soil. Li et al. [31] applied feature level and decision level fusion using hyperspectral and PolSAR data. They combined the parameters of scattering mechanism of the PolSAR data and the features of the hyperspectral image to create a concatenation of features. The classification results of multiple classifiers are then merged using decision fusion. An application of spilled oil detection was studied by Dabbiru et al. [32] using hyperspectral and PolSAR data. They applied pixel level concatenation on the data, and employed support vector machine (SVM) as the classifier. Hu et al. [33] proposed a framework for fusing hyperspectral and PolSAR data based on the segmented objects that provide spatial information. A two-stream convolutional neural network (CNN) was introduced in [34] that takes advantage of the feature extraction power of CNN.

Among the existing fusion methods, including manifold learning-based [35,36], kernel-based [37], ensemble-based [38,39], tensor-based [40,41], and matrix factorization [42], manifold learning is one of most celebrated techniques. However, although it has been proven as a powerful technique in the field of data fusion, it is barely studied in the scope of fusing hyperspectral and PolSAR data. Generally, manifold-based fusion techniques attempt to find a shared latent space where the original data sets can be fused or aligned. Wang and Mahadevan [35,43–45] proposed several manifold-based techniques to find the correspondence of data sets which describe the same object from different aspects via the latent space. A kernel based manifold alignment [46] searches the latent space from a kernel space of the original data, because the kernel space has a better representation of the data than the original feature space of the data. In remote sensing, it was introduced in [36,47] that the manifold latent space is able to align multiple optical data sets and improve the LULC classification. A manifold-based data alignment technique was introduced in [48] for the fusion of hyperspectral and Lidar data with application to classification. Besides data fusion, various manifold techniques can be found in remote sensing field for detection [49], visualization [50], and dimension reduction [51].

1.2. Scope of This Paper

When fusing data with manifold techniques, one technical question is that: in which domain should the fusion be carried on? We categorized the existing techniques into two types: (1) data alignment-based approach and (2) manifold alignment-based approach. As shown in the left of Figure 1, the data alignment approach carries out the fusion in the data domain. As the simplest example, it fuses the data by concatenation, and carries out a manifold-based dimension reduction. Essentially, this approach assumes that an intrinsic manifold exists in the concatenated data. Representatives of this approach are the locality preservation projection (LPP) [52] and the generalized graph fusion (GGF) [48]. On the contrary, the manifold alignment-based approach carries out the fusion on manifolds which are separately derived from different data sources. This is demonstrated in the right of Figure 1. The assumption of this approach is that different manifold exists in each data source. Those manifolds can be aligned in a latent space. Representative algorithms are the manifold alignment (MA) [36] and the MAPPER-Induced manifold alignment (MIMA) [53].

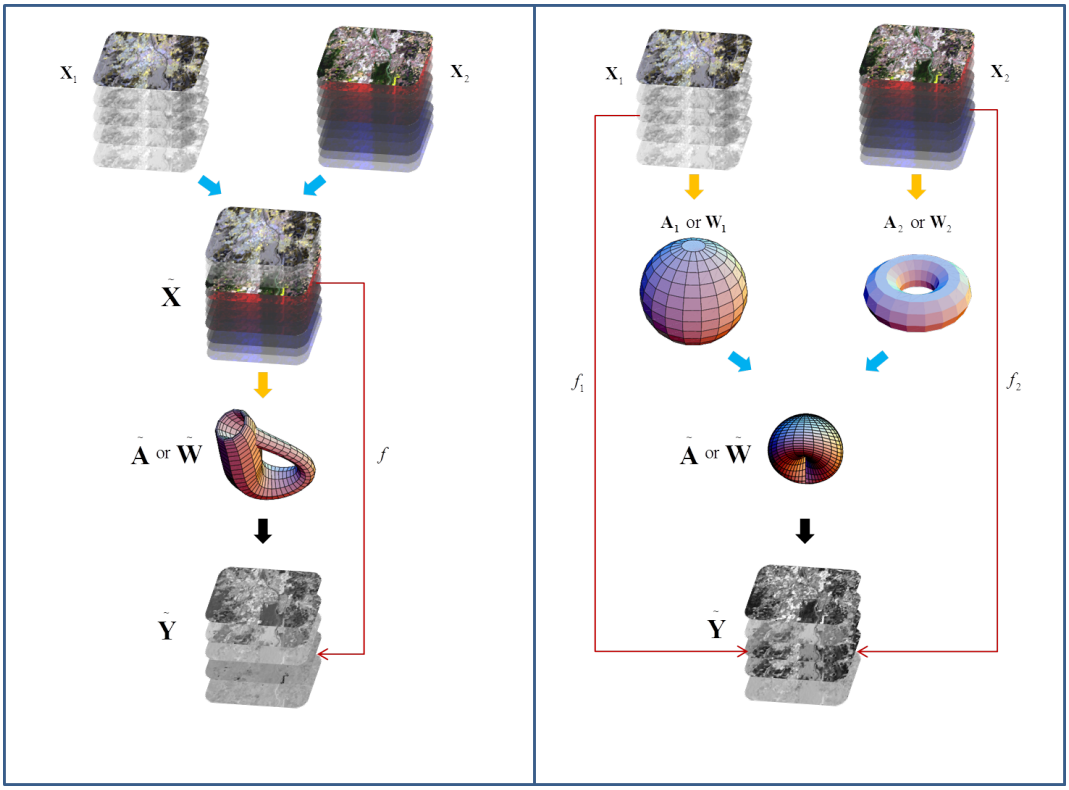


Figure 1. Frameworks of manifold learning fusion techniques. Left: the data alignment fusion; Right: the manifold alignment fusion. The blue arrow indicates the fusion step. The yellow arrow indicates where the modeling of manifold takes place. The black arrow indicates the feature extraction. $\mathbf{X}_i$: ith data source; **A** or **W**: mathematical modeling of manifolds; **Y**: fused feature; **f**: the learned projection. $\tilde{\cdot}$: a fusion of certain form.

The other essential research question of manifold-based fusion is: how should the manifold be extracted? We categorize the existing techniques into three learning strategies in terms of the usage of labeled data: unsupervised, semi-supervised, and supervised. When modeling a manifold, a general assumption is that, hidden in the data representation, there exists an underlying lower dimensional manifold where the data truly distributes [54]. Early studies [54–57] model manifolds by following a geometric assumption that the Riemannian manifold can be locally approximated by Euclidean measures. The geometric assumption models the manifold in an unsupervised manner by using k-nearest-neighbor (kNN). With the presence of labeled data, the manifold can be jointly modeled by the Riemannian manifold and the labeled data. For example, one can construct the manifold in a semi-supervised fashion by using both kNN and the labeled data [35,36]. The manifold can also be modeled in a supervised manner by using only the labeled data. One of the main goals of this paper is to investigate the impact of these learning strategies on the classification performance on the fused data.

1.3. Contribution of This Paper

This paper investigates the performance of manifold learning technique on the fusion of hyperspectral and PolSAR data, based on four state-of-art algorithms, locality preservation projection (LPP) [52], generalized graph fusion (GGF) [48], manifold alignment (MA) [36,44], and MAPPER-induced manifold alignment (MIMA) [53]. We implemented 16 variants of the four algorithms which involve the above-mentioned two alignment approaches and the three manifold learning strategies. These algorithms were tested on two study areas for a LULC classification task with five classifiers: one nearest neighbor (*1NN*) [58], linear SVM (*LSVM*) [59,60], Gaussian kernel SVM (*KSVM*) [59,60], random forest (*RF*) [61], and canonical correlation forest (*CCF*) [62]. We avoided

any deep network classifiers, because the goal of this article is to solely evaluate the performance of multi-sensory data fusion. In total, 80 classification maps were produced for each study area, based on which a comprehensive discussion was carried out. The main contributions of this paper are as follows:

- An exhaustive investigation of existing manifold learning techniques. A sufficient number of manifold techniques and classifiers were tested on the fusion of hyperspectral and PolSAR data in terms of classification. It provides a reliable demonstration on the performance of the manifold technique regarding hyperspectral and PolSAR data fusion.
- An objective comparison of the performance of different manifold data fusion algorithms. To avoid any fortuity, five classifiers were applied for the classification. A grid search was applied to all tunable hyperparameters of those algorithms. The best classification accuracies are compared.
- A comprehensive analysis of the results. The experiment results were analyzed in regard to two fusion approaches, three manifold learning strategies, four basic algorithms, and five classifiers.

1.4. Structure of This Paper

The second section recalls the theory of manifold technique and the four selected state-of-art algorithms. The third section describes the data sets used in the experiments, introduces the experiment setting, and carries out the discussion. The fourth section concludes the paper. Table 1 also lists the symbols used in this article for a better understanding of the content of the article.

Table 1. The notations used in this article.

Notation	Explanation
$\mathbf{X}_i$	The ith data source
$\mathcal{M}_i$	The manifold of the $\mathbf{X}_i$
$\mathbf{x}_i^p$	The pth instance of the $\mathbf{X}_i$
m_i	The number of dimensions of the $\mathbf{X}_i$
$\mathbf{E}_i$	The labeled subset of the $\mathbf{X}_i$
$\mathbf{y}_i^p$	The pth instance of the $\mathbf{Y}_i$
l_i	The number of dimensions of the $\mathbf{Y}_i$
$\mathbf{F}$	The filter function in MAPPER
$\mathbf{W}$	The weight matrix that models a manifold
$\mathbf{D}$	The degree matrix of a graph
$\tilde{}$	The fusion at certain form
$\mathcal{L}$	The loss fuction
dn	The dimension of underlying manifold
b	The number of bins in MAPPER
λ	The eigenvalue of generalized eigenvalue decomposition
K	The total number of data sources
$\mathbf{Y}_i$	The data representation of the $\mathcal{M}_i$
$\mathbf{x}_i^q$	The qth instance of the $\mathbf{X}_i$
n_i	The number of instances of the $\mathbf{X}_i$
n_i^*	The number of instances of the $\mathbf{E}_i$, $n_i^* < n_i$
$\mathbf{y}_i^q$	The qth instance of the $\mathbf{Y}_i$
$\mathbf{f}$	The projection $\mathbf{Y} = \mathbf{f}^T\mathbf{X}$
$\mathbf{A}$	The binary matrix that models a manifold
σ	The filtering parameter of weight matrix
$\mathbf{L}$	The Laplacian matrix of a graph
$\mathcal{D}$	The pairwise distance matrix
k	The number of local neighbors
μ	The weighting of topology structure in MA
c	The overlap rate in MAPPER

2. Materials and Methods

In this section, the general concept of the manifold technique is introduced, with the help of necessary mathematical notations. Meanwhile, the theoretical impact of different learning strategies

to the fusion result is discussed. Then, the following up sub-sections recall the principles of the four selected state-of-art manifold fusion techniques, namely LPP [52], GGF [48], MA [36,44], and MIMA [53]. Pseudo-codes of these four algorithms are listed in the Appendix A–D, which provides the technical details.Finally, the data sets and the experiment settings are introduced in detail.

2.1. Manifold Technique, Learning Strategy, and Notations

Let $\mathbf{X}_i = [\mathbf{x}_i^1, ..., \mathbf{x}_i^p, ..., \mathbf{x}_i^{n_i}] \in \mathbb{R}^{m_i \times n_i}$ be a matrix representing the ith data source, with m_i dimensions by n_i instances. For simplicity, the subscript i is omitted in the following content when only one data source is involved. The m_i-dimensional data space is named as the feature space of data $\mathbf{X}_i$ in this paper. The term $\mathbf{x}_i^p$ denotes the pth instance of the ith data source. Let K denote the total number of data sources.

A manifold $\mathcal{M}$ is a smooth hyper-surface embedded in a higher dimensional space [56], e.g., the surface of a sphere is a 2D manifold in a 3D space. The underlying assumption of the manifold technique is that, for a data $\mathbf{X} \in \mathbb{R}^{m \times n}$ of redundant m dimensions, there exists a low dimensional intrinsic manifold $\mathcal{M}$ where the data distributes [54,57,63,64]. The goal of a manifold technique is to pursue a representation, realized by a projection $\mathbf{Y} = [\mathbf{y}^1, ..., \mathbf{y}^p, ..., \mathbf{y}^n] \in \mathbb{R}^{l \times n}, l < m$ of the original data, of the manifold $\mathcal{M}$. In order to approximate $\mathbf{Y}$, the bridging property is that the data point $\mathbf{y}^p$ on the manifold is locally homeomorphic to its counterpart $\mathbf{x}^p$ in the feature space [56]. It means that a data point has identical local structures in its intrinsic manifold and in its feature space. With this property, a variety of methods [54,55,57,65] extract the local structure of a data [52,66–69] in its feature space as an estimation to the local structure in its intrinsic manifold, with different locality criterion. All those methods pursue an optimized projection $\mathbf{f}$ which maps data from the feature space to a representation ($\mathbf{Y} = \mathbf{f}^T\mathbf{X}$) of the intrinsic manifold $\mathcal{M}$. In terms of the manifold technique for data fusion [36,44,48], the aim is to find the projection which maps multiple data sources $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_K\}$ into a fused manifold $\tilde{\mathcal{M}}$ where the fused data locates.

The centerpiece of the abovementioned algorithms is the modeling of the manifold. Usually, an intrinsic manifold of the data is modeled by an $n \times n$ symmetric binary matrix $\mathbf{A}$ that describes the connection among the data points. $\mathbf{A}(p, q) = 1$ for a confirmed connection between $\mathbf{x}_i^p$ and $\mathbf{x}_i^q$ while $\mathbf{A}(p, q) = 0$ otherwise. $\mathbf{A}$ can be generalized to an $n \times n$ symmetric weight matrix $\mathbf{W}$. Different from $\mathbf{A}$, $\mathbf{W}(p, q)$ takes a real value in $[0, 1]$, which describes the strength of the connection between $\mathbf{x}_i^p$ and $\mathbf{x}_i^q$. Essentially, $\mathbf{A}$ and $\mathbf{W}$ are the adjacency and weight matrix of a graph that captures the topology of the manifold. As introduced in [52], the manifold structure ($\mathbf{A}$ or $\mathbf{W}$) can be defined from different perspectives. In this paper, we would like to categorize these perspectives based on how the labeled data is utilized for modeling the manifold, namely the unsupervised learning, supervised learning, and semi-supervised learning.

- The unsupervised learning takes the original geometric assumption that the manifold and the original data space share the same local property. Besides the geometric measure, model-based similarity measurement can also be used to build up the structure of the manifold. The key point is that the definition of the similarity measurement is capable of revealing the underlying distribution of the data or the physical information in the data.
- The supervised learning assumes that a given set of labeled data includes sufficient amount of inter- and intra-class connections among the data points, so that they can well capture the topology of the manifold. As a result, the underlying manifold is directly defined by the label information. Thus, the quality of the label has a great impact.
- The semi-supervised learning pursues a manifold where the data distribution partially correlates to the label information and partially associates to the distribution predefined by a similarity measurement. This manifold implicitly propagates the label information to the unlabeled data.

2.2. Locality Preservation Projection (LPP)

LPP aims to find a lower dimensional representation $\mathbf{Y}$ of the original data $\mathbf{X}$ which reflects the intrinsic manifold $\mathcal{M}$. According to the geometric assumption that the intrinsic manifold and the original data share the same local properties, the lower dimensional representation $\mathbf{Y}$ achieved by LPP preserves the local structure of the original data $\mathbf{X}$. The locality is defined by either the k-nearest-neighbor or the ϵ-neighborhoods [52] and is mathematically described in a weight matrix $\mathbf{W}$ as Equation (1).

$$\mathbf{W}(p,q) = \begin{cases} e^{-\frac{||x^p - x^q||^2}{\sigma}} & x^p \text{ and } x^q \text{ are local neighbors} \\ 0 & x^p \text{ and } x^q \text{ are not local neighbors} \end{cases} \tag{1}$$

where σ is a filtering parameter.

LPP pursues an optimized projection $\mathbf{f}$ which maps the data $\mathbf{X}$ to a lower dimensional representation $\mathbf{Y} = \mathbf{f}^T\mathbf{X}$. As the local structure of the intrinsic manifold is modeled by Equation (1), minimizing the objective function expressed by Equation (2) encourages the preservation of the derived local structure in the intrinsic manifold.

$$\mathcal{L} = \sum_{pq}(y^p - y^q)\mathbf{W}(p,q) = \sum_{pq}(\mathbf{f}^T x^p - \mathbf{f}^T x^q)\mathbf{W}(p,q). \tag{2}$$

Thus, the optimization is formulated as follow:

$$\min_{\mathbf{f}} \sum_{pq}(\mathbf{f}^T x^p - \mathbf{f}^T x^q)\mathbf{W}(p,q). \tag{3}$$

Proven in [52], the solution that minimizes the objective function $\mathcal{L}(\mathbf{f})$ is given by the minimum eigenvalue solution to the generalized eigenvalue problem expressed in Equation (4).

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{f} = \lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{f}, \tag{4}$$

where $\mathbf{D}$ is the degree matrix; if $p = q$, $\mathbf{D}(p,q) = \sum_{p=1}^{p=n}\mathbf{W}(p,q)$, otherwise, $\mathbf{D}(p,q) = 0$; and the $\mathbf{L}$ is the Laplacian matrix, $\mathbf{L} = \mathbf{D} - \mathbf{W}$.

As brief described above, LPP is originally designed to as a dimension reduction algorithm, instead of data fusion. However, it is essential to include this algorithm in the scope of this paper, because (1). When conducting manifold fusion, the dimension reduction is also accomplished as a side effect. Due to the well-known curse-of-dimensionality [70], classification on selective subset of dimensions could result in better performance than using the data with all dimensions [71]. LPP can serve as a baseline algorithm to reduce the dimension of the data; and (2). LPP is essentially a manifold learning technique. Some data fusion algorithms [48,72] are developed on the idea of data alignment using the LPP.

2.3. Generalized Graph-Based Fusion (GGF)

GGF is originally proposed to fuse hyperspectral data and LiDAR data for land cover classification [48]. Its fusion strategy comprises a joint LPP dimension reduction and an additional constraint that captures the common local structure that exists in all data sources.

Technically, GGF concatenates K data sources ($\mathbf{X}_i = [\mathbf{x}_i^1, ..., \mathbf{x}_i^p, ..., \mathbf{x}_i^n] \in \mathbb{R}^{m_i \times n}, i \in \{1, 2, ..., K\}$) into a stack ($\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}^1, ..., \tilde{\mathbf{x}}^p, ..., \tilde{\mathbf{x}}^n] \in \mathbb{R}^{(m_1+m_2+...+m_K)\times n}$) which are treated as one data sources in its high dimensional feature space. Therefore, GGF is essentially a LPP carried out on the data stack $\tilde{\mathbf{X}}$, with an additional constraint. The constraint assumes that the connectivity $\tilde{\mathbf{A}}$ of the fused intrinsic manifold $\tilde{\mathcal{M}}$ should be a complete subset of the connectivity matrices of the manifolds $\mathcal{M}_i$ of the individual data sources $\mathbf{X}_i, i \in \{1, 2, ..., K\}$. Thus, the assumption is formulated as Equation (5).

$$\tilde{\mathbf{A}} = \mathbf{A}_1 \odot \mathbf{A}_2 \odot, ..., \odot \mathbf{A}_K \tag{5}$$

where $\odot$ indicates element-wise multiplication.

The manifold constraint $\tilde{\mathbf{A}}$ is embedded into a n by n pairwise distance matrix $\tilde{\mathcal{D}}$ ($\tilde{\mathcal{D}}(p,q) = ||\tilde{\mathbf{x}}^p - \tilde{\mathbf{x}}^q||$), which is expressed by Equation (6) where $\neg$ means logical operator negative, and $max(\cdot)$ means the maximum value of all elements in $'\cdot'$. The distance between any two data points that are not connected according to $\tilde{\mathbf{A}}$ is penalized with the maximum distance value of $\tilde{\mathcal{D}}$. The final distance matrix is named as $\tilde{\mathcal{D}}_{\mathbf{GGF}}$.

$$\tilde{\mathcal{D}}_{\mathbf{GGF}} = \tilde{\mathcal{D}} + (\neg\tilde{\mathbf{A}})max(\tilde{\mathcal{D}}) \tag{6}$$

The weight matrix $\tilde{\mathbf{W}}$ of the intrinsic manifold is then as follows.

$$\tilde{\mathbf{W}}(p,q) = \begin{cases} e^{-\tilde{\mathcal{D}}_{GGF}(p,q)} & x^p \text{ and } x^q \text{ are local neighbors} \\ 0 & x^p \text{ and } x^q \text{ are not local neighbors} \end{cases} \tag{7}$$

After achieving the weight matrix $\tilde{\mathbf{W}}$, similar to the LPP, the optimized projection $\mathbf{f}$ is given by the minimum eigenvalue solution to the generalized eigenvalue problem in Equation (8).

$$\tilde{\mathbf{X}}\tilde{\mathbf{L}}\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}\tilde{\mathbf{D}}\tilde{\mathbf{X}}^T\mathbf{f} \tag{8}$$

where $\tilde{\mathbf{D}}$ is the degree matrix. If $p = q$, $\tilde{\mathbf{D}}(p,q) = \sum_{p=1}^{p=n}\tilde{\mathbf{W}}(p,q)$, otherwise, $\tilde{\mathbf{D}}(p,q) = 0$. $\tilde{\mathbf{L}}$ is the Laplacian matrix, $\tilde{\mathbf{L}} = \tilde{\mathbf{D}} - \tilde{\mathbf{W}}$.

2.4. Manifold Alignment (MA)

Manifold alignment [35,36,44] aims to learn a set of projections $\{\mathbf{f}_1, ..., \mathbf{f}_K\}$ that (1) apply to individual data sources $\mathbf{X}_i$ in order to obtain their individual manifolds $\mathcal{M}_i$, and (2) align those obtained manifolds $\{\mathcal{M}_1, ..., \mathcal{M}_K\}$ to each other.

Designed in [36,44], three properties hold in the fused manifold: (***a***) data of the same class should locate close to each other; (***b***) data of different classes should locate far from one another; and (***c***) the intrinsic manifolds of individual data are preserved. These three properties are respectively formulated by the following three connection matrices $\tilde{\mathbf{A}}_s$ (9), $\tilde{\mathbf{A}}_d$ (10), and $\tilde{\mathbf{A}}_g$ (11).

$$\tilde{\mathbf{A}}_s = \begin{pmatrix} \mathbf{A}_s^{1,1} & ... & \mathbf{A}_s^{1,K} \\ ... & ... & ... \\ \mathbf{A}_s^{K,1} & ... & \mathbf{A}_s^{K,K} \end{pmatrix} \tag{9}$$

The connection matrix of similarity (9) is computed by the labeled information to pursue property (***a***).

$$\tilde{\mathbf{A}}_d = \begin{pmatrix} \mathbf{A}_d^{1,1} & ... & \mathbf{A}_d^{1,K} \\ ... & ... & ... \\ \mathbf{A}_d^{K,1} & ... & \mathbf{A}_d^{K,K} \end{pmatrix} \tag{10}$$

The connection matrix of dissimilarity is modeled as (10) to accomplish property (***b***), which is also computed from the label.

$$\tilde{\mathbf{A}}_g = \begin{pmatrix} \mathbf{A}_g^{1,1} & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & \mathbf{A}_g^{K,K} \end{pmatrix} \tag{11}$$

The connection matrix (11) describes the manifolds of individual data sources by using kNN, which aims at the property ***(c)***. All of the matrices (9)–(11) have the size of $(n_1 + n_2 + ... + n_k) \times (n_1 + n_2 + ... + n_k)$. In each matrix, the superscript i, j, e.g., $\mathbf{A}^{i,j}$, represents the relationship between the ith and jth data sources.

With connection matrices (9)–(11), three terms are formulated as Equations (12)–(14) to preserve the three properties, respectively.

$$\mathcal{A} = \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{p=1}^{n_i}\sum_{q=1}^{n_j} \|\mathbf{f}_i^T x_i^p - \mathbf{f}_j^T x_j^q\|^2 \tilde{\mathbf{A}}_s^{i,j}(p,q). \tag{12}$$

Minimizing Equation (12) pulls data of the same class together, which meets property (***a***).

$$\mathcal{B} = \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{p=1}^{n_i}\sum_{q=1}^{n_j} \|\mathbf{f}_i^T x_i^p - \mathbf{f}_j^T x_j^q\|^2 \tilde{\mathbf{A}}_d^{i,j}(p,q). \tag{13}$$

Maximizing Equation (13) pushes data of different classes away, which meets property (***b***).

$$\mathcal{C} = \sum_{i=1}^{K}\sum_{p=1}^{n_i}\sum_{q=1}^{n_i} \|\mathbf{f}_i^T x_i^p - \mathbf{f}_i^T x_i^q\|^2 \tilde{\mathbf{A}}_g^{i,i}(p,q). \tag{14}$$

Minimizing Equation (14) preserves the geometric structure of individual data sources, which corresponds to property (***c***). The terms (12)–(14) jointly construct the objective unction (15):

$$\mathcal{L}(\mathbf{f}_1, ..., \mathbf{f}_K) = (\mathcal{A} + \mathcal{C})/\mathcal{B}, \tag{15}$$

and hence an optimization problem (16) can be written as

$$\min_{\mathbf{f}_1,...,\mathbf{f}_K} \mathcal{L}(\mathbf{f}_1, ..., \mathbf{f}_K). \tag{16}$$

Proven in [35], the solution $\{\mathbf{f}_1, ..., \mathbf{f}_K\}$ that minimizing the cost function $\mathcal{L}(\mathbf{f}_1, ..., \mathbf{f}_K)$ is given by the smallest non-zero eigenvectors of the generalized eigenvalue decomposition of (17).

$$\tilde{\mathbf{X}}(\mu\tilde{\mathbf{L}}_g + \tilde{\mathbf{L}}_s)\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}\tilde{\mathbf{L}}_d\tilde{\mathbf{X}}^T\mathbf{f}, \tag{17}$$

where

$$\tilde{\mathbf{X}} = \begin{pmatrix} \mathbf{X}_1 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & ... & 0 & \mathbf{X}_K \end{pmatrix},$$

$$\tilde{\mathbf{L}}_{\{s,d,g\}} = \tilde{\mathbf{A}}_{\{s,d,g\}} - \tilde{\mathbf{D}}_{\{s,d,g\}},$$

$$\tilde{\mathbf{D}}_{\{s,d,g\}}(p,q) = \begin{cases} \sum_{q=1}^{m_1+...+m_k} \tilde{\mathbf{A}}_{\{s,d,g\}}(p,q) & p = q \\ 0 & p \neq q \end{cases}.$$

The matrices $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{L}}$ with subscript s, d, and g are the degree matrices and the Laplacian matrices, respectively.

2.5. MAPPER-Induced Manifold Alignment (MIMA)

MIMA is designed to fuse optical and PolSAR data for the purpose of LULC classification [53]. It follows the framework of MA [36,44] yet introduces a novel constraint term which originates from a recent field of topological data analysis (TDA). TDA has emerged as a new mathematical sub-field of big data analysis that aims to derive relevant information from the topological property of a data [73–77]. One TDA tool, named MAPPER [78], has been proven capable of revealing unknown insights in medical studies, by interpreting topological structures of data sets [79–82]. As a brief introduction, the MAPPER requires a filter function as an input which projects the data into a parameter space.

The original data is sorted into overlapping bins guided by the projected parameter. MAPPER carries out clustering of data points in each of the data bins, respectively. Afterwards, MAPPER models a graph where a node represents a cluster and an edge links two clusters that share common data points. Finally, a simplified graph is built up to represent the shape of the data. Such graph is an approximation of the Reeb graph [83] .

Technically, MIMA pursues the solution $\{\mathbf{f}_1, ..., \mathbf{f}_K\}$ by solving the same generalized eigenvalue decomposition as in Equation (17), except the the connection matrix of geometry $\tilde{\mathbf{A}}_g$ (Equation (11)) is replace by the MAPPER-derived connection matrix $\tilde{\mathbf{A}}_{MIMA}$ where $\mathbf{A}^{i,i}_{MIMA}(p,q) = 1$ if x_i^p and x_i^q belongs to the same cluster or belongs to two separated but linked clusters; $\mathbf{A}^{i,i}_{MIMA}(p,q) = 0$ elsewhere. Comparing to $\tilde{\mathbf{A}}_g$, $\tilde{\mathbf{A}}_{MIMA}$ introduces some unique properties that are listed as follows:

- Field knowledge. An expertise knowledge is introduced by the selection of the filter function. It defines a perspective of viewing the data while deriving the structure.
- A regional-to-global structure. Clustering in each data bin provides a regional structure. The design of overlapping bins combines the regional structures into a global one. It makes the derived structure more robust to outliers than the one derived by *kNN*.
- A data-driven regional structure. A spectral clustering is applied in the step, which is capable of detecting the number of clusters by the concept of eigen-gap [84]. It allows the derived structure constraining to the data distribution.

2.6. Data Description

Two sets of real data were used to investigate the manifold learning techniques for the fusion of hyperspectral and PolSAR data. The two data sets are in city of Berlin, Germany, and Augsburg, Germany.

2.6.1. The Berlin Data Set

In the Berlin data set, the hyperspectral image is a synthetic spaceborne EnMAP scene synthesized from airborne HyMap data. It has a size of 817 by 220 pixels, a 30-m ground sampling distance (GSD), and 244 spectral bands ranging from 400 nm to 2500 nm [85]. The dual-channel PolSAR data is a VH-VV polarized Sentinel-1 data acquired in interferometric wide swath mode. The Sentinel-1 SLC data is preprocessed using ESA SNAP toolbox and filtered by a non-local mean filter [86]. The PolSAR data has a GSD of 13 m and a size of 1723 by 476 pixels. The ground truth is a land use land cover map derived from Open Street Map (OSM) data [87]. The ground truth labels are spatially separated into a training data set and a testing data set shown in Figure 2. The details of the training and testing data sets are summarized in Table 2.

Table 2. Summary of the training data and the testing data for the scene of city Berlin.

Class	# of Training Sample	# of Testing Sample
Forest	298	52,455
Residential area	756	262,903
Industrial area	296	17,462
Low plants	344	56,683
Soil	428	14,505
Allotment	281	11,322
Commercial area	560	20,909
Water	153	5539

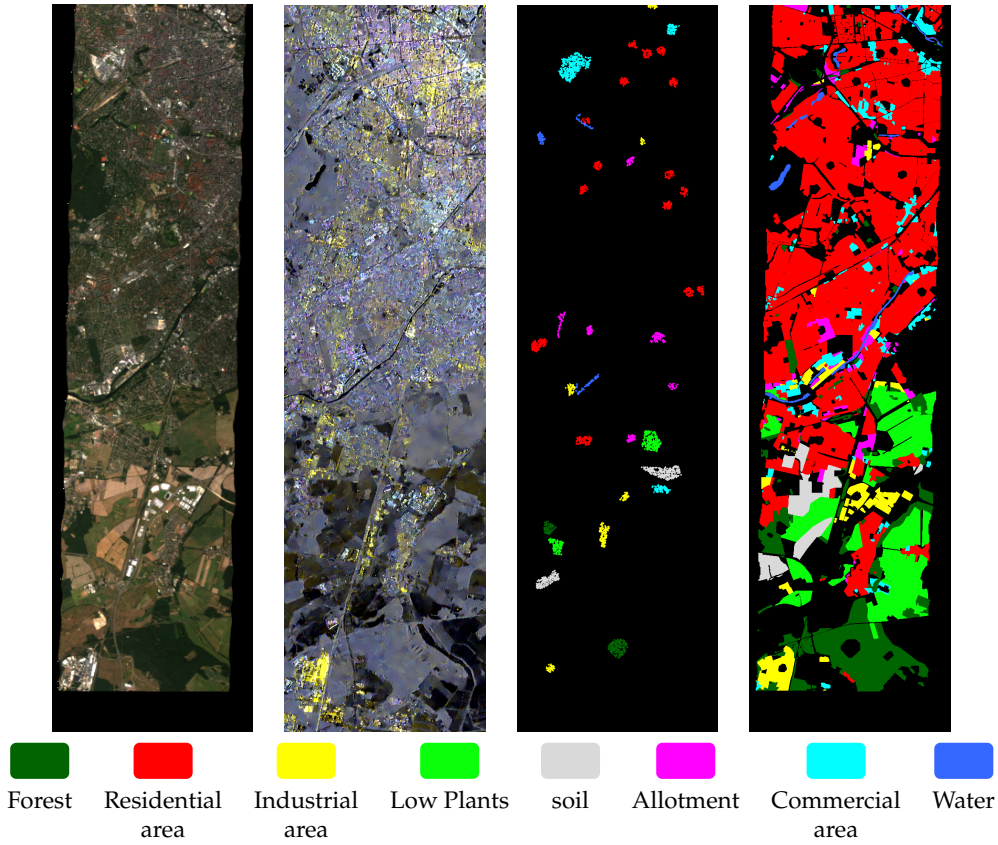


Figure 2. The Berlin data set. From left to right: RGB components of the simulated EnMAP data; Sentinel-1 dual-Pol data; The training data; The testing data.

2.6.2. The Augsburg Data Set

Similar to the Berlin data set, the hyperspectral image in the Augsburg data set is a synthetic spaceborne imagery simulated based on an airborne HySpex data. It has a GSD of 30 m, a size of 332 by 485 pixels, and 180 bands ranging from 400 nm to 2500 nm. Same as the Berlin data set, the PolSAR data is a VH-VV polarized Sentinel-1 image with a GSD of 10 m and a size of 997 by 1456 pixels. The training data and the testing data are shown in Figure 3 which are spatially separated. The details of the training and testing data sets are summarized in Table 3.

Table 3. Summary of the training data and the testing data for the scene of city Augsburg.

Class	# of Training Sample	# of Testing Sample
Forest	200	4100
Residential area	200	4100
Industrial area	200	4100
Low plants	200	4100
Soil	-	-
Allotment	200	4100
Commercial area	200	4100
Water	200	4100

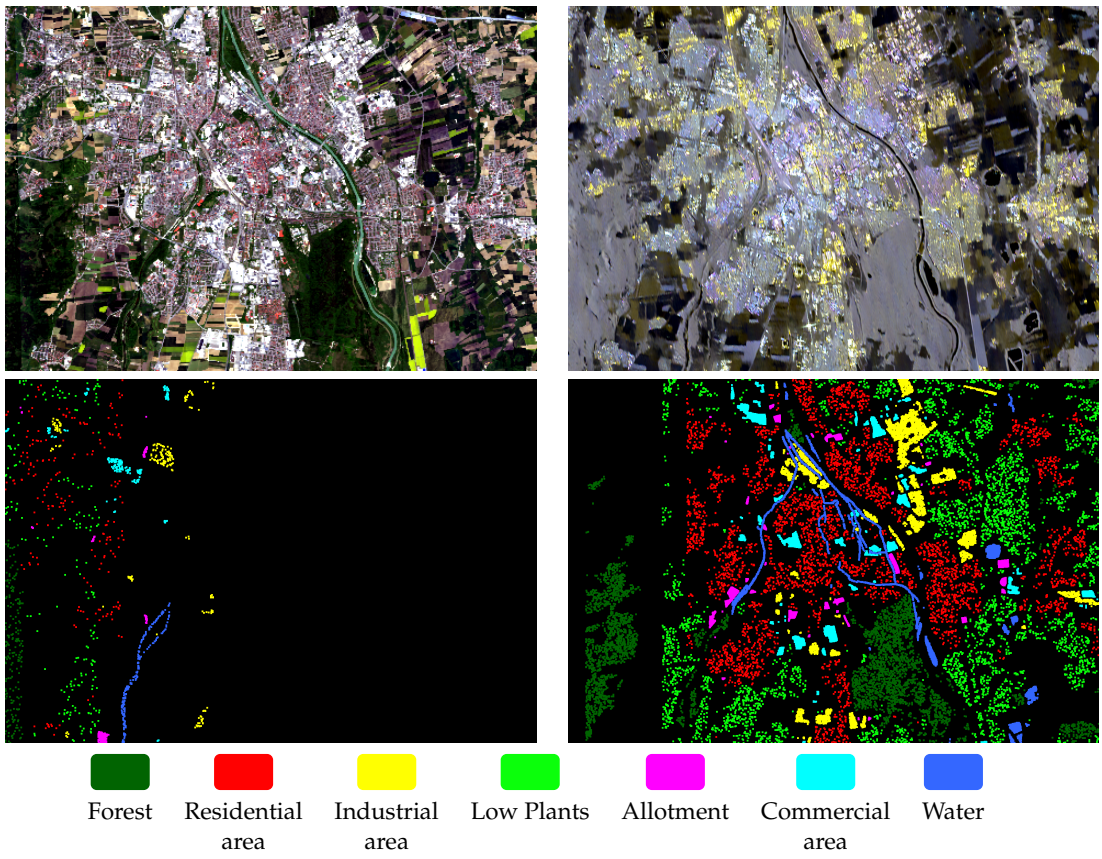


Figure 3. The Augsburg data set. From left to right, top to bottom: RGB components of the hyperspectral image; Sentinel-1 dual-Pol data; The training data; The testing data.

2.7. Experiment Setting

We start with a reasonable feature selection and extraction strategy from the original data, since it is well known that feature selection and extraction promote the classification performance of remote sensing data. The spectral-spatial feature extraction was employed for the hyperspectral image because of its excellent performance on classification tasks [88–91]. Specifically, the first four and six principal components (PCs) which occupy 99% of the variance of the data were extracted from the hyperspectral images of Berlin and Augsburg, respectively. The morphological profiles with radius of one, two, and three were employed to extract the spatial information on each PC. Thus, in total, 28 features and 42 features were extracted from the hyperspectral images of Berlin and Augsburg, respectively. For the feature extraction of Sentinel-1 dual-Pol data, four polarimetric features were extracted. They are the intensity of the VH channel, the intensity of the VV channel, the coherence of VV and VH, and the intensity ratio of VV and VH. Since the morphological profile was proven to promote classification of PolSAR [92,93], it is also used to extract spatial information from the four polarimetric features with radius equal to one, two, and three. In addition, the local statistics including the mean and standard deviation were extracted using a sliding window of 11 by 11 pixel on those four polarimetric features. In total, 36 features were extracted from the dual-Pol SAR data for both data sets of Berlin and Augsburg, respectively.

To carry out a comprehensive comparison of the fusion algorithms, in total 16 algorithms were implemented. Listed in Table 4, they are (1) PolSAR data only (POL), (2) hyperspectral image only (HSI), (3) feature stacking of hyperspectral and PolSAR data (HSI+POL), (4) data alignment using the original locality preserving projections (LPP) [52], (5) supervised version of LPP (LPP_SU), (6) semi-supervised version of LPP (LPP_SE), (7) the generalized graph-based fusion (GGF) [48], (8) supervised version of GGF (GGF_SU), (9) semi-supervised version of GGF (GGF_SE), (10) manifold

alignment (MA) [36,44], (11) unsupervised version of MA (MA_UN), (12) supervised version of MA (MA_SU), (13) MAPPER-Induced manifold alignment with first two principal components as filter functions (MIMA) [53], (14) unsupervised MIMA (MIMA_UN), (15) MIMA with local density as filter function (MIMA-D), and (16) unsupervised MIMA with local density as filter function (MIMA-D_UN).

Table 4. Technical summary of the selected algorithms. 'SU', 'UN', and 'SE' represent the learning strategy of supervised, unsupervised, or semi-supervised, respectively. **W** and **A** represent the weight matrix and the connection matrix, respectively. The hyperparameter set $\{k, dn, \mu, b\}$ indicates the number of neighbors, the number of dimensions, the topology weighting parameter, and the number of bins.

	Algorithm	Data		Learning Strategy			Fusion Concept	Manifold	Hyper-Parameter
		HSI	POL	SU	UN	SE			
1	POL	-	✓	-	-	-	-	-	-
2	HSI	✓	-	-	-	-	-	-	-
3	HSI+POL	✓	✓	-	-	-	Concatenation	-	-
4	LPP	✓	✓	-	✓	-	data alignment	**W**	$\{k, dn\}$
5	LPP_SU	✓	✓	✓	-	-	data alignment	**W**	$\{k, dn\}$
6	LPP_SE	✓	✓	-	-	✓	data alignment	**W**	$\{k, dn\}$
7	GGF	✓	✓	-	✓	-	data alignment	**W**	$\{k, dn\}$
8	GGF_SU	✓	✓	✓	-	-	data alignment	**W**	$\{k, dn\}$
9	GGF_SE	✓	✓	-	-	✓	data alignment	**W**	$\{k, dn\}$
10	MA	✓	✓	-	-	✓	manifold alignment	**A**	$\{\mu, k, dn\}$
11	MA_UN	✓	✓	-	✓	-	Constrained dimension reduction	**A**	$\{k, dn\}$
12	MA_SU	✓	✓	✓	-	-	manifold alignment	**A**	$\{dn\}$
13	MIMA	✓	✓	-	-	✓	manifold alignment	**A**	$\{\mu, b, dn\}$
14	MIMA_UN	✓	✓	-	✓	-	Constrained dimension reduction	**A**	$\{b, dn\}$
15	MIMA-D	✓	✓	-	-	✓	manifold alignment	**A**	$\{\mu, b, dn\}$
16	MIMA-D_UN	✓	✓	-	✓	-	Constrained dimension reduction	**A**	$\{b, dn\}$

These manifold algorithms listed in Table 4 are categorized into the two approaches (data alignment or manifold alignment) mentioned in Section 1.2. LPP and GGF belong to the category of data alignment which concatenates data as a stack, and applies manifold learning on the stacked data. MA and MIMA belong to the category of manifold alignment which independently project K data sources to a latent space where the data are aligned.

The hyperparameters of each algorithm were tuned via a grid search, so that each algorithm reaches its best performance. The k was set in a range of 10 to 120 with an interval of 10. The number of dimension dn is set in a range of 5 to 50 with an interval of 5. The topology weighting parameter μ is set in a range of 0.5 to 3 with an interval of 0.5. The number of bins b is set in a range of 5 to 55 with an interval of 5.

After the data being fused, five different shallow classifiers were applied to the fused data set in the classification step. They are: one nearest neighbor (*1NN*) [58], linear SVM (*LSVM*) [59,60], Gaussian kernel SVM (*KSVM*) [59,60], random forest (*RF*) [61], and canonical correlation forest (*CCF*) [62]. The parameter tuning of *LSVM* is done in a heuristic procedure [60]. LIBSVM [94] is employed for the implementation of the *KSVM*. The number of trees was set as 40 for both *RF* and *CCF*.

3. Experiment Results

The discussion of experiment result mainly focus on the following three aspects:

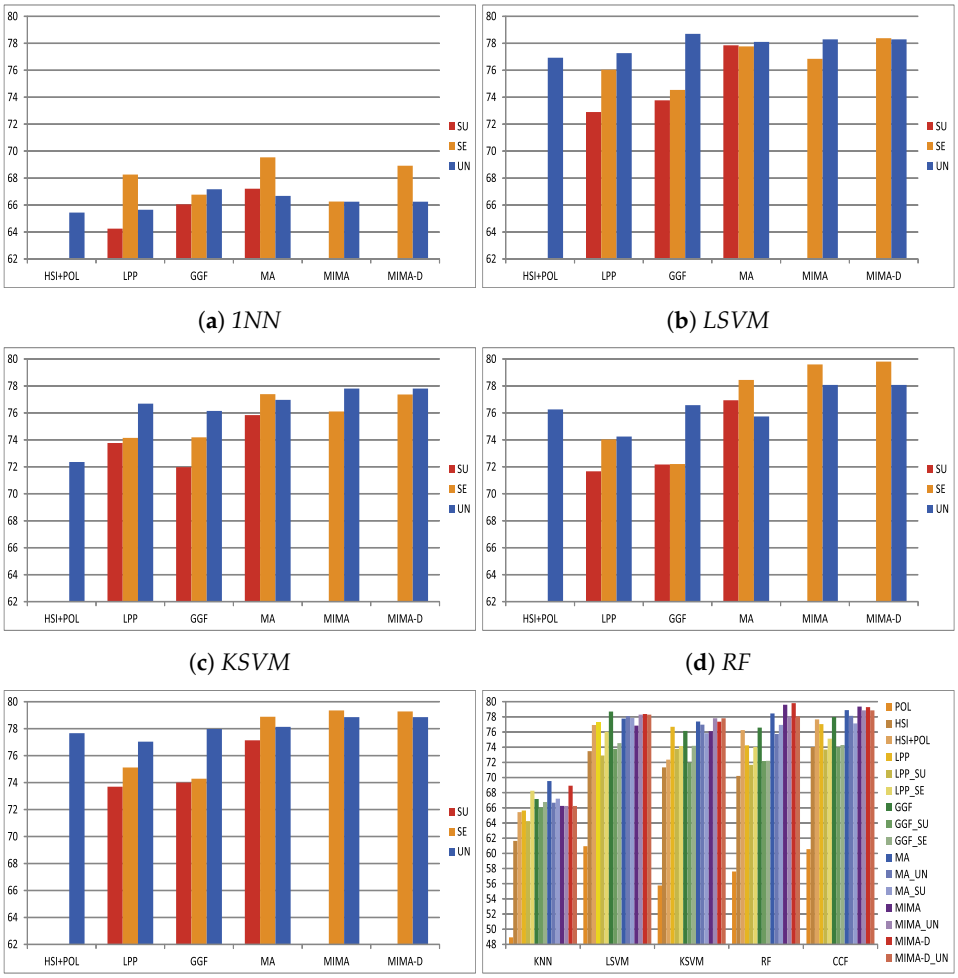
- Manifold learning strategy. The experiment result supports the discussion of the impact that causes by different learning strategies, the unsupervised learning, the supervised learning, and the semi-supervised learning.
- Data fusion approach. The result supports the discussion of the two fusion approaches, the data alignment-based and the manifold alignment-based, for the fusion of the hyperspectral image and PolSAR data.

- Performance on classification. The experiment result reveals how manifold techniques perform on fusing hyperspectral images and PolSAR data and how different these manifold techniques perform.

The classification result is quantitatively evaluated by the class-specific accuracy, the average accuracy, the overall accuracy, and the kappa coefficient. The class-specific accuracy provides the percentage of correct predictions for a specific class. The average accuracy is the mean value of class-specific accuracy. The overall accuracy indicates the percentage of correctness for all predictions. And kappa coefficient also evaluates the overall correctness, yet is more robust than the overall accuracy [95].

3.1. Experiment on the Berlin Data Set

As shown in Figure 4 and Table 5, for the data alignment-based fusion algorithms (LPP and GGF), the unsupervised versions outperform the supervised and the semi-supervised versions. However, for the manifold alignment-based fusion algorithms (MA, MIMA, and MIMA-D), the semi-supervised versions have the best performance comparing to the supervised and the unsupervised ones. Surprisingly, in both type of fusion algorithms, the fully supervised strategy performs the worst.



(**a**) *1NN* (**b**) *LSVM*

(**c**) *KSVM* (**d**) *RF*

(**e**) *CCF* (**f**) All experiments

Figure 4. Comparison of the classification accuracies of different classifiers applied on the Berlin data set. Each chart is resulted from a corresponding classifier. The right bottom chart demonstrates all the overall accuracy resulted by applying five classifiers on every fused data achieved from the selected 16 algorithms. The y-axis report the overall accuracy in percentage (%). The 'SU', 'SE', and 'UN', represent supervised, semi-supervised, and unsupervised, respectively.

Taking the result of the simple concatenation (HSI+POL in Table 5) as reference, the data alignment-based fusion algorithms (LPP and GGF) marginally improve the classification accuracy. Sometimes the performance even drops below the reference accuracy. On the contrary, the manifold alignment-based fusion algorithms (MA, MIMA, and MIMA-D) have a more consistent improvement of the classification accuracy by ca. 3%. In fact, MIMA and MIMA-D have a considerable improvement comparing to LPP, GGF, and MA, especially when *RF* and *CCF* are employed as the classifier. This can be seen in Figure 4. Among all the algorithms, MIMA and MIMA-D have the best overall performance. Shown in Table 5, their best performance reach over 0.66, 65%, and 79%, for the kappa coefficient, the average accuracy, and the overall accuracy, respectively. For a visual comparison, Figure 5 plots the ground truth and the classification maps predicted by the 16 algorithms with *CCF*.

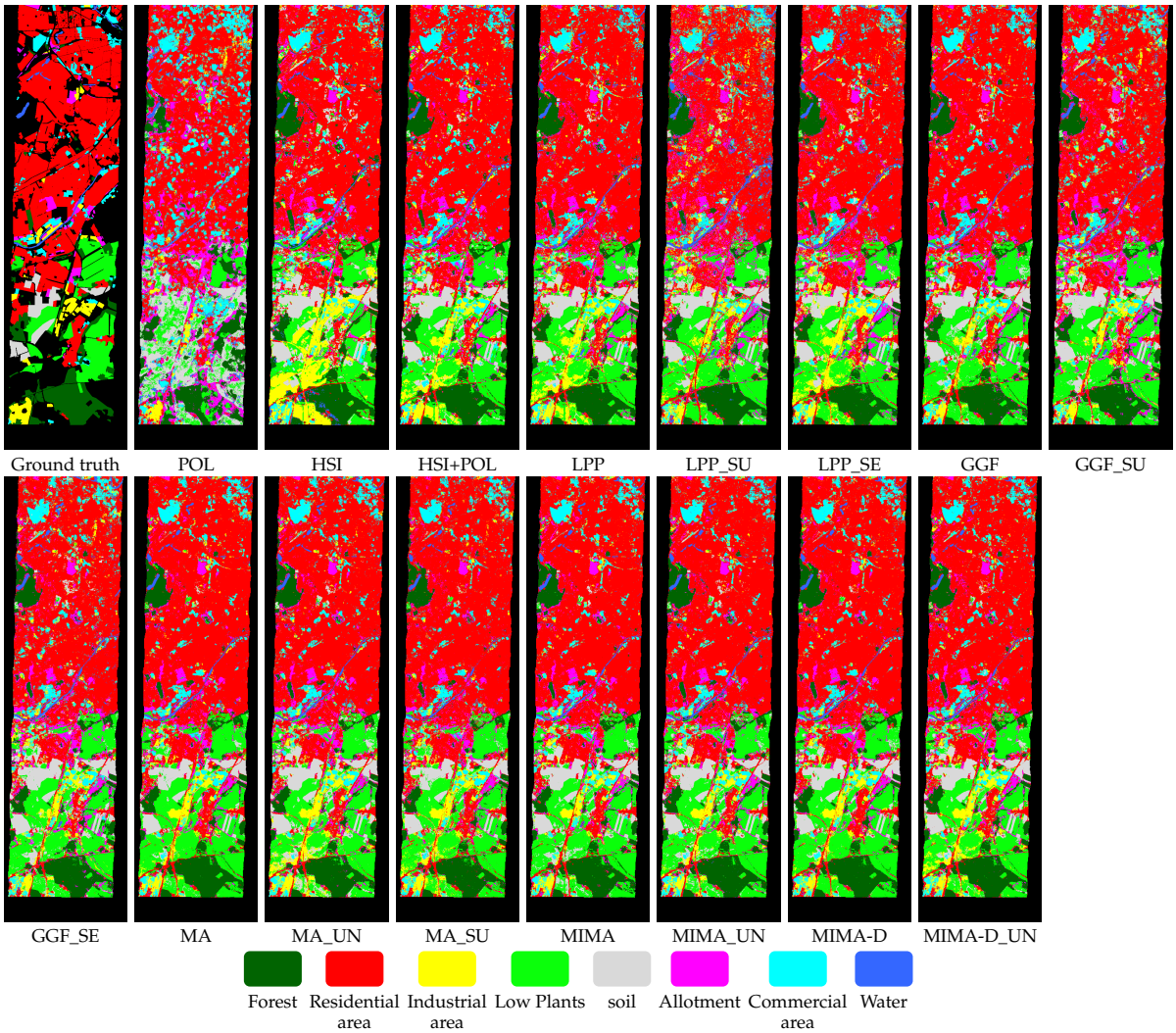


Figure 5. Visualization of the classification maps and the ground truth. The 16 classification maps are provided by applying *CCF* on fused data of the 16 algorithms, for the Berlin data set. Classification maps achieved by manifold alignment fusion methods are more accurate than the maps achieved by data alignment fusion methods.

Table 5. Quantitative performance comparison on the Berlin data, in terms of class-specific accuracy, kappa coefficient, average accuracy, overall accuracy, and mean overall accuracy. The mean overall accuracy is calculated based on the overall accuracies achieved by the five classifiers. The listed indications are achieved after hyperparameter tunning. The hyperparameters of each algorithm are listed under the name of the algorithm and their values are listed in the table. The kappa coefficient, average accuracy, and the overall accuracy that larger than 0.66, 65%, and 79% are marked in bold. And the three highest mean overall accuracies are also marked in bold.

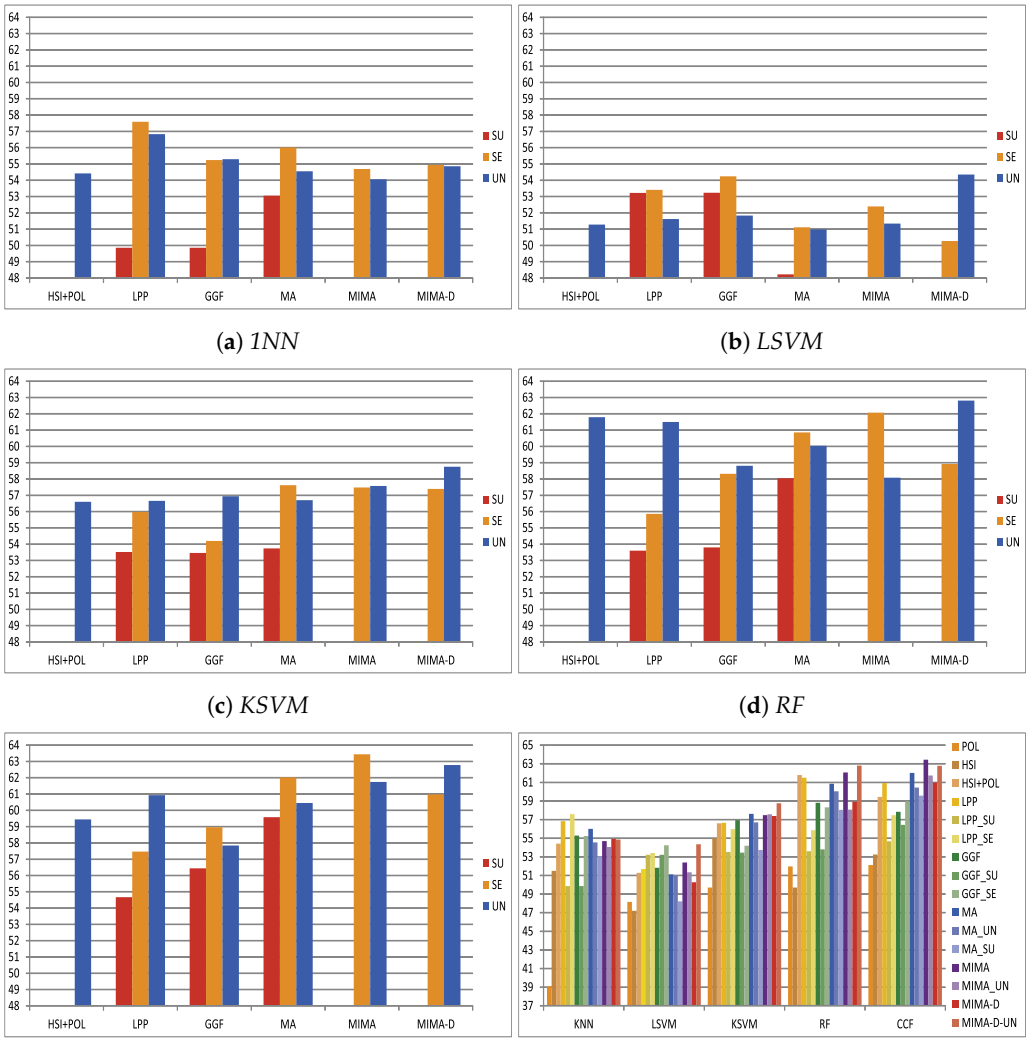
Algorithm	Parameter	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Soil	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
POL	-	*1NN*	40.64	57.67	25.14	32.94	56.88	32.19	30.37	33.85	0.2927	38.71	48.92	56.76
		LSVM	33.02	77.92	13.85	36.46	72.6	40.64	32.23	37.68	0.4012	43.05	60.94	
		KSVM	34.36	69.94	20.38	30.61	68.27	38.62	32.79	42.82	0.3566	42.23	55.76	
		RF	35.61	72.3	25.63	28.66	66.38	43.9	37.87	45.39	0.3789	44.47	57.61	
		CCF	37.96	76.87	24.87	30.69	64.72	38.82	36.88	41.34	0.4035	44.02	60.56	
HSI	-	*1NN*	68.78	63.87	30.01	57.58	90.73	55.76	32.86	73.89	0.4599	59.18	61.64	70.14
		LSVM	69.2	82.5	18.55	65.7	79.06	53.59	44.77	72.81	0.585	60.77	73.48	
		KSVM	72.58	78.68	35.43	63.74	74.18	56.87	31.58	74.29	0.5625	60.92	71.34	
		RF	66.65	79.64	30.25	57.44	75.33	47.77	35.17	78.1	0.5437	58.79	70.21	
		CCF	71	81.86	31.54	68.95	81.36	53.47	38.35	74.81	0.597	62.67	74.03	
HSI+POL	-	*1NN*	64.83	69.7	32.89	65.27	83.81	54.77	34.59	63.51	0.4975	58.67	65.44	73.73
		LSVM	66.57	86.24	30.48	75.3	79.61	53.52	40.12	76.11	0.6329	63.49	76.93	
		KSVM	67.27	80.93	41.78	64.02	72.37	57.58	33	74.6	0.5764	61.44	72.36	
		RF	63.46	84.99	37.79	74.38	82.72	56.26	40.61	82.09	0.6266	**65.29**	76.26	
		CCF	71.51	86.27	34.05	72.03	83.24	56.3	44.33	77.7	0.6445	**65.68**	77.67	
LPP $\{k, dn\}$	{60, 15}	*1NN*	69.53	69.07	34.56	66.09	80.27	57.51	32.18	64.56	0.5009	59.22	65.65	74.18
	{20, 30}	*LSVM*	70.1	87.05	32.52	70.97	79.26	58.88	36.48	72.61	0.6354	63.48	77.27	
	{30, 25}	*KSVM*	71.19	85.77	41.43	70.95	82.36	53.97	30.77	72.68	0.6297	63.64	76.69	
	{10, 20}	*RF*	56.2	85.87	28.9	69.28	76	49.9	38.64	67.07	0.5874	58.98	74.25	
	{10, 15}	*CCF*	68.41	86.68	34.35	71.96	80.07	54.07	37.54	75.93	0.6325	63.63	77.04	
LPP_SU $\{dn\}$	{10}	*1NN*	63.86	67.04	34.79	71.42	79.06	54.39	28.17	72.32	0.4817	58.88	64.25	71.26
	{30}	*LSVM*	64.41	81.51	34.12	70.1	81.56	56.74	29.1	71.38	0.578	61.11	72.9	
	{50}	*KSVM*	67.06	81.6	43.96	72.17	82.34	57.81	25.04	69.69	0.5908	62.46	73.77	
	{25}	*RF*	64.71	80.89	30.98	65.55	72.26	55.27	32.9	69.36	0.5596	58.99	71.67	
	{25}	*CCF*	64.25	81.99	33.72	74.47	75.59	55.89	33.77	69.76	0.5883	61.18	73.7	
LPP_SE $\{k, dn\}$	{80, 10}	*1NN*	68.22	72.17	38.92	73.21	73.43	58.09	30.65	74.02	0.5327	61.09	68.26	73.52
	{120, 40}	*LSVM*	64.68	85.37	38.15	74.36	79.63	59.18	29.75	77.41	0.6194	63.57	76.04	
	{120, 40}	*KSVM*	69.02	81.93	41.67	70.74	77	59.76	30.77	76.17	0.6001	63.38	74.15	
	{120, 30}	*RF*	66.96	83.15	29.66	72.12	66.45	56.39	34.17	74	0.5919	60.36	74.03	
	{120, 25}	*CCF*	64.86	85.09	34.63	71.85	66.83	56.05	34.33	75.05	0.6044	61.09	75.12	
GGF $\{k, dn\}$	{20, 30}	*1NN*	69.28	71.37	36.65	66.54	83.51	56.94	31.34	63.82	0.5186	59.93	67.17	75.31
	{90, 30}	*LSVM*	68.11	88.76	34.14	76.11	79.29	54.93	36.54	75.14	0.655	64.13	78.7	
	{20, 30}	*KSVM*	72.18	84.64	37.08	70.29	81.88	57.25	34.49	74.44	0.6254	64.03	76.15	
	{10, 20}	*RF*	68.97	86.55	29.13	70.39	81.23	49.45	41.85	62.88	0.6242	61.31	76.58	
	{10, 25}	*CCF*	70.53	87.51	31.29	76.34	70.86	51.95	42.06	67.95	0.6448	62.31	77.98	

Table 5. *Cont.*

Algorithm	Parameter	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Soil	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
GGF_SU $\{dn\}$	{10}	*1NN*	65.57	69.99	37.73	68.89	80.13	51.96	28.71	76.62	0.5013	59.95	66.05	71.59
	{50}	*LSVM*	63.6	82.87	36.49	69.8	82.34	56.58	29.62	76.22	0.5906	62.19	73.77	
	{50}	*KSVM*	69.99	80.63	46.43	60.43	77.21	53.92	25.15	78.77	0.5695	61.57	71.98	
	{50}	*RF*	62.01	81.42	32.09	67.3	74.08	53.3	38.83	65.17	0.5678	59.28	72.17	
	{40}	*CCF*	65.54	83.4	31.38	70.58	72.26	51.15	37.02	68.24	0.5906	59.95	74	
GGF_SE $\{k, dn\}$	{10, 15}	*1NN*	66.96	70.63	36.07	69.65	80.62	55.65	29.49	76.35	0.5119	60.68	66.77	72.40
	{120, 45}	*LSVM*	63.06	83.52	37.69	73.01	81.94	55.48	29.11	79.87	0.6007	62.96	74.54	
	{40, 40}	*KSVM*	70.19	82.26	41.52	67.92	80.35	54.38	31.22	82.51	0.5988	63.79	74.19	
	{20, 40}	*RF*	65.27	80.56	34.49	67.01	75.85	54.57	38.72	66.98	0.5716	60.43	72.21	
	{70, 30}	*CCF*	60.15	83.94	35.07	74.18	74.3	51.22	35.51	68.64	0.5942	60.37	74.29	
MA $\{\mu, k, dn\}$	{2, 90, 10}	*1NN*	69.83	73.8	38	75.68	69.64	60.09	29.41	72.27	0.5474	61.09	69.54	**76.40**
	{2.5, 20, 25}	*LSVM*	65.49	86.97	37.63	79.08	80.06	55.63	34.46	73.37	0.6445	64.09	77.77	
	{2.5, 90, 35}	*KSVM*	69.38	85.81	37.49	78.3	80.54	55.42	33.29	73.21	0.6405	64.18	77.39	
	{2, 10, 50}	*RF*	64.5	90.08	30.25	77.68	65.58	49.41	36.85	67.95	0.644	60.29	78.45	
	{2, 10, 20}	*CCF*	66.66	89.12	33.05	79.51	68.95	54.91	39.47	71.01	0.6557	62.84	78.89	
MA_UN $\{k, dn\}$	{120, 15}	*1NN*	68.46	69.61	32.87	72.87	78.51	54.88	34.76	67.95	0.5159	59.99	66.68	75.13
	{90, 30}	*LSVM*	66.86	87.58	35.97	77.55	78.59	55.44	36.3	76.15	0.649	64.3	78.1	
	{40, 50}	*KSVM*	70.55	85.61	36.23	74.18	79.83	57.57	35.55	73.14	0.6346	64.08	76.97	
	{100, 30}	*RF*	58.91	87.37	26.35	69.77	80.7	53.14	41.94	60.71	0.6079	59.86	75.74	
	{30, 30}	*CCF*	67	88.05	33.05	74.11	81.85	55	41.91	70.52	0.6467	63.94	78.14	
MA_SU $\{dn\}$	{5}	*1NN*	69.88	71.34	34.87	68.69	71.01	57.88	32.38	73.52	0.5199	59.94	67.21	75
	{50}	*LSVM*	67.56	86.73	38.76	79.67	77.21	56.87	32.27	75.45	0.6457	64.31	77.85	
	{50}	*KSVM*	71.6	83.96	35.72	75.92	61.57	59.59	37.1	72.65	0.6204	62.26	75.84	
	{50}	*RF*	60.53	87.82	33.22	77.13	70.16	52.42	38.82	63.66	0.6242	60.47	76.94	
	{50}	*CCF*	64.09	88.37	30.57	76.73	62.56	51.86	36.99	59.9	0.6257	58.89	77.14	
MIMA $\{\mu, b, dn\}$	{1, 15, 5}	*1NN*	69.91	70.2	33.39	69.63	61.94	53.49	35.07	68.62	0.5055	57.78	66.26	**76.22**
	{1, 15, 15}	*LSVM*	67.76	84.97	36.22	78.36	79.08	57.74	38	70.25	0.6328	64.05	76.85	
	{1, 15, 15}	*KSVM*	71.06	84.24	41.01	76.11	69.87	55.82	32.97	68.97	0.6233	62.51	76.11	
	{1.5, 25, 40}	*RF*	65.1	90.31	32.54	80	82.77	50.79	35.08	71.01	**0.6642**	63.45	**79.6**	
	{2, 25, 20}	*CCF*	70.86	88.06	36.54	80.42	76.88	57.21	39.61	73.21	**0.667**	**65.35**	**79.36**	
MIMA_UN $\{b, dn\}$	{10, 20}	*1NN*	72.57	68.39	35.96	70.18	79.27	62.58	30.73	67.41	0.513	60.89	66.25	75.85
	{10, 35}	*LSVM*	68.21	88.59	36.62	74.6	80.79	55.87	29.86	76.08	0.6495	63.83	78.29	
	{10, 35}	*KSVM*	71.78	87.1	36.85	73.13	82.31	58.05	31.79	73.14	0.6449	64.27	77.81	
	{55, 30}	*RF*	67.92	88.44	27.36	77.22	81.32	50.9	35	61.04	0.6417	61.15	78.08	
	{30, 20}	*CCF*	71.06	88.19	29.72	77.55	79.81	55.71	39.99	69.67	0.658	63.96	78.86	
MIMA-D $\{\mu, b, dn\}$	{1.5, 30 ,15}	*1NN*	71.31	72.3	35.31	74.51	76.66	57.37	33.48	71.84	0.5423	61.6	68.92	**76.75**
	{1.5, 45, 20}	*LSVM*	67.59	86.85	36.8	81.07	78.3	56.4	38.97	75.88	0.6549	**65.23**	78.38	
	{2.5, 55, 30}	*KSVM*	70.01	85.33	36.79	78.84	78.52	56.83	36.44	76.08	0.6425	64.86	77.37	
	{1, 30, 30}	*RF*	67.02	89.85	33.09	80.46	83.21	50.61	37.95	74.27	**0.6698**	64.56	**79.81**	
	{1, 45, 30}	*CCF*	68.91	89.18	34.79	78.63	75.48	51.74	39.85	69.45	**0.6628**	63.5	**79.28**	
MIMA-D_UN $\{b, dn\}$	{55, 15}	*1NN*	72.57	68.39	35.96	70.18	79.27	62.58	30.73	67.41	0.513	60.89	66.25	75.52
	{55, 25}	*LSVM*	68.21	88.59	36.62	74.6	80.79	55.87	29.86	76.08	0.6495	63.83	78.29	
	{40, 20}	*KSVM*	71.78	87.1	36.85	73.13	82.31	58.05	31.79	73.14	0.6449	64.27	77.81	
	{45, 30}	*RF*	67.92	88.44	27.36	77.22	81.32	50.9	35	61.04	0.6417	61.15	78.08	
	{45, 25}	*CCF*	71.06	88.19	29.72	77.55	79.81	55.71	39.99	69.67	0.658	63.96	78.86	

3.2. Experiment on Augsburg Data Set

The findings of the Augsburg data set are consistent with that of the Berlin data set. For the data alignment-based fusion algorithms, the unsupervised learning strategy works the best among the three learning strategies. For the manifold alignment-based fusion algorithms, the semi-supervised learning strategy performs the best. Comparing the results to that of simple concatenation (HSI+POL), the data alignment-based fusion (LPP and GGF) barely has any improvement, while the manifold alignment fusion has a 2% improvement comparing to the LPP and GGF. These findings can be seen in Figure 6, and Table 6. Among all the algorithms, combining MIMA or MIMA-D with *RF* or *CCF* provide the best classification performance. Their kappa coefficient, the average accuracy, and the overall accuracy, reach 0.56, 62.5%, and 62.5% respectively. A visual comparison of the results is shown in Figure 7. Similar to Figure 5, the classification maps were predicted by *CCF*.



(**a**) *1NN* (**b**) *LSVM*

(**c**) *KSVM* (**d**) *RF*

(**e**) *CCF* (**f**) *All experiments*

Figure 6. Comparison of the classification accuracies of different classifiers applying on the Augsburg data set. Each chart is resulted from a corresponding classifier. The right bottom chart demonstrates all the overall accuracy resulted by applying five classifiers on the fused data achieved from each of the 16 algorithms. The y-axis is the overall accuracy in percentage (%). The 'SU', 'SE', and 'UN', represent supervision, semi-supervision, and un-supervision, respectively.

Table 6. Quantitative performance comparison on the Augsburg data, in terms of class-specific accuracy, kappa coefficient, average accuracy, overall accuracy, and mean overall accuracy. The mean overall accuracy is calculated based on the overall accuracies achieved by the five classifiers. The listed indications are achieved after hyperparameter tunning. The hyperparameters of each algorithm are listed under the name of the algorithm and their values are listed in the table. The kappa coefficient, average accuracy, and the overall accuracy that larger than 0.56, 62.5%, and 62.5% are marked in bold. And the three highest mean overall accuracies are also marked in bold.

Algorithm	Parameter	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
POL	-	*1NN*	64	35.88	38.8	55.02	22.54	38.9	18.66	0.2897	39.11	39.11	48.21
		LSVM	86.93	46.44	39.15	73.17	25.37	44.29	21.8	0.3952	48.16	48.16	
		KSVM	86.51	64.49	31.41	81.98	22.39	41.98	19.12	0.4131	49.7	49.7	
		RF	81.88	63.44	47.76	88.46	28.88	38.71	14.63	0.4396	51.97	51.97	
		CCF	82.29	61.85	47.8	88.37	30.34	38.07	16.1	0.4414	52.12	52.12	
HSI	-	*1NN*	27.9	52.49	61.1	78.2	60.66	24.9	55.24	0.4341	51.5	51.5	51.33
		LSVM	25.44	50.22	75.93	67.46	38.32	15.15	57.93	0.3841	47.21	47.21	
		KSVM	31.2	65.2	70.71	86.37	55.98	20.8	54.63	0.4748	54.98	54.98	
		RF	25.59	58.29	70.29	84.34	40.41	15.98	52.98	0.4131	49.7	49.7	
		CCF	27.29	64.56	75.71	84.68	48.29	16.54	55.66	0.4546	53.25	53.25	
HSI+POL	-	*1NN*	34.76	58.17	55.93	84.56	57.73	34.9	54.88	0.4682	54.42	54.42	56.71
		LSVM	31	65.95	73.29	83.85	36.9	25.07	42.85	0.4315	51.28	51.28	
		KSVM	40.59	67.83	67.07	92.59	45.24	27.1	55.78	0.4937	56.6	56.6	
		RF	61.27	73.88	70.1	94.98	47.51	25.63	59.17	0.5542	61.79	61.79	
		CCF	46.07	75.63	78.05	95.51	58.07	18.49	44.22	0.5267	59.44	59.44	
LPP $\{k, dn\}$	{10, 40}	*1NN*	44.9	60.61	53.29	86.56	61.37	34.76	56.32	0.4963	56.83	56.83	57.42
	{20, 20}	*LSVM*	28.17	64.93	76.63	81.54	38.27	17.88	53.93	0.4356	51.62	51.62	
	{40, 50}	*KSVM*	40.98	67.98	73.49	92.32	45.49	22.68	53.66	0.4943	56.66	56.66	
	{10, 30}	*RF*	73.66	66.15	65.8	89.54	51.24	25.78	55.17	0.5456	61.05	61.05	
	{10, 35}	*CCF*	59.63	70.71	72.8	92.2	51.9	22.78	56.51	0.5442	60.93	60.93	
LPP_SU $\{dn\}$	{5}	*1NN*	31.93	55.83	56.95	78.51	49.07	33.98	42.76	0.415	49.86	49.86	52.97
	{10}	*LSVM*	40.85	63.1	63.29	87.46	49.17	32.61	36.05	0.4542	53.22	53.22	
	{40}	*KSVM*	54.24	63.93	66.32	87.2	45.05	28.49	29.41	0.4577	53.52	53.52	
	{35}	*RF*	44.46	60.93	62.78	90.07	44.15	30.95	41.88	0.4587	53.6	53.6	
	{35}	*CCF*	52.07	62.15	64.66	90.17	44.24	28.88	40.51	0.4711	54.67	54.67	
LPP_SE $\{k, dn\}$	{20, 45}	*1NN*	49.76	59.15	53	85.05	60.98	40.05	55.15	0.5052	57.59	57.59	56.06
	{10, 10}	*LSVM*	43.49	65.51	77.22	85.07	40.76	20.8	41.05	0.4565	53.41	53.41	
	{120, 35}	*KSVM*	37.66	71.27	75.22	93.22	48.44	20.54	45.49	0.4864	55.98	55.98	
	{30, 15}	*RF*	27.17	63.22	72.2	91.78	54.46	26.54	55.66	0.485	55.86	55.86	
	{80, 40}	*CCF*	47.2	66.46	73.22	90.93	56.07	23.27	45.17	0.5039	57.47	57.47	
GGF $\{k, dn\}$	{20, 50}	*1NN*	41.37	57.22	49.68	82.63	61.61	38.2	56.32	0.4784	55.29	55.29	55.81
	{30, 15}	*LSVM*	29.17	63.76	74.83	82.12	36.54	19.71	56.71	0.438	51.83	51.83	
	{20, 15}	*KSVM*	34.51	69.22	73.71	92.34	45.32	23.9	59.61	0.4977	56.94	56.94	
	{40, 45}	*RF*	60.22	65.61	61.29	89.73	46.46	31.78	56.56	0.5194	58.81	58.81	
	{40, 35}	*CCF*	47.9	70.9	72.22	92.44	43.34	23.05	55	0.5081	57.84	57.84	

Table 6. *Cont.*

Algorithm	Parameter	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
GGF_SU $\{dn\}$	{5}	*1NN*	31.93	55.83	56.95	78.51	49.07	33.98	42.76	0.415	49.86	49.86	53.36
	{10}	*LSVM*	40.85	63.15	63.29	87.46	49.2	32.61	36.05	0.4543	53.23	53.23	
	{35}	*KSVM*	51.17	64.07	65.46	86.78	44.37	30.9	31.49	0.4571	53.46	53.46	
	{45}	*RF*	44.93	61.05	60.15	89.93	42.88	32.68	45.02	0.4611	53.8	53.8	
	{45}	*CCF*	51.2	62.61	66.93	90.24	46.78	28.76	48.59	0.4918	56.44	56.44	
GGF_SE $\{k, dn\}$	{120, 10}	*1NN*	44.8	58.17	63.32	84.54	56.05	32.95	46.83	0.4778	55.24	55.24	56.19
	{20, 30}	*LSVM*	53.02	66.54	66.95	84.61	47.27	29.41	31.88	0.4661	54.24	54.24	
	{10, 50}	*KSVM*	67.54	68.24	66.8	87.12	41.32	23.98	24.41	0.4657	54.2	54.2	
	{90, 15}	*RF*	42.88	64.9	68.07	92.56	56.68	26.63	56.54	0.5138	58.32	58.32	
	{120, 40}	*CCF*	47	65.83	67.88	92.51	57.29	27.07	55.15	0.5212	58.96	58.96	
MA $\{\mu, k, dn\}$	{2, 70, 35}	*1NN*	30.88	58.68	61.39	82.05	77.27	27.78	54.02	0.4868	56.01	56.01	**57.52**
	{2.5, 60, 35}	*LSVM*	26.22	66.63	78.2	72.44	42.9	16.1	55.27	0.4296	51.11	51.11	
	{2, 70, 25}	*KSVM*	31.44	69.54	78.8	93	59.05	17.73	53.76	0.5055	57.62	57.62	
	{1, 110, 45}	*RF*	75.34	72.15	64.66	91.61	48.88	30.12	43.24	0.5433	60.86	60.86	
	{1, 110, 45}	*CCF*	65.85	73.24	72.61	93.61	55	23.8	50.05	0.557	62.02	62.02	
MA_UN $\{k, dn\}$	{100, 25}	*1NN*	31.61	56.85	57.29	80.71	73.98	26.61	54.83	0.4698	54.55	54.55	56.54
	{100, 30}	*LSVM*	26.51	67.12	76.8	73.78	41.07	15.71	55.83	0.428	50.98	50.98	
	{100, 20}	*KSVM*	32.56	68.2	74	89.29	58.63	18.88	55.32	0.4948	56.7	56.7	
	{20, 25}	*RF*	75.15	67.93	63.17	87.93	44.29	31	50.83	0.5338	60.04	60.04	
	{20, 40}	*CCF*	75.27	69.07	60.95	89.83	50.07	32.41	45.56	0.5386	60.45	60.45	
MA_SU $\{dn\}$	{50}	*1NN*	26.71	52.78	61.15	80.22	69.93	26.07	54.46	0.4522	53.05	53.05	54.53
	{50}	*LSVM*	25.2	57.2	77.56	70.29	36.85	16.68	53.76	0.3959	48.22	48.22	
	{50}	*KSVM*	28.68	60.68	74.83	87.9	56.2	17.46	50.39	0.4602	53.74	53.74	
	{50}	*RF*	49.76	67.1	67.12	91.9	47.27	28.85	54.32	0.5105	58.05	58.05	
	{45}	*CCF*	64.07	69.12	66.78	92.41	52.39	27.63	44.63	0.5284	59.58	59.58	
MIMA $\{\mu, b, dn\}$	{2.5, 35, 35}	*1NN*	27.68	57.07	62.56	81.39	72.17	26.46	55.51	0.4714	54.69	54.69	**58.01**
	{3, 25, 5}	*LSVM*	23.61	71.93	78.63	79.98	44.29	13.76	54.51	0.4445	52.39	52.39	
	{1.5, 35, 15}	*KSVM*	34.15	68.12	72.9	92.27	53.51	22.07	59.34	0.5039	57.48	57.48	
	{0.5, 40, 35}	*RF*	66.22	76.88	65.51	92.8	47.78	26.27	59.02	0.5575	62.07	62.07	
	{0.5, 55, 40}	*CCF*	76.78	77.49	65.12	92.73	50	28.78	53.15	**0.5734**	**63.44**	**63.44**	
MIMA_UN $\{b, dn\}$	{5, 35}	*1NN*	34.34	55.24	54.85	80.76	71.41	28.44	53.39	0.4641	54.06	54.06	56.56
	{10, 50}	*LSVM*	28.39	67.78	76.05	74.17	40.73	19.22	53.02	0.4323	51.34	51.34	
	{5, 30}	*KSVM*	31.9	68.12	74.78	90.78	59.44	20.1	57.9	0.505	57.57	57.57	
	{20, 30}	*RF*	58.95	66.54	71.76	89.68	51.68	25.24	42.68	0.5109	58.08	58.08	
	{15, 45}	*CCF*	83.71	67.61	68.07	89.73	57.98	27.68	37.37	0.5536	61.74	61.74	
MIMA-D $\{\mu, b, dn\}$	{3, 50, 35}	*1NN*	28.76	57.63	62.68	80.22	74.83	24.98	55.49	0.4743	54.94	54.94	56.5
	{3, 40, 15}	*LSVM*	25.27	67.44	78.46	73.29	39.85	15.41	52.17	0.4198	50.27	50.27	
	{3, 40, 15}	*KSVM*	33.12	68.95	70.41	92.8	55.8	20.05	60.59	0.5029	57.39	57.39	
	{3, 35, 30}	*RF*	52.54	72.27	73.56	92.24	49.83	24.29	47.71	0.5207	58.92	58.92	
	{2.5, 30, 40}	*CCF*	55.51	73.66	72.98	92.85	53.61	24.05	54.22	0.5448	60.98	60.98	
MIMA-D_UN $\{b, dn\}$	{20, 30}	*1NN*	34.93	56.02	56.54	80.46	75.29	26.39	54.41	0.4734	54.86	54.86	**60.29**
	{55, 5}	*LSVM*	87.22	55.24	48	57.41	36.2	41.95	54.41	0.4674	54.35	54.35	
	{15, 30}	*KSVM*	35.27	67.93	77.54	91.83	66.56	17.37	54.73	0.5187	58.75	58.75	
	{20, 50}	*RF*	82.95	65.17	58.12	88.41	54.32	34.05	56.63	**0.5661**	**62.81**	**62.81**	
	{20, 30}	*CCF*	78.54	72.29	65.85	92.63	49.88	26.68	53.56	**0.5657**	**62.78**	**62.78**	

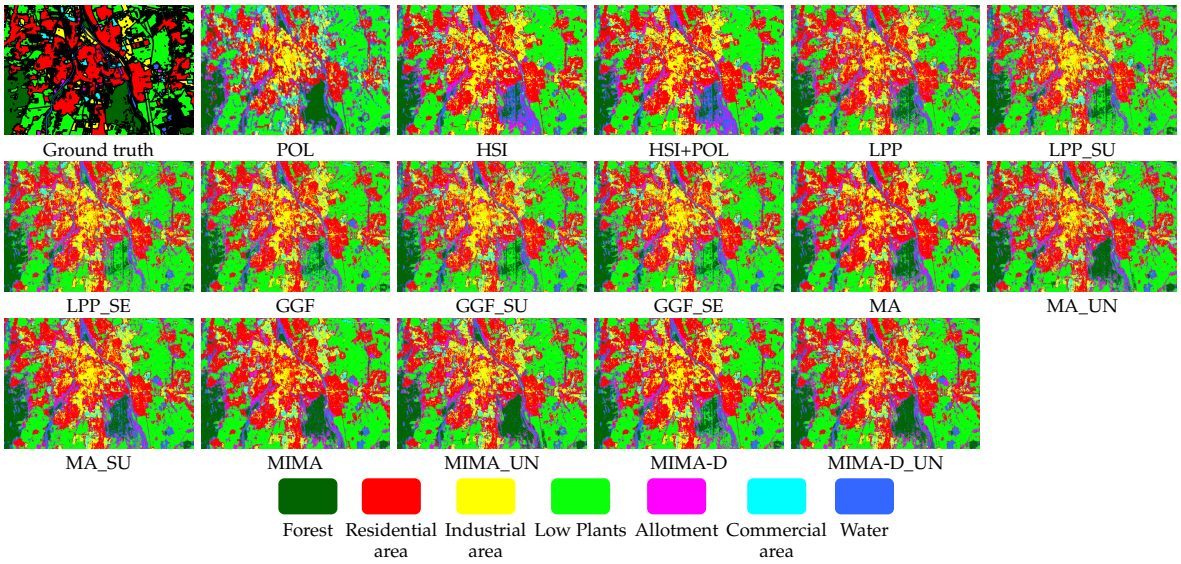


Figure 7. Visualization of the achieved classification maps and the ground truth. The 16 classification maps are provided by applying *CCF* on fused data of the 16 algorithms, for the Augsburg data set. Classification maps achieved by manifold alignment fusion methods are more accurate than the maps achieved by data alignment fusion methods, especially MIMA based methods.

4. Discussion

4.1. The Setting of the Training and Testing Samples

As shown in Figures 2 and 3, the training and testing samples are spatially separated as a standard machine learning practice. However, the distribution of training samples of the Berlin and the Augsburg data set are slightly different. For each class of the Berlin data set, the training data are block-wisely scattered over the whole area. For the Augsburg data set, the training data only covers on the western part of the area. There is no sample from the eastern half of the site where the testing data distribute. Both scenarios are common in remote sensing applications. The latter one is naturally more challenging. This is why the overall accuracy of the Augsburg data set fluctuates around 56%, while it is around 76% for the Berlin data set.

4.2. The Data Alignment Fusion

An unsupervised data alignment-based fusion in this article pursues an intrinsic manifold of a concatenation of the hyperspectral and PolSAR data. Intuitively, making use of the additional label information in the manifold learning (semi-supervision) should be improve the classification accuracy. However, we observed the exact opposite in our experiments. We believe it is due to the misalignment of image pixels of optical and SAR images caused by their distinct imaging geometry. This pixel misalignment leads to extra difficulty in learning a joint manifold. Adding one more manifold defined by the misaligned label will only lead to destructive effects. Therefore, the data alignment-based fusion algorithm is not competent for fusing hyperspectral and PolSAR data with the resolution similar to our dataset. This finding should also be able to generalized to high resolution optical and SAR data, although we have not conducted any experiment.

4.3. The Manifold Alignment Fusion

Different to the data alignment-based fusion, the semi-supervised manifold alignment-based fusion outperforms the unsupervised manifold alignment fusion. This fusion concept is able to introduce the advantage of label information while pursuing the intrinsic manifold. The reason is that

this fusion concept models the manifold of individual data sources independently which suits the fact that hyperspectral and PolSAR data are severely dissimilar in geometry and content. The label information is merged into the two manifolds in a way that the two manifolds are separately link to the label and are then aligned to each other by the label. In such manner, the advantage of label data appears on the classification results. Comparing to the data alignment-based fusion, the manifold alignment-based fusion introduces considerable improvements to the classification accuracy, which shows its competence for the fusion of hyperspectral and PolSAR data.

4.4. The Filter Function of MIMA

As introduced in Section 2.5, the filter function of MIMA introduces an expertise knowledge while deriving the manifold structure of data. MIMA and MIMA-D in this paper employed PCA and a density estimation as the filter function, respectively. The principal components are frequently used in classification. It is proven to be effective [82,96]. The density function is an important property for classification or clustering tasks [97,98]. However, from the experiment in this paper, it is inconclusive which choice is more suitable to serve as the filter function.

5. Conclusion and Outlook

This paper compares 16 variants of four state-of-the-art multi-sensory data fusion algorithms based on manifold learning. The comparison was done via a rigorous evaluation of the performance of the 16 algorithms on land use land cover classification on two sets of spaceborne hyperspectral images and PolSAR data. To carry out an objective comparison, the hyperparameters of the 16 algorithms were optimized via a grid search. Five different shallow classifiers were applied on the data sets fused by the 16 algorithms. We avoided any deep network classifiers, because the goal of this article is to solely evaluate the performance of multi-sensory data fusion algorithms. The experiments conclude that (1) data alignment-based (data concatenation) manifold techniques are less competent for the fusion of hyperspectral images and PolSAR data, or in general optical and SAR images fusion, because a concatenation of the two data sets with distinct imaging geometries causes difficulty even destructive effects when optimizing the target manifold; On the contrary, manifold alignment-based techniques are more competent for the task of optical and SAR images fusion, because the manifolds of the two data are separatly modeled and aligned; (2) Among the manifold alignment-based manifold techniques, semi-supervised methods are able to effectively make use of both the structure of data and the existing label information; (3) the MIMA algorithm cooperating with the *CCF* classifier provides the best classification accuracy among all the algorithms.

Based on our current research, our future research directions can include:

- In the current algorithms, the learned manifold is specific to the very input data sets. We would like to study the generalization of such manifold on data sets of the same sensors. Eventually, we aim at big data processing where one common manifold can be applied to all the data sets of the same type.
- Graph CNN has been an emerging filed in deep lerning. It is also of great interest to combine it with the traditional manifold learning techniques described in this article.
- Because of the data availability of spaceborne hyperspectral and PolSAR data, they have not been extensively applied to real world problems. We would like to address more real world applications especially those for social good using those two types of data, for example, contributing to the monitoring of Unite Nation's sustainable development goals.

Author Contributions: Conceptualization, J.H., D.H., and X.X.Z..; methodology, J.H., and D.H.; software, J.H.; validation, Y.W..; formal analysis, J.H.; investigation, J.H.; resources, J.H.; data curation, J.H.; writing—original draft preparation, J.H.; writing—review and editing, D.H. and Y.W.; visualization, J.H.; supervision, X.X.Z.; project administration, J.H.; funding acquisition, X.X.Z.

Funding: This research was funded by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program with the grant number ERC-2016-StG-714087 (Acronym: So2Sat,

project website: www.so2sat.eu), and the Helmholtz Association under the framework of the Young Investigators Group Signal Processing in Earth Observation (SiPEO) with the grant number VH-NG-1018 (project web-site: www.sipeo.bgu.tum.de).

Acknowledgments: The author would like to thank Claas Grohnfeldt for providing the Augsburg data, Wenzhi Liao for releasing the GGF code, Devis Tuia for releasing the MA code.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Pseudo-Code of LPP

Algorithm A1: LPP($\mathbf{X}$,k,σ)

Input:

$\mathbf{X}$: the data source $\mathbf{X} = [\mathbf{x}^1, ..., \mathbf{x}^p, ..., \mathbf{x}^n] \in \mathbb{R}^{m \times n}$ with n instances and m dimensions

k: the number of local neighbors

σ: the filtering parameter

Output:

$\mathbf{Y}$: the representation of data $\mathbf{X}$ on the intrinsic manifold $\mathcal{M}$.

$\mathbf{f}$: the projection maps data $\mathbf{X}$ to $\mathbf{Y}$

1 construct the n by n weight matrix with Equation (1)

2 construct the degree matrix $\mathbf{D}$

3 construct the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$

4 solve the generalized eigenvalue decomposition $\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{f} = \lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{f}$

5 construct $\mathbf{Y}$: $\mathbf{Y} = \mathbf{f}^T\mathbf{X}$

6 **Return** $\mathbf{Y}$ and $\mathbf{f}$

Appendix B. Pseudo-Code of GGF

Algorithm A2: GGF($\mathbf{X}_1$,$\mathbf{X}_2$,k,σ)

Input:

$\mathbf{X}_1$: the data source $\mathbf{X}_1 \in \mathbb{R}^{m_1 \times n}$ with n instances and m_1 dimensions

$\mathbf{X}_2$: the data source $\mathbf{X}_2 \in \mathbb{R}^{m_2 \times n}$ with n instances and m_2 dimensions

k: the number of local neighbors

σ: the filtering parameter

Output:

$\tilde{\mathbf{Y}}$: the fused data.

$\mathbf{f}$: the projection maps data $\tilde{\mathbf{X}}$ to $\tilde{\mathbf{Y}}$

1 stacking data sources on the feature dimension: $\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = [\tilde{\mathbf{x}}^1, ..., \tilde{\mathbf{x}}^p, ..., \tilde{\mathbf{x}}^n] \in \mathbb{R}^{(m_1+m_2) \times n}$

2 construct binary matrices $\mathbf{A}_i (i \in \{1, 2\})$ to model manifolds of $\mathbf{X}_i$:

$$\mathbf{A}_i(p,q) = \begin{cases} 1 & x_i^p \text{ is one of the } k \text{ nearest neighbor to } x_i^q \\ 0 & \text{otherwise} \end{cases}$$

3 construct a fused binary matrix $\tilde{\mathbf{A}}(p,q) = \mathbf{A}_1(p,q) * \mathbf{A}_2(p,q)$

4 calculate a n by n pairwise distance matrix $\tilde{\mathcal{D}}$

5 construct a GGF pairwise distance matrix $\tilde{\mathcal{D}}_{GGF}$ as Equation (6)

6 calculate the n by n weight matrix: $\tilde{\mathbf{W}}$ as Equation (7)

7 calculate the degree matrix $\tilde{\mathbf{D}}$

8 calculate the Laplacian matrix $\tilde{\mathbf{L}} = \tilde{\mathbf{D}} - \tilde{\mathbf{W}}$

9 solve the generalized eigenvalue decomposition $\tilde{\mathbf{X}}\tilde{\mathbf{L}}\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}\tilde{\mathbf{D}}\tilde{\mathbf{X}}^T\mathbf{f}$

10 calculate $\tilde{\mathbf{Y}} = \mathbf{f}^T\tilde{\mathbf{X}}$

11 **Return** $\tilde{\mathbf{Y}}$ and $\mathbf{f}$

Appendix C. Pseudo-Code of MA

Algorithm A3: MA($\mathbf{X}_1$,$\mathbf{X}_2$,$\mathbf{E}_1$,$\mathbf{E}_2$,k)

Input:

$\mathbf{X}_1$: the data source $\mathbf{X}_1 \in \mathbb{R}^{m_1 \times n_1}$ with n_1 instances and m_1 dimensions
$\mathbf{X}_2$: the data source $\mathbf{X}_2 \in \mathbb{R}^{m_2 \times n_2}$ with n_2 instances and m_2 dimensions
$\mathbf{E}_1$: $\mathbf{E}_1 \in \mathbb{R}^{1 \times n_1^*}$ with $n_1^* < n_1$, labels for the first n_1^* instances of $\mathbf{X}_1$
$\mathbf{E}_2$: $\mathbf{E}_2 \in \mathbb{R}^{1 \times n_2^*}$ with $n_2^* < n_2$, labels for the first n_2^* instances of $\mathbf{X}_2$
k: the number of local neighbors

Output:

$\tilde{\mathbf{Y}}_1$: the projected data of $\mathbf{X}_1$.
$\tilde{\mathbf{Y}}_2$: the projected data of $\mathbf{X}_2$.
$\mathbf{f}_1$: the projection maps data $\mathbf{X}_1$ to $\tilde{\mathbf{Y}}_1$
$\mathbf{f}_2$: the projection maps data $\mathbf{X}_2$ to $\tilde{\mathbf{Y}}_2$

1 construct $(n_1 + n_2)$ by $(n_1 + n_2)$ binary matrices $\tilde{\mathbf{A}}_s$ (Equation (9)) and $\tilde{\mathbf{A}}_d$ (Equation (10)) using $\mathbf{E}_1$ and $\mathbf{E}_2$
2 construct $(n_1 + n_2)$ by $(n_1 + n_2)$ binary matrix $\tilde{\mathbf{A}}_g$ (Equation (11)) using k-nearest-neighbor with the given k
3 construct degree matrices $\tilde{\mathbf{D}}_s$, $\tilde{\mathbf{D}}_d$, and $\tilde{\mathbf{D}}_g$ with $\tilde{\mathbf{A}}_s$, $\tilde{\mathbf{A}}_d$, and $\tilde{\mathbf{A}}_g$, respectively
4 construct Laplacian matrices $\tilde{\mathbf{L}}_s$, $\tilde{\mathbf{L}}_d$, and $\tilde{\mathbf{L}}_g$ as instructed in Equation (17)
5 organize the data matrix $\tilde{\mathbf{X}}$ as instructed in Equation (17)
6 solve the generalized eigenvalue decomposition $\tilde{\mathbf{X}}(\mu\tilde{\mathbf{L}}_g + \tilde{\mathbf{L}}_s)\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}\tilde{\mathbf{L}}_d\tilde{\mathbf{X}}^T\mathbf{f}$ so that $\mathbf{f}_1$ and $\mathbf{f}_2$ are achieved, $\mathbf{f} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix}$.
7 calculate $\tilde{\mathbf{Y}}_1 = \mathbf{f}_1^T\mathbf{X}_1$ and $\tilde{\mathbf{Y}}_2 = \mathbf{f}_2^T\mathbf{X}_2$
8 **Return** $\tilde{\mathbf{Y}}_1$, $\tilde{\mathbf{Y}}_2$, $\mathbf{f}_1$, $\mathbf{f}_2$

Appendix D. Pseudo-Code of MIMA

Algorithm A4: MIMA-MAPPER(X,b,c,F)

Input:

X: the data source $\mathbf{X} \in \mathbb{R}^{m \times n}$ with n instances and m dimensions
b: the number of data bins
c: the overlapping rate
F: the filtering function

Output:

$\mathbf{A}_{MIMA}$: the connection matrix

1 calculate the parameter space **XF**
2 divide **XF** into **b** intervals with **c**% overlap of adjacent intervals
3 divide data **X** into **b** bins corresponding to intervals achieved in 2
4 **for** (each data bin):
5 Spectral clustering
6 **end for**
7 Construct topological matrix $\mathbf{A}_{MIMA}(p,q) = \begin{cases} 1, & \text{if } p \text{ and } q \text{ in the same cluster;} \\ 1, & \text{if } p \text{ and } q \text{ in the linked clusters;} \\ 0, & \text{otherwise.} \end{cases}$
8 **Return** $\mathbf{A}_{MIMA}$

Algorithm A5: MIMA($\mathbf{X}_1$,$\mathbf{X}_2$,$\mathbf{E}_1$,$\mathbf{E}_2$,k)

Input:

$\mathbf{X}_1$: the data source $\mathbf{X}_1 \in \mathbb{R}^{m_1 \times n_1}$ with n_1 instances and m_1 dimensions

$\mathbf{X}_2$: the data source $\mathbf{X}_2 \in \mathbb{R}^{m_2 \times n_2}$ with n_2 instances and m_2 dimensions

$\mathbf{E}_1$: $\mathbf{E}_1 \in \mathbb{R}^{1 \times n_1^*}$ with $n_1^* < n_1$, labels for the first n_1^* instances of $\mathbf{X}_1$

$\mathbf{E}_2$: $\mathbf{E}_2 \in \mathbb{R}^{1 \times n_2^*}$ with $n_2^* < n_2$, labels for the first n_2^* instances of $\mathbf{X}_2$

k: the number of local neighbors

Output:

$\tilde{\mathbf{Y}}_1$: the projected data of $\mathbf{X}_1$.

$\tilde{\mathbf{Y}}_2$: the projected data of $\mathbf{X}_2$.

$\mathbf{f}_1$: the projection maps data $\mathbf{X}_1$ to $\tilde{\mathbf{Y}}_1$

$\mathbf{f}_2$: the projection maps data $\mathbf{X}_2$ to $\tilde{\mathbf{Y}}_2$

construct $(n_1 + n_2)$ by $(n_1 + n_2)$ binary matrices $\tilde{\mathbf{A}}_s$ (Equation (9)) and $\tilde{\mathbf{A}}_d$ (Equation (10)) using $\mathbf{E}_1$ and $\mathbf{E}_2$

for(i=1:2)

  $\mathbf{A}_{MIMA}^{i,i} =$ **MIMA-MAPPER**($\mathbf{X}_i$,b,c)

end

construct matrix $\tilde{\mathbf{A}}_{MIMA} = \begin{pmatrix} \mathbf{A}_{MIMA}^{1,1} & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & \mathbf{A}_{MIMA}^{K,K} \end{pmatrix}$

construct degree matrices $\tilde{\mathbf{D}}_s$, $\tilde{\mathbf{D}}_d$, and $\tilde{\mathbf{D}}_{MIMA}$ with $\tilde{\mathbf{A}}_s$, $\tilde{\mathbf{A}}_d$, and $\tilde{\mathbf{A}}_{MIMA}$, respectively

construct Laplacian matrices $\tilde{\mathbf{L}}_s$, $\tilde{\mathbf{L}}_d$, and $\tilde{\mathbf{L}}_{MIMA}$ as instructed in Equation (17)

organize the data matrix $\tilde{\mathbf{X}}$ as instructed in Equation (17)

solve the generalized eigenvalue decomposition $\tilde{\mathbf{X}}(\mu\tilde{\mathbf{L}}_g + \tilde{\mathbf{L}}_s)\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}\tilde{\mathbf{L}}_{MIMA}\tilde{\mathbf{X}}^T\mathbf{f}$ so that $\mathbf{f}_1$ and $\mathbf{f}_2$ are achieved, $\mathbf{f} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix}$

calculate $\tilde{\mathbf{Y}}_1 = \mathbf{f}_1^T\mathbf{X}_1$ and $\tilde{\mathbf{Y}}_2 = \mathbf{f}_2^T\mathbf{X}_2$

Return $\tilde{\mathbf{Y}}_1$, $\tilde{\mathbf{Y}}_2$, $\mathbf{f}_1$, $\mathbf{f}_2$

References

1. Zhang, J. Multi-source remote sensing data fusion: status and trends. *Int. J. Image Data Fusion* **2010**, *1*, 5–24.
2. Dalla Mura, M.; Prasad, S.; Pacifici, F.; Gamba, P.; Chanussot, J.; Benediktsson, J.A. Challenges and opportunities of multimodality and data fusion in remote sensing. *Proc. IEEE* **2015**, *103*, 1585–1601.
3. Yokoya, N.; Grohnfeldt, C.; Chanussot, J. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 29–56.
4. Hong, D.; Yokoya, N.; Chanussot, J.; Zhu, X.X. CoSpace: Common Subspace Learning from Hyperspectral-Multispectral Correspondences. *arXiv* **2018**, arXiv:1812.11501.
5. Dalponte, M.; Bruzzone, L.; Gianelle, D. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1416–1427.
6. Swatantran, A.; Dubayah, R.; Roberts, D.; Hofton, M.; Blair, J.B. Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. *Remote Sens. Environ.* **2011**, *115*, 2917–2930.
7. Khodadadzadeh, M.; Li, J.; Prasad, S.; Plaza, A. Fusion of hyperspectral and LiDAR remote sensing data using multiple feature learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2971–2983.
8. Merkle, N.; Auer, S.; Müller, R.; Reinartz, P. Exploring the Potential of Conditional Adversarial Networks for Optical and SAR Image Matching. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1811–1820, doi:10.1109/JSTARS.2018.2803212.
9. Joshi, N.; Baumann, M.; Ehammer, A.; Fensholt, R.; Grogan, K.; Hostert, P.; Jepsen, M.R.; Kuemmerle, T.; Meyfroidt, P.; Mitchard, E.T.; et al. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sens.* **2016**, *8*, 70.

10. Wang, Y.; Zhu, X.X.; Zeisl, B.; Pollefeys, M. Fusing Meter-Resolution 4-D InSAR Point Clouds and Optical Images for Semantic Urban Infrastructure Monitoring. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 14–26, doi:10.1109/TGRS.2016.2554563.
11. Schmitt, M.; Hughes, L.H.; Zhu, X.X. THE SEN1-2 Dataset for DEEP LEARNING IN SAR-OPTICAL DATA FUSION. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *IV-1*, 141–146, doi:10.5194/isprs-annals-IV-1-141-2018.
12. Hong, D.; Yokoya, N.; Ge, N.; Chanussot, J.; Zhu, X.X. Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 193–205.
13. Chang, Y.L.; Han, C.C.; Ren, H.; Chen, C.T.; Chen, K.S.; Fan, K.C. Data fusion of hyperspectral and SAR images. *Opt. Eng.* **2004**, *43*, 1787–1798.
14. Koch, B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 581–590.
15. Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M.; et al. GMES Sentinel-1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24.
16. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36.
17. Stuffler, T.; Kaufmann, C.; Hofer, S.; Förster, K.; Schreier, G.; Mueller, A.; Eckardt, A.; Bach, H.; Penne, B.; Benz, U.; et al. The EnMAP hyperspectral imager—An advanced optical payload for future applications in Earth observation programmes. *Acta Astronaut.* **2007**, *61*, 115–120.
18. Wu, X.; Hong, D.; Ghamisi, P.; Li, W.; Tao, R. MsRi-CCF: Multi-scale and rotation-insensitive convolutional channel features for geospatial object detection. *Remote Sens.* **2018**, *10*, 1990.
19. Wu, X.; Hong, D.; Tian, J.; Chanussot, J.; Li, W.; Tao, R. ORSIm Detector: A Novel Object Detection Framework in Optical Remote Sensing Imagery Using Spatial-Frequency Channel Features. *arXiv* **2019**, arXiv:1901.07925.
20. Hong, D.; Zhu, X.X. SULoRA: Subspace unmixing with low-rank attribute embedding for hyperspectral data analysis. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 1351–1363.
21. Drumetz, L.; Veganzones, M.A.; Henrot, S.; Phlypo, R.; Chanussot, J.; Jutten, C. Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability. *IEEE Trans. Image Process.* **2016**, *25*, 3890–3905.
22. Hong, D.; Yokoya, N.; Chanussot, J.; Zhu, X.X. Learning a low-coherence dictionary to address spectral variability for hyperspectral unmixing. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 235–239.
23. Ceamanos, X.; Waske, B.; Benediktsson, J.A.; Chanussot, J.; Fauvel, M.; Sveinsson, J.R. A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *Int. J. Image Data Fusion* **2010**, *1*, 293–307.
24. Hong, D.; Yokoya, N.; Chanussot, J.; Zhu, X.X. An augmented linear mixing model to address spectral variability for hyperspectral unmixing. *IEEE Trans. Image Process.* **2019**, *28*, 1923–1938.
25. Lee, J.S.; Pottier, E. *Polarimetric Radar Imaging: From Basics to Applications*; CRC Press: Boca Raton, FL, USA, 2009.
26. Cloude, S.R.; Pottier, E. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 68–78.
27. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43.
28. Schmitt, A.; Wendleder, A.; Hinz, S. The Kennaugh element framework for multi-scale, multi-polarized, multi-temporal and multi-frequency SAR image preparation. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 122–139.
29. Hu, J.; Ghamisi, P.; Zhu, X. Feature Extraction and Selection of Sentinel-1 Dual-Pol Data for Global-Scale Local Climate Zone Classification. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 379.
30. Jouan, A.; Allard, Y. Land use mapping with evidential fusion of features extracted from polarimetric synthetic aperture radar and hyperspectral imagery. *Inf. Fusion* **2004**, *5*, 251–267.

31. Li, T.; Zhang, J.; Zhao, H.; Shi, C. Classification-oriented hyperspectral and PolSAR images synergic processing. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, Australia, 21–26 July 2013; pp. 1035–1038.
32. Dabbiru, L.; Samiappan, S.; Nobrega, R.A.A.; Aanstoos, J.V.; Younan, N.H.; Moorhead, R.J. Fusion of synthetic aperture radar and hyperspectral imagery to detect impacts of oil spill in Gulf of Mexico. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015.
33. Hu, J.; Ghamisi, P.; Schmitt, A.; Zhu, X. Object Based Fusion of Polarimetric SAR and Hyperspectral Imaging for Land Use Classification. In Proceedings of the 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, USA, 21–24 August 2016.
34. Hu, J.; Mou, L.; Schmitt, A.; Zhu, X.X. FusioNet: A two-stream convolutional neural network for urban scene classification using PolSAR and hyperspectral data. In Proceedings of the 2017 Joint Urban Remote Sensing Event (JURSE), Dubai, UAE, 6–8 March 2017; pp. 1–4.
35. Wang, C.; Mahadevan, S. A General Framework for Manifold Alignment. In Proceedings of the AAAI Fall Symposium: Manifold Learning and Its Applications, 2009; pp. 53–58.
36. Tuia, D.; Volpi, M.; Trolliet, M.; Camps-Valls, G. Semisupervised manifold alignment of multimodal remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7708–7720.
37. Ghamisi, P.; Benediktsson, J.A.; Phinn, S. Land-cover classification using both hyperspectral and LiDAR data. *Int. J. Image Data Fusion* **2015**, *6*, 189–215.
38. Xia, J.; Yokoya, N.; Iwasaki, A. Hyperspectral image classification with canonical correlation forests. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 421–431.
39. Yokoya, N.; Ghamisi, P.; Xia, J.; Sukhanov, S.; Heremans, R.; Tankoyeu, I.; Bechtel, B.; Le Saux, B.; Moser, G.; Tuia, D. Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS Data Fusion Contest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1363–1377.
40. Makantasis, K.; Doulamis, A.; Doulamis, N.; Nikitakis, A.; Voulodimos, A. Tensor-Based Nonlinear Classifier for High-Order Data Analysis. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; 2 pp. 2221–2225.
41. Makantasis, K.; Doulamis, A.D.; Doulamis, N.D.; Nikitakis, A. Tensor-based classification models for hyperspectral data analysis. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6884–6898.
42. Yokoya, N.; Chanussot, J.; Iwasaki, A. Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1430–1437.
43. Wang, C.; Mahadevan, S. Manifold Alignment without Correspondence. In Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI), Pasadena, CA, USA, 11–17 July 2009; Volume 2, p. 3.
44. Wang, C.; Mahadevan, S. Heterogeneous domain adaptation using manifold alignment. *IJCAI Proc. Int. Jt. Conf. Artif. Intell.* **2011**, *22*, 1541.
45. Wang, C.; Mahadevan, S. Manifold Alignment Preserving Global Geometry. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI), Beijing, China, 3–9 August 2013; pp. 1743–1749.
46. Tuia, D.; Camps-Valls, G. Kernel manifold alignment for domain adaptation. *PLoS ONE* **2016**, *11*, e0148655.
47. Tuia, D.; Munoz-Mari, J.; Gómez-Chova, L.; Malo, J. Graph matching for adaptation in remote sensing. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 329–341.
48. Liao, W.; Pižurica, A.; Bellens, R.; Gautama, S.; Philips, W. Generalized graph-based fusion of hyperspectral and LiDAR data using morphological features. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 552–556.
49. Volpi, M.; Camps-Valls, G.; Tuia, D. Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis. *ISPRS J. Photogramm. Remote Sens.* **2015**, *107*, 50–63.
50. Liao, D.; Qian, Y.; Zhou, J.; Tang, Y.Y. A manifold alignment approach for hyperspectral image visualization with natural color. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3151–3162.
51. Hong, D.; Yokoya, N.; Zhu, X.X. Learning a Robust Local Manifold Representation for Hyperspectral Dimensionality Reduction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2960–2975.
52. He, X.; Niyogi, P. Locality preserving projections. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2004; pp. 153–160.

53. Hu, J.; Hong, D.; Zhu, X.X. MIMA: MAPPER-Induced Manifold Alignment for Semi-Supervised Fusion of Optical Image and Polarimetric SAR Data. *IEEE Trans. Geosci. Remote Sens.* **2019**, under review.
54. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326.
55. Tenenbaum, J.B.; De Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323.
56. Hatcher, A. *Algebraic Topology*; Tsinghua University Press: Beijing, China, 2005.
57. Lin, T.; Zha, H. Riemannian manifold learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 796–809.
58. Friedman, J.H.; Bentley, J.L.; Finkel, R.A. An algorithm for finding best matches in logarithmic time. *ACM Trans. Math. Softw.* **1976**, *3*, 209–226.
59. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.
60. Schölkopf, B.; Smola, A.J.; Bach, F. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.
61. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
62. Rainforth, T.; Wood, F. Canonical correlation forests. *arXiv* **2015**, arXiv:1507.05444.
63. Hong, D.; Yokoya, N.; Zhu, X.X. The K-LLE algorithm for nonlinear dimensionality ruduction of large-scale hyperspectral data. In Proceedings of the 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, USA, 21–24 August 2016; pp. 1–5.
64. Hong, D.; Yokoya, N.; Zhu, X.X. Local manifold learning with robust neighbors selection for hyperspectral dimensionality reduction. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 40–43.
65. Belkin, M.; Niyogi, P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2002; pp. 585–591.
66. Hong, D.; Pan, Z.; Wu, X. Improved differential box counting with multi-scale and multi-direction: A new palmprint recognition method. *Opt.-Int. J. Light Electron Opt.* **2014**, *125*, 4154–4160.
67. He, X.; Cai, D.; Yan, S.; Zhang, H.J. Neighborhood preserving embedding. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–21 October 2005; Volume 2, pp. 1208–1213.
68. Hong, D.; Liu, W.; Su, J.; Pan, Z.; Wang, G. A novel hierarchical approach for multispectral palmprint recognition. *Neurocomputing* **2015**, *151*, 511–521.
69. Hong, D.; Liu, W.; Wu, X.; Pan, Z.; Su, J. Robust palmprint recognition based on the fast variation Vese–Osher model. *Neurocomputing* **2016**, *174*, 999–1012.
70. Donoho, D.L.; others. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Chall. Lect.* **2000**, *1*, 32.
71. Farrell, M.D.; Mersereau, R.M. On the impact of PCA dimension reduction for hyperspectral detection of difficult targets. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 192–195.
72. Debes, C.; Merentitis, A.; Heremans, R.; Hahn, J.; Frangiadakis, N.; van Kasteren, T.; Liao, W.; Bellens, R.; Pižurica, A.; Gautama, S.; et al. Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2405–2418.
73. Chintakunta, H.; Robinson, M.; Krim, H. Introduction to the special session on Topological Data Analysis, ICASSP 2016. In Proceedings o fthe 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 6410–6414. doi:10.1109/ICASSP.2016.7472911.
74. Chazal, F.; Michel, B. An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv* **2017**, arXiv:1710.04019.
75. Zomorodian, A.; Carlsson, G. Computing persistent homology. *Discret. Comput. Geom.* **2005**, *33*, 249–274.
76. Edelsbrunner, H.; Letscher, D.; Zomorodian, A. Topological persistence and simplification. In Proceedings of the Proceedings 41st Annual Symposium on Foundations of Computer Science, Redondo Beach, CA, USA, 12–14 November 2000; pp. 454–463.
77. Edelsbrunner, H. *A Short Course in Computational Geometry and Topology*; Springer: Cham, Switzerland, 2014.
78. Singh, G.; Mémoli, F.; Carlsson, G.E. Topological methods for the analysis of high dimensional data sets and 3D object recognition. In Proceedings of the Eurographics Symposium on Point-Based Graphics (SPBG), 2007; pp. 91–100.

79. Nicolau, M.; Levine, A.J.; Carlsson, G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7265–7270.
80. Nielson, J.L.; Paquette, J.; Liu, A.W.; Guandique, C.F.; Tovar, C.A.; Inoue, T.; Irvine, K.A.; Gensel, J.C.; Kloke, J.; Petrossian, T.C.; et al. Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. *Nat. Commun.* **2015**, *6*, 8581.
81. Li, L.; Cheng, W.Y.; Glicksberg, B.S.; Gottesman, O.; Tamler, R.; Chen, R.; Bottinger, E.P.; Dudley, J.T. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci. Transl. Med.* **2015**, *7*, 311ra174.
82. Lum, P.Y.; Singh, G.; Lehman, A.; Ishkanov, T.; Vejdemo-Johansson, M.; Alagappan, M.; Carlsson, J.; Carlsson, G. Extracting insights from the shape of complex data using topology. *Sci. Rep.* **2013**, *3*, srep01236.
83. Carriere, M.; Michel, B.; Oudot, S. Statistical analysis and parameter selection for Mapper. *J. Mach. Learn. Res.* **2018**, *19*, 478–516.
84. Ng, A.Y.; Jordan, M.I.; Weiss, Y. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2002; pp. 849–856.
85. Okujeni, A.; Van Der Linden, S.; Hostert, P. *Berlin-Urban-Gradient dataset 2009—An EnMAP Preparatory Flight Campaign (Datasets)*; GFZ Data Services: 2016.
86. Hu, J.; Guo, R.; Zhu, X.; Baier, G.; Wang, Y. Non-local means filter for polarimetric SAR speckle reduction-experiments using TerraSAR-x data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 71.
87. Haklay, M.; Weber, P. Openstreetmap: User-generated street maps. *IEEE Pervasive Comput.* **2008**, *7*, 12–18.
88. Benediktsson, J.A.; Palmason, J.A.; Sveinsson, J.R. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 480–491.
89. Ghamisi, P.; Benediktsson, J.A.; Ulfarsson, M.O. Spectral—Spatial classification of hyperspectral images based on hidden Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2565–2574.
90. Liao, W.; Chanussot, J.; Dalla Mura, M.; Huang, X.; Bellens, R.; Gautama, S.; Philips, W. Taking Optimal Advantage of Fine Spatial Resolution: Promoting partial image reconstruction for the morphological analysis of very-high-resolution images. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–28.
91. Rasti, B.; Ghamisi, P.; Gloaguen, R. Hyperspectral and lidar fusion using extinction profiles and total variation component analysis. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3997–4007.
92. Zhu, Z.; Woodcock, C.E.; Rogan, J.; Kellndorfer, J. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. *Remote Sens. Environ.* **2012**, *117*, 72–82.
93. Wurm, M.; Taubenböck, H.; Weigand, M.; Schmitt, A. Slum mapping in polarimetric SAR data using spatial features. *Remote Sens. Environ.* **2017**, *194*, 190–204.
94. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 27.
95. Banerjee, M.; Capozzoli, M.; McSweeney, L.; Sinha, D. Beyond kappa: A review of interrater agreement measures. *Can. J. Stat.* **1999**, *27*, 3–23.
96. Hong, D.; Yokoya, N.; Xu, J.; Zhu, X. Joint and progressive learning from high-dimensional data for multi-label classification. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2018; pp. 478–493.
97. Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496.
98. Chazal, F.; Guibas, L.J.; Oudot, S.Y.; Skraba, P. Persistence-based clustering in riemannian manifolds. *J. ACM (JACM)* **2013**, *60*, 41.

Appendices

D Zhu X X, Hu J, Qiu C, Shi Y, Kang J, Mou L, Bagheri H, Häberle M, Hua Y, Huang R, Hughes L, Li H, Sun Y, Zhang G, Han S, Schmitt M, and Wang Y (2019). So2Sat LCZ42: A Benchmark Dataset for GlobalLocal Climate Zones Classification. IEEE Geoscience and Remote Sensing Magazine, doi: 10.1109/MGRS.2020.2964708.

So2Sat LCZ42: A Benchmark Dataset for Global Local Climate Zones Classification

Xiao Xiang Zhu, *Senior Member, IEEE*, Jingliang Hu, Chunping Qiu, Yilei Shi, Jian Kang, Lichao Mou, Hossein Bagheri, Matthias Häberle, Yuansheng Hua, Rong Huang, Lloyd Hughes, Hao Li, Yao Sun, Guichen Zhang, Shiyao Han, Michael Schmitt, *Senior Member, IEEE*, and Yuanyuan Wang, *Member, IEEE*

***Abstract*—Access to labeled data is of vital importance in supervised machine learning endeavors. This is especially true for an automatic analysis of remote sensing images on a global scale, which enables us to address global challenges such as urbanization and climate change, and assess the progress of UN sustainable development goals using state-of-the-art machine learning techniques. To meet these pressing needs, especially in urban research, we provide open access to a valuable benchmark dataset named "*So2Sat LCZ42*," which consists of local climate zone (LCZ) labels of about half a million Sentinel-1 and Sentinel-2 image patches in 42 urban agglomerations (plus 10 additional smaller areas) across the globe. This dataset was labeled by 15 domain experts following a carefully designed labeling work flow and evaluation process over a period of six months. Despite this intensive labor, we would like to share this dataset, which is intrinsically different from many other semantic land use and land cover classifications because LCZs provide an objective measure of the morphology, compactness, and height of urban areas. These measures are human and culturally independent, which is fundamental in creating an unbiased globally-distributed dataset for urban growth monitoring using machine learning methods. This dataset can be accessed from http://doi.org/10.14459/2018mp1483140.**

***Index Terms*—enchmark dataset, classification, deep learning, Earth observation, local climate zones (LCZs), machine learning, multi-spectral, remote sensing, SAR, Sentinel-1, Sentinel-2, urban areasenchmark dataset, classification, deep learning, Earth observation, local climate zones (LCZs), machine learning, multi-spectral, remote sensing, SAR, Sentinel-1, Sentinel-2, urban areasB**

I. Introduction

THE production of land use/land cover (LULC) maps at large or even global scale is an essential task in the field of remote sensing. These maps can provide valuable input to a large number of societal questions, such as understanding human poverty or climate change, supporting the conservation of biodiversity and ecosystems, and providing stakeholder information for disaster management and sustainable urban development [1]. Urbanization is undoubtedly the most important mega-trends in the 21st century, after climate change. Currently, half of humanity – 3.5 billion people – lives in cities. Shockingly, 1 billion of them still live in slums. Therefore, sustainable urban development has become one of the 17 sustainable development goals (SDGs) of the United Nations. Today, sustainable development increasingly depends on the successful management of urban growth, especially in developing countries where the pace of urbanization is projected to be the fastest, according to *World Urbanization Prospects: The 2018 Revision* [2]. LULC maps enable us to describe, track, and manage urban growth in an objective and consistent manner.

Examples of global LULC products created by the remote sensing community include the Global Urban Footprint (GUF) [3], [4], produced from synthetic aperture radar (SAR) data acquired by the TanDEM-X (TD-X) mission; the Global Human Settlement Layer (GHSL) produced from global, multi-temporal archives of fine-scale satellite imagery, census data, and volunteered geographic information [5]; and the Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) and GlobeLand30 datasets, produced from 30m-resolution Landsat data [6]. This list is not exhaustive. However, these products all provide semantic labels of urban/non-urban, or even finer classes. These semantic labels are often subjective (to human interpretation), and culture-dependent. For example, the definition of urban and non-urban areas might be drastically different in Europe and Africa, and from person to person.

A. Advantage of LCZ classification for global urban mapping

For a consistent analysis of the urban areas across the globe, an objective and culture-independent classification scheme of urban areas that is applicable worldwide is pressingly needed. After extensive research, we turned to Local Climate Zones (LCZs). LCZs were originally developed for metadata communication of observational urban heat island studies, and have also been shown to support the assessment of social inequalities [7]. There are a total of 17 classes in the LCZ classification scheme, where 10 are built classes and 7 are natural classes. They are based on climate-relevant surface properties on local scale, which are mainly related to 3D surface structure (e.g., height and density of buildings and

This work is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. ERC-2016-StG-714087, acronym: So2Sat, www.so2sat.eu).

X. Zhu is with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), as well as Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM).

J. Hu, M. Häberle, Y. Hua, Y. Sun are with DLR-IMF.

C. Qiu, J. Kang, L. Mou, H. Bagheri, L. Hughes, M. Schmitt, and Y. Wang are with TUM-SiPEO.

Y. Shi is with the Chair of Remote Sensing Technology, TUM.

R. Huang, H. Li, G. Zhang, and S. Han were with TUM-SiPEO when conducting the work of this paper.

(Correspondence: Xiao Xiang Zhu; E-mail: xiaoxiang.zhu@dlr.de)

Manuscript received March XX, 2019; revised XX XX, 2019.

trees), surface cover (e.g., vegetation or paving), as well as anthropogenic parameters (such as human-based heat output). A schematic drawing of the 17 classes is shown in the left of Fig. 1. The middle of the figure shows the LCZ classification of Vancouver, Canada, created by the authors. As can be seen the 10 urban classes describe the morphology of the area, including the density and the height of the buildings, as well as the percentage of the impervious surface. The urban classes are mostly coded by red, with decreasing intensities as the building density and height decreases from compact high-rise to open low-rise. For example, the dark red part marked by the yellow rectangle is downtown Vancouver, where most of Vancouver's high-rise buildings are located. As a reference, the Google image of this area is shown in the right of Fig. 1. The light red part of the classification map is mostly low-rise residential houses. Because the LCZ classes are defined by their physical properties, they are generic and applicable to cities over the world, offering the potential to compare different areas of different cities with trenchant distinctions representing the heterogeneous thermal behavior within an urban environment [8]. Besides the increasing impact on worldwide climatological studies, such as the cooling effect of green infrastructure and micro-climatic effects on town peripheries [9]–[18], researchers have recently started to use the LCZ approach to classify the internal structure of urban areas, providing promising information for various applications such as infrastructure planning, disaster mitigation, health and green space planning, and population assessment [19], [20] in this increasingly urbanized world [1]. The remote sensing community also addressed its particular attention to this topic by organizing the 2017 IEEE data fusion contest with the goal of LCZ classification [21].

B. Related work in LCZ classification

Recently, various promising classifiers for LCZ have been proposed by different research groups. They include random forests, support vector machines [23], canonical correlation forests [24], [25], rotation forests [21], gradient boosting machines [26], and ensembles of multiple classifiers [27]. Deep learning certainly played an important role in LULC using remote sensing data [28]. Multiple algorithms based on convolutional neural networks [29]–[34] have been developed.

A significant part of the existing development of LCZ classification is community-based large-scale LCZ mapping using freely available Landsat data and softwares [35]–[37]. World Urban Database and Portal (WUDAPT) [22], a community-driven initiative, was organized by researchers to map high-quality LCZ maps worldwide through the contribution of the community. Within this framework, currently almost 100 cities worldwide have been mapped with a moderate quality, providing sufficient detail for certain model applications [38]. LCZ maps of tens of cities, after undergoing quality assessment and generation of metadata, are now openly available in the WUDAPT portal. More recently, LCZs of Europe is being mapped as part of the WUDAPT project, with data including Sentinel-1, Sentinel-2, and the Defense Meteorological Program (DMSP) Operational Linescan System (OLS) night-time lights product [39]. These LCZ maps and tools can benefit various studies, such as modelling the effects of changes to the energy infrastructure and simulating the urban climate [38].

The procedures of community-based LCZ mapping mainly consist of two steps: 1) labeling ground truth data in Google Earth, and 2) classification using shallow learning algorithms such as random forest in GIS software, a process that is detailed in [8] and Section II-A. These community-based efforts mark the first step toward a more synergetic cooperation among researchers. Yet, multiple studies have reported that the quality of the produced LCZ maps is inconsistent [40], [41], as the procedures strongly rely on the knowledge of individual volunteers.

Therefore, it still requires a significant development toward a global LCZ mapping because of the lack of high quality labels, and transferable classifiers for global deployment. According to [8], [38], [42], regional variations in vegetation and artificial materials, as well as significant variations in cultural and physical environmental factors, cause large intra-class variability of spectral signatures. As a result, transferability, i.e., classifying one city with data from a different city, is challenging and not possible to result in a large-scale product in a short time, as demonstrated in [42]. Existing effort to further improve LCZ classification results includes fusing multi-source data such as SAR images and Google Street View, as well as developing advanced models with high generalization ability to facilitate efficient up-scaling [33], [39], [42]. In order to fully exploit the state-of-the-art machine learning algorithms for global LCZ mapping, it is necessary to create a large quantity of reliable reference data as a first step. This task will be addressed in this article.

C. Contribution of this paper

To answer the pressing need for LCZ training datasets, we carefully selected 42 urban agglomerations plus 10 additional smaller areas across all the inhabited continents (except Antarctica) around the globe. Their geographic distribution can be seen in Fig. 2. A large quantity of polygons in those cites were manually labeled by the authors. We find that only following the definition of LCZs in [9] and the labeling process mentioned in WUDAPT is not optimal for a joint labeling activity by a group of people. To ensure the highest possible quality of the result, we designed a rigorous labeling work flow, shown in Fig. 4. Meetings were conducted before and during the labeling process to calibrate our understanding of the definition of the 17 classes. Afterwards, the labeling results from each member of the labeling crew were visually inspected by a different person to spot obvious errors. Last but not least, we conducted a quantitative evaluation of 10 cities in the dataset by having a group of remote sensing experts cast 10 independent votes on each labeled polygon, in order to identify possible errors and assess the human labeling accuracy. This rigorous labeling processing took approximately 15 person-months. By projecting these labels to coregistered Sentinel-1 and Sentinel-2 images, we obtained $400{,}673$ pairs of corresponding Sentinel-1 SAR and Sentinel-2 multi-spectral image patches with LCZ labels. An impression of the Sentinel

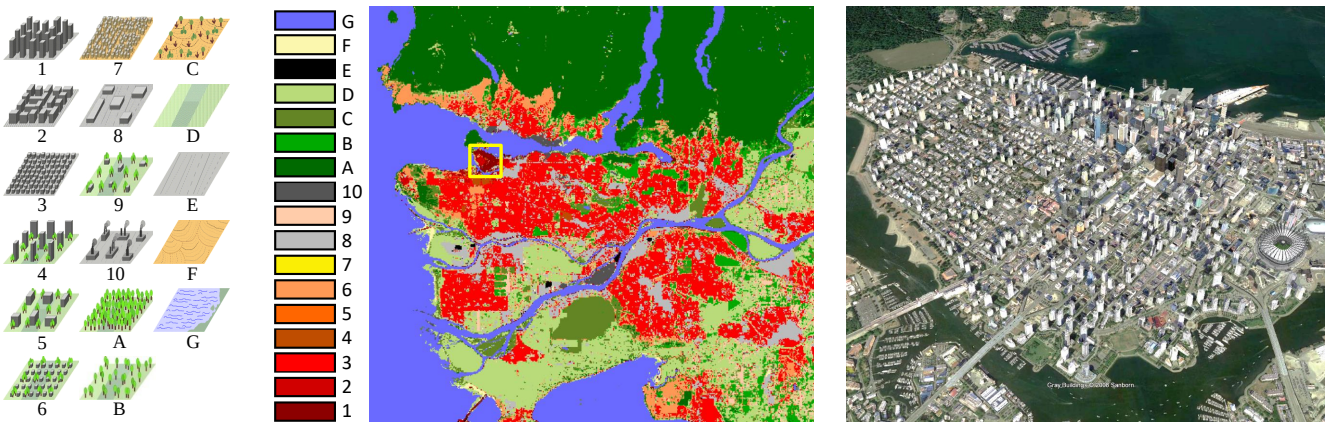


Fig. 1: Left: the schematic drawing of the 17 LCZ classes; middle: the LCZ classification map of Vancouver, Canada, created by the authors; and right: the Google image of downtown Vancouver; where most of the high-rise buildings are located. The yellow rectangle in the LCZ map marks the downtown areas. The left subfigure was modified from the WUDAPT [22].

image patch pairs in the dataset can be seen in Fig. 3. However, the actual patches in the dataset have a dimension of 320m by 320m, which is smaller than the visualization in Fig. 3. In this paper, we provide open access to this high quality *So2Sat LCZ42* dataset to the research community. It is meant to foster the development of fully automatic classification pipelines based on modern machine learning approaches, and support the accelerated use of LCZ mapping at a global scale.

- Provide the community a large dataset of
- Designed a LCZ labeling procedure that
-

D. Structure of this paper

The remainder of this paper is structured as follows: Section II describes the detailed procedure of creating the So2Sat LCZ42 dataset. Section III summarizes the label validation endeavour carried out by the authors, in order to assess the quality of our manual labels. Section IV highlights some baseline classification results obtained using conventional machine learning algorithms on the dataset. Finally, Section VI concludes the paper, and provides our insights into the future development of this topic.

II. Creating the So2Sat LCZ42 Dataset

The quality of training labels is of great importance to machine learning approaches. Yet, annotating remote sensing data remains a challenge for a variety of reasons, including intense labor cost, requirement of domain expertise, and personal discrepancy in understanding of each class. Thus, to ensure the quality of the So2Sat LCZ42 dataset, a four-phase labeling process was designed. It was designed to maximize label consistency and minimize human error. The four phases are: learning, labeling, visual validation, and quantitative validation. They can be seen in Fig 4 as blocks A, B, C, and D, respectively. The detailed procedures in each phase are introduced in this section. In addition, we also downloaded the corresponding Sentinel-1 and Sentinel-2 images of the 52 areas. Proper preprocessing procedures were performed on the two types of images. By projecting the label to the Sentinel images, we extracted coregistered Sentinel-1 and Sentinel-2 image pairs with corresponding LCZ labels.

A. Creating the labels

1) Learning phase: The learning phase aims at creating a standard for the labeling crew. There are two reasons of doing so. First, according to the criteria of LCZ classes given in [9] and listed in Table. I, the LCZ classes definition are not mutually disjoint (e.g. class 3 compact low-rise and 8 large low-rise), and their union also does not describe the whole Earth land surface. That is to say that some areas do not fall into any of the LCZ classes, and some can be labeled to multiple classes. Second, the interpretation of the definition by different persons still differs from each other. We started by building a visual impression of different LCZ classes in Google Earth, then moved toward a quantitative understanding of each class. As a result, we constructed a quantitative labeling decision rule according to the literal definition. This is shown in Fig. 7 in the appendix. A labeling ”examination” was conducted before the actual labeling started, where everyone in the labeling crew cast a vote on many selected scenes. Ambiguous scenes were selected and discussed, in order to calibrate everyone’s understanding.

2) Labeling phase: The labeling phase follows a standard procedure defined in the WUDAPT project [22]. First, each person on the labeling crew claimed a few cities among the 42 cities, and defined a region of interest (ROI) within each selected city by drawing a rectangle of approximately 50×50 kilometers around the city center in Google Earth. Second, polygons enclosing different LCZ classes were manually delineated in Google Earth. These polygons are the preliminary labels. Afterwards, Landsat 8 images were prepared for the ROI of the city. With the help of the SAGA GIS (http://www.saga-gis.org/en/index.html) software, a random forest classifier was trained using the Landsat 8 images and the preliminary LCZ labels, in order to produce a LCZ classification map of the specific city. This classification map and the satellite image

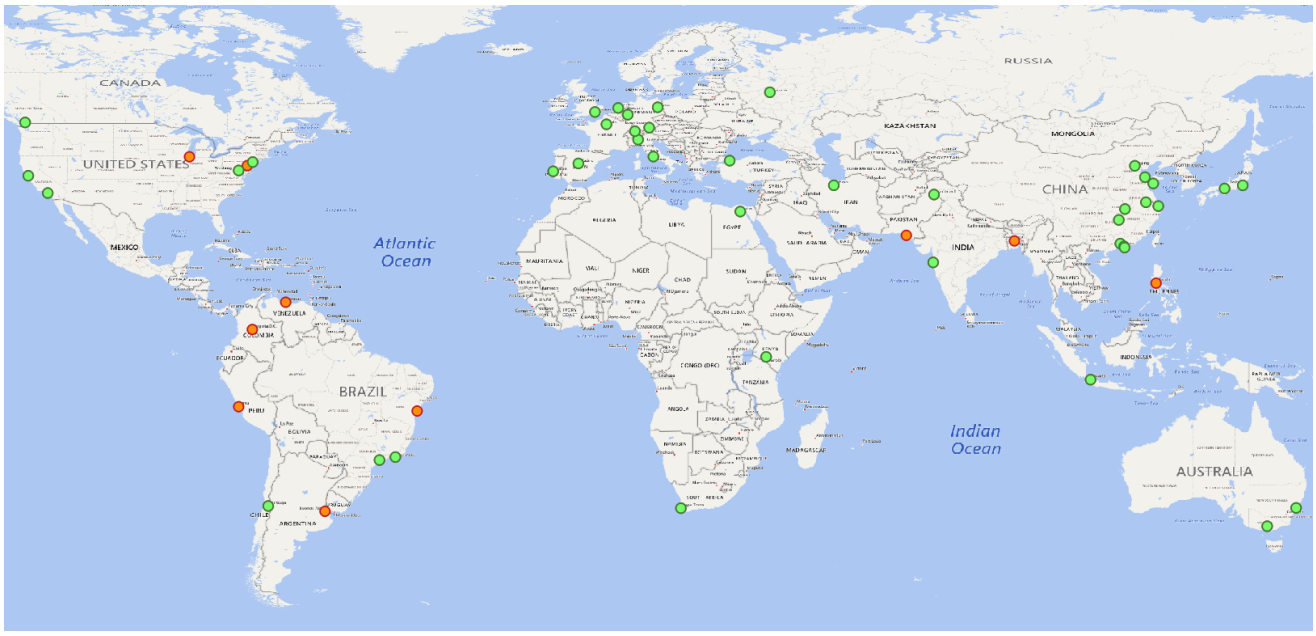


Fig. 2: The location of the 42 main cities (green dot) plus the 10 additional cities (orange dot) included in the So2Sat LCZ42 dataset.

TABLE I: Fractions of building surface, pervious surface, and impervious surface in percentage (%) of each class [9], as well as their height above the ground in meters.

Class		Building Surface Fraction [%]	Pervious Surface Fraction [%]	Impervious Surface Fraction [%]	Height above ground [m]
Compact high-rise	1	40-60	0-10	40-60	>25
Compact mid-rise	2	40-70	0-20	30-50	10 - 25
Compact low-rise	3	40-70	0-30	20-50	2 - 10
Open high-rise	4	20-40	30-40	30-40	>25
Open mid-rise	5	20-40	20-40	30-50	10 - 25
Open low-rise	6	20-40	30-60	20-50	2 - 10
Lightweight low-rise	7	60-90	0-30	0-20	2 - 10
Large low-rise	8	30-50	0-20	40-50	2 - 10
Sparsely built	9	10-20	60-80	0-20	2 - 10
Heavy industry	10	20-30	40-50	20-40	2 - 10
Dense trees	A	0-10	90-100	0-10	>3
Scattered tree	B	0-10	90-100	0-10	>3
Bush, scrub	C	0-10	90-100	0-10	1 - 2
Low plants	D	0-10	90-100	0-10	<1
Bare rock or paved	E	0-10	0-10	90-100	0
Bare soil or sand	F	0-10	90-100	0-10	0
Water	G	0-10	90-100	0-10	0

on Google Earth served as auxiliary data to cross-check the *correctness* and *completeness* of the LCZ labels. Regarding the correctness of the label, the crew visually inspected the discrepancies between the classification result and the label of the polygons. If a label mismatch was found for a labeled polygon, the crew inspected the satellite image on Google Earth, and corrected the given label if necessary. This process was repeated until no noticeable discrepancy between the classification result and the label was found. Regarding the completeness of the label, the labeling crew cross-checked the classification result with the satellite image on Google Earth in unlabeled areas, in order to find negative samples. For example, dense forest might be classified as water because it lacked the dense forest label. The labeling crew then labeled those negative samples of dense forest and included them in the whole label dataset. This hard negative mining procedure was carried out iteratively until no noticeable discrepancies between the classification map and the Google Earth image in unlabeled areas were found. It is important to point out that the classification maps produced during the manual labeling process were only employed to provide guidance to the labeling crew, and were not used in the final data. All LCZ labels in the final provided reference data fully relied on manual human annotation.

3) Visual quality control phase: Despite a clear quantitative definition that was conveyed to the labeling crew in the learning phase, personal bias and outliers still existed in the labeling result. A manual inspection was thus required before a quantitative validation to adjust personal biases, as well as decrease the inevitable human mistakes. Therefore, after

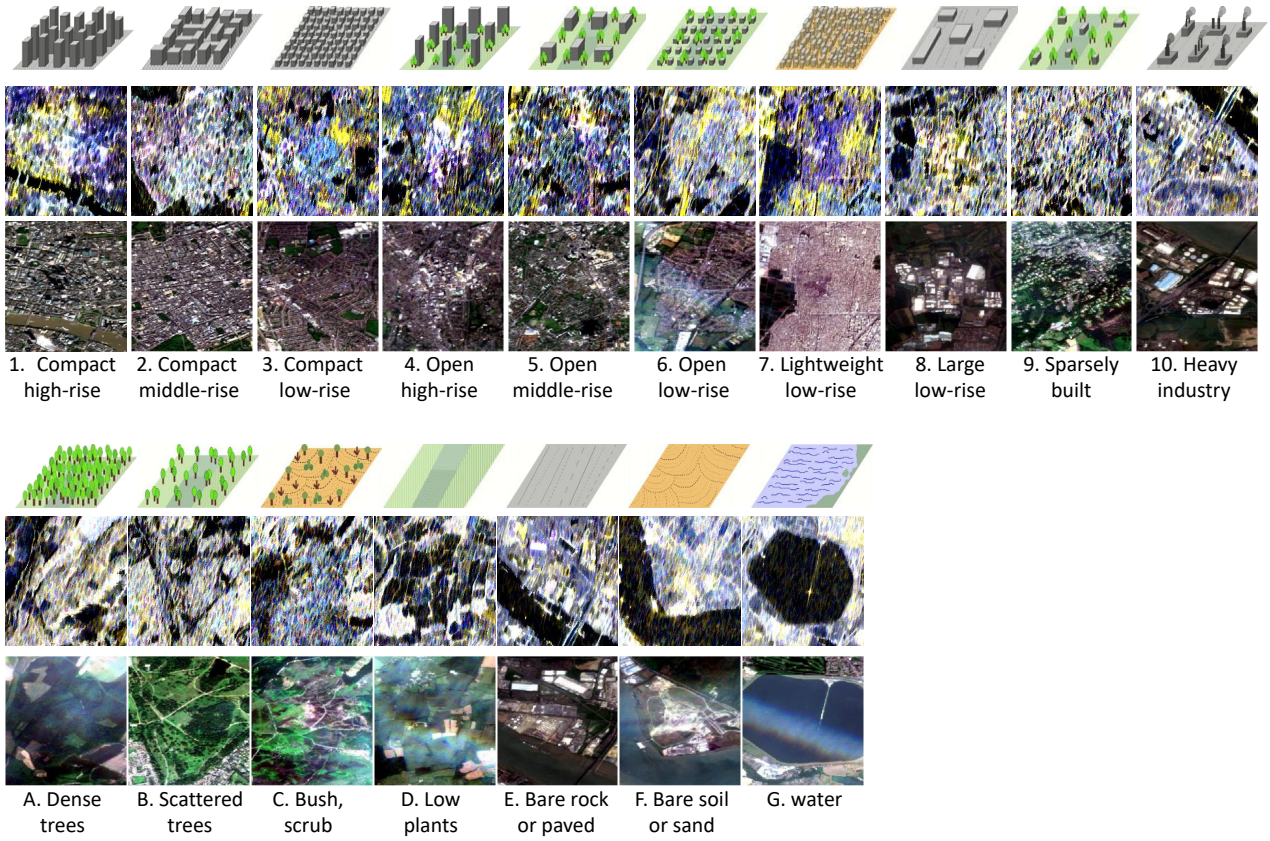


Fig. 3: Examples of the Sentinel-1 and Sentinel-2 image scenes of the 17 LCZ classes. In each LCZ, the upper image is the Sentinel-1 scene in PauliRGB visualization, and lower one is the corrsponding Sentinel-2 scene in RGB. For visualization purposes, the image scenes are much larger than the actual patches (32*32 pixel) in the So2Sat LCZ42 dataset.

the labeling phase, two persons other than the expert who labeled the polygons sequentially and independently validated the labels, as demonstrated in block C of Fig. 7. The two persons were responsible for visually inspecting two types of signals in the classification map: 1) obvious outliers, such as water being classified as a dense high-rise building, and 2) a normal compactness-centric pattern of urban areas, i.e., the compactness of urban buildings decreases as you go from the city center to the suburbs. If the obvious outliers cover a comparative large area, a polygon with the correct label has to be added. If an abnormal compactness pattern appears, the validation requires a detailed inspection, which often leads to adding polygons or correcting the labels of existing polygons. We found that visual validation already give us a significant indication of label quality.

4) Label post-processing: After obtaining the labeled LCZ polygons, we discovered the following post-processing procedures were necessary:

- Polygon shrinking
  Although all the polygons were correctly labeled, some polygons in given LCZ class were drawn in a close proximity to another LCZ class. This might cause erroneous labels on the pixels close to the borders of the polygon when the polygon is rasterized, especially when using a large ground sampling distance (GSD). For example, the GSD of a LCZ label map defined in our research is 100 meters. A pixel in the label map that is too close to the boundary of two LCZs may cover both LCZ classes. To avoid this, shrinking the polygon of all non-urban LCZ classes except water (i.e., A to F) by 160m was carried out. We chose a distance of 160m because this corresponds to half of the patch size (16 pixels) of the Sentinel-1 and Sentinel-2 image patches in the So2Sat LCZ42 dataset. For class G (water), the shrinking distance is only 10m, given that the width of many rivers is in the order of hundreds of meters.
- Class balancing
  To use those vector-format polygon labels in machine learning of Earth observation images, they need to be rasterized into image format in certain geographic coordinate systems. We used geotiff and local UTM coordinates. However, the polygons of the non-urban LCZ classes (i.e., classes A to G) tend to be much larger in area than those of the urban classes, because the percentage of nonurban areas are naturally larger, and they are certainly much easier for humans to label. This results in many more pixels (samples) for nonurban classes. In order to balance the number of samples among all the LCZ classes, for each city, we reduced the number of samples of each of classes A to G to N_m, where N_m is the maximum number of samples from the urban classes (i.e., classes 1 to 10). If the number of

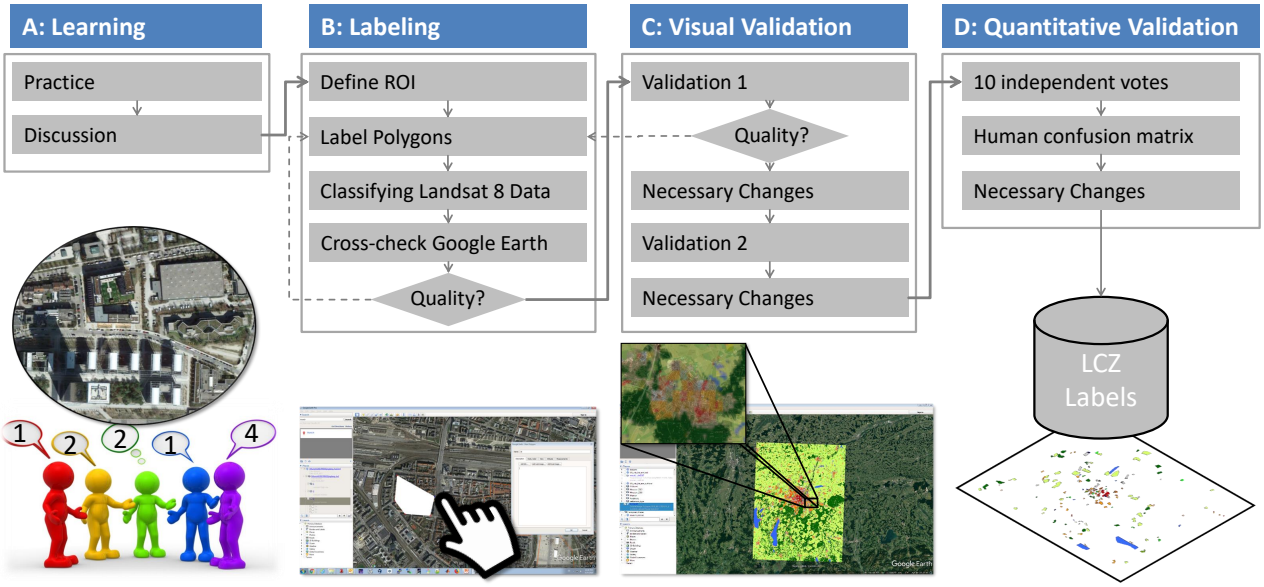


Fig. 4: Flowchart of four-phase labeling project. Block A: Learning phase; Block B: Labeling phase; Block C: First validation phase; Block D: Second validation phase.

samples of certain classes was less than N_m, those classes remained untouched. The samples of the urban classes were not reduced, because they are difficult to label. To this end, we were able to balance the different LCZ classes.

5) Quantitative quality control and validation phase: It is known that the maximum accuracy achievable by any supervised learning procedure depends not only on the chosen algorithm, but also on the quality of the training data. Therefore, we conducted quantitative evaluation on 10 European cities in the dataset by having a group of remote sensing experts cast 10 independent votes on each labeled polygons, in order to assess the human labeling accuracy, and identify possible remaining errors. Despite the huge labor cost, we believe this is essential for Earth observation, because it is important to establish an error bar for any product in Earth observation, which is unique and distinctly different from the computer vision community. This label evaluation procedure will be discussed in detail in section III.

B. Preparing the Sentinel-1 data

The Sentinel-1 mission provides an open access global SAR dataset. We accessed the Sentinel-1 VV-VH dual-Pol single look complex (SLC) Level-1 data via the Copernicus Open Access Hub (https://scihub.copernicus.eu/) using an automatic script developed by the authors based on SentinelSat (https://github.com/sentinelsat/sentinelsat).

A series of preprocessing steps were applied to the Level-1 Sentinel-1 dual-Pol data by the graph processing tool in the ESA SNAP toolbox. The detailed configurations of the preprocessing are listed as follows.

- *Apply orbit profile*: This module downloads the latest released orbit profile so that a precisely geocoded product can be achieved.
- *Radiometric calibration*: Radiometric computes the backscatter intensity using sensor calibration parameters in the metadata. The output is set to complex-valued image, in order to preserve the relative phase between VV and VH channels.
- *TOPSAR deburst*: For each polarization channel, the Sentinel-1 IW product has three swaths. Each swath image consists of a series of bursts. The TOPSAR deburst merges all these bursts and swaths into a single SLC image.
- *Polarimetric speckle reduction*: Speckle reduction was conducted by using the SNAP-integrated refined Lee filter. An unfiltered version is also included in the dataset.
- *Terrain correction*: Terrain correction eliminates the distortion introduced by topographical variations. To accomplish the correction, the SRTM was used as the DEM to provide height information. The data was re-sampled to a 10m GSD by the nearest-neighbor interpolation. The data was geocoded into the WGS84/UTM coordinate system of the corresponding city with a GSD of 10m.

To summarize, the Sentinel-1 data in the So2Sat LCZ42 dataset contain the following 8 real-valued bands:

1) the real part of the unfiltered VH polarization channel,
2) the imaginary part of the unfiltered VH polarization channel,
3) the real part of the unfiltered VV polarization channel,
4) the imaginary part of the unfiltered VV polarization channel,
5) the intensity of the refined LEE filtered VH polarization channel,
6) the intensity of the refined LEE filtered VV polarization channel,
7) the real part of the refined LEE filtered covariance matrix off-diagonal element, and
8) the imaginary part of the refined LEE filtered covariance

matrix off-diagonal element.

C. Preparing the Sentinel-2 data

The other satellite images employed in the So2Sat LCZ42 dataset are globally available Sentinel-2 multispectral images.

Google Earth Engine (GEE), a web- and cloud-based platform for large-scale scientific analysis and visualization of geospatial data, was used to create the cloud-free Sentinel-2 images [43]. The overall workflow, based on the GEE Python API, consisted of three main steps.

- The *querying step* for loading Sentinel-2 images from the catalogue,
- The *scoring step* for the calculation of a cloud related-quality score of each loaded image, and
- The *mosaicing step* for mosaicing the selected images based on the meta-information generated in the preceding modules.

More details can be found in [44].

Sentinel-2 images contain bands B2, B3, B4, B8 with 10m GSD, bands B5, B6, B7, B8a, B11, B12 with 20m GSD, and bands B1, B9, B10 with 60m GSD. In the So2Sat LCZ42 dataset, the 20m bands were upsampled to 10m GSD, and the bands B1, B9, and B10 were discarded because they mostly contain data related to the atmosphere and thus bear little relevance to LCZ classification. To summarize, the Sentinel-2 data in the So2Sat LCZ42 dataset contain the following 10 real-valued bands:

1) Band B2, 10m GSD
2) Band B3, 10m GSD
3) Band B4, 10m GSD
4) Band B5, upsampled to 10m from 20m GSD
5) Band B6, upsampled to 10m from 20m GSD
6) Band B7, upsampled to 10m from 20m GSD
7) Band B8, 10m GSD
8) Band B8a, upsampled to 10m from 20m GSD
9) Band B11, upsampled to 10m from 20m GSD
10) and Band B12, upsampled to 10m from 20m GSD

D. Content of the So2Sat LCZ42 dataset

By projecting the labels to the coregistered Sentinel-1 and Sentinel-2 images, we can extract Sentinel-1 and Sentinel-2 image patch pairs with corresponding LCZ labels. We define the dimension of the image patches in the So2Sat LCZ42 dataset as 32 by 32 pixels, which corresponds to a physical dimension of 320m by 320m. In order to create non-overlapping patches, we sampled the labeled polygons with a 320m by 320m grid, where the grid nodes are the center of each image patch. We obtained 400,673 pairs of Sentinel image patches. The whole dataset is about 56GB.

For machine learning purposes, the dataset was split into a training set, a testing set, and a validation set. They consist of 352,366, 24,188, and 24,119 pairs of image patches, respectively. The training set comprises all the image patches of 32 cities plus the 10 add-on areas in the city list (please see Appendix B for the full list of cities). The remaining 10 cities are distributed across all the continents and culture regions over the world. For each of them, we split the labels of each LCZ class into the west and east halves of each city, to form the testing and validation sets, respectively. Therefore, all three sub-datasets are geographically separated from each other, despite having drawn the testing and validation sets from the same list of cities.

TABLE II: 10 European cities selected for the evaluation.

City	Country
Amsterdam	The Netherlands
Berlin	Germany
Cologne	Germany
London	United Kingdom
Madrid	Spain
Milan	Italy
Munich	Germany
Paris	France
Rome	Italy
Zurich	Switzerland

III. Label Evaluation

It is well known that the maximum accuracy achievable by any supervised learning procedure depends not only on the chosen algorithm, but also on the quality of the training data [45]. In the context of the *HUMINEX* experiment, Bechtel et al. [40] have recently shown the difficulties associated with the assignment of LCZ classes by human experts. Therefore, evaluating the labels as a result of human expert knowledge is of vital importance for further use of the dataset in the training of classification algorithms for large-scale automatic LCZ mapping.

A. The Evaluation Set

For the evaluation, we have chosen a subset of 10 European cities (shown in Table II) from the group of cities we labeled. The choice was based on three rationales:

1) All our labeling experts have lived in Europe for a significant number of years. This ensures familiarity with the general morphological appearance of European cities.
2) Google Earth provides detailed 3D models for the 10 cities, which is of great help in determining the approximate height of urban objects. This is necessary to be able to distinguish between low-rise, mid-rise, and high-rise classes.
3) As previously mentioned, LCZ labeling is very labor-intensive. Reducing the evaluation set to 10 cities allowed us to generate more individual votes per polygon for better statistics.

Unfortunately, not many European cities contain LCZ class 7 (light-weight low-rise), which mostly describes informal settlements (e.g., slums). Therefore, we included the polygons of class 7 for an additional 9 cities that are representative of the 9 major non-European geographical regions of the world (see Table III).

B. Evaluation Strategy and Results

For the evaluation experiment, 10 remote sensing experts (hereafter referred to as the label validation crew), who were

TABLE III: Additional 9 cities whose polygons of class 7 (light-weight low-rise) were used for the evaluation.

City	Geographic Region
Guangzhou, China	East Asia
Islamabad, Pakistan	Middle East
Jakarta, Indonesia	South-East Asia
Los Angeles, USA	North America
Melbourne, Australia	Oceania
Moscow, Russia	Eastern Europe
Mumbai, India	Indian Subcontinent
Nairobi, Kenya	Sub Saharan Africa
Rio de Janeiro, Brazil	Latin America

already trained in applying the LCZ scheme to annotate urban areas, were provided with `.kml`-files containing the polygons of the original So2Sat LCZ42 dataset, but without labels. They were then asked to reassign an LCZ class to every polygon, using Google Earth as the labeling environment. After all the relabeled `.kml`-files were submitted, both a polygon-wise and a pixel-wise evaluation between the original labels and the votes newly cast by the label validation crew was carried out in the form of confusion matrices, which combine the validation results of the 10 European validation cities (cf. Table II) and the slum areas of the additional 9 non-European validation cities (cf. Table III). These confusion matrices are displayed in Fig. 5.

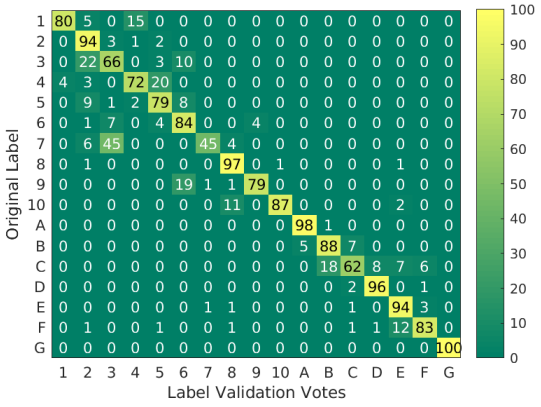
In addition, majority voting was carried out for each polygon, i.e., each polygon was reassigned to the class for which a majority of the label validation crew had voted, although we kept the original label in case there was a draw between this original class and another major class. The polygon-wise and pixel-wise confusion matrices between these final labels and the votes of the label validation crew can be seen in Fig. 6.

C. Interpretation of the Evaluation Results

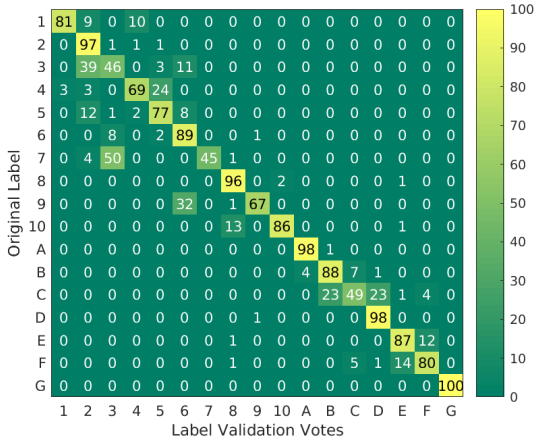
The confusion matrices in Figs. 5 and 6 show that:

- There is no significant difference between the polygon-wise and the pixel-wise results, which indicates that the polygons are evenly distributed with respect to size.
- The majority voting step helped to slightly improve the label confidences: Before the refinement, 11 of the 17 LCZ classes provided a confidence of more than 80%; after the refinement, this confidence level held for 13 classes.
- In general, confusion among the urban classes is slightly higher than among the non-urban classes.
- The most confident classes are 8 (large low-rise), A (dense trees), D (low plants), and G (water), with classes 2 (compact mid-rise) and E (bare rock/paved) following close behind.
- The least confident classes are classes 3 (compact low-rise), 7 (lightweight low-rise), and C (bush, scrub), with classes 4 (open high-rise) and 9 (sparsely built) following behind. The main sources of confusion for these classes are summarized in Table IV.

These experiences go hand-in-hand with the findings of [40], who also found that LCZ classes A (dense trees), D (low plants), G (water), 2 (compact mid-rise), 6 (open low-rise),



(a)



(b)

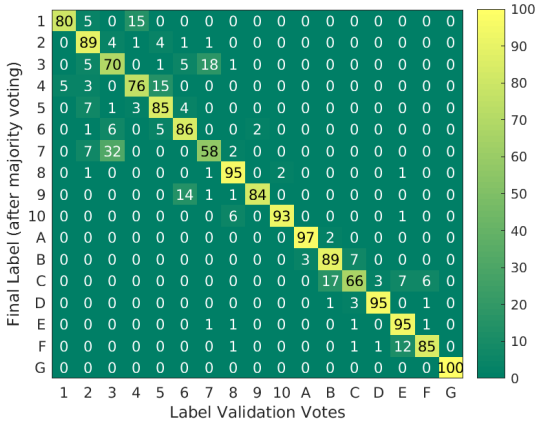
Fig. 5: Confusion matrices (values in %) of the original labels vs. votes cast by the label validation crew for the polygons of the evaluation cities selected in Tables II and III: (a) polygon-wise assessment, (b) pixel-wise assessment.

TABLE IV: Main sources of confusion for the less confident LCZ classes.

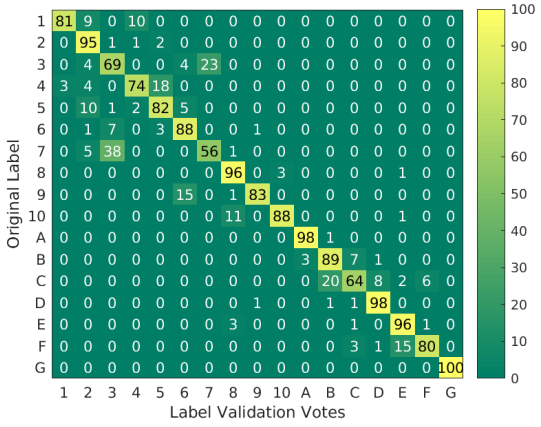
Low confidence class	Major confusion classes
3 (compact low-rise)	2 (compact mid-rise), and 6 (open low-rise)
4 (open high-rise)	5 (open mid-rise)
7 (lightweight low-rise)	3 (compact low-rise)
9 (sparsely built)	6 (open low-rise)
C (bush, scrub)	B (scattered trees), and D (low plants)

and 8 (large low-rise) were recognized consistently well by all operators, while classes 9 (sparsely built) and B (scattered trees) were reported as difficult to classify. Classes 1 (compact high-rise), 4 (open high-rise), 7 (lightweight low-rise), and C (bush, scrub) were not present in most of their study cities and thus not discussed in detail.

Looking at the major sources of confusion as summarized in Table IV, all these confusions appear fairly reasonable: Apparently, it is difficult even for human experts to distinguish the vaguely defined characteristics *open* and *compact*, as well as *mid-rise* and *high-rise*. In addition, sparsely built environ-



(a)



(b)

Fig. 6: Confusion matrices (values in %) of the final labels (refined by majority voting) vs. votes cast by the label validation crew for the polygons of the evaluation cities selected in Tables II and III: (a) polygon-wise assessment, (b) pixel-wise assessment.

ments are understandably frequently confused with open low-rise neighborhoods, as is bush/scrubland with scattered trees and low plants.

Given the accordance with the findings of [40], the semantic subtleties of the LCZ classification scheme, and a mean class confidence of about 80% before refinement by majority voting and 85% after refinement, the So2Sat LCZ42 dataset can be considered a reliable source of labels for the training of machine learning procedures aiming at automated LCZ mapping at a larger scale.

IV. Baseline Methods

In order to provide a baseline for the achievable LCZ classification accuracy, we performed classification on the So2Sat LCZ42 dataset using popular classifiers, including the classical random forests (RF), support vector machines (SVM) [23], and an attention-based ResNeXt as proposed in [46] and [47]. The employed RF consists of 200 trees, and the max_depth is set to 10, with the other parameters set to the default. A radial basis function kernel is chosen for SVM in the experiment. The depth of the ResNeXt is 29 and the Convolutional Block Attention Module is plugged into each of the residual blocks. For RF and SVM, the pixel values of the patches are converted into vectors, using the statistical measures (maximum, minimum, standard deviations and mean) of each band. All the classifiers are trained using the training set and tested on the validation set.

The resulting accuracy based on the Sentinel-2 images in the So2Sat LCZ42 dataset can be seen in Table V. The accuracy measures include overall accuracy (OA), averaged accuracy (AA), and kappa coefficient. In addition, weighted accuracy (WA) introduced in [40] is also considered, because it gives user-defined weights to confusions between different classes. For example, misclassifying compact high-rise as compact middle-rise is less critical than the confusion between compact high-rise and water, and should thus be penalized less.

TABLE V: Classification accuracy from three baseline methods, with the Sentinel-2 images in the proposed dataset.

	OA	WA	AA	Kappa
RF	0.51	0.87	0.31	0.46
SVM	0.54	0.88	0.36	0.49
ResNeXt-CBAM	0.61	0.92	0.51	0.58

V. Discussion

The goal of this paper is to provide documentation about a large benchmark dataset for local climate zone classification from Sentinel-1 and Sentinel-2 satellite data. Since the Sentinel data is openly available for the whole globe, the main intention of the dataset is to enable the training of models that generalize to any unseen areas across the world. This is ensured by sampling the data from altogether 52 cities located on all inhabited continents. In spite of these promising characteristics, two major challenges have to be noted:

1) *Local climate zones are sometimes hard to distinguish* As the label validation results shown in Section III illustrate, it is extremely hard to distinguish some of the local climate zone classes, even if human experts investigate several data sources (such as high-resolution optical imagery and 3D building models as available in Google Earth). This holds especially for the distinction of different height levels in compact areas, but also for open areas, which comprise both open land / vegetation and building structures. This has to be acknowledged as a natural limitation when tackling local climate zone mapping with remote sensing data and can possibly only be solved by combining remote sensing data with other data sources, e.g. information from social media data.
2) *Learning a generic LCZ prediction model is challenging* As described in Section II-D, the test set and the training set are completely disjunct, with the test cities being distributed across the ten major cultural regions of the inhabited world. Therefore, results achieved on this dataset can be considered a good measure of how well the trained model will generalize to completely unseen data. In this regard, overall accuracies between 50% and 60% can

already be considered promising – especially for a target scheme comprised of 17 difficult-to-distinguish classes. Nevertheless, there is still room for improvement, as usually an accuracy of at least about 85% to 90$ is required for land cover mapping purposes according to [48].

We hope that the community is eager to tackle those challenges and puts the So2Sat LCZ42 dataset to good use in order to achieve significant progress in the global mapping of cities into local climate zones.

VI. Conclusion and Outlook

This paper introduces a unique dataset that contains manually labeled LCZs reference data, as well as coregistered Sentinel-1 and Sentinel-2 image patch pairs over 42 cities plus 10 smaller areas across the six inhabited continents on this planet. The paper describes the carefully designed labeling process and a rigorous evaluation procedure that ensures the quality of the dataset. Despite the fact that each LCZ class is quantitatively defined in the original paper, we discovered that several LCZ classes can be easily confused with each other, because the height and percentage of pervious surface of these classes cannot be easily distuiguished by the human eye from aerial images during labeling. This renders the whole labeling process highly labor-intensive. Still, we were able to achieve an average class confidence of 85% through our human evaluation procedure with independent voting by 10 experts. Hence, this dataset is a reliable source for the training of machine learning procedures, and can be considered a challenging and large-scale data fusion and classification benchmark dataset for cutting-edge machine learning methodological developments. Examples for possible research directions include:

- Since we have provided the label confusion matrix, the question of how to introduce such prior knowledge into machine learning, deep learning models in particular, is an interesting direction;
- Due to culture-induced diversity existing in the data, transferablity of the models will be a key to achieving good classification results on a global scale;
- Radar and optical data possess completely different yet partially complementary characteristics. Developing methods to fuse them in an optimal way or select appropriate features from such diverse data sources is of general interest to the remote sensing community;
- Thanks to the large scale of the proposed benchmark data set, it can serve as a test bed for the development of efficient training techniques.

Our vision in the near future is to produce the world's first global LCZ classification map using multi-sensory remote sensing images, which will be made available to the community. This will strongly boost urban geographic research and help us develop a better understanding of global urbanization. As an example, Fig. 8 shows the LOD1 building model of Munich, Germany overlaid on the LCZ classification map. The left subfigure is the center of Munich, which is mostly shown as compact mid-rise class. The right subfigure is mostly an open mid-rise area next to the city center, which mostly consists of middle density residential blocks. As we can see, the height and density of the buildings in city center are significantly greater than those in the open mid-rise area. Such matching of LCZs and building models seems trivial in developed countries, because precise 3D building models are readily available. However, for developing areas, a 3D model cannot always be taken for granted. For example, the city of Lagos, Nigeria (population 21 million) does not have a quality 3D city model. Therefore, a quality LCZ classification map will become the firsthand information of urban building volume and distribution. For this purpose, we invite everybody to contribute by using this dataset and developing new, sophisticated algorithms.

Appendix A
Decision rule of the LCZ labeling

Please see Fig. 7.

Appendix B
City list of the So2Sat LCZ42 dataset

Training: Amsterdam, Beijing, Berlin, Bogota (addon), Buenos Aires (addon), Cairo, Cape Town, Caracas (addon), Changsha, Chicago (addon), Cologne, Dhaka (addon), Dongying, Hong Kong, Islamabad, Istanbul, Karachi (addon), Kyoto, Lima (addon), Lisbon, London, Los Angeles, Madrid, Manila (addon), Melbourne, Milan, Nanjing, New York, Paris, Philadelphia (addon), Qingdao, Rio De Janeiro, Rome, Salvador (addon), Sao Paulo, Shanghai, Shenzhen, Tokyo, Vancouver, Washington D.C., Wuhan, Zurich

Testing and validation: Guangzhou, Jakarta, Moscow, Mumbai, Munich, Nairobi, San Francisco, Santiago de Chile, Sydney, Tehran

Acknowledgment

This research was funded by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program with the grant number ERC-2016-StG-714087 (Acronym: So2Sat, project website: www.so2sat.eu), and the Helmholtz Association under the framework of the Young Investigators Group Signal Processing in Earth Observation (SiPEO) with the grant number VH-NG-1018 (project website: www.sipeo.bgu.tum.de).

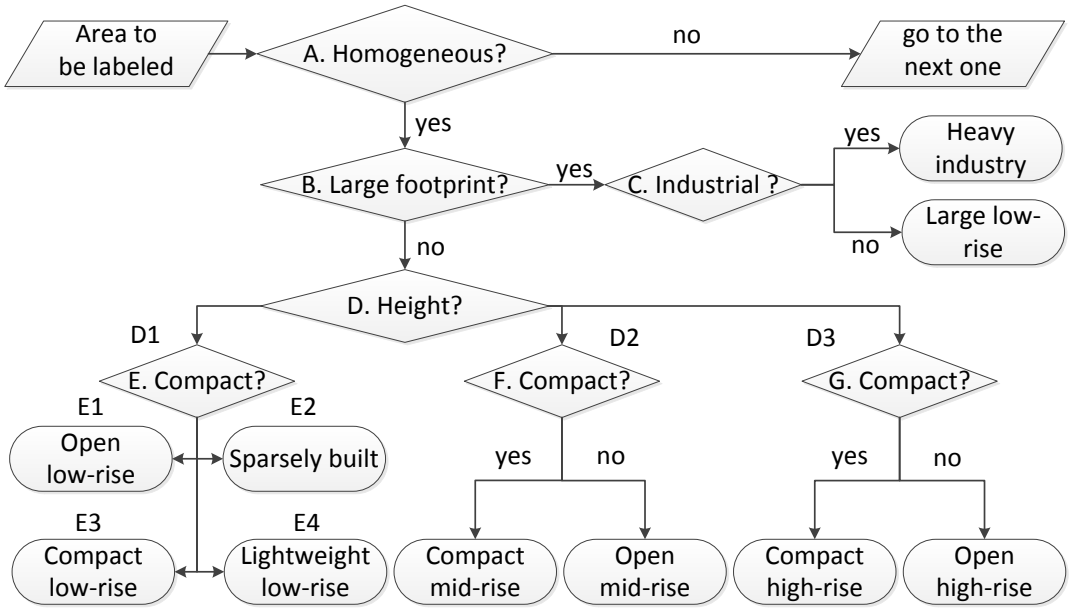


Fig. 7: Flowchart of the labeling decision rule, which labels one scene with seven decisions. They are: A: Is it homogeneous for at least five pixels of 100-by-100 meters? B: Is the building footprint large? C: Does any obvious industrial feature exist (such as oil tanks, cranes, or conveyor belts)? D1: Buildings with up to three floors; D2: Buildings with three to ten floors; D3: Buildings with ten floors and higher; E1: Building surface fraction between 20% and 40%; E2: Building surface smaller than 20%; E3: Building surface fraction between 40% and 70%; E4: Light material built with surface fraction larger than 60%; F: Is building surface fraction larger than 40%?; G: Is building surface fraction larger than 40%? The percentage is estimated by experts with a 100-by-100-meter polygon drawn on Google Earth. The building height is decided by experts using any available information, such as a 3D model, street view, or photo.

REFERENCES

[1] H. Taubenböck, T. Esch, A. Felbier, M. Wiesner, A. Roth, and S. Dech, "Monitoring urbanization in mega cities from space," *Remote sensing of Environment*, vol. 117, pp. 162–176, 2012.

[2] United Nations, *World Urbanization Prospects: 2018 Revision*. United Nation, 2018.

[3] T. Esch, H. Taubenböck, A. Roth, W. Heldens, A. Felbier, M. Schmidt, A. A. Mueller, M. Thiel, and S. W. Dech, "TanDEM-X mission–new perspectives for the inventory and monitoring of global settlement patterns," *Journal of Applied Remote Sensing*, vol. 6, no. 1, p. 061702, 2012.

[4] T. Esch, M. Marconcini, A. Felbier, A. Roth, W. Heldens, M. Huber, M. Schwinger, H. Taubenböck, A. Müller, and S. Dech, "Urban footprint processor–Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 6, pp. 1617–1621, 2013.

[5] M. Pesaresi, G. Huadong, X. Blaes, D. Ehrlich, S. Ferri, L. Gueguen, M. Halkia, M. Kauffmann, T. Kemper, L. Lu, *et al.*, "A global human settlement layer from optical HR/VHR RS data: Concept and first results," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 5, pp. 2102–2131, 2013.

[6] Y. Ban, P. Gong, and C. Gini, "Global land cover mapping using earth observation satellite data: Recent progresses and challenges," *ISPRS journal of photogrammetry and remote sensing (Print)*, vol. 103, no. 1, pp. 1–6, 2015.

[7] I. D. Stewart, "Local climate zones: Origins, development, and application to urban heat island studies," *Paper presented at the Annual Meeting of the American Association of Geographers*, 2011.

[8] B. Bechtel, P. J. Alexander, J. Böhner, J. Ching, O. Conrad, J. Feddema, G. Mills, L. See, and I. Stewart, "Mapping local climate zones for a worldwide database of the form and function of cities," *ISPRS International Journal of Geo-Information*, vol. 4, no. 1, pp. 199–219, 2015.

[9] I. D. Stewart and T. R. Oke, "Local climate zones for urban temperature studies," *Bulletin of the American Meteorological Society*, vol. 93, no. 12, pp. 1879–1900, 2012.

[10] I. D. Stewart, T. R. Oke, and E. S. Krayenhoff, "Evaluation of the 'local climate zone' scheme using temperature observations and model simulations," *International Journal of Climatology*, vol. 34, no. 4, pp. 1062–1080, 2014.

[11] D. Fenner, F. Meier, B. Bechtel, M. Otto, and D. Scherer, "Intra and inter local climate zone variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany," *Meteorologische Zeitschrift*, vol. 26, pp. 525–547, 2017.

[12] S. J. Quan, F. Dutt, E. Woodworth, Y. Yamagata, and P. P.-J. Yang, "Local climate zone mapping for energy resilience: A fine-grained and 3D approach," *Energy Procedia*, vol. 105, pp. 3777–3783, 2017.

[13] J. A. Quanz, S. Ulrich, D. Fenner, A. Holtmann, and J. Eimermacher, "Micro-scale variability of air temperature within a local climate zone in Berlin, Germany, during summer," *Climate*, vol. 6, no. 1, p. 5, 2018.

[14] R. Kotharkar and A. Bagade, "Evaluating urban heat island in the critical local climate zones of an Indian city," *Landscape and Urban Planning*, vol. 169, pp. 92–104, 2018.

[15] J. Geletič, M. Lehnert, S. Savić, and D. Milošević, "Inter-/intra-zonal seasonal variability of the surface urban heat island based on local climate zones in three central european cities," *Building and Environment*, vol. 156, pp. 21–32, 2019.

[16] C. B. Koc, P. Osmond, A. Peters, and M. Irger, "Understanding land surface temperature differences of local climate zones based on airborne remote sensing data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 8, pp. 2724–2730, 2018.

[17] C. B. Koc, P. Osmond, A. Peters, and M. Irger, "Mapping local climate zones for urban morphology classification based on airborne remote sensing data," in *2017 Joint Urban Remote Sensing Event (JURSE)*, pp. 1–4, IEEE, 2017.

[18] J. Geletič, M. Lehnert, P. Dobrovolnỳ, and M. Žuvela-Aloise, "Spatial modelling of summer climate indices based on local climate zones: expected changes in the future climate of brno, czech republic," *Climatic Change*, vol. 152, no. 3-4, pp. 487–502, 2019.

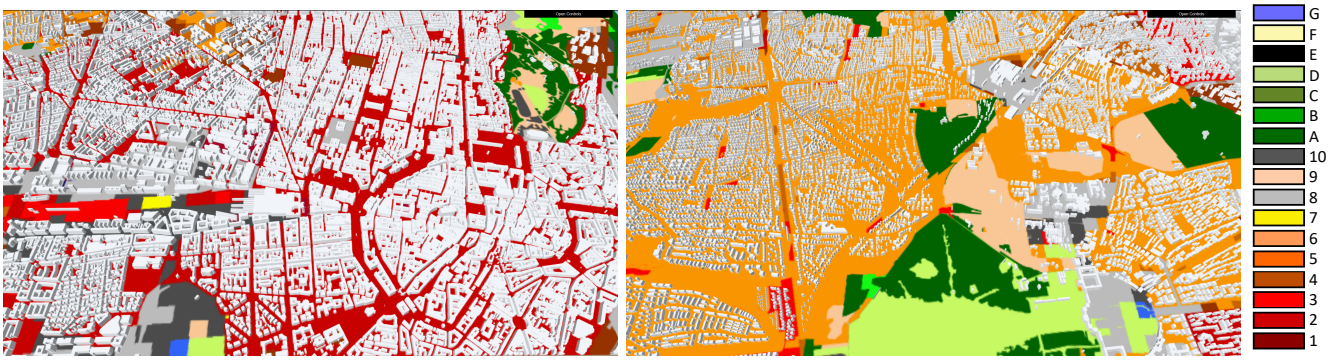


Fig. 8: LOD1 building model of Munich, Germany overlaid on the LCZ classification map. The left subfigure is Munich center that is mostly shown as compact mid-rise class (class 2). The right subfigure is mostly a open mid-rise (class 5) area next to the city center, which are mostly made of middle density residential blocks. The legend of LCZ classes is also shown in the figure. As we can see, the height and density of the buildings in city center are significantly higher than those in the open mid-rise area. Such matching of LCZs and building models seems trivial in developed countries, because precise 3D building models are readily available. However, a 3D model is not always available when comes to developing areas. For example, the city of Lagos, Nigeria with population of 21 million does not have a quality 3D city model. Therefore, a high quality LCZ classification map is the first hand information of urban building volume and distribution.

[19] A. Wicki and E. Parlow, "Attribution of local climate zones using a multitemporal land use/land cover classification scheme," *Journal of Applied Remote Sensing*, vol. 11, no. 2, p. 026001, 2017.

[20] H. C. Ho, K. K.-L. Lau, R. Yu, D. Wang, J. Woo, T. C. Y. Kwok, and E. Ng, "Spatial variability of geriatric depression risk in a high-density city: A data-driven socio-environmental vulnerability mapping approach," *International journal of environmental research and public health*, vol. 14, no. 9, p. 994, 2017.

[21] N. Yokoya, P. Ghamisi, J. Xia, S. Sukhanov, R. Heremans, I. Tankoyeu, B. Bechtel, B. L. Saux, G. Moser, and D. Tuia, "Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS data fusion contest," *IEEE J-STARS*, 2018, DOI: 10.1109/JSTARS.2018.2799698.

[22] "WUDAPT project website." http://www.wudapt.org/lcz/lcz-framework/. Accessed on: 2018-11-16.

[23] Y. Xu, C. Ren, M. Cai, N. Y. Y. Edward, and T. Wu, "Classification of local climate zones using ASTER and Landsat data for high-density cities," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 7, pp. 3397–3405, 2017.

[24] C. Qiu, M. Schmitt, P. Ghamisi, and X. X. Zhu, "Effect of the training set configuration on Sentinel-2-based urban local climate zone classification," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2018. in press.

[25] J. Hu, P. Ghamisi, and X. Zhu, "Feature extraction and selection of Sentinel-1 dual-pol data for global-scale local climate zone classification," *ISPRS International Journal of Geo-Information*, vol. 7, no. 9, p. 379, 2018.

[26] S. Sukhanov, I. Tankoyeu, J. Louradour, R. Heremans, D. Trofimova, and C. Debes, "Multilevel ensembling for local climate zones classification," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*, pp. 1201–1204, IEEE, 2017.

[27] N. Yokoya, P. Ghamisi, and J. Xia, "Multimodal, multitemporal, and multisource global data fusion for local climate zones classification based on ensemble learning," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*, pp. 1197–1200, IEEE, 2017.

[28] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[29] C. Qiu, M. Schmitt, L. Mou, and X. X. Zhu, "Urban local climate zone classification with a residual convolutional neural network and multi-seasonal Sentinel-2 images," in *10th IAPR Workshop on Pattern Recognition in Remote Sensing*, The International Association of Pattern Recognition (IAPR), 2018. in press.

[30] C. Qiu, M. Schmitt, P. Ghamisi, L. Mou, and X. X. Zhu, "Feature importance analysis of Sentinel-2 imagery for large-scale urban local climate zone classification," in *Geoscience and Remote Sensing Symposium (IGARSS), 2018 IEEE International*, IEEE, 2018. in press.

[31] C. Qiu, L. Mou, M. Schmitt, and X. X. Zhu, "LCZ-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network," *ISPRS J. Photogramm. Remote Sens.*, vol. 154, pp. 151–162, 2019.

[32] C. Yoo, D. Han, J. Im, and B. Bechtel, "Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using landsat images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 157, pp. 155–170, 2019.

[33] G. Xu, X. Zhu, N. Tapper, and B. Bechtel, "Urban climate zone classification using convolutional neural network and ground-level images," *Progress in Physical Geography: Earth and Environment*, p. 0309133319837711, 2019.

[34] Y. Fu, K. Liu, Z. Shen, J. Deng, M. Gan, X. Liu, D. Lu, and K. Wang, "Mapping impervious surfaces in town–rural transition belts using chinas gf-2 imagery and object-based deep cnns," *Remote Sensing*, vol. 11, no. 3, p. 280, 2019.

[35] G. Mills, J. Ching, L. See, B. Bechtel, and M. Foley, "An introduction to the WUDAPT project," in *Proceedings of the 9th International Conference on Urban Climate, Toulouse, France*, pp. 20–24, 2015.

[36] B. Bechtel, M. Foley, G. Mills, J. Ching, L. See, P. Alexander, M. OConnor, T. Albuquerque, M. de Fatima Andrade, M. Brovelli, *et al.*, "CENSUS of cities: LCZ classification of cities (Level 0)–Workflow and initial results from various cities," in *Proceedings of the 9th International Conference on Urban Climate, Toulouse, France*, 2015.

[37] J. Hidalgo, G. Dumas, V. Masson, G. Petit, B. Bechtel, E. Bocher, M. Foley, R. Schoetter, and G. Mills, "Comparison between local climate zones maps derived from administrative datasets and satellite observations," *Urban Climate*, vol. 27, pp. 64–89, 2019.

[38] B. Bechtel, P. J. Alexander, C. Beck, J. Böhner, O. Brousse, J. Ching, M. Demuzere, C. Fonte, T. Gál, J. Hidalgo, *et al.*, "Generating wudapt level 0 data–current status of production and evaluation," *Urban climate*, vol. 27, pp. 24–45, 2019.

[39] M. Demuzere, B. Bechtel, A. Middel, and G. Mills, "Mapping europe into local climate zones," *PloS one*, vol. 14, no. 4, p. e0214474, 2019.

[40] B. Bechtel, M. Demuzere, P. Sismanidis, D. Fenner, O. Brousse, C. Beck, F. Van Coillie, O. Conrad, I. Keramitsoglou, A. Middel, *et al.*, "Quality of crowdsourced data on urban morphology–The human influence experiment (HUMINEX)," *Urban Science*, vol. 1, no. 2, p. 15, 2017.

[41] C. Ren, R. Wang, M. Cai, Y. Xu, Y. Zheng, and E. Ng, "The accuracy of lcz maps generated by the world urban database and access portal tools (WUDAPT) method: A case study of Hong Kong," in *4th Int. Conf. Countermeasure Urban Heat Islands, Singapore*, 2016.

[42] M. Demuzere, B. Bechtel, and G. Mills, "Global transferability of local climate zone models," *Urban Climate*, vol. 27, pp. 46–63, 2019.

[43] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, vol. 202, pp. 18–27, 2017.
[44] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "Aggregating cloud-free Sentinel-2 images with google earth engine," 2019. to appear.
[45] C. E. Brodley and M. A. Friedl, "Identifying mislabeled training data," *Journal of artificial intelligence research*, vol. 11, pp. 131–167, 1999.
[46] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1492–1500, 2017.
[47] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "Cbam: Convolutional block attention module," in *Proc. European Conference on Computer Vision*, pp. 3–19, 2018.
[48] J. R. Anderson, "Land-use classification schemes – used in selected recent geographic applications of remote sensing," *Photogrammetric Engineering*, vol. 37, no. 4, pp. 379–387, 1971.

Appendices

E Hu J, Wang Y, and Zhu X X. (2019) Analyzing the Sustainability of Urban Land Consumption using Morphological Information of Local Climate Zone Scheme. (submitted to Remote Sensing of Environment)

LAND CONSUMPTION AND POPULATION DENSITY: ANALYZING URBAN SUSTAINABILITY USING SENTINEL SATELLITE IMAGERY

A PREPRINT

Jingliang Hu
Signal Processing in Earth Observation (SiPEO)
Technical University of Munich (TUM)
80333 Munich, Germany
jingliang.hu@tum.de

Yuanyuan Wang
Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
82234 Wessling, Germany
Signal Processing in Earth Observation (SiPEO)
Technical University of Munich (TUM)
80333 Munich, Germany
y.wang@tum.de

Xiao Xiang Zhu
Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
82234 Wessling, Germany
Signal Processing in Earth Observation (SiPEO)
Technical University of Munich (TUM)
80333 Munich, Germany
xiaoxiang.zhu@dlr.de

November 18, 2019

ABSTRACT

Urban sustainable development has an influence on the livelihood of 55% inhabitants on Earth, according to the United Nations' report of *'The World's Cities in 2018'*. The urban sustainable development is crucial enough to be listed in the Sustainable Development Goals (SDGs) announced by the United Nations General Assembly. One of its essential contents of the goals is to balance the trade-off of efficient land consumption and demands of current inhabitants. However, due to related countless variables, such as historical impact, cultural influence, economic status, and future plan, it is difficult to locate the balancing pivot, and it is even harder to have a standard ruler measuring all cities around the world in terms of sustainable development. In this study, we developed a methodology to produce morphological maps of cities by using Sentinel satellite images, and to categorize cities based on their urban morphological formation and population distributions. This methodology is applied to 42 cities selected around the world and resulted in eight clusters. Coupling with the population density, we analyzed the status of land consumption within each cluster where cities have similar urban structure and functionality. This paper concludes in a statistical perspective that, New York and Sao Paulo are at the balanced spot for the trade-off between land consumption and inhabitants demands for a group of cities that are super compact; London and Berlin are in the same position for openly built cities; Santiago de Chile is the city among compact cities; And lastly, Zurich is the one for cities of open and green formations. Additionally, the city Dongying is found to be severely under efficient in land consumption, while Mumbai, Hong Kong, and Cairo are heavily populated.

***K*eywords** Local climate zone (LCZ) · Population density · Sentinel-1 · Sentinel-2 · Sustainable Development Goals (SDGs) · Topology enhanced ensemble classification

1 Introduction

The sustainable development goals (SDGs) set by the United Nation General Assembly aims at achieving a prosper world in the year of 2030 [1]. The 17 goals of the SDG framework involve eliminating poverty, improving education quality, reduce inequality and so on. As reported in an UN survey [2], 55.3% of the world's population lived in urban area in the year of 2018, which is estimated to be 60% in the year of 2030. With more than half of the global population are urban inhabitants and the number is still increasing, the sustainable development in general wouldn't be successful without a dedicated plan for the development of the city. Therefore, the 11^{th} goal of the SDG framework, 'make cities and human settlements inclusive, safe, resilient and sustainable', aims at achieving a sustainable urban environment by creating job, providing housing, building resilient societies, investing public infrastructures, and extending green public space [1]. In order to accomplish these goals, necessary information of current urban status needs to be acquired in an unbiased manner. This is where remote sensing technique appears on the stage.

The remote sensing community has already been serving the urban sustainable development for decades. A tremendous number of studies have been published, such as urban local climate [3, 4, 5] and urban land monitoring [6, 7, 8, 9]. Besides numerous research groups, European Space Agency (ESA) has also put an huge effort on the urban sustainable development. It initiated an association called Earth observation for Sustainable Development (EO4SD)[1] which aims to utilize the remote sensing technique on benefiting sustainable development. The top-priority thematic of EO4SD is the urban sustainable development which is addressed by a group named as EO4SD-urban[2]. The EO4SD-urban has provided strategic spatial information on the status of the urban by a portfolio [10] of information services which consists of urban extent detection [11, 12, 13], land cover land use classification [14, 15], urban heat islands [16, 17], detecting land use changes [8, 18], and so on. As the SDGs has been announced as the strategic goal of the human kind by UN, the Group on Earth Observation (GEO)[3] initiated an organization named as Earth Observation for the Sustainable Development Goals (EO4SDG)[4] who aims to advance the 2030 Agenda and enable societal benefits of the SDGs by applying Earth observation data and geospatial information. Regarding technical development of the urban aspect of SDGs, according to an official documentation [19], EO4SDG focuses on monitoring the urban growth by using the product of Global Urban Footprint (GUF) [13, 20, 21]. Another major contributor of SDGs in remote sensing is Committee on Earth Observation Satellites (CEOS)[5]. One of its three Ad Hoc teams is the Ad Hoc Team on Sustainable Development Goals (AHT-SDG) who aims to support SDGs in the content of Earth observation. It is reported in their documentation [22] that, in terms of urban aspect, they detect land consumption by using the Global Human Settlement Layer (GHSL) [23, 24] and the World Settlement Footprint 2015 (WSF2015) [22, 21]. As the GHSL was a global settlement layer for the years of 1975,1990,2000,and 2015, it enabled temporal analysis of urbanization on the global scale [25], and it directly supported the SDGs goal 11.3.1 for 10,000 cities [26].

The SDGs goal 11.3.1 is described as: "ratio of land consumption rate to population growth rate". According to UN explanation [27], the indicator is designed to reveal the efficiency of land consumption. The indicator itself focuses on monitoring the efficiency of land consumption in terms of dynamic changing along time which answers the question, 'whether the change of land consumption is proportionate to the change of population?'. This changing is of course important. Besides measuring whether the changing quantity is reasonable, it is also important to answer, does the current land consumption of the whole city fall into a status of sustainability? Unfortunately, there are countless variables that make it too complex to quantify a city being sustainable or not. For example, should different standards applied to industrial cities and non-industrial cities? And the higher the population density, the more efficient of the land consumption, and the less well-being of the inhabitant. Where is the balance? How to quantify it? In this work, we attempted to find the cluster of city and locate the balance of land consumption and population density within each cluster in a statistical manner. In order to achieve our goal, we take the advantages of a recent classification scheme, local climate zone (LCZ) [28, 29, 30].

Being the available state-of-the-art global scale urban mappings, the GUF [13] and the GHSL [24] often appear as the remote sensing product that supports the urban SDGs. The GUF delineates the built-up area as a binary mask, and the GHSL characterizes the urban area as a four-category thematic map considering the density of built-ups and population [26]. Both products follow a relatively easy classification scheme which benefits describing cities from regions of transcultural, transnational, and cross-environmental in a consistent manner. Meanwhile, the detailed information within an urban area is missing from them, such as the functional type of a region in a city. The LCZ classification scheme also aims classify urban areas in a consistent manner across the global, yet attempts to detail urban areas into 17 classes [28]. Although the LCZ classification scheme was originally designed for urban climatic

[1] http://eo4sd.esa.int/

[2] http://www.eo4sd-urban.info/

[3] http://www.earthobservations.org/index.php

[4] http://eo4sdg.org/

[5] http://ceos.org/

studies [28, 29], it essentially classifies the urban area by morphological criteria, mainly including the land surface material, the compactness of built-ups, and the height of built-ups. Additionally, according to our experiences of annotating an half million LCZ labels [31] and related studies [32, 33], LCZ scheme is also able to map functional types of regions and to support applications as urban planning. Thus, comparing to the current state-of-the-art global urban mappings, the LCZ scheme could also theoretically suit the role of supporting urban SDGs with more detailed defined classes within a city extent. Therefore, we would like to investigate the potential of the LCZ scheme in supporting the urban SDGs in this paper.

Since the initiation of LCZ, researchers have focus on the development of classification algorithms. As first introduced in [30, 34], LCZ map is produced for individual city by annotating labels, preparing LandSat-8 multi-spectral imagery, training a random forest classifier, and predicting the LCZ map with the trained classifier. To explore the potential of different types of remote sensing data sources on the task, researches have carried out studies on varies optical data [35, 36, 37], digital surface model (DSM) [38, 39], Interferometric Synthetic Aperture Radar (InSAR) [39], geographical information system (GIS) data [5, 40, 41], and polarimetric synthetic aperture radar (PolSAR) data [42, 43]. To sum up, the optical data source is the common choice for its easy access and good performance. The other data sources are often servers as additional ones. The DSM provides height information which is crucial in the LCZ scheme and is complementary to the spectral information provided by the optical data, but it has limited accessibility. The GIS benefits the classification procedure, but also limited by the availability. The InSAR data is not perfect for this task because of its height information in urban area is still not precise enough. The PolSAR data benefits this task by providing spatial information and back scatterings of built-ups, and it even more available than the optical data for being independent of weather condition. After the comprehensive literature review on the data sources, the Sentinel-1 dualPol SAR data and the Sentinel-2 multi-spectral imagery were chosen as data sources in this paper, for their performance, accessibility, and global coverage.

Beside investigating data sources, in terms of methodology, the aforementioned studies are successful on producing LCZ map for a city by a classifier trained with data of that city. However, those trained models are not applicable to other cities [44], due to the challenge of transferability. To sum up the findings of studies that attempt to fill the gap of transferring [45, 46, 47, 43], it is still the most precise way to producing LCZ map for a city by a classifier trained with data of that city, if the training data is available. In order to demonstrate the potential of LCZ maps on applications, we followed the strategy of producing LCZ maps for selected 42 major cities where LCZ labels are available. We developed an automatic classification system in this paper, which fuses Sentinel-1 and Sentinel-2 data in a semi-supervision strategy [48, 49] and produces a classification map with a topological enhanced ensemble strategy. With the LCZ maps describe the morphological characteristics of these 42 cities, coupling a population dataset [50], we categorized those cities into different urban morphological clusters. Finally, we carried out a statistical analysis of the sustainable development of land consumption among those 42 cities.

In this study, we found out that the lightweight built-up area defined in LCZ scheme is over populated than the other defined classes, as long as it exists in a city. And the city of Dongying is severely under efficient in land consumption, while Mumbai, Hong Kong, and Cairo are heavily populated in general. It is also found in this study there are eight clusters among the 42 selected cities, which are different in morphology. The city that having a median population density within a cluster is treated as a statistical balanced spot of land consumption and well-being of inhabitants.

This paper describes the data sets and explains the proposed methodology in section 2, demonstrates the results and carries out a detailed discussion in section 3, and concludes the work in section 4.

2 Data and Method

2.1 Study Area

The study area includes 42 major cities that are listed in Table 1. They geographically distribute on every continents except Antarctica, as shown in Fig. 1. In order to carry out a comprehensive analysis of the urban morphology, these 42 cities are selected from transcultural, transnational, and cross-environmental regions. While selecting the 42 cities, the population is also considered. Each one of these cities is known as either highly populated or rapidly expanded. In total, according to UN statistic [51, 52], the 42 cities are homeland to about 400 million urban inhabitants, in the year of 2016.

2.2 Data

The Sentinel-1 data is the SAR data source used in this paper. A level-1 product, VV-VH dual-Pol single look complex (SLC) data, is freely downloaded via the Copernicus Open Access Hub (https://scihub.copernicus.eu) using a programme

Figure 1: The geographic locations of the 42 cities of interest.

developed by the author. Thereafter, the data is prepared as an analysis-ready data via a group of preprocessing using the ESA SNAP toolbox. The procedure is detailed as follows:

- Apply Orbit Profile: A latest released orbit metadata will be downloaded which includes accurate satellite position and velocity information. It leads to a precisely geocoded product.
- Radiometric Calibration: A radar backscatter of the scene surface is achieved from a digital number of the pixel. It reveals a quantitative measure of the surface.
- TOPSAR Deburst: A complete single look complex image is achieved by combining bursts and swaths.
- Polarimetric Speckle Reduction: A speckle-free dual-Pol SAR data is achieved by using the SNAP-integrated refined Lee filter.
- Terrain Correction: A analysis-ready data is achieved which has a ground sampling distance of 10 meter and is geocoded in the WGS84/UTM coordinate system. The topographical variation is eliminated with the help of the SRTM.

After the preprocessing, the feature extraction was carried out on the analysis-ready data. It is designed to derive features representing polarimetric information, local neighborhood information, and spatial information [15, 43]. The features related to polarimetric information are the intensity of VV, the intensity of VH, the coherence between VV and VH, and the intensity ratio of VV and VH. The intensity features are converted into dB. The local neighborhood information includes the mean and the standard deviation of an 11-by-11-pixel local area for each of the four polarimetric features. The spatial information is derived by the morphological profile [53, 54], with radius of 1, 2, and 3 pixels, for each of the four polarimetric features. The feature extraction results in 36 features in total.

The Sentinel-2 imagery was preprocessed by the Google Earth Engine (GEE) [55, 56] for eliminating clouds. Afterwards, the feature extraction is carried out on the Sentinel-2 analysis-ready data. The principal components of the data that holds 99% of the total spectrum information is derived to represent the whole dataset. The morphological profile is utilized to derive the spatial information on each of the principal component, with radius of 1, 2, and 3 pixels. Thus, the number of extracted S2 features varies from city to city depending on the numbers of principal components.

The population data used in this paper is a spatial raster dataset depicting the distribution of global population of the year 2015 at a resolution of 250 meters [50]. It was disaggregated from census or administrative units to grid cells, constrained to the distribution of built-up that mapped in the Global Human Settlement Layer (GHSL) global layer [57].

The LCZ ground reference data used in this paper is the benchmark data set, the So2Sat LCZ42 [31]. It was annotated by a group of 15 experts who followed an elaborately designed labeling protocol that maximizes the consistence, completeness, and correctness of labeling, meanwhile minimizes inevitable human error. For each of the 42 cities, the reference data is formatted as a raster image with a ground sampling distance of 100 meter. It is important to mention

Table 1: 42 cities of interest, with information on the regions and populations [51, 52].

Region	City	Population at Year		
		2000	2016	2030
Australia	Melbourne	3,461,000	4,258,000	5,071,000
	Sydney	4,052,000	4,540,000	5,301,000
Eastern Asia	Beijing	10,162,000	21,240,000	27,706,000
	Nanjing	6,160,000	8,270,000	9,750,000
	Wuhan	6,638,000	7,979,000	9,442,000
	Hong Kong	6,835,000	7,365,000	7,885,000
	Shanghai	13,959,000	24,484,000	30,751,000
	Changsha	2,182,000	3,882,000	5,013,000
	Dongying	628,000	967,000	1,278,000
	Guangzhou	7,330,000	13,070,000	17,574,000
	Kyoto	18,660,000	20,337,000	19,976,000
	Qingdao	2,940,000	4,686,000	5,920,000
	Shenzhen	6,550,000	10,828,000	12,673,000
	Tokyo	34,450,000	38,140,000	37,190,000
Southern Asia	Jakarta	8,390,000	10,483,000	13,812,000
	Mumbai	16,367,000	21,357,000	27,797,000
	Islamabad	597,000	1,433,000	2,275,000
Western Asia	Tehran	7,128,000	8,516,000	9,990,000
	Istanbul	8,744,000	14,365,000	16,694,000
Africa	Cairo	13,626,000	19,128,000	24,502,000
	Nairobi	2,214,000	4,070,000	7,140,000
	Cape Town	2,715,000	3,698,000	4,322,000
Europe	Amsterdam	1,005,000	1,099,000	1,213,000
	Berlin	3,384,000	3,578,000	3,658,000
	London	8,613,000	10,434,000	11,467,000
	Paris	9,737,000	10,925,000	11,803,000
	Zurich	1,078,000	1,259,000	1,494,000
	Milan	2,985,000	3,104,000	3,162,000
	Rome	3,385,000	3,738,000	3,842,000
	Lisbon	2,672,000	2,902,000	3,192,000
	Moscow	10,005,000	12,260,000	12,200,000
	Cologne	963,000	1,042,000	1,095,000
	Munich	1,202,000	1,454,000	1,548,000
	Madrid	5,014,000	6,264,000	6,707,000
North America	Washington DC	3,949,000	5,013,000	5,690,000
	Los Angeles	11,798,000	12,317,000	13,257,000
	San Francisco	3,230,000	3,299,000	3,615,000
	Vancouver	1,959,000	2,523,000	2,930,000
	New York	17,813,000	18,604,000	19,885,000
South America	Rio de Janeiro	11,307,000	12,981,000	14,174,000
	Santiago de Chile	5,658,000	6,544,000	7,122,000
	Sao Paulo	17,014,000	21,297,000	23,444,000

that, instead of the original 17 classes of LCZ, this paper aims to classify a modified version of LCZ with 11 classes. The 17 classes of original LCZ scheme were categorised by three elements, the covering material, the compactness of neighborhood, and the vertical height. Since the height element is not crucial in this study and the data used have no information on height, the modified version of LCZ eliminated the element of height by merging the compact high rise, compact middle rise, and compact low rise as compact buildup area, the open high rise, open middle rise, and open low rise as open buildup area. From our labeling experience, the large low rise class is very often warehouses locating in industry area so that we merged the large low rise with heavy industry as industry area. The sparsely buildup is ignored

because it frequently locates in scattered tree area or low plant area. Therefore, we came up with a modified version of LCZ with 11 classes. The main reason of modifying the LCZ scheme is that there is no access to a global scale data set that provides reliable and accurate height information in urban area.

2.3 Method

The proposed methodology includes two modules, a classification module, and a clustering module. The classification module is an automatic classification system consisting of a data fusion section and a topology enhanced ensemble classification section, which are detailed in subsection 2.3.1 and 2.3.2, respectively. The workflow of the automatic classification is demonstrated in Fig. 2. The clustering module categorizes the 42 cities into clusters of different urban morphological properties. It is detailed in subsection 2.3.3.

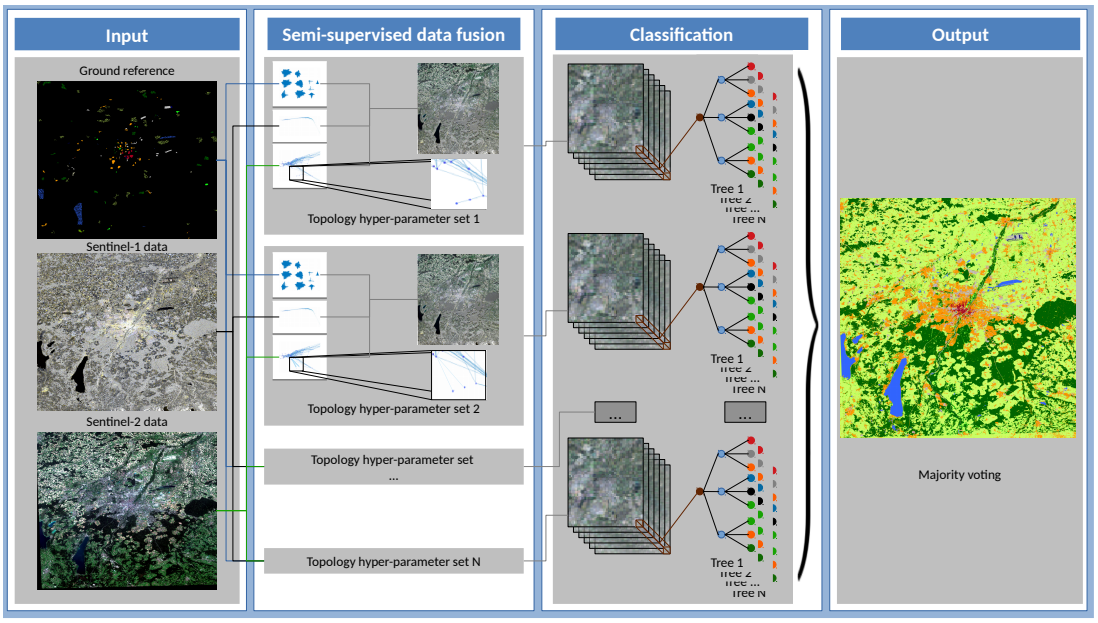


Figure 2: The workflow of the proposed automatic classification system.

2.3.1 Semi-supervised data fusion

This section explains the technical details of the data fusion algorithm. It is demonstrated as the grey block in the Semi-supervised data fusion column in Fig. 2. One might notice that there are multiple grey blocks in the column of the Fig. 2. All those blocks indicate the same data fusion procedure, except with different hyper-parameters for deriving topological structure. The reason of doing so will be detailed later in section 2.3.2.

The data fusion algorithm maps the multi-sensory data sources, the Sentinel-1 dual-Pol SAR data and the Sentinel-2 multispectral imagery, into a latent space where (I) the data of the same class locate close to each other, and (II) the topological structure of individual data source is preserved.

To precisely explain the methodology in the paper, the terminology has to be defined. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ represent the Sentinel-1 and Sentinel-2 data, respectively. $\mathbf{X}_i = [x_i^1, ..., x_i^p, ..., x_i^{n_i}] \in \mathbb{R}^{m_i \times n_i}$, where $i \in \{1, 2\}$, p indicates the p^{th} instance ($p \in \{1, 2, 3, ..., n_i\}$), n is the number of data instances, and m is the dimension of data. Let $\mathbf{X}_i^a \in \mathbb{R}^{m_i \times n_i^a}$ represents the annotated subset of $\mathbf{X}_i$ and $\mathbf{X}_i^{\bar{a}} \in \mathbb{R}^{m_i \times n_i^{\bar{a}}}$ represents the subset without annotation, so that $\mathbf{X}_i^{\bar{a}} \cup \mathbf{X}_i^a = \mathbf{X}_i$, $\mathbf{X}_i^{\bar{a}} \cap \mathbf{X}_i^a = \emptyset$, and $n_i^a + n_i^{\bar{a}} = n_i$. Let the annotation of instance $x_i^{a,p}$ indicated by $y_i^{a,p}$, thus, the annotated data set is $\{\mathbf{X}_i^a, \mathbf{Y}_i^a\}$, where $\mathbf{Y}_i^a \in \mathbb{R}^{1 \times n_i^a}$.

In order to find the latent space that carries the desired properties (I) and (II), two mathematical terms have to be defined. Firstly, a $(n_1 + n_2)$ by $(n_1 + n_2)$ similarity matrix $\mathbf{S}$ aims to present whether two data instances belong to the same class or not.

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{1,1} & \mathbf{S}_{1,2} \\ \mathbf{S}_{2,1} & \mathbf{S}_{2,2} \end{pmatrix} \tag{1}$$

where $\mathbf{S}_{i,j}^{p,q} = 1$, if x_1^p, the p^{th} instances of data $\mathbf{X}_1$, and x_2^q, the q^{th} instances of data $\mathbf{X}_2$, belong to the same class, otherwise, $s_{i,j}^{p,q} = 0$.

The second term, $\mathbf{T}$, is also a $(n_1 + n_2)$ by $(n_1 + n_2)$ binary matrix, but presents the topological structures of the two data sources.

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{2,2} \end{pmatrix} \tag{2}$$

where $\mathbf{T}_{1,1}$ and $\mathbf{T}_{2,2}$ represent the topological structures of the Sentinel-1 and Sentinel-2 data, respectively. If x_i^p and x_i^q are topologically connected, $t_{i,i}^{p,q} = 1$, otherwise, $t_{i,i}^{p,q} = 0$. The topological structure is the implementation of the semi-supervision concept. It builds up the connection of data with and without annotation via topology closeness. Conventionally, the topology closeness is practically calculated by k-nearest-neighbor [49, 58]. In this work, we applied a recent topological data analysis tool to derive the topology closeness, which has been proven to outperform the k-nearest-neighbor [15]. For more details, please refers to [59, 60, 61].

Let $\mathbf{f}_1^T$ and $\mathbf{f}_2^T$ stand for the two projections that respectively brighes the Sentinel-1 and Sentinel-2 data to the latent space. Therefore, minimizing the cost function (3) pursues the property (I), data of the same class locate close to each other in the latent space.

$$\mathcal{C}_s = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{p=1}^{n_1}\sum_{q=1}^{n_2} \|\mathbf{f}_i^T x_i^p - \mathbf{f}_j^T x_j^q\|^2 \mathbf{S}^{i,j}(p,q). \tag{3}$$

Minimizing the cost function (4) perserves the topological structures of the two data sources.

$$\mathcal{C}_t = \sum_{i=1}^{2}\sum_{p=1}^{n_1}\sum_{q=1}^{n_2} \|\mathbf{f}_i^T x_i^p - \mathbf{f}_i^T x_i^q\|^2 \mathbf{T}^{i,i}(p,q). \tag{4}$$

To minimize the two cost functions simutanensly, we only need to minimize $\mathcal{L} = \mathcal{C}_s + \mathcal{C}_t$, which turns into an optimization problem $\min_{\mathbf{f}_1,\mathbf{f}_2} \mathcal{L}$. Therefore, the solution to the optimization problem is the smallest non-zero eigenvectors of the generalized eigenvalue decomposition of (5).

$$\tilde{\mathbf{X}}(\mathbf{L}_s + \mathbf{L}_t)\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}(\mathbf{D}_s + \mathbf{D}_t)\tilde{\mathbf{X}}^T\mathbf{f}, \tag{5}$$

where $\tilde{\mathbf{X}} = \begin{pmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{pmatrix}$, $\mathbf{D}$ and $\mathbf{L}$ are the degree matrix and the Laplacian matrix, respectively. After achieving the projections $\mathbf{f}_1$ and $\mathbf{f}_2$, the Sentinel-1 and Sentinel-2 data are mapped into the latent space and are ready for the classification.

2.3.2 Topology enhanced ensemble classification

Random forest is chosen as the classifier in our automatic classification system, since it is a state-of-the-art, non-parametric, and time efficient classifier that is widely used in global scale classification tasks [43, 62] and local climate zone classification tasks [30, 45, 46].

To briefly recall random forest, the classifier predicts a label for an instance by counting predictions of a number of decision trees. In order to avoid decision trees, a deterministic classifier, producing identical predictions, random subset of features and random subset of training samples are utilized to train each of the decision trees. Therefore, every single decision tree considers slightly different information for the classification task and predict different labels for an instance. This configuration creates a diversity that empowers random forest the most. The diversity makes the random forest statistically follows the *Law of Large Numbers* so that it has robust performance, avoids overfitting that the decision tree suffers, and meanwhile preserves the computational efficiency of the decision tree [63, 64, 65].

Topology ensemble is a practical strategy that the authors utilize to tackle the issue of hyper-parameter optimization while deriving topological structure. The problem is that, no matter which method is chosen to obtain the topological structure of the Sentinel-1 and Sentinel-2 data, one would inevitably confront with the issue of hyper-parameter optimization. To handle this practical issue, a conventional method took an empirical value [49], and a recently algorithm [60] reported no statistically significant optimal solution was found. However, the authors found that varying values of the hyper-parameter leads to achieve topological structures with slight differences, as shown in the zoomed-in parts of Fig. 2. Inspired by the *Law of Large Numbers* concept in random forest, the solution in this paper is to fuse the two data multiple times with varying hyper-parameter. The variations of topological structures introduce an extra diversity on the aspect of topology.

As demonstrated in the classification module of Fig. 2, the topology enhanced ensemble classification includes diversity on three aspects, the derived different topological structure, the random subset of features, and the random subset of training samples.

2.3.3 Urban clustering module

The automatic classification module produces classification maps for the 42 cities, which categorizes the urban land consumption as compact built-up, open built-up, light weight built-up, industrial structure, or one of the seven natural types. In this urban clustering module, it first extracts the percentages of each land consumption type within a city and the percentages of population that distributes on each land consumption type as features. Then, the 42 cities are clustered by the classic K-mean algorithm. In the clustering, we have grouped all seven natural types of LCZ as one class, natural space.

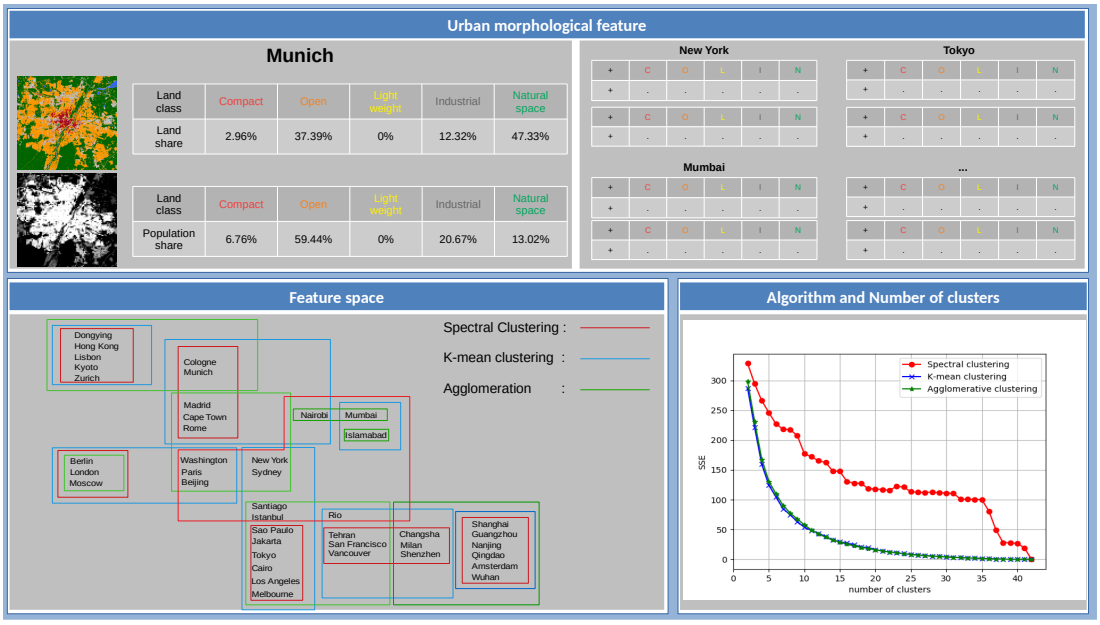


Figure 3: The workflow of the clustering module. Upper block: feature extraction based on the LCZ map and the population data; Bottom left: three clustering algorithms with number of clusters as input parameter are chosen as candidate clustering algorithms; Bottom right: the SSE is applied as the indicator for choosing algorithm and deciding the number of clusters.

Feature extraction. As shown in Fig. 3, we first calculate the percentages of the five urban land consumption types, including the compact built-up, the open built-up, the light weight built-up, the industrial structure, and the natural space. The water body was excluded here. These percentages describes how a city is formed in terms of morphological characteristics. Accordingly, we also extract the percentages of population that distributes on each of the five land consumption types. These population percentages describes how population distribution is related to the urban morphology. For the population data, we need to point out that, the current stat-of-the-art population data with spatial distribution is produced by census data and remote sensing human settlement layer. Population census data was evenly distributed over human settlement layer with constrain to administrative boundaries. It results a strong correlation between the amount of population and the area of built-up types. Nevertheless, there still some interesting phenomenon were found, which will be introduced in detail in Section 3.2 and 3.3 . Therefore, for each city, we extracted in total 10 features which are the input for the following clustering.

Clustering. To chose the clustering algorithm and the number of clusters, the error sum of squares (SSE) was applied as a quantitative indicator. $\mathbf{SSE}$ is the sum of squared error between each data record and the mean center of its cluster, $\mathbf{SSE} = \sum_{i=1}^{n}(\mathbf{C}_i - \tilde{\mathbf{C}})^2$, where n is the number of data records, $\mathbf{C}_i$ is a 10 by 1 vector representing the feature of the i^{th} city, $\tilde{\mathbf{C}}$ is the mean center of the cluster that $\mathbf{C}_i$ belongs [66]. When the number of clusters is fixed, the smaller value of SSE, the more concentrate of these clusters. Regarding deciding the number of clusters, there are generally two strategies among literature. First, the number of clusters is an input parameter of a clustering algorithm. Such as K-mean clustering [67], spectral clustering [68, 69], and agglomerating clustering. Second, the number of clusters is decided by a given criteria, such as DBSCAN [70], mean shift [71], modified spectral clustering [72], and persistence-based clustering [73]. For the sake of simplicity, we chose the K-mean clustering, spectral clustering, and the agglomerating clustering as candidate algorithms. They are evaluated by the indicator $\mathbf{SSE}$ with the number of clusters given as 2 to 42. With the same clustering algorithm, the larger the number of clusters, the smaller of the SSE. However, a large

number of clusters means over-clustering. A simple way of choosing the number of clusters is named as the rule of elbow [66]. It infers the point where increasing the number of clusters brings uninterestingly small reduce of SSE. Depend on the experiment results shown in Fig. 3, the K-mean clustering has a better performance in terms of $\mathbf{SSE}$ so that it was chosen as the clustering algorithm. The number of clusters equals to 8 according to the rule of elbow.

2.3.4 Population density

In section 2.3.3, the percentages of population that distributes on land consumption types reveal the distribution of population within a city. It normalizes the population differences between cities for the clustering. For land consumption analysis, it is necessary to access the absolute value of population density. We have calculated the population density for each city, on the land of different morphological types, the densities is demonstrated in Table. 3 in the discussion. It is simply the number of population that reside on a certain land type dividing the area of the land type. The unit of the density is persons per 100 square meters. The absolute population density directly quantifies the status of land consumption. It is one of the crucial quantity that was discussed in section 3.2 and 3.3.

3 Results and discussion

In this section, first, we perform a comprehensive evaluation on the performance of the proposed automatic classification system, on three cities, Munich, Mumbai and New York. Second, we discuss the eight morphological clusters in details which separates the 42 cities in terms of their morphological characteristics. At last, a sustainable development analysis is carried out by statistically considering the population density and the dominant characteristics of morphological clusters.

3.1 Evaluation of the automatic classification system

With three exemplary cities, Munich, Mumbai, and New York, this section carries out a comprehensive evaluations of the proposed automatic classification system.

3.1.1 Ground reference configuration

To evaluate the classification system, the ground reference data needs to be separated as the training data and testing data. We block-wisely and equally separate the ground reference data into ten subsets, as shown in Fig. 4.

It is preferred by the authors to block-wisely separate the ground reference data, because training data are block-wisely delineated for real remote sensing applications. The random selection of training data that often appears in literature would make the training and the testing samples strictly follow the same distribution, resulting in the trained algorithm overfits the testing data and its performance is overestimated. This kind of evaluation can not summarize the performance of an algorithm for real remote sensing applications [74].

The ground reference data is separated into ten subsets for two reasons. First, in real remote sensing applications, the training samples normally occupies a very small fraction of data of the whole study area, as the example shown in Fig. 4. It means the training data can not fully represent the data distribution. Therefore, training with 90% of ground reference data and testing with 10% of it often overestimates the performance of an algorithm in terms of the whole study area. Thus, in our evaluation, instead of only taking 90% of the ground reference as training samples, it also takes 10% and 50% of the ground reference as training samples. With the ten subsets, it is easy to pick up 10%, 50%, or 90% of the ground reference data as the training data by respectively selecting one, five, or nine subsets out of the ten. The rest subsets that are not involved in training will then be used as the testing data. These three cases with 10%, 50%, and 90% in this evaluation are coded as TR10, TR50, and TR90, respectively, as shown in Fig. 5. The authors believe that evaluation with training data occupying different percentages of reference data would provide a possible range of accuracy as a reliable indication for a real application. Second, with the ten subsets, the authors carried out the evaluation in a cross-validation style. For example, with 10% ground reference data as the training data, the cross-validation evaluation first trains an algorithm with the first subset as training data and tests it with the rest nine subsets. It repeats the procedure nine more times with the second to tenth subset as training data respectively. Therefore, a mean accuracy over the ten experiments is statistically more accurate for the evaluation comparing to any one of the ten accuracy. This evaluation is able to mitigate the impact over the evaluation indicator that is caused by fortuitous factors hidden in training and testing data sets, such as different training data sets might represent slightly different data distributions due to sampling.

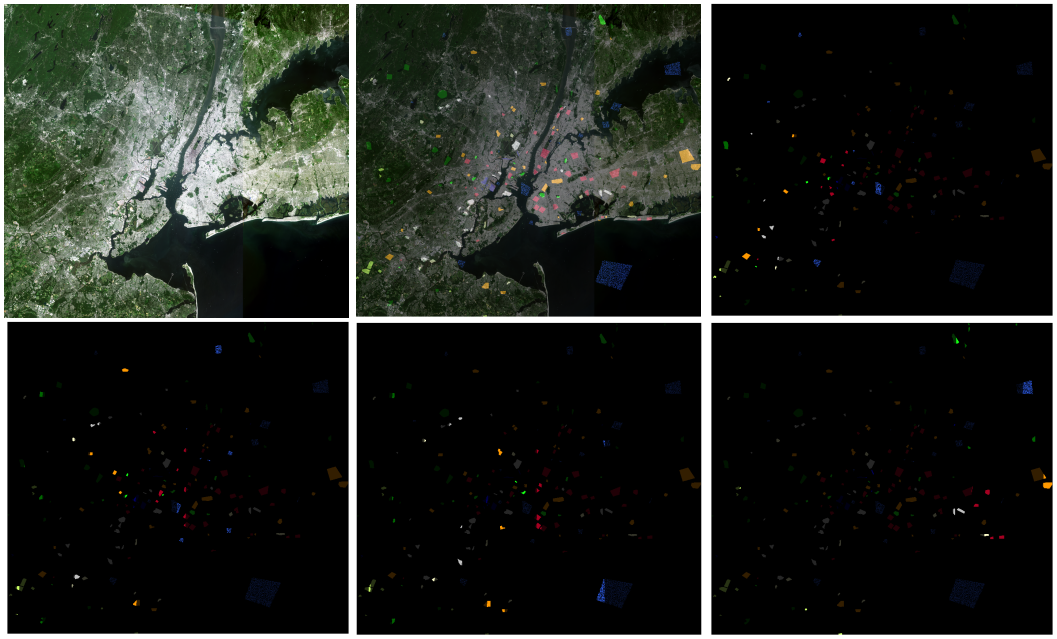


Figure 4: The ground reference data of the city New York is spatially block-wisely separated into ten subsets, sequentially from west to east. The first subset includes the most western 10% samples of each class; the second subset includes the second western 10% samples of each class; and so on so forth. From top to bottom, left to right: the RGB components of Sentinel-2 data for the city New York; the ground reference data overlapped with the RGB imagery; the ground reference data with the first subset highlighted; the ground reference data with the second subset highlighted; the ground reference data with the third subset highlighted; the ground reference data with the tenth subset highlighted.

3.1.2 Comparison configuration

To demonstrate the performance of the proposed classification system, three other classification systems are chosen for comparison. They are: (1) classifying Sentinel-1 data using random forest (S1); (2) classifying Sentinel-2 data using random forest (S2); (3) classifying the concatenation of Sentinel-1 and Sentinel-2 data using random forest (CON). The three classification systems are chosen for comparison because similar strategies have appeared on large scale classification tasks in literature [45, 62, 46, 12, 75, 13, 76]. For all of the four classification systems, they are all operated in an automatic way. There are two technical details should be mentioned. The first is the input data. Each system takes the same input data as described in section 2.2. The second detail is setting the number of trees for random forest. In our proposed system, the topology diversity results in 81 fused data. And 100 trees were used in random forest for classifying each of the 81 fused data. Therefore, it results in 8100 trees in the topology enhanced ensemble classification. To make a fair comparison, the number of trees in random forest is set as 8100 for all the three classification systems in comparison. The proposed classification system is named as the TEEC in the evaluation, short for topology enhanced ensemble classification.

3.1.3 Evaluation

As shown in Fig. 5, the proposed classification system outperform the other systems in terms of the mean overall accuracy of a 10-folds cross-validation evaluation, under all scenario of different training sets and for all three cities. More importantly, by comparing the accuracy differences of systems, one can find that the proposed system is specially good at those cases where the amount of testing data is much larger than the amount of training data. This is exactly the real case in remote sensing practice. With the sophisticated evaluation strategy, the authors conclude the proposed classification system is a good option among state-of-the-art techniques to accomplish the task of classifying 42 cities in this study.

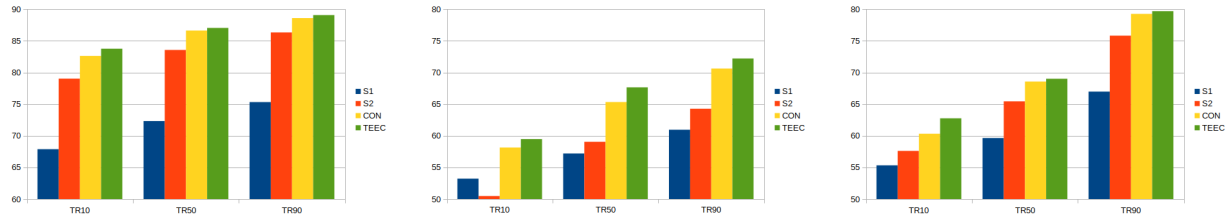


Figure 5: Evaluation of four classification systems, S1, S2, CON, and TEEC. The x-axis depicts the percentages of training samples over the ground reference, TR10: training samples occupies 10% data of ground reference data; TR50: 50%; TR90: 90%. The y-axis denotes the mean overall accuracy of the 10-folds cross-validation evaluation.

3.2 Urban pattern

With the method described in section 2.3.3, the selected 42 cities have been categorized into eight clusters, as shown in Fig. 6. Inspecting the classification maps visually in Fig. 6 and the cluster centers in Table. 2, we found each of the eight clusters has an outstanding characteristic in urban morphology. And some are related to population distribution.

Table 2: This table presents the mean cluster centers of the ten features for the eight morphological clusters. The features are explained in section 2.3.3 in details. The unit of the numbers is percentage (%).

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Share of compact area	63	10	5	31	15	50	29	13
Share of open area	8	27	35	11	55	11	10	40
Share of lightweight area	0	0	0	9	0	0	0	5
Share of industrial area	17	31	8	2	8	12	6	12
Share of natural area	12	31	52	47	22	27	55	30
Share of population in compact area	79	18	9	40	24	69	64	25
Share of population in open area	5	35	59	5	59	9	14	38
Share of population in lightweight area	0	0	0	43	0	0	0	13
Share of population in industrial area	13	36	14	3	9	12	8	16
Share of population in natural area	4	12	18	8	8	9	14	7

The first cluster concludes a group of cities which has a super compact urban morphology. This cluster has on average 63% of land having been consumed as compact built-up. And 79% of their population resides within the compact built-up area. The second cluster represents an industrial urban morphology. It includes cities whose 31% of land are occupied by industrial structure. Seven out of the total nine cities in this cluster are from China. By cross check with optical images, the industrial structures are mostly large and low buildings functioning as factories and warehouses. For the city of Amsterdam, tanks and warehouses near the harbor were also identified. The third cluster mainly includes European cites that has an open and green urban morphology. They are characterized by having more than 50% of land as natural cover and nearly 60% population is distributed in open area. The fourth cluster has the morphological characteristics that a large portion of informal settlement exists. With the 42 cities, it includes only the city of Islamabad which has 9% of land constructed as light weight built-up and has 43% of population residing in light weight built-ups. According to our labeling experiences and visual inspection, the light weight built-up mostly reflects the distribution of informal settlements. 9% is a very high percentage among the selected 42 cities. The 9% land area of light weight build-up in Islamabad is home to 43% of the population in this city, which makes those areas heavily populated. The fifth cluster owns an open urban morphology. This cluster includes eight cities whose 55% of land were constructed into open built-ups. A very interesting point is that all those eight cities in this cluster are the capital city to eight different countries. The sixth cluster has a compact urban morphology. It gathered cities that have high percentages of compact built-up and population distributed in there, namely 50% and 69% respectively. Comparing to the first cluster which also has a high percentage of compact area, there exists 10% shortage in the sixth cluster. And the sixth cluster has 27% of land being natural cover, which is 15% higher than it in the first cluster. The seventh cluster has a compact and green urban morphology. It has 29% and 55% of their land covered by compact built-up and natural cover, respectively. However, an average 64% of their population resides in the 29% compact built-up areas. The last cluster has mixture urban morphology of light weight built-up and open built-up. In this cluster, the open built-up occupies 40% of the land and the light weight built-ups accounts for about 5%.

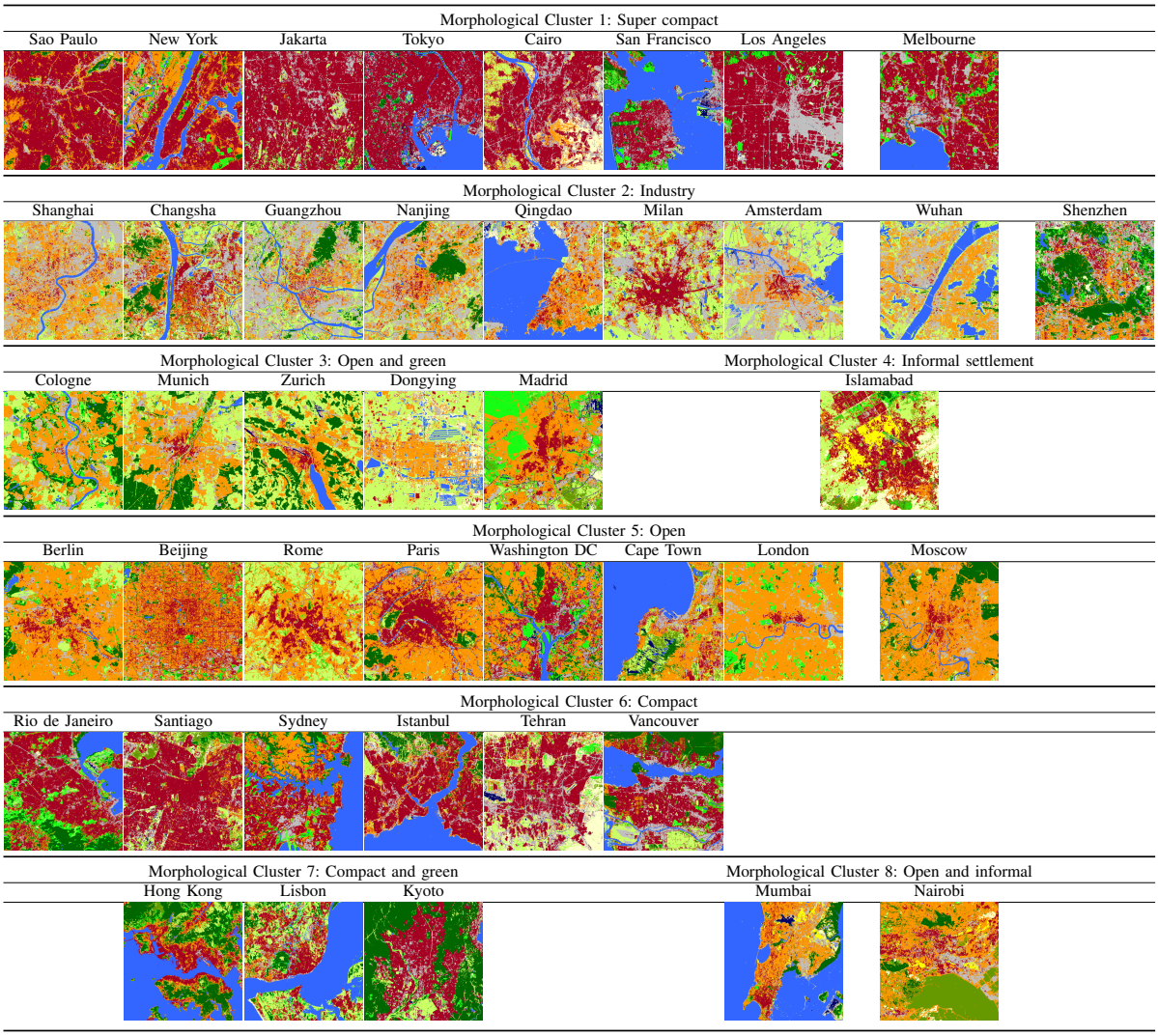


Figure 6: Visualization of the LCZ classification maps for the 42 cities. They are provided by the proposed classification system and organized by the clustering results.

To sum up, we have identified eight different urban morphological clusters among the 42 cities around the world. Those urban morphological clusters are: super compact, industry, open-green, informal settlement, open, compact, compact-green, and open-informal.

3.3 Discussion of sustainable development

United Nations Human Settlements Program explains the SDG 11.3.1 in [27] as an indicator to reflect whether the urban area and population change proportionally or not. It essentially evaluates the land consumption sustainability of individual city by focusing on the changing along time. Besides the changing part, it is also important to analyze whether an entire city suits the demand of the land consumption sustainability or not. However, an obvious contradiction in the sustainability of land consumption locates at the trade-off of efficient land consumption and the well-being of human kind. The population density is a quantitative measure for the trade-off. The higher value of population density, the more efficient of land consumption, and the higher probability of less well-being for the inhabitants. Even with the measurement, it is very complex to locate a balanced spot of the trade-off for a city. Furthermore, it might be reasonable to apply different standards on different cities, due to historical, economical, and cultural impacts.

To deal with the challenge, we first attempt to identify some extreme cases where cities are clearly not following a pattern of sustainable development, in terms of urban morphology and population density. Afterwards, within in each morphological cluster, we discuss the balance of the trade-off between efficient land consumption and human well-being in a statistical way.

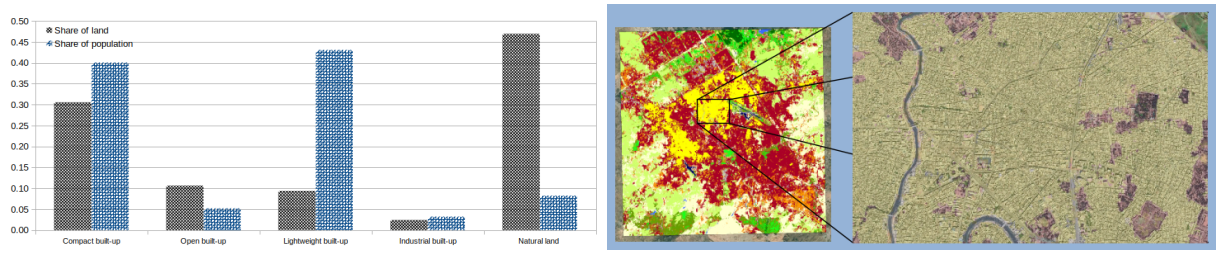


Figure 7: Left: the shares of land and population corresponding to the five different built-ups for the city of Islamabad; Right: A zoomed in area of lightweight built-up in the city of Islamabad.

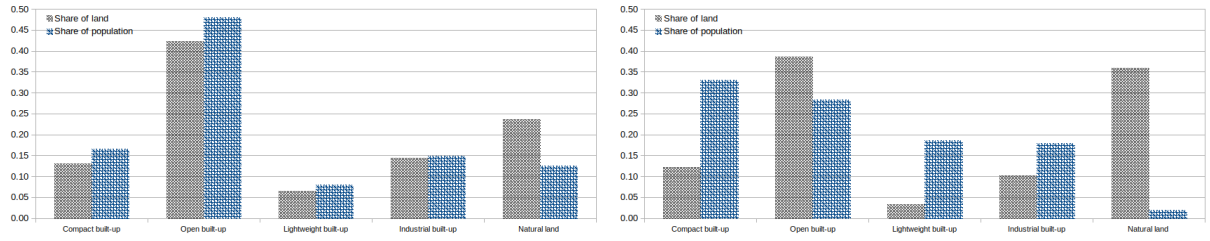


Figure 8: Left: the shares of land and population corresponding to the five different built-ups for the city of Mumbai; Right: the shares of land and population corresponding to the five different built-ups for the city of Nairobi.

Among the urban morphological clusters, the city Islamabad is isolated as one cluster. As shown in the bar graph in Fig. 7, 43% population of the city Islamabad resides in the lightweight built-up area which occupies 9% land of the city. According to Table. 3, it leads to an heavily populated lightweight built-up area with a density of 3.22 persons per 100 square meters, which is almost three times denser than the overall population density of Islamabad, 1.21 persons per 100 square meters. This severely uneven population distribution within a city is not an outcome of sustainable land consumption. More importantly, as shown in the zoom-in lightweight built-up area of Islamabad in Fig. 7, the evidence of arbitrary building sizes, disordered building orientations, and super compactness of the man-made structures strongly suggest that the lightweight built-up area in Islamabad is a slum area [77, 78]. It was also found in [33] that the lightweight built-up is highly related to slum districts. Since there is a high probability that the lightweight built-up is related to slum area, we looked into the eighth cluster, open-informal, where the lightweight built-up exists. Unlike the case of Islamabad where a dominated proportion of population live in lightweight built-up, the majority inhabitants in Mumbai and Nairobi reside in compact and open built-up area. However, as shown in Table. 3, the population densities in the lightweight built-up area are 4.13 and 4.01 persons per 100 spuare meters for Mumbai and Nairobi, respectively. Both densities are even higher than it in Islamabad. According to the Table. 3, besides Nairobi, Mumbai, and Islamabad, three other cities, Cairo, Shanghai, and Rio de Janeiro, also have heavily populated lightweight built-up areas with population density of 6.34, 5.38, and 4.30 persons per 100 square meters. For all those six cities, their lightweight built-up area is heavily populated both in absolute population density number and by comparing to the densities of other built-up areas of the same city. To sum up, the prior task of these six cities, Mumbai, Cairo, Shanghai, Rio de Janeiro, Islamabad, and Nairobi, for achieving sustainable land consumption is to decrease the population density in lightweight built-up area.

As shown in Table. 3 and Fig. 9, the city Hong Kong has an overall population density of 3.37 persons per 100 square meters, which is five times denser than Lisbon and Kyoto (0.80 and 0.77) which fall into the same morphological cluster of Hong Kong. The large population density differences within one urban morphological cluster draw our attention. For the three cities, they have on average 64% population resides in 29% of their land that was compact built-up, as shown in Table. 2. These percentages have already demonstrated that inhabitants concentrate in the compact areas. However, since Hong Kong has a much larger number of population, the absolute population density of the compact built-up is 4.38 persons per 100 square meters in Hong Kong, which is also the most dense compact built-up area among the 42 selected cities. Statistically within the 42 cities, one can concludes that it is too crowed to claim that Hong Kong is in the track of sustainable development of land consumption.

Table 3: Population densities in the 42 selected cities in terms of overall built-up areas, compact built-up areas, open built-up areas, lightweight built-up areas, and the industrial built-up areas. The unit is the number of persons per 100 square meters.

City	Population density (unit: persons per 100 square meters)				
	Overall	Compact built-up	Open built-up	Lightweight built-up	Industrial built-up
Mumbai	3.83	4.24	3.79	4.13	3.42
Hong Kong	3.37	4.38	2.34	0.00	1.71
Cairo	3.27	3.97	0.72	6.34	2.32
Shanghai	2.53	3.50	2.35	5.38	2.55
Guangzhou	2.15	3.35	2.41	2.00	1.82
Istanbul	2.12	2.22	1.53	0.00	2.14
Beijing	1.98	2.08	1.99	0.00	1.55
Wuhan	1.72	2.61	1.59	0.00	1.84
Tehran	1.67	1.77	1.20	0.00	1.61
Tokyo	1.50	1.54	1.38	0.00	1.31
Jakarta	1.46	1.46	1.16	1.66	1.45
Shenzhen	1.42	1.59	1.47	0.00	1.29
Rio De Janeiro	1.39	1.58	0.97	4.30	0.70
Moscow	1.33	1.25	1.35	0.00	1.22
Nanjing	1.32	2.72	1.27	0.00	1.12
New York	1.31	1.82	0.83	0.00	0.28
Paris	1.30	1.88	1.05	0.00	1.01
Sao Paulo	1.25	1.44	0.74	0.00	0.60
Islamabad	1.21	0.92	0.34	3.22	0.94
Qingdao	1.19	1.41	1.37	0.00	0.83
Nairobi	1.06	1.88	0.51	4.01	1.23
London	1.02	0.75	1.07	0.00	0.66
Changsha	0.97	1.32	0.91	0.00	0.72
Santiago De Chile	0.95	1.00	0.60	0.00	0.90
Madrid	0.93	1.16	0.87	0.00	0.94
Lisbon	0.80	0.92	0.66	0.00	0.46
Kyoto	0.77	0.78	0.53	0.00	0.73
Berlin	0.59	0.73	0.57	0.00	0.66
Milan	0.59	0.72	0.52	0.00	0.53
San Francisco	0.58	0.72	0.14	0.00	0.34
Los Angeles	0.55	0.64	0.27	0.00	0.30
Munich	0.54	0.75	0.52	0.00	0.55
Rome	0.47	0.62	0.41	0.00	0.52
Sydney	0.46	0.56	0.36	0.00	0.32
Vancouver	0.39	0.49	0.17	0.00	0.24
Zurich	0.39	0.60	0.36	0.00	0.41
Cologne	0.38	0.48	0.38	0.00	0.40
Amsterdam	0.36	0.60	0.39	0.00	0.33
Washington DC	0.36	0.56	0.26	0.00	0.29
Cape Town	0.35	0.60	0.27	0.97	0.31
Melbourne	0.33	0.38	0.20	0.00	0.24
Dongying	0.11	0.09	0.12	0.00	0.12

According to Fig. 9 and Fig. 10, from statistical perspective within morphological types or over 42 cities, Mumbai, Hong Kong, and Cairo are over populated for their excessive margin of population densities. Similarly on the other end, Dongying has the issue of under efficient land consumption.

It is worthy to point out that we didn't detail the sustainable discussion relating to the cities of industrial type. The state-of-the-art global population data is produced by evenly distributing census data to the built-up area of the administrative unit [57, 79]. It results in the number of inhabitants in a district is linear correlated with the area of the district.

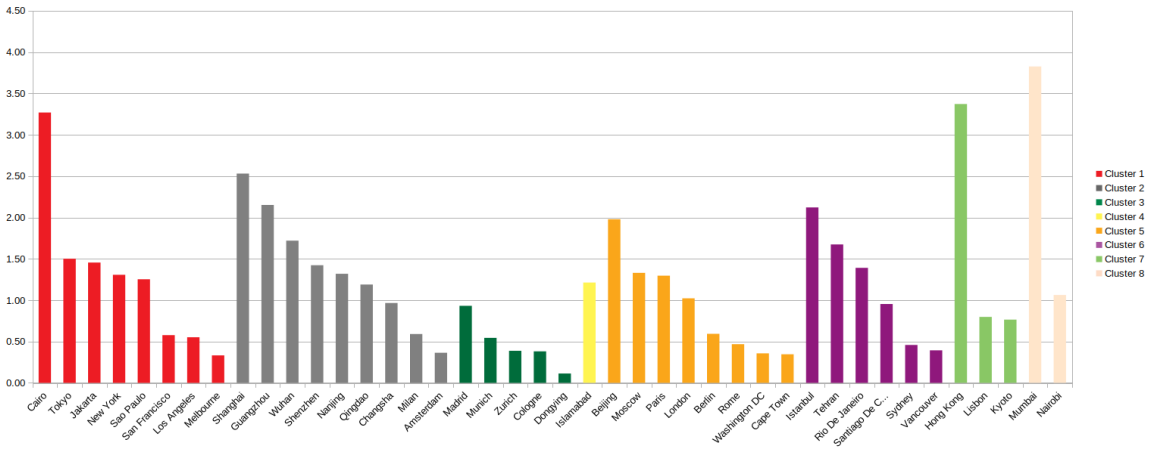


Figure 9: A bar graph of overall population densities of the selected 42 cities organized in clusters. The colors represent clusters.

Therefore, the industrial area is also assigned with a number of inhabitant who actually live in the non-industrial area of the same administrative unit. For those cities of industrial type, these industrial area were filled with population that resides somewhere else and the population density of residential areas was heavily underestimated.

As each city cluster reveals its morphological characteristics and partially functionalities, we tend to believe different standards should be applied to different city clusters. Because city has its own morphological characteristics, functionality, historical impact, future plan, economic status, and cultural influence. Therefore, we nominate the cities having the median population density of each cluster and having not suffered from any extreme case as a potential model of sustainable development for cities belonging to the same cluster. New York and Sao Paulo are the models for the super compact cluster. London and Berlin are the models for the open cluster. Santiago de Chile is the model for the compact cluster. Although the cities in open-green cluster, mainly European cities, are generally under efficient in land consumption by comparing the population density to all other clusters, Zurich is the model for this cluster. The compact-green cluster includes cities of, Hong Kong, Lisbon, and Kyoto, where statistical analysis would not be convincing for a sampling number of three. Hong Kong is evidently over populated which persuades us Lisbon and Kyoto are closer to the goal of sustainable development, in this cluster.

4 Conclusion and outlook

In this work, we have developed an automatic classification system that fuses the Sentinel-1 and Sentinel-2 data with a state-of-the-art semi-supervised technique and produces a LCZ classification map by a topological enhanced ensemble classification technique. With the classification system, we have produced LCZ maps for 42 selected cities that distributes around the world in trans-cultural, transcontinental, and trans-environmental regions. The LCZ map categorized the urban area into five classes, compact built-up area, open built-up area, lightweight built-up area, industrial built-up area, and natural land cover. Coupling with the freely accessible global population data provided by Joint Research Centre of European Commission, we have derived the percentages of land that fall into the five classes and the percentages of population that reside in the five classes as features upon which eight morphological clusters were derived from the 42 cities. Those eight clusters reveal the morphological characteristics and the functionalities. Considering the absolute value of population density as an important additional factor, we have analyzed the land consumption efficiency in a statistical manner within each cluster and among all 42 cities.

In the discussion, we have analyzed the sustainable development of land consumption for the 42 cities, from the perspectives of city clusters and the population density. We first conclude that, on the way to sustainable development, the issue for Mumbai, Cairo, Shanghai, Rio de Janeiro, Islamabad, and Nairobi is the unevenly and heavily populated lightweight built-up area being an obstacle for human well-being. Second, Hong Kong suffers from super populated compact built-up area. Third, statistically among the 42 cities, the cities of Mumbai, Hong Kong, and Cairo are heavily populated in general, and the city of Dongying is severely under efficient in land consumption. At last, besides those extreme cases, we name the cities of New York and Sao Paulo as the model of sustainability of land consumption for the super compact cities; London and Berlin as the models for the openly built cities; Santiago de Chile as the model for the compact cities; and Zurich as the model for open-green cities.

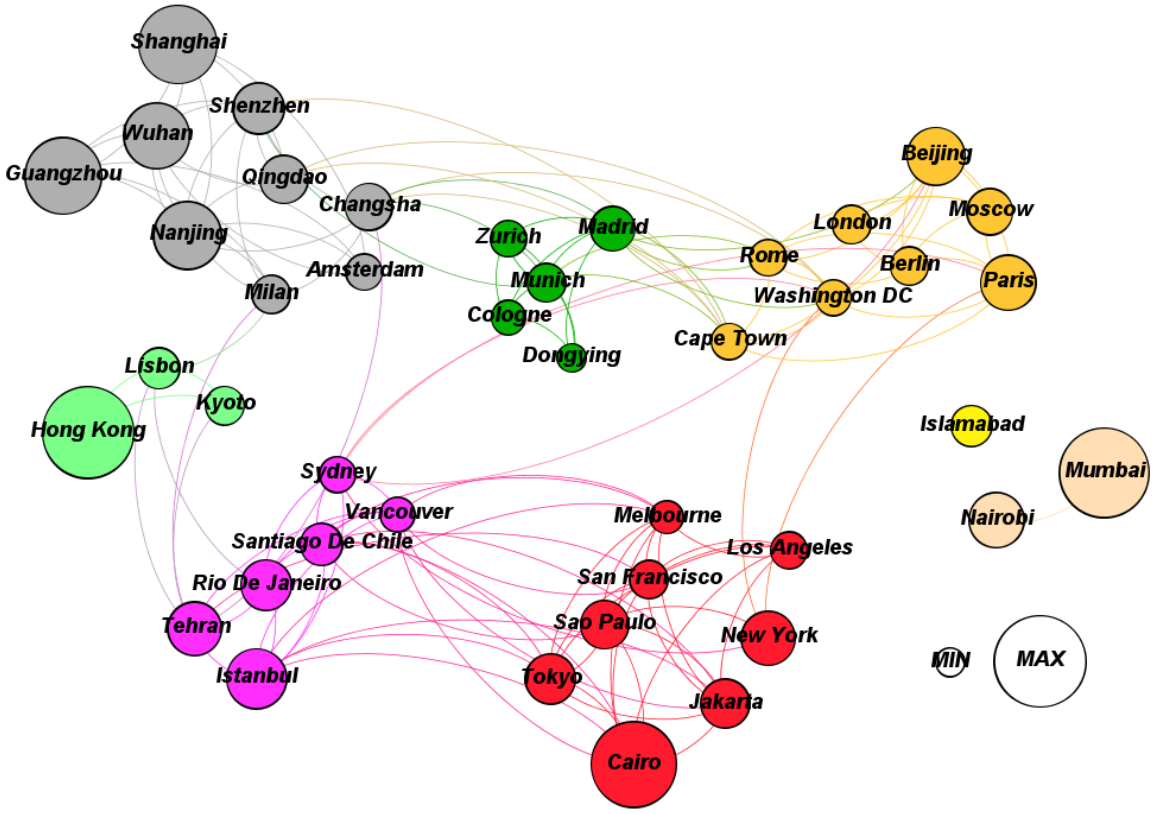


Figure 10: A graph visualization of the 42 cities in terms of their urban morphological types. This size of the node represents the overall population density of the city. The MIN and MAX are designed to show the minimum and maximum sizes, respectively. Within one urban morphological type, city with smaller population density locates closer to the center.

For the future work, instead of a binary settlement layer, our future global LCZ layer will categorize the urban area into detailed classes which reveal the functionality and the morphology of the land. As having been discussed in the section 3.3, the binary settlement layer is not accurate enough to estimate the spatial distribution of population. Therefore, our future global LCZ layer is able to improve the spatial distribution accuracy of population, within an urban area. For example, the urban inhabitant will not appear in an industrial area.

As this work aimed at exploring the possibility of using LCZ maps to study the sustainable development of land consumption statistically, 42 representative cities from the world were chosen, yet not statistically significant. We plan to carry out a much more statistically meaningful study in the future with about 1700 cities around the world. And several technical details will be improved in the future study, such as how to decide the city boundary.

References

[1] United Nations. Resolution adopted by the general assembly on 25 september 2015. In *Transforming Our World: The 2030 Agenda for Sustainable Development*. United Nations New York, NY, 2015.

[2] United nations, the world's cities in 2018. `https://www.un.org/en/events/citiesday/assets/pdf/the_worlds_cities_in_2018_data_booklet.pdf`. Accessed: 2019-06-14.

[3] Cathy Aniello, Ken Morgan, Arthur Busbey, and Leo Newland. Mapping micro-urban heat islands using landsat tm and a gis. *Computers & Geosciences*, 21(8):965–969, 1995.

[4] George Xian and Mike Crane. An analysis of urban thermal characteristics and associated land cover in tampa bay and las vegas using landsat satellite data. *Remote Sensing of environment*, 104(2):147–156, 2006.

[5] Enikő Lelovics, János Unger, Tamás Gál, and Csilla V Gál. Design of an urban monitoring network based on local climate zone mapping and temperature pattern modelling. *Climate research*, 60(1):51–62, 2014.

[6] JG Masek, FE Lindsay, and SN Goward. Dynamics of urban growth in the washington dc metropolitan area, 1973-1996, from landsat observations. *International Journal of Remote Sensing*, 21(18):3473–3486, 2000.

[7] Martin Herold, Noah C Goldstein, and Keith C Clarke. The spatiotemporal form of urban growth: measurement, analysis and modeling. *Remote sensing of Environment*, 86(3):286–302, 2003.

[8] Qihao Weng. Land use change analysis in the zhujiang delta of china using satellite remote sensing, gis and stochastic modelling. *Journal of environmental management*, 64(3):273–284, 2002.

[9] H Taubenböck, M Weigand, T Esch, J Staab, M Wurm, J Mast, and S Dech. A new ranking of the world's largest cities—do administrative units obscure morphological realities? *Remote Sensing of Environment*, 232:111353, 2019.

[10] GAF AG. Earth observation for sustainable development: Urban development - service portfolio.

[11] Paolo Gamba and Gianni Lisini. Fast and efficient urban extent extraction using asar wide swath mode data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(5):2184–2195, 2013.

[12] Yifang Ban, Alexander Jacob, and Paolo Gamba. Spaceborne sar data for global urban mapping at 30 m resolution using a robust urban extractor. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:28–37, 2015.

[13] Thomas Esch, Wieke Heldens, Andreas Hirner, Manfred Keil, Mattia Marconcini, Achim Roth, Julian Zeidler, Stefan Dech, and Emanuele Strano. Breaking new ground in mapping human settlements from space–the global urban footprint. *ISPRS Journal of Photogrammetry and Remote Sensing*, 134:30–42, 2017.

[14] Dengsheng Lu and Qihao Weng. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5):823–870, 2007.

[15] Jingliang Hu, Danfeng Hong, Yuanyuan Wang, and Xiao Xiang Zhu. A comparative review of manifold learning techniques for hyperspectral and polarimetric sar image fusion. *Remote Sensing*, 11(6):681, 2019.

[16] Qihao Weng, Dengsheng Lu, and Jacquelyn Schubring. Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. *Remote sensing of Environment*, 89(4):467–483, 2004.

[17] Huanfeng Shen, Liwen Huang, Liangpei Zhang, Penghai Wu, and Chao Zeng. Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: A 26-year case study of the city of wuhan in china. *Remote Sensing of Environment*, 172:109–125, 2016.

[18] Peijun Du, Sicong Liu, Paolo Gamba, Kun Tan, and Junshi Xia. Fusion of difference images for change detection over urban areas. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(4):1076–1086, 2012.

[19] GEO Group on Earth Observations. Earth observations in support of the 2030 agenda for sustainable development, 2017.

[20] M Klotz, T Kemper, C Geiß, T Esch, and H Taubenböck. How good is the map? a multi-scale cross-comparison framework for global settlement layers: Evidence from central europe. *Remote Sensing of Environment*, 178:191–212, 2016.

[21] Thomas Esch, Felix Bachofer, Wieke Heldens, Andreas Hirner, Mattia Marconcini, Daniela Palacios-Lopez, Achim Roth, Soner Üreyen, Julian Zeidler, Stefan Dech, et al. Where we live—a summary of the achievements and planned evolution of the global urban footprint. *Remote Sensing*, 10(6):895, 2018.

[22] Marc Paganini, Ivan Petiteville, Stephen Ward, George Dyke, Matthew Steventon, Jennifer Harry, Flora Kerblat, and CEOS. Satellite earth observations in support of the sustainable development goals. *Euorpean Space Agency*, 2018.

[23] Martino Pesaresi, Daniele Ehrlich, Ivano Caravaggi, Mayeul Kauffmann, and Christophe Louvrier. Toward global automatic built-up area recognition using optical vhr imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(4):923–934, 2011.

[24] Martino Pesaresi, Guo Huadong, Xavier Blaes, Daniele Ehrlich, Stefano Ferri, Lionel Gueguen, Matina Halkia, Mayeul Kauffmann, Thomas Kemper, Linlin Lu, et al. A global human settlement layer from optical hr/vhr rs data: concept and first results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(5):2102–2131, 2013.

[25] Michele Melchiorri, Aneta Florczyk, Sergio Freire, Marcello Schiavina, Martino Pesaresi, and Thomas Kemper. Unveiling 25 years of planetary urbanization with remote sensing: Perspectives from the global human settlement layer. *Remote Sensing*, 10(5):768, 2018.

[26] Michele Melchiorri, Martino Pesaresi, Aneta J Florczyk, Christina Corbane, and Thomas Kemper. Principles and applications of the global human settlement layer as baseline for the land use efficiency indicator—sdg 11.3. 1. *ISPRS International Journal of Geo-Information*, 8(2):96, 2019.

[27] United Nations Human Settlements Programme (UN-Habitat). Goal 11. make cities and human settlements inclusive, safe, resilient and sustainable. `https://unstats.un.org/sdgs/metadata/?Text=&Goal=&Target=11.3`, 2016.

[28] Ian D Stewart and Tim R Oke. Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, 93(12):1879–1900, 2012.

[29] Iain D Stewart, Timothy R Oke, and E Scott Krayenhoff. Evaluation of the 'local climate zone'scheme using temperature observations and model simulations. *International journal of climatology*, 34(4):1062–1080, 2014.

[30] Benjamin Bechtel, Paul Alexander, Jürgen Böhner, Jason Ching, Olaf Conrad, Johannes Feddema, Gerald Mills, Linda See, and Iain Stewart. Mapping local climate zones for a worldwide database of the form and function of cities. *ISPRS International Journal of Geo-Information*, 4(1):199–219, 2015.

[31] X. X. Zhu, J. Hu, C. Qiu, Y. Shi, J. Kang, L. Mou, H. Bagheri, M. Häberle, Y. Hua, R. Huang, Lloyd H. Hughes, H. Li, Y. Sun, G. Zhang, S. Han, M. Schmitt, and Y. Wang. So2sat lcz42: A benchmark dataset for globallocal climate zones classification. *submitted*.

[32] Benjamin Bechtel, M Pesaresi, L See, G Mills, J Ching, PJ Alexander, JJ Feddema, AJ Florczyk, and I Stewart. Towards consistent mapping of urban structure-global human settlement layer and local climate zones. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41:1371–1378, 2016.

[33] NGR Perera and R Emmanuel. A "local climate zone" based approach to urban planning in colombo, sri lanka. *Urban climate*, 23:188–203, 2018.

[34] Wudapt. `http://www.wudapt.org/`. Accessed: 2018-04-10.

[35] Benjamin Bechtel. Multitemporal landsat data for urban heat island assessment and classification of local climate zones. In *Urban Remote Sensing Event (JURSE), 2011 Joint*, pages 129–132. IEEE, 2011.

[36] Olha Danylo, Linda See, Benjamin Bechtel, Dmitry Schepaschenko, and Steffen Fritz. Contributing to wudapt: A local climate zone classification of two cities in ukraine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5):1841–1853, 2016.

[37] Yong Xu, Chao Ren, Meng Cai, Ng Yan Yung Edward, and Tianjun Wu. Classification of local climate zones using aster and landsat data for high-density cities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(7):3397–3405, 2017.

[38] Zhigang Xu, Jike Chen, Junshi Xia, Peijun Du, Hongrui Zheng, and Le Gan. Multisource earth observation data for land-cover classification using random forest. *IEEE Geoscience and Remote Sensing Letters*, 2018.

[39] Benjamin Bechtel and Christian Daneke. Classification of local climate zones based on multiple earth observation data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(4):1191–1202, 2012.

[40] Tamás Gál, Benjamin Bechtel, and János Unger. Comparison of two different local climate zone mapping methods. 2015.

[41] Jan Geletič and Michal Lehnert. Gis-based delineation of local climate zones: The case of medium-sized central european cities. *Moravian Geographical Reports*, 24(3):2–12, 2016.

[42] Benjamin Bechtel, Linda See, Gerald Mills, and Mícheál Foley. Classification of local climate zones using sar and multispectral data in an arid environment. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(7):3097–3105, 2016.

[43] Jingliang Hu, Pedram Ghamisi, and Xiao Zhu. Feature extraction and selection of sentinel-1 dual-pol data for global-scale local climate zone classification. *ISPRS International Journal of Geo-Information*, 7(9):379, 2018.

[44] N. Kaloustian, M. Tamminga, and B. Bechtel. Local climate zones and annual surface thermal response in a mediterranean city. In *2017 Joint Urban Remote Sensing Event (JURSE)*, pages 1–4, March 2017.

[45] Naoto Yokoya, Pedram Ghamisi, and Junshi Xia. Multimodal, multitemporal, and multisource global data fusion for local climate zones classification based on ensemble learning. In *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*, pages 1197–1200. IEEE, 2017.

[46] Matthias Demuzere, Benjamin Bechtel, and Gerald Mills. Global transferability of local climate zone models. *Urban climate*, 27:46–63, 2019.

[47] Chunping Qiu, Lichao Mou, Micheal Schmitt, and Xiao Xiang Zhu. LCZ-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network. *ISPRS J. Photogramm. Remote Sens.*, 154:151–162, 2019.

[48] Chang Wang and Sridhar Mahadevan. A general framework for manifold alignment. In *AAAI fall symposium: manifold learning and its applications*, pages 53–58, 2009.

[49] Devis Tuia, Michele Volpi, Maxime Trolliet, and Gustau Camps-Valls. Semisupervised manifold alignment of multimodal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 52(12):7708–7720, 2014.

[50] Marcello Schiavina, Sergio Freire, and Kytt MacManus. Ghs population grid multitemporal (1975, 1990, 2000, 2015). *European Commission, Joint Research Centre (JRC)[Dataset] PID: DOI: 10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218 PID: http://data.europa.eu/89h/0c6b9751-a71f-4062-830b-43c9f432370f*, 2019.

[51] United nations, the world's cities in 2016. `http://www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2016_data_booklet.pdf`. Accessed: 2018-04-10.

[52] Department of Economic United Nations and Population Division Social Affairs. World urbanization prospects: The 2014 revision. `https://population.un.org/wup/Publications/Files/WUP2014-Report.pdf`, 2015. Accessed: 2019-08-09.

[53] Jon Atli Benediktsson, Martino Pesaresi, and Kolbeinn Amason. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1940–1949, 2003.

[54] Mathieu Fauvel, Jón Atli Benediktsson, Jocelyn Chanussot, and Johannes R Sveinsson. Spectral and spatial classification of hyperspectral data using svms and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 46(11):3804–3814, 2008.

[55] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27, 2017.

[56] M. Schmitt, Lloyd H. Hughes, C. Qiu, and X. X. Zhu. Aggregating cloud-free Sentinel-2 images with google earth engine. 2019. to appear.

[57] Sergio Freire, Kytt MacManus, Martino Pesaresi, Erin Doxsey-Whitfield, and Jane Mills. Development of new open and free multi-temporal global population grids at 250 m resolution. 2016.

[58] Wenzhi Liao, Aleksandra Pižurica, Rik Bellens, Sidharta Gautama, and Wilfried Philips. Generalized graph-based fusion of hyperspectral and lidar data using morphological features. *IEEE Geoscience and Remote Sensing Letters*, 12(3):552–556, 2015.

[59] Gurjeet Singh, Facundo Mémoli, and Gunnar E Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *SPBG*, pages 91–100, 2007.

[60] Jingliang Hu, Danfeng Hong, and Xiao Xiang Zhu. Mima: Mapper-induced manifold alignment for semi-supervised fusion of optical image and polarimetric sar data. *arXiv preprint arXiv:1906.05512*, 2019.

[61] Li Li, Wei-Yi Cheng, Benjamin S Glicksberg, Omri Gottesman, Ronald Tamler, Rong Chen, Erwin P Bottinger, and Joel T Dudley. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Science translational medicine*, 7(311):311ra174–311ra174, 2015.

[62] Congcong Li, Peng Gong, Jie Wang, Zhiliang Zhu, Gregory S Biging, Cui Yuan, Tengyun Hu, Haiying Zhang, Qi Wang, Xuecao Li, et al. The first all-season sample set for mapping global land cover with landsat-8 data. *Science Bulletin*, 62(7):508–515, 2017.

[63] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[64] Tom Rainforth and Frank Wood. Canonical correlation forests. *arXiv preprint arXiv:1507.05444*, 2015.

[65] Haytham Elghazel, Alex Aussem, and Florence Perraud. Trading-off diversity and accuracy for optimal ensemble tree selection in random forests. In *Ensembles in Machine Learning Applications*, pages 169–179. Springer, 2011.

[66] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[67] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[68] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.

[69] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[70] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[71] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (5):603–619, 2002.

[72] Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. In *Advances in neural information processing systems*, pages 1601–1608, 2005.

[73] Frédéric Chazal, Leonidas J Guibas, Steve Y Oudot, and Primoz Skraba. Persistence-based clustering in riemannian manifolds. *Journal of the ACM (JACM)*, 60(6):41, 2013.

[74] Ronny Hänsch, Andreas Ley, and Olaf Hellwich. Correct and still wrong: The relationship between sampling strategies and the estimation of the generalization error. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3672–3675. IEEE, 2017.

[75] Gianni Lisini, Andreas Salentinig, Peijun Du, and Paolo Gamba. Sar-based urban extents extraction: from envisat to sentinel-1. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(8):2683–2691, 2017.

[76] Zhe Zhu, Curtis E Woodcock, John Rogan, and Josef Kellndorfer. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using landsat and sar data. *Remote Sensing of Environment*, 117:72–82, 2012.

[77] Hannes Taubenböck and NJ Kraff. The physical face of slums: a structural comparison of slums in mumbai, india, based on remotely sensed data. *Journal of Housing and the Built Environment*, 29(1):15–38, 2014.

[78] Mike Davis. Planet of slums. *New Perspectives Quarterly*, 23(2):6–11, 2006.

[79] Martino Pesaresi, Daniele Ehrlich, Stefano Ferri, Aneta Florczyk, Sergio Freire, Matina Halkia, Andreea Julea, Thomas Kemper, Pierre Soille, Vasileios Syrris, et al. Operating procedure for the production of the global human settlement layer from landsat data of the epochs 1975, 1990, 2000, and 2014. *Publications Office of the European Union*, pages 1–62, 2016.