TUM School of Engineering and Design
Technische Universität München

Population estimation utilizing Earth Observation data

Sugandha Doda

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz:

Prof. Dr. rer. nat. Martin Werner

Prüfende der Dissertation:

1. Prof. Dr.-Ing. habil. Xiaoxiang Zhu
2. Prof. Dr.-Ing. Yuanyuan Wang
3. Prof. Dr. Monika Kuffer

Die Dissertation wurde am 14.11.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 24.04.2024 angenommen.

Abstract

The rapid growth in population is putting a lot of pressure on the environment in terms of the demand for additional infrastructure, food and water, healthcare, schools, etc. A thorough understanding of current population distribution and future forecasts could support the government in many decision-making processes involving urban planning and policies, effectively allocating socio-technical supply and ensuring a good standard of living for all. The traditional method of collecting population data is through the census, in which the population data is collected and compiled across the census units. However, it is expensive, time-consuming and results in low spatial resolution population data. Alternatively, statistical and machine-learning methods have become more prevalent in population estimation studies to develop more up-to-date and spatially improved population data.

Despite these advancements, a thorough literature study undertaken as part of this Ph.D. research indicated that the majority of existing large-scale gridded products disaggregate known census counts to grids. Therefore, their accuracy is dependent on the quality of the census data. Other approaches that could predict the population in addition to population disaggregation are typically based on regional data obtained from a few cities. This limits their transferability to a different geographical region. The application of deep learning approaches in population estimation enables more efficient population maps. Nevertheless, due to the blackbox nature of these models, understanding and explaining their findings, as well as revealing the key features employed by the model to obtain its outcome, becomes critical. The resolution of the population data sets continues to be a challenge in various disciplines that demand high spatial-resolution population data such as infectious disease containment or disaster management. Overall, it has been determined that the constraints observed in population estimating studies are either a) due to the unavailability of large-scale data sets, b) transferable and interpretable methods, and c) enhancements in the spatial resolution of the existing population data.

In this context, this Ph.D. thesis investigates the various input data sources that correlate well with the population and creates a large-scale benchmark data set for population estimation using publicly available data sets. This data collection would aid in overcoming the limitation a) described above. It will be a valuable addition to the current literature and research community for developing new sophisticated methods in population estimation studies and comparing various approaches established utilizing regional data sets.

Using this data set, an end-to-end deep learning-based framework is developed that predicts the population rather than population disaggregation. Since the method can infer the population even in the absence of census data and only uses openly avail-

able large-scale data sets, it is easily transferable. To assess the method's performance, a comparison with another community standard product is performed for a few cities in Europe and the United States. And to further promote the trustworthiness of the method, an explainable AI technique is employed to unwire the blackbox model and comprehend its performance and limitations. These methods help to identify the important features that the model employed to reach its decision. The emphasis is on the method's accuracy, its transferability and interpretability to address the limitation b).

Finally, this thesis focuses on improving the resolution of the population data and thus addresses limitation c) from above. It employs a hybrid deep learning approach that generates gridded population maps and then disaggregates the estimated population counts to buildings, resulting in building level population maps. This study integrates the building data, such as building functions, building areas, and heights, to improve the accuracy and resolution of the population estimates. It also conducts a qualitative comparison of building data sets gathered from publicly accessible large-scale sources and regional open-access administrative sources, as well as examines how data quality impacts population estimation on a fine scale.

Zusammenfassung

Das rasche Bevölkerungswachstum setzt die Umwelt stark unter Druck. durch den Bedarf an zusätzlicher Infrastruktur, Nahrung und Wasser, Gesundheitsversorgung, Schulen usw. Ein gründliches Verständnis der aktuellen Bevölkerungsverteilung und der zukünftigen Prognosen könnte die Regierung bei vielen Entscheidungsprozessen in der Stadtplanung und -politik unterstützen der Stadtplanung und -politik, der effektiven Verteilung des soziotechnischen Angebots und der Gewährleistung eines guten Lebensstandards für alle Lebensstandard für alle. Die traditionelle Methode zur Erhebung von Bevölkerungsdaten ist die Volkszählung, bei der die Bevölkerungsdaten in den einzelnen Volkszählungseinheiten gesammelt und zusammengestellt werden. Allerdings, Sie ist jedoch teuer, zeitaufwändig und führt zu Bevölkerungsdaten mit geringer räumlicher Auflösung. Alternativ dazu haben sich statistische und maschinelle Lernmethoden durchgesetzt in Studien zur Bevöl-kerungsschätzung durchgesetzt, um aktuellere und räumlich verbesserte Bevölkerungsda-ten zu entwickeln.

Trotz dieser Fortschritte hat eine gründliche Literaturstudie, die im Rahmen dieser Doktorarbeit durchgeführt wurde, gezeigt, dass die Mehrheit der bestehenden groß angelegten, gerasterten Produkte die bekannten Volkszählungsdaten in Raster aufschlüsselt. Daher hängt ihre Genauigkeit von der Qualität der Volkszählungsdaten ab. Andere Ansätze, die die Bevölkerung zusätzlich zur Bevölkerungsdisaggregation vorhersagen könnten, basieren in der Regel auf regionalen Daten, die von einigen wenigen Städten stammen. Dies schränkt ihre Übertragbarkeit auf eine andere geografische Region ein. Die Anwendung von Deep Learning-Ansätzen bei der Bevölkerungsschätzung ermöglicht effizientere Bevölkerungskarten. Aufgrund des Blackbox-Charakters dieser Modelle ist es jedoch von entscheidender Bedeutung, ihre Ergebnisse zu verstehen und zu erklären sowie die Schlüsselmerkmale aufzudecken, die das Modell verwendet, um sein Ergebnis zu erzielen. Die Auflösung der Bevölkerungsdatensätze ist nach wie vor eine Herausforderung in verschiedenen Disziplinen, die Bevölkerungsdaten mit hoher räumlicher Auflösung erfordern, wie etwa die Eindämmung von Infektionskrankheiten oder das Katastrophenmanagement. Insgesamt wurde festgestellt, dass die bei Bevölkerungsschätzungsstudien beobachteten Einschränkungen entweder a) auf die Nichtverfügbarkeit großer Datensätze, b) auf übertragbare und interpretierbare Methoden und c) auf Verbesserungen der räumlichen Auflösung der vorhandenen Bevölkerungsdaten zurückzuführen sind.

In diesem Zusammenhang untersucht diese Doktorarbeit die verschiedenen Eingangsdatenquellen, die gut mit der Bevölkerung korrelieren, und erstellt einen groß angelegten Benchmark-Datensatz für Bevölkerungsschätzungen unter Verwendung öffentlich verfügbarer Datensätze. Diese Datensammlung würde dazu beitragen, die oben beschriebene Einschränkung a) zu überwinden. Sie wird eine wertvolle Ergänzung der aktuellen Literatur und der Forschungsgemeinschaft sein, um neue anspruchsvolle Methoden für

Bevölkerungsschätzungsstudien zu entwickeln und verschiedene Ansätze zu vergleichen, die unter Verwendung regionaler Datensätze entwickelt wurden.

Unter Verwendung dieses Datensatzes wird ein auf Deep Learning basierendes End-to-End-Rahmenwerk entwickelt, das die Bevölkerung vorhersagt und nicht die Bevölkerungs disaggregation. Da die Methode die Bevölkerung auch ohne Volkszählungsdaten ableiten kann und nur offen zugängliche, groß angelegte Datensätze verwendet, ist sie leicht übertragbar. Um die Leistung der Methode zu bewerten, wird ein Vergleich mit einem anderen Community-Standardprodukt für einige Städte in Europa und den Vereinigten Staaten durchgeführt. Und um die Vertrauenswürdigkeit der Methode weiter zu fördern, wird eine erklärbare KI-Technik eingesetzt, um das Blackbox-Modell zu entschlüsseln und seine Leistung und Grenzen zu verstehen. Diese Methoden helfen dabei, die wichtigen Merkmale zu identifizieren, die das Modell für seine Entscheidung verwendet hat. Der Schwerpunkt liegt dabei auf der Genauigkeit der Methode, ihrer Übertragbarkeit und Interpretierbarkeit, um die Einschränkung b) zu beheben.

Schließlich konzentriert sich diese Arbeit auf die Verbesserung der Auflösung der Populationsdaten und geht damit die oben genannte Einschränkung c) an. Sie verwendet einen hybriden Deep-Learning-Ansatz, der gerasterte Bevölkerungskarten erzeugt und dann die geschätzten Bevölkerungszahlen auf Gebäude aufschlüsselt, was zu Bevölkerungskarten auf Gebäudeebene führt. Diese Studie integriert die Gebäudedaten, wie Gebäudefunktionen, Gebäudeflächen und -höhen, um die Genauigkeit und Auflösung der Bevölkerungsschätzungen zu verbessern. Sie führt auch einen qualitativen Vergleich von Gebäudedatensätzen durch, die aus öffentlich zugänglichen, groß angelegten Quellen und regionalen, frei zugänglichen administrativen Quellen stammen, und untersucht, wie sich die Datenqualität auf die Bevölkerungsschätzung auf feiner Ebene auswirkt.

Acknowledgements

First and foremost, I express my sincere gratitude to Professor Xiaoxiang Zhu for allowing me to pursue a doctorate under her guidance and supervision. I am thankful for her unwavering academic support over the past four years. Her assistance and advice carried me through every phase of my work. Second, I want to express my sincere gratitude to Prof. Yuanyuan Wang and Dr. Matthias Kahl, for their invaluable help, support and guidance throughout the entire thesis work. They continuously monitored my work progress, gave insightful feedback on all my research articles, and provided me professional and personal guidance. This thesis won't be the same without their expertise, feedback, and proofreading. Also, I would like to thank Prof. Hannes Taubenböck for being my mentor and for his constant encouragement and insightful feedback. Next, I would like to thank Prof. Monika Kuffer for serving as an external examiner on my thesis committee and Prof. Dr. rer. nat. Martin Werner for acting as the chair of my doctoral defense. I appreciate their interest and effort in reviewing and examining this thesis.

Throughout my Ph.D. research, I was fortunate to collaborate with smart and dedicated individuals from a variety of disciplines. I am thankful to my colleagues at TUM SiPEO and DLR's IMF DAS departments for enhancing my Ph.D. life by exchanging great ideas during post-lunch walks, coffee breaks, virtual meetup and chit-chat during the pandemic. Special thanks to Dr. Andres Camero Unzueta for consistently organizing the opportunities to present and discuss the ideas within DLR. In the event of any organizational obstacles, Dr. Anja Rösel, Vasiliki Karasmanaki, Irene Danhofer, and Yingjie Schreiber-Liu were always available to help and special thanks to TUM management team, Dr. Julia Kollofrath and Dr. Simon Schneider for their endless patience and support.

Last, I want to thank my grandparents for their blessings, my parents, Asha and Ashok Doda, for their love and support in all aspects of life, my brother Sushant, and the rest of my family for believing in me. Finally, I want to express my profound gratitude to my partner, Parag, who had my back when I was anxious during this time, helped me make decisions, and provided me with unwavering support in all situations.

Contents

List of Figures

List of Tables

Acronyms

CD	Class Distance.
CE	Cross Entropy.
CIAT	Centro Internacional de Agricultura Tropical.
CIESIN	Center for Iternational Earth Science Information Network.
CNN	Convolutional Neural Network.
CPU	Central Processing Unit.
CRS	Coordinate Reference Systems.
DEM	Digital Elevation Model.
DMSP	Defense Meteorological Satellite Program.
EFTA	European Free Trade Association.
EO	Earth Observation.
ESSnet	European Statistical System.
FCL	Facebook Connectivity Lab.
FL	Focal Loss.
GHS-POP	Global Human Settlement Layer Population.
GHSL	Global Human Settlement Layer.
GPU	Graphics Processing Units.
GPW	Gridded Population of the World.
GRUMP	Global Rural-Urban Mapping Project.
GSD	ground sample distance.
GUF	Global Urban Footprint.
HL	Huber Loss.
HRSL	High Resolution Settlement Layer.
IFPRI	International Food Policy Research Institute.
IG	Integrated Gradients.
JRC	Joint Research Centre.
LCZ	Local Climate Zones.

LiDAR	Light Detection and Ranging.
MACD	Mean Absolute Class Distance.
MAE	Mean Absolute Error.
MMU	Memory Management Unit.
MSE	Mean Squared Error.
NPP	National Polar-orbiting Partnership.
OLS	Operational Linescan System.
ORNL	Oak Ridge National Laboratory.
OSM	OpenStreetMap.
PCI	Peripheral Component Interconnect.
PCIe	Peripheral Component Interconnect Express.
REE	Relative Estimation Error.
ReLu	Rectified Linear unit.
RF	Random Forest.
RMSE	Root Mean Squared Error.
SAR	Synthetic Aperture Radar.
SDGs	Sustainable Development Goals.
SEDAC	Socioeconomic Data and Application Center.
UN	United Nations.
UNPF	United Nations Population Fund.
UTM	Universal Transverse Mercator.
VIIRS	Visible Infrared Imaging Radiometer Suite.
WGS84	World Geodetic System.
xAI	explainable Artificial Intelligence.

1 Introduction

The world's population exceeded 8 billion in 2022, followed by 9.7 billion in 2050 [7]. Due to this rapid population growth, the proportion of the world's population living in urban areas is expanding dramatically [8]. This worldwide emerging trend of urbanization is changing the world, while new megacities are emerging in developing countries, some cities in Eastern Europe, including Poland, Romania, the Russian Federation, and Ukraine, are seeing population declines [8]. As seen in Figure 1.1, 70 % of the world's urban population presently resides in developing nations and the percentage will rise to 80 % by 2050 if the current trend holds. On the other side, the global urban population is declining in urbanized developed nations [1, 9]. A strategic response to this unprecedented urban population growth is imperative for humankind.

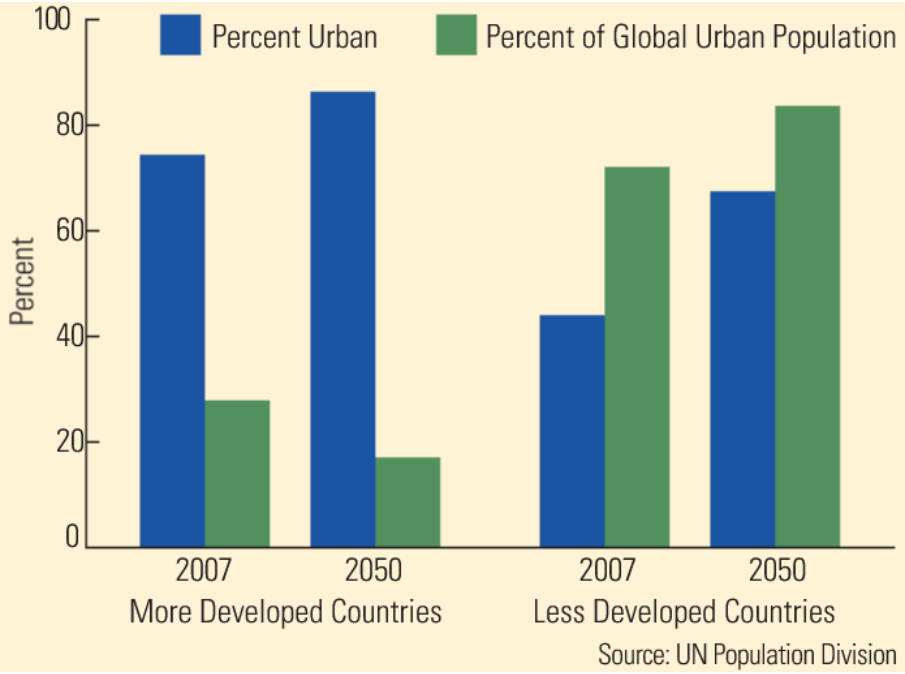


Figure 1.1: Percent Urban vs. Percent of Global Urban Population in developed and developing countries for the years 2007 and 2050 [1].

1.1 Motivation

Cities' population growth has changed as they have grown. This momentum of population growth puts a lot of pressure on the environment, particularly because of the increased demand for food, water, infrastructure, health facilities, and other resources. In 2015, the United Nations (UN) adopted 17 Sustainable Development Goals (SDGs) [10] to protect the planet and ensure good health, basic facilities, and prosperity for everyone. Population growth directly impacts the SDGs and could pose hurdles in achieving them. However, a good understanding of population distribution and projection could aid their accomplishment. Population estimation and distribution is crucial to many other disciplines, such as managing catastrophes, disease control, civil protection, etc., and influence a government's policy-making, planning, and fund allocation [4]. The national census has traditionally been undertaken to obtain population count, distribution, and demographic information. The government has extensively used this data to prepare for a region's future and essential services. Census data is generally collected once every decade, and in certain nations, once every few decades [11, 12]. Particularly in low-income nations with erratic governments, where the population is expanding quickly and unevenly, it becomes extremely difficult to conduct the national cenus [12]. For example, the first and only national census in Afghanistan was conducted in 1979 and due to security reasons, only 67 % of districts were covered [12]. The fast population growth rate means these estimates could be incomplete and outdated by a decade or possibly sooner. Furthermore, census enumeration zones limit the spatial resolution of population distribution, making it unsuitable for use in some applications such as earthquakes and floods [13]. Another challenge is the cost-effectiveness of the complete census process. The U.S. Census Bureau estimated the 2020 census cost around $14.2 billion [14]. According to the United Nations Population Fund (UNPF), "censuses are one of the most complex and massive peacetime exercises a nation undertakes" [15].

Remote sensing techniques, particularly optical satellite images, have grown in popularity recently and are an important independent source of data for studying population disaggregation and, more importantly, population prediction. More precise and rapid population estimation has been rendered possible by the availability of high-resolution satellite images and an upsurge in machine learning techniques [16, 17, 18, 19]. However, the actual ground truth data often do not exist [16]. As a result, most approaches have yet to be tested on a large-scale and instead have been built using data from only a few cities. Across all regions, they are lacking in sustainability and external review. Only a few studies have compared their estimations with the other population products [20, 21], however, the census data must be gathered and processed for these investigations. As a result, it is challenging and time-consuming to replicate the findings or contrast the methodologies [4].

Therefore, this thesis aims to develop a large-scale population estimation data set and fine-resolution population estimation framework that solely employs publicly available data sets to estimate the population on a large-scale with unparalleled quality.

1.2 Objectives

To accomplish the above-mentioned overall objective of this research, the following individual goals are noted:

- **Population Estimation Data Set:** This dissertation shall provide a large-scale population estimation benchmark data set created from publicly available data sets. This data set would save the cost of gathering and processing a new data set to develop and validate the methods in this domain. It would pave the way for the development of powerful statistical and machine-learning methods for population estimation. It would also lay the foundation for future research into a variety of additional urban related applications.
- **Deep Learning for population estimation:** This dissertation aims to design a deep learning method that only uses openly available Earth Observation (EO) data for population estimation. The method should improve the accuracy of the state-of-the-art large-scale gridded population products. Additionally, explainable AI methods are utilized to enhance the transparency and confidence of the proposed model.
- **Building level population estimation:** Additionally, this dissertation aims to improve the resolution of the existing population products by proposing a method for mapping the population to individual buildings using only the open-access data set. The approach is being developed and evaluated in two rapidly expanding Bavarian cities. This work also investigates and compares the suitability of regional vs. large-scale open-access data sets for fine-scale population estimation.

1.3 Structure of Thesis

While this chapter highlighted the need for population estimation as well as the thesis' objectives, the next chapter, Chapter 2, gives background information that aids in comprehending the fundamentals of population estimation methods and deep learning. Using this information, Chapter 3 provides a summary of related research in the areas of population estimation, existing population products, remote sensing, and machine learning in urban environments. Our benchmark data set for population estimation is introduced in Chapter 4. The deep learning techniques for reliable, large-scale population estimation using Earth observation data are described in Chapter 5. In Chapter 6, a study for estimating the population at the building level is provided. The techniques, outcomes, and opportunities for further research are discussed in Chapter7.

2 Fundamentals

This chapter provides an overview of the methods employed in this thesis. It begins with an overview of remote sensing and its applications in urban environments. Following that, population estimation methods are discussed. Finally, the deep learning approaches employed in this study are described.

2.1 Remote Sensing in Urban environment

For the majority of us, the word 'remote sensing' brings up the images from satellites. However, it is more than that. In an environmental context, it usually refers to a process of collecting electromagnetic radiation from distant objects, be it on the Earth's surface, oceans, or atmosphere and then analyzing its physical characteristics [22, 23]. The two main categories are active and passive remote sensing. In active remote sensing, the sensor directs its own radiation at the target and captures the radiation that is reflected from the surface. These sensors are important, for example, to assess the topography of the sea surface, ice, precipitation, and winds, among other things, and have the ability to penetrate the atmosphere under most circumstances. Synthetic Aperture Radar (SAR) systems like Sentinel-1 [24] and Light Detection and Ranging (LiDAR) sensors are typical examples of active systems. In passive remote sensing, electromagnetic radiation is sourced from surface reflections of the sun or ground-based emissions such as thermal or night lights, etc., and is measured by detectors on a remote sensing platform [22]. These sensors could gauge physical characteristics like vegetation characteristics, cloud and aerosol characteristics, land and sea surface temperatures, and more [25]. The Sentinel-2 [26] and Landsat [27] optical satellites are two popular examples that work passively.

The majority of remote sensing data consists of digital images acquired by active and passive sensors. A digital image comprises of a two-dimensional array of discrete pixels, each of which has an intensity value and a geographical address. In terms of computer science, a pixel's intensity is a numerical value representing the physical amount measured over the entire ground area that the pixel is covering. A pixel's logical address is a one-to-one relationship between its column-row address and the location's coordinates (such as longitude and latitude) [28]. Therefore, each pixel is associated with a specific geographic area on the ground, and this measurement of the geographic area on the ground that a pixel represents [29] is known as spatial resolution or ground sample distance (GSD). For instance, Landsat-7 [30] acquires imagery at a 15 m resolution, which means that each pixel in its imagery corresponds to a 15 $\times$ 15 m grid cell on the ground [31]. The higher the spatial resolution of the imagery, the smaller the area it covers on the ground and the better its ability to discriminate between objects.

The physical characteristics of objects on Earth determined from remote sensing data could be used in a variety of research areas, including sustainable development [32], natural hazards research [33], environmental study [34], the impact of climate change [35], and land use mapping [36], among many others. Remote sensing could be utilized in urban settings to monitor and manage the urban environment, supporting the planning processes for resilient and sustainable cities. High-resolution imagery from Earth observation satellites, for example, could aid in the identification, monitoring, and capture of a variety of urban environmental variables [37], the morphology of formal and informal urban settlements [38], detect the presence of human settlement [39], and so on. This information is critical for estimating urban population and contributing to achieving the UN's sustainability goals. Following the increase in remote sensing data, new methods especially deep learning algorithms are becoming more accepted in the field of urban remote sensing [40, 41, 42]. More information and insights into deep learning are provided in section 2.3.

2.2 Population estimation techniques

A population estimate calculates the number of people living in a census or administrative unit. The population has traditionally been measured through a national census. Censuses have a long history, dating back to 3800 BC when they were first used to count the number of people, cattle, quantities of butter, honey, milk, wool, and vegetables, to the modern population census, which took place in Canada in the years 1665-1666 [43]. Aside from the primary goal of giving a total enumeration of a nation's population, it also provides critical information on its spatial distribution, age and gender structure, and other vital social and economic features [44]. Population estimation is sometimes mistaken for population projection. Population projection predicts the future population size based on multiple factors such as current population count, the other population growth metrics such as the birth and mortality rate, immigration, and urbanization, etc. [45].

Census data is often provided at the aggregated level to protect people's privacy, which is typically an administrative unit [46]. Hence, the data quality depends on the number and size of administrative units, which vary significantly across and within nations. Because of this lack of consistency, census data cannot be readily integrated with other data sets and used in large-scale analyses. Alternatively, much work has been done to convert vector-based census data to grids with uniform spatial resolution [47] using an interpolation algorithm. It has two key advantages: first, it makes it simple to integrate population data with other geospatial data, and second, it makes population data comparable and uniform across regions. Interpolation algorithms usually translate the data from coarse, high-level geographic areas to consistent fine-scale or areas with comparable scales but different boundaries [48]. The accuracy of the interpolation method is determined by the relative size and homogeneity of the two zonal sets, the method's generalization, and the correlation between the variables used, among other factors [49]. Many studies used the areal interpolation method for population estimation. Areal in-

terpolation is the process of making estimations from one spatial unit, the source zone (known values), to another, known as the target zone (unknown values) [50]. Areal interpolation algorithms are further classified into two types based on whether auxiliary information is employed [49] or not for the interpolation.

Areal Interpolation without Ancillary Information: It is an interpolation method used when population data from the source zone is available, and there are no constraints for its spatial reallocation into the target zones [51]. The areas of intersection between the source zone and the target zone are used to proportionally reallocate the population counts from the source units to target units [50]. It keeps the population count or volume preserved, which implies that the aggregate of population counts from all target units equals the original population total of the source unit. One drawback is that it makes an assumption that is rarely true in reality: geographic homogeneity, which means that the same number of persons would be assigned to each target unit inside a source unit [52]. Since the interpolation is based only on the geometric properties of the source and target zone, therefore, its quality depends on the accuracy and spatial resolution of the source zone population data [53, 54]. However, in the lack of auxiliary information, it is still a viable option and could be easily integrated with other geospatial data sets without restrictions [55]. Figure 2.1(a) depicts areal interpolation without the use of auxiliary data, in which the population counts from the source zones are uniformly redistributed to the target zones, assuming spatial homogeneity.

Areal Interpolation with Ancillary Information: This method tries to overcome the limitation of the areal weighting by using an auxiliary layer that constrains how source zone data is allocated to the target zone. A correlation between the additional ancillary data and the information being interpolated results in a more accurate allocation and depiction of the real distribution [56]. The population allocation within target zones is generally guided by remote sensing data such as satellite imagery, digital surface models, night lights, land use and land cover data, other socioeconomic characteristics and more, which significantly correspond with the population. Figure 2.1(b) depicts the impervious layer in the middle being utilized as auxiliary data to mask out the unpopulated target cells (water and road) and distribute the population counts to the rest of the target zones depending on their residential proportions. Despite the fact that this mapping might offer a more spatially informed interpolation, the application of such methodologies places additional demands on the availability of ancillary data [57]. However, the availability of publicly accessible new forms of data, such as volunteered geographic information and social media data, opens up fresh opportunities for performing informed areal interpolation [48]. One of the most used interpolation methods using ancillary information is dasymetric mapping [46] and based on the modeling technique, it could be further divided into two categories: binary and intelligent dasymetric mapping [58].

Binary dasymetric mapping [59] is a basic method that divides a source zone into two zones, usually populated and unpopulated. The binary layers that are widely used

include water bodies, natural or protected areas, build-up regions that mask out the unpopulated zones and all of the source zone population is assigned to the populated areas. A major disadvantage of this strategy is that it allocates the people equally to all populated regions, whether rural or urban, and there will be no population allocation if the populated area is misclassified, which will have an impact on the quality of the population distribution.

Intelligent dasymetric mapping [60] uses one or more ancillary data to calculate a weighted layer that determines how much population to allocate to each grid cell within a source zone. These weights reflect the correlation between population counts and geospatial data. While preserving volume, it allocates population counts to target cells in a heterogeneous manner, resulting in more realistic spatial and accurate distributions [61]. Usually, this complex relationship is modeled using a variety of statistical [62] and machine-learning methods [63]. Currently, deep learning approaches [16, 18] are being utilized to automatically identify abstract features from auxiliary data and optimize the dasymetric distribution of the population from source zones to target zones. The complexity rises as more variables are incorporated. As a result, it's crucial to use the right ancillary data when choosing population spatial distribution indicators [58].

2.3 Deep Learning

Deep learning is a subfield of machine learning that is motivated by the structure of a human brain. It automatically extracts the low and high-level features to learn the multiple levels of representations [64]. Recently, they have been great success and used in various fields such as healthcare [65, 66], natural language processing [67, 68], self-driving cars [69, 70], virtual assistants [71], and many more.

Computer vision is one of the most well-known fields where deep learning techniques are successful. It has been employed on various computer vision tasks, including semantic segmentation [72, 73], pose estimation [74, 75], face recognition [76, 77], and object detection [78, 79]. With the increased availability of data and computational resources, deep learning is finally taking off in remote sensing as well. The handcrafted geometrical and textural aspects of the images were not as robust in traditional remote sensing methodologies [80]. As a result, deep learning has proven to be a novel and intriguing method for handling large-scale raw big data in remote sensing applications [40, 41, 81].

While the majority of deep learning methods are well known for classification or detection tasks, most remote sensing problems aim to predict continuous values [40]. The deep learning architectures used for classification typically have numerous convolutional layers, often followed by a few fully connected layers and a classification softmax layer [82]. The overall architecture is referred to as a Convolutional Neural Network (CNN). For regression analysis, a fully connected regression layer with linear or sigmoid activations is frequently used instead of the softmax layer [82]. A brief overview of the basic components of neural networks is presented in the following section. More theoretical

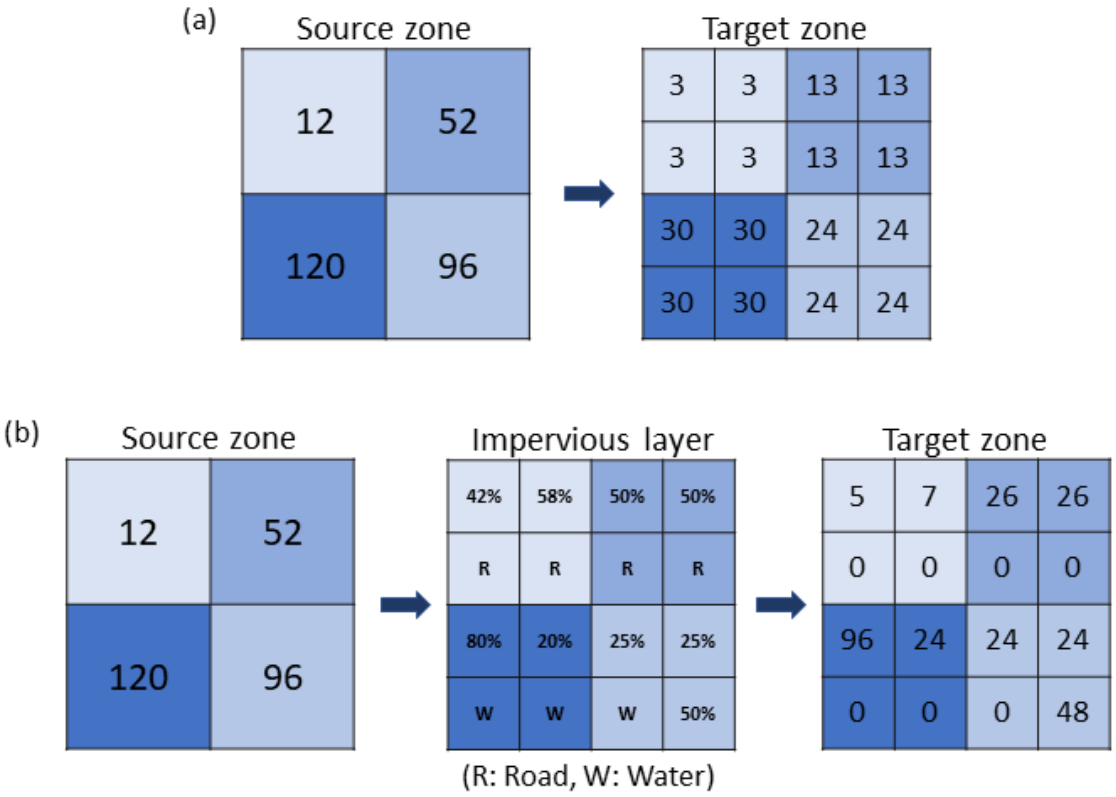


Figure 2.1: population mapped from source to target zones (a) without any ancillary data and assuming an equal distribution to all the target zones (b) with ancillary data, the impervious layer in the middle indicates the unpopulated (water and road) and populated cells with residential proportions. This information is used for the weighted redistribution of population counts from source to target zones.

details about deep learning can be found in Goodfellow et al. [83]. Next, discussed the state-of-the-art deep architectures used in this work.

2.3.1 Components of Neural Network

2.3.1.1 Layers

Convolutional Layer: Our visual cortex served as inspiration for a convolutional layer. It utilizes various filters to slide over the whole input image to generate a feature map. A model can extract and learn the underlying features from a previous input by employing a number of these filters. Filters in the first layers detect simple features, and as the network progresses, filters may detect increasingly complex features. The convolution operation is a critical component of deep learning architectures.

Fully connected Layer: It is a fundamental layer of deep learning architecture, often known as the dense or feed-forward layer. It connects each input of the preceding layer

to each output of the subsequent layer, as the name implies. It flattens the output received from the previous layer and passes it on to generate the final output.

Pooling Layer: The pooling operations, also called subsampling or downsampling, are used to reduce the spatial dimensions of the data. It aggregated the information of nearby features to a single feature as they are likely to contain the same information. The most common strategies are average pooling and max pooling. Max pooling uses the maximum value, and average pooling takes the average of the values in the feature map to keep only the relevant features.

2.3.1.2 Activation Functions

It is an additional function that neural networks use to learn complicated patterns from data instead of just learning linear relationships. Using a non-linear function, it can transform the output from the previous cell into different forms that can be utilized as input to the following cell. It is applied to all neural network nodes except the input and output nodes, and the activation function selection significantly impacts the model's performance. In different parts of the network, different activation functions could be applied. For example, the most typical activation function for hidden layers is Rectified Linear unit (ReLu). For the output layer, the right activation function depends on the predictions made by the network.

2.3.1.3 Objective Functions

A cost or loss function are two common names for the objective function. Configuring a neural network requires an objective function that calculates the model error. The objective function could maximize or minimize while optimizing the network's learning. The methodology used to determine the error has a significant impact on the loss function that is used. The literature has put out a number of different objective functions for the classification and regression tasks. Here are a few that have been used in this work.

Classification functions: Classification involves discrete and mutually exclusive predictions, which could be labels or categories. The following are two objective functions used for classification analyses in this study:

1. Cross Entropy (CE) Loss: It is also known as softmax loss and is typically used for classification tasks. In classification, the neural network outputs a vector of probabilities over the pre-defined categories (classes) and then the category with the highest probability is selected for the given input. It measures the loss by calculating how far away the actual distribution is from the target distribution and mathematically, it is defined as follows, where Y is the true class and p is the predicted probability vector.

$$CE = -\sum_{i=1}^{n} Y_i log(p_i) \tag{2.1}$$

2. Focal Loss (FL): Cross entropy loss does not perform well when there is a class imbalance by favoring the majority class and failing to pay attention to the hard examples. Focal loss solves this problem by introducing down weighting, which ensures the model improves over hard examples over time rather than on the ones it can predict confidently. Mathematically, it is achieved by adding a tunable focus parameter called gamma (γ) to the CE.

$$FL = -\sum_{i=1}^{n} (1 - p_i)^{\gamma} log(p_i) \tag{2.2}$$

Regression functions: Regression makes continuous and real-valued predictions. Therefore, the goal of regression analysis is to reduce the difference between the predicted and the actual value. The loss functions used in this study's regression analyses are listed below.

1. Mean Squared Error (MSE): It is a default loss function used while training the linear regression models. As all error terms are squared, it results in a substantially more severe penalty for large errors than for small ones. It measures the mean of squared differences between the actual value and the estimated value, mathematically defined as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\ y_i - \hat{y_i})^2 \tag{2.3}$$

2. Mean Absolute Error (MAE): To evaluate the efficacy of a regression model, MAE averages the absolute differences between the actual and anticipated outputs. When there are many outliers or extreme values in the training data, MAE is helpful because it pays more attention to the little errors and does not penalize the large errors as harshly as MSE. It is described as following where y_i donates the ground truth and $\hat{y_i}$ denote the prediction:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\ y_i - \hat{y_i}| \tag{2.4}$$

2.3.1.4 Regularization

Regularization is a technique to improve the generalization of the deep learning model. Model generalization is the ability of deep learning models to perform good not only on the observed data but also on real-world unseen data. Overfitting transpires when a model performs better on training data than on testing data and consequently does not

generalize well to new data. Deep neural networks are especially prone to overfitting because of their large number of parameters. In the literature, several strategies have been presented to handle the overfitting [84, 85, 86, 87]. Here are some commonly used strategies:

Data Augmentation: Increasing the number of training data is the easiest technique to prevent overfitting, but labeled data acquisition is usually difficult and expensive. In this case, incorporating transformations into the training data, such as scaling, random variation, noise, rotation, and flipping of the image could aid in increasing the size of the data and preventing overfitting on a particular set of observed data.

Early Stopping: The training loss may continue to diminish after a certain point while the validation loss rises. Therefore, in order to avoid overfitting, the model training should be halted right away using the early stopping when the performance of the model does not further improve on the validation set.

Weight decay: It is also known as L2 regularization as it penalizes the model parameters with the L2 norm. By penalizing the large weights in the network, it prevents overfitting and pushes the model to learn simpler functions and makes it less likely to overfit the training set of the data.

Dropout: It is one of the most popular regularization methods in the deep learning community. This method randomly removes neurons from the neural network during each iteration in training, together with all their incoming and outgoing connections. Each iteration produces a unique set of outputs, resulting in different architectures in parallel. This probabilistic node dropping introduces randomness into the model, making it more robust.

2.3.1.5 Evaluation Metrics

To evaluate the performance of a trained neural network, different types of metrics could be used depending on the task, Classification or Regression.

Classification Metrics: As classification tasks produce discrete results, classification metrics are designed to compare discrete classes; nevertheless, they can be evaluated differently.

1. Confusion Matrix: Confusion Matrix is a table-based depiction of actual labels versus predicted. Each row of the confusion matrix represents an example in a predicted class, while each column represents an occurrence in an actual class. The table is filled by counting how the test set samples are predicted. Figure 2.2 represents a confusion matrix for binary classification, implying two classes. All the successfully predicted samples are on the diagonal, whereas all incorrect predictions are in the other cells.

- True Positive (TP_{c_i}): how many samples of a class c_i are correctly predicted.
- True Negative (TN_{c_i}): how many samples of second class c_j are correctly predicted.
- False Positive (FP_{c_i}): how many samples of other class c_j are incorrectly predicted as class c_i.
- False Negative (FN_{c_i}): how many samples of class c_i are incorrectly predicted as other class c_j.

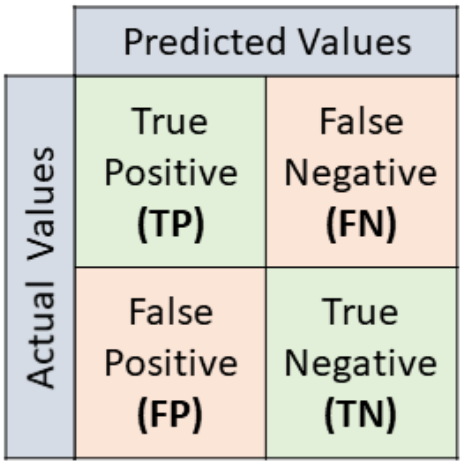


Figure 2.2: Example of a confusion matrix

These counts could be used to determine the precision and recall measures.

Precision indicates how often it correctly predicts a class for the sample. Therefore, it is calculated by taking the total number of correctly predicted samples (true positives) divided by the column sum (true positives & false positives). In the example above, the precision (P_{c_i}) of a class c_i is defined as:

$$P_{c_i} = \frac{TP_{c_i}}{TP_{c_i} + FP_{c_i}} \tag{2.5}$$

Recall measures how accurately a model is able to predict a correct class and is calculated as the number of correctly predicted samples divided by the row sum (true positives & false negatives). Again, from the example above, recall (R_{c_i}) of class c_i is defined as:

$$R_{c_i} = \frac{TP_{c_i}}{TP_{c_i} + FN_{c_i}} \tag{2.6}$$

2. Accuracy: It is the simplest and most commonly used metric to assess the overall performance of the neural network. It is calculated by dividing the number of correct predicted labels by the total number of predictions [88].

$$Accuracy_{c_i} = \frac{TP_{c_i} + TN_{c_i}}{TP_{c_i} + FP_{c_i} + FN_{c_i} + TN_{c_i}} \tag{2.7}$$

3. F1-score: The F1-score utilizes a classifier's precision and recall and is calculated by taking their harmonic mean. It limits the number of false positives and false negatives. Its range is [0,1]; the greater the F1 score, the better the performance.

$$F1_{c_i} = \frac{TP_{c_i}}{TP_{c_i} + \frac{1}{2}(FP_{c_i} + FN_{c_i})} = 2\left(\frac{precision_{c_i} \times recall_{c_i}}{precision_{c_i} + recall_{c_i}}\right) \tag{2.8}$$

4. Balanced Accuracy: It is a better metric to use with imbalanced data when the instances of one of the target classes are much more than the others.

$$BalancedAccuracy_{c_i} = \frac{1}{2}\left(\frac{TP_{c_i}}{TP_{c_i} + FN_{c_i}} + \frac{TN_{c_i}}{TN_{c_i} + FP_{c_i}}\right) \tag{2.9}$$

Regression Metrics : Regression models produce a continuous result. Therefore, to evaluate the performance of a regression model, there is a need of a metric that gauges the difference between predicted and observed values. Following are some population regression metrics that have been employed in this work:

1. Mean Absolute Error: It measures the average discrepancy between the actual values and the forecasts. As MAE does not differentiate between high or low errors, it is more resistant to outliers. It increases linearly as the magnitude of the error increases and since it uses the absolute value of the error without considering their direction, it cannot determine whether the model is under- or over-predicting. It is mathematically expressed as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\ y_i - \hat{y_i}| \tag{2.10}$$

where y_i donates the ground truth and $\hat{y_i}$ denote the prediction.

2. Root Mean Squared Error: It is one of the regression model's primary performance indicators. By calculating the square root of the errors, it is an extension of MSE. Root Mean Squared Error (RMSE) measures the deviation of the predictions from the true values; the larger the difference, the higher the penalty. Mathematically, it is calculated by taking the square root of the mean of all the squared errors and formulated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.11}$$

where y_i donates the ground truth and $\hat{y}_i$ denote the prediction.

3. Coefficient of determination (R^2): R^2 measures how much of a dependent variable's variance is explained by the independent variable in a regression model [89]. It gives a squared correlation between the predicted values and actual values. The higher the R-squared value, the better the model fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{2.12}$$

where y_i donates the ground truth, $\hat{y}_i$ denote the prediction and $\bar{y}_i$ is the mean of y.

2.3.2 State-of-the-art architectures

The neural network's history began in 1943 when Walter Pitts and Warren McCulloch developed a mathematical model of a neural network [90], followed by the introduction of convolutional operations in neural networks in 1980 [91]. The true evolutionary stride for deep learning occurred in 1999 when computers began to process data faster, and Graphics Processing Units (GPU) were introduced, increasing computational speeds by 1000 times [92]. AlexNet [93], a CNN architecture that won multiple international competitions in 2011 and 2012, is one early successful example. However, the architectures are still refining and evolving. This section explains the two state-of-the-art architectures used in this thesis.

2.3.2.1 VGG16

VGG-16 is a 16-layer deep CNN, which means it has sixteen learnable layers [94]. It improves on AlexNet by adding depth and replacing large filters with small convolution filters (3×3) with a stride of one pixel. The advantage of using multiple convolution layers with smaller filters over one with bigger filters is that there will also be multiple non-linear activation layers with convolutional layers rather than just one, allowing the network to converge faster. The Figure 2.3 illustrates the architecture. Following the convolution stacks are three fully connected layers, two of size 4096 and one of size

1000, corresponding to the total number of ImageNet classes. The final output layer has softmax activation.

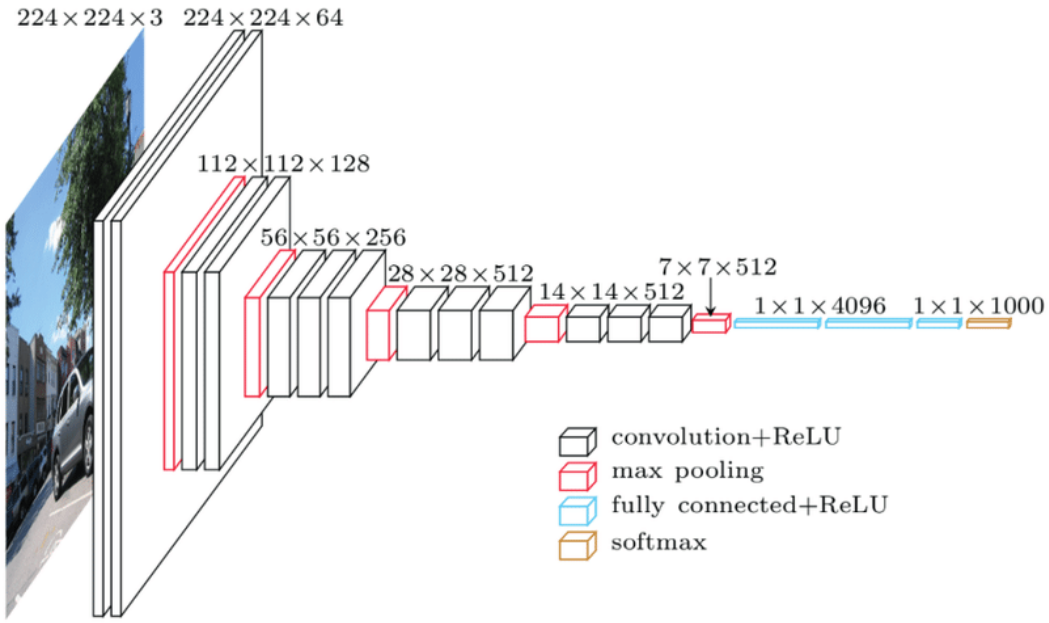


Figure 2.3: VGG16 network architecture. Source [2].

2.3.2.2 ResNets

Beyond a certain number of layers, the idea of increasing depth with smaller convolutional filters and improving performance does not scale. Every step of backpropagation makes the gradients from the loss function smaller as the depth is increased. This issue is known as the vanishing or exploding gradient problem, which is fixed in ResNet [3]. ResNet is constructed from Residual Blocks, which use skip connections between the layers to give the gradients a different and faster route to flow. The skip connection makes it possible to train significantly deeper networks by adding the outputs from the previous block to the current block. Figure 2.4 illustrates a residual block.

The most significant change to understand in the Figure 2.4 is the skip connection or identity mapping. As can be seen, the output from the preceding layer flows not just to the layer ahead but also makes a hop and is fed to another layer down the architecture, resulting in the residual block taking the input x and generating f(x) + x as output. In a residual block, the most commonly utilized activation function is ReLu.

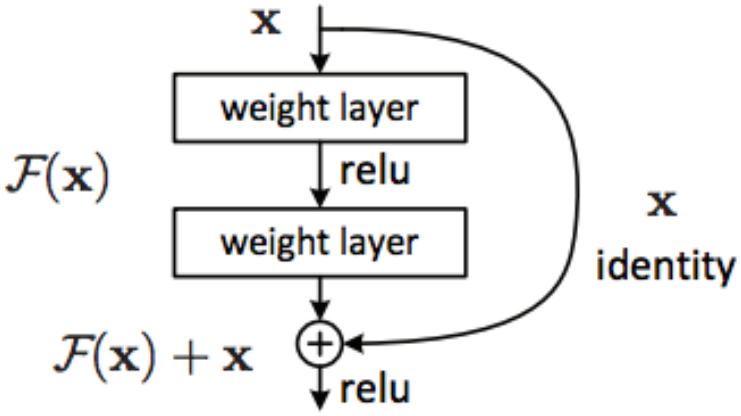


Figure 2.4: Residual block architecture. Source [3].

3 Related Work

This chapter gives an overview of the relevant work that has been done in the field of population estimation. First, the review of different population products is discussed, followed by the role of remote sensing in population estimation and different methods based on that.

3.1 Review of population data sources

A wide number of disciplines, including good governance, policy formulation, development planning, and many other SDGs, depend on population data. Traditionally, population estimation has been carried out through a census, in which information about a nation's populace and its spatial distribution within an administrative unit is carefully collected and assembled. However, censuses are not frequently conducted as they are costly and time-consuming [95]. Additionally, it still lacks complete information for some nations. There is an estimate that 350 million individuals globally remain unaccounted for in national censuses, particularly in poorer nations where informal settlements are not tallied [96]. Agarwal et al. in India demonstrated that the official population estimates for many Indian cities are false because they underestimated to take into consideration unrecognized informal settlements [97].

A city, district, or province is usually the census unit that serves as the spatial resolution for a national census. As a result, the majority of census data are provided as low-resolution data, which masks the fine-scale population dynamics and limits their use for extensive applications. Statistics that are based on the nation as a whole may conceal the fact that those who need help the most often fall behind [98, 99]. Despite these limitations, censuses are conducted and regarded as one of the most expensive government statistics that could go up to several billion US dollars for a country. Efforts have been made to create gridded maps at higher resolution out of the national census data by using different algorithms and variables, resulting in varying-quality population distribution maps. A few large-scale gridded data packages have been outlined in the subsections that follow.

3.1.1 GPW

Gridded Population of the World (GPW), now in its fourth version (v4), is a gridded data product on population count and density which are consistent with the national population census or alternative sources where no census data is available. It is developed by Center for Iternational Earth Science Information Network (CIESIN), Columbia University and accessible on a worldwide scale with the World Geodetic System (WGS84)

geographic reference system at a spatial resolution of 1 km. The population disaggregation method employed by GPW is unmodeled and based on the water mask. It simply uses the areal weighing scheme to disaggregate the population only to the land pixels.

Since the census units were utilized as the input, the quality of the data differs significantly among countries. In another version, the population counts are matched with the UN adjusted population totals from the World Population Prospects: The 2015 Revision. Data is available as raster data and contains the population counts over the years 2000, 2005, 2010, 2015, and 2020. It could be freely downloaded and accessible in both GeoTIFF and ASCII (text) formats at `https://sedac.ciesin.columbia.edu/data/collection/gpw-v4`. Technical details and input data sources utilized in different regions are also available.

3.1.2 GRUMP

Global Rural-Urban Mapping Project (GRUMP) data set is produced by CIESIN in collaboration with International Food Policy Research Institute (IFPRI), The World Bank, and Centro Internacional de Agricultura Tropical (CIAT) to create the population count grids for the years 1990, 1995, and 2000 [100]. It is built on the GPW population data collection. It uses the binary dasymetric approach to allocate the population counts to urban and rural areas as derived from the night light imagery [101]. To identify urban settlement points, the night-time satellite images of a city from the National Oceanic and Atmospheric Administration [102] are used. It is also available in WGS84 at a spatial resolution of 1 km. The data is openly available at https://sedac.ciesin.columbia.edu/data/collection/grump-v1

3.1.3 GHS-POP

To provide the population count and density distribution at a finer grid level, the European Commission Joint Research Centre (JRC) and CIESIN, Columbia University created GHS-POP. The population information used to disaggregate the population count to grid cells is utilized from the CIESIN-produced GPWv4. It employs the weighted dasymetric approach, which distributes the population proportionally to the grid cell's built-up density in relation to the entire cell area. The GHS-BUILT data collection is used to calculate built-up density. It aims to refine the dasymetric mapping by including built-up regions that are directly related to the population, producing precise population distributions. It is accessible in the global Mollweide projection at two spatial resolutions, 1 km (GHS-POP 1 km), and 250 m (GHS-POP 250 m). The grids are built using open and publicly available input data and are, thus, freely downloadable at `https://ghsl.jrc.ec.europa.eu/ghs_pop2023.php`. Their detailed disaggregation algorithm can be found in their scientific paper [103].

3.1.4 LandScan

The ambient (average day/night) population distribution at a worldwide level is represented by LandScan from Oak Ridge National Laboratory (ORNL) [104]. LandScan

disaggregates census counts to the grid cells at a geographical resolution of 30 arc-seconds (approximately 1 km) as a function of census demographic data, additional geographic data, and remote sensing imagery using a multivariate dasymetric modeling framework [105]. It's a smart interpolation approach that is dynamically adaptive to the various input data available in each country. For each country, it uses the set of sub-national level census data, high resolution satellite imagery, and local supplementary data which includes land cover and land use, urban and suburban regions, and street networks as key indicators to map the population. The population distribution model uses these regional statistics to calculate the probability coefficient for each cell and the population of the area is distributed to each cell proportionally to the determined probable population coefficient. LandScan Global is freely available for educational purposes and downloadable at `https://landscan.ornl.gov/`.

3.1.5 WorldPop

The WorldPop project by the University of Southampton generates gridded population count products as well as a range of other demographic data on a worldwide scale [106]. By combining the three regional population mapping products: AfriPop [107], AsiaPop [108], and AmeriPop [109] - the WorlPop program was launched in 2013. It uses the weighted intelligent dasymetric mapping to create a weighting layer that dasymetrically redistributes the recent census-based population counts to 100 x 100 m grid cells in the geographic projection WGS84.

The most recent official census data on population is one of the input variables for WorldPop, along with a wide range of supplementary spatial data sets. The spatial databases include information on the whereabouts and sizes of settlements, as well as nighttime lights, maps of buildings and roads, locations of medical facilities, vegetation, and refugee camps. Then, a predictive weighting layer is created using a RF [110] regression tree-based mapping technique to reallocate population counts into gridded pixels. The data set offers gridded annual population data for the years 2000 to 2020 and is freely downloadable on their project website (`https://www.worldpop.org/project/list`). All the input data and covariates utilized in model prediction are also openly accessible.

3.1.6 HRSL

High Resolution Settlement Layer (HRSL) is a Facebook Connectivity Lab (FCL) and CIESIN project that provides human population distribution across 140 countries. The population grids are provided in the WGS84 geographic reference system with a spatial resolution of approximately 30 m. To identify the populated regions, DigitalGlobe's very high resolution satellite images (0.5 m) served as an input to machine learning models that identified the settlements in the region. Population counts from the most recent national census have been distributed proportionally to the identified settlements in the grid cell. More technical information about their technique and accuracy analysis

can be found in their scientific article [111]. The population grids are accessible and downloadable at `https://www.ciesin.columbia.edu/data/hrsl/`.

3.1.7 GHS-POP-EUROSTAT

European Commission Joint Research Centre put an effort to produce the detailed population grids in Europe, GHS-POP-EUROSTAT [112]. This data set illustrates the residential population distribution and density at a spatial resolution of 100 m in equal-area projection (LAEA ETRS89). It employs intelligent dasymetric mapping [60] to disaggregate each country's 2011 census count to built-up areas informed by European Settlement Map and scaled by the land use and land cover extracted from Corine Land Cover Refined 2006 [113], as well as the distribution and density of settlements as defined in the European Settlement Map layer [114, 115]. This data product can be found and downloaded for free at `http://data.jrc.ec.europa.eu/dataset/jrc-ghsl-ghs_pop_eurostat_europe_r2016a`.

Table 3.1 outlines the similarities and differences between these data products based on their corresponding methods, resolution and sources.

Global			
Product	**Resolution**	**Source**	**Method**
GPWv4	1 km	CIESIN	areal weighting
GRUMP	1 km	CIESIN, IFPRI, The World Bank, CIAT	dasymetric
GHS-POP	250 m	JRC, CIESIN	refined dasymetric
LandScan	30 arcsec	ORNL	smart interpolation
WorldPOP	100 m	University of Southampton	dasymetric
Regional			
HRSL	30 m	FCL, CIESIN	binary dasymetric
GHS-POP-EUROSTAT	100 m	JRC	intelligent daysmetric

Table 3.1: Summary of the popular gridded products at the global and regional levels, together with their methods and sources utilized.

3.2 Remote Sensing in population estimation

The use of remote sensing has been continuously explored in the population estimation literature due to its ability to gather information about large geographic areas quickly and efficiently. The widely used remote sensing sources in the population estimation studies are discussed below:

a) Optical Satellite Imagery: Optical sensors record the reflected infrared and visible light from the Earth's surface. These images are frequently employed in population estimation studies since they offer high resolution visual representations of the landscape [116, 117, 118]. Optical satellite data is particularly helpful for locating built-up areas, urban centers, and human settlements. Some prominent sources consistently used in the literature include Landsat Program [119, 120, 121], Sentinel-2 [53, 122], Maxar [123, 124], and SPOT [125, 126].

b) Synthetic Aperture Radar: SAR sensors utilize microwave signals to penetrate clouds and provide all-weather, day-and-night imaging. SAR data is useful for tracking changes in land cover and urban growth, two important factors in population estimation. SAR data has been used in several research studies for population estimation. Henderson et al. demonstrated the potential of SAR data in urban studies focusing on settlement detection and population estimation [127]. Esch et al. investigated the TanDEM-X SAR satellite capability to analyze and monitor human settlement patterns [128].

c) Nighttime Light Data: Nighttime lights are often acquired by satellites equipped with specialized sensors. It has been widely used in the literature to estimate population density by correlating light intensity with human activities. The intensity of nighttime lights serves as a proxy for human presence and economic activities, particularly in urban areas. Areas with brighter nighttime lights tend to have higher population densities, while areas with darker or less intense lights indicate lower population densities. Several works utilized different night light data sources. Defense Meteorological Satellite Program (DMSP)-Operational Linescan System (OLS) and the National Polar-orbiting Partnership (NPP)-Visible Infrared Imaging Radiometer Suite (VIIRS) are two widely used nighttime lights data in population estimation studies [129, 130, 131].

e) LiDAR: LiDAR data offers three-dimensional information that helps characterize and identify structures and vegetation. It has mostly been used to extract building footprints and heights, which are essential for precise population estimation [129]. Although LiDAR data has much to offer for population estimation, there are some drawbacks, such as the expense of data collecting and processing. Additionally, not all places may have easy access to LiDAR data, particularly in developing nations [132]. Despite these difficulties, combining LiDAR data with additional remote sensing and demographic data sources can boost the accuracy and utility of population estimation models, especially in urban settings and places with complicated topography.

3.3 Machine Learning in population estimation

Machine learning approaches have lately sparked renewed interest in remote sensing [40] and in a variety of other related fields, including population estimation, because of their ability to handle complex data, adaptability, and learning non-linear patterns in the data. One often used method in the literature is the RF model to determine the weighted dasymetric mapping scheme for population disaggregation. Recent research has employed deep learning models, particularly CNN to directly predict the population counts from the remote sensing data instead of disaggregating the known census counts. They have demonstrated promising results in extracting significant characteristics from a variety of data sources such as remote sensing imagery, ancillary data, and other socioeconomic indicators to create more precise and dynamic population estimates.

Stevens et al. [63] estimated the population of Vietnam, Cambodia, and Kenya at $\sim 100\,\mathrm{m}$ resolution using a RF approach. They used census data from Cambodia's National Institute of Statistics, Vietnam's National Statistics Office, and Kenya's National Bureaus of Statistics, as well as a variety of other remotely sensed and geospatial data sets such as nighttime lights, road network, health facilities, elevation models, land cover, vegetation, and built-up regions. Hara et al. [133] trained a RF regressor using social media data to estimate the population in a specific region. Doupe et al. [19] proposed a method that combined Landsat-7 satellite data with (DMSP/OLS) nighttime lights to estimate the population using a CNN. They trained their model with data from Tanzania at a resolution of 250 m and approximated Kenya's population at an 8 km resolution. Robinson et al. [16] proposed yet another CNN-like method. They estimated the population in US counties at a 1 km resolution using US census summary grids and Landsat data. Hu et al. [18] suggested a deep learning technique for determining population density in India at village and subdistrict level by combining Landsat-8 and Sentinel-1 satellite data with the Socio-Economic Caste Census survey. Huang et al. [134] trained a deep learning model with existing population grids from LandScan [104] to map population changes in two US cities using a variety of different state-of-the-art architectures. Metzger et al. [135] employed a deep learning model to perform population distribution at the spatial resolution of 100 m. When census data is unavailable, they could alternatively anticipate the population count using open geodata. Similarly, Georganos et al. [136] proposed a deep learning-based methodology to estimate the population in three Sub-Saharan nations at 100 m resolution using open building footprints and high resolution satellite data. Most of the above mentioned methods provide the population estimates at a coarser-scale and many researchers have revealed that their spatial resolution might not be sufficient for reaching a well-informed conclusion [20, 54].

As a result, recent research investigates population estimation at relatively fine resolutions. Fine-resolution gridded population data are critical for a variety of domains, including urban planning, resource allocation optimization [137, 138], natural disaster management [139, 140], public health [141], and as a foundation for various other applications. Zhou et al. [142] estimated the population distribution in Chongqing, Southwest China at 30 m spatial resolution using the RF regression approach with numerous data sources. Balakrishnan et al. [143] build a population estimation technique that gen-

erates population density maps at 30 m resolution using building information, known census data, and other socioeconomic variables. Considering where people live has a significant correlation with the buildings, building level population estimation would be the finest level source [144]. Only some studies have recently evaluated methodologies for calculating population counts at the building level [129, 132, 145, 146, 147]. These studies used high resolution satellite imagery or other supplementary data sources such as land use/land cover maps, night lights, and other socioeconomic indicators to map census population counts to buildings. Some of those methods rely on building volumes obtained from LiDAR [129] or digital surface models (DSM) [147], both of which are not always available [132]. Also, the majority of these studies rely on handcrafted features to dissagregate the available census data to buildings [20, 132, 145, 146, 147] and this limits their transferability. Additionally, due to the diversity in input data, each has its own framework, making standardization and comparison of the approaches difficult [132]. Nonetheless, their preliminary findings pave the way for further research into fine-scale population estimation.

3.4 Summary

Remote sensing could provide extensive and consistent coverage over large geographic areas, including remote and inaccessible regions which is very important in the field of population estimation where traditional data collection approaches are difficult and costly. Also, remote sensing data could be gathered and updated frequently. It allows researchers to track the population change due to urbanization and other migration events over time. However, integrating the remote sensing data with the population counts needs robust methodology which is an ongoing research to improve the accuracy and utility of remote sensing in population estimation. Several efforts have been made in the past to generate large-scale gridded population products. The field of large-scale population grid modeling is developing, leading to more precise and spatially refined population grids. Additionally, incorporating deep learning and machine learning techniques leads to the development of more sophisticated methods. Despite these developments, the majority of these data sets still have several limitations due to the methodology utilized, the completeness of the census, and the supplementary data used.

According to the accuracy assessment done by Thomsan et al., the lack of ancillary data on building use and built-up densities causes the gridded population products, such as the GPWv4, GHS-POP, WordPOP, and LandScan, to significantly underestimate population counts in slums and densely populated areas of Kenya and Nigeria [54]. In general, a lack of built-up environment data leads to an overestimation in unpopulated areas (with industry and commercial complexes) and an underestimation in high-rise residential structures. On the one hand, including multiple variables in weighted dasymetric mapping, such as in Worldpop and Landscan, helps to make more informed decisions; on the other hand, inconsistencies and disagreements in the multi source data may add bias to the results and make frequent product updates difficult [101]. Some other studies highlight the limitation of spatial resolution in some specific applications. Smith et al.

demonstrated that the resolution of the WorldPop (100 m) and LandScan (1 km) data sets limits their integration with other high resolution data sets, such as flood hazard data (90 m) and suggest improving the resolution of existing gridded population data sets [148]. It has been identified that lower resolution creates erroneous population estimates and these errors grow worse as the geographic resolution required for the analysis decreases [20, 149].

Therefore, several studies focus on fine-resolution population estimates. The majority of these research used their own methodologies, regional data and evaluated solely in comparable geographical areas. Therefore, it is important to build approaches that rely on freely accessible worldwide data sets that may be compared in different geographic regions. Another challenge is the availability of reference population data, which is usually lacking or out-of-date. Even when it is available, the resolution is typically at the scale of census enumeration zones, which are relatively coarse, making validation of the methods at fine resolution difficult. Therefore, aggregating fine-resolution predictions to the next level to compare them with the available reference population data has become the most commonly accepted practice in population estimation studies. With this indirect assessment setting, it is critical to impose the trustworthiness of the method developed. With explainable Artificial Intelligence (xAI), methods may become more apparent and interpretable. Beyond quantitative performance, this unboxing of black box models would aid in better understanding and comparing the operation of deep learning algorithms in population estimation.

4 Population Estimation Data Set

Population distribution is a key to study the spatiality of our landscapes. Undergoing rapid urbanization in cities is leading to environmental concerns such as climatic changes, food and water scarcity, poor air quality, deforestation, and so on [150, 151, 152, 153, 154]. In the past few years, machine learning and statistical methods have been used to estimate population distributions directly from remote sensing data [16, 18, 19, 47, 155]. In most of these studies, either the data is not available for download or could be reconstructed only for a few cities. Also, due to the lack of a large-scale benchmark population estimation data set, these methodologies have been applied to a smaller region or require collecting and processing census data. This makes the overall development of the new methods complex and time-consuming. As a result, this work aims to build a comprehensive data set for large-scale population estimation. The data set provides a systematic regression and classification scheme by fusing multi-source Earth observation data over 98 European cities. These cities serve 28 European Union (EU) member countries and four European Free Trade Association (EFTA) countries, representing a diverse variety of topography, demography, and architectural designs [4].

4.1 Study area

The study area is spread over Europe, the orange dots in Figure 4.1 depict the selected cities. The cities are chosen based on the total number of inhabitants. First, all cities in Europe with a population of 300000 or more in 2014, according to the UN World Urbanization Prospects - The 2014 Revision [156] are selected. Then, depending on the availability of the reference population data, 106 cities are chosen. The extraction of the city's geographic area using the administrative boundary could be difficult because administrative census tracts split or combine over time. Furthermore, due to rapid city growth, cities expand well beyond their official bounds [157, 158]. So, an algorithm is employed to determine the city's extent, considering city growth over time. The city center coordinates from the UN World Urbanization Prospects - The 2014 Revision [156] are utilized as a starting point, along with the Global Urban Footprint (GUF) [159], which gives a binary mask of urban versus non-urban regions. A rectangle that is centered at the extracted coordinates of each city is adaptively expanded outward until half of its area is no longer built up [4] according to the GUF. To further account for rising urbanization, each side of the rectangle is increased by a factor of two (a factor of four in the area).

Considering that the resulting rectangles of two neighboring cities may intersect, a set of rules is used to assign the intersecting region to one of the two cities and to ensure that each city's extent covers a distinct area. The algorithm is summarized in Algorithm

Figure 4.1: The orange dots on the figure above indicate the location of selected EU cities in our study. Image is taken from our own publication [4].

1. It iterates in descending order, beginning with the cities with the biggest overlap. Depending on the proportional size of the overlapping region, $city_a$ is either merged into $city_b$ or the overlapping area is allotted to $city_a$ and removed from $city_b$. After resolving the overlaps, the number of cities was merged to 98. Based on the defined extent of each city, the following data were collected and processed for the data set.

Figure 4.2 shows the three cases that are presented in Algorithm 1. In the first scenario (4.2a), Birken ($city_a$) and Liverpool ($city_b$) have an overlapping area larger than the threshold, which is equal to 50 % of Birken's total area. Additionally, Birken is an entire subset of Liverpool. Thus, Birken was removed. In the second case (4.2b), again, the overlapping area between Coventry ($city_a$) and Birmingham ($city_b$) is greater than the 50 % of Coventry's total area, however, Coventry is not a subset of Birmingham. Therefore, Coventry is merged with Birmingham. In the third scenario (4.2c), the overlapping area between the New Port ($city_a$) and Cardiff ($city_b$) is less than the threshold

Algorithm 1 Allocation of intersecting areas - Pseudocode.

Require: $city_a$ and $city_b$ with biggest overlap and $city_b \geq city_a$

```
for city_a ∩ city_b in data set do
    while overlap > 0.5 * city_a do
        if city_a ⊂ city_b then
            Remove city_a
        else
            Merge overlapping area of city_a into city_b
        end if
    end while
    Remove overlapping area from city_b
end for
```

(50 % of New Port's total area), thus, the overlapping area is merged to the New Port and removed from Cardiff.

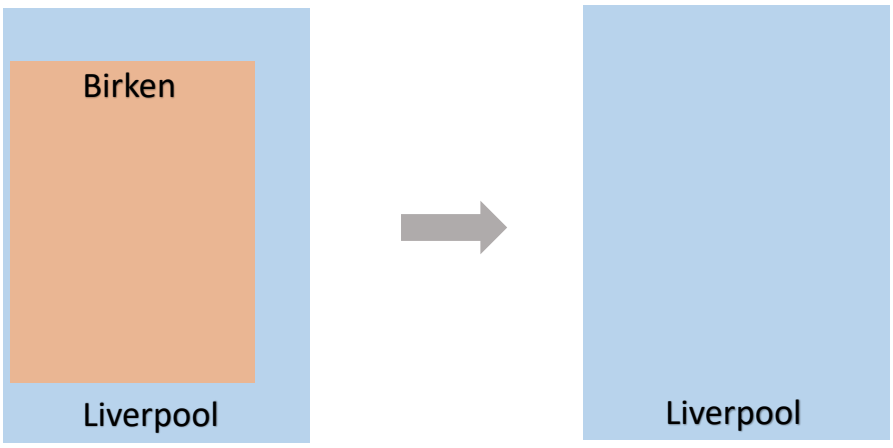
4.2 Data

The data set comprises reference population data, multi-spectral Sentinel-2 imagery (SEN2), Digital Elevation Model (DEM), Local Climate Zones (LCZ), VIIRS Nighttime lights, and data from the OpenStreetMap (OSM) initiative.

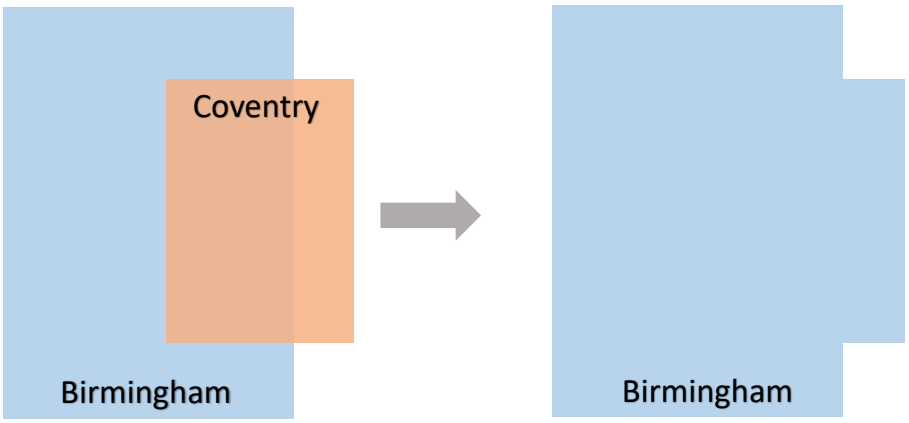
4.2.1 Population data

The census bureau of each country publishes population estimates on its administrative units. However, the size of administrative units varies from region to region, making these data difficult to use for analysis. In such a case, gridded population data, in which the uniformly sized grid cell represents the people living in that grid area, offers consistent population data.

In Europe, the European Statistical System (ESSnet) project, in collaboration with the European Forum for Geography and Statistics (EFGS), developed and published such population grids using census data at a resolution of 1 km. Their methodology includes aggregation, disaggregation, and a hybrid approach, depending on the availability of the data. Typically, aggregation (bottom-up) is considered the most accurate approach for creating population grids [160] and in their project for approximately 18 nations, aggregation or a hybrid approach is used to produce the population grids. The disaggregation method is used for the remainder due to a lack of detailed data [161]. As a result of the differences in methods, the quality of the output differs. In disaggregation, for example, the misplacement of persons is proportional to the size of the census unit; the larger the census unit, the greater the misplacement. The positional accuracy for each building and address ranges from 0.1 m in Austria to 100 m in Estonia [162]. The population grids are publicly available for non-commercial use via Eurostat and cover approximately 4.3 million km^2 with 480 million inhabitants [160]. The GEO-

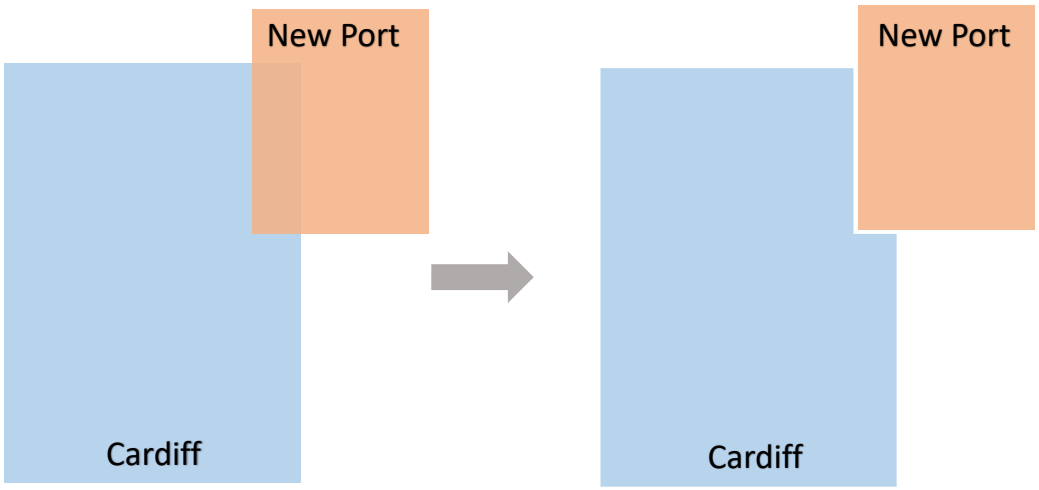


(a) Case 1: the overlapping area between Birken ($city_a$) and Liverpool ($city_b$) is greater than the threshold (50 % of Birken's total area) and Birken is an entire subset of Liverpool, therefore, Birken is removed.



(b) Case 2: the overlapping area between Coventry ($city_a$) and Birmingham ($city_b$) is greater than the threshold (50 % of Coventry's total area) and Coventry is not a subset of Birmingham, therefore, Coventry is merged to Birmingham.

Figure 4.2: Illustration of Algorithm 1 with three use cases regarding the allocation of intersecting areas among the selected European cities.



(c) Case 3: the overlapping area between New Port ($city_a$) and Cardiff ($city_b$) is less than the threshold (50 % of New Port's total area), therefore, the overlapping area merged to New Port and removed from Cardiff.

Figure 4.2: Illustration of Algorithm 1 with three use cases regarding the allocation of intersecting areas among the selected European cities.

STAT 1B project website contains additional information about the product standards, methodology, and quality assessments [163].

4.2.2 Sentinel-2

The Sentinel-2 mission was launched in June 2015 [164], consisting of two identical satellites (2A & 2B) in a sun-synchronous orbit phased at 180 degrees to each other. Sentinel-2 satellites offer multi-spectral optical images spanning 13 spectral bands at spatial resolutions of 10 m, 20 m, and 60 m. Thus, it has an enormous potential for fine-scale mapping of human populations, however, obtaining could-free mosaics are always a challenge. Google Earth Engine (GEE) is employed to create the cloud-free Sentinel-2 images [165], following three main steps: querying, scoring and mosaicing. In the query step, the Sentinel-2 images are loaded from the catalogue. The quality score for the loaded image is calculated in the scoring step based on pixel-by-pixel cloud analysis and in the mosaicing step, the selected images are mosaiced based on the meta-information calculated in the previous steps. The complete details of this algorithm can be found at [166]. To capture seasonal variations in the data, all four seasonal sets of Sentinel-2 imagery are used. We processed Sentinel-2 imagery from 10-2016 until 09-2017, covering Winter, Spring, Summer, and Autumn.

4.2.3 TanDEM-X Digital Elevation Model

An accurate 3D topographic map of Earth could be very useful in urban studies and other land use-related downstream applications. In view of this, the TanDEM-X mission is

launched to create a high-quality, homogeneous, three-dimensional image of Earth. The data for the global DEM product was collected between December 2010 and January 2015, and in September 2016, the global DEM was completed. It is considered to be one of the most accurate digital elevation models available on a global scale [167]. Thus, it is suited for many environmental studies such as land use and cover analysis, urban planning, climate change, etc. with a coverage of 150 million km^2 of the entire landmasses of the Earth and 10 m absolute height accuracy (90 % linear error) [168]. We used the publicly available TanDEM-X 90 m (3 arcsec) global DEM product [169], which provides a final Digital Elevation Model of the Earth's landmasses.

4.2.4 Local climate zones

While local climatic zones (LCZ) are explicitly created to standardize urban heat island research, they are currently utilized to categorize urban areas in many environmental-related studies [170]. It is divided into 17 structural categories based on the land surface and properties, with 10 defining built-up zones ranging from compact high-rise to open low-rise and 7 describing natural zones ranging from dense vegetation to bare fields. As a result, the built-up and land cover properties distinguish each zone. Our work utilizes the So2SatLCZ v1.0 urban local climatic zone classifications at a resolution of 100 m. They were created by fusing the LCZ classification results from freely available Sentinel-1 and Sentinel-2 data using deep learning [171]. The patches in this data set are hand-labeled by 15 domain experts according to the local climate zone categorization methodology, followed by a well-defined visual and quantitative evaluation process. As an outcome, this benchmark data collection could be useful to urbanologists, demographers, climatologists, and a variety of other studies.

4.2.5 Nighttime lights

Nighttime lights have been widely used in population estimation studies because of their strong correlation with the spatial distribution of human population [56, 172, 173, 174]. DMSP-OLS and NPP-VIIRS are the two most commonly utilized nighttime light data. With finer spatial resolution NPP-VIIRS offers a better potential for modeling socio-economic indicators than DMSP-OLS [175]. For each year, 2012 to 2020, global VIIRS nighttime lights have been produced using the monthly cloud-free mean radiance. In this work, the average-masked radiance version of VNL V2 with a resolution of 500 m is used. It is a preprocessed version free of outliers from transitory events [176] and freely available at (`https://eogdata.mines.edu/products/vnl/`)

4.2.6 OpenStreetMap

OpenStreetMap (OSM) (`http://www.openstreetmap.org`) was launched in London in 2004 as a wiki-style collaborative mapping project with approximately 10 million registered participants. Based on aerial images and field research, its contributors correct and insert the geographical location data on a very detailed level. This data could be entered as nodes or relations and characterized with informative tags. The locations cover,

for example, types of streets, buildings, boundaries, water bodies, etc. [177]. OSM is openly available on a large-scale under the Open Data Commons Open Database License (ODbL) (`http://www.opendatacommons.org/licenses/odbl/1.0/`). In this work, low-level and high-level features from the OSM data are extracted. The OSM features that have a strong correlation with the local population are included in the low-level features. For example, the statistics computed for a location with a high number of certain nodes (such as stores, gas stations, dwellings, schools, etc.) is a reliable indicator and can be used as a feature vector to estimate population density. High-level features define urban land use by extracting the building functions from OSM building tags. These features together illustrate the interplay between human activities and the environment [178].

4.3 Data Preparation

For all the data sources, a two-stage preprocessing is employed. In the first stage, data is gathered from all the sources mentioned above, cropped using extended city extents, and then processed. In the second stage, the 1 x 1 km patches have been generated for each city. Figure 4.3 depicts all of the preprocessing steps performed for each input data in the first stage of data preprocessing. Data collected from all the sources has been cropped using the extended city boundaries predefined by our algorithm. The DEM mean is subtracted from DEM data to standardize it and scaled to a unit variance. Since the input data is gathered from various sources, they are at different spatial resolutions and in multiple Coordinate Reference Systems (CRS). For example, the nighttime lights data is in WGS84 (EPSG:4326) while the LCZ, DEM, and Sentinel-2 data are in Universal Transverse Mercator (UTM) zones, and the population grid is in EPSG:3035 - ETRS89-extended / LAEA Europe. All input data have been reprojected from their respective coordinate system to the EPSG:3035 coordinate reference system in order to align with the population grid. As the sentinel-2 data has the highest spatial resolution (10 m) among others, all input data has been upsampled to match it.

Low-level features have been extracted from the OSM planet dump (`https://planet.osm.org/planet/2017/`) downloaded from 2017 to match the year with the Sentinel-2 data. Each city's bounding box is utilized to extract the OSM dump using the command-line tool Osmosis (`https://github.com/openstreetmap/osmosis`), and the node statistics for each 1 x 1 km patch of all cities are computed using the OSMnx python library [179]. The chosen OSM tags for which the statistical counter has been computed are displayed in the Table 4.1.

High-level features represent the land use. In this work, the three different OSM tags: *building*, *amenity*, and *shop* are extracted. According to OSM guidelines, each of the three tags could have a variety of values. These three tags have a total of 341 potential values that are mapped to a unified and reduced system of classifying land uses: *commercial*, *industrial*, *residential*, and *other*. Due to the possibility of the three tags occurring simultaneously, it has been ensured that they do not conflict with one another and any buildings with inconsistent values are excluded. Additionally, because the tags

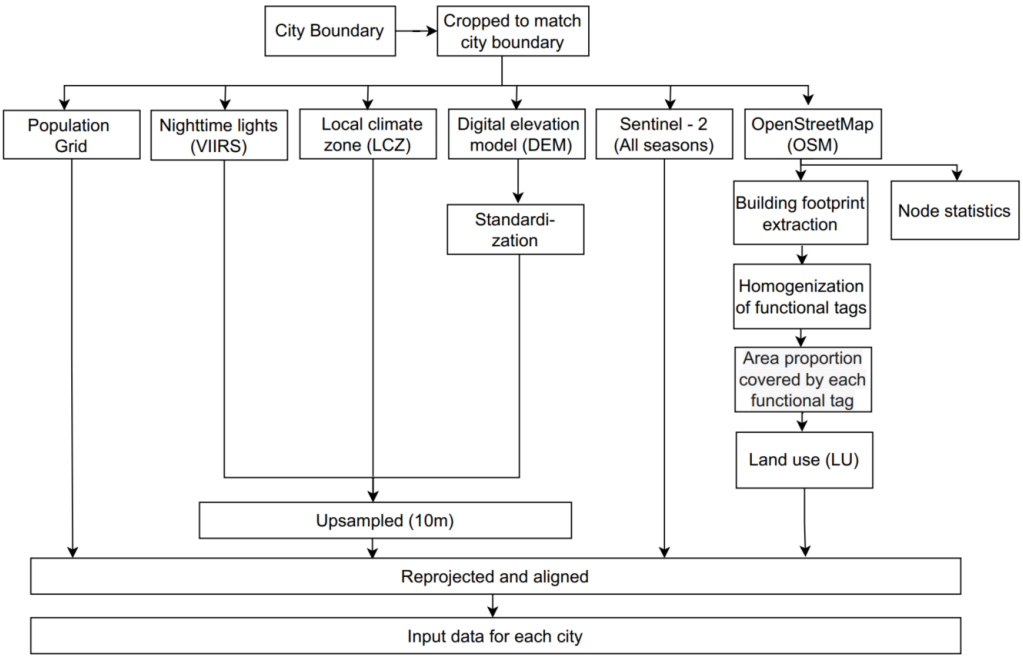


Figure 4.3: Step-by-step preprocessing of all the input data sources to prepare the corresponding input data for each city. Image is taken from our own publication [4].

are recorded as free-form text fields, OSM contributors are not limited to using them and can instead input any text, so their semantic information needs to be homogenized. After homogenizing, the building polygons are converted into raster data. The values of the raster represent the area covered by building polygons that lie within a raster pixel. A four-band raster with associated land use proportions is produced by using this approach for each land use class. Figure 4.4 displays the results of the first data preprocessing stage for the city of Munich.

All the input data processed in the first stage are utilized to create patches in the second stage. The population grid with the grid cell size of 1 x 1 km is used as a reference to crop all the other input data. Since the grid cells at the border of the city boundary

aerialway	building	historic	natural	restrictions	water
aeroway	craft	landuse	office	route	waterway
amenity	emergency	leisure	place	shop	
addr:housenumber	geological	man_made	power	sport	
barrier	healthcare	other: True	public_transport	telecom	
boundary	highway	military	railway	tourism	

Table 4.1: Nodes with these OSM tags are considered for the statistical analysis/counting of the corresponding 1 x 1 km patch. Table is taken from our own publication [4].

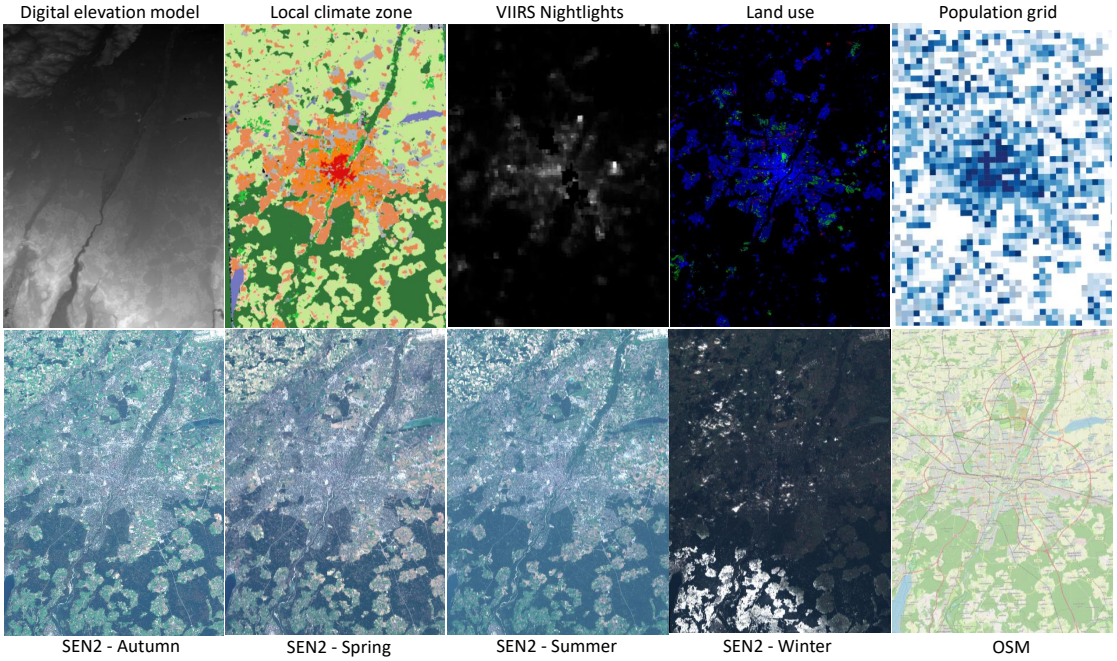


Figure 4.4: All the input data for the Munich city which is created using the first step of data preprocessing. Image is taken from our own publication [4].

might be cut in between, an area threshold is applied to each cell. Only cells larger than $0.9\,\mathrm{km}^2$ or $90\,\%$ of typical cell size are considered part of the city. This eliminates the remaining cells while including the edge cells that lie mostly within the city boundary.

The reference population grid, which serves as our ground truth contains only populated grid cells, which means that the uninhabited regions of a city are not well represented in these grids. These missing grid cells were randomly examined in a few cities, and it was discovered that they covered green spaces and water bodies. In the development of population estimation methods, to understand the geographical pattern of the regions where people live, it is also important to understand the spatial features that do not correlate with the residential settlements. Therefore, these cells have been added to the data set. All input data is cropped to patches measuring $1\,\mathrm{x}\,1\,\mathrm{km}$ using all grid cells, whether populated or not. The procedure of creating a patch is shown in Figure 4.5. As a result, a total of 9 patches, one from each input data source, are produced for each population grid cell.

There are some applications where an approximation of people living in a region is sufficient such as climate change or post-impact studies of natural disasters. Therefore, the absolute population counts are binned into population classes based on the specified range in which the population count falls. Following Robinson et al. [16] discretization method, the population class of a grid cell C_{cell} is a function of its absolute population count P_{cell} defined as follows:

$$C_{\text{cell}} = \begin{cases} 1 & \text{if } 2^0 \leq P_{\text{cell}} < 2^1 \\ 2 & \text{if } 2^1 \leq P_{\text{cell}} < 2^2 \\ 3 & \text{if } 2^2 \leq P_{\text{cell}} < 2^3 \\ \dots & \\ k+1 & \text{if } 2^k \leq P_{\text{cell}} < 2^{k+1} \end{cases}$$

This can be simplified as follows, where C_{cell} denotes the population class and P_{cell} the absolute population of the corresponding *cell*:

$$C_{\text{cell}} = \lfloor \log_2(P_{\text{cell}}) \rfloor + 1$$

This leads to the following population range and their class associations:

Class	Population Range
Class 0	0
Class 1	1
Class 2	2 – 3
Class 3	4 – 7
Class 4	8 – 15
Class 5	16 – 31
Class 6	32 – 63
Class 7	64 – 127
Class 8	128 – 255
Class 9	256 – 511
Class 10	512 – 1023
Class 11	1024 – 2047
Class 12	2048 – 4095
Class 13	4096 – 8191
Class 14	8192 – 16383
Class 15	16384 – 32767
Class 16	32768 – 65536

As per the collected population data, k has a maximum value of 16, thus 17 classes in total. Including population class, in addition to population counts, would provide end-users additional freedom to create either a regression or a classification model for the task based on the application's requirements.

After both preprocessing steps, Figure 4.6 illustrates the odd-numbered class patch-sets from the data set, as well as their respective population class and population count. Lower-class patches mainly consist of green fields, water bodies, bare grounds, and sparsely inhabited areas. As the class number increases, patches represent low to high-populated regions. In other words, patches ranging from lower to higher classes indicate rural to urban areas.

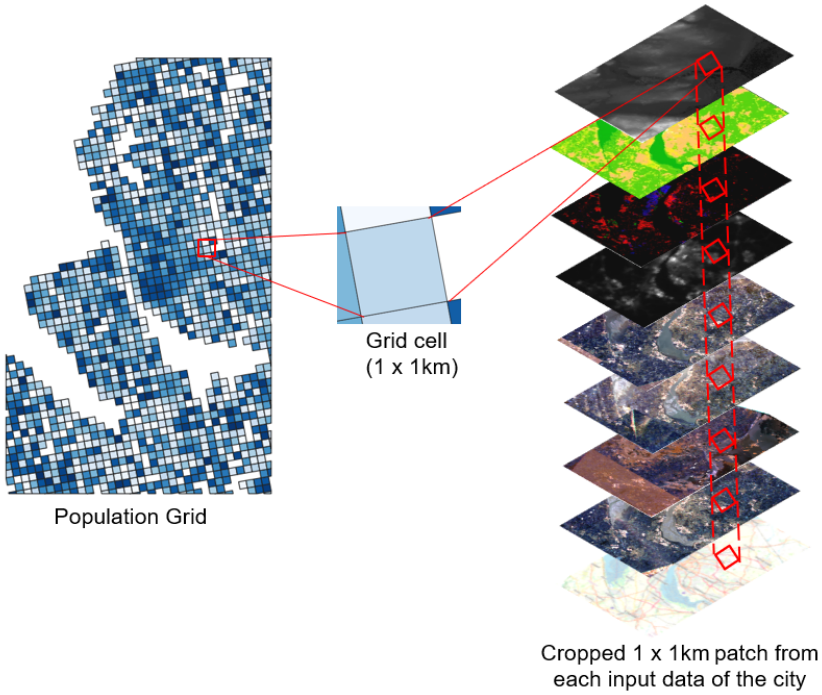


Figure 4.5: Patch creation process, the second step of data preprocessing. All input data sources have been cropped for each cell in the population grid. The size of each patch is 1 x 1 km.

4.4 Data Structure

The raw OSM data is available as OSM XML files, the statistical features extracted from the OSM are available as Comma Separated Value (CSV) files and rest of the data as GeoTiff files. Due to different licensing requirements, the data set has been split into two parts: So2Sat-POP Part1 and So2Sat-POP Part2. So2Sat-POP Part1 is distributed under the Creative Commons Attribution 4.0 International License (`http://creativecommons.org/licenses/by/4.0/`), and So2Sat-POP Part2 is distributed under the Creative Commons Attribution Share-Alike International License (`http://creativecommons.org/licenses/by-sa/4.0/`). So2Sat-POP Part 1 includes patches from all the seasons of Sentinel-2, local climate zones, nighttime lights, land use, and OSM features, yielding a total of 1104688 patches. So2Sat-POP Part2 consists of 276172 patches from the digital elevation model and raw OSM patches. So2Sat-POP Part1 requires $\sim$ 98 GB of storage, while So2Sat-POP Part2 requires $\sim$ 5.20 GB.

The data set has a predefined train and test split in both the parts. For the training set around 80 % of the data (80 cities) have been randomly selected and the rest 20 % of the data (20 cities) constitutes the test set. In addition to the input data folders, all city folders in So2Sat-POP Part1 have a comma-separated value (*.csv) file with the referenced ground truth, population count and a population class for each patch. All data

folders follow a standard folder structure, with class sub-folders labeled Class_x, where x is the class value. The number of class folders varies in each city due to differences in population distribution. Malaga, for example, has the highest class folder of 16 because the largest population count in the city's 1 x 1 km area is 39535, whereas Riga has the highest population count of 15839, hence the highest class folder in the Riga city folder is 14.

The city folder is named as `xxxx_xxxxx_city_name`, where `xxxx_xxxxx` is a randomly generated identification number and the city's postal code. Each patch name has a unique identification code that matches the naming convention of its associated population grid cell [161]. Zero-count patches that do not constitute population grids have been issued a numeric identification number. The data set is freely accessible through the official media library of the Technical University of Munich (TUM). So2Sat-POP Part 1 is available for download at (`https://mediatum.ub.tum.de/1633792`), and So2Sat-POP Part 2 is available at (`https://mediatum.ub.tum.de/1633795`).

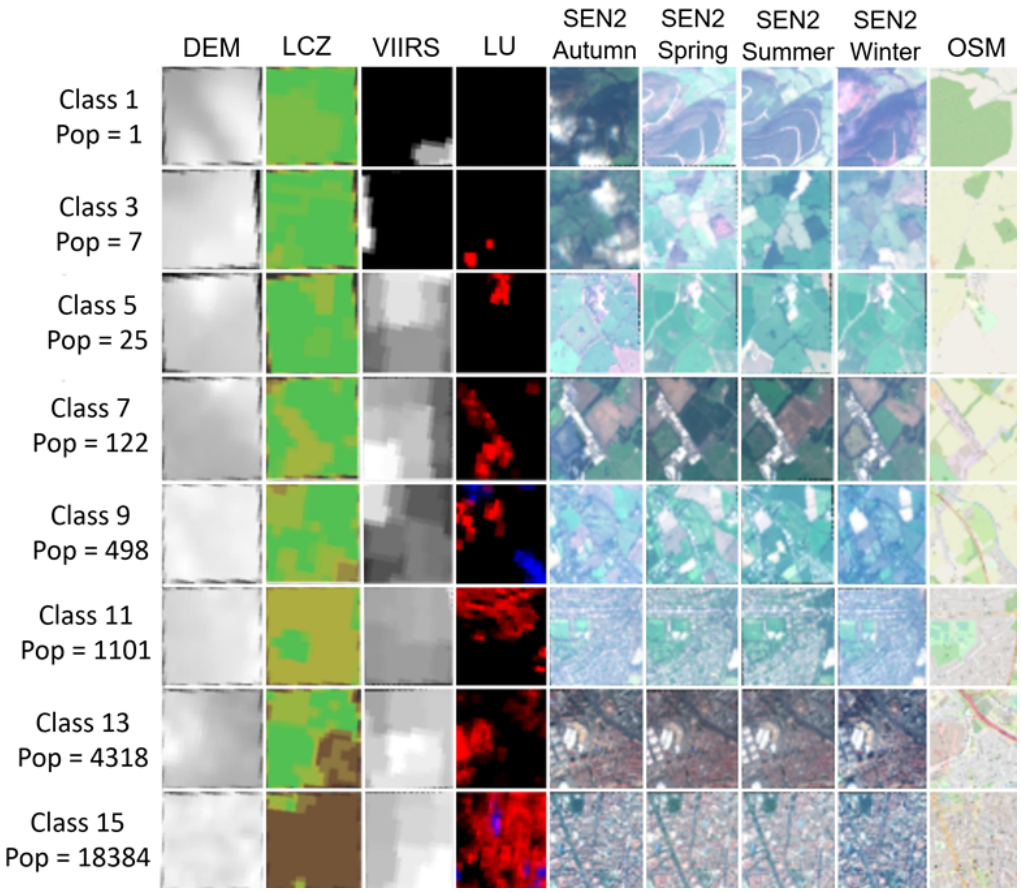


Figure 4.6: Sample patches from the odd-numbered classes of our data set. Lower classes depict sparsely populated regions while higher classes depict densely populated regions.

4.5 Technical Validation

The RF proposed by Breiman [110] is a sort of ensemble learning that consists of a set of randomly generated decision trees that can be used for classification and regression and has been utilized successfully in numerous population estimation studies [63, 146, 180, 181]. Besides its ability to handle noisy data and impervious to overfitting, it is simple within the RF algorithm to assess the relative relevance of each feature on prediction. Therefore, in this study to demonstrate the suitability of the data for population estimation, a RF algorithm is implemented in Python for both regression and classification. The two important hyper-parameters in the RF model, the number of trees to grow and the maximum number of features are automatically fine-tuned using the grid search combined with 10-fold cross-validation to avoid the impact of the data partition.

In order to train the model, different features have been extracted from all the input data of the training set. These features include the mean, median, min, max, and standard deviation from the Sentinel-2 imagery's RGB band, the mean and max for the DEM and nightlights, the total area covered by each class of land use, the majority class of the LCZ and OSM-based statistical features such as street density, the presence of highways, railways, and other [4]. Using this method, 125 features in total were extracted. While class labels are employed as the ground truth in classification, the absolute population count is the response variable in regression.

On the 18 unseen test cities, the trained model has been assessed. For regression, the model is evaluated using RMSE, MAE and R^2 as indicated in Table 4.2. For each grid cell in the test data, its actual population count versus the predicted population count is plotted to visually assess the model fit. Figure 4.7 (a) shows that the model understates the population counts for the high population density patches. However, it is a fair fit for patches with a population count of less than 15000. The classification model's performance is assessed using balanced accuracy and macro-averaged Precision, Recall, and F1-score due to the data imbalance caused by a higher percentage of low to medium-population density patches than high-population density patches. The classification results are listed in a Table 4.3. A normalized confusion matrix is presented to show the performance of the classification model for each class. Figure 4.7 (b) shows that the model is more capable of accurately forecasting the upper classes than the lower classes, particularly in the first three populated classes (Class 1, 2, and 3). These three classes indicate areas where the population count is between 1 and 8. It is possible that it is difficult to tell if these three classes are apart due to their similar features. A few Sentinel-2 samples from each of these three classes are examined visually to determine whether or not they can be distinguished visibly. As seen in the Figure 4.8, the randomly selected patches from Class 1, 2 and 3 appear to be quite similar visually, making it exceedingly challenging for the model to distinguish among them. Merging these indistinguishable three classes is one possible taxonomy modification that retains all vital information while assisting the model in better classifying these very low populated regions.

The RF algorithm's built-in feature importance describes which attributes are more significant for the predictions. It also aids in understanding the model's learning. The

importance of each feature is computed by calculating how each feature on the internal nodes across the decision trees in the forest reduces the impurity of the split in classification or decreases the variance in regression. The method is included in the RF scikit-learn implementation. Figure 4.9 only shows the twelve most important features chosen by the RF algorithm and utilized to estimate the population count and population class for the cities in the test data set. Land use area proportions, LCZ classes, nightlights, and statistical variables retrieved from OSM, such as count on buildings, shops, and highways, are ranked as the most essential features for regression and classification.

Regression			
Method	**RMSE**	**MAE**	**R^2**
RF	1276.26	463.35	0.827

Table 4.2: Evaluation of RF model to estimate the population counts on the test data set. The experimental results have been directly taken from our own publication [4].

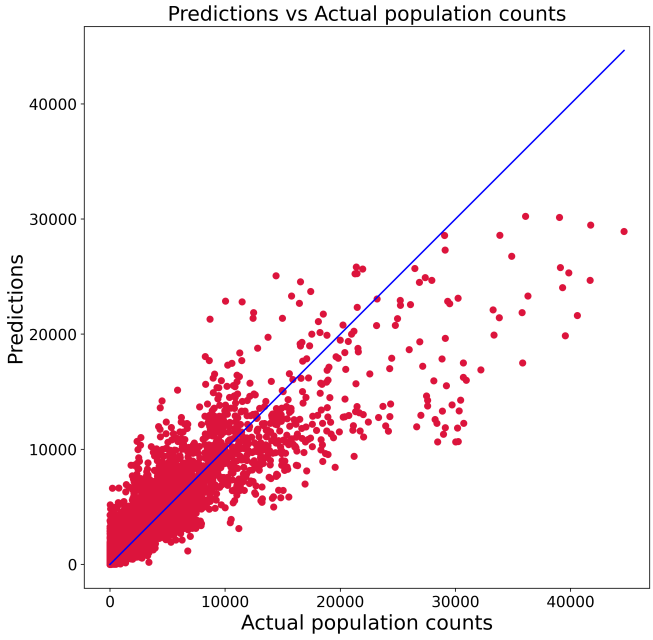
Classification					
Method	**Accuracy**	**Balanced Accuracy**	**F1 score**	**Precision**	**Recall**
RF	0.5913	0.3795	0.3833	0.4533	0.3795

Table 4.3: Evaluation of RF model to predict the population class on the test data set. The experimental results have been directly taken from our own publication [4].

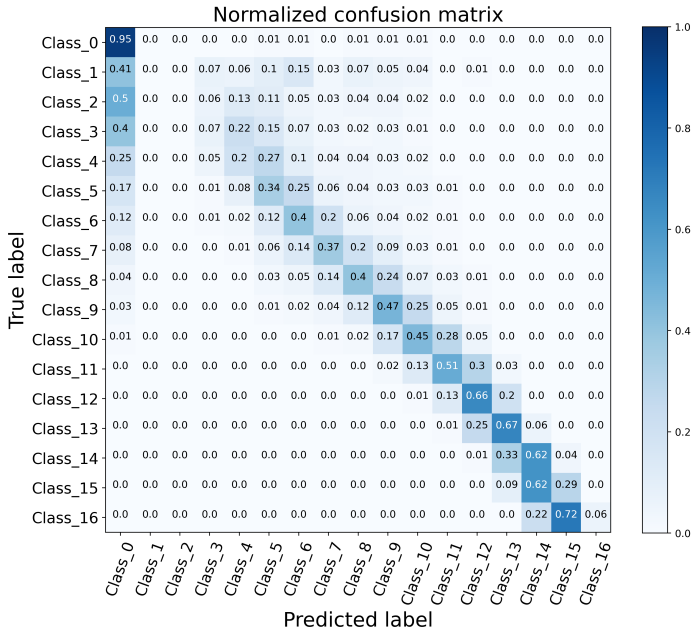
4.6 Summary

The development and evaluation of novel approaches in population estimation studies have always been a challenge owing to the lack of data. For most population estimation studies, data utilized is either developed on a very small scale or not openly available. There are some gridded population products such as GHS-POP, LandScan, Worldpop, HRSL available on a large-scale as discussed in section 3.1. However, most of these products yield different results due to differences in their methodology and use of regional data sources. There have been some comparison studies to evaluate them. But, these studies also necessitate the collection of accurate reference population data. Thus, reproducing the results, comparing methodologies, and developing new methods becomes difficult and time-consuming. In this work, an attempt has been made to fill this gap by offering a systematic population estimation data set.

It is a large-scale data set spanning 98 European cities, including different landscapes, demography, and geography and integrates data from previously unexplored data sources



(a)



(b)

Figure 4.7: (a) Predicted vs. Actual Values for regression, the model fits well except for the high population counts where the points appeared dispersed from the regressed diagonal line (b) Confusion matrix for classification, normalized by class support size (number of patches in each class). Confusion among the non-urban classes is higher than among the urban classes.

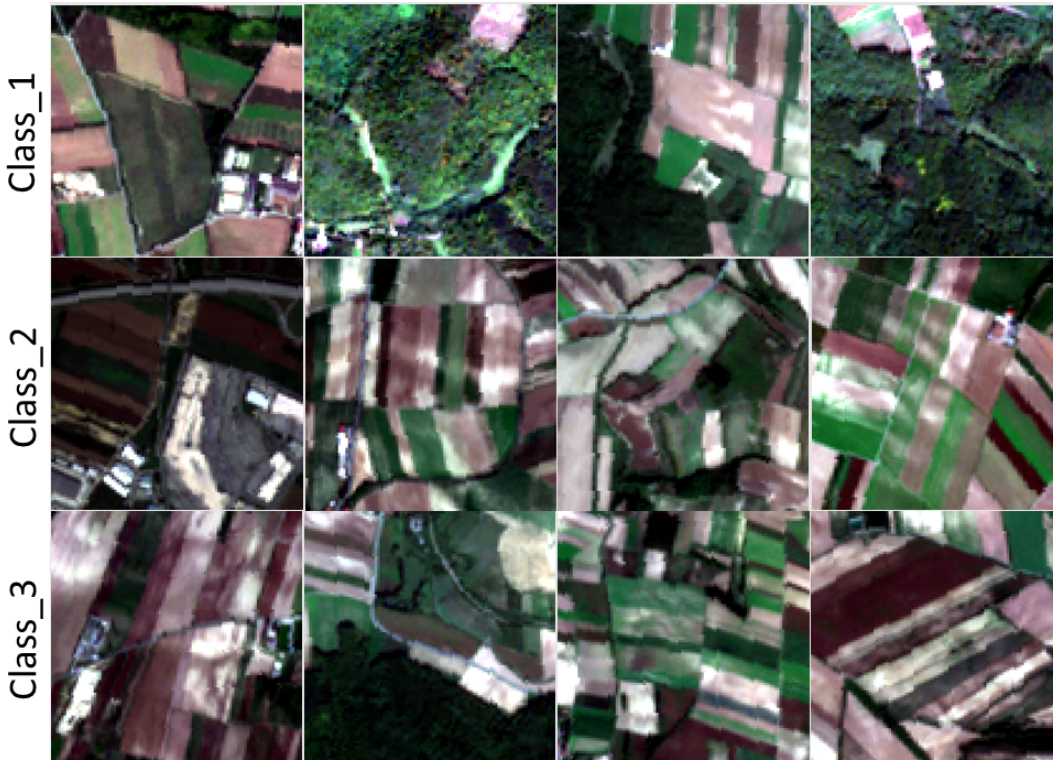
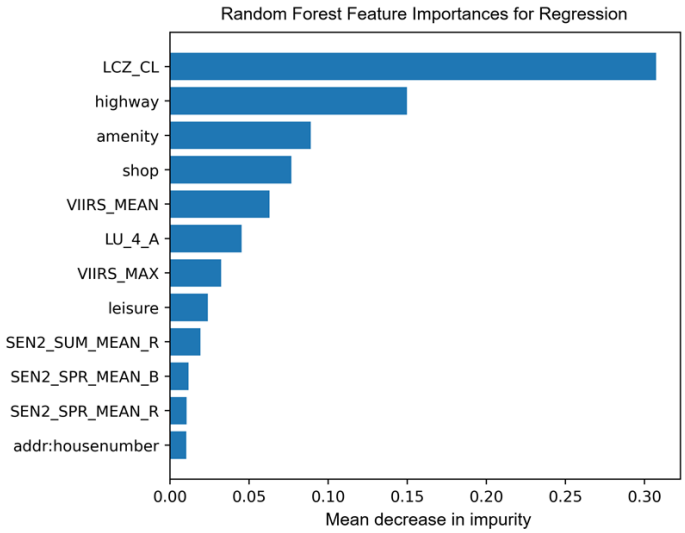


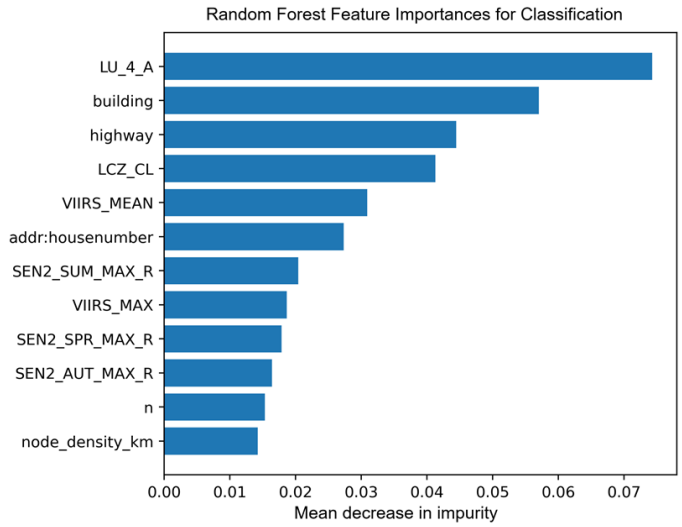
Figure 4.8: Visualization of a few randomly selected Sentinel-2 patches from Class 1, 2 and 3 to determine their distinguishability. These examples appear to be visually similar.

in population estimation studies. The reference population data grids utilized in this data set are accessible throughout Europe via the ESSnet project at a consistent resolution of 1 km, however, the quality varies by country depending on the data available and the method used. Furthermore, the population grids have been created using the 2011 population and housing census while the data collected from other input sources are from a different time period. For example, Sentinel-2 data is from 2017, as the mission was initiated in 2015. The time lag between collecting population data and other associated input data may induces noise of uncertain quality. However, because population data is typically collected once every decade, obtaining data from different sources from the same year becomes extremely challenging. Nonetheless, this data set will be useful in developing new statistical and machine-learning approaches for estimating population at a consistent spatial resolution which is often missing in today's population data sets across countries.

In addition to population estimation studies, this data set could serve as a foundation for future comparative studies in a variety of applications. This improved population distribution database collected from multiple sources could provide substantial data in both academic and non-academic domains such as disaster risk analysis, urban planning, provision of socio-technical infrastructure, updating of census, or understanding of changing population dynamics and urbanization trends.



(a)



(b)

Figure 4.9: RF feature importance based on the mean decrease in impurity (MDI). The higher the value the more important the feature. Plot shows only the twelve most relevant features for both regression (a) and classification (b)

5 Deep Learning for population estimation

Recent advances in deep learning approaches and the availability of high resolution satellite imagery have enabled more accurate and up-to-date population estimation [16, 17, 18, 19, 182]. The majority of these approaches rely on an external settlement layer to assign known census population numbers to grid cells. Recent studies attempt to address this issue by designing algorithms that could directly predict the population counts from remote sensing data. Metzger et al. [135], for example, employed a deep learning model to predict population without relying solely on the known census. Similarly, Georganos et al. [136] proposed a deep learning-based methodology for predicting the population in three Sub-Saharan African countries without always relying on census data. However, differences in employed input data result in varied outcomes and their methods have been evaluated only in a comparable geographic area, which does not reflect their generalizability. Therefore, in this work a deep learning approach on a large-scale is developed and also comprehensive analyses in various geographies is conducted. Another major drawback of these methods is their lack of transparency. The black-box nature of deep learning models makes it impossible to thoroughly understand the methods' outcomes, thus makes difficult for end users to trust the results [183, 184]. To improve the transparency of these models an explainable AI technique is investigated and integrated which highlights the important features identified by the model while making the predictions. This interpretable framework has the potential to increase the usability and reliability of deep learning algorithms for population estimation.

5.1 Data

5.1.1 So2Sat-POP data set

The experiments are based on the So2Sat-POP data set developed in the chapter 4. This data set covers 98 cities in Europe, with 80 serving as the training set and the remaining 18 being used as the test set. It's a multi-source data set that includes the digital elevation model, local climate zone classifications, land use, nighttime light emissions, Sentinel-2 imagery from all seasons, and OpenStreetMap data. This data set could be used to build regression and classification population models. Section 4.2 contains more information on the input data sources utilized in the So2Sat-POP data set. Figure 5.1 depicts a sample patch taken from all input sources, with a reference population count of 755 and a population class of 10. There are a total of 276172 patches. The overall area covered by test cities is $\sim 18292\,\mathrm{km}^2$, whereas train cities encompass $\sim 119794\,\mathrm{km}^2$.

Figure 5.1: A patch-set from Class 10 and a reference population count of 755 as the ground-truth labels. Each such patch set consists of 9 patches, one from each input data source. Image is taken from our own publication [5].

5.1.2 Supplementary data set

A popular community standard product for population estimation, GHS-POP[103] is collected for comparison analysis. The European Union Joint Research Center created this gridded population data product utilizing remote sensing data. It offers a worldwide population count at two different resolutions of 250 m and 1 km [149]. In this comparison analysis, the ESSnet project's population grids are again used as reference population data [162]. It is same source as utilized in the So2Sat-POP data set, but it is unseen in our training, validation, and test data.

To demonstrate the transferability of our model, an additional data set is prepared using the three randomly selected cities in the United States (US): New York City, San Jose, and Denver. To account for ongoing urbanization, the geographic extent of the cities has been determined by the expansion algorithm from section 4.1. For each of these cities, the data is gathered from all of the input data sources used in the So2Sat-POP data set, preprocessed, and cropped to create 1×1 km patches. Census grids from the Socioeconomic Data and Application Center (SEDAC) are collected at a resolution of 1 km as the reference population data in the US. A quick overview of different data sets used in this chapter is presented in the Table 5.1.

5.1.3 Data preparation

The So2Sat-POP data set has a predefined train and test split. Only the training set is analyzed and used in data preprocessing step. The test set is left untouched, so it is unknown to the model. Figure 5.2 shows the distribution of population counts in the training data set. The right-skewed distribution indicates that high-populated samples are underrepresented in the data set. Often remote sensing data is affected by the noise introduced due to strong light sources on the ground such as in event of fire or reflections from sun and ice. These outliers could affect the data normalization. Therefore statistics such as mean, maximum, minimum, standard deviation and the 99.9th percentile of the pixel values have been computed for each input data source and shown in Table 5.2. In case of multi-channel data sources, statistics were calculated separately for each channel. While the 99.9th percentile for all channels is around 3200 in the Sentinel-2 imagery, the maximum lies in the range of 30000. Similarly, the maximum pixel value for the VIIRS is approximately six times its 99.9th percentile. As a result, pixels in the 99.9th percentile are regarded as outliers and the channels are clipped to its 99.9th percentile. Following

Data set	Year	Resolution	Purpose
So2Sat-POP			
Sentinel-2	2017	10 m	So2Sat-POP is a collection of multi-data sources. It is used as the input data for our training pipeline.
Digital Elevation Model (TanDEM-X)	2016	10 m	
Local climate zones (So2Sat LCZv1.0)	2017	10 m	
Nighttime lights (NPP-VIIRS)	2016	10 m	
OpenStreetMap (OSM)	2017	10 m	
Population Grids (GEOSTAT- EU)	2011	1 km	
SEDAC Census Grids (US)	2010	1 km	Reference population grid for the comparison study in the US.
GHS-POP	2015	1 km	Gridded population poduct for comparison in EU & US.

Table 5.1: A summary of all the data sets used in our work for training and comparison analysis.

the removal of outliers, the Sentinel-2 images are normalized in accordance with F. Li et al.'s [185] recommended preprocessing for images while training a ResNet or ResNet-like architecture, such that the channel-wise mean is zero and the channel-wise standard deviation is one. The VIIRS and DEM values are normalized to the range [0, 1]. Land use is a four band raster in which each pixel value represents the area covered by the respective land use classes (commercial, industrial, residential, and other) within that pixel. Given that it is a land use proportion percentage, it should theoretically add up to 1 for each pixel. However, it has been noticed in Table 5.2 that these values sometimes exceed because the buildings are stacked on top of each other and are in mixed-use, such as residential buildings on top of a commercial complex. Such values are normalized so that the maximum value for a pixel summed across all channels is equal to 1.

LCZ is categorical data with classes ranging from 1 to 17, classes 1–10 representing built classes, and classes 11–17 representing natural classes [186]. To simplify this categorization system, all-natural classes are mapped to 0 so that the model treats them the same and does not need to differentiate between them. All of the build classes are assigned a value between 0.1 and 1, ranging from lightly to heavily built-up areas. Table 5.3 shows the mapping of the LCZ categories to new processed values. The low-level features prepared in chapter 4 from OSM data are the vector values in different ranges. Therefore, a min-max normalization is employed to bring all values on the same scale [187].

Data Source	Mean	Standard Deviation	Maximum	Minimum	99.9th Percentile
SEN-2 (Autumn)					
Channel-R	708.69	474.74	32563.00	0.00	3597.00
Channel-G	839.14	440.73	31634.00	0.00	3185.00
Channel-B	1005.07	486.69	31914.00	0.00	3243.00
SEN-2 (Spring)					
Channel-R	747.21	492.73	29986.00	0.00	3371.00
Channel-G	838.89	442.19	28499.00	0.00	2976.00
Channel-B	931.09	442.19	29134.00	0.00	2926.00
SEN-2 (Summer)					
Channel-R	770.02	549.05	25484.00	0.00	3534.00
Channel-G	881.28	476.76	23252.00	0.00	3160.00
Channel-B	949.80	481.49	23290.00	0.00	3104.00
SENl-2 (Winter)					
Channel-R	740.05	493.04	32304.00	0.00	5387.00
Channel-G	838.21	457.98	32791.00	0.00	4715.00
Channel-B	1067.63	536.39	31532.00	0.00	3104.00
Land use					
Commercial	0.002	0.02	3.90	0.00	0.48
Industrial	0.002	0.02	2.00	0.00	0.54
Residential	0.004	0.03	1.64	0.00	0.36
Other	0.023	0.08	3.99	0.00	0.88
DEM	153.70	152.46	2431.60	-976.44	1414.87
VIIRS	2.86	7.83	535.94	0.00	86.48

Table 5.2: Pixel-level statistics of the input data sources in the training data set.

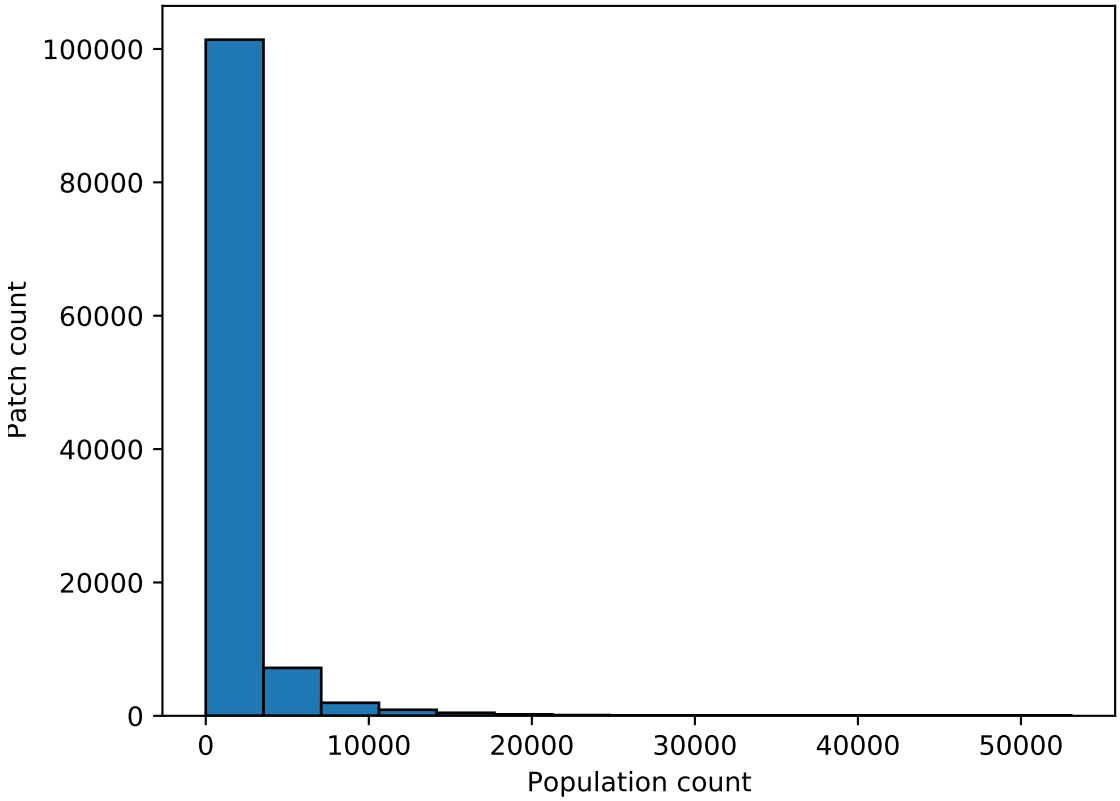


Figure 5.2: Population distribution of the training data set. The right-skewed distribution indicates that high-populated samples are underrepresented in the data set.

5.2 Method

VGG and ResNet have been the most widely used deep learning models in population estimation studies [16, 18, 19, 135]. Therefore, the two architectures are compared using a subset of the So2Sat-POP data set in an initial experiment. In this subset, Germany's northern cities serve as the train and validation set, while the southern cities serve as the test set. All the image data is concatenated and fed as input, and the vector data is added directly to the fully connected layers. ResNets outperformed the VGG-16 on this So2Sat-POP subset and was thus chosen for further experiments.

Due to the difference in dimensionality, the input data is divided into two categories. The first is two-dimensional raster data, which includes Sentinel-2 (RGB), VIIRS, LCZ, land use, DEM, and the second is one-dimensional feature vectors extracted from OSM data. The ResNet architecture is modified to handle both data categories simultaneously. The custom architecture, as depicted in the Figure 5.3 consists of two branches: the

Category	LCZ class	Mapped value
Compact highrise	1	1.0
Compact midrise	2	0.9
Compact lowrise	3	0.8
Open highrise	4	0.7
Open midrise	5	0.6
Open lowrise	6	0.5
Lightweight lowrise	7	0.4
Large lowrise	8	0.3
Sparsely built	9	0.2
Heavy industry	10	0.1
Dense trees	A	0.0
Scattered trees	B	0.0
Bush, scrub	C	0.0
Low plants	D	0.0
Bare rock or paved	E	0.0
Bare soil and sand	F	0.0
Water	G	0.0

Table 5.3: Mapping of LCZ categories from their corresponding classes to new processed values.

upper branch handles image data, called the image branch, and the bottom branch handles vector data, called the vector branch, and both are concatenated before the first fully connected layer. Both branches make use of a modified ResNet-50 [3] architecture. The image branch is modified to handle inputs of size $10 \times 100 \times 100$ (channels $\times$ width $\times$ height), whereas the lower branch uses a ResNet-50-like architecture for tabular data [188], where the convolutional layers of the ResNet-50 architecture are replaced with fully connected layers. The two branches are fused using intermediate fusion protocol [189].

5.2.1 Experimental setup

The training set is split into training (80 %) and a validation set (20 %). For all the experiments, the normal Xavier initialization [190] is used to initialize weights and biases. An ADAM optimizer [191] is used with an initial learning rate of 1 x 10^{-4} and decayed by a factor of 0.1 whenever the training loss did not improve for five subsequent epochs. Each experiment is run for a maximum of 50 epochs with a batch size of 32. Regularization techniques include batch normalization [190] and weight decay [192]. In

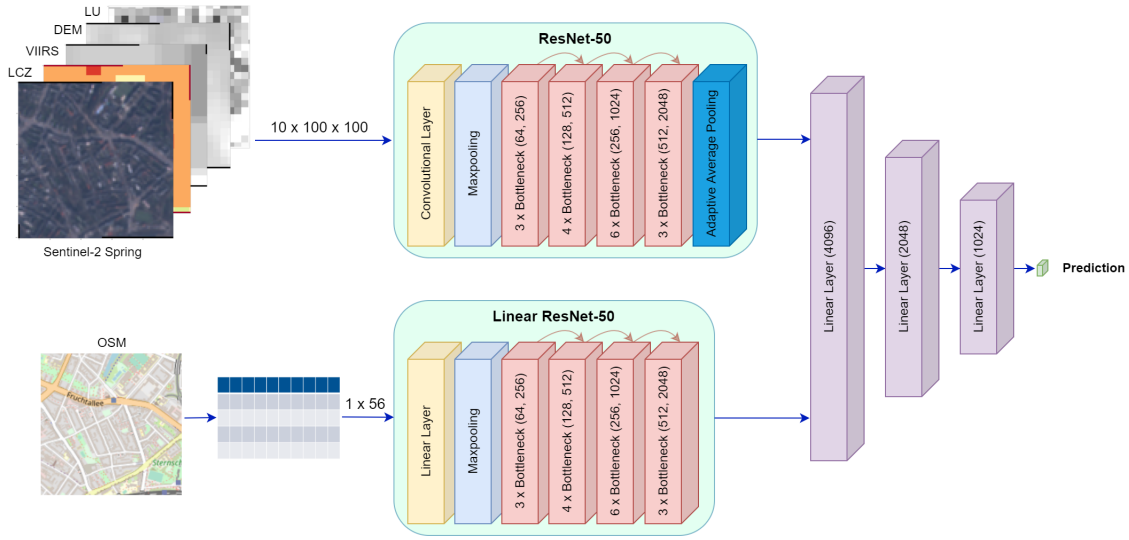


Figure 5.3: The proposed interpretable deep learning framework for population estimation. Image is taken from our own publication [5].

addition to random flipping and rotations, random brightness and gamma adjustments are also used to add randomized natural luminance shifts and scalings to the images with $beta = [0.8, 1.2]$ and $gamma = [0.8, 1.2]$, respectively [193, 194]. All data augmentation approaches are used with a 50 % probability of being applied to the image, resulting in a differently augmented image each time. The model output size for classification is set to 17 as indicated by the data set, whereas the model output for regression is a single value population count. For regression, the loss function is set to Mean Squared Error (MSE) and for classification, it is set to Focal Loss [195]. The complete method is written in Python 3.8 and implemented with the PyTorch 1.10 framework [196]. All models are trained on a single NVIDIA RTX 3090 GPU and 24GB of RAM.

5.2.2 Evaluation metrics

For regression, the commonly used evaluation metrics RMSE and MAE are employed. To calculate the proportion of variance in population counts captured by the model, R^2 is used. For classification, in addition to accuracy, balanced accuracy, and macro-averaged F1-score, another metric called class distance is used. Class Distance (CD) measures the distance between the actual and the predicted class label. The metric takes into account that a misclassification to a "nearby" class has a lower error than to a "far away" class due to the underlying regression task. The CD is calculated using the equation 5.1. The Mean Absolute Class Distance (MACD), which is basically the average of the class distances across all samples (N) is calculated using equation 5.2 .

$$CD = reference\_class_i - predicted\_class_i \tag{5.1}$$

Regression			
Sentinel-2 season	**RMSE**	**MAE**	**R^2**
Autumn	1579.59	548.93	0.747
Spring	1501.62	545.12	0.785
Summer	1776.90	562.74	0.680
Winter	1453.54	613.46	0.781

Table 5.4: Evaluation of different Sentinel-2 seasons on the test set for regression.

Classification				
Sentinel-2 season	**Accuracy(%)**	**Bal. Accuracy(%)**	**F1 score**	**MACD**
Autumn	56.52	35.43	0.355	1.01
Spring	57.63	37.34	0.378	0.977
Summer	57.58	36.32	0.369	0.981
Winter	55.27	36.85	0.377	1.25

Table 5.5: Evaluation of different Sentinel-2 seasons on the test set for classification.

$$MACD = \frac{1}{n}\sum_{i=1}^{N} |reference\_class_i - predicted\_class_i| \tag{5.2}$$

5.3 Experiments & Results

5.3.1 Relevance of input data sources

The initial experiments are carried out to determine which Sentinel-2 season to employ for further studies as the So2Sat-POP data set includes four seasons of Sentinel-2: autumn, spring, summer, and winter. This experiment setup omitted the vector branch and only the individual Sentinel-2 images for each season are fed to the model. Table 5.4 shows that Sentinel-2's spring season yields better performance on the MAE than the autumn (by 0.7 %), summer (by 3 %), and winter (by 11 %) and ~ 6 % on an average on R^2. Also, the classification achieved better-balanced accuracy, MACD and F1 scores in the spring season as indicated in Table 5.5. Therefore, in the following experiments, the Sentinel-2 spring season is utilized as satellite imagery.

To examine the relevance of other input data sources in the data set, different experiments based on various combinations of the data sources have been conducted using both the branches of proposed deep learning architecture. For regression and classification, models are trained following the "leave-one-out" principle in cross-validation except for

Regression			
Excluded data source	**RMSE**	**MAE**	**R^2**
None	1164.39	394.38	0.863
OSM	1216.65	422.18	0.849
DEM	1181.46	404.87	0.858
LU	1270.64	437.71	0.836
LCZ	1224.74	428.46	0.847
VIIRS	1168.89	386.64	0.861

Table 5.6: Evaluation of the relevance of input data sources using the test set by omitting each input data source once, except Sentinel-2 (spring) for regression [5].

Classification				
Excluded data source	**Acc.(%)**	**Bal. Acc.(%)**	**F1 score**	**MACD**
None	61.40	45.25	0.449	0.781
OSM	61.68	44.25	0.442	0.791
DEM	61.71	43.68	0.444	0.778
LU	58.44	37.68	0.374	0.887
LCZ	60.63	40.55	0.413	0.833
VIIRS	61.69	42.87	0.436	0.779

Table 5.7: Evaluation of the relevance ofinput data sources using the test set by omitting each input data source once, except Sentinel-2 (spring) for classification [5].

the Sentinel-2 spring season. Using this strategy, all data sources are included in the initial experiment and then each input data source is removed once for successive trials to see if its removal affects the outcomes. Each trained model is evaluated on the 18 unseen test cities. The results of this set of experiments are shown in Table 5.6 for regression and 5.7 for classification. In both studies, land use was found to be the most important input by improving balanced accuracy by 7.5 % in the case of classification and decreasing mean absolute error by 11 % in the case of regression. While the results deteriorate slightly in the absence of each input data source, the impact is the least when excluding VIIRS in regression and DEM in classification. In fact, the MAE slightly improved in regression when VIIRS was excluded. Similarly, in classification, the absence of DEM slightly improved the accuracy and MACD. When none of the data sources were removed, the best results were achieved on the majority of metrics. Therefore, it has been concluded that all input data sources are important for both regression and classification models.

Regression				
Model	**RMSE**		**MAE**	**R^2**
Random Forest	1276.26		463.35	0.827
Custom ResNet-50	1164.39		394.38	0.863
Classification				
	Acc.(%)	**Bal. Acc.(%)**	**F1 score**	**MACD**
Random Forest	59.13	37.95	0.383	0.896
Custom ResNet-50	61.40	45.25	0.449	0.781

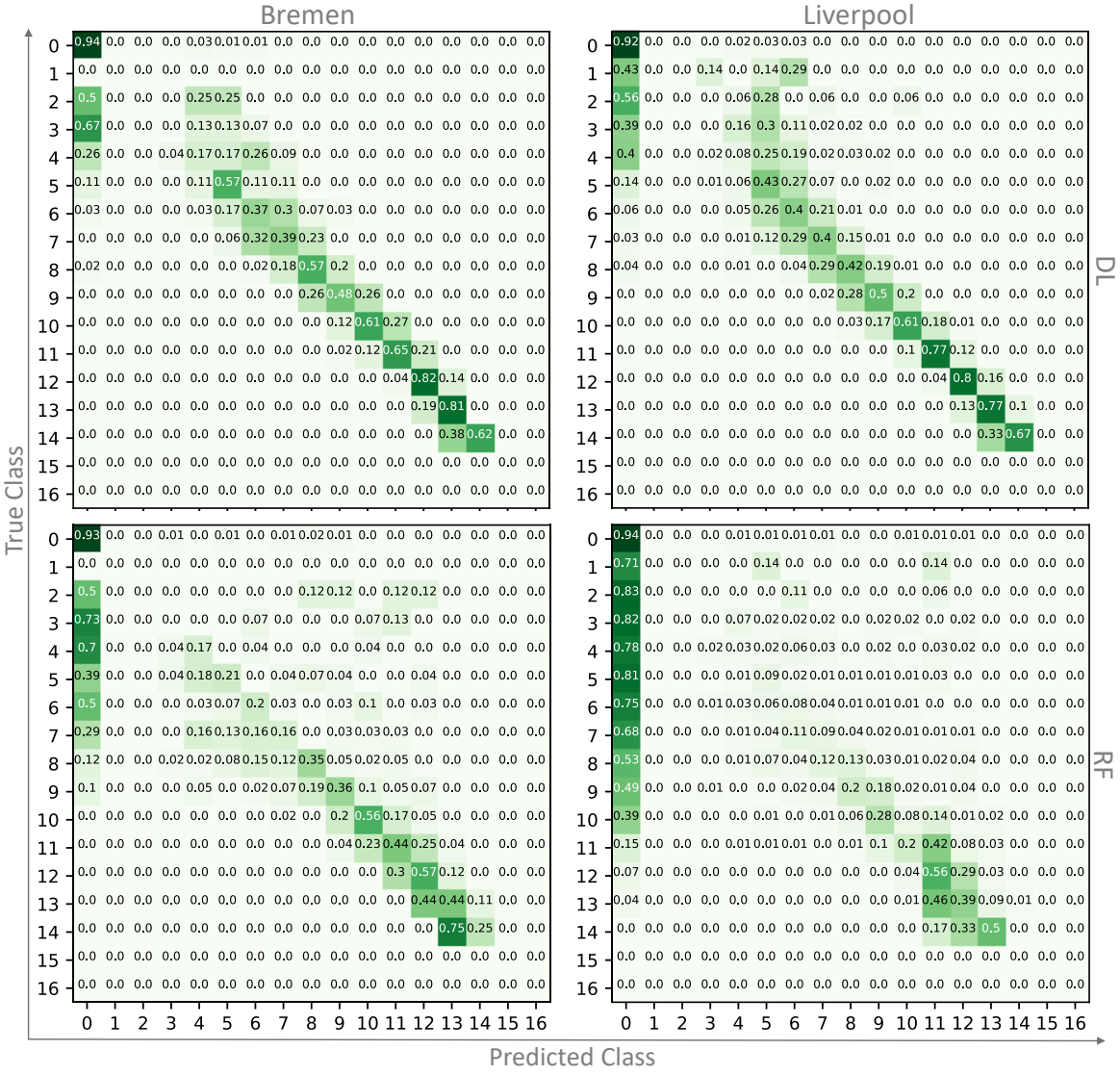
Table 5.8: Comparison of the best deep learning-based models with the baseline RF model. Across all criteria, the deep learning model outperforms the RF model [5].

5.3.2 Comparison with Random Forest

In chapter 4, RF is utilized as the baseline model on the So2Sat-POP data set [4]. Therefore, the best regression and classification model achieved in the experiments above is compared with the RF model. Table 5.8 shows that the baseline RF results have been improved across all metrics. There has been roughly an improvement of 7.5 % in balanced accuracy, as well as 15 % and 8 % improvements in RMSE and MAE, respectively. For the visual comparison, two top-performing, two average, and two worst-performing test cities are selected. In Figure 5.4 the normalized confusion matrix for each of these six selected cities are visualized. The confusion matrices show that, while the model is confident in predicting classes with moderate to high population densities, it performs poorly on classes with very low population densities (Class 1, 2, and 3) and very high population densities (Class 15 and 16). As highlighted in Figure 4.8 of section 4.5, patches from these very low population density classes are difficult to distinguish and frequently misclassified among themselves. On the other hand, as illustrated in Figure 5.2, there are very few samples in the classes with the highest population density, which leads to their poor classification performance. The RF model, for most cities, overestimates in low-population classes and underestimates in high population classes. For regression, the scatter plots of the predicted population counts versus the actual population count for each of these cities are plotted, as shown in Figure 5.5. The scattering in the deep learning model predictions is closer to the ideal fitting line for low to moderate population values and more dispersed from the ground truth values for higher population counts. On the other hand, for the RF model, a similar pattern has been observed as in the case of its classification results, with an over-prediction over low population values and an under-prediction across higher population ranges.

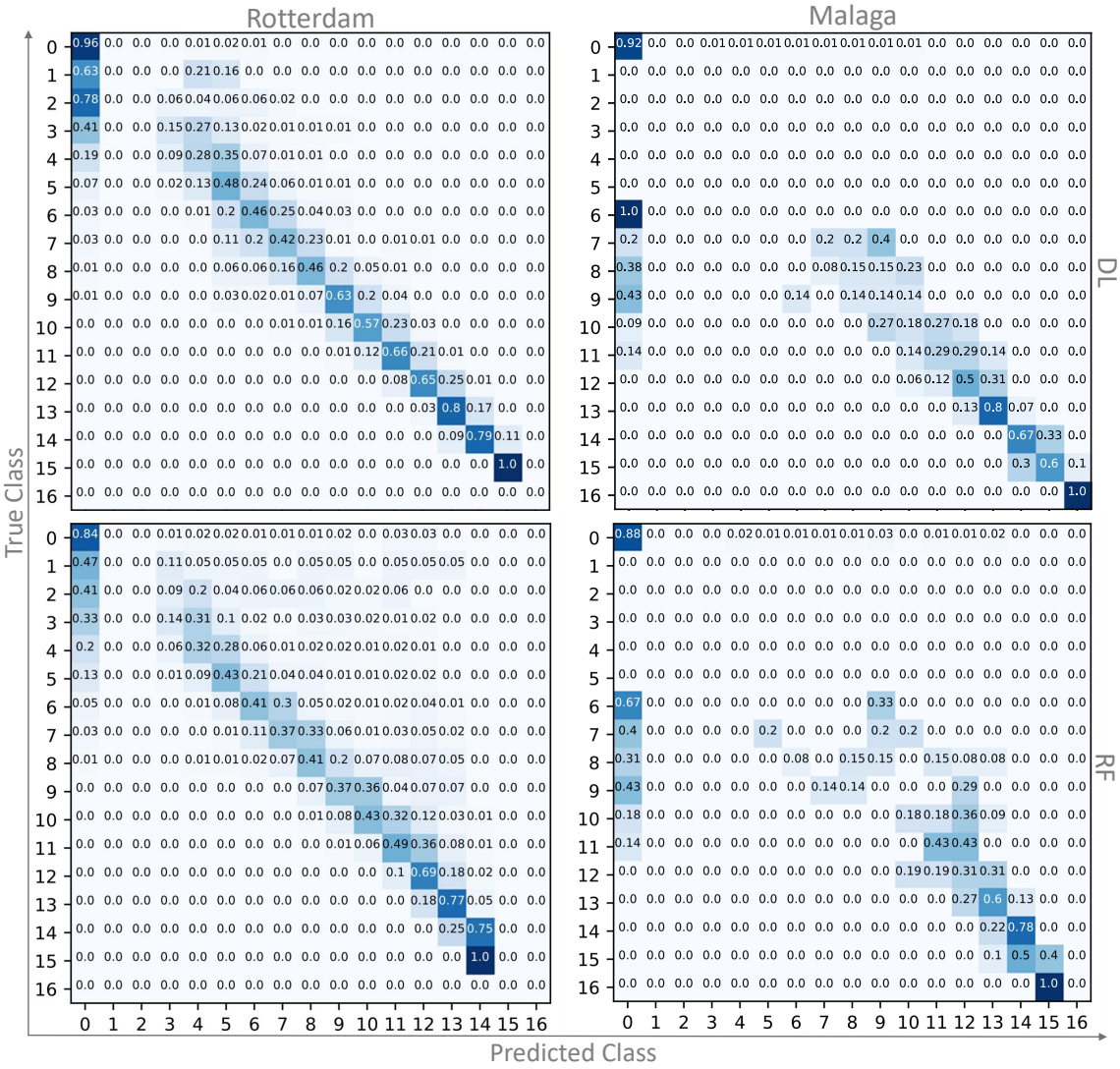
5.3.3 Comparison with GHS-POP

Using these six cities from above, a comparative study is conducted with GHS-POP, which has been collected and processed as supplementary data in section 5.1.2. Table



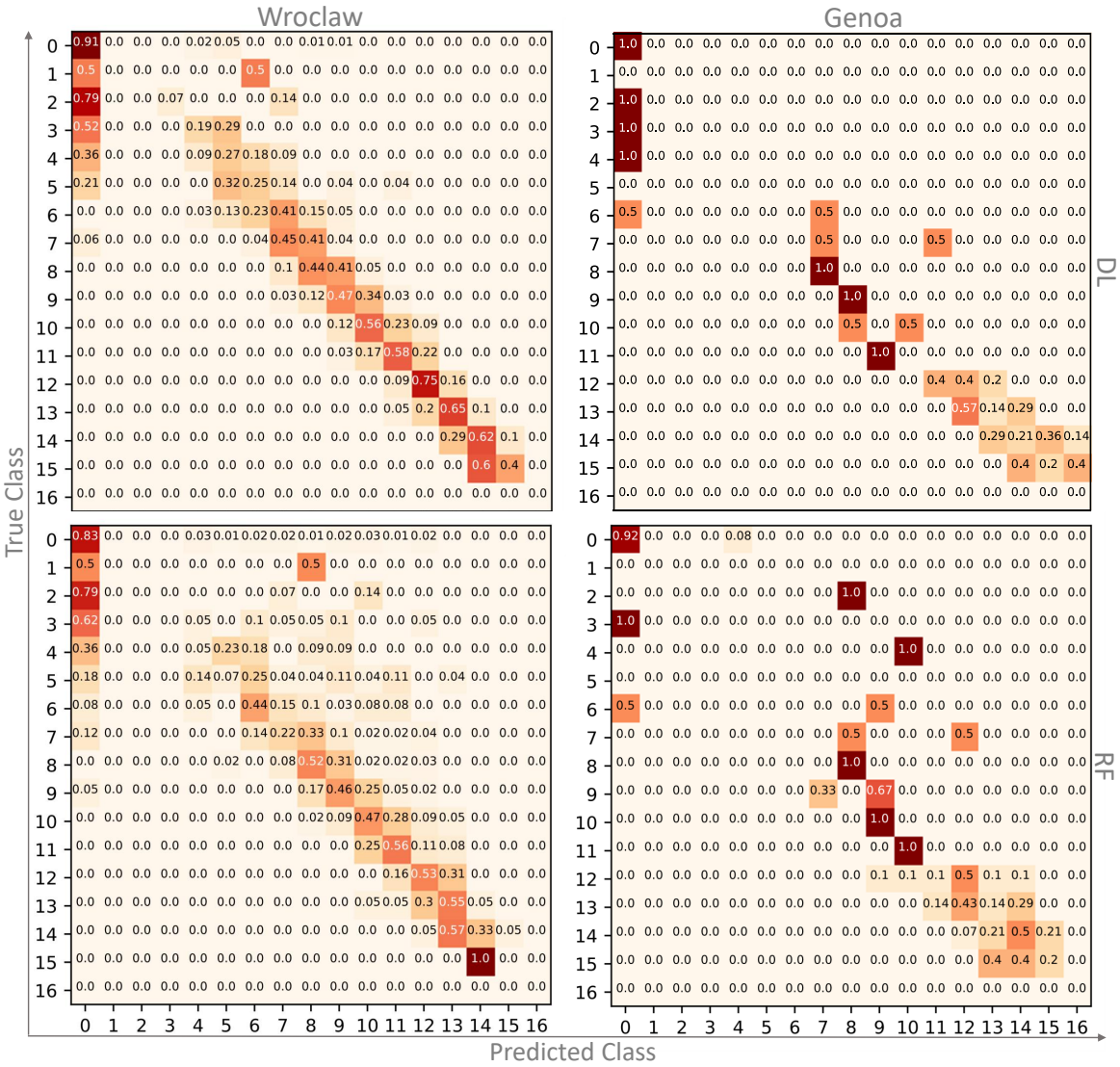
(a) Top-performing

Figure 5.4: Normalized confusion matrix of two top-performing cities (Bremen, Liverpool 5.4a), two average (Rotterdam, Malaga 5.4b), and two worst-performing cities (Wroclaw, Genoa 5.4c) test cities for our deep learning and RF model. Image is taken from our own publication [5].



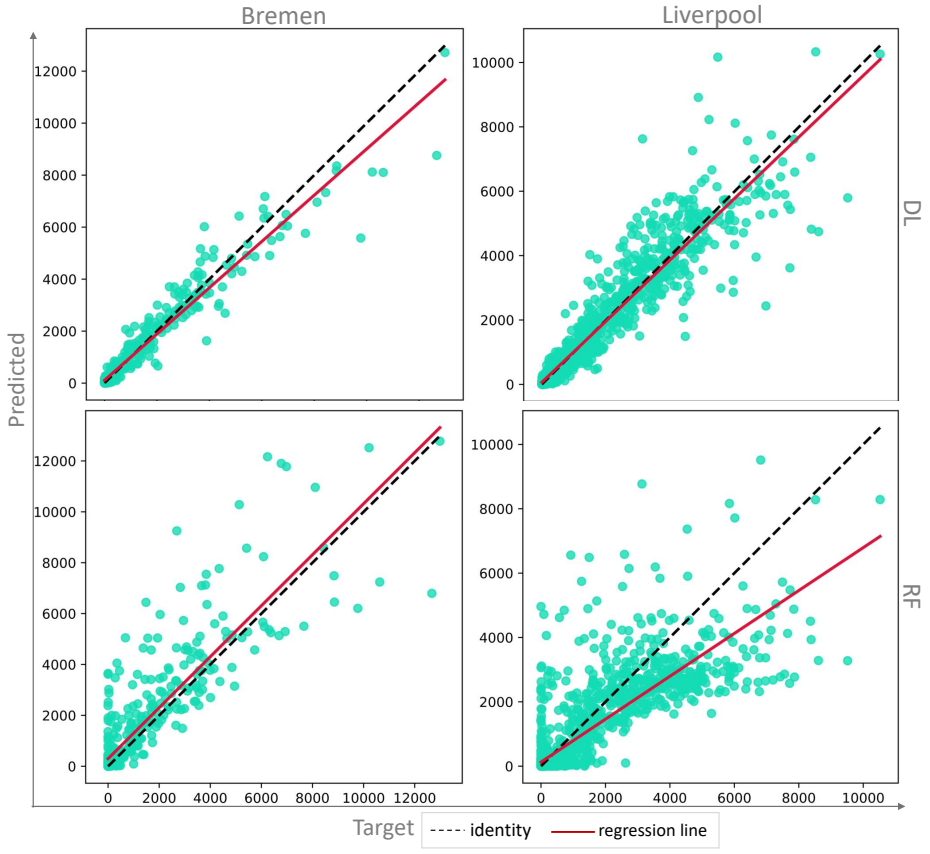
(b) Average-performing

Figure 5.4: Normalized confusion matrix of two top-performing cities (Bremen, Liverpool 5.4a), two average (Rotterdam, Malaga 5.4b), and two worst-performing cities (Wroclaw, Genoa 5.4c) test cities for our deep learning (DL) and RF model. Image is taken from our own publication [5].

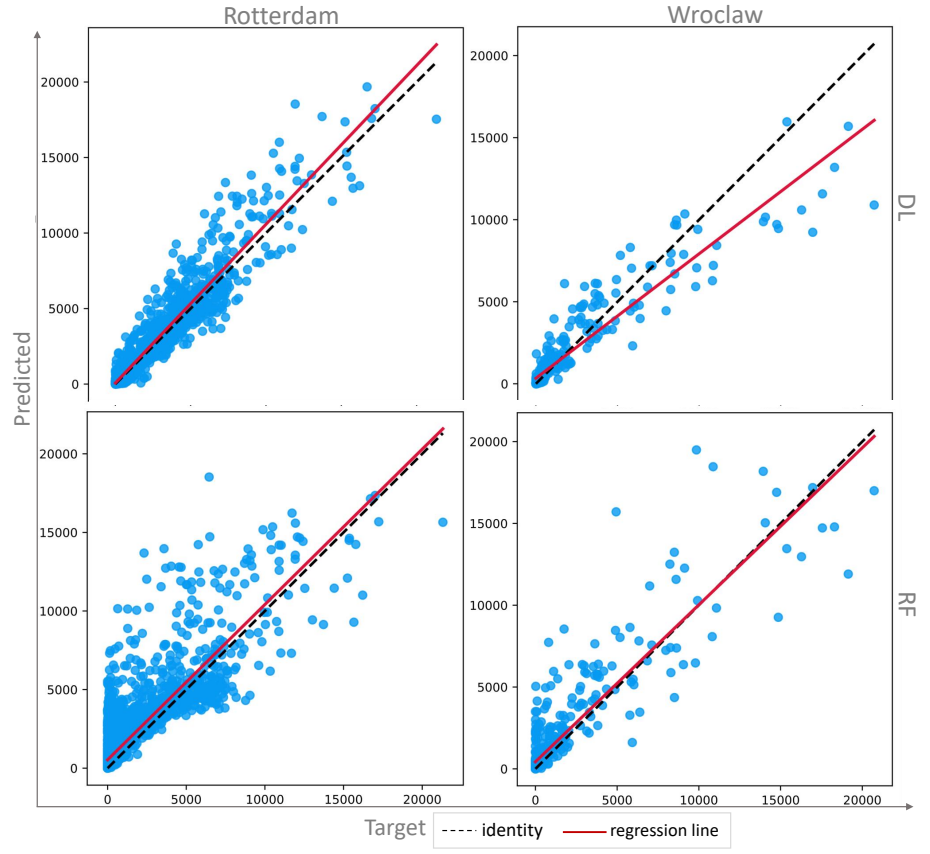


(c) Worst-performing

Figure 5.4: Normalized confusion matrix of two top-performing cities (Bremen, Liverpool 5.4a), two average (Rotterdam, Malaga 5.4b), and two worst-performing cities (Wroclaw, Genoa 5.4c) test cities for our deep learning (DL) and RF model. Image is taken from our own publication [5].

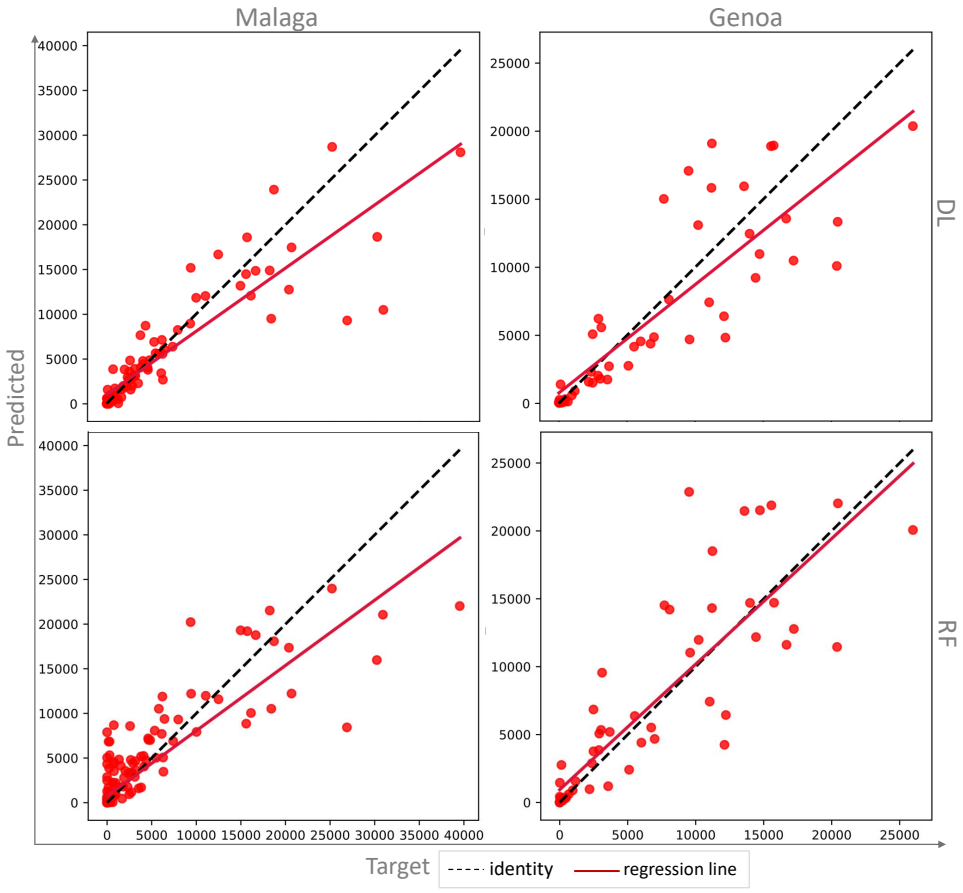


(a) Top-performing



(b) Average-performing

Figure 5.5: Scatter plots of our deep learning (DL) model predictions and RF model at the grid level for two top-performing cities (Bremen, Liverpool 5.5a), two average (Rotterdam, Wroclaw 5.5b), and two worst-performing cities (Malaga, Genoa 5.5c) test cities. The black dotted line, identity, represents the perfect fitting line and regression line, in red, indicates the trend in the model predictions [5].



(c) Worst-performing

Figure 5.5: Scatter plots of our deep learning model (DL) predictions and RF model at the grid level for two top-performing cities (Bremen, Liverpool 5.5a), two average (Rotterdam, Wroclaw 5.5b), and two worst-performing cities (Malaga, Genoa 5.5c) test cities. The black dotted line, identity, represents the perfect fitting line and regression line, in red, indicates the trend in the model predictions. Image is taken from our own publication [5].

Regression						
Cities	**Ours**			**GHS-POP**		
	RMSE	**MAE**	$\mathbf{R^2}$	**RMSE**	**MAE**	$\mathbf{R^2}$
Bremen	537.67	272.89	0.933	1530.27	892.60	0.461
Liverpool	616.86	308.88	0.887	1039.60	607.16	0.679
Rotterdam	884.05	427.06	0.890	1871.76	1097.83	0.509
Wroclaw	1184.88	518.58	0.856	2199.25	1096.29	0.515
Malaga	3758.04	1710.72	0.777	6492.98	3901.72	0.334
Genoa	3724.55	2652.69	0.686	2732.62	1906.80	0.831

Table 5.9: Quantitative comparison of our best regression model with GHS-POP on two top-performing (Bremen, Liverpool), two average (Rotterdam, Wroclaw), and two worst-performing-performing (Malaga, Genoa) test cities [5].

5.9 and 5.10 show a quantitative comparison of our deep learning-based model predictions versus GHS-POP estimations for regression and classification, respectively. On most of the observed evaluation metrics, our method outperforms GHS-POP. In regression, the improvements in RMSE for Bremen and Liverpool are up to 65 %, and for our two worst-performing cities, while Genoa performed poor than GHS-POP, Malaga still outperforms the GHS-POP with a 42 % improvement in RMSE. For classification, again, our model outperformed the GHS-POP across every evaluation metric for all of these cities except Genoa. The improvements in balanced accuracy ranges from 33 % in Rotterdam to 16 % in Malaga. A visual comparison is shown in Figure 5.6 and 5.7. As seen in these plotted population maps, GHS-POP does not capture heavily populated urban centers well. It under-counts the population in the city's densely populated central parts and does not discriminate between dense and sparsely populated areas very well.

5.3.4 Evaluation and comparison on inter-regional cities

Since the model was trained on the So2Sat-POP data set, which only contains European cities, it was tested for transferability and generalizability in a new geographical region. For this evaluation, a subset of three randomly selected US cities generated as supplementary data in section 5.1.2 is utilized. The model predicts a population class and a population count over all 1 x 1 km patches of each city. The estimations are again compared with GHS-POP and the reference population data from SEDAC. The results in Tables 5.11 and 5.12 for classification and regression, respectively, show that our model, trained using data only from European cities, does not clearly outperform the GHS-POP in US. In New York, our model outperformed the GHS-POP by 21 % improvement in the RMSE, but in San Jose, it underperformed than GHS-POP and nearly doubled the RMSE. A similar pattern is observed in classification. Nevertheless, our model performed in line with the GHS-POP, though it was never trained in the US.

Classification						
Cities	**Ours**			**GHS-POP**		
	Acc.(%)	**Bal. Acc.(%)**	**MACD**	**Acc.(%)**	**Bal. Acc.(%)**	**MACD**
Bremen	54.15	46.67	0.096	23.42	17.28	1.89
Liverpool	53.03	41.78	0.179	19.13	13.56	2.38
Rotterdam	51.19	48.88	0.071	15.52	11.74	2.39
Malaga	42.26	41.19	0.989	25.77	20.66	1.63
Wroclaw	36.06	42.03	- 0.08	19.91	15.46	2.23
Genoa	22.00	15.05	0.340	56.00	31.90	0.92

Table 5.10: Quantitative comparison of our best Classification model with GHS-POP on two top-performing (Bremen, Liverpool), two average (Rotterdam, Malaga), and two worst-performing (Wroclaw, Genoa) test cities [5].

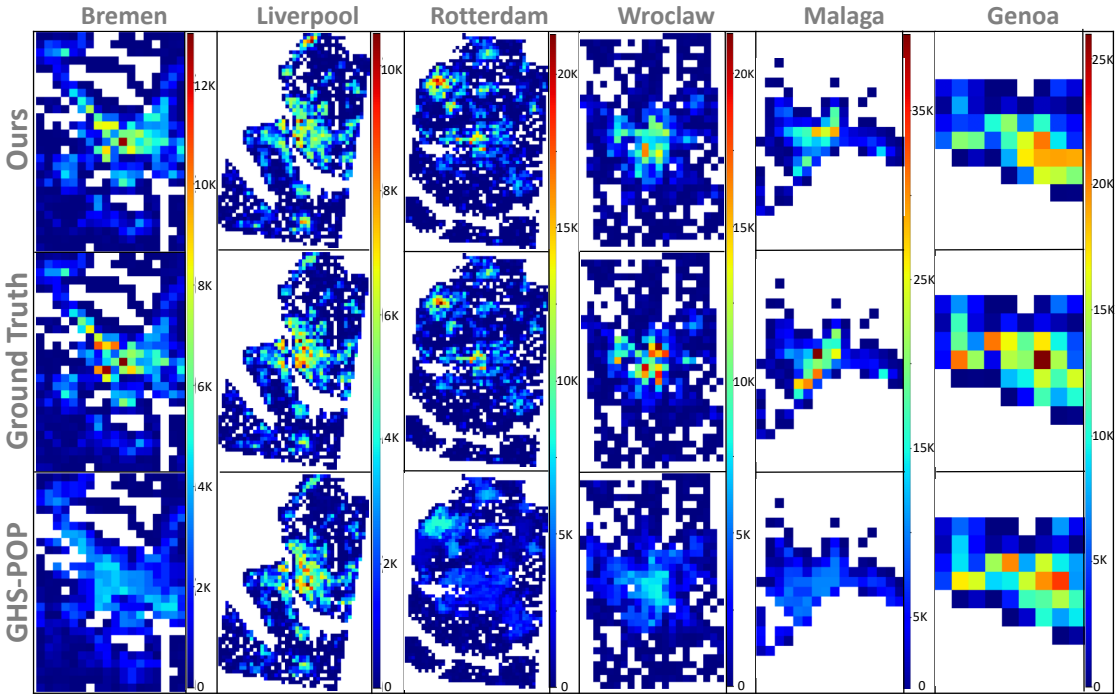


Figure 5.6: Comparison of two top-performing cities (Bremen, Liverpool), two average (Rotterdam, Wroclaw), and two worst-performing cities (Malaga, Genoa) with GHS-POP for regression. Please note that the population counts are in thousands. Image is taken from our own publication [5].

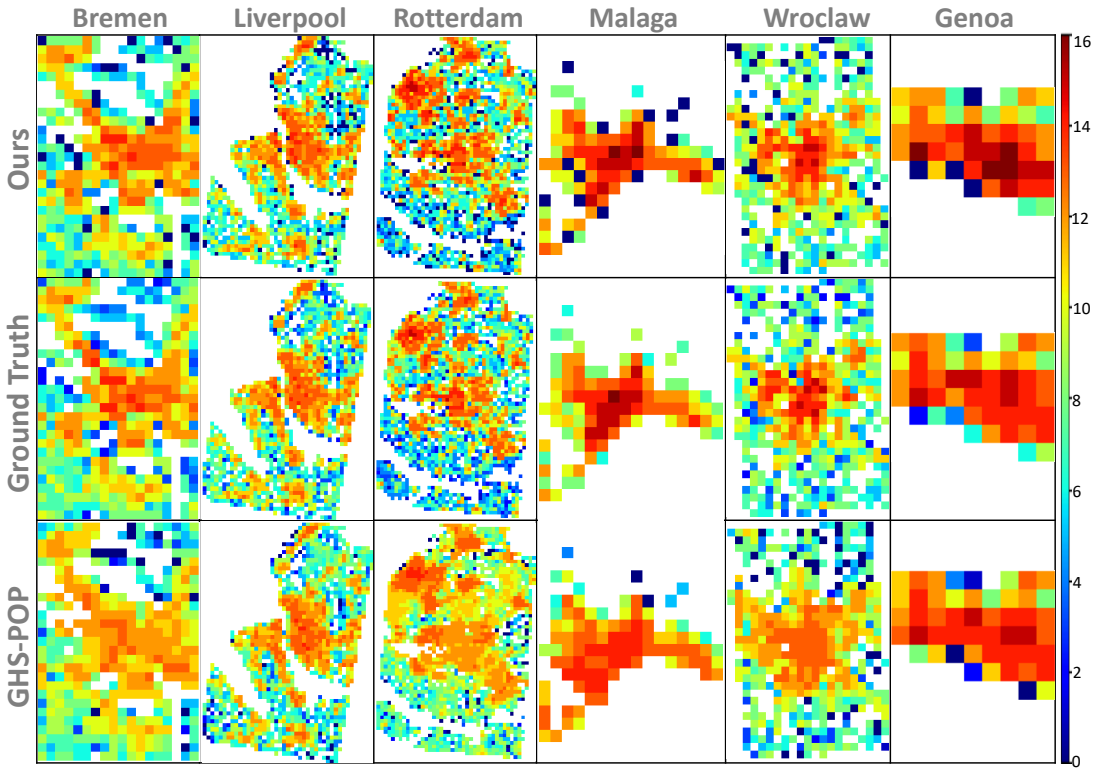


Figure 5.7: Comparison of two top-performing cities (Bremen, Liverpool), two average (Rotterdam, Malaga), and two worst-performing cities (Wroclaw, Genoa) with GHS-POP for classification. Image is taken from our own publication [5].

Regression						
Cities	**Ours**			**GHS-POP**		
	RMSE	**MAE**	**R^2**	**RMSE**	**MAE**	**R^2**
New York	1615.96	674.10	0.60	2042.18	718.14	0.38
San Jose	1550.16	611.40	0.16	761.54	338.71	0.35
Denver	420.69	264.32	0.21	447.03	174.62	0.17

Table 5.11: Quantitative comparison of our best Regression model (trained with European cities only) with GHS-POP on three random US test cities [5].

Classification						
Cities	**Ours**			**GHS-POP**		
	Acc.(%)	**Bal. Acc.(%)**	**MACD**	**Acc.(%)**	**Bal. Acc.(%)**	**MACD**
New York	18.70	21.10	2.04	12.23	8.19	3.53
San Jose	38.20	15.46	0.84	41.29	25.73	2.04
Denver	23.48	7.14	2.92	34.02	30.48	1.46

Table 5.12: Quantitative comparison of our best Classification model (trained with European cities only) with GHS-POP on three random US test cities [5].

These preliminary quantitative results suggest that our methodology has the potential to be applied to different geographic regions, implying transferability. Of course, by fine-tuning the model using a local micro-census, the model's performance in a new region could be considerably improved.

5.4 Interpretability

In the previous experiments, the results demonstrated that the deep learning models could be used to reliably estimate the population and has the potential to help data-driven decision-making. However, in order to ensure its acceptance and use by the key stakeholders, its transparency must be improved by revealing its inner workings beyond predicted performance. Therefore, an explainable AI (xAI) module is fused with the proposed deep learning framework which examines the outcomes of the blackbox model. Figure 5.8 depicts the complete framework. The xAI module is based on the Integrated Gradients (IG) saliency method [197] to reveal the relevant features used by the model for population estimation. While most explainability approaches are specific for image inputs, the IG method can attribute multi-model inputs, which is also the case for the proposed deep learning model, which has both image and tabular input data. Furthermore, IG satisfies two fundamental axioms, namely sensitivity and implementation invariance. The sensitivity axiom guarantees that if an input feature changes the model

scores in any way, then the attribution to that feature should be non-zero. On the other hand, the implementation invariance guarantees the feature attributions of two functionally equivalent models should also be identical.

The feature attribution for an input example x is defined as the integral of the gradients for the model predictions on examples that lie along the path from x' to x, where x' is the baseline input that signifies the absence of a feature in an input. In practice, the integral is approximated with a summation, and the importance $I_d(x)$ for a feature d of the input example x is computed with the following equation [197]:

$$I_d(x) = \frac{(x_d - x'_d)}{m} \sum_{k=1}^{m} \frac{\partial f\left(x' + \frac{k}{m}(x - x')\right)}{\partial x_d} \tag{5.3}$$

In the framework for population estimation, x is the multi-modal input example consisting of images and tabular data, x' is the baseline input consisting of a black image and a zero OSM vector, and f is the prediction of the ResNet-50 model for population estimation. And, m is the number of interpolation steps in the path from x' to x.

Applying this method, the feature attribution maps for a few examples from the test data are visualized in Figure 5.9. In this example, the feature attribution map and the corresponding Sentinel-2, LCZ, land use, and OSM features are plotted. The first two instances are selected where the estimations are entirely correct. The predicted population in the first instance is 4747, as is the actual population count. The model correctly identified the settlements, as demonstrated in the feature attribution map. Similarly, in the second instance, the actual and predicted population count is 3. The model focuses on the built-up areas in the upper-left corner and distinguishes them from the natural surroundings, which in this case are vegetation and bare soil. The street statistics such as length, count, and proportions are among the most important OSM elements. In the third case, the reference population count is 11 and the predicted population count is 216. Although the information from the land use data is absent in this case, a built-up region is clearly visible in the corresponding Sentinel-2 and LCZ patches. Despite the fact that the predicted population count does not match the reference population count, the feature attribution map reveals that the model accurately recognized the settlements in the areas. This mismatch could be due to discrepancies in the reference data as a result of the time lag between the acquisition of reference population data and the other corresponding input data. The fourth instance represents the port of Genoa with a reference population count of 126, however, the population count was significantly overestimated to 1400. Its feature attribution map shows that the model is looking at the upper left corner of the image, which appears to be a built-up region in the Sentinel-2, LCZ, and land use data. Furthermore, the model discovered meaningful features on the right side, resulting in an over-prediction by the model. This area has been investigated and it was discovered to be a dock container terminal. It is densely packed with big containers that may easily be misidentified as house roofs in satellite imagery. As a result, the model incorrectly interprets it as settlements and overestimates the population in this patch. This example illustrates some of the limitations of utilizing satellite data

to estimate the population as in some instances, the physical characteristics of built-up areas in satellite images are not discernible even to humans.

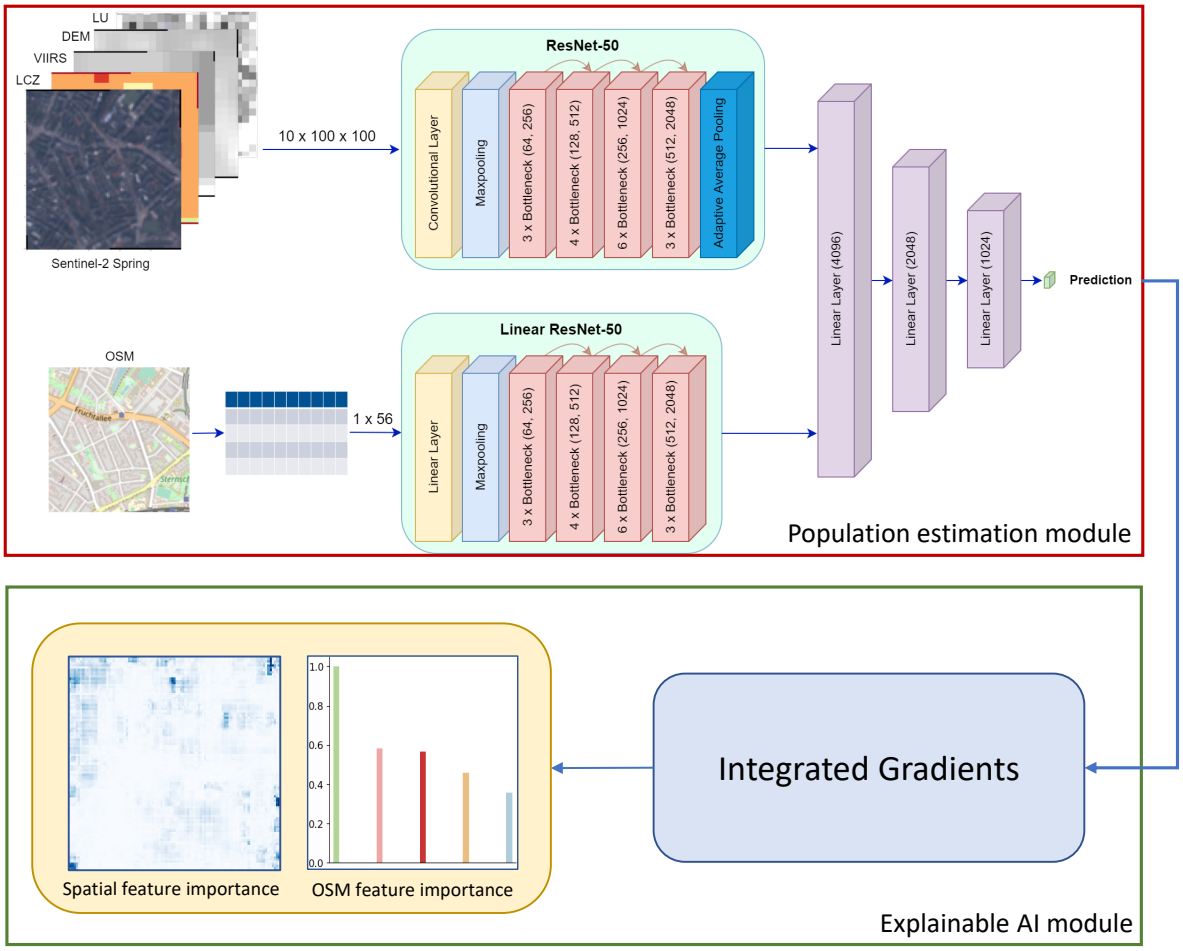


Figure 5.8: The deep learning based population estimation module integrated with the explainable AI module [5].

5.5 Summary

In this work, an interpretable deep learning framework is proposed for estimating the population at a consistent resolution of 1 km using only publicly available data sources. The deep learning architecture is customized to handle raster and vector data at the same time and predicts the population by generalizing across countries. The emphasis has been placed on the method's transparency so that it may be used in real-world applications such as urban planning, infrastructure development, risk assessments, and so on. The model is trained using the So2Sat-POP data set that has been mentioned in chapter 4. The evaluations done on the 18 unknown test cities showed promising

results. A comparative analysis has also been carried out in a few US and European cities with a popular community standard product, GHS-POP. In most European cities, a better performance than the GHS-POP has been achieved. However, due to a lack of non-European training data, our estimations did not clearly outperform GHS-POP in US cities. Nevertheless, our model performed in line with GHS-POP and this geographically heterogeneous evaluation illustrates the method's transferability.

Using an explainable AI technique, the most relevant features used by the model to reach an outcome are assessed. This additional interpretation of the model's decisions would not only promote trustworthiness but may also be used as a reference to compare the functioning of deep learning models beyond their predictive performance. The feature attribution maps in the explainability analysis reveal that the model is capable of detecting the built-up areas and used them as the most relevant features. This also supports the intuition that land use data is a crucial predictor of population. However, in some cases, the model could not distinguish between the residential and non-residential built-up regions. The addition of high resolution satellite data and detailed building function maps could further help to improve the results.

Also, the model is trained only on European cities, which means that the model's estimation could suffer a high bias in very densely or sparsely populated regions such as India, China, and Mongolia or regions with very different architectural or cultural peculiarities compared to Europe. Including these cities would also help to balance the So2Sat-POP data set since there are only a few samples from the high population density regions. Thus, expanding the training data or fine-tuning with the local micro-census might boost the generalizability and performance of the model. Nevertheless, even in the absence of census data, the framework could be utilized to generate more accurate and interpretable population estimation maps at a large-scale.

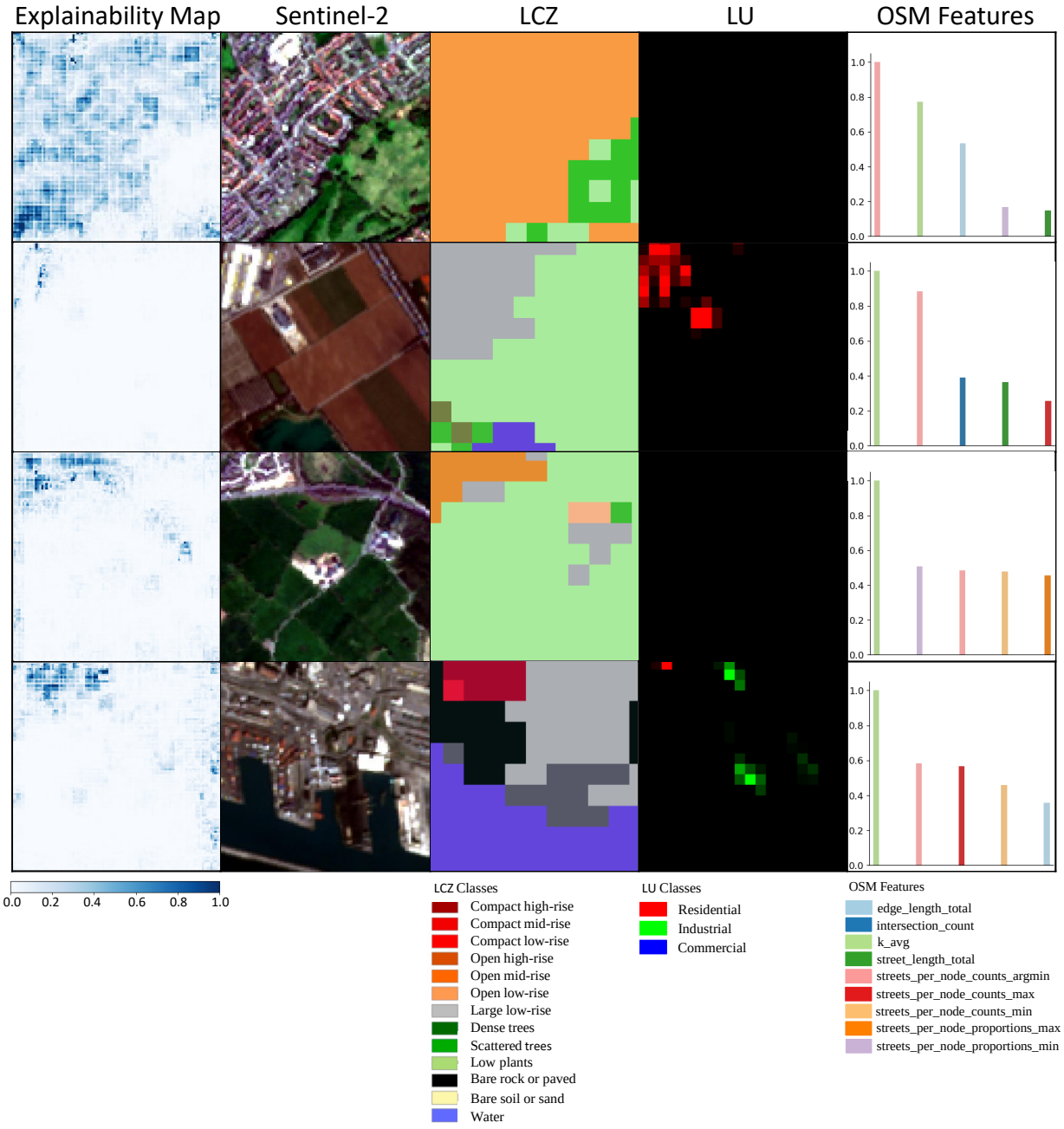


Figure 5.9: Explainability maps for a few examples from the test set, which represents the calculated attribution score. Image is taken from our own publication [5]. A higher feature attribution score implicitly indicates the higher importance of that feature for the model's prediction. Only Sentinel-2, LCZ, Land Use (LU), and OSM patches as they allow for visual interpretation of the semantically significant features. Detailed documentation about the OSM geometric and topological network features can be found at OSMnx [6] user reference (`https://osmnx.readthedocs.io/en/stable/osmnx.html`)

6 Building level population estimation

One of the primary issues of census data, as previously stated in section 2.2, is its resolution is usually at census administrative units, making it impractical for certain applications [129]. One example of an application being impacted by such national census constraints is the Global Polio Eradication Initiative (GPEI) effort in Nigeria. GPEI regularly undertakes vaccination programs intending to immunize every child under the age of five [198]. Nigeria's most recent national census, conducted in 2006, provides the population counts and distribution at the Local Government Area (LGA) level. This level of aggregation did not allow the identification of every vaccine-eligible child under five and the campaign might miss some. The scarcity of available comprehensive geo-demographic data and its limited access makes such ambitious projects extremely difficult to implement and execute efficiently [199]. Fine-grained population estimation could be beneficial in a variety of domains, including urban planning, resource allocation optimization [137, 138], natural disaster management [139, 140], public health [141], and as a foundation for various other applications.

Some studies have tried to improve the resolution of the available population data by employing deep learning approaches [135, 136, 143], however, the spatial resolution remains low. Where people live correlates strongly with the buildings; therefore, building level population estimation would be the most precise and finest level source [144]. Recently, few studies have examined different methods for estimating population counts at the building level [129, 132, 145, 146, 147]. These studies mapped census population counts to buildings using high resolution satellite imagery or ancillary data sources such as land use/land cover maps, night lights, and other socioeconomic indicators. Some of these methods solely rely on building volumes acquired from LiDAR [129] or digital surface models (DSM) [147], which are not available everywhere [132]. The majority of these works used handcrafted features to disaggregate the available census data to buildings [132, 145, 146, 147], which limits their transferability. Additionally, due to the diversity in regional input data, each study has its own framework, making it difficult to standardize and compare the methodologies [132]. Nonetheless, their preliminary findings open the field for fine-scale population estimation for further exploration.

In contrast to the handcrafted features used in the related building population estimation studies [129, 132, 145, 146, 147], in this work, a hybrid deep learning-based approach is employed. The method uses the high resolution satellite imagery and building-related data to predict the population at the 1 km and subsequently redistributes the estimated population counts to the buildings. The predicted building population maps are aggregated to 100 m for evaluation purposes and compared with the popular state-of-the-art population data set, WorldPOP.

While satellite imagery is widely accessible, building data sets are often constrained by their spatial coverage and quality. Therefore, in this work, using a case study in two Bavarian cities, the significance and impact of data quality in the context of fine-scale population estimation is investigated. Besides the widely available OSM open data, high-quality regional data from open-access official sources is also collected. The building-related data collected from sources that are readily accessible on a large-scale is called *coarse* level data and the more comprehensive data collected from local government agencies is referred to as *granular* level data. For the first time, a population estimation study uses the Bavarian Surveying Administration's recently released open geodata and compares them quantitatively and qualitatively to crowd-sourced platform data sets. The comparative results indicate a clear advantage of using regional governmental sources over crowd-sourced platforms. This could encourage the government administrative offices to openly publish the geodata, which could be very useful for urban research and development. To the best of our knowledge, this is the first effort toward investigating a deep learning method coupled with OSM and Bayern open-access EO data to estimate the population at the building level and study the impact of varying data quality.

6.1 Study area

The method has been demonstrated using the two largest cities in Bavaria, Munich and Nuremberg (Figure 6.1a). The city of Munich is the most densely populated municipality in Germany, with a population density of 4777 inhabitants per km^2 [200] and Nuremberg is the second-largest city in Bavaria. Since 2000, Munich and Nuremberg have been Germany's two fastest-growing metropolitan regions, making them excellent examples of thriving cities in Bavaria [201]. Due to the city's tremendous expansion outwards, an expansion algorithm suggested in section 4.1 is utilized. This algorithm expands the city to accommodate the city's ongoing urbanization. Table 6.1b shows the overall statistics for the expanded version of the cities. This extension technique expands the administrative area of both cities by approximately six times, resulting in a good balance of urban and suburban areas in the study.

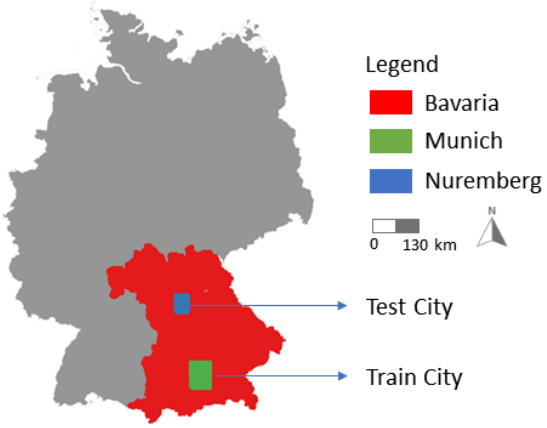
6.2 Data

For both cities, the reference population data, high resolution satellite imagery and the building data sets, which include building functions, building areas and building heights, are collected.

6.2.1 Input data

Population data

A good reference population data is crucial to develop an accurate population estimation method. Therefore, the population grids are collected from official statistical offices



(a) Study area in Germany. Munich and Nuremberg are used as train and test cities, respectively.

City	Area(km^2)	Population
Munich	2,052	2,016,101
Nuremberg	962	892,874

(b) Overview of the extended cities in Bavaria, Germany.

Figure 6.1: Location and statistics of the study area.

at spatial resolutions of 1 km and 100 m in order to assess the grid level and building level population estimates, respectively. The 1 km-resolution population grids is from the ESSnet project in partnership with the European Forum for Geography and Statistics (EFGS) and are collected at a spatial resolution of 1 km. The 100 m-resolution population grids are collected from Germany's Federal Statistical Office. These grids were generated by utilizing information on people, buildings, and apartments for every address that is allocated to a grid cell size 100 x 100 m based on its geographic coordinates [202]. Thus, the population count for a grid cell is obtained by aggregating the counts from all of its allocated addresses, yielding a population count at 100 m. More information regarding its method could be obtained on the project website [202], in the document titled "Explanatory notes on Demographie am 100 Meter-Gitter".

Satellite imagery

Satellite imagery has a strong potential to track the physical environment and human footprint. Several studies have used satellite imagery to estimate populations [16, 17, 18, 203]. While high resolution satellite imagery is expensive and may not be publicly available, alternatively, Google Earth satellite imagery is utilized. It is a large-scale, freely accessible imagery subject to some restrictions for commercial purposes. It provides a clear view of buildings, streets, vegetation, etc., and therefore could be sufficiently utilized for urban-related applications. The noise caused by the heterogeneity in the image acquisition dates of this imagery lies in the range of heterogeneity in the remaining data sources as well as the ground truth census data. The satellite imagery is automatically retrieved, within the limitations of Google's usage agreements by using the freely available Geographic Information System (GIS) software product *SASPlanet*. The bounding box coordinates of the retrieved city extents have been used as region of

interest in selection for the batch download in the zoom level 20 of the Google satellite layer.

Building function

Building-related data offers a reasonably accurate representation of population trends [204] and could be particularly useful for urban studies. OSM (http://www.openstreetmap.org) is a large-scale open and crowd-sourced mapping tool where millions of contributors gathered the geographic and descriptive information for buildings with good coverage for the urban areas of Germany [205, 206]. Therefore, the building footprints from OSM are extracted for the study area using the Geofabrik [207] OSM service provider. The downloaded OSM data comprises buildings as polygons and the function of the buildings as building tags. For the *coarse* level, the OSM data is utilized as it is. While the building footprint is reasonably complete in OSM, building function tags are not [132]. Therefore, for the *granular* level, the building functions from the official German cadastral land register ALKIS (Amtliches Liegenschaftskatasterinformationssystem [208] are collected. The "Actual Use" (Tatsächliche Nutzung) layer, a component of ALKIS, published by the Bavarian Surveying Administration, describes the utilization of the buildings in detail [209].

Building heights

Building heights are a key element in reflecting the varied pattern of buildings within a region. Due to a lack of data, information on urban building height across vast areas is currently scarce [210]. A few modeled products are available in the literature. However, their quality is still debatable because they have been generated utilizing multiple regional input data and methodologies, thus, also not available on large-scale [210, 211, 212, 213]. Therefore, for *coarse* level, a large-scale open database called EUBUCCO is utilized. This data set is based on publicly available government data sets and OSM that have been gathered, harmonized and partly validated [214]. Milojevic-Dupont et al. built the EUBUCCO scientific database of individual building footprints for nearly 206 million buildings across the 27 European Union countries and Switzerland, with the completeness of 74 % for the building heights attribute [214]. For the *granular* level, again the ALKIS, which provides the open-source CityGML 3D building models (3D-Gebäudemodelle (LoD2)) [215] is utilized.

6.2.2 Data Preparation

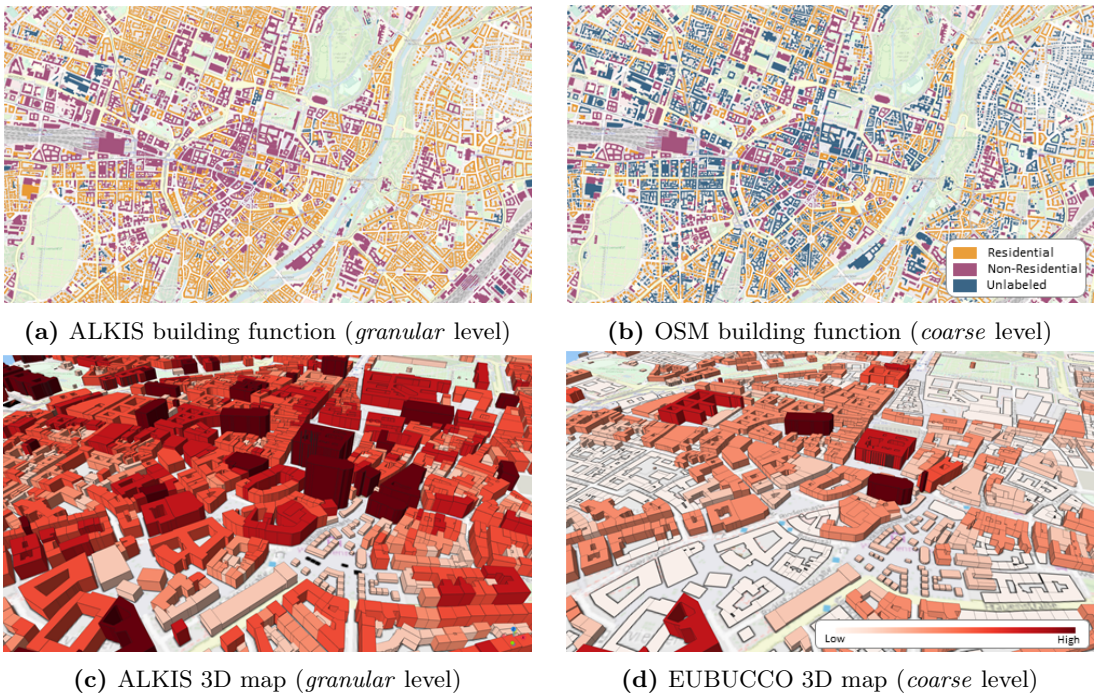
All of the aforementioned input data sources for Munich and Nuremberg have been gathered. Each input data has been cropped using the city extents established by the expansion algorithm. Since the data is gathered from different sources, they are represented in distinct CRS. The reference population grid is represented as EPSG:3035 - ETRS89-extended/LAEA Europe. Therefore, all input data from their associated coordinate systems are reprojected to the EPSG:3035 CRS in order to align them with the population grid.

Building function data gathered from OSM and ALKIS contains a variety of land use tags. OSM has 147 distinct land use tags for buildings in the study area, while ALKIS has 103 collected from two land use attributes: usage type (Nutzungsart) and detailed descriptor (Bezeichnung). A mapping scheme has been devised to homogenize the land use categorization of buildings, with each tag mapped to a reduced building use classification: commercial, industrial, residential, and other, with certain buildings remaining unlabeled due to the lack of any building use tag. Detailed mapping scheme can be found in Appendix Tables A.1, A.2, and A.3. Table 6.1 summarizes the completeness of OSM and ALKIS building function tags in Munich and Nuremberg. Using the reduced building function tags, data is further processed to create the binary residential/non-residential building masks with their corresponding built-up area.

Building Function	**Munich**		**Nuremberg**	
	OSM	ALKIS	OSM	ALKIS
Commercial	1.47 %	3.21 %	0.92 %	3.14 %
Industrial	0.57 %	3.96 %	0.37 %	4.40 %
Residential	30.16 %	68.92 %	17.85 %	72.22 %
Other	19.52 %	22.24 %	15.57 %	17.87 %
Unlabelled	48.27 %	1.67 %	65.29 %	2.34 %

Table 6.1: The percentage of buildings that fall into each of our simplified building use classification schemes, as well as those that remain unlabeled.

3D building models are available in ALKIS as CityGML data with Level of Detail 2 (LoD2). CityGML is a data modeling standard for semantic 3D cities and landscape models based on the Geography Markup Language (GML) [216]. The GML files are parsed to extract the building polygons and their corresponding height values. The height values represent the vertical disparity between the lowest terrain intersection and the highest point of the roof. In EUBUCOO, the data is provided as a geo-package (GPKG), which is another Open Geospatial Consortium (OGC) standard for the exchange of geospatial data. Similarly, the GPKG files are parsed to extract and assign height values to all buildings in the study area. However, due to missing data, around 3.9 % and 80.9 % of the buildings in ALKIS and EUBUCCO, respectively, remain without a height value. Figure 6.2 shows the building footprint extracted from OSM for a few buildings in the Munich city center, as well as their associated building usage, as indicated by (a) ALKIS and (b) OSM data and 3D building map generated using the (c) ALKIS and (d) EUBUCCO data sets. In this example, residential buildings refer to the predefined simplified building use classification scheme's residential class. In contrast, non-residential buildings comprise commercial, industrial, and other classes, with some unlabeled buildings in both scenarios. As can be seen, the completeness of the data in crowd-sourced platforms and official sources differ significantly.



(a) ALKIS building function (*granular* level)

(b) OSM building function (*coarse* level)

(c) ALKIS 3D map (*granular* level)

(d) EUBUCCO 3D map (*coarse* level)

Figure 6.2: Building footprint for a few buildings of Munich city center and building function maps generated using the (a) ALKIS and (b) OSM building use data, building-height map created using (c) ALKIS and (d) EUBUCCO data set.

After all input data has been processed and cleaned, ESSnet project population grid cells are used as a reference to crop the 1 x 1 km patches from the corresponding input data. The complete preprocessing pipeline is shown step-by-step in Figure 6.3.

6.3 Method

Being excellent at computer vision tasks, convolutional neural networks (CNN) have been effectively used in state-of-the-art deep learning-based methods for population estimation [16, 17, 18, 19]. As a result, in this study, experiments are conducted with a modified version of ResNet18 [217], pre-trained on the ImageNet data set as Xie et al. noted that pre-training on standard data sets could be useful for satellite imagery as well [218]. ResNet has proven to be a reliable architecture for population mapping [17, 53, 136], but previously has not been utilized in building level population estimation studies. The input layer of the model is modified to handle inputs of size 334 x 334 x C (width $\times$ height $\times$ channels) using the RGB bands from satellite imagery and the single bands from residential masks, building area, and height rasters. Also, the output layer is adapted for regression and reduced to a single value to predict a population count.

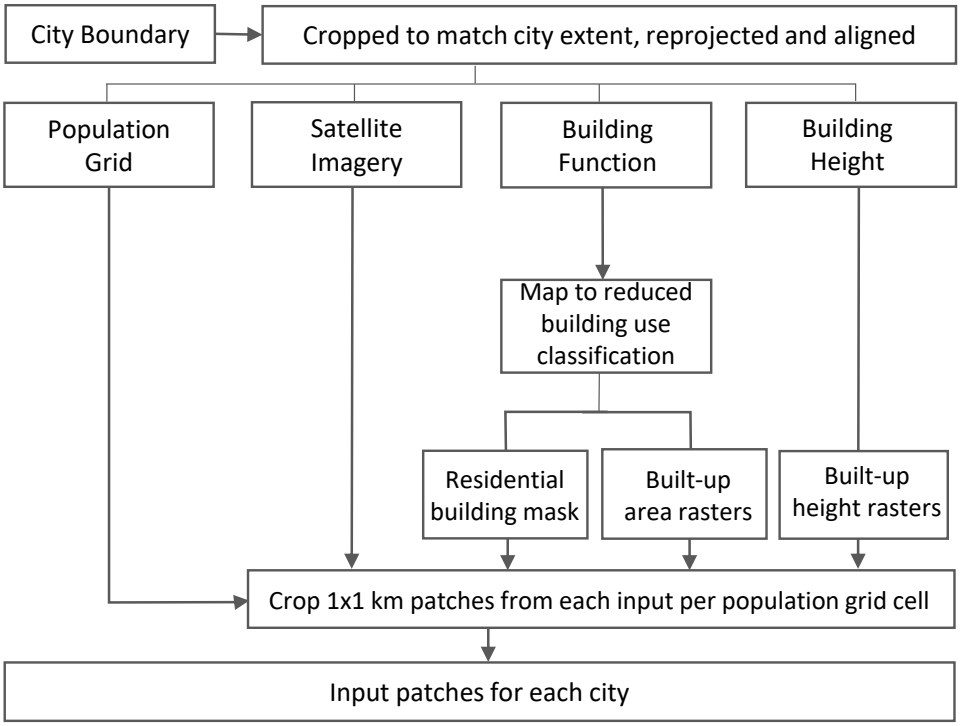


Figure 6.3: Step-by-step preprocessing of all the input data sources to create the corresponding input data for each city.

6.3.1 Experimental setup

The model is trained to predict a population count at 1 km^2 using four different scenarios. In each setup, a new input layer is added to determine the significance of each input data. Table 6.2 shows these four configurations, together with the data used in each case. The outcomes are evaluated for each of these described scenarios to quantify the added value achieved by adding an additional level of detail in each case. In addition to the satellite imagery used in Scenario 1, all experiments in other scenarios are carried out at two alternative building data quality levels, *coarse* and *granular*, in order to compare and contrast the results obtained for each set of studies. For all the experiments, 80 % of the grid cells in Munich are utilized for training and the remaining 20 % for validation. All of the grid cells in Nuremberg are used only for testing. All models are trained for a maximum of 75 epochs with a batch size of eight. The ADAM optimizer is used with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and an initial learning rate of 1 x 10^{-4}, which is decayed by a factor of 0.1 if the training loss did not improve for five subsequent epochs. Random flipping, random rotations, random brightness and gamma adjustments with $beta = [0.8, 1.2]$ and $gamma = [0.8, 1.2]$ are employed as data augmentation techniques to improve the models' robustness and performance [193, 194]. All data augmentation approaches are used with a 50 % probability of being applied to the image, resulting in a differently augmented image each time. The implementation was done using Python 3.8 and the

Setup	Data utilized
Scenario 1	Satellite Imagery
Scenario 2	Satellite Imagery + Residential Masks
Scenario 3	Satellite Imagery + Residential Masks + Building Area
Scenario 4	Satellite Imagery + Residential Masks + Building Area + Building Heights

Table 6.2: Overview of four scenarios and the corresponding data utilized for grid level population estimation.

PyTorch 1.10 framework [196]. All models are trained on a single NVIDIA RTX 3090 GPU with 24 GB RAM.

6.3.2 Evaluation Metrics

To evaluate the models' predictions, the root-mean-square error (RMSE) and the mean absolute error (MAE) are employed. To study the error distribution pattern at two different data quality levels, calculate the percent Relative Estimation Error (REE) [61]. The absolute difference between the actual (y_i) and estimated ($\hat{y_i}$) population counts, divided by the actual population count for each grid cell, is used to compute the REE.

$$REE_i = \frac{|\hat{y_i} - y_i|}{y_i} * 100$$

6.3.3 Mapping population to buildings

To map the estimated grid level population counts to buildings, a post-processing pipeline, as shown in Figure 6.4 is employed. To assign the population only to the residential buildings in a grid cell, first, the residential mask is applied to extract the residential buildings, followed by adding its height information as per the availability to construct the residential area or volume masks. The estimated population of the grid cell is then distributed among its residential buildings in proportion to their residential area, which is then further refined when the residential volume is available such that no building is missed owing to a lack of 3D information. Since building level population estimates are typically not publicly available, the conventional technique of aggregating and analyzing the results at the next possible fine resolution is followed [18, 129, 219, 220, 221].

The reference population data collected and processed at 100 m is used to validate the building level population estimates.

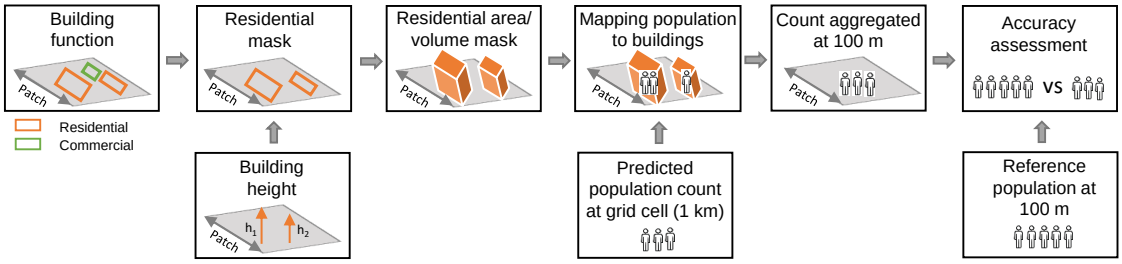


Figure 6.4: The subsequent processing to map population to buildings and its accuracy assessment at 100 m.

6.4 Experimental Results

This section first presents the findings for the population estimation at grid level using the *coarse* and *granular* data and then assesses how well the estimated population is mapped among the buildings.

6.4.1 Grid level Population Estimation

For grid level population estimation, the model is trained to predict a population count at a resolution of 1 km in all four scenarios, adding one level of detail at a time. Table 6.3 presents the RMSE and MAE for the population estimation in each scenario. Since Scenario 1 only uses identical satellite imagery, the outcomes are the same at the *coarse* and *granular* levels. In scenario 2, the residential masks extracted from OSM (*coarse*) and ALKIS (*granular*) are added as input data in addition to the satellite imagery, which tremendously improved the results across all metrics. The MAE of the population prediction improved by 13 % at the *coarse* level and 24 % at the *granular* level. The addition of a building area improved the outcomes slightly in scenario 3. In scenario 4, the addition of height information further improved the MAE by 3 % at the *coarse* level and 12 % at the *granular* level. It has been believed that the completeness of the height data results in multi-fold improvement at the *granular* level compared to the *coarse* level. The scatter plots for all scenarios are shown in Figure 6.5. In these plots, the dotted line represents the identity or line of equality with slope one and the solid line represents the regression line fitted by our model. In general, while the model underpredicts at the (a) *coarse* level, it slightly over-predicts at the (b) *granular* level. As a new input layer is added for each scenario at both levels, not only does the dispersion for population values below 2500 get closer to the identity line, but the estimates for higher values also become closer to the actual population counts. The addition of building areas and building heights improved the values for moderately populated values ranging between 2500 and 5000 at the *granular* level, while at the *coarse* level, it improved the

predictions for the population values greater than 15000. These improvements highlight the learning ability and potential of the proposed method.

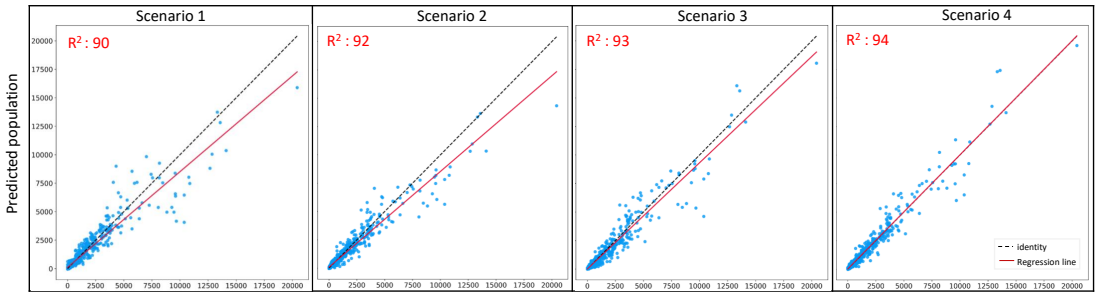
Setup	***Coarse***		***Granular***	
	RMSE	*MAE*	*RMSE*	*MAE*
Scenario 1	653.59	255.90	653.59	255.90
Scenario 2	550.65	220.88	446.17	194.25
Scenario 3	529.97	218.76	423.52	189.22
Scenario 4	511.58	211.89	385.82	166.07

Table 6.3: Evaluation metrics for the population estimation at 1 km resolution for all four scenarios using the *coarse* and *granular* data.

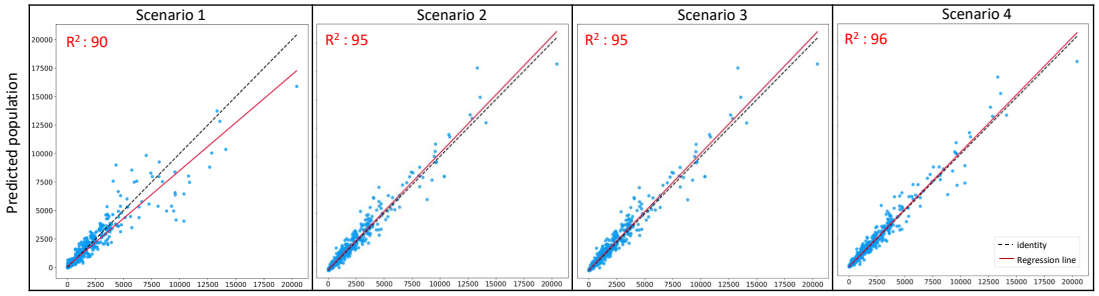
The addition of residential masks in Scenario 2 significantly improved the MAE, approximately 13 % at the *coarse* level and 24 % at the *granular* level. Despite being considered very important information, the improvements with the addition of building areas and heights are not as large as in Scenario 2. Therefore, the distribution of residential building's heights vs. non-residential building's heights and residential building's areas vs. non-residential building's areas are analyzed to ascertain whether their distribution separates residential and non-residential buildings. As shown in Figure 6.6(a), residential building height values at both data quality levels substantially overlap with non-residential building heights. In Figure 6.6(b), though the non-residential buildings appear to have larger areas than the residential buildings, the residential building area values again coincide with the non-residential building area values. As a result, distinguishing between residential and non-residential buildings purely on their heights or areas seems difficult.

6.4.2 Building level Population Estimation

To map the population to buildings, as illustrated in Figure 6.4, the estimated population from the best-trained models is allocated to buildings as a function of building use and building area or volume based on the availability of the data. Figure 6.7 shows an example of the distribution of the predicted population among residential buildings without and with height information. The population estimates for the buildings are aggregated and compared to the reference population at 100 m, shown in Table 6.4. Around four people, on average, are misplaced at the *granular* level, whereas around nine are misplaced at the *coarse* level by their corresponding 100 m grid cell. Figure 6.8 depicts population maps at 100 m for (a) *coarse* level, (b) *granular* level, (c) WorldPop, and (d) reference population data. The population map generated at the *granular* level is the most accurate among others. At *coarse* level, while the population map is comparable to the reference population map in densely populated urban sections of the city, it appears empty as it moves farther from the city, supporting the notion that OSM data



(a) *Coarse* level



(b) *Granular* level

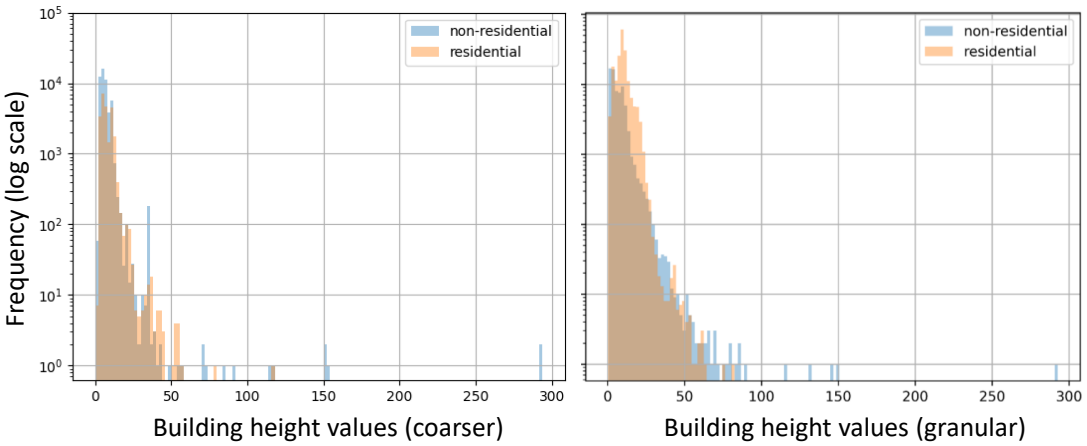
Figure 6.5: Scatter plots of predicted vs. actual population counts for all four scenarios at (a) *coarse* and (b) *granular* levels. The dotted line represents the identity or line of equality and the solid line represents the regression line fitted by our model.

is more complete in densely populated urban regions within the city. The WorldPop population map appears to be anticipating a similar population count for adjacent grid cells and overestimates in the less populated outskirts of the city, resulting in a higher mean absolute error.

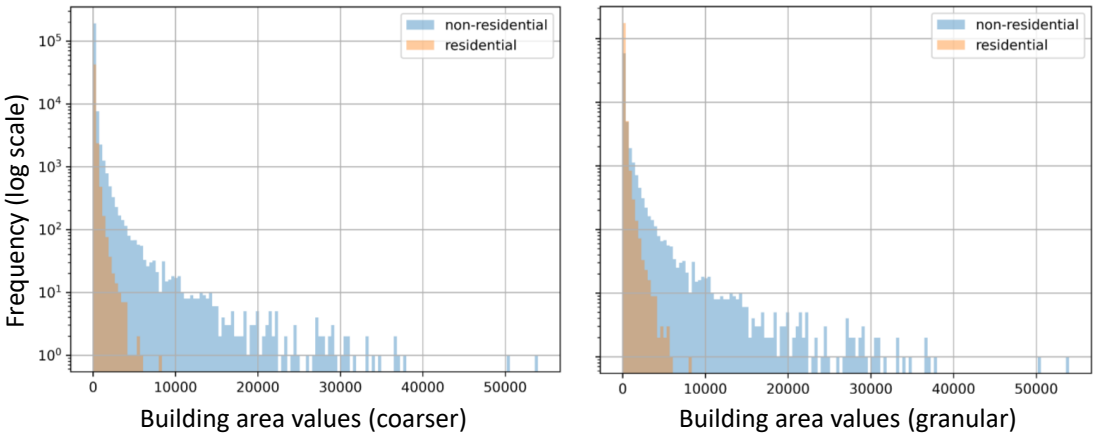
Setup	*RMSE*	*MAE*
Coarse level	34.79	8.99
Granular level	14.12	3.85

Table 6.4: Evaluation metrics for population estimation at the building level for the city of Nuremberg aggregated at 100 m for both *granular* and *coarse* levels.

To further visualize the error distribution pattern, the REE for each populated grid cell is calculated, and its histogram normalized by the total number of grids is plotted in Figure 6.9. It can be seen that around 65 % of the patches at the *coarse* level have estimation errors that are larger than 100 %, indicating that more than half of the predictions are off by over twice the absolute population counts which include over-predictions as well as under-predictions. For the Worldpop, the majority of error lies between 40 % and 80 %. On the other hand, at the *granular* level, the predicted population counts



(a) Residential vs Non-residential building heights



(b) Residential vs Non-residential building area

Figure 6.6: The distribution of (a) Residential vs. Non-residential building heights and (b) Residential vs. Non-residential building areas at both *coarse* and *granular* levels to analyze whether their distribution separates residential and non-residential buildings.

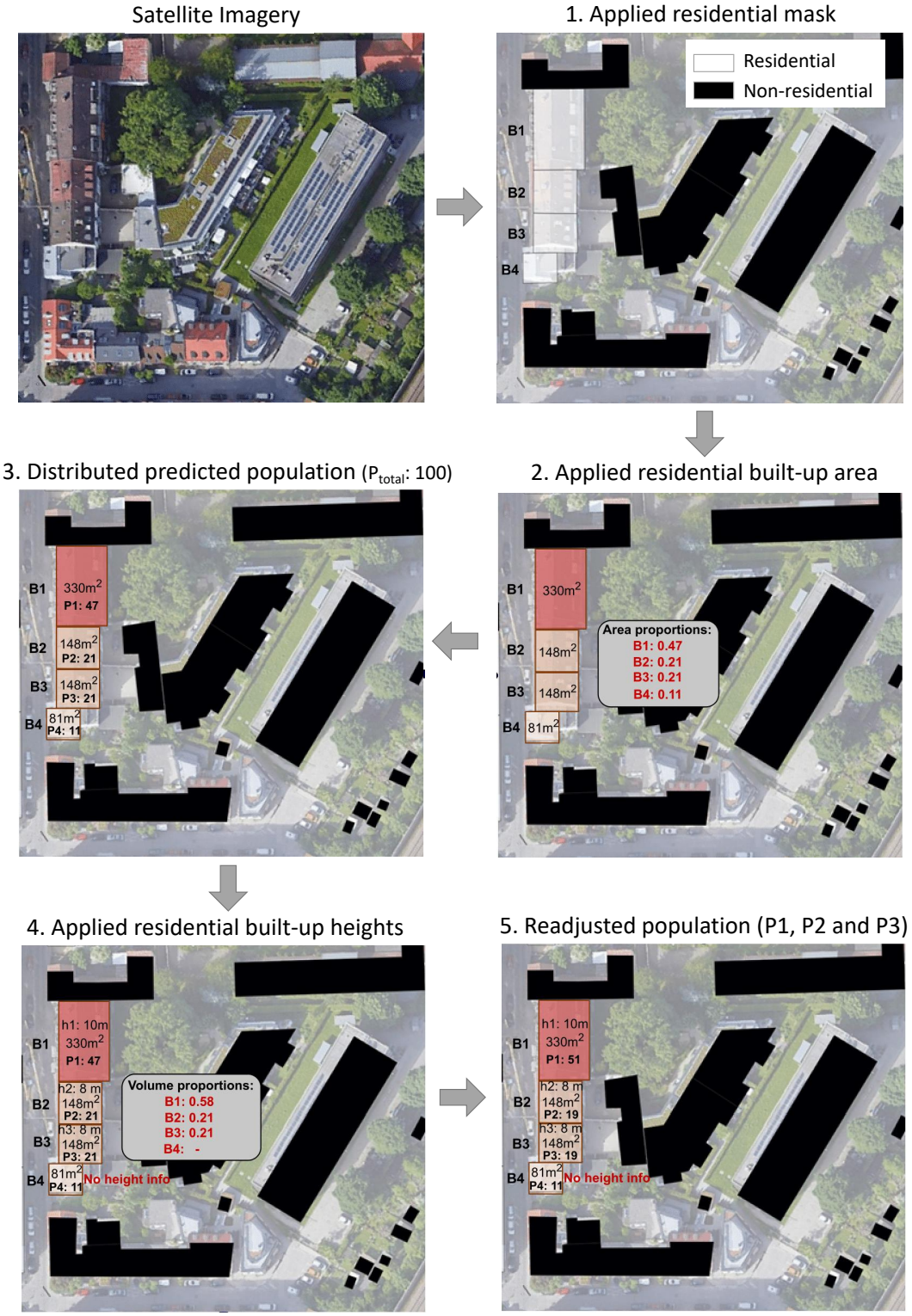


Figure 6.7: An example that illustrates the population mapping to individual residential buildings as a function of building area, subsequently refined for the buildings with height information.

are closer to the actual population counts for the majority of patches with REE smaller than 40 %.

The association between the distribution of REEs at both data levels and the related population counts is studied further. Figure 6.10 depicts the relationship between the REE distribution within a specific population range. At the *coarse* level, the majority of the large errors are concentrated in low-population regions with a population of less than 100. REE continues to be high within populations ranging from 100 to 200; however, REE improves in moderately populated regions. A similar pattern is observed at the *granular* level, with high estimation errors for population areas up to 200, and the REE remains less than 40 % for moderately populated regions and again increases for more populated regions due to the small sample size in this range. Figure 6.11 presents (a) building population maps of Nuremberg at *coarse* and (b) *granular* level, (c) and (d) represent the Nuremberg city center, and (e) and (f) the Nuremberg suburb at *coarse* and *granular* levels, respectively. Similar to population maps at 100 m, the *coarse* level building population maps are substantially poorer in the suburbs due to a lack of OSM data.

6.5 Conclusion

This work offers a deep learning-based end-to-end pipeline for estimating the population at buildings. It has been accomplished by first predicting the population at the grid level (1 km) and then mapping the predicted population count into individual buildings as a function of building use, building area, and building volume whenever this information is available. Since building population counts are frequently unavailable, the building population counts are aggregated to compare them to the reference population data accessible at 100 m. In Nuremberg, with the MAE of 4, the granular level population map at 100 m is highly accurate.

The input data is generated at two levels to highlight the impact of data quality on population estimation: one with large-scale open-access sources (*coarse* level) and another with comprehensive open-access data from official administrative sources (*granular* level). Along with comparing their quality, all experiments are conducted at both levels. It has been discovered that the completeness of building function data is critical in population estimation. The use of a residential mask generated from building function data significantly enhanced the results. Also, at *coarse* level, while the population maps at both 100 m and building level look good in urban parts of the city, they appear empty in Nuremberg city suburbs due to a lack of OSM building function tags. However, with an active OSM community, its quality is continually improving. Nonetheless, population maps at both *coarse* and *granular* scale results in better accuracy than Worldpop.

The encouraging results at the *coarse* level also demonstrate that the method is easily adaptable and could yield acceptable outcomes even when only employing publicly available large-scale data sources. Therefore, this method could reasonably work in a remote setting where extensive regional data from official sources is not available. However, because the data was considerably more complete at the *granular* level, the best results are

achieved at the *granular* level. These findings could stimulate and encourage regional federal offices to develop and offer open-access detailed data that many downstream applications can use.

The population estimates for the buildings are evaluated only at 100 m due to the scarcity of available building level population data. Evaluating the results at the building level itself could help to further understand the method's capabilities and limitations. Exploring and integrating other data sources with high correlations to the population, such as night light data, would be a worthwhile future effort to improve performance. Monitoring the population dynamics at a fine level or building level as a result of urbanization could be another interesting future topic.

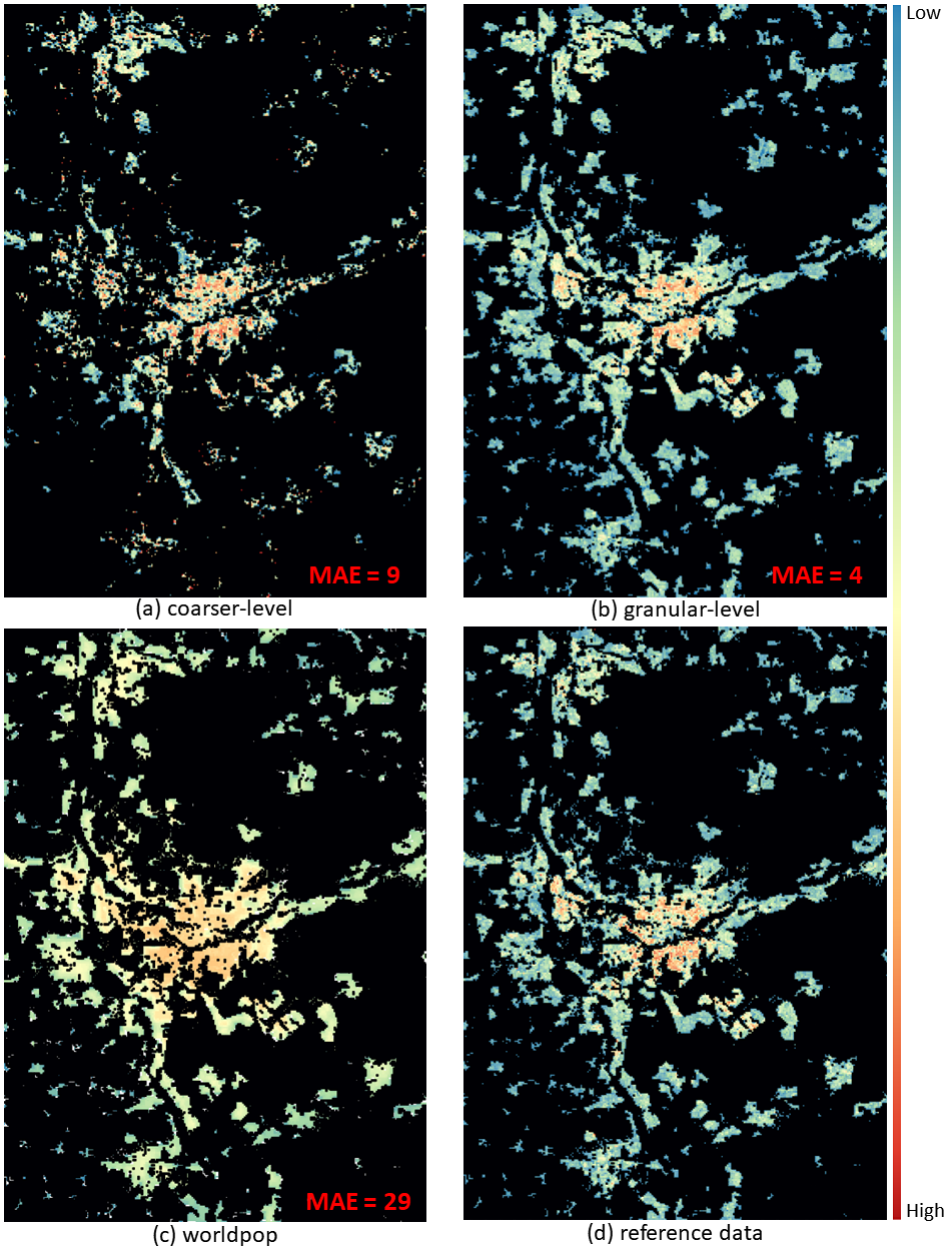


Figure 6.8: Population maps at 100 m for (a) *coarse* level, (b) *granular* level, (c) Worldpop, and (d) reference population data.

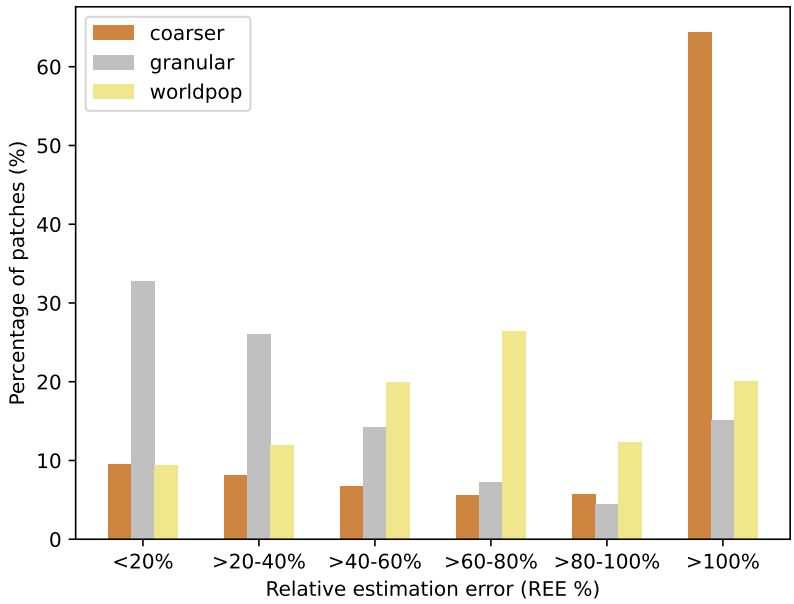
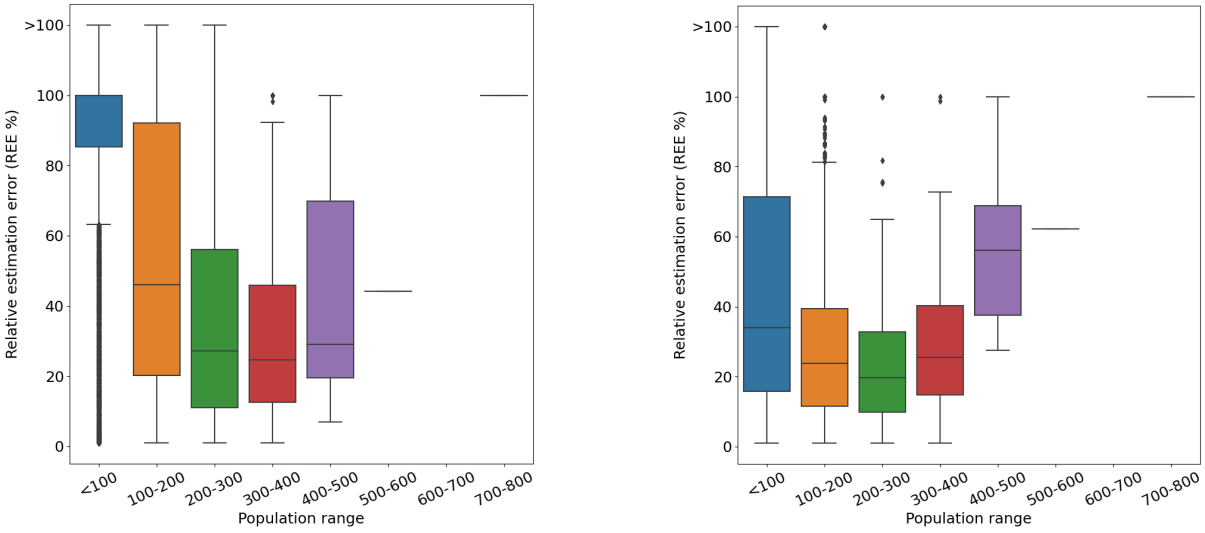
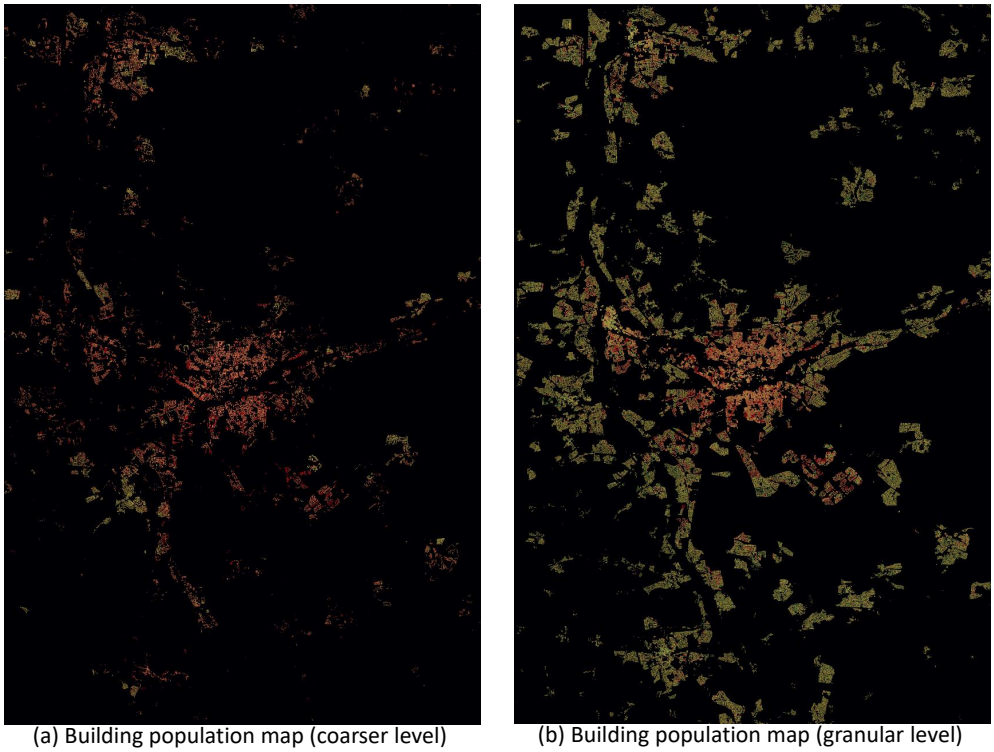


Figure 6.9: Relative estimation error distribution at 100 m for *granular* and *coarse* level. It displays the proportion of patches (y-axis) that fall into a particular relative error range.



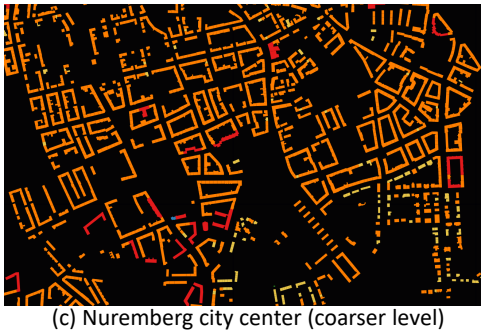
(a) *coarse* **(b)** *granular*

Figure 6.10: Boxplots showing the relationship between the relative estimation error percentage (REE%) and population range at different *coarse* and *granular* levels.

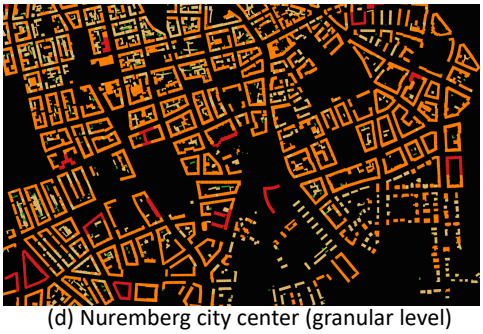


(a) Building population map (coarser level)

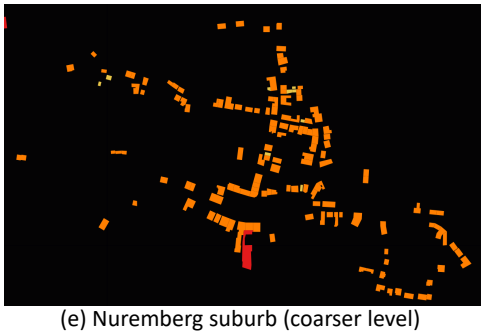
(b) Building population map (granular level)



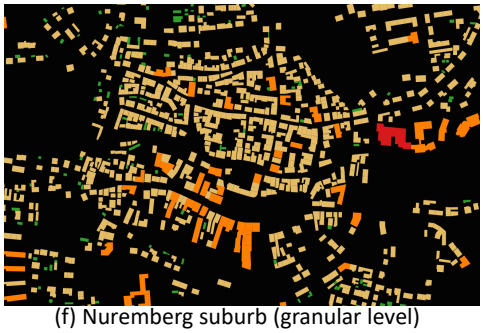
(c) Nuremberg city center (coarser level)



(d) Nuremberg city center (granular level)



(e) Nuremberg suburb (coarser level)



(f) Nuremberg suburb (granular level)

Figure 6.11: Building population maps for Nuremberg (a) and (b), Nuremberg city center (c) and (d), and Nuremberg suburbs shown in (e) and (f) at *coarse* and *granular* levels, respectively.

7 Summary

Remote sensing has propelled the population estimating literature because of its ability to acquire information about vast geographic areas rapidly and efficiently. It enabled the large-scale study as well as the investigation of new data sets and methodologies. This chapter presents a brief review of the various input data sources investigated in this thesis, methodological advancements achieved, the results from different efforts, and future work that could improve and extend the existing approaches.

7.1 Conclusion

Accurate and frequently updated population information is critical for achieving the UN's SDGs. A population census is the standard method of collecting population data, but it comes with its own set of challenges. With the availability of large amounts of EO data sets, population distribution may now be studied on a regular basis, across multiple spatial resolutions, and at a far lesser cost than census data. This thesis investigates openly available EO data that has a strong correlation with people, as well as advancements in deep learning to create interpretable high resolution population maps.

In Chapter 4, this thesis proposed So2Sat-POP, a novel data set that incorporates multi-input geospatial data that were previously unexplored at the cross-country level in the domain of population estimation. The data set covers 98 European cities, thus providing a diverse mix of topography, demography, and architectural styles. The data set includes digital elevation models, local climate zones, land use, nighttime lights, Sentinel-2 imagery and OpenStreetMap data. All input data for each city is collected, processed and cropped using the reference population grids at a resolution of 1 km, yielding a total of 276172 patches. This data set would eliminate the need to acquire and process the data from scratch in order to develop and evaluate the population estimation algorithms. The RF model is trained as a baseline to demonstrate the potential capabilities of the data set, and preliminary findings show that the So2Sat-POP data set has the potential to support the development of sophisticated algorithms for population estimation.

Using this data set, an end-to-end interpretable deep learning framework is developed to predict the population at a resolution of 1 km (Chapter 5). The deep learning architecture is based on an adaptation of the widely used ResNet-50 architecture, which has shown promising results in population mapping. In European cities, results indicate that our population maps are more accurate than GHS-POP. Furthermore, the significance of different input data sources included in the So2Sat-POP data set was explored. The findings indicate that land use data is the most important indicator of population.

To enhance the trustworthiness of the model's inferences, IG, a popular xAI method, is employed. It highlights the most relevant features considered by the model for population estimation, thus sheds light on the workings of the model. It also highlights some specific challenges when relying solely on remote sensing data for population estimation, such as its inability to differentiate between distinct kinds of built-up regions.

To improve the spatial resolution of the existing population data, an extensive study is being carried to map the population to the buildings, with application to two Bavarian cities. A hybrid deep learning approach is developed that first predicts the population at the grid level and subsequently redistributes the predicted population to individual buildings. The method relies on publicly available high resolution satellite imagery and building data such as building functions, area, and volumes, allowing it to be easily applied to a remote situation.

While various deep architectures for population estimation have been studied in the literature, less effort has been made to investigate the impact of data quality in population mapping. To investigate this impact, building data sets at two different quality levels are collected, one from crowd-sourced platforms and the other from regional administrative sources. Their comparative analysis reveals the incompleteness of the crowd-sourced platforms, particularly in the city's suburbs, and the models developed using data from regional governmental sources are substantially more accurate than those trained on data from crowd-sourced platforms. This could encourage the government administrative offices to openly publish the geodata, which could be very useful for urban research and development. It is an initial effort to combine open-access EO data with a deep learning method for building level population estimation, as well as to investigate the influence of variable data quality.

7.2 Outlook

There are numerous possibilities for further investigation. In Chapter 4, a benchmark data set, So2Sat-POP, for population estimation is proposed. Since only European cities are included in this data set, methods developed using this data set may be biased in denser or more sparsely populated areas, such as India and China, or areas with very different architectural or cultural quirks from Europe, such as modern US cities. Expanding the data set to include these missing regions would be good future work to further assist the researchers. Also, another challenge in the population data is noisy labels. In Chapter 5, it was discovered that the reference population count of a region does not always agree with features extracted from satellite imagery and other supplementary input data. As a result, in addition to developing robust algorithms, utilizing uncertainty to detect a certain form of label noise would be highly beneficial, and this is a subject that has not previously been investigated in population estimation studies.

Typically, bias in the reference population counts affects the work in this field. For example, in Europe, the samples from the very high population counts were fewer than the ones from the lower population counts. Chapter 5 seeks to overcome it, for example,

by applying cost-sensitive loss functions and data augmentation. One area of future research could be exploring more advanced approaches to handle imbalanced data and be more resistant to outliers, such as dynamic sampling [222] or distribution smoothing [223]. Additionally, using the "Leave One-Out" principle, Chapter 5 examines the importance of different input data sources for modeling population counts using the suggested deep learning architecture. The same setup is used for each configuration. This principle does not investigate other possible combinations of input data and fusion strategies, which could be expanded in this study.

Although the method is assessed and compared with GHS-POP in a few European cities and three additional US cities, it remains open to compare the method with other state-of-the-art large-scale gridded population products across different regions of the world. This comparison could aid in further understanding the generalizability of our model and explore transfer learning methods in population estimation. However, the availability of reference population data will always be a constraint. As a result, less data intensive methodologies, such as semi-supervised learning, could be investigated. For example, if the population data from micro-census could be collected and processed at a global scale, semi-supervised methods could be utilized for developing population estimation models worldwide.

In Chapter 6, the creation of population maps at the building level helped to visualize the detailed population distribution. However, the estimates were only analyzed at the aggregated resolution of 100 m due to the lack of available building population counts. In this situation, the model's effectiveness might be further assessed by applying population estimates to specific downstream tasks for which reference data is available and directly related to building occupancy, such as building electricity or water consumption. This would help to understand the model's performance at the building level and highlight the significance of fine-scale population estimates in real-world applications.

Another exciting direction could be the illustration of growth of the human population at a global scale. As the population continues to expand rapidly, monitoring the population dynamics and patterns becomes increasingly crucial in maintaining a sustainable living environment. With the availability of multi-temporal high resolution satellite programs such as LandScan, new methods for multi-temporal population mapping could be investigated.

Bibliography

[1] L. Jiang, M. H. Young, and K. Hardee. Population, urbanization and the environment. *World Watch*, (5):34–39, 2008.

[2] T. Sugata and C. Yang. Leaf app: Leaf recognition with deep convolutional neural networks. In *IOP Conference Series: Materials Science and Engineering*, volume 273, page 012004. IOP Publishing, 2017.

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

[4] S. Doda, Y. Wang, M. Kahl, E. J. Hoffmann, K. Ouan, H. Taubenböck, and X. X. Zhu. So2sat pop-a curated benchmark data set for population estimation from space on a continental scale. *Scientific data*, 9(1):715, 2022.

[5] S. Doda, M. Kahl, K. Ouan, I. Obadic, Y. Wang, H. Taubenböck, and X. X. Zhu. Interpretable deep learning for consistent large-scale urban population estimation using earth observation data. *International Journal of Applied Earth Observation and Geoinformation*, 128:103731, 2024.

[6] G. Boeing. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Syst.*, 65:126–139, 2017.

[7] United nations department of economic and social affairs, population division (2022). world population prospects 2022: Summary of results. un desa/pop/2022/tr/no. 3.

[8] Un. department of economic and social affairs. `https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html`. Accessed on: 2023-03-09. URL: `https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html`.

[9] P. C. Bhattacharya. Urbanisation in developing countries. *Economic and Political Weekly*, 37(41):4219–4228, 2002. URL: `http://www.jstor.org/stable/4412720`.

[10] W. Huck. Sustainable development goals. In *Sustainable Development Goals*. Nomos Verlagsgesellschaft mbH & Co. KG, 2022.

[11] Demographic and social statistics. https://unstats.un.org/unsd/demographic-social/census/censusdates/. Accessed on: 2023-03-09. URL: https://unstats.un.org/unsd/demographic-social/census/censusdates/.

[12] N. Wardrop, W. Jochem, T. Bird, H. Chamberlain, D. Clarke, D. Kerr, L. Bengtsson, S. Juran, V. Seaman, and A. Tatem. Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences*, 115(14):3529–3537, 2018.

[13] D. Balk, U. Deichmann, G. Yetman, F. Pozzi, S. Hay, and A. Nelson. Determining global population distribution: Methods, applications and data. In S. I. Hay, A. Graham, and D. J. Rogers, editors, *Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications*, volume 62 of *Advances in Parasitology*, pages 119–156. Academic Press, 2006. URL: https://www.sciencedirect.com/science/article/pii/S0065308X05620040, doi:https://doi.org/10.1016/S0065-308X(05)62004-0.

[14] 2020 census. https://www.gao.gov/products/gao-23-105819, 2023. Accessed on: 2023-03-09. URL: https://www.gao.gov/products/gao-23-105819.

[15] United nations population fund. census. https://www.unfpa.org/census, 2022. Accessed on: 2023-03-09. URL: https://www.unfpa.org/census.

[16] C. Robinson, F. Hohman, and B. Dilkina. A deep learning approach for population estimation from satellite imagery. In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*, pages 47–54, 2017.

[17] K. Klemmer, G. Yeboah, J. P. de Albuquerque, and S. A. Jarvis. Population mapping in informal settlements with high-resolution satellite imagery and equitable ground-truth. *arXiv preprint arXiv:2009.08410*, 2020.

[18] W. Hu, J. H. Patel, Z.-A. Robert, P. Novosad, S. Asher, Z. Tang, M. Burke, D. Lobell, and S. Ermon. Mapping missing population in rural india: A deep learning approach with satellite imagery. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 353–359, 2019.

[19] P. Doupe, E. Bruzelius, J. Faghmous, and S. G. Ruchman. Equitable development through deep learning: The case of sub-national population density estimation. In *Proceedings of the 7th Annual Symposium on Computing for Development*, pages 1–10, 2016.

[20] R. Chen, H. Yan, F. Liu, W. Du, and Y. Yang. Multiple global population datasets: Differences and spatial distribution characteristics. *ISPRS International Journal of Geo-Information*, 9(11):637, 2020.

[21] R. Sliuzas, M. Kuffer, and T. Kemper. Assessing the quality of global human settlement layer products for kampala, uganda. In *2017 Joint Urban Remote Sensing Event (JURSE)*, pages 1–4. IEEE, 2017.

[22] P. J. Gibson and C. H. Power. *Introductory remote sensing: Principles and concepts*. Psychology Press, 2000.

[23] S. Liang, A. H. Strahler, M. J. Barnsley, C. C. Borel, S. A. Gerstl, D. J. Diner, A. J. Prata, and C. L. Walthall. Multiangle remote sensing: Past, present and future. *Remote Sensing Reviews*, 18(2-4):83–102, 2000.

[24] Sentinel-1. `https://sentinel.esa.int/web/sentinel/missions/sentinel-1`. Accessed on: 2023-04-26. URL: `https://sentinel.esa.int/web/sentinel/missions/sentinel-1`.

[25] What is remote sensing? `https://www.earthdata.nasa.gov/learn/backgrounders/remote-sensing`. Accessed on: 2023-04-26. URL: `https://www.earthdata.nasa.gov/learn/backgrounders/remote-sensing`.

[26] Sentinel-2 mission guide. `https://sentinel.esa.int/web/sentinel/missions/sentinel-2`. Accessed on: 2023-04-26. URL: `https://sentinel.esa.int/web/sentinel/missions/sentinel-2`.

[27] Landsat satellite missions. `https://www.usgs.gov/landsat-missions/landsat-satellite-missions`. Accessed on: 2023-04-26. URL: `https://www.usgs.gov/landsat-missions/landsat-satellite-missions`.

[28] Analog and digital images. `https://crisp.nus.edu.sg/~research/tutorial/image.htm`. Accessed on: 2023-04-26. URL: `https://crisp.nus.edu.sg/~research/tutorial/image.htm`.

[29] S. Liang. *Comprehensive remote sensing*. Elsevier, 2017.

[30] Landsat 7. `https://landsat.gsfc.nasa.gov/satellites/landsat-7/`. Accessed on: 2023-07-09. URL: `https://landsat.gsfc.nasa.gov/satellites/landsat-7/`.

[31] W. Fu, J. Ma, P. Chen, and F. Chen. Remote sensing satellites for digital earth. *Manual of digital earth*, pages 55–123, 2020.

[32] R. Avtar, A. A. Komolafe, A. Kouser, D. Singh, A. P. Yunus, J. Dou, P. Kumar, R. D. Gupta, B. A. Johnson, H. V. T. Minh, et al. Assessing sustainable development prospects through remote sensing: A review. *Remote sensing applications: Society and environment*, 20:100402, 2020.

[33] K. E. Joyce, S. E. Belliss, S. V. Samsonov, S. J. McNeill, and P. J. Glassey. A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. *Progress in physical geography*, 33(2):183–207, 2009.

[34] P. Holmgren and T. Thuresson. Satellite remote sensing for forestry planning—a review. *Scandinavian Journal of Forest Research*, 13(1-4):90–110, 1998.

[35] J. Yang, P. Gong, R. Fu, M. Zhang, J. Chen, S. Liang, B. Xu, J. Shi, and R. Dickinson. The role of satellite remote sensing in climate change studies. *Nature climate change*, 3(10):875–883, 2013.

[36] J. Rogan and D. Chen. Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in planning*, 61(4):301–325, 2004.

[37] R. B. Miller and C. Small. Cities from space: potential applications of remote sensing in urban environmental research and policy. *Environmental Science & Policy*, 6(2):129–137, 2003.

[38] H. Taubenböck, N. J. Kraff, and M. Wurm. The morphology of the arrival city-a global categorization based on literature surveys and remotely sensed data. *Applied Geography*, 92:150–167, 2018.

[39] S. Srivastava, J. E. Vargas-Muñoz, D. Swinkels, and D. Tuia. Multilabel building functions classification from ground pictures using convolutional neural networks. In *Proceedings of the 2nd ACM SIGSPATIAL international workshop on AI for geographic knowledge discovery*, pages 43–46, 2018.

[40] 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*, 5(4):8–36, 2017.

[41] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152:166–177, 2019.

[42] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang, et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241:111716, 2020.

[43] 2006 census: Census through the ages. `https://www.abs.gov.au/websitedbs/d3310114.nsf/51c9a3d36edfd0dfca256acb00118404/eadaffffb171cab6ca257161000a78d7!OpenDocument`. Accessed on: 2023-04-26. URL: `https://www.abs.gov.au/websitedbs/d3310114.nsf/51c9a3d36edfd0dfca256acb00118404/eadaffffb171cab6ca257161000a78d7!OpenDocument`.

[44] B. Baffour, T. King, and P. Valente. The modern census: evolution, examples and evaluation. *International Statistical Review*, 81(3):407–425, 2013.

[45] P. J. Hardin, M. W. Jackson, and J. M. Shumway. Intraurban population estimation using remotely sensed imagery. *Geo-spatial technologies in urban environments: policy, practice, and pixels*, pages 47–92, 2007.

[46] M. Langford. An evaluation of small area population estimation techniques using open access ancillary data. *Geographical Analysis*, 45(3):324–344, 2013.

[47] S.-s. Wu, X. Qiu, and L. Wang. Population estimation methods in gis and remote sensing: A review. *GIScience & Remote Sensing*, 42(1):80–96, 2005.

[48] W. Zeng and A. Comber. Using household counts as ancillary information for areal interpolation of population: Comparing formal and informal, online data sources. *Computers, Environment and Urban Systems*, 80:101440, 2020.

[49] C. Thomas-Agnan, A. Vanhems, et al. Accuracy of areal interpolation methods for count data. *Spatial Statistics*, 14:412–438, 2015.

[50] N. S.-N. Lam. Spatial interpolation methods: a review. *The American Cartographer*, 10(2):129–150, 1983.

[51] A. Comber and W. Zeng. Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations. *Geography Compass*, 13(10):e12465, 2019.

[52] M. F. Goodchild and N. S.-N. Lam. Areal interpolation: A variant of the traditional spatial problem. *Geo-processing*, 1(3):297–312, 1980.

[53] S. Hafner, S. Georganos, T. Mugiraneza, and Y. Ban. Mapping urban population growth from sentinel-2 msi and census data using deep learning: A case study in kigali, rwanda. In *2023 Joint Urban Remote Sensing Event (JURSE)*, pages 1–4. IEEE, 2023.

[54] D. R. Thomson, A. E. Gaughan, F. R. Stevens, G. Yetman, P. Elias, and R. Chen. Evaluating the accuracy of gridded population estimates in slums: a case study in nigeria and kenya. *Urban Science*, 5(2):48, 2021.

[55] A. F. Tapp. Areal interpolation and dasymetric mapping methods using local ancillary data sources. *Cartography and Geographic Information Science*, 37(3):215–228, 2010.

[56] X. Liu, P. C. Kyriakidis, and M. F. Goodchild. Population-density estimation using regression and area-to-point residual kriging. *International Journal of geographical information science*, 22(4):431–447, 2008.

[57] R. G. Cromley, D. M. Hanink, and G. C. Bentley. A quantile regression approach to areal interpolation. *Annals of the Association of American Geographers*, 102(4):763–777, 2012.

[58] Y. Qiu, X. Zhao, D. Fan, S. Li, and Y. Zhao. Disaggregating population data for assessing progress of sdgs: methods and applications. *International Journal of Digital Earth*, 15(1):2–29, 2022.

[59] C. L. Eicher and C. A. Brewer. Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science*, 28(2):125–138, 2001.

[60] J. Mennis and T. Hultgren. Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science*, 33(3):179–194, 2006.

[61] D. Palacios-Lopez, F. Bachofer, T. Esch, M. Marconcini, K. MacManus, A. Sorichetta, J. Zeidler, S. Dech, A. J. Tatem, and P. Reinartz. High-resolution gridded population datasets: Exploring the capabilities of the world settlement footprint 2019 imperviousness layer for the african continent. *Remote Sensing*, 13(6):1142, 2021.

[62] J. Mennis. Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55(1):31–42, 2003.

[63] F. R. Stevens, A. E. Gaughan, C. Linard, and A. J. Tatem. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS one*, 10(2):e0107042, 2015.

[64] F. Q. Lauzon. An introduction to deep learning. In *2012 11th international conference on information science, signal processing and their applications (ISSPA)*, pages 1438–1439. IEEE, 2012.

[65] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24–29, 2019.

[66] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6):1236–1246, 2018.

[67] D. W. Otter, J. R. Medina, and J. K. Kalita. A survey of the usages of deep learning for natural language processing. *IEEE transactions on neural networks and learning systems*, 32(2):604–624, 2020.

[68] T. Young, D. Hazarika, S. Poria, and E. Cambria. Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine*, 13(3):55–75, 2018.

[69] Q. Rao and J. Frtunikj. Deep learning for self-driving cars: Chances and challenges. In *Proceedings of the 1st international workshop on software engineering for AI in autonomous systems*, pages 35–38, 2018.

[70] J. Ni, Y. Chen, Y. Chen, J. Zhu, D. Ali, and W. Cao. A survey on theories and applications for self-driving cars based on deep learning methods. *Applied Sciences*, 10(8):2749, 2020.

[71] G. Iannizzotto, L. L. Bello, A. Nucita, and G. M. Grasso. A vision and speech enabled, customizable, virtual assistant for smart environments. In *2018 11th International Conference on Human System Interaction (HSI)*, pages 50–56. IEEE, 2018.

[72] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017.

[73] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[74] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah. Deep learning-based human pose estimation: A survey. *arXiv preprint arXiv:2012.13392*, 2020.

[75] A. Toshev and C. Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660, 2014.

[76] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua. A convolutional neural network cascade for face detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5325–5334, 2015.

[77] S. S. Farfade, M. J. Saberian, and L.-J. Li. Multi-view face detection using deep convolutional neural networks. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pages 643–650, 2015.

[78] C. Szegedy, A. Toshev, and D. Erhan. Deep neural networks for object detection. *Advances in neural information processing systems*, 26, 2013.

[79] A. Diba, V. Sharma, A. Pazandeh, H. Pirsiavash, and L. Van Gool. Weakly supervised cascaded convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 914–922, 2017.

[80] J. A. Benediktsson, M. Pesaresi, and K. 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.

[81] L. Zhang, L. Zhang, and B. Du. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and remote sensing magazine*, 4(2):22–40, 2016.

[82] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud. A comprehensive analysis of deep regression. *IEEE transactions on pattern analysis and machine intelligence*, 42(9):2065–2081, 2019.

[83] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.

[84] P. Baldi and P. J. Sadowski. Understanding dropout. *Advances in neural information processing systems*, 26, 2013.

[85] X. Ying. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, volume 1168, page 022022. IOP Publishing, 2019.

[86] C. Shorten and T. M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.

[87] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015.

[88] P. Kowalek, H. Loch-Olszewska, and J. Szwabiński. Classification of diffusion modes in single-particle tracking data: Feature-based versus deep-learning approach. *Physical Review E*, 100(3):032410, 2019.

[89] N. J. Nagelkerke et al. A note on a general definition of the coefficient of determination. *biometrika*, 78(3):691–692, 1991.

[90] D. Salunke, R. Joshi, P. Peddi, and D. Mane. Deep learning techniques for dental image diagnostics: A survey. In *2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, pages 244–257. IEEE, 2022.

[91] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

[92] C. McClanahan. History and evolution of gpu architecture. *A Survey Paper*, 9:1–7, 2010.

[93] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[94] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[95] Census. `https://www.unfpa.org/census`. Accessed on: 2023-04-26. URL: `https://www.unfpa.org/census`.

[96] R. Carr-Hill. Missing millions and measuring development progress. *World Development*, 46:30–44, 2013.

[97] S. Agarwal. The state of urban health in india; comparing the poorest quartile to the rest of the urban population in selected states and cities. *Environment and Urbanization*, 23(1):13–28, 2011.

[98] A. J. Tatem, N. Campiz, P. W. Gething, R. W. Snow, and C. Linard. The effects of spatial population dataset choice on estimates of population at risk of disease. *Population health metrics*, 9(1):1–14, 2011.

[99] S. Sabry. How poverty is underestimated in greater cairo, egypt. *Environment and Urbanization*, 22(2):523–541, 2010.

[100] L. Warszawski, K. Frieler, V. Huber, F. Piontek, O. Serdeczny, X. Zhang, Q. Tang, M. Pan, Y. Tang, Q. Tang, et al. Center for international earth science information network—ciesin—columbia university.(2016). gridded population of the world, version 4 (gpwv4): Population density. palisades. ny: Nasa socioeconomic data and applications center (sedac). doi: 10. 7927/h4np22dq. *Atlas of Environmental Risks Facing China Under Climate Change*, page 228, 2017.

[101] D. L. Balk, U. Deichmann, G. Yetman, F. Pozzi, S. I. Hay, and A. Nelson. Determining global population distribution: methods, applications and data. *Advances in parasitology*, 62:119–156, 2006.

[102] C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, and E. R. Davis. Mapping city lights with nighttime data from the dmsp operational linescan system. *Photogrammetric Engineering and Remote Sensing*, 63(6):727–734, 1997.

[103] S. Freire, K. MacManus, M. Pesaresi, E. Doxsey-Whitfield, and J. Mills. Development of new open and free multi-temporal global population grids at 250 m resolution. *Population*, 250, 2016.

[104] J. E. Dobson, E. A. Bright, P. R. Coleman, R. C. Durfee, and B. A. Worley. Landscan: a global population database for estimating populations at risk. *Photogrammetric engineering and remote sensing*, 66(7):849–857, 2000.

[105] B. Bhaduri, E. Bright, P. Coleman, and J. Dobson. Landscan. *Geoinformatics*, 5(2):34–37, 2002.

[106] A. J. Tatem. Worldpop, open data for spatial demography. *Scientific data*, 4(1):1–4, 2017.

[107] C. Linard, M. Gilbert, R. W. Snow, A. M. Noor, and A. J. Tatem. Population distribution, settlement patterns and accessibility across africa in 2010. *PloS one*, 7(2):e31743, 2012.

[108] A. E. Gaughan, F. R. Stevens, C. Linard, P. Jia, and A. J. Tatem. High resolution population distribution maps for southeast asia in 2010 and 2015. *PloS one*, 8(2):e55882, 2013.

[109] A. Sorichetta, G. M. Hornby, F. R. Stevens, A. E. Gaughan, C. Linard, and A. J. Tatem. High-resolution gridded population datasets for latin america and the caribbean in 2010, 2015, and 2020. *Scientific data*, 2(1):1–12, 2015.

[110] L. Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[111] T. G. Tiecke, X. Liu, A. Zhang, A. Gros, N. Li, G. Yetman, T. Kilic, S. Murray, B. Blankespoor, E. B. Prydz, et al. Mapping the world population one building at a time. *arXiv preprint arXiv:1712.05839*, 2017.

[112] S. Freire, M. Halkia, and M. Pesaresi. Ghs population grid, derived from eurostat census data (2011) and esm r2016. *European Commission. Joint Research Centre (JRC)*, 2016.

[113] S. Freire and M. Halkia. Ghsl application in europe: Towards new population grids. In *European Forum For Geography And Statistics, Krakow, Poland*, 2014.

[114] M. Pesaresi, D. Ehrlich, S. Ferri, A. Florczyk, S. Freire, M. Halkia, A. Julea, T. Kemper, P. Soille, V. 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.

[115] A. J. Florczyk, S. Ferri, V. Syrris, T. Kemper, M. Halkia, P. Soille, and M. Pesaresi. A new european settlement map from optical remotely sensed data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5):1978–1992, 2015.

[116] J. Kubanek, E.-M. Nolte, H. Taubenböck, F. Wenzel, and M. Kappas. Capacities of remote sensing for population estimation in urban areas. *Earthquake Hazard Impact and Urban Planning*, pages 45–66, 2014.

[117] E.-M. Nolte. *The application of optical satellite imagery and census data for urban population estimation: A case study for Ahmedabad, India*. PhD thesis, Karlsruher Inst. für Technologie, Diss., 2010, 2010.

[118] B. Tellman, J. A. Sullivan, C. Kuhn, A. J. Kettner, C. S. Doyle, G. R. Brakenridge, T. A. Erickson, and D. A. Slayback. Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, 596(7870):80–86, 2021.

[119] G. Li and Q. Weng. Using landsat etm+ imagery to measure population density in indianapolis, indiana, usa. *Photogrammetric Engineering & Remote Sensing*, 71(8):947–958, 2005.

[120] C. Wu and A. T. Murray. Population estimation using landsat enhanced thematic mapper imagery. *Geographical Analysis*, 39(1):26–43, 2007.

[121] J. Iisaka and E. Hegedus. Population estimation from landsat imagery. *Remote Sensing of Environment*, 12(4):259–272, 1982.

[122] C. S. Fibæk, C. Keßler, J. J. Arsanjani, and M. L. Trillo. A deep learning method for creating globally applicable population estimates from sentinel data. *Transactions in GIS*, 26(8):3147–3175, 2022.

[123] G. Boo, E. Darin, D. R. Leasure, C. A. Dooley, H. R. Chamberlain, A. N. Lázár, K. Tschirhart, C. Sinai, N. A. Hoff, T. Fuller, et al. High-resolution population estimation using household survey data and building footprints. *Nature communications*, 13(1):1330, 2022.

[124] I. Neal, S. Seth, G. Watmough, and M. S. Diallo. Census-independent population estimation using representation learning. *Scientific Reports*, 12(1):5185, 2022.

[125] C. Lo. Automated population and dwelling unit estimation from high-resolution satellite images: a gis approach. *Remote Sensing*, 16(1):17–34, 1995.

[126] N. Mudau, W. Mapurisa, T. Tsoeleng, and M. Mashalane. Towards development of a national human settlement layer using high resolution imagery: a contribution to sdg reporting. *South African Journal of Geomatics*, 9(1):1–12, 2020.

[127] F. M. Henderson and Z.-G. Xia. Sar applications in human settlement detection, population estimation and urban land use pattern analysis: a status report. *IEEE transactions on geoscience and remote sensing*, 35(1):79–85, 1997.

[128] T. Esch, H. Taubenböck, A. Roth, W. Heldens, A. Felbier, M. Thiel, M. Schmidt, A. Müller, and S. Dech. Tandem-x mission—new perspectives for the inventory and monitoring of global settlement patterns. *Journal of Applied Remote Sensing*, 6(1):061702–061702, 2012.

[129] H. Chen, B. Wu, B. Yu, Z. Chen, Q. Wu, T. Lian, C. Wang, Q. Li, and J. Wu. A new method for building-level population estimation by integrating lidar, nighttime light, and poi data. *Journal of Remote Sensing*, 2021.

[130] B. Wu, C. Yang, Q. Wu, C. Wang, J. Wu, and B. Yu. A building volume adjusted nighttime light index for characterizing the relationship between urban population and nighttime light intensity. *Computers, Environment and Urban Systems*, 99:101911, 2023.

[131] Z. Chang, Y. Wu, L. Liu, J. Shen, and K. Shi. Exploring the correlations between snpp-viirs nighttime light data and population from a multiplescale perspective. *IEEE Geoscience and Remote Sensing Letters*, 2023.

[132] E. Pajares, R. Muñoz Nieto, L. Meng, and G. Wulfhorst. Population disaggregation on the building level based on outdated census data. *ISPRS International Journal of Geo-Information*, 10(10):662, 2021.

[133] H. Hara, Y. Fujita, and K. Tsuda. Population estimation by random forest analysis using social sensors. *Procedia Computer Science*, 176:1893–1902, 2020.

[134] X. Huang, D. Zhu, F. Zhang, T. Liu, X. Li, and L. Zou. Sensing population distribution from satellite imagery via deep learning: Model selection, neighboring effects, and systematic biases. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:5137–5151, 2021.

[135] N. Metzger, J. E. Vargas-Muñoz, R. C. Daudt, B. Kellenberger, T. T.-T. Whelan, F. Ofli, M. Imran, K. Schindler, and D. Tuia. Fine-grained population mapping from coarse census counts and open geodata. *Scientific Reports*, 12(1):20085, 2022.

[136] S. Georganos, S. Hafner, M. Kuffer, C. Linard, and Y. Ban. A census from heaven: Unraveling the potential of deep learning and earth observation for intra-urban population mapping in data scarce environments. *International Journal of Applied Earth Observation and Geoinformation*, 114:103013, 2022.

[137] R. Welch. Monitoring urban population and energy utilization patterns from satellite data. *Remote sensing of Environment*, 9(1):1–9, 1980.

[138] D. Ehrlich, S. Freire, M. Melchiorri, and T. Kemper. Open and consistent geospatial data on population density, built-up and settlements to analyse human presence, societal impact and sustainability: A review of ghsl applications. *Sustainability*, 13(14):7851, 2021.

[139] J. L. Garb, R. G. Cromley, and R. B. Wait. Estimating populations at risk for disaster preparedness and response. *Journal of Homeland Security and Emergency Management*, 4(1), 2007.

[140] C. Aubrecht, D. Özceylan, K. Steinnocher, and S. Freire. Multi-level geospatial modeling of human exposure patterns and vulnerability indicators. *Natural Hazards*, 68:147–163, 2013.

[141] S. I. Hay, A. M. Noor, A. Nelson, and A. J. Tatem. The accuracy of human population maps for public health application. *Tropical medicine & international health*, 10(10):1073–1086, 2005.

[142] Y. Zhou, M. Ma, K. Shi, and Z. Peng. Estimating and interpreting fine-scale gridded population using random forest regression and multisource data. *ISPRS International Journal of Geo-Information*, 9(6):369, 2020.

[143] K. Balakrishnan. A method for urban population density prediction at 30m resolution. *Cartography and Geographic Information Science*, 47(3):193–213, 2020. URL: `https://doi.org/10.1080/15230406.2019.1687014`, `arXiv:https://doi.org/10.1080/15230406.2019.1687014`, `doi:10.1080/15230406.2019.1687014`.

[144] L. Zhuo, Q. Shi, C. Zhang, Q. Li, and H. Tao. Identifying building functions from the spatiotemporal population density and the interactions of people among buildings. *ISPRS International Journal of Geo-Information*, 8(6):247, 2019.

[145] S. Shang, S. Du, S. Du, and S. Zhu. Estimating building-scale population using multi-source spatial data. *Cities*, 111:103002, 2021.

[146] M. Wang, Y. Wang, B. Li, Z. Cai, and M. Kang. A population spatialization model at the building scale using random forest. *Remote Sensing*, 14(8):1811, 2022.

[147] S. Ural, E. Hussain, and J. Shan. Building population mapping with aerial imagery and gis data. *International Journal of Applied Earth Observation and Geoinformation*, 13(6):841–852, 2011.

[148] A. Smith, P. D. Bates, O. Wing, C. Sampson, N. Quinn, and J. Neal. New estimates of flood exposure in developing countries using high-resolution population data. *Nature communications*, 10(1):1814, 2019.

[149] B. Calka and E. Bielecka. Ghs-pop accuracy assessment: Poland and portugal case study. *Remote Sensing*, 12(7):1105, 2020.

[150] R. I. McDonald, P. Green, D. Balk, B. M. Fekete, C. Revenga, M. Todd, and M. Montgomery. Urban growth, climate change, and freshwater availability. *Proceedings of the National Academy of Sciences*, 108(15):6312–6317, 2011.

[151] A. J. Tatem. Mapping the denominator: spatial demography in the measurement of progress. *International health*, 6(3):153–155, 2014.

[152] G. McGranahan, D. Balk, and B. Anderson. The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and urbanization*, 19(1):17–37, 2007.

[153] X. Zhang, L. Han, H. Wei, X. Tan, W. Zhou, W. Li, and Y. Qian. Linking urbanization and air quality together: A review and a perspective on the future sustainable urban development. *Journal of Cleaner Production*, page 130988, 2022.

[154] S. Szabo. Urbanisation and food insecurity risks: Assessing the role of human development. *Oxford Development Studies*, 44(1):28–48, 2016.

[155] S. Leyk, A. E. Gaughan, S. B. Adamo, A. de Sherbinin, D. Balk, S. Freire, A. Rose, F. R. Stevens, B. Blankespoor, C. Frye, et al. The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth System Science Data*, 11(3):1385–1409, 2019.

[156] U. United Nations. *World Urbanization Prospects: 2014 Revision*. United Nation, 2014.

[157] U. Habitat. *State of the world's cities 2012/2013: Prosperity of cities*. Routledge, 2013.

[158] 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.

[159] T. Esch, W. Heldens, A. Hirner, M. Keil, M. Marconcini, A. Roth, J. Zeidler, S. Dech, and E. 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.

[160] F. J. Gallego. A population density grid of the european union. *Population and Environment*, 31(6):460–473, 2010.

[161] Efgs - essnet project geostat 1b - final report. https://www.efgs.info/wp-content/uploads/geostat/1b/GEOSTAT1B-final-technical-report.pdf. Accessed on: 2022-10-05. URL: https://www.efgs.info/wp-content/uploads/geostat/1b/GEOSTAT1B-final-technical-report.pdf.

[162] Eurostat gisco geostat $1\,km^2$ population grid. https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat, 2011. Accessed on: 2022-10-05. URL: https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat.

[163] Efgs - essnet project geostat 1b - geostat 2011 quality assessment. http://www.efgs.info/wp-content/uploads/geostat/1b/GEOSTAT1B-Appendix17-GEOSTAT-grid-POP-1K-ALL-2011-QA.pdf. Accessed on: 2022-10-05. URL: http://www.efgs.info/wp-content/uploads/geostat/1b/GEOSTAT1B-Appendix17-GEOSTAT-grid-POP-1K-ALL-2011-QA.pdf.

[164] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, et al. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, 120:25–36, 2012.

[165] 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*, 202:18–27, 2017.

[166] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu. Aggregating cloud-free Sentinel-2 images with Google Earth Engine. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume IV-2/W7, pages 145–152, 2019. doi:https://doi.org/10.5194/isprs-annals-IV-2-W7-145-2019.

[167] T. Esch, J. Zeidler, D. Palacios-Lopez, M. Marconcini, A. Roth, M. Mönks, B. Leutner, E. Brzoska, A. Metz-Marconcini, F. Bachofer, et al. Towards a large-scale 3d modeling of the built environment—joint analysis of tandem-x, sentinel-2 and open street map data. *Remote Sensing*, 12(15):2391, 2020.

[168] B. Wessel, M. Huber, C. Wohlfart, U. Marschalk, D. Kosmann, and A. Roth. Accuracy assessment of the global tandem-x digital elevation model with gps data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 139:171–182, 2018.

[169] German aerospace center (dlr) (2018): Tandem-x - digital elevation model (dem) - global, 90m. URL: https://doi.org/10.15489/ju28hc7pui09, doi:10.15489/ju28hc7pui09.

[170] I. D. Stewart and T. R. Oke. Local climate zones: Origins, development, and application to urban heat island studies. In *Proceedings of the Annual Meeting of the American Association of Geographers, Seattle, WA, USA*, pages 12–16, 2011. Accessed on: 2022-06-21.

[171] X. X. Zhu, J. Hu, C. Qiu, Y. Shi, J. Kang, L. Mou, H. Bagheri, M. Haberle, Y. Hua, R. Huang, L. Hughes, H. Li, Y. Sun, G. Zhang, S. Han, M. Schmitt, and Y. Wang. So2sat lcz42: A benchmark data set for the classification of global local climate zones [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, 8(3):76–89, 2020. doi:http://doi.org/10.14459/2018mp1483140.

[172] D. Stathakis and P. Baltas. Seasonal population estimates based on night-time lights. *Computers, Environment and Urban Systems*, 68:133–141, 2018.

[173] X. Chen. Nighttime lights and population migration: Revisiting classic demographic perspectives with an analysis of recent european data. *Remote Sensing*, 12(1):169, 2020.

[174] A. Bruederle and R. Hodler. Nighttime lights as a proxy for human development at the local level. *PloS one*, 13(9):e0202231, 2018.

[175] K. Shi, B. Yu, Y. Huang, Y. Hu, B. Yin, Z. Chen, L. Chen, and J. Wu. Evaluating the ability of npp-viirs nighttime light data to estimate the gross domestic product and the electric power consumption of china at multiple scales: A comparison with dmsp-ols data. *Remote Sensing*, 6(2):1705–1724, 2014.

[176] C. D. Elvidge, M. Zhizhin, T. Ghosh, F.-C. Hsu, and J. Taneja. Annual time series of global viirs nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, 13(5):922, 2021. doi:10.3390/rs13050922.

[177] Map features documentation wiki. https://wiki.openstreetmap.org/wiki/Map_features. Accessed on: 2021-08-21. URL: https://wiki.openstreetmap.org/wiki/Map_features.

[178] X. Li, Y. Wang, J. Li, and B. Lei. Physical and socioeconomic driving forces of land-use and land-cover changes: A case study of wuhan city, china. *Discrete Dynamics in Nature and Society*, 2016, 2016.

[179] G. Boeing. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126–139, 2017. URL: https://www.sciencedirect.com/science/article/pii/S0198971516303970, doi:https://doi.org/10.1016/j.compenvurbsys.2017.05.004.

[180] T. Ye, N. Zhao, X. Yang, Z. Ouyang, X. Liu, Q. Chen, K. Hu, W. Yue, J. Qi, Z. Li, et al. Improved population mapping for china using remotely sensed and points-of-interest data within a random forests model. *Science of the total environment*, 658:936–946, 2019.

[181] S. Georganos, T. Grippa, A. Niang Gadiaga, C. Linard, M. Lennert, S. Vanhuysse, N. Mboga, E. Wolff, and S. Kalogirou. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, 36(2):121–136, 2021.

[182] M. Sapena, M. Kühnl, M. Wurm, J. E. Patino, J. C. Duque, and H. Taubenböck. Empiric recommendations for population disaggregation under different data scenarios. *Plos one*, 17(9):e0274504, 2022.

[183] A. Adadi and M. Berrada. Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE Access*, 6:52138–52160, 2018.

[184] D. Tuia, R. Roscher, J. D. Wegner, N. Jacobs, X. Zhu, and G. Camps-Valls. Toward a collective agenda on ai for earth science data analysis. *IEEE Geosci. Remote Sens. Mag.*, 9(2):88–104, 2021. doi:10.1109/MGRS.2020.3043504.

[185] F.-F. Li, R. Krishna, and D. Xu. Cs231n: Convolutional neural networks for visual recognition - lecture 7: Training neural networks, part 1. *Stanford University*, 2021.

[186] I. D. Stewart and T. R. Oke. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.*, 93(12):1879–1900, 2012.

[187] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 12:2825–2830, 2011.

[188] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko. Revisiting deep learning models for tabular data. *Adv. Neural Inf. Process. Syst.*, 34, 2021.

[189] S.-C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, and M. P. Lungren. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ digital medicine*, 3(1):136, 2020.

[190] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[191] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[192] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[193] I. Sirazitdinov, M. Kholiavchenko, R. Kuleev, and B. Ibragimov. Data augmentation for chest pathologies classification. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 1216–1219. IEEE, 2019.

[194] X. Sun, H. Fang, Y. Yang, D. Zhu, L. Wang, J. Liu, and Y. Xu. Robust retinal vessel segmentation from a data augmentation perspective. In *International Workshop on Ophthalmic Medical Image Analysis*, pages 189–198. Springer, 2021.

[195] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. In *Proc. IEEE. Int. Conf. Comput. Vis.*, pages 2980–2988, 2017.

[196] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[197] M. Sundararajan, A. Taly, and Q. Yan. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR, 2017.

[198] E. M. Weber, V. Y. Seaman, R. N. Stewart, T. J. Bird, A. J. Tatem, J. J. McKee, B. L. Bhaduri, J. J. Moehl, and A. E. Reith. Census-independent population mapping in northern nigeria. *Remote sensing of environment*, 204:786–798, 2018.

[199] A. Tatem and C. Linard. Population mapping of poor countries. *Nature*, 474(7349):36–36, 2011.

[200] Statistische Ämter des bundes und der länder. gemeindeverzeichnis — statistikportal.de. `http://www.statistikportal.de/de/gemeindeverzeichnis`. Accessed on: 2023-04-21. URL: `http://www.statistikportal.de/de/gemeindeverzeichnis`.

[201] Nuremberg and munich lead german economic growth in 2014. `https://www.brookings.edu/blog/the-avenue/2014/11/17/nuremberg-and-munich-lead-german-economic-growth-in-2014/`. Accessed on: 2023-04-21. URL: `https://www.brookings.edu/blog/the-avenue/2014/11/17/nuremberg-and-munich-lead-german-economic-growth-in-2014/`.

[202] Explanatory notes on "demographie im 100 meter-gitter". `https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html`. Accessed on: 2023-04-21. URL: `https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html`.

[203] J. Harvey. Estimating census district populations from satellite imagery: Some approaches and limitations. *International journal of remote sensing*, 23(10):2071–2095, 2002.

[204] X. Huang, C. Wang, Z. Li, and H. Ning. A 100 m population grid in the conus by disaggregating census data with open-source microsoft building footprints. *Big Earth Data*, 5(1):112–133, 2021.

[205] F. Biljecki, Y. S. Chow, and K. Lee. Quality of crowdsourced geospatial building information: A global assessment of openstreetmap attributes. *Building and Environment*, 237:110295, 2023.

[206] D. Zielstra and A. Zipf. Quantitative studies on the data quality of openstreetmap in germany. In *Proceedings of GIScience*, number 3, 2010.

[207] Openstreetmap data extracts. geofabrik. Accessed on: 2023-04-21. URL: `http://download.geofabrik.de`.

[208] Amtliches liegenschaftskatasterinformationssystem. `https://www.adv-online.de/AdV-Produkte/Liegenschaftskataster/ALKIS/`. Accessed on: 2023-04-21. URL: `https://www.adv-online.de/AdV-Produkte/Liegenschaftskataster/ALKIS/`.

[209] Alkis®-tatsächliche nutzung (tn). `https://geodaten.bayern.de/opengeodata/OpenDataDetail.html?pn=tatsaechlichenutzung`. Accessed on: 2023-04-21. URL: `https://geodaten.bayern.de/opengeodata/OpenDataDetail.html?pn=tatsaechlichenutzung`.

[210] X. Li, Y. Zhou, P. Gong, K. C. Seto, and N. Clinton. Developing a method to estimate building height from sentinel-1 data. *Remote Sensing of Environment*, 240:111705, 2020.

[211] Q. Li, L. Mou, Y. Hua, Y. Shi, S. Chen, Y. Sun, and X. X. Zhu. 3dcentripetalnet: Building height retrieval from monocular remote sensing imagery. *International Journal of Applied Earth Observation and Geoinformation*, 120:103311, 2023.

[212] H. G. Kamath, M. Singh, L. A. Magruder, Z.-L. Yang, and D. Niyogi. Globus: Global building heights for urban studies. *arXiv preprint arXiv:2205.12224*, 2022.

[213] C. Yang and S. Zhao. A building height dataset across china in 2017 estimated by the spatially-informed approach. *Scientific Data*, 9(1):76, 2022.

[214] N. Milojevic-Dupont, F. Wagner, F. Nachtigall, J. Hu, G. B. Brüser, M. Zumwald, F. Biljecki, N. Heeren, L. H. Kaack, P.-P. Pichler, et al. Eubucco v0. 1: European building stock characteristics in a common and open database for 200+ million individual buildings. *Scientific Data*, 10(1):147, 2023.

[215] 3d-gebäudemodelle (lod2). `https://geodaten.bayern.de/opengeodata/OpenDataDetail.html?pn=lod2`. Accessed on: 2023-04-21. URL: `https://geodaten.bayern.de/opengeodata/OpenDataDetail.html?pn=lod2`.

[216] G. Gröger and L. Plümer. Citygml–interoperable semantic 3d city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71:12–33, 2012.

[217] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition (2015). *arXiv preprint arXiv:1512.03385*, 730, 2020.

[218] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon. Transfer learning from deep features for remote sensing and poverty mapping. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

[219] Y. Yao, X. Liu, X. Li, J. Zhang, Z. Liang, K. Mai, and Y. Zhang. Mapping fine-scale population distributions at the building level by integrating multisource

geospatial big data. *International Journal of Geographical Information Science*, 31(6):1220–1244, 2017.

[220] D. Palacios-Lopez, T. Esch, K. MacManus, M. Marconcini, A. Sorichetta, G. Yetman, J. Zeidler, S. Dech, A. J. Tatem, and P. Reinartz. Towards an improved large-scale gridded population dataset: A pan-european study on the integration of 3d settlement data into population modelling. *Remote Sensing*, 14(2):325, 2022.

[221] L. Cheng, L. Wang, R. Feng, and J. Yan. Remote sensing and social sensing data fusion for fine-resolution population mapping with a multimodel neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:5973–5987, 2021.

[222] M. Lin, K. Tang, and X. Yao. Dynamic sampling approach to training neural networks for multiclass imbalance classification. *IEEE Transactions on Neural Networks and Learning Systems*, 24(4):647–660, 2013.

[223] Y. Yang, K. Zha, Y. Chen, H. Wang, and D. Katabi. Delving into deep imbalanced regression. In *International Conference on Machine Learning*, pages 11842–11851. PMLR, 2021.

A Appendix

Table A.1: Mapping of Bezeichnung values from the Bayernatlas to the reduced classification scheme used in this thesis. The value represents the land use value in German directly derived from the ALKIS Tatsächliche Nutzung data, translation represents its translation in English, and class represents its corresponding mapped class in the reduced classification scheme.

Value	Translation	Class
Abfallbehandlungsanlage	Waste treatment plant	2
Ackerland	Farmland	2
Anlegestelle	Jetty	1
Ausstellung, Messe	Exhibition, fair	1
Autokino, Freilichtkino	Drive-in cinema, open-air cinema	1
Baumschule	Tree nursery	2
Botanischer Garten	Botanical garden	4
Campingplatz	Camping ground	1
Deponie (oberirdisch)	Landfill (above ground)	2
Deponie (untertägig)	Landfill (below ground)	2
Entsorgung	Waste disposal	2
Erholungsfläche	Recreation area	4
Fähranlage	Ferry facility	1
Festplatz	Fairground	1
Fischereiwirtschaftsfläche	Fishery area	2
Forstwirtschaftliche Betriebsfläche	Forestry operational area	2
Förderanlage	Conveyor plant	2
Freilichtmuseum	Open-air museum	1
Freilichttheater	Open-air theater	1
Freizeitanlage	Recreational area	1

Continued on next page

Table A.1 – continued from previous page

Value	Translation	Class
Freizeitpark	Amusement park	1
Funk- und Fernmeldeanlage	Radio and telecommunication system	2
Fußgängerzone	Pedestrian zone	5
Garten	Garden	4
Gaswerk	Gas plant	2
Gärtnerei	Gardening shop	1
Gebäude- und Freifläche, Mischnutzung mit Wohnen	Building and open space, mixed use with housing	3
Gebäude- und Freifläche Land- und Forstwirtschaft	Building and open space agriculture and forestry	2
Golfplatz	Golf course	1
Grünanlage	Green area	4
Grünland	Grassland	2
Hafenanlage (Landfläche)	Port facility (land area)	1
Handel und Dienstleistung	Trade and services	1
Heizwerk	Heating plant	2
Hopfen	Hops	2
Hubschrauberflugplatz	Helicopter airfield	1
Industrie und Gewerbe	Industry and commerce	5
Internationaler Flughafen	International airport	1
Kanal	Water duct	4
Kläranlage, Klärwerk	Sewage plant, sewage treatment plant	2
Kleingarten	Allotment garden	4
Kraftwerk	Power plant	2
Kultur	Culture	4
Lagerplatz	Storage yard	2
Landeplatz, Sonderlandeplatz	Landing field, special landing field	1
Landwirtschaftliche Betriebsfläche	Agricultural farmland	2
Marktplatz	Market place	1
Medien und Kommunikation	Media and Communication	4
Modellflugplatz	Airfield for model planes	1

Continued on next page

Table A.1 – continued from previous page

Value	Translation	Class
nan	-	5
Null	-	5
Obstplantage	Orchard	2
Öffentliche Zwecke	Public purposes	4
Park	Park	4
Parkplatz	Parking lot	4
Raffinerie	Refinery	2
Rastplatz	Rest area (next to a motorway)	4
Raststätte	Roadhouse	1
Regionalflughafen	Regional airport	1
Safaripark, Wildpark	Safari park, game park	1
Schleuse (Landfläche)	Sluice (land area)	1
Schwimmbad, Freibad	Swimming pool, outdoor pool	1
Segelfluggelände	Glider site	1
Speicherbecken	Reservoir	4
Spielplatz, Bolzplatz	Playground, football field	4
Sportanlage	Sports facility	1
Stausee	Reservoir lake	4
Umspannstation	Electrical substation	2
Verkehrslandeplatz	Airfield	1
Versorgungsanlage	Supply facility	2
Wasserwerk	Waterworks	2
Weihnachtsbaumkultur	Christmas tree nursery	2
Weingarten	Winegarden	2
Werft	Shipyard	2
Wochenendplatz	Weekendplace	4
Wochenend- und Ferienhausfläche	Weekend- and cottage area	1
Zoo	Zoo	1

Table A.2: Mapping of Nutzungsart values from the Bayernatlas to the reduced classification scheme used in this thesis. value represents the land use value in German directly derived from the ALKIS Tatsächliche Nutzung data, translation represents its translation in English, and class represents its corresponding mapped class in reduced classification scheme.

Value	Translation	Class
Bahnverkehr	Railroad traffic	4
Bergbaubetrieb	Mining area	2
Fläche besonderer funktionaler Prägung	Area of special functional character	4
Fläche gemischter Nutzung	Area of mixed use	5
Fließgewässer	Running water	4
Flugverkehr	Air traffic	1
Friedhof	Cemetery	4
Gehölz	Woodland	4
Hafenbecken	Dock	4
Haide	Heath	4
Heide	Heath	4
Industrie- und Gewerbefläche	Industrial and commercial area	5
Landwirtschaft	Agriculture	2
Moor	Moor	4
Platz	Square	5
Schiffsverkehr	Shipping traffic	1
Sport-, Freizeit- und Erholungsfläche	Sports, recreational and rest area	5
Stehendes Gewässer	Standing water	4
Straßenverkehr	Road traffic	5
Sumpf	Swamp	4
Tagebau, Grube, Steinbruch	Open pit, mine, quarry	2
Unkultivierte Fläche	Uncultivated area	4
Unland/Vegetationslose Fläche	Vegetation free area	4
Wald	Forest	4
Weg	Road	5
Wohnbaufläche	Residential area	3

Table A.3: Mapping of OSM tag values to the reduced classification scheme used in this work. value represents the land use tag in OSM and class represents its corresponding mapped class in reduced classification scheme.

Value	Class	Value	Class
allotment_house	3	conservatory	4
allotments	4	construction	4
apartments	3	container	4
arts_centre	4	court	4
attachment	4	cowshed	4
barn	4	demolished	4
bicycle_parking	4	detached	3
boathouse	1	disused	4
brewery	2	dormitory	3
bridge	4	elevator	4
brothel	1	exhibition_hall	4
building_passage	4	farm	4
bungalow	3	farm_auxiliary	4
bunker	4	farmland	2
cabin	4	farmyard	3
carport	4	fga	4
castle	4	film_set	4
cemetery	4	fire_station	4
chapel	4	flo	4
chimney	4	forest	4
church	4	G	4
cinema	1	GA	4
civic	4	garage	4
collapsed	4	garages	4
college	4	garbage_shed	4
columbarium	4	gazebo	4
commercial	1	ger	3

Continued on next page

Table A.3 – continued from previous page

Value	Class	Value	Class
grandstand	4	parking	4
grass	5	pavilion	4
greenhouse	4	power_station	4
gymnasium	4	prefabricated	4
hangar	4	presbytery	4
heath	4	proposed	4
horseshed	4	public	4
hospital	1	public_transport	4
hotel	1	quarry	2
house	3	recreation_ground	4
hut	3	rectory	4
industrial	2	religious	4
kindergarten	4	residential	3
kiosk	1	restaurant	1
loading_ramp	4	retail	1
manufacture	4	riding_hall	4
meadow	4	roof	3
military	4	roofed_ramp	4
monastery	4	ruins	4
mosque	4	school	4
museum	1	scrub	4
nan	5	semidetached_house	3
nature_reserve	4	service	4
no	4	shed	4
Null	5	sheepfold	4
nursing_home	3	shelter	4
office	1	ship	4
orchard	2	shrine	4
parish_hall	4	silo	4
park	4	skyscraper	1

Continued on next page

Table A.3 – continued from previous page

Value	**Class**	**Value**	**Class**
social_facility	4	terrace	3
sports_centre	1	Tiefgarage	4
sports_hall	4	toilets	4
stable	4	tower	4
stadium	1	train_station	4
staircase	4	transformer_tower	4
stairs	4	transportation	4
static_caravan	4	triumphal_arch	4
storage_tank	4	university	4
street_cabinet	4	warehouse	1
substation	4	waste	4
supermarket	1	water	4
synagogue	4	water_tower	4
temple	4	workshop	4
tent	4	yes;industrial	2
terminal	1	yes;apartments	3