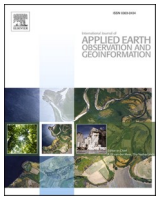
International Journal of Applied Earth Observations and Geoinformation 105 (2021) 102568

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



UCL: Unsupervised Curriculum Learning for water body classification from remote sensing imagery

Nosheen Abid [a,b,c,*], Muhammad Shahzad [b,c,d], Muhammad Imran Malik [b,c], Ulrich Schwanecke [e], Adrian Ulges [e], György Kovács [a], Faisal Shafait [b,c]

[a] Machine Learning Group, Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Sweden
[b] School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan
[c] Deep Learning Lab, National Center of Artificial Intelligence, National University of Sciences and Technology, Pakistan
[d] Data Science in Earth Observation, Department of Aerospace and Geodesy, Technical University of Munich (TUM), Munich, Germany
[e] RheinMain University of Applied Sciences, Germany

ARTICLE INFO

Keywords:
Sentinel-2
Aircraft Imagery
Remote Sensing
Water classification
Deep Learning
Unsupervised Curriculum Learning
Multi-scale Classification

ABSTRACT

This paper presents a Convolutional Neural Networks (CNN) based Unsupervised Curriculum Learning approach for the recognition of water bodies to overcome the stated challenges for remote sensing based RGB imagery. The unsupervised nature of the presented algorithm eliminates the need for labelled training data. The problem is cast as a two class clustering problem (water and non-water), while clustering is done on deep features obtained by a pre-trained CNN. After initial clusters have been identified, representative samples from each cluster are chosen by the unsupervised curriculum learning algorithm for fine-tuning the feature extractor. The stated process is repeated iteratively until convergence. Three datasets have been used to evaluate the approach and show its effectiveness on varying scales: (i) SAT-6 dataset comprising high resolution aircraft images, (ii) Sentinel-2 of EuroSAT, comprising remote sensing images with low resolution, and (iii) PakSAT, a new dataset we created for this study. PakSAT is the first Pakistani Sentinel-2 dataset designed to classify water bodies of Pakistan. Extensive experiments on these datasets demonstrate the progressive learning behaviour of UCL and reported promising results of water classification on all three datasets. The obtained accuracies outperform the supervised methods in domain adaptation, demonstrating the effectiveness of the proposed algorithm.

1. Introduction

Wetlands are the core source of life on Earth. Each wetland can be considered as an individual ecosystem considering the geographical location, climatic variability, exposure to the type of landscape, and many other factors (Bhowmik, 2020). They are having a significant ecological and economic importance (Mishra, 2020) and need to be preserved. An essential prerequisite for this is to estimate the presence of wetlands at different locations. United Nations (UN) water report 2019 states that the consumption of water from various sources (i.e., reservoirs, lakes, and rivers) increases following the economic expansion, urbanization, and growth of the population, causing an imbalance between supply and demand. In parallel, the global water cycle is intensifying due to climate change, where wet regions are generally decreasing, and the drier regions are becoming even drier. Other global changes such as deforestation, intensification of agriculture, and urbanization pose further challenges. The lack of adequate water management systems in many countries leads to a waste of water with an economic value of billions of dollars every year (UNE, 2019). This deficiency raises the need for efficient wetland monitoring and management systems. Typically, the water body monitoring systems rely on rich remote sensing spatial and temporal data to detect water bodies (Mahdavi et al., 2018).

Waterbody detection from the water surface is a fundamental module in many remote sensing studies such as land cover (Zhang et al., 2018) and land use (Buma et al., 2018), estimating water scarcity (Guo et al., 2017), controlling flood hazard (Kyriou and Nikolakopoulos, 2015), predicting aquatic widespread disease, and measuring water quality (Sadeghi et al., 2017). In this paper, we cast water detection as a two-class problem where input images are divided into smaller sized

* Corresponding author at: Machine Learning Group, Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Sweden. School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan
E-mail addresses: nosheen.abid@ltu.se (N. Abid), muhammad.shehzad@seecs.edu.pk (M. Shahzad).

https://doi.org/10.1016/j.jag.2021.102568
Received 15 May 2021; Received in revised form 27 August 2021; Accepted 28 September 2021
Available online 29 October 2021

image patches, and each image patch is classified as water or a non-water patch. To solve this classification problem, we present an efficient and robust deep Unsupervised Curriculum Learning (UCL) based algorithm. Specifically, in this work, we propose unsupervised curriculum learning such that representative sample selection for water and non-water category is done from unlabelled data. Extraction of representative samples for each cluster to fine-tune the deep network is thus a key step in our approach. For this, we take inspiration from curriculum learning (Bengio et al., 2009) in which the learning algorithm is fed by training samples in the order of their difficulty. Difficulty depends upon the complexity of the samples' features. Easy samples are fed in the starting iterations and the complexity is gradually increased in the subsequent iterations. It has been shown that feeding the learner in this manner improves the learning speed as deep learning algorithm has to gradually learn complex examples. The major challenge in curriculum learning is how to define the difficulty level of a sample. Different heuristics have been used successfully in different domains to establish the complexity of a data sample. Here, we use proximity from cluster centroids as the criteria for the selection of representative "easy" samples. In the beginning, the deep learning model used for feature extraction is pre-trained on the ImageNet dataset, which represents a different domain. Hence, the resulting model may not extract good features of the water bodies and non-water bodies from remote sensing imagery, leading to loose clusters. When the clusters are loosely packed, only a few easy samples are selected. The deep learning model is fine-tuned on these samples using their pseudo labels. With every iteration, the fine-tuned deep learning model extracts better features resulting in improved clusters. Progressively, we increase the complexity of the chosen samples via curriculum learning by allowing more samples to be selected for fine-tuning deep model on water bodies using pseudo labels. This progressive-leaning behaviour can be called Unsupervised Curriculum Learning. This idea has been exploited in natural images (Fan et al., 2018; Caron et al., 2018) and the proposed framework adopted this idea for the classification of remote sensing imagery.

In this context, the contributions proposed in this paper are twofold. (1) An easy-to-implement unsupervised progressive deep learning model for water body classification from RGB remote sensing imagery. The integration of clustering with curriculum learning leads to unsupervised learning of the deep model by using pseudo labels. (2) The evaluation of the proposed strategy is performed using three datasets, out of which two are benchmark datasets, space-borne EuroSAT (Helber et al., 2019) and air-bone Sat-6 (Basu et al., 2015), and our newly introduced custom space-borne dataset, PakSAT, which shall be made open for the public. The statistics of the PakSAT dataset are detailed in Section 4.3.

The rest of the paper is structured as follows: In Section 2 we present a brief review of the techniques related to our approach and waterbodies detection using remote sensing data. In Section 3 we describe the designed unsupervised methodology. Then, in Section 4 we discuss the datasets used "as-is" as well as the dataset generated for this paper. Following this, in Section 5 we present the experiments conducted and give a detailed analysis of the obtained results. Lastly, in Section 6 we provide our conclusions and give an outlook to future work.

2. Related work

Numerous techniques have been introduced to detect water utilizing radar and optical imagery (Guo et al., 2017). Radar data have the advantage of capturing the information in almost every weather and day-night condition. However, the prominent features of vegetation (Huang et al., 2018), waves (Töyrä and Pietroniro, 2005; Marti-Cardona et al., 2013), sand (Martinis et al., 2018) and radar shadows produced by landscape features (Giustarini et al., 2013) deterrent the effective separation of water from the land surface. Therefore, the extraction of water bodies from remotely sensed data is more effective from optical imagery than from radar data (Schumann et al., 2009).

2.1. Sensing Modalities

2.1.1. Satellite Imagery (Space-borne)

The satellite imagery (spaceborne) of the Landsat Program (launched in 1972) gradually became the most popular source of optical imagery in remote sensing. The optical sensor of the Landsat satellites has a typical spatial resolution of 30 m which was better than other freely available coarse sensors like MERIS with spatial resolution 300 m, NOAA/AVHRR with spatial resolution 1100 m, and MODIS with spatial resolution 250–1000 m. Therefore, most of the global digital maps for water are designed using Landsat imagery, some of the datasets are Global Inland Water Body (GIW) (Veloso et al., 2017), Global Surface Water (Markert et al., 2018), and Global Water Bodies Database (GLOWABO) (Pahlevan et al., 2017). Similarly, most of the water detection based applications are designed using Landsat Imagery (Manakos et al., 2020; Davranche et al., 2013; Puliti et al., 2018). The Sentinel-2 which was launched by the European Space Agency (ESA) in 2015 became an alternative to Landsat. Compared to the Landsat imagery, Sentinel-2 provides 10 meter spatial resolution imagery of RGB bands with less revisit time and wider swath.

The data captured from satellites has a low ground resolved distance. This leads to a blurring of the images and making the detailed land features like the boundaries of water bodies difficult or almost impossible to identify. The lack of detailed land feature analysis impacts the results of environmental assessment (Sanga-Ngoie et al., 2012). Despite the availability of high-resolution sensors like GeoEye (0.46–1.84 m), WorldView (0.31–2.40 m), and IKONOS (1–4 m) (Tu et al., 2004) that are used in different applications, the difficulty of accessing the data creates hindrance in the development of pervasive applications. Alternative solutions are required to overcome the stated limitation to capture the land information in detail for analyzing inter and intra class variation.

2.1.2. Aerial Imagery (Air-borne)

The multispectral satellite imagery allows reliable extraction of water using various water indices (McFeeters, 1996; Domenikiotis et al., 2003; Kordelas et al., 2019) and specific bands based threshold methods. However, usage of optical imagery in the presence of clouds prevents the observation of the earth's surface (Shen et al., 2019). For this, some approaches have used radar data during the period of intense cloud cover to overcome the limitations of optical imagery (Markert et al., 2018; Manakos et al., 2020). On the contrary, most air-crafts and Unmanned Aerial Vehicles capture the data from very low elevation and provide high-resolution air-borne imagery when required considering the limitations of law regulation and weather condition (Iizuka et al., 2018; Pásler et al., 2015). Air-borne imagery is actively used to detect land cover and land use changes (Salamí et al., 2014; Feng et al., 2015), and natural disasters like floods and earthquakes (Abdelkader et al., 2013; Casado et al., 2018). Baker and Kamgar-Parsi (2010) presented autonomous shoreline navigation using UAVs. The use of UAVs to capture airborne imagery is comparatively cheap and fast for detailed land feature analysis of a specific area.

2.2. Relevant Approaches

2.2.1. Threshold Methods

In remote sensing studies, most of the water detection algorithms are based on water indices. In 1996, McFeeters (McFeeters, 1996) firstly designed a popular water index, the Normalized Difference Water Index (NDWI) for water mapping from satellite imagery. He has used the near-infrared (NIR) and the green bands of the Landsat Thematic Mapper (TM) for depicting water features. Xu (Xu, 2006) modified the NDWI by replacing the NIR band with shortwave-infrared (SWIR) and named it Modified Normalized Difference Water Index (MNDWI). MNDWI partly reduced the error rate generated by soil, vegetation, and urbanized areas. Feyisa (Feyisa et al., 2014) introduced the Automated Water

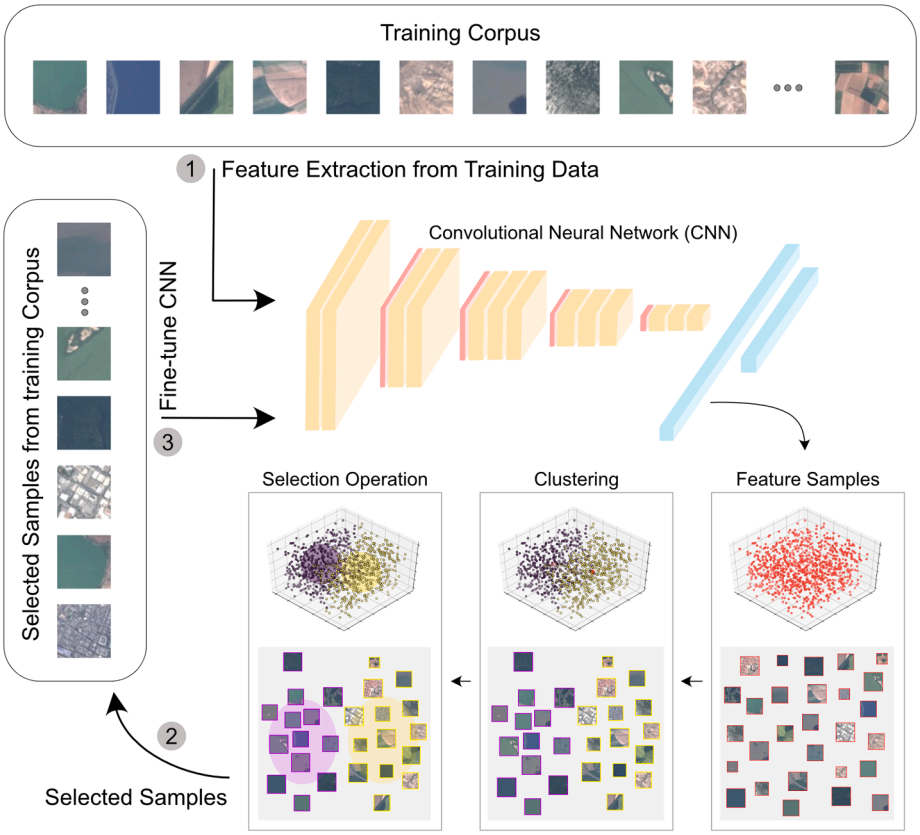


Fig. 1. UCL: Deep learning based Unsupervised Curriculum Learning for water classification. UCL (1) extracts the features of a training corpus using Convolutional Neural Network (CNN). It clusters the features into two classes. (2) It applies a selection operation to remove the noisy samples from the clusters. (3) The selected samples from the training corpus are used to fine-tune the CNN model. Once, the CNN model is fine-tuned, the steps are repeated until the CNN model has learned the patterns in the training corpus.

Extraction Index (AWEI) to cater for the misclassification of shadow as water by using multispectral bands. A new water index was created with linear discriminant analysis which Fisher (Fisher et al., 2016) revised by using five surface reflectance (SR) bands of Landsat. She also provided a thorough comparison of water indices for Landsat imagery.

Water detection requires rich spatial information to design threshold methods like Near Infra-Red (NIR) band that separates the water from the land. Most of the low-cost off-the-shelf UAVs are only equipped with conventional cameras that provide only the spectral bands red, green, and blue. RGB based threshold methods are only used for vegetation detection and observing its growth, e.g., Colour Index of Vegetation Extraction (CIVE) (Kataoka et al., 2003), Excess Green (ExG) (Woebbecke et al., 1995), Excess Red (ExR) (Meyer et al., 3543), Green Leaf Index (GLI) (Louhaichi et al., 2001), Normalized Green–Red Difference Index (NGRDI) (Tucker, 1979), Red-Green–Blue Vegetation Index (RGBVI) (Bendig et al., 2015), and many others. The lack of rich spectral information in imagery provided by most UAVs limits the use of indices for water body classification (Komarkova et al., 2019). In such scenarios, an effective machine learning algorithm is needed based on water classification for remote sensing imagery having limited spectral information but substantial scale variations.

2.2.2. *Machine Learning Methods*

Machine learning algorithms for water estimation can be categorized into supervised and unsupervised methods. Many waterbody classification algorithms have been designed using supervised methods, such as Support Vector Machines (SVM) Wang et al. (2018), Decision Trees (Davranche et al., 2013; Lefebvre et al., 2019; Acharya et al., 2016), Random Forests (Huang et al., 2002; Ko et al., 2015; Vries et al., 2017), Gradient Boosting (Mahdavi et al., 2018), and Deep Neural Networks (Isikdogan et al., 2017). In the past few years, bag-of-visual-words based methods employing K-Means and SVM have been used in several classification and target detection techniques leading to better accuracy (Yang and Newsam, 2010; Hu et al., 2015). It is noteworthy that the essential semantic information is stored within the spatial relationship of pixels instead of individual pixel intensity values. Many methods have been introduced involving image context to make the class information more explicit (Feng et al., 2015). Luo et al. (2014) proposed a hierarchical generative model, the Author-Genre-Topic Model (AGTM), to introduce context information. It was designed to perform annotation of satellite images. Recently, Generative Adversarial Neural networks (GANs) are being used in a semi-supervised manner to address the problem of cross-domain adaptation in semantic segmentation of remote sensing imagery Zhu et al. (2019), Li et al. (2021).

Deep learning has become a state-of-the-art method to extract more abstract features from lower layers to higher layers of the model. Comparing deep learning methods with shallow classification methods like SVM, deep learning solutions result in better learning models (Hinton et al., 2015). Cheng et al. (2017) replaced hand-crafted features with CNN for water bodies segmentation. Lin et al. (2017) used Fully Convolutional Network (FCN) to add multi-scale information. Noh et al. (2015) designed a multi-layer deconvolutional network to address the scale challenge. Wei et al. (2017) and Miao et al. (2018) used auto-encoders to extract high-level feature maps from high-resolution images. Fang et al. (2019) used the ResNet model to identify global water reservoirs. Yagmur et al. (2019) combined residual blocks in the inception network to detect shallow water areas. In spite of their excellent performance, supervised methods have some limitations. 1) If labelled data is not available, supervised methods require the collection of data and expert knowledge for data labelling which is a time-consuming and tedious task. 2) Supervised methods are domain specific. Their accuracy often decreases drastically when applied to different domain data about the same problem. Some researchers have also explored and used the concept of Curriculum Learning to efficiently train supervised deep neural networks (Ul-Hasan et al., 2015). The focus of our work is to overcome the stated issues by introducing an

unsupervised deep learning solution using the concept of curriculum learning for water body classification for RGB data.

3. Methodology

The proposed deep learning based UCL model learns the features of water bodies from remote sensing imagery in an unsupervised manner using the pseudo-labels generated by the clustering technique. The outline of the proposed method is shown in Fig. 1. UCL is composed of two main modules (i) A pre-trained deep learning architecture (CNN) to extract and learn features from the remote sensing data, (ii) an unsupervised clustering technique to cluster the extracted features. A UCL based selection operation is added between clustering and fine-tuning to extract the samples present near the clusters' centroids, called "reliable samples". The deep learning model is fine-tuned on the extracted reliable samples. UCL is composed of the following steps:

Step 1. Extract features of remote sensing imagery of water bodies and non-water bodies using a pre-trained deep learning architecture.
Step 2. Create two clusters on the extracted features of remote sensing imagery, assigning them pseudo-labels of water-bodies and non-water-bodies clusters.
Step 3. From each cluster, select the reliable images using the UCL based selection operation.
Step 4. Fine-tune the deep learning module on reliable samples using pseudo labels given by the clusters.
Step 5. Extract features of the whole training corpus of remote sensing imagery using the pre-trained model of the previous step.
Step 6. Repeat steps 2 to 5 until the deep learning model is converged.

In the beginning, the CNN, pre-trained on a different domain (ImageNet) is used to extract features from remote sensing imagery. These features are clustered into two, assuming them to be of water and non-water category. As the clusters are created from the features generated from a deep learning model trained on a different domain, we may obtain noisy clusters (for water classification of the remote sensing imagery). To filter out the reliable samples from the clusters, a UCL based selection operation is used to extract a small number of reliable samples. These are the samples present near the centroid of the clusters, containing the prominent features. The CNN model is fine-tuned on these reliable samples with pseudo labels assigned by clustering. The reliable samples restrict the CNN to learn only the prominent features of the clusters by avoiding unnecessary noise. The updated CNN is used for feature extraction in the preceding iteration. With every iteration, the model learns the features of remote sensing imagery with pseudo labels of clusters resulting in comparatively better clusters than the previous iteration. The process iterates until the CNN model has converged. This process is called unsupervised because it only needs the pseudo labels of the clusters to fine-tune the CNN model.

3.1. Deep Learning Model

In UCL, a deep learning module is used for feature extraction from remote sensing imagery. Later, this module is fine-tuned on selected reliable samples, say, of water bodies and non-water bodies. Several deep learning models like VGG-16, ResNet-50, DenseNet, Inception Net, and Xception Net have been explored in this work to demonstrate the generalization of the introduced framework. VGG-16 and ResNet-50 outperformed the other networks for water bodies classification from remote sensing imagery. Hence, we have used VGG-16 for our final UCL algorithm due to its lower computational complexity compared to ResNet-50. The described training process is general and independent of the CNN used. It will work with VGG as well as with Inception Net, and Xception Net. The training process of VGG-16 in UCL can be decomposed into two parts:

1. Feature Extraction: VGG-16, pre-trained on the ImageNet dataset is used for feature extraction of remote sensing imagery in the first iteration. The output of the last convolutional layer is extracted to get feature maps of each sample in the dataset. The extracted feature maps are flattened to get the feature vectors. These feature vectors are clustered into two, assuming them to be of water and non-water bodies. From these clusters, reliable feature samples are selected using the UCL based selection operation for fine-tuning the deep model.
2. Fine-tuning of VGG-16 with reliable images: The model is fine-tuned on the training set of reliable samples considering their cluster as their pseudo label. We have modified the input layer of VGG-16 according to our aerial image patch size and the output layer to the number of clusters we generate. In the current scenario, the considered two classes for training are; water-body and non-water-body.

3.2. Clustering

The features extracted from the deep learning module are clustered using an unsupervised clustering technique. We have explored three different types of clustering techniques, K-Means, hierarchical clustering, and Fuzzy C-Means (FCM). These techniques are suitable for the considered problem of water and non-water classification of remote sensing imagery patches as they have a fixed number of classes. FCM and K-Means clustering generated better clusters than hierarchical clustering (see Section 5). We have deployed K-Means clustering as it has lower time complexity than FCM.

Suppose the features extracted from remote sensing imagery patches $\{x_i\}_{i=1}^{N}$ using the deep learning model $\phi(.,\theta_i)$are represented by $\{f_i\}_{i=1}^{N}$.

$$\{f_i\}_{i=1}^{N} \leftarrow \phi(\{x_i\}_{i=1}^{N}, \theta) \tag{1}$$

These features are clustered such that each feature vector is assigned a cluster label $\{y_i\}_{i=1}^{N}$ where $y_i \in \{1, \ldots, k\}$on the basis of a minimum distance from the centroid c_k, where c is the centroid of k^{th} cluster. In the current scenario, $k = 2$ to generate two clusters, assuming them to be of water bodies and non-water bodies.

$$\{y_i\}_{i=1}^{N} \leftarrow \min \sum_{i=1}^{N} \sum_{k=1}^{2} |f_i - c_k| \tag{2}$$

These generated pseudo-labels, $\{y_i\}_{i=1}^{N}$, are later used for fine-tuning the CNN model.

3.3. UCL based Selection Operation

Using the concept of CL we want to extract reliable samples for fine-tuning the CNN. We achieve this in an unsupervised manner, by selecting features near the centroid of a cluster. More specifically, we select all features that are at a distance λ from the centroid of the cluster (see Selection Operation in Fig. 1). The parameter λ is a constant value that can be adjusted according to the requirement. We have used $\lambda = 0.85$ after empirical evaluations.

The closest feature vector to the centroid is considered as a centroid feature vector, $\{f_k\}_{k=1}^{2}$where k represents the cluster.

$$\{f_k\}_{k=1}^{2} \leftarrow \min\{|f_{ik} - c_k|\}_{i=1}^{N} \tag{3}$$

We calculate the similarity between a specific feature vector f_i belonging to a cluster k and the centroid feature vector f_k using the inner product, i.e.

$$\{\gamma_i\}_{i=1}^{N} \leftarrow \{f_{ik} \cdot f_k\}_{i=1}^{N} \tag{4}$$

If the calculated similarity is greater than the λ, the sample of the

considered feature vector is declared as a reliable sample $x_i^{'}$ and the cluster label is considered as the pseudo sample label for the next training cycle.

$$\{x_i^{'}\}_{i=1}^{M} \leftarrow \{\gamma_i\}_{i=1}^{N} < \lambda \quad (5)$$

The number of extracted reliable samples vary at every iteration of fine-tuning.

3.4. Progressive Learning

Initially, when the model is not trained on remote sensing imagery, the features extracted from this model may result in loose clusters. These clusters are less dense and result in a set of only a few reliable images. In the beginning, the network is fine-tuned on this set of a few reliable images considering their cluster as their pseudo label, either water or non-water. Then the fine-tuned network is used to extract features of the whole training corpus and new clusters are generated. The selection operation is performed on these clusters to extract reliable samples. This time the clusters might be comparatively dense than they were in the previous iteration, and we get more reliable images. Progressively, the model gets stronger by learning the patterns from the data and the set of reliable images iteratively grows leading to self-paced learning.

The proposed UCL technique is an unsupervised binary classifier where the model has learnt the features and is able to create clusters of water and non-water. Now the question is how to know which cluster indicates which class? The automation of mapping the pseudo-labels to true labels is beyond the scope of this work and can be considered as one of the future directions.

3.5. Implementation Details

The experiments were conducted on a GPU machine having an NVIDIA Titan-X GPU for training and fine-tuning with 32 GB RAM and Linux operating system. It took about 4.5 h for training the model on the considered datasets. We used a Stochastic Gradient Descent (SGD) optimizer and categorical cross-entropy loss. Learning rate was set to 0.0001, momentum to 0.9 and batch size to 16 images. The training dataset with pseudo labels is split into 80% for training and 20% for validation for progressive learning of UCL. The input layer of the deep learning model is set to 64x64x3. As the considered three datasets have a difference in their image patch sizes (28x28, 61x61, and 64x64), the patches of all the datasets were interpolated to patch size 64x64 to make them suitable as an in input to the deep learning model.

4. Datasets

UCL takes the input in the form of an image patch and classifies it either to be of water or non-water category. Therefore, huge tiles of aerial images were broken down into smaller patches on the basis of the requirement for better classification. Only RGB bands of the images were considered to model the situation where only these bands are available, like high-resolution UAV data which usually does not contain multi-spectral bands. We have used two publicly available datasets, namely, EuroSAT (Helber et al., 2019) and SAT-6 (Basu et al., 2015), and our newly created PakSat dataset to demonstrate the effectiveness of the proposed architecture.

This study has considered the imagery from different parts of the world to address variations in the spectral responses of water depending on the region of the globe. EuroSAT carries Sentinel-2 imagery patches of 34 European countries, SAT-6 covers the area of California, and PakSAT is composed of Sentinel-2 imagery patches of Pakistan.

4.1. SAT-6 Dataset

SAT-6 (Basu et al., 2015) is a high-resolution dataset captured from an aircraft providing 1 meter GSD pixel resolution covering different parts of California. It is composed of small patches of size 28 × 28 divided into six classes. Initially, it has four bands; red, green, blue, and infrared. The dataset was preprocessed to remove the infrared band from the corpus to make it suitable for RGB input. The patches were divided into two categories i.e. water and non-water class. Both classes contain an almost equal number of patches, having a corpus of 7500 image patches for fine-tuning and 3000 for testing. This high-resolution dataset was used to evaluate the robustness of UCL for scale variation and to prove the hypothesis of the progressive learning behaviour of the model.

4.2. EuroSAT Dataset

EuroSAT (Helber et al., 2019) is composed of image patches from different regions of 34 European countries. The image parches are of sentinel-2 satellite having a resolution of 10 m per pixel for red, green, and blue bands. It consists of 10 classes of land cover and land use. We have only used the red, green, and blue bands of the data and divided the dataset into two classes by considering the “river” and “sea & lake” as water class and the rest of the classes as the non-water class in such a way that there are almost 50% samples of water and 50% of non-water category. The purpose behind balancing the two classes is the unbiased training of the deep learning model. The size of each patch is in EuroSAT dataset is 64 × 64. We have considered 8000 image patches for fine-tuning and 3000 for testing.

4.3. PakSAT Dataset

With the PakSat dataset, we have developed the dataset for water bodies segmentation from Sentinel-2 satellite imagery. This dataset is composed of water bodies of Pakistan. We have applied the threshold methods to get a roughly estimated mask of water pixels in the image. Later, the correction of the generated mask is done manually. The PakSat dataset is composed of 61 × 61 sized patches of water, non-water, and mixed classes. Each patch has 14 bands, the first 13 bands are of Sentinel-2 and the 14 th is the generated mask. In this study, we have only considered the red, green, and blue bands of the data having a pixel resolution of 10 m. The dataset consists of three classes; namely water, non-water, and mixed. The patches containing more than 75% of water pixels are declared as water patches. Patches with less than 75% of water are placed in the mixed category. The patches with no water pixel were classified as non-water patches. Lets have a look into the process of PakSAT dataset creation.

4.3.1. Process for creating the PakSAT dataset

The creation of PakSAT dataset was started with the downloading Sentinel-2 tiles of water reservoirs and part of the Indus river of Pakistan from June 2015 till October 2019. After downloading the required tiles of Sentinel-2 imagery, the following steps were taken one by one to create the ground truth of the PakSAT dataset.

4.3.1.1. Resampling. The Sentinel-2 bands come in varying spatial resolution of 10 m, 20 m, and 60 m. For overlying data, the cell resolution should be the same. For this, each band was upsampled to achieve 10 m resolution using the Bilinear Resampling technique. Bilinear is an interpolation technique that considers the values of the four nearest pixels to calculate the value for the current pixel on the output image. The calculated new values on the output raster are the weighted average of the considered four nearest values. The four values are considered on the basis of their distance from the center of the output pixel. Resampling is processed by Data Management Tools in ArcGIS. All bands of the spatial resolution of 20 m e.g. Band5, Band6, Band7, Band8A, Band11, and Band12 of Sentinel-2 and bands of 60 m resolution e.g. Band1, Band9, and Band10 of Sentinel-2 resolution was resampled to 10 m by using the Bilinear Resampling technique. We have performed this step in

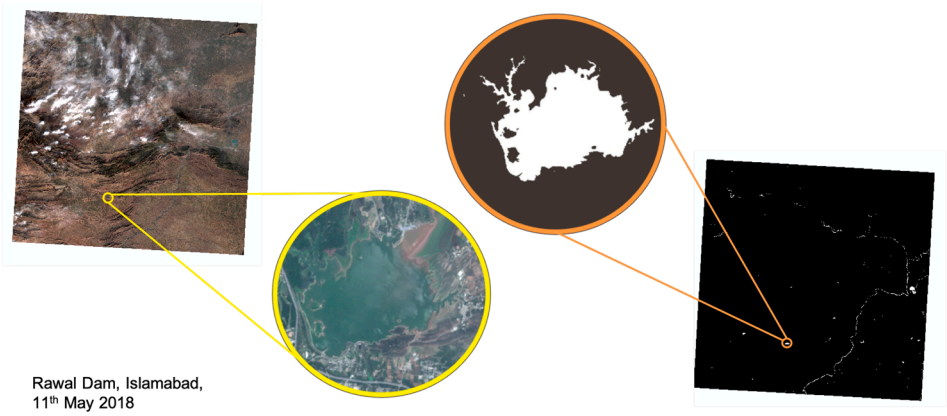


Fig. 2. Sentinel-2 tile captured on 11th May 2018 containing Rawal Dam along with its approximated binary mask generated by using NDVI and NDWI.

the creation of the PakSAT data to make it more practical for more interested researchers who intended to use the dataset with multispectral bands.

4.3.1.2. Labelling Water Features. Clear and greenish reflection water bodies show high reflectance values for the green band than red and blue. Whereas muddy water bodies show high reflectance on the red band than green and blue. Hence, the Normalized Difference Water Index (NDWI), uses the green band along with the near-infrared (NIR) band, whereas the Normalized Difference Vegetation Index (NDVI), uses the red band with near-infrared (NIR). Both of the approaches map the water bodies differently according to the difference in their color. We have used both of the indices to estimate the water bodies in Sentinel-2 imagery.

Normalized Water Index (NDWI) is another threshold method to detect water from remote sensing data. NDWI absorbs the NIR reflectance and emphasizes the green band reflectance to detect water bodies. Thus, water pixels become prominent having positive values, and other categories like soil and vegetation carry zero or negative values. A threshold on these values gives an estimated binary mask of water bodies.

$$NDWI = \frac{Green - NIR}{Green + NIR} \tag{6}$$

The values of NDWI range between −1 and 1. In most cases, NDWI was used to detect water features but in some tiles, where there are some problems in detecting features due to shadows, clouds, or muddy water, NDVI was deployed. The values of NDVI as well range between −1 and 1 as both are normalized indices. For NDVI, the negative values close to −1 represent water/ The values close to zero usually correspond to barren areas like rocky or sandy land or snow category. Low positive values roughly up to 0.4 classify green lands like grass and shrubs. Whereas positive values approaching 1 represent tropical and temperate rainforests.

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{7}$$

The pixels of water were labelled by reclassifying the pixels. The pixel values representing water were labelled as 1 and 0 otherwise. This resulted in a binary mask for water bodies. These auto-generated binary masks contained some errors like cloud cover, shadows, and boundary pixels of the water bodies.

4.3.1.3. Manual Correction. Manual Correction included the step of locating the errors by visualizing the satellite image and the corresponding generated binary mask for a specific area of interest containing water body (see Fig. 2). Wherever the pixel(s) was misclassified, it was manually corrected by changing the binary mask value from 0 to 1 and vice versa.

Table 1
The count of patches in each category generated from Sentinel-2 tile captured on a specified date containing water body.

Water Bodies	Capturing Date	Water	Land	Mixed	Total
Chashma Reservior	2016–09-05	877	160	-	1037
	2017–11-06	315	296	-	611
Darawat Dam	2019–10-19	118	510	131	759
Ghazi Barotha Reservior	2019–04-06	9	778	137	924
Gomal zam Dam	2016–08-15	372	-	-	372
	2019–06-08	58	189	76	323
Indus River	2016–07-22	287	-	-	287
Manchar Lake	2017–06-08	650	215	-	865
	2017–08-27	570	1651	173	2394
	2018–08-01	554	-	-	554
	2019–10-16	719	1694	242	2655
Mangla Dam	2016–02-06	455	3186	228	3869
	2016–10-20	673	-	-	673
	2017–12-07	206	2299	402	2907
	2018–08-19	492	3336	660	4488
	2018–12-11	452	-	-	452
Rawal Dam	2016–11-12	75	-	-	75
	2018–10-07	73	-	-	73
	2019–06-10	48	951	171	1170
Tarbela Dam	2016–07-04	347	-	-	347
	2017–07-07	235	2102	212	2549
	2019–04-06	389	1954	335	2678
Total Count		7974	19321	2767	29389

4.3.1.4. Bands Composition. Band composition is similar to layer stacking in which after resampling into the same resolution of 10 m, all the 13 bands are stacked together in the standard order of Sentinel-2. Lastly, the generated ground-truth mask is added to it as a 14th band and stored as a single raster file.

4.3.1.5. Splitting into Patches. As each raster file was huge in size, they should be divided into smaller patches to make it suitable for UCL. Each raster file composed of 14 channels (13 Sentinel-2 bands and 14th ground-truth binary mask) is split into smaller non-overlapping patches of size $61 \times 61 \times 14$. The size of the patch is an important parameter as UCL takes the patch and does classification at the patch level.

4.3.1.6. Categorization of Patches. Some of the generated patches contain only water, some do not contain water or contain the coastline of the water having a portion of water and non-water part. We segmented the patches into three different categories on the basis of the water pixels

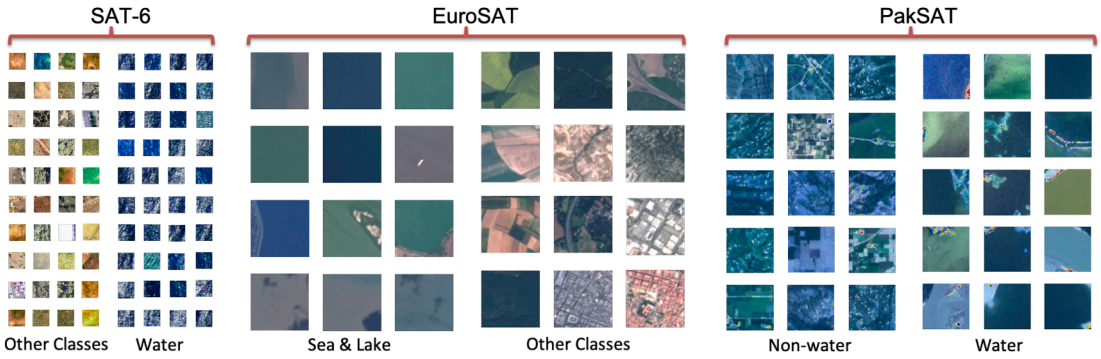


Fig. 3. Some patches of Sat-6, EuroSAT and PakSAT Datasets.

Table 2
Describes the direct inference of EuroSAT, PakSAT, and SAT-6 datasets on VGG-16 considering two different techniques. Random FC indicates the results with random initialization of the FC layer. K-Means shows the results for clustering the features extracted from VGG-16 and classified by K-Means clustering.

Initial Weights	Test Set	F1-Score (%)	
		Random FC	K-Means
ImageNet	EuroSAT	44.46	54.45
ImageNet	PakSAT	51.98	57.13
ImageNet	Sat-6	58.20	65.50

Table 3
F1-Scores of VGG-16 supervised fine-tuned and tested on each dataset; EuroSAT, PakSAT, and SAT-6. The last column reports the results of UCL fine-tuned and tested on each considered dataset.

Fine-Tuned & Tested	F1-Score (%)	
	Supervised	UCL
EuroSAT	99.49	84.05
SAT-6	99.53	90.89
PakSAT	96.07	87.66

they contain. If a patch has no water pixel, it belongs to the "Land" category. If a patch has 75% or more water pixels then it belongs to the "Water" category. If a patch has water less than 75%, it is categorized as "Mixed" containing both water and non-water regions.

In total 29389 patches were created. Out of which 7974 patches belong to the Water category, 19321 belong to the Land category, and 2767 belong to the Mixed category. The details of the water patches can be seen in Table 1.

In this work, we have used the water and non-water patches with RGB bands for unsupervised progressive learning of water classification from satellite imagery. Fig. 3 provides a visualization of the patches of the considered three datasets.

The size of the input patch has great significance. If the patch size is large, there are chances of mixing the boundary pixels of multiple land covers. The smaller the patch size the better the results. Our main focus is to develop an unsupervised deep learning based approach considering the prominent features of each class. This could be done by avoiding the patches containing the features of both classes significantly like 50% of each class in the patch. It may confuse the network at the time of training for binary classification. If the network is designed for multiple classes and a new class can be added called "mix class", the network will be able to learn such patches containing features of multiple classes which can be considered as a future direction of the work. Our major focus is on learning the prominent features for binary classification.

5. Experimental Validation

The progressive learning behaviour of UCL was analyzed by conducting multiple experiments on the considered datasets, namely Sat-6, EuroSAT, and PakSAT. The experiments are divided into 4 subsections; (i) direct testing of the considered dataset on ImageNet weights, (ii) fine-tuning of VGG-16 and UCL supervised, (iii) their cross domain adaptation, (iv) clustering analysis and (v) error analysis.

5.1. Direct Testing on ImageNet Weights

Before conducting the experiments of UCL progressive learning and supervised comparison, direct testing on VGG-16 with ImageNet weights of EuroSAT, SAT-6 and PakSAT is done (see Table 2).

VGG-16 with ImageNet weights is directly tested on EuroSAT, PakSAT, and SAT-6 datasets. Assuming the dataset to be unlabelled, two strategies were followed; (1) the classification layer of VGG-16 is randomly initiated with 0 mean and 0.001 standard deviations, and (2) VGG-16 with ImageNet weights are used as a feature extractor for remote sensing data. In Table 2, Random FC represents the results of randomly initializing the fully connected binary classification layer. As the random initialization of the classification layer is not aware of the considered datasets, it makes the classification quite challenging leading to unsatisfactory results. Whereas, in the second approach, K-Means has been used for the classification of deep extracted features which lead to comparatively better performance for all three datasets. The EuroSAT gives the lowest performance, it is because we have combined the River class with Sea & Lake. Sea & Lake class carries the Sentinel-2 patches of only water. Whereas, in River class patches we can observe a great part of the land with the river stream. This intermixes the river patches with non-water class leading to poor intra class variation for clustering and compromised F1-Score. The PakSAT and SAT-6 datasets carry prominent patches of water and non-water categories.

5.2. Fine-tuning of VGG-16 and UCL

In general, supervised deep learning models have better performance than unsupervised models as they are trained using ground-truth labels. UCL's performance has been analyzed considering the supervised model's performance as the benchmark. VGG-16 with ImageNet weights is fine-tuned in a supervised manner on EuroSAT, SAT-6 and PakSAT datasets that reported the F1-Score of 99.49%, 99.53% and 96.07%, respectively, see Table 3. All the fine-tuned models have learned the remote sensing features to classify the water patches for the respective datasets.

We have analyzed the progressive learning behaviour of UCL that is capable of learning the variations in the new dataset, progressively. For UCL training, we assume that there are no labels available for the training process. UCL uses the clustering technique to generate the pseudo labels to train the deep learning model. The CNN based model of UCL with ImageNet weights has been fine-tuned with EuroSAT, PakSAT, and SAT-6 datasets in unsupervised progressive learning behaviour. The

Table 4
F1-Scores of VGG-16 fine-tuned in a supervised manner on each dataset and tested on the other two datasets. The last column reports the results of UCL fine-tuned on each dataset and tested on the other two datasets. The considered datasets are EuroSAT, PakSAT, and SAT-6.

Fine-Tuned	Tested	F1-Score (%)	
		Supervised	UCL
EuroSAT	PakSAT	70.68	78.00
	SAT-6	32.84	75.76
SAT-6	PakSAT	63.16	68.97
	EuroSAT	42.83	72.43
PakSAT	EuroSAT	62.70	79.00
	SAT-6	59.92	71.72

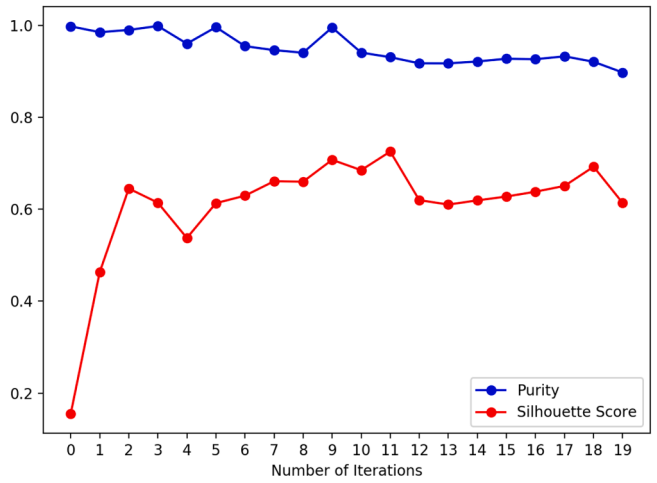


Fig. 4. Graph showing the purity and Silhouette Score of generated clusters for water bodies and other regions of SAT-6 over every iteration of fine-tuning of VGG16.

UCL reported promising results on all the three datasets by giving F1-Score above 80%, see Table 3. The supervised model performed better than UCL but there is a huge trade-off of data labelling which is quite an exhaustive task.

Considering the dataset with no labels, the direct testing on ImageNet weights with K-Mean clustering performed comparatively better than the random classification layer (see Table 2). It reported the F1-Score of 54.45% for EuroSAT, 57.13% for PakSAT, and 65.5% for SAT-6 dataset. Later, deploying the UCL for progressive unsupervised learning from the data, it is able to learn the features from the unlabelled data by reporting the considerable improvement in the F1-Score. The evaluated F1-Score on UCL is 84.05% for EuroSAT, 87.66% for PakSAT, and 90.89% for SAT-6 dataset, see Table 3. The F1-Score is improved by around 20% for EuroSAT and PakSAT datasets, and 25% for SAT-6 dataset.

5.3. Cross-domain Adaptation

To analyze the domain adaptation, we have tested the model trained on one dataset on the other two datasets (see Table 4). It has been observed that the test accuracy for supervised fine-tuned VGG-16 is quite low on the other datasets, indicating that the supervised models are data-specific and lack adaptability in the cross-domain adaptation scenario. Here, the variation among the datasets is due to different data acquisition platforms (space-borne/air-borne), resulting in different image properties with varying image resolution.

To further analyze the progressive learning behaviour of the proposed model and its adaptation to the new dataset, each UCL model

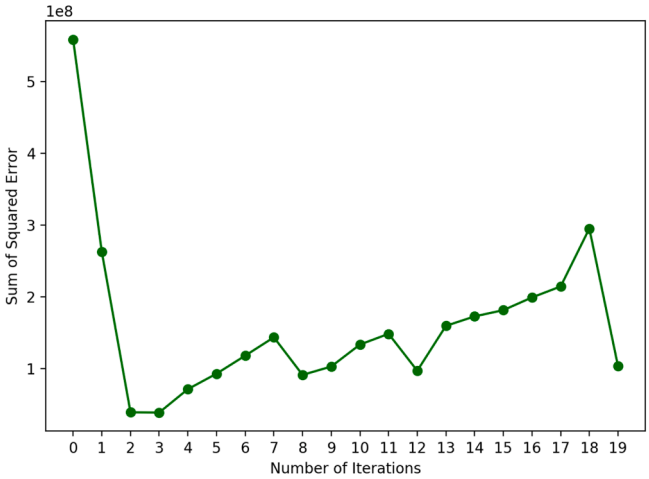


Fig. 5. Graph showing Sum of Squared Error of generated clusters for water bodies and other regions of SAT-6 over every iteration of fine-tuning of VGG16.

trained on one dataset was tested on the other two considered datasets. It can be seen in Table 4, the supervised model trained on EuroSAT does not perform that well on PakSAT and SAT-6 datasets with an F1-Score of 70.68% and 32.84%, respectively. Whereas, UCL fine-tuned with EuroSAT achieved comparatively better F1-Scores on the other datasets i.e., 78.00% for PakSAT and 75.76% for SAT-6 datasets. Similarly, the supervised trained models and the UCL trained models of SAT-6 and PakSAT were tested on the other two datasets and a similar trend was observed (as reported in Table 4).

5.4. Clustering Analysis

The UCL results of the SAT-6 dataset are further analyzed by observing the clusters generated using K-Means during training. We have observed the training procedure of UCL for 20 iterations of fine-tuning. Initially, the clusters are created using the features extracted from a model trained on a different domain, ImageNet, which does not contain any remote sensing data. Consequently, the generated clusters are not compact and loosely packed for remote sensing imagery. To evaluate the compactness of the clusters, purity and Silhouette Score are computed, see Fig. 4. The purity is a supervised measure that calculates the correctly classified samples over the total number of samples in the cluster. Whereas, Silhouette Score is an unsupervised measure that calculates the compactness of the clusters on the basis of the distance between each sample within the cluster and the neighboring clusters.

It can be seen in Fig. 4 that the Silhouette Score is as low as 0.15 at the start, showing the lack of compactness in the clusters. Later, with iterations of fine-tuning, the value grows indicating the saturation in the compactness. Whereas, the purity remains at a good score in the range of 0.90 to 1.00 over iterations.

The Sum of Squared Error (SSE) is evaluated for the clusters generated from SAT-6 dataset which is also an objective function of K-Means clustering. In Fig. 5, the graph represents the values of SSE for generated clusters at every fine-tuning iteration. The vertical axis represents the values of SSE and the horizontal axis are the fine-tuning iterations of the model. The graph shows a quite high value of SEE at the start when clustering is done with features extracted from a pre-trained model of ImageNet. As soon as the model is fine-tuned on SAT-6 patches, the SEE score is exponentially decreased. We can see a small peak at the end of the curve showing that the progressive fine-tuning of the model has led to somewhat over-fitting and fine-tuned models of these iterations can be ignored. This graph helps to choose the candidate fine-tuned models over iterations. The fine-tuning iterations having minimum SSE can be the potential fine-tuned model for the deployment.

As the clusters are loosely packed at the start, we extract the reliable

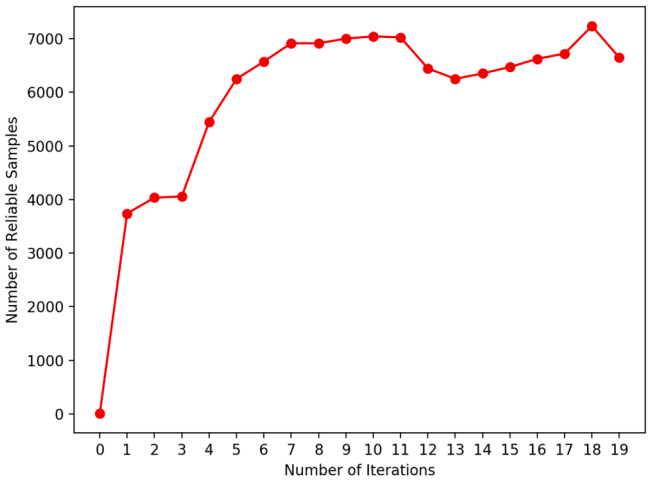


Fig. 6. Graph showing the count of reliable samples of SAT-6 over every iteration of fine-tuning of VGG16.

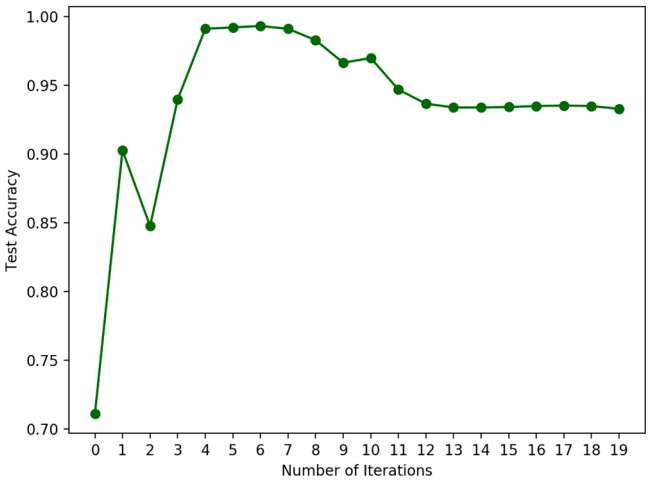


Fig. 8. Graph showing the test accuracy of SAT-6 over every iteration of fine-tuning of VGG16.

samples present near the centroid of the clusters using the selection operation of UCL. The deep learning model is fine-tuned on these samples. This step avoided the noisy samples of the clusters and restricted the model to learn random features.

It can be seen in Fig. 6 that only a few reliable samples (exactly 4) are extracted at the start using the pre-trained model further indicating that the samples are loosely packed in the clusters. With iterations of fine-tuning, the count of reliable samples is significantly growing. After a few iterations, the growth in the count of the reliable samples seemed saturated. After 4 iterations, the count of reliable samples remains above 6000 indicating that the reliable set contained the majority of the samples from the corpus of size 7000.

In UCL, VGG-16 is trained with pseudo-labels generated by the clustering technique, K-Means. The model has reported almost 0% error over cross-validation on every iteration of fine-tuning. The cross-validation corpus is a fraction of the training corpus where the generated pseudo-labels are used for cross-validation purpose as well. The model is fine-tuned end-to-end to the classification layer. Fig. 7 shows a visualization of centroid patches of the clusters at every iteration of fine-tuning for the SAT-6 dataset. It can be seen that the centroids of both clusters at every iteration are belonging to the water and non-water class indicating that the model is able to distinguish both classes.

The fine-tuned models of 20 iterations are evaluated on the test corpus, see Fig. 8. It can be seen that the model has converged well on the 4th iteration, reporting the F1-Score of 0.99. The accuracy remained consistent for the next two iterations. After that, the accuracy tends to decrease indicating a sign of overfitting. It may be because of the multiple interactions of fine-tuning the model over the same dataset. The model reported the highest accuracy 99.31% at the 6th iteration. Table 5 shows the confusion matrix of the 6th iteration of fine-tuning. It can be seen that most of the patches are correctly classified with the exception of 20 false-negative patches that are belonging to other regions and are declared as water by the model.

To analyze the change of pseudo-labels among the patches, we have evaluated the count of the patches whose pseudo-labels are changed in the next iteration, see Fig. 9. It can be seen that at the 0th iteration, all the patches are predicted with a specific class. Later, with every iteration, the count of change in the labels of patches tends to decrease and gets converged after a few iterations.

Table 5
Confusion Matrix for SAT-6 test corpus on best iteration of fine-tuning of VGG-16.

	Predicted →		
Actual ↓	Water	Non-water	Total
Water	1,460	0	1,460
Non-water	20	1,440	1,460
Total	1,480	1,440	2,920

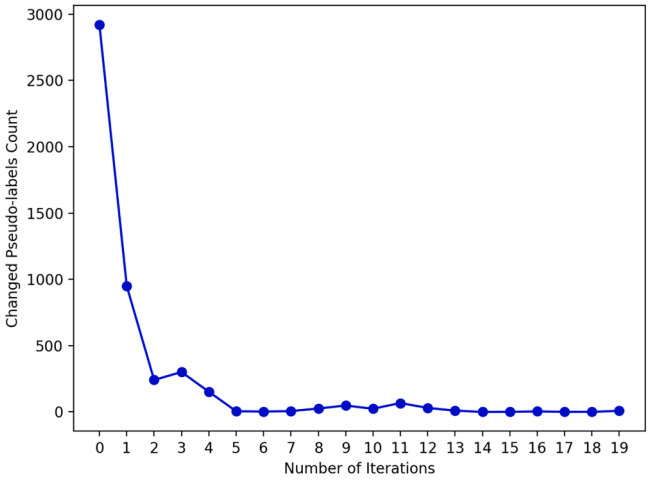


Fig. 9. Graph showing the count of the patches whose pseudo-labels are changed in the next iteration of fine-tuning of VGG16.

5.5. Error Analysis

The extracted reliable samples of the best fine-tuning iteration of EuroSAT demonstrate that the visual features of residential, industry,

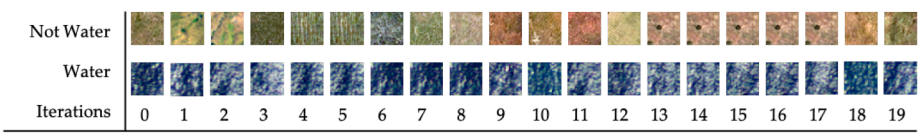


Fig. 7. The centroids of the clusters of SAT-6 at each iteration of fine-tuning of VGG-16.

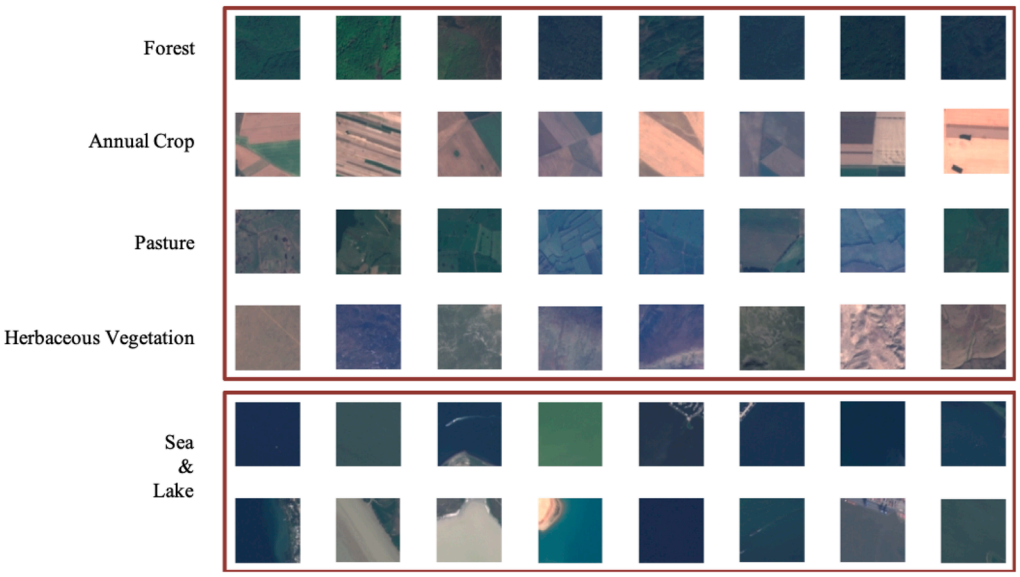


Fig. 10. These are some failure cases that where intermixed with water class. The generated clusters were intermixing forest, annual crop, pasture and herbaceous vegetation with water class because of resemblance in the appearance.

highways, annual and permanent crops were properly classified as non-water categories. Whereas, the forest patches show a similar colour as the reflection of some lakes and seawater which resulted in intermixing of some forest patches with water and vice versa. Herbaceous vegetation patches also showed a resemblance to water. It could be because of the low resolution of Sentinel-2 to appropriately distinguish the intra-class variation in different lands. Some of the failure cases have been visualized in Fig. 10.

6. Conclusion and Outlook

In this paper, we have introduced an unsupervised method UCL to categorize water bodies from remote sensing imagery and showed that the unsupervised deep learning approach can learn the desired features and have the tendency to outperform the supervised model with respect to domain adaptation. The supervised models of deep learning need a massive dataset of labelled images to train the architecture, which is a tedious, exhausting, and time-consuming task. The unsupervised algorithm removes the requirement of a labelled dataset for training the architecture and perform classification. The datasets used to prove the hypothesis are, EuroSAT (covering 34 European countries), PakSAT (covering Pakistan), and SAT-6 (covering California). This paper has shown the efficiency of unsupervised architecture by reporting the F1-Score around 85% to 91% for considered datasets (see Table 3). We have also analyzed the domain adaptation of UCL to check its robustness using EuroSAT, PakSAT, and SAT-6 datasets. We have trained the UCL on one dataset and tested its performance on the other datasets. UCL was able to give better domain adaptation performance on other datasets than supervised models with a considerable difference in the F1-Score from 8% to 42% (see Table 4). However, a big room of work is still to be done in the field. We worked on the patch-wise classification of the images considering red, green, and blue bands only which can be extended to pixel-based segmentation. This work also focused only on binary classification and could be further extended to the multi-class classification of aerial photographs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

United Nations World Water Development Report 2020: Water and Climate Change, 2019.

Abdelkader, M., Shaqura, M., Claudel, C.G., Gueaieb, W. A UAV based system for real time flash flood monitoring in desert environments using Lagrangian microsensors. In: 2013 International Conference on Unmanned Aircraft Systems, ICUAS 2013 - Conference Proceedings. 2013. doi:10.1109/ICUAS.2013.6564670.

Acharya, T.D., Lee, D.H., Yang, I.T., Lee, J.K. Identification of water bodies in a landsat 8 oli image using a j48 decision tree. Sensors (Switzerland) 2016;16. doi:10.3390/s16071075.

Baker, P., Kamgar-Parsi, B. Using shorelines for autonomous air vehicle guidance. Computer Vision and Image Understanding 2010;114. doi:10.1016/j.cviu.2010.01.009.

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., Nemani, R., 2015. DeepSat - A learning framework for satellite imagery. In: GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems. https://doi.org/10.1145/2820783.2820816.

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M.L., Bareth, G., 2015. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. Int. J. Appl. Earth Obs. Geoinf. 39 https://doi.org/10.1016/j.jag.2015.02.012.

Bengio, Y., Louradour, J., Collobert, R., Weston, J. Curriculum learning. In: ACM International Conference Proceeding Series. volume 382; 2009. doi:10.1145/1553374.1553380.

Bhowmik, S., 2020. Ecological and economic importance of wetlands and their vulnerability: a review. Current State and Future Impacts of Climate Change on Biodiversity 95–112.

Buma, W.G., Lee, S.I., Seo, J.Y. Recent surface water extent of lake Chad from multispectral sensors and GRACE. Sensors (Switzerland) 2018;18. doi:10.3390/s18072082.

Caron, M., Bojanowski, P., Joulin, A., Douze, M. Deep clustering for unsupervised learning of visual features. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). volume 11218; 2018. doi:10.1007/978-3-030-01264-9_9.

Casado, M.R., Irvine, T., Johnson, S., Palma, M., Leinster, P., 2018. The use of unmanned aerial vehicles to estimate direct tangible losses to residential properties from flood events: A case study of Cockermouth Following the Desmond Storm. Remote Sensing 10. https://doi.org/10.3390/rs10101548.

Cheng, G., Li, Z., Yao, X., Guo, L., Wei, Z. Remote Sensing Image Scene Classification Using Bag of Convolutional Features. IEEE Geoscience and Remote Sensing Letters 2017;14. doi:10.1109/LGRS.2017.2731997.

Davranche, A., Poulin, B., Lefebvre, G. Mapping flooding regimes in Camargue wetlands using seasonal multispectral data. Remote Sensing of Environment 2013;138. doi:10.1016/j.rse.2013.07.015.

Domenikiotis, C., Loukas, A., Dalezios, N.R. The use of NOAA/AVHRR satellite data for monitoring and assessment of forest fires and floods. Natural Hazards and Earth System Science 2003;3. doi:10.5194/nhess-3-115-2003.

Fan, H., Zheng, L., Yang, Y., 2018. Unsupervised person re-identification: Clustering and fine-tuning. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 14 (4). https://doi.org/10.1145/3243316.

Fang, W., Wang, C., Chen, X., Wan, W., Li, H., Zhu, S., Fang, Y., Liu, B., Hong, Y., 2019. Recognizing global reservoirs from Landsat 8 images: A deep learning approach.

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (9). https://doi.org/10.1109/jstars.2019.2929601.
Feng, Q., Liu, J., Gong, J., 2015. UAV Remote sensing for urban vegetation mapping using random forest and texture analysis. Remote Sensing 7. https://doi.org/10.3390/rs70101074.
Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sensing of Environment 2014;140. doi:10.1016/j.rse.2013.08.029.
Fisher, A., Flood, N., Danaher, T. Comparing Landsat water index methods for automated water classification in eastern Australia. Remote Sensing of Environment 2016;175. doi:10.1016/j.rse.2015.12.055.
Giustarini, L., Hostache, R., Matgen, P., Schumann, G.J., Bates, P.D., Mason, D.C. A change detection approach to flood mapping in Urban areas using TerraSAR-X. IEEE Transactions on Geoscience and Remote Sensing 2013;51. doi:10.1109/TGRS.2012.2210901.
Guo, M., Li, J., Sheng, C., Xu, J., Wu, L., 2017. A review of wetland remote sensing. Sensors 17 (4), 777.
Guo, Q., Pu, R., Li, J., Cheng, J., 2017. A weighted normalized difference water index for water extraction using landsat imagery. Int. J. Remote Sens. 38 https://doi.org/10.1080/01431161.2017.1341667.
Helber, P., Bischke, B., Dengel, A., Borth, D. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2019;12. doi:10.1109/JSTARS.2019.2918242.
Hinton, G., LeCun, Y., Bengio, Y., 2015. Deep learning. Nature.
Hu, F., Xia, G.S., Wang, Z., Huang, X., Zhang, L., Sun, H. Unsupervised Feature Learning Via Spectral Clustering of Multidimensional Patches for Remotely Sensed Scene Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2015;8. doi:10.1109/JSTARS.2015.2444405.
Huang, C., Davis, L.S., Townshend, J.R. An assessment of support vector machines for land cover classification. International Journal of Remote Sensing 2002;23. doi:10.1080/01431160110040323.
Huang, W., DeVries, B., Huang, C., Lang, M.W., Jones, J.W., Creed, I.F., Carroll, M.L., 2018. Automated extraction of surface water extent from Sentinel-1 data. Remote Sensing 10. https://doi.org/10.3390/rs10050797.
Iizuka, K., Itoh, M., Shiodera, S., Matsubara, T., Dohar, M., Watanabe, K. Advantages of unmanned aerial vehicle (UAV) photogrammetry for landscape analysis compared with satellite data: A case study of postmining sites in Indonesia. Cogent Geoscience 2018;4. doi:10.1080/23312041.2018.1498180.
Isikdogan, F., Bovik, A.C., Passalacqua, P. Surface water mapping by deep learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2017;10. doi:10.1109/JSTARS.2017.2735443.
Kataoka, T., Kaneko, T., Okamoto, H., Hata, S. Crop growth estimation system using machine vision. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM. volume 2; 2003. doi:10.1109/AIM.2003.1225492.
Ko, B.C., Kim, H.H., Nam, J.Y. Classification of potential water bodies using landsat 8 OLI and a combination of two boosted random forest classifiers. Sensors (Switzerland) 2015;15. doi:10.3390/s150613763.
Komarkova, J., Cermakova, I., Sedlak, P., Jech, J., 2019. Spectral Enhancement of Imagery for Small Inland Water Bodies Monitoring: Utilization of UAV-Based Data. Journal of Information Systems Engineering & Management 4. https://doi.org/10.29333/jisem/6346.
Kordelas, G.A., Manakos, I., Lefebvre, G., Poulin, B., 2019. Automatic inundation mapping using sentinel-2 data applicable to both camargue and do'ana biosphere reserves. Remote Sensing 11. https://doi.org/10.3390/rs11192251.
Kyriou, A., Nikolakopoulos, K. Flood mapping from Sentinel-1 and Landsat-8 data: a case study from river Evros, Greece. In: Earth Resources and Environmental Remote Sensing/GIS Applications VI. volume 9644; 2015. doi:10.1117/12.2194449.
Lefebvre, G., Davranche, A., Willm, L., Campagna, J., Redmond, L., Merle, C., Guelmami, A., Poulin, B., 2019. Introducing WIW for detecting the presence of water in wetlands with Landsat and Sentinel satellites. Remote Sensing 11. https://doi.org/10.3390/rs11192210.
Li, Y., Shi, T., Zhang, Y., Chen, W., Wang, Z., Li, H. Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation. ISPRS Journal of Photogrammetry and Remote Sensing 2021;175. doi:10.1016/j.isprsjprs.2021.02.009.
Lin, H., Shi, Z., Zou, Z. Fully Convolutional Network with Task Partitioning for Inshore Ship Detection in Optical Remote Sensing Images. IEEE Geoscience and Remote Sensing Letters 2017;14. doi:10.1109/LGRS.2017.2727515.
Louhaichi, M., Borman, M.M., Johnson, D.E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. Geocarto International 2001;16. doi:10.1080/10106040108542184.
Luo, W., Li, H., Liu, G., Zeng, L. Semantic annotation of satellite images using author-genre-topic model. IEEE Transactions on Geoscience and Remote Sensing 2014;52. doi:10.1109/TGRS.2013.2250978.
Mahdavi, S., Salehi, B., Granger, J., Amani, M., Brisco, B., Huang, W., 2018. Remote sensing for wetland classification: A comprehensive review. GIScience & Remote Sensing 55 (5), 623–658.
Mahdavi, S., Salehi, B., Granger, J., Amani, M., Brisco, B., Huang, W. Remote sensing for wetland classification: a comprehensive review. GIScience and Remote Sensing 2018b;55. doi:10.1080/15481603.2017.1419602.
Manakos, I., Kordelas, G.A., Marini, K. Fusion of Sentinel-1 data with Sentinel-2 products to overcome non-favourable atmospheric conditions for the delineation of inundation maps. European Journal of Remote Sensing 2020;53. doi:10.1080/22797254.2019.1596757.
Markert, K.N., Chishtie, F., Anderson, E.R., Saah, D., Griffin, R.E. On the merging of optical and SAR satellite imagery for surface water mapping applications. Results in Physics 2018;9. doi:10.1016/j.rinp.2018.02.054.
Marti-Cardona, B., Dolz-Ripolles, J., Lopez-Martinez, C. Wetland inundation monitoring by the synergistic use of ENVISAT/ASAR imagery and ancilliary spatial data. Remote Sensing of Environment 2013;139. doi:10.1016/j.rse.2013.07.028.
Martinis, S., Plank, S., Cwik, K., 2018. The use of Sentinel-1 time-series data to improve flood monitoring in arid areas. Remote Sensing 10. https://doi.org/10.3390/rs10040583.
McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing 1996;17. doi:10.1080/01431169608948714.
Meyer, G.E., Hindman, T.W., Laksmi, K. <title>machine vision detection parameters for plant species identification</title>. In: Precision Agriculture and Biological Quality. volume 3543; 1999. doi:10.1117/12.336896.
Miao, Z., Fu, K., Sun, H., Sun, X., Yan, M. Automatic Water-Body Segmentation from High-Resolution Satellite Images via Deep Networks. IEEE Geoscience and Remote Sensing Letters 2018;15. doi:10.1109/LGRS.2018.2794545.
Mishra, N.B., 2020. Wetlands: remote sensing. In: Wetlands and Habitats. CRC Press, pp. 201–212.
Noh, H., Hong, S., Han, B., 2015. Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. https://doi.org/10.1109/ICCV.2015.178.
Pahlevan, N., Sarkar, S., Franz, B.A., Balasubramanian, S.V., He, J. Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: Demonstrations and validations. Remote Sensing of Environment 2017;201. doi:10.1016/j.rse.2017.08.033.
Pásler, M., Komárková, J., Sedlák, P., 2015. Comparison of possibilities of UAV and Landsat in observation of small inland water bodies. In: International Conference on Information Society, i-Society 2015. https://doi.org/10.1109/i-Society.2015.7366855.
Puliti, S., Saarela, S., Gobakken, T., Ståhl, G., Næsset, E. Combining UAV and Sentinel-2 auxiliary data for forest growing stock volume estimation through hierarchical model-based inference. Remote Sensing of Environment 2018;204. doi:10.1016/j.rse.2017.10.007.
Sadeghi, M., Babaeian, E., Tuller, M., Jones, S.B. The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. Remote Sensing of Environment 2017;198. doi:10.1016/j.rse.2017.05.041.
Salamí, E., Barrado, C., Pastor, E., 2014. UAV flight experiments applied to the remote sensing of vegetated areas. Remote Sensing 6. https://doi.org/10.3390/rs61111051.
Sanga-Ngoie, K., Iizuka, K., Kobayashi, S., 2012. Estimating CO2 sequestration by forests in oita prefecture, Japan, by combining LANDSAT ETM+ and ALOS satellite remote sensing data. Remote Sensing 4. https://doi.org/10.3390/rs4113544.
Schumann, G., Baldassarre, G.D., Bates, P.D. The utility of spaceborne radar to render flood inundation maps based on multialgorithm ensembles. IEEE Transactions on Geoscience and Remote Sensing 2009;47. doi:10.1109/TGRS.2009.2017937.
Shen, X., Wang, D., Mao, K., Anagnostou, E., Hong, Y., 2019. Inundation extent mapping by synthetic aperture radar: A review. Remote Sensing 11. https://doi.org/10.3390/RS11070879.
Tu, T.M., Huang, P.S., Hung, C.L., Chang, C.P. A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery. IEEE Geoscience and Remote Sensing Letters 2004;1. doi:10.1109/LGRS.2004.834804.
Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 1979;8. doi:10.1016/0034-4257(79)90013-0.
Töyrä, J., Pietroniro, A. Towards operational monitoring of a northern wetland using geomatics-based techniques. Remote Sensing of Environment 2005;97. doi:10.1016/j.rse.2005.03.012.
Ul-Hasan, A., Shafaity, F., Liwicki, M. Curriculum learning for printed text line recognition of ligature-based scripts. In: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR. 2015. doi:10.1109/ICDAR.2015.7333912.
Veloso, A., Mermoz, S., Bouvet, A., Toan, T.L., Planells, M., Dejoux, J.F., Ceschia, E. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. Remote Sensing of Environment 2017;199. doi:10.1016/j.rse.2017.07.015.
Vries, B.D., Huang, C., Lang, M.W., Jones, J.W., Huang, W., Creed, I.F., Carroll, M.L., 2017. Automated quantification of surface water inundation in wetlands using optical satellite imagery. Remote Sensing 9. https://doi.org/10.3390/rs9080807.
Wang, Z., Liu, J., Li, J., Zhang, D.D. Multi-SpectralWater Index (MuWI): A Native 10-m Multi-SpectralWater Index for accuratewater mapping on sentinel-2. Remote Sensing 2018;10. doi:10.3390/rs10101643.
Wei, X., Guo, Y., Gao, X., Yan, M., Sun, X., 2017. A new semantic segmentation model for remote sensing images. In: International Geoscience and Remote Sensing Symposium (IGARSS). https://doi.org/10.1109/IGARSS.2017.8127319.
Woebbecke, D.M., Meyer, G.E., Bargen, K.V., Mortensen, D.A. Color indices for weed identification under various soil, residue, and lighting conditions. Transactions of the American Society of Agricultural Engineers 1995;38. doi:10.13031/2013.27838.
Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing 2006;27. doi:10.1080/01431160600589179.
Yagmur, N., Musaoglu, N., Taskin, G. Detection of shallow water area with machine learning algorithms. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. volume 42; 2019. doi:10.5194/isprs-archives-XLII-2-W13-1269-2019.

Yang, Y., Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. In: GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems. https://doi.org/10.1145/1869790.1869829.

Zhang, F., Li, J., Zhang, B., Shen, Q., Ye, H., Wang, S., Lu, Z., 2018. A simple automated dynamic threshold extraction method for the classification of large water bodies from landsat-8 OLI water index images. Int. J. Remote Sens. 39 https://doi.org/10.1080/01431161.2018.1444292.

Zhu, R., Yan, L., Mo, N., Liu, Y. Semi-supervised center-based discriminative adversarial learning for cross-domain scene-level land-cover classification of aerial images. ISPRS Journal of Photogrammetry and Remote Sensing 2019;155. doi:10.1016/j.isprsjprs.2019.07.001.