INTERNATIONAL JOURNAL OF REMOTE SENSING
2023, VOL. 44, NO. 14, 4417–4440
https://doi.org/10.1080/01431161.2023.2235643

Taylor & Francis
Taylor & Francis Group

Perturbed peer model with joint confidence for semi-supervised object-based classification in urban area

Wenye Wang, Xueliang Zhang, Pengfeng Xiao and Qi Su

Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing, Jiangsu, China

ABSTRACT

The application of deep learning (DL) improves the accuracy of object-based classification in urban area, but the huge numbers of labelled samples required for training DL models are difficult to obtain. To address that, we propose a novel semi-supervised object-based classification method that combines pseudo-labelling method and consistency regularization method, named as perturbed peer model (PPM). To evaluate the quality of unlabelled object samples, a new object-level joint confidence is designed to assess the confidence of the samples' prediction, which provides guidance for selecting unlabelled object samples in semi-supervised object-based classification. Instead of only using the high-confidence samples as usual, both the high-confidence samples with discriminative power and the low-confidence samples with a large range of structural features are exploited using pseudo-labelling method and consistency regularization method, respectively, facilitating the fusion of the two methods in the proposed PPM. In addition, two types of perturbations, dropout and difference augmentation, are integrated into the model to drive the consistency regularization method as well as to enhance the difference to facilitate the pseudo-labelling method. Experimental results show that the classification accuracy of PPM is better than the widely-used self-training, co-training, noisy-student, unsupervised data augmentation for consistency training, and FreeMatch methods in urban area.

ARTICLE HISTORY
Received 2 April 2023
Accepted 2 July 2023

KEYWORDS
Object-based classification; semi-supervised learning; convolutional neural networks; high-spatial resolution remote sensing; urban area

1. Introduction

Geographic Object-Based Image Analysis (GEOBIA) paradigm (Blaschke et al. 2014) has played an important role in urban mapping (Georganos et al. 2018; Hamedianfar and Shafri 2015; Momeni, Aplin, and Boyd 2016), change detection (Chen et al. 2012; Tan et al. 2019), vegetation monitoring (Chabot et al. 2018; Whyte, Ferentinos, and Petropoulos 2018), and agricultural monitoring (Ozelkan, Chen, and Ustundag 2016; Torres-Sánchez, López-Granados, and Pena 2015). Especially after meeting Deep Learning (DL), the use of the this technique eliminates the feature selection and de-redundancy operations,

CONTACT Xueliang Zhang zxl@nju.edu.cn Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing, Jiangsu 210023, China

making the whole process of GEOBIA easier and more accurate, and thus bringing it to a new stage (Chen et al. 2018; Kucharczyk et al. 2020).

GEOBIA performs image segmentation at first to obtain objects, and then the features obtained from the object (e.g. spectral, shape, texture) are used for classification (Blaschke 2010). However, these features may add up to nearly a hundred. Excessive features may actually lead to a decrease in classification accuracy, therefore requiring feature selection and de-redundancy operations (Sun et al. 2020, 2021). Moreover, these features have limited representation of semantic information (Zhang et al. 2018). DL eliminates the complexity of these operations and makes GEOBIA easier to carry out, with its automatic feature and semantic information extraction (Janiesch, Zschech, and Heinrich 2021). Thanks to its spatial information extraction ability (Rousset et al. 2021; Wang et al. 2021), convolutional neural networks (CNNs) have been proven effective for a variety of satellite image analysis tasks, including classification (Tang et al. 2021; Zhao, Du, and Emery 2017), segmentation (Gao et al. 2021; Zheng et al. 2021) and change detection (Shi et al. 2021). Good progress and high accuracy have also been achieved using CNNs in the GEOBIA framework, which is valuable in a variety of practical applications (Hossain and Chen 2019). However, CNNs are generally data hungry (Mahajan et al. 2018), and the labelling of remote sensing images is time-consuming and laborious. The semi-supervised methods, which utilize unlabelled samples to learn features, could reduce the need for labelled samples.

Semi-supervised methods can take advantage of massive unlabelled object samples obtained by image segmentation. In general, slight over segmentation is usually performed during segmentation for better object-based classification accuracy (Fu et al. 2018), which results in a relatively large number of small-size object samples. Although these samples usually contain a single class by maintaining homogeneity within the samples, it becomes difficult for semi-supervised methods to utilize them. This is because the features contained in a small object are sparse and partial, making it challenging to assess the quality of an object sample. Furthermore, it is tough to select suitable high-quality samples from the large numbers of unlabelled samples while avoiding those of low-quality and redundancy. Therefore, we pay attention to designing object-level joint confidence to perform quality assessment of object samples to cope with this problem.

There are currently four common paradigms for semi-supervised methods: pseudo-labelling methods (Qiao et al. 2018), consistency regularization methods (Sajjadi, Javanmardi, and Tasdizen 2016), generative methods (Springenberg 2015), and graph-based methods (Kipf and Welling 2017). The first two can be well applied to the GEOBIA framework because of the reasons as followed. Pseudo-labelling methods learn the features of unlabelled samples by assigning the high-confidence samples with pseudo labels (Zhai et al. 2019), which can help to select high-confidence samples among large numbers of object samples. Consistency regularization methods learn structural features of unlabelled samples by adding perturbations between teacher and student models while restricting the output of both models to be the same (Tarvainen and Valpola 2017), which can conveniently perform structural features learning of massive object samples. Generative methods learn implicit features of the data by generating new data with similar distribution to the training data. However, for large numbers of object samples, generative methods can cause huge computational expenses and inefficiencies. The graph-based methods are also difficult

to apply because it is difficult to reconstruct the entire sample distribution by finding samples with similar features in the feature space due to the sparse features of the object samples.

Although both the pseudo-labelling methods and the consistency regularization methods can be utilized to deal with a large number of unlabelled object samples, they act in different ways. The pseudo-labelling methods reinforce the features of samples in high-density regions in the sample space, but ignore the features of samples in low-density regions near the decision boundary. This causes the model to always reinforce the features already learned and prevents the model from generalizing to the entire feature space (Zhang and Rudnicky 2006). Especially, when the initial sample has certain errors, such errors may be reinforced and amplified (Wu et al. 2020). Consistency regularization, on the other hand, learns the features of the entire samples without discrimination, treating high-density and low-density regions equally. Without focusing on the most discriminative high-density regions of the sample, it does not strengthen the core features needed for the model.

An ideal semi-supervised algorithm should explore the entire feature space to obtain better accuracy globally (Wu, Li, and Wang 2018). Hence, the hybrid method of pseudo-labelling methods and consistency regularization methods have been applied to remote sensing image analysis. Wang et al. (2020) concatenated the two methods by performing consistency regularization training followed by pseudo-labelling self-training. In the framework of mean-teacher, the output between the student and the teacher models is constrained by pseudo labels obtained by randomly pasting part of the ground truth label into the predicted segmentation map (Wang et al. 2022). In the generative adversarial network framework, the model is optimized by pseudo labels generation and consistent self-training to learn the distribution of labelled and unlabelled data (Li et al. 2021). However, on the one hand, these hybrid methods are not designed for GEOBIA and cannot be used directly in the GEOBIA framework. On the other hand, among these semi-supervised methods, the pseudo-labelling methods and the consistency regularization methods are not fused according to each other's complementary characteristics and thus cannot effectively take advantage of each other.

We propose the perturbed peer model (PPM) for semi-supervised object-based classification. The following technical contributions are achieved in this study.

(1) We propose object-level joint confidence to evaluate the quality of unlabelled object samples, which provides guidance for selecting unlabelled object samples in semi-supervised object-based classification.
(2) The pseudo-labelling method and the consistency regularization method respectively utilize the high-confidence samples with discriminative power and the low-confidence samples with a large range of structural features, facilitating the fusion of the two methods in the proposed PPM.
(3) In addition, two types of perturbations are integrated into PPM to enhance both the pseudo-labelling method and the consistency regularization method.
(4) The proposed PPM model is demonstrated with better object-based classification performance than the widely-used semi-supervised methods of self-training, co-training, noisy-student, and consistency training in urban area.

2. Study area and data

2.1. Study area

The study areas include two cities: Mr Andreessen (S1) and Nanjing (S2), which are located in the west coast of South Korea and the lower reaches of the Yangtze River in eastern China, respectively, as shown in Figure 1.

Mr Andreessen is an emerging idyllic industrial city, one of the top 10 cities in Korea, with a temperate monsoon climate. The city's green area accounts for 55% of the total area and has the highest greening rate in South Korea. Meanwhile, Mr Andreessen is a seaside city with a rich waterfront system. Nanjing, the seventh largest city in China, is located in a subtropical monsoon climate zone with a rich water system. Nanjing is a National Ecological Garden City in China and the city with the most forest coverage in Jiangsu Province.

Both study areas are rich in surface features and involve both urban and suburban areas that are highly heterogeneous and distinctive from each other in land cover characteristics, which are therefore suitable for testing the generalization capability of the proposed PPM.

2.2. Data

The data of S1 come from the SuperView-1 satellite. The image contains 18,249 × 14654 pixels, and has four multispectral bands (red, green, blue, and near infrared) with a spatial resolution of 50 cm. Nine land cover categories are found in S1, including building, impervious surface, bare land, farmland, woodland, tidal flat, water, and shadow. The data of S2 come from the GF-2 satellite. The image has four multispectral bands (red,

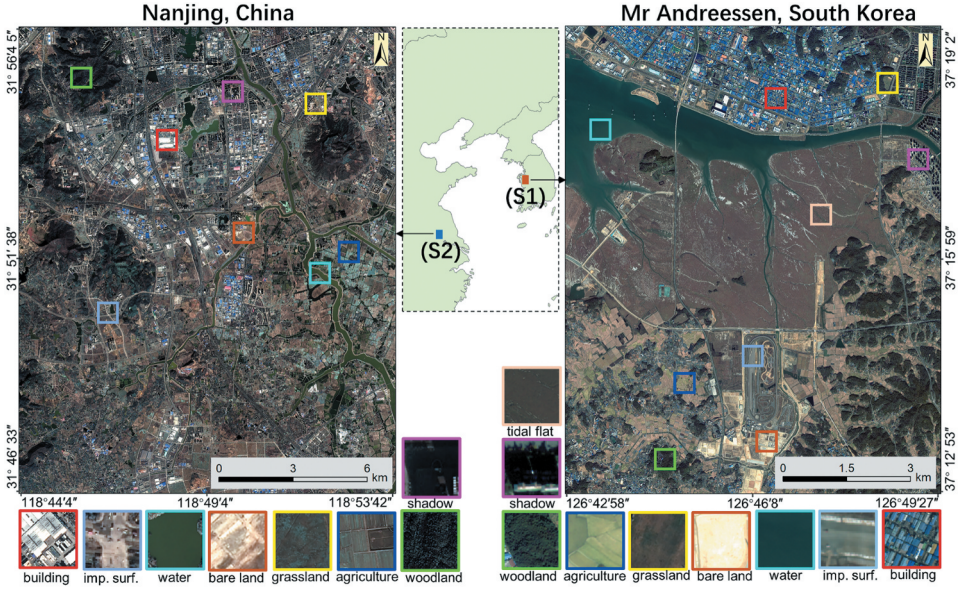

Figure 1. The two study areas: S1 (Mr Andreessen) and S2 (Nanjing) with highlighted regions representing the majority of land cover categories.

Table 1. Numbers of labelled samples per class at the two study sites.

Class (Site S1)	Training sample pool	Test samples	Class (Site S2)	Training sample pool	Test samples
building	50	404	building	50	273
impervious surface	50	335	impervious surface	50	272
grassland	50	338	grassland	50	282
bare land	50	257	bare land	50	271
farmland	50	292	farmland	50	278
woodland	50	262	woodland	50	265
water	50	266	water	50	276
shadow	50	263	shadow	50	281
tidal flat	50	310			

green, blue, and near infrared) with a spatial resolution of 80 cm and with a spatial extent of 19,862 × 15949 pixels. Eight dominant land cover categories are identified within S2, including building, impervious surface, bare land, farmland, woodland, water, and shadow.

We use the multiresolution segmentation (MRS) method for image segmentation (Benz et al. 2004). The segmentation scale parameter is set as 60 by using trial-and-error method to achieve a slight over-segmentation.

After performing the MRS method, a large number of irregularly shaped objects are obtained, with 354,632 for S1 and 419,787 for S2. Then, the barycentre of each object is calculated. The barycentre refers to the object's centre of mass. Specifically, the x-coordinate of the barycentre is calculated as the mean x-coordinate of all the pixels in the object, and the y-coordinate of the barycentre refers to the mean y-coordinate of all the pixels in the object. With this barycentre as the centre, the first patch of 36 × 36 pixels is generated, and the second patch of 48 × 48 pixels is generated and resized to 36 × 36 pixels. The second patch is used as auxiliary data to expand the multi-scale information of the object sample and is used for the subsequent object-level joint confidence calculation. These patches will be used as unlabelled samples.

The labelled samples are obtained by manually selecting the 36 × 36 patches in the image, and the number of labelled samples in two sites are presented in Table 1. Among the labelled samples, 50 samples of each class are randomly selected as the training sample pool, and the remaining ones were used as the test samples. During training, 10, 20, or more samples are randomly selected from the training sample pool, depending on the experimental design.

3. Method

3.1. Overview

The proposed method is a two-channel model, as shown in Figure 2. Two CNNs with different network structures, called peer models, are the targets of the training optimization. The labelled and unlabelled samples are sent together to the peer models for training. For the labelled samples, cross entropy is used for optimization:

$$L_{ce} = -\sum_{i,k} y^{l}_{i,k} \log p^{l}_{i,k} \tag{1}$$

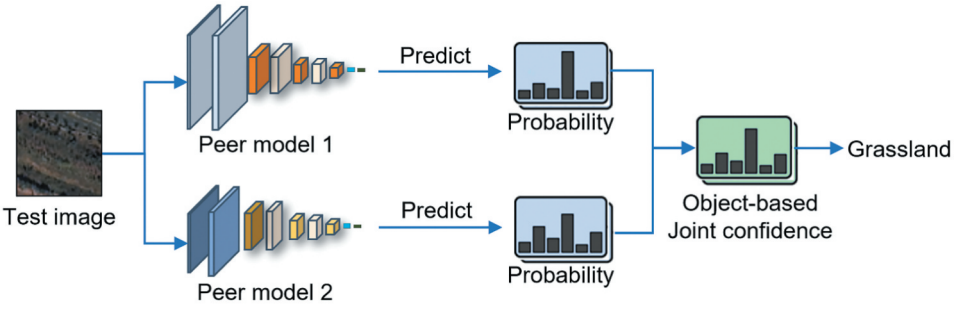

Figure 2. Flowchart of prediction process of the proposed method.

Figure 3. Flowchart of obtaining object patches.

where $y^{l}_{i,k}$ represents the ground truth of the k-th category of the *i*th labelled sample, $p^{l}_{i,k}$ represents the probability that the *i*-th labelled sample is predicted to be in category k, and the upper index l indicates that this parameter is for labelled samples.

For the unlabelled samples, the number of samples obtained through the segmentation algorithm is enormous. Therefore, a random sampling approach is employed to select a portion for utilization. Then, the peer models output their classification probabilities and generate object-level joint confidence. The high-confidence samples are assigned as pseudo labels, while the low-confidence samples are optimized with consistency regularization. For these high-confidence pseudo labelled samples, the model optimization in subsequent iterations will utilize the L_{pl}, specifically cross entropy between prediction and pseudo labels. As for low-confidence samples, the model optimization will directly employ the L_{cr}, namely consistency regularization loss, which is calculated using mean square error. In addition, two perturbations, dropout and different augmentation (diff-aug), are introduced in the model to improve the model accuracy and robustness. Diff-aug is an operation of differential enhancement for pseudo labelled samples. By using different augmentation methods, two augmented versions of the pseudo labelled samples are constructed and then added to labelled dataset 1 and labelled dataset 2, respectively. These two sets of samples will be used to train two peer models. Labelled dataset 1 and labelled dataset 2 are sampled from the labelled dataset, which are independent with each other. Initially, the samples inside labelled dataset 1 and labelled dataset 2 are the same, but after adding pseudo labelled samples using diff-aug, they become different. The PPM is iteratively trained and the training process stops after 10 iterations at default.

The prediction process of PPM is shown in Figure 3. The peer models output the category probabilities of the test image and generate object-level joint confidence, and the final prediction category is determined as that with the highest joint confidence.

3.2. Object-level joint confidence

It is hoped that the correct pseudo labels can be assigned to unlabelled samples so that they are used by the model for improving semi-supervised training. However, the wrong choice of samples always occurs in the sample selection process of the pseudo-labelling methods. These errors will be propagated to the downstream classification tasks, lowering the final classification accuracy. This drawback is also known as the problem of confirmation bias in pseudo-labelling methods (Arazo et al. 2020). Therefore, it is crucial to assess the quality of the unlabelled object samples and to give them the correct pseudo labels.

Usually, the probability of model prediction is directly used as a confidence measure for assigning pseudo labels (Li et al. 2021). However, using probability as confidence causes the model to become confident in predicting the easier classes, which usually dominates the training process (Han et al. 2018). In addition, the model may be too confident in misclassifying samples to avoid introducing too many incorrect pseudo labels. To overcome the above problems, we propose object-level joint confidence by using the probability of the peer models. It uses multi-scale information of the object samples and combines the output between peer models to make it more reliable.

Object samples have both within-object information and between-object information (Zhang et al. 2018). The within-object information can be obtained by the CNN's ability to automatically extract features. In addition, we obtain the object's context information by adding a larger-size auxiliary patch to the original patches, as shown in Figure 4.

At first, one-tenth of all unlabelled samples are randomly selected to generate $\mathbf{u}$ in each training iteration, where $\mathbf{u} = \{u_i\}_{i=1}^{l}$ and u_i refers to the i-th sample in $\mathbf{u}$.

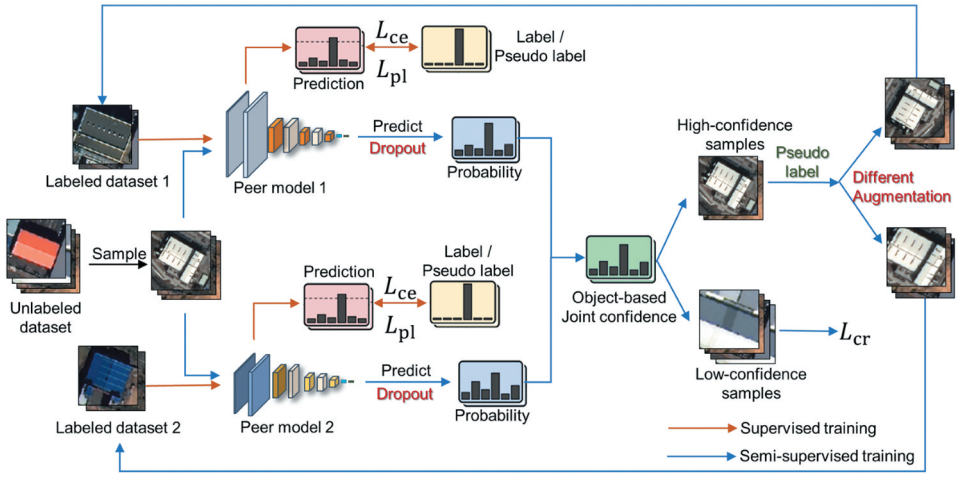

Figure 4. Flowchart of training process of the proposed method. Orange arrows indicate supervised training, blue arrows indicate semi-supervised training, and black arrow indicates sample process. Labeled dataset 1 and labeled dataset 2 are sampled from the labeled dataset. L_{ce} denotes cross entropy between prediction and ground truth labels. L_{pl} denotes cross entropy between prediction and pseudo labels. L_{cr} denotes the consistency regularization loss of the unlabeled samples, which is calculated using the mean square error.

We input $\mathbf{u}$ to peer model $M^j (j \in \{1,2\})$, where M^j refers to the peer model 1 or peer model 2. The output vectors of softmax layers from M^j is $\mathbf{F}^j = \left\{p_i^j\right\}_{i=1}^{l}, j \in \{1,2\}$. $p_i^j = \left\{p_{i,1}^j, p_{i,2}^j, \ldots, p_{i,\mathrm{k}}^j, \ldots, p_{i,\mathrm{K}}^j\right\}$, where $p_{i,\mathrm{k}}^j$ represents the probability that unlabelled sample u_i belongs to class k predicted by model M^j, where k$\in$ {1, … , K}, and K refers to the total number of classes.

Each sample corresponds to two patches, one with the original size and the other with the auxiliary larger size that provides context information. Both patches are sent to M^j and get the corresponding $\mathbf{F}^j{}_{\text{origin}}$ and $\mathbf{F}^j{}_{\text{auxiliary}}$.

Thus, the final predicted probability of a sample is:

$$\mathbf{F}^j = \alpha\, \mathbf{F}^j{}_{\text{origin}} + (1-\alpha)\mathbf{F}^j{}_{\text{auxiliary}} \tag{2}$$

where $\alpha = 0.7$ by default.

Using $\mathbf{F}^1$ and $\mathbf{F}^2$ to calculate the object-level joint confidence, the formula is as follows:

$$\mathbf{C} = \frac{\mathbf{F}^1 \times \mathbf{F}^2}{\mathbf{F}^1 + \mathbf{F}^2} \tag{3}$$

where $\mathbf{C} = \{c_i\}_{i=1}^{l}$, $c_i = \left\{c_{i,1}, c_{i,2}, \ldots, c_{i,\mathrm{k}}, \ldots, c_{i,\mathrm{K}}\right\}$, and $c_{i,\mathrm{k}}$ represents the object-level joint confidence that unlabelled sample u_i belongs to class k. It is noted that we use the harmonic mean to calculate the object-level joint confidence, as in Equation (3). The harmonic mean is more sensitive to small values compared to the arithmetic mean and geometric mean. Therefore, we chose it to ensure that a high value of object-based joint confidence corresponds to both high prediction values ($\mathbf{F}^1$ and $\mathbf{F}^2$) from the peer models.

Subsequently, the set of object-level joint confidence of class k is ranked, and the top x samples are considered as high-confidence samples of that class and are labelled with the corresponding pseudo labels. x is set to 10 in the experiment. After the high-confidence samples of each class are selected, the remaining samples are considered as low-confidence samples, and these samples are optimized by consistency regularization method.

3.3. Differential treatment of high-confidence and low-confidence samples

The unlabelled samples are divided into high-confidence and low-confidence samples according to the object-level joint confidence and are utilized by applying the pseudo-labelling method and the consistency regularization method, respectively. These high-confidence samples contain more discriminative features, thus they are given the pseudo labels and treated with the same weight as the labelled samples to better learn the crucial features. As for the low-confidence samples, although their features lack discriminative power, the structural features in them can help the model to restore the whole feature space and improve the robustness, thus the consistency regularization method is used to learn their structural features. Consistency regularization method learns structural features of unlabelled samples by adding perturbations between two models while restricting the output of both models to be the same. The proposed hybrid method is designed to fully exploit the value of both high-confidence and low-confidence samples by taking advantage of both semi-supervised methods.

The adopted pseudo-labelling method is an extension of co-training, which is the semi-supervised method based on divergence. Co-training is originally designed for two independent sufficient views which trained two classifiers separately and chose unlabelled samples for each other. It is usually driven by constructing different classifiers (e.g. different algorithms, different structures, or different parameters) (Wang and Zhou 2007). We used two CNNs with different structures to construct the initial difference. During training, we continue to maintain the differences between the classifiers by adding perturbations: dropout and diff-aug. Dropout is an operation in the fully connected layer of CNNs that randomly makes some parameters not participate in the calculation. Diff-aug makes the difference between the two sample sets larger by using different data augmentation methods. If the differences between the models are not maintained, the classifiers will converge during training due to mutual learning, which will lead to ineffective co-training (Qiao et al. 2018).

In co-training, two models select samples that they consider to be high-confidence for each other, enabling the exchange of information. However, the samples predicted by a single model may be not credible and may pass on the error to each other. Different from co-training, the proposed method does not select samples for each other. As shown in Figure 5, PPM uses the probabilities predicted by the peer models to calculate object-level joint confidence at first, and then selects the high-confidence samples together by the two peer models, which makes the selected pseudo labels more credible.

High-confidence samples are utilized by the pseudo-labelling method. They are augmented using the diff-aug strategy and then added to the labelled sample sets. The optimization of these pseudo labelled samples occurs in the next iteration, where they are treated as labelled samples and optimized using cross-entropy:

$$L_{\mathrm{pl}} = -\sum\nolimits_{i,\mathrm{k}} \widetilde{y^{\mathrm{u}}_{i,\mathrm{k}}} \log p^{\mathrm{u}}_{i,\mathrm{k}} \tag{4}$$

where $\widetilde{y^{\mathrm{u}}_{i,\mathrm{k}}}$ represents the pseudo labels of the k-th category of the *i*-th unlabelled sample, $p^{\mathrm{u}}_{i,\mathrm{k}}$ represents the probability that the *i*-th unlabelled sample is predicted to be in category k, and the upper index u indicates that this parameter is for unlabelled samples.

For unlabelled samples with low-confidence, consistency regularization is used for optimization. Consistency regularization believes that optimizing the relative entropy between predictions of unlabelled data with different perturbations can improve model performance. Accordingly, the two perturbations of dropout and diff-aug are used. By

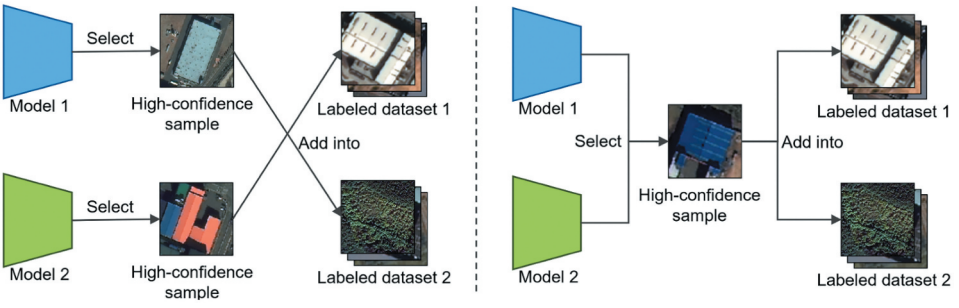

Figure 5. Two ways of selecting process of pseudo labeled samples: co-training (left) and PPM (right).

limiting the distance between peer model outputs, a consistency regularization constraint can be implemented to learn the structural features of unlabelled samples. The distance is measured using mean square error:

$$L_{cr} = MSE\left(\mathbf{F}^1, \mathbf{F}^2\right) = ||\mathbf{F}^1 - \mathbf{F}^2||^2 \tag{5}$$

In summary, the total loss function is as follows:

$$L = L_{ce} + \lambda L_{pl} + \beta L_{cr} \tag{6}$$

where λ and β are the weight parameters.

3.4. Perturbations

We use two types of perturbations, dropout and difference augmentation, which benefit the pseudo-labelling method and consistency regularization method at the same time, specifically by increasing the differences between the peer models to drive pseudo-labelling method and by adding perturbations to enable the consistency regularization method to be carried out.

Dropout is an operation in the fully connected layer of CNNs that randomly makes some parameters not participate in the calculation (Srivastava et al. 2014). Using dropout during training can reduce overfitting and improve the generalization ability of the model. Usually, the dropout of the CNNs is locked when predicting, so that it no longer functions and thus a stable output can be obtained. In this study, the CNNs no longer lock the dropout when making predictions. It continues to play a role, thus introducing perturbations. Specifically, in the prediction, if the dropout is equal to 0, all parameters of the fully connected layer are involved and the output is stable, no perturbation is introduced. If the dropout is equal to 0.1, 10% of the parameters are not involved in the prediction at this time, and this 10% of the parameters are random in the peer models, thus introducing perturbations between models. By limiting the distance between peer model outputs, a consistency regularization constraint can be implemented to learn the structural features of unlabelled samples.

Diff-aug was proposed to differentially augment the high-confidence samples. There is a pool of augmentation methods that includes flip, rotation, scaling, gaussian noise, salt, pepper noise, and gaussian filtering. After finding high-confidence samples and labelling them with pseudo labels, two different augmentation methods are randomly selected from the pool and used on these samples to form two pseudo labelled datasets. The two datasets are then used to train the two peer models, respectively. The diff-aug creates continuous input-type perturbations for peer models that are first reflected in the network parameters and then make the models better learn the features of the pseudo labelled samples under the consistency regularization constraint.

The two types of perturbations act differently on the pseudo-labelling method and the consistency regularization method. For the pseudo-labelling method, the perturbations increase the difference between the peer models, allowing difference-driven co-training to implement. For the consistent regularization method, structural features of unlabelled samples can be learned by restricting the output of peer models to be consistent under perturbations.

4. Experimental setup

4.1. Cnns for PPM

CNN models typically consist of three main types of layers: convolutional layer, pooling layer, and fully connected layer. The convolutional layer is used as a hierarchical feature extractor, the pooling layer performs spatial down-sampling of the feature maps, and the fully connected layer is used as a classifier to generate predicted classification probabilities of the input data.

In our experiments, we design two lightweight models to better apply to small sample size. By neural architecture search (Zoph and Le 2017), two models are tested as optimal and identified as peer models among a large set of artificially set models with different structures. Figure 6 shows the detailed structure of CNNs we employ. Both peer models are pre-trained in UCMerced_LandUse dataset (Yang and Newsam 2010).

4.2. Parameter settings

In the training process, we use stochastic gradient descent (SGD) with momentum of 0.9 to optimize the network. We set the batch size as 3000. The epoch is set to 5. A total of 10 iterations are conducted. The learning rate decays from 0.001 to 0.0001 with step 0.0001. The weight parameter λ for the pseudo-labelling method is set to 1 and the weight parameter β for the consistency regularization method is set to 0 in the first four iterations and continued to grow to 0.3 with step 0.05 in the following iterations. At the end of each iteration, 10 high-confidence unlabelled samples for each category are selected and assigned as pseudo labels to add into the labelled sample sets. All experiments run on an NVIDIA GTX3080 GPU.

4.3. Comparison with other semi-supervised methods

We compared 5 representative semi-supervised methods to demonstrate the effectiveness of the proposed method.

Self-training (Lee 2013) selects samples with high prediction probability and labels them with pseudo labels for utilization. The structure of the CNN used in self-training is set the same as peer model 1. The probabilities are ranked at the end of each iteration and the 10 high-confidence samples per category are selected to assign pseudo labels.

Co-training (Qiao et al. 2018) has two models selecting pseudo samples with high probability for each other for semi-supervised training. The structures of the two CNNs

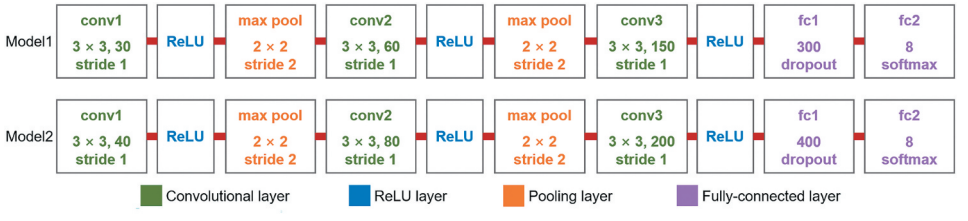

Figure 6. Architecture of two peer models.

used in co-training are set the same as peer model 1 and peer model 2. The approach of assigning pseudo labels is similar to self-training, where 10 high-confidence samples of each category are selected and added to each other's sample set.

Noisy-student (Xie et al. 2020) is an extension of the self-training, forcing the student model to learn more information by adding noise. It uses three types of noise, RandAugment data augmentation (Cubuk et al. 2019), dropout, and stochastic depth (Huang et al. 2016). In our experiments, we used the first two perturbations and the third one was not used due to the lightweight feature of the network structure. The structure of the CNN used in noisy-student is set the same as peer model 1.

Unsupervised data augmentation for consistency training (UDA) (Xie et al. 2020) is a consistency regularization method that adds perturbations by RandAugment data augmentation. The structure of the CNN used in UDA is set the same as peer model 1.

FreeMatch (Wang et al. 2023) is a pseudo-labelling method which can adjust the confidence threshold in an adaptive manner based on the learning state of the model. Further adaptive class fairness regularization penalties are introduced to encourage models to make diverse predictions in the early training phase. The structure of the CNN used in FreeMatch is set the same as peer model 1.

The training parameters of the above methods are the same as those of the proposed method in the training process for fair comparison.

5. Results

5.1. Object-based classification accuracies in the two study sites

The classification performance of the proposed method is investigated in both S1 and S2. Visual inspection and quantitative accuracy assessment, including overall accuracy (OA), Kappa coefficient (κ), and precision of every class, are adopted to evaluate the classification results.

5.1.1. Site S1

The PPM achieves the highest OA of 86.84% with a κ of 0.856, apparently higher than the co-training (82.64% OA and κ of 0.805), the noisy-student (81.17% OA and κ of 0.785), the UDA (83.85% OA and κ of 0.818), the FreeMatch (85.83% OA and κ of 0.842), and the self-training (79.28% OA and κ of 0.766), as shown in Table 2. The superiority of the proposed method is further demonstrated by the precision of every class. Three categories with simple characteristics, namely water, tidal flat, and shadow, have the best classification accuracy, which can reach 95.57%, 94.37%, and 93.11% by PPM, respectively. In contrast, buildings with complex features and impervious surfaces are not as well classified, with accuracies of 74.23% and 80.41%, respectively. The classification accuracies of the three types of vegetation are also moderate, with a classification accuracy of 88.33% for grassland, 80.44% for farmland, and 85.99% for woodland.

To provide a visualization, three subsets from S1 classified by different semi-supervised methods are presented in Figure 7. The proposed method has better performance in the extraction of urban buildings and impervious surface. In addition, linearly shaped roads and channels are identified with high geometric fidelity and coherence. In Figure 7(a), PPM does a good job of extracting rivers and roads, rather than misclassifying them into

Table 2. Classification accuracy comparison among self-training, co-training, noisy-student, UDA, FreeMatch, and PPM for S1 using precision of each class (%), overall accuracy (OA, %), and Kappa coefficient (κ).

Class	Self-training	Co-training	Noisy-student	UDA	FreeMatch	PPM
Building	59.80	65.55	57.01	66.06	72.12	74.23
Impervious surface	74.30	74.27	75.15	72.56	77.58	80.41
Water	92.07	91.36	94.14	92.43	94.87	95.57
Bare land	81.25	83.15	86.04	90.53	**90.79**	89.54
Grassland	87.06	86.84	86.70	84.33	87.54	88.33
Farmland	68.90	77.46	75.50	78.80	78.95	80.44
Woodland	83.80	84.71	82.66	83.74	84.66	85.99
Tidal flat	84.02	90.33	88.42	91.20	93.36	94.37
Shadow	85.55	92.67	83.41	**93.65**	93.47	93.11
Overall accuracy (OA)	79.28	82.64	81.17	83.85	85.83	86.84
Kappa coefficient (κ)	0.766	0.805	0.785	0.818	0.842	0.856

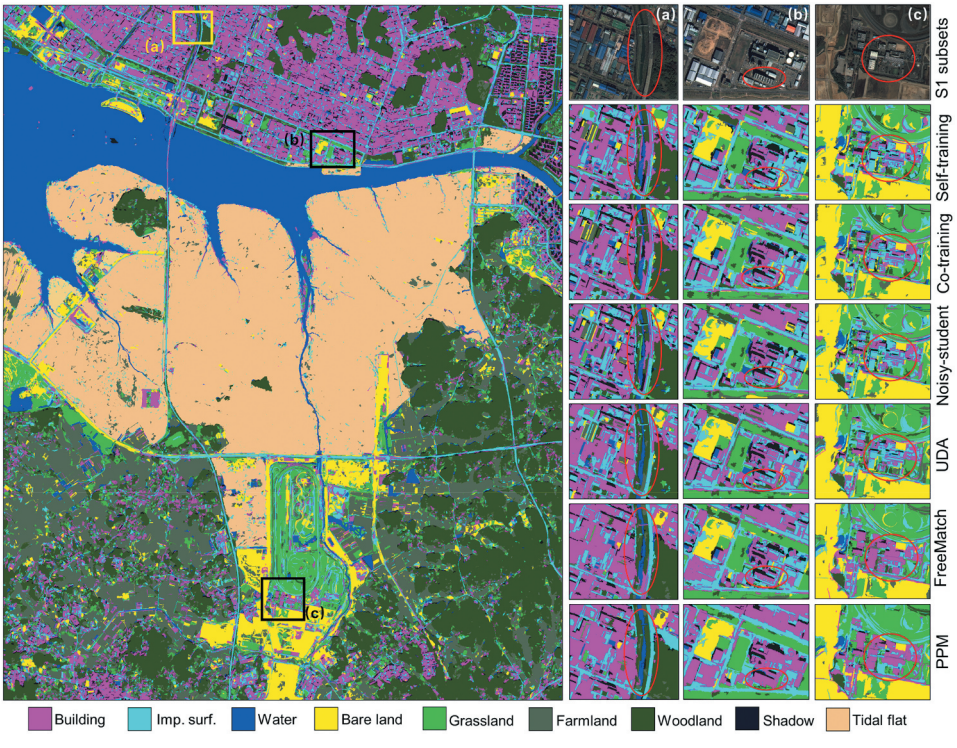

Figure 7. Classification result predicted by PPM in S1, as well as three subsets and their classification results by self-training, co-training, noisy-student, UDA, FreeMatch, and PPM. The positions of the three subsets are marked in the whole classification map.

woodlands or buildings. In Figure 7(b), buildings with special textures are easily classified as a mixture strips of buildings and impervious surface, as can be seen in UDA and noisy-student, while PPM achieves a more complete and accurate extraction of the buildings in the factory area. In Figure 7(c), PPM also has a good extraction effect for relatively small buildings.

Table 3. Classification accuracy comparison among self-training, co-training, noisy-student, UDA, FreeMacth, and PPM for S2 using precision of each class (%), overall accuracy (OA, %), and Kappa coefficient (κ).

Class	Self-training	Co-training	Noisy-student	UDA	FreeMatch	PPM
Building	61.72	64.55	62.40	67.46	72.32	75.86
Impervious surface	64.91	69.17	67.30	67.04	72.89	76.23
Water	91.20	93.14	92.07	93.20	93.08	93.38
Bare land	72.40	76.03	77.50	76.90	78.25	80.90
Grassland	81.12	84.82	83.71	82.23	85.57	88.31
Farmland	76.64	76.05	77.20	77.90	**78.03**	77.75
Woodland	87.10	88.80	89.75	90.36	90.75	91.24
Shadow	90.32	92.43	90.97	91.55	93.56	93.93
Overall accuracy (OA)	77.65	79.82	79.41	80.34	82.87	84.36
Kappa coefficient (κ)	0.754	0.776	0.771	0.782	0.811	0.834

5.1.2. Site S2

The proposed method also achieves the highest classification accuracies in S2, as illustrated in Table 3. It can be seen that PPM obtains the highest OA of 84.36% with a κ of 0.834, apparently higher than the co-training (OA of 79.82% and κ of 0.776), the noisy-student (OA of 79.41% with κ of 0.771), the UDA (OA of 80.34% and κ of 0.782), the FreeMatch (OA of 82.87% and κ of 0.811), and the self-training (OA of 77.65% with κ of 0.754). The effectiveness of the PPM is also demonstrated by the precision of each class.

The visualized classification results in S2 are presented in Figure 8. In Figure 8(a), open parking lots with densely parked cars are easily misclassified as buildings, and PPM reduces this misclassification to a small amount with its excellent feature learning capabilities. In Figure 8(b), the fallow farmland is not distinctly characterized and has unclear boundaries with the surrounding grassland, making classification difficult. PPM has better extraction performance compared to other methods, and rarely misclassifies it as grassland or bare land. In Figure 8(c), it can be seen that PPM is more capable of distinguishing between bare ground and grassland compared to other methods.

5.2. Influence of the number of labelled and unlabelled samples on accuracy

To further demonstrate the advantages of PPM, we carried out multiple groups of experiments with labelled sample numbers of 10, 20, 30, 40, and 50, taking S1 as an example. As shown in Figure 9, the proposed method has obvious accuracy advantages in the case of each labelled sample number. It can be seen that with the increase in the number of labelled samples, the accuracy of the beginning model also increases rapidly, which is more obvious when the number of labelled samples is relatively small. With the improvement in the accuracy of the beginning model, the gaps between semi-supervised methods and the beginning model are getting smaller, while PPM still maintains a large accuracy lead.

It can be seen from Figure 10 that the improved overall accuracy (semi-supervised model accuracy minus beginning model accuracy) is most obvious when the sample is the smallest. With the increase in the number of labelled samples, the improved accuracy of the model continues to decline. When the initial sample size is 10, the improved accuracy of the proposed method is 14.46%, which is apparently higher than that of other methods co-training (10.34%), noisy-student (8.87%), UDA (11.55%), and self-training (6.98%). The

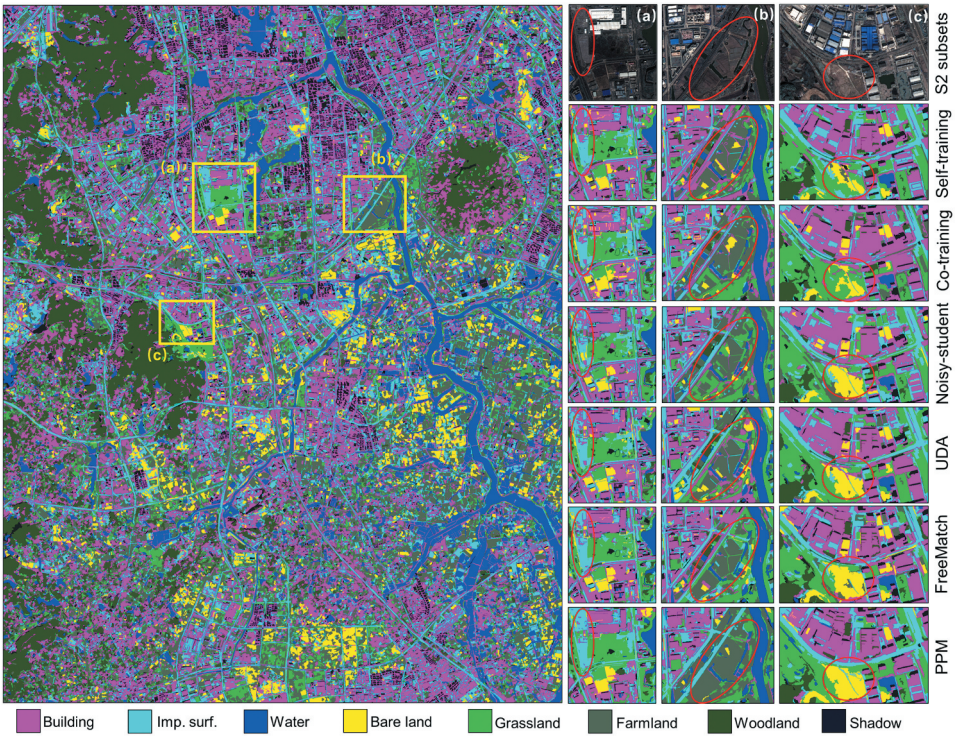

Figure 8. Classification result predicted by PPM in S2, as well as three subsets and their classification result by self-training, co-training, noisy-student, UDA, FreeMatch, and PPM. The positions of the three subsets have been marked in the complete classification map.

proposed method maintains the advantage of maximum accuracy improvement. Co-training performs well when there are fewer labelled samples. Noisy-student improves the model accuracy more when there are more labelled samples.

Furthermore, we discuss the influence of the quantitative relationship between the number of labelled samples and pseudo labelled samples on the accuracy. As found in (Xie et al. 2020; Zhu et al. 2016), there is a correlation between the number of labelled and pseudo labelled samples to maximize accuracy. As shown in Table 4, when the numbers of labelled samples are 10, 15, 20, and 25, the model achieves the highest accuracy when the number of pseudo labelled samples is 10, 15, 25, and 25, respectively. In general, the number of pseudo labelled samples should increase with the increase in the number of labelled samples to achieve the maximum accuracy of the model. In other words, an approximate proportion needs to be maintained between the number of labelled samples and pseudo labelled samples in order to give full play of the model. In this study, this proportion is approximately 1:1. This is because the selected pseudo labelled samples are not always correct. The small number of labelled samples would result in a weak model. The errors introduced by incorrect pseudo labelled samples can be fatal to the weak model. Therefore, in this case, fewer pseudo labelled samples are beneficial for model training. With the increase in labelled samples, the ability

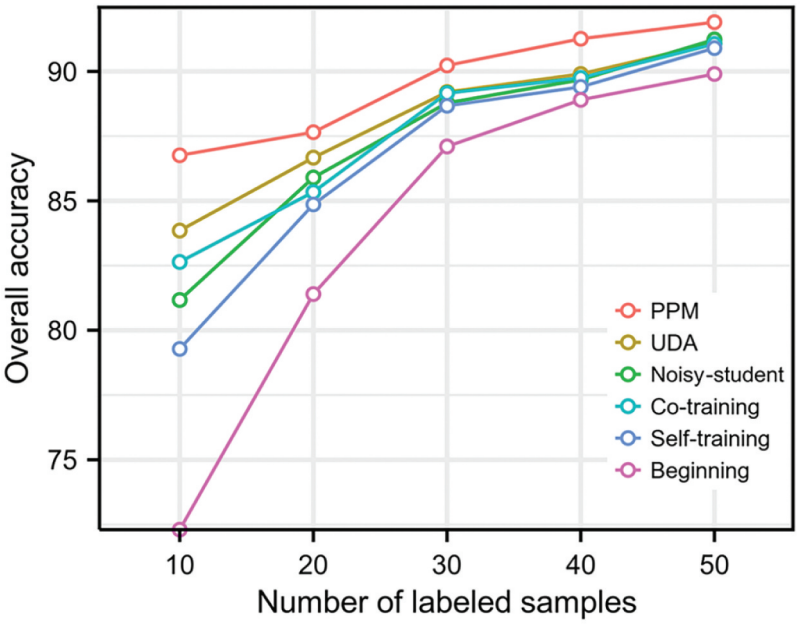

Figure 9. Overall accuracy of each semi-supervised method versus number of labeled samples for training. Beginning represents the accuracy of the model that was trained using only labeled samples.

of the model improves. It can assign more correct pseudo labels and carry more incorrect pseudo labelled samples. Meanwhile, more pseudo labelled samples in turn better improve the model ability. It is worth noting that PPM has multiple iterative processes, and new pseudo labelled samples are added in each iteration. Even though there are not many pseudo labelled samples updated in a single iteration, the number of pseudo labelled samples is considerable when multiple iterations are added up.

5.3. Ablation study

In the ablation experiments, we explored the effects of dropout, diff-aug, and consistency regularization on the accuracy of the model. The number of labelled samples is set to 10 and the number of high-confidence pseudo labelled samples is set to 10. The data of S1 is used for present the results. Since the pseudo-labelling method is the basis of the framework, it is not taken into consideration. For the two perturbations, dropout and diff-aug, the results in Table 5 show that diff-aug has a greater impact on the model accuracy. Eliminating the use of diff-aug decreases the model accuracy by 0.53%, while eliminating the use of dropout decreases the model accuracy by 0.11%. This can be explained by the limited role of dropout, which adds perturbations only at the time of prediction, while diff-aug acts on the whole training process and improves for the pseudo-labelling method as well. After eliminating the consistency regularization, the model accuracy decreases by 1.42%, which shows the important contribution of consistency regularization to the overall model.

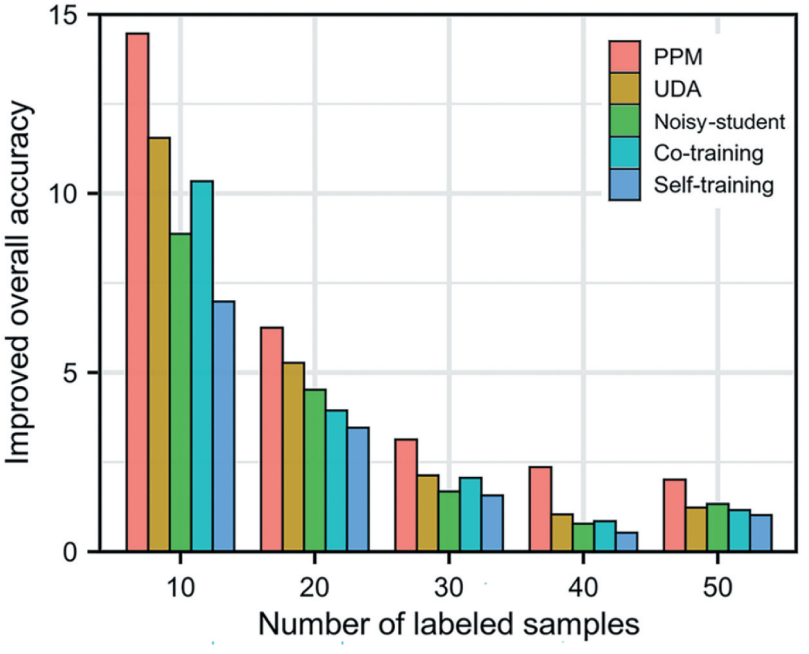

Figure 10. Improved overall accuracy of each semi-supervised method versus number of labeled samples. Improved overall accuracy is obtained by subtracting the accuracy using only labeled samples from the final accuracy.

Table 4. Classification accuracy with different numbers of pseudo labelled (Np) and labelled (Nl) samples. The highest classification accuracies under every number of labelled samples are highlighted in bold.

Np/Nl	10	15	20	25
25	86.35	87.04	**87.93**	**88.87**
20	86.54	87.15	87.89	88.82
15	86.74	**87.26**	87.77	88.79
10	**86.76**	87.09	87.65	88.63

Table 5. Ablation study on the effectiveness of dropout, different augmentation, and consistency regularization.

Dropout	Different augmentation	Consistency regularization	Overall accuracy (%)
√	√	√	86.76
	√	√	86.65
√		√	86.23
√	√		85.34

5.4. Sensitivity analysis for perturbations

Furthermore, we discuss the effect of the intensity of the two types of perturbations on the accuracy of the model. The data of S1 is also used for presenting the results. First, the influence of diff-aug is discussed. The intensity of diff-aug is set from 0.2 to 2, which

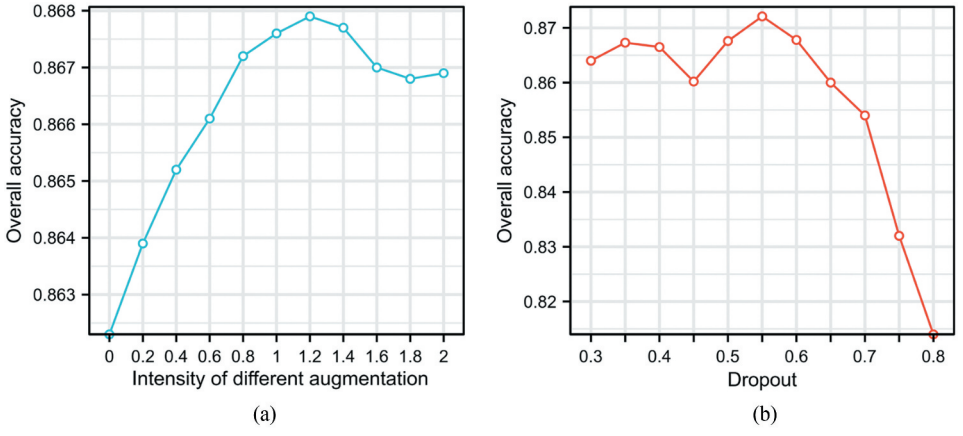

Figure 11. Impact of perturbations on model accuracy. (a) Impact of intensity of different augmentation on overall accuracy. (b) Impact of dropout on overall accuracy.

means that the data are expanded by 20% to 200% of the original. As shown in Figure 11(a), it can be roughly seen that the model accuracy increases and then decreases with the strength of diff-aug. The maximum accuracy of the model is reached when the data are enhanced by 120%.

Then, the influence of dropout on the model accuracy is discussed. As shown in Figure 11(b), when the dropout value increases from 0.3 to 0.5, the accuracy of the model changes little. The model accuracy is maximized at a dropout of 0.55. When the dropout is greater than 0.55, the accuracy of the model decreases continuously. Dropout is to throw away some parameters in the whole connection layer. The value represents the proportion of throwing away parameters. When the dropout is not too large, the CNN has the ability to tolerate some of the parameters not to participate in the prediction. At this point, the dropout introduces perturbations that are beneficial to the model. When the dropout rate is large, there are too few parameters involved in the training process, so that the training effect of the model is not good, and the accuracy rate decreases.

The introduction of perturbations creates differences between peer models. The differences are the driving force for model training in divergence-based semi-supervised approaches (Wang and Zhou 2007). To understand intuitively the relationship between differences and model accuracy, we introduce Q statistics (Kuncheva and Whitaker 2003) to calculate the difference (D) between the models with the following equation.

$$D = 1 - Q \tag{7}$$

where a larger D indicates a larger difference between the peer models.

We plotted the relation between D and OA in PPM and co-training, as shown in Figure 12. As training progresses, the difference between the classifiers of co-training decreases from 0.152 to 0.072, with a decrease of 0.08, and the difference between the classifiers of PPM decreases from 0.156 to 0.088, with a decrease of 0.068, which is smaller than that of co-training. At the same time, with the iteration, the accuracy of co-training increased from 72.9% to 82.64%, with an increase of 9.74%, and the accuracy of PPM

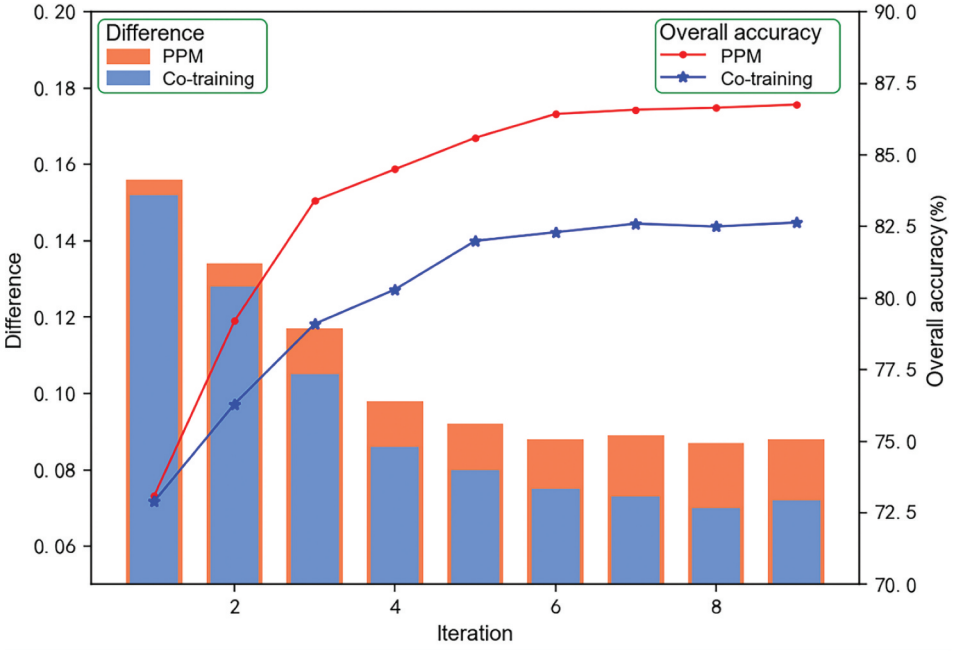

Figure 12. Relation between the difference and overall accuracy in PPM and co-training. The histogram represents the variation in difference between the two models from PPM and co-training. The folded line represents the average accuracy of the two models with iteration.

increased from 73.1% to 86.76%, with an increase of 13.66%. Moreover, the accuracy of co-training stops improving at iteration 7 while the accuracy of PPM continues to improve at iteration 9 because the high difference that is still maintained in PPM.

5.5. Sensitivity analysis for semi-supervised loss parameters

The proposed method uses two semi-supervised methods, pseudo-labelling method and consistent regularization method, for learning unlabelled samples. To further investigate the contributions of the two semi-supervised methods to the model, we conducted a sensitivity analysis on the parameters λ and β, which refer to the weights of the pseudo-labelling method and the consistency regularization method in the loss function, respectively. The data of S1 is used for this experiment. The number of labelled samples is set to 10 and the number of high-confidence pseudo labelled samples is also set to 10. The parameter λ is set from 0 to 1.0 in steps of 0.1, and the parameter β is set from 0 to 0.5 in steps of 0.05.

As shown in Figure 13, the overall accuracy of model increases with increasing λ regardless of the parameter β. This indicates that the pseudo labelled samples selected by the model are of high quality, and increasing the weight of the pseudo-labelling method can make the model learn better. As for the parameter β, there is a pattern that the overall accuracy first increases and then decreases as β increases. This suggests that the consistent regularization method is useful within a certain range of weighting parameters, and that the too large or too small weighting parameters

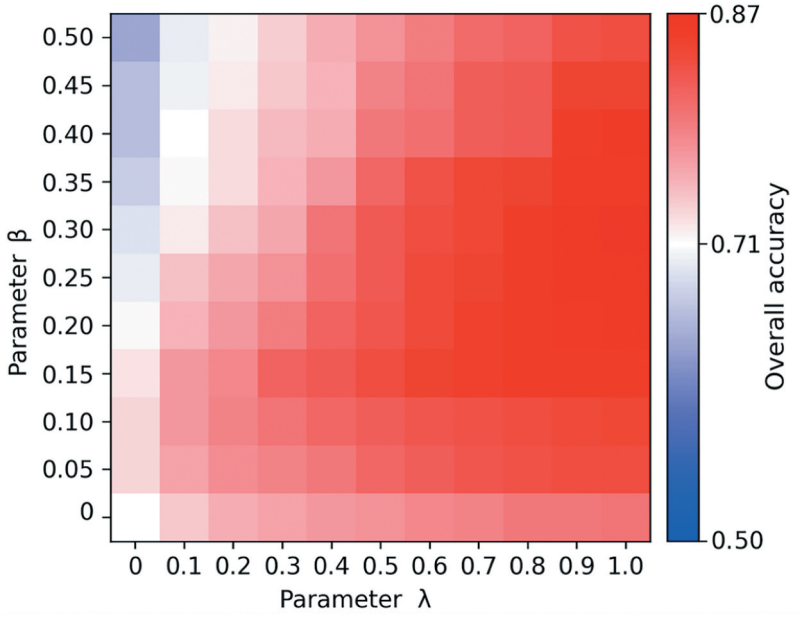

Figure 13. Impact of different parameter λ and parameter β values on the overall accuracy. λ and β refer to the weights of the pseudo-labeling method and the consistency regularization method in the loss function.

reduce the accuracy of the model. When λ is larger, the β corresponding to the optimal accuracy of the model becomes larger, which indicates that the learning of pseudo labelled samples with larger weights improves the model's ability to tolerate larger consistency regularization weight parameters and thus learn more structural features.

6. Conclusion

In this study, we propose a novel semi-supervised object-based classification method. It combines pseudo-labelling method and consistency regularization method. Object-level joint confidence is designed to assess the confidence of the unlabelled object samples. The high-confidence and low-confidence samples are divided by calculating the object-level joint confidence, and then utilized by applying the pseudo-labelling method and the consistency regularization method, respectively. In addition, two types of perturbations are used to make the model more accurate and robust. The experimental results show that the classification accuracy of PPM is better than that of self-training, co-training, noisy-student, UDA, and FreeMatch. Using 10 labelled samples per class, an accuracy of 86.84% was achieved by PPM in S1 and an accuracy of 84.36% in S2. The reasonable choice of perturbations intensity facilitates the training of the model. The proposed method can generate differences well and maintain them, thus the peer model can perform feature learning based on these differences and achieve high classification accuracy. Although this method is designed for GEOBIA, the ideas behind this method could be generalized to other tasks.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study is supported by the National Natural Science Foundation of China [Grant No. 42071297].

ORCID

Pengfeng Xiao http://orcid.org/0000-0003-2739-3302

References

Arazo, E., D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness. 2020. "Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning." *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 1–8. Glasgow, Scotland, UK: IEEE https://doi.org/10.1109/IJCNN48605.2020.9207304.

Benz, U. C., P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. 2004. "Multi-Resolution, Object-Oriented Fuzzy Analysis of Remote Sensing Data for GIS-Ready Information." *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3–4): 239–258. https://doi.org/10.1016/j.isprsjprs.2003.10.002.

Blaschke, T. 2010. "Object Based Image Analysis for Remote Sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1): 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004.

Blaschke, T., G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Queiroz Feitosa, et al. 2014. "Geographic Object-Based Image Analysis – Towards a New Paradigm." *ISPRS Journal of Photogrammetry and Remote Sensing* 87:180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014 .

Chabot, D., C. Dillon, A. Shemrock, N. Weissflog, and E. P. Sager. 2018. "An Object-Based Image Analysis Workflow for Monitoring Shallow-Water Aquatic Vegetation in Multispectral Drone Imagery." *ISPRS International Journal of Geo-Information* 7 (8): 294. https://doi.org/10.3390/ijgi7080294.

Chen, G., G. J. Hay, L. M. Carvalho, and M. A. Wulder. 2012. "Object-Based Change Detection." *International Journal of Remote Sensing* 33 (14): 4434–4457. https://doi.org/10.1080/01431161.2011.648285 .

Chen, G., Q. Weng, G. J. Hay, and Y. He. 2018. "Geographic Object-Based Image Analysis (GEOBIA): Emerging Trends and Future Opportunities." *GIScience & Remote Sensing* 55 (2): 159–182. https://doi.org/10.1080/15481603.2018.1426092.

Cubuk, E. D., B. Zoph, J. Shlens, and Q. V. Le. 2019. "Randaugment: Practical Data Augmentation with No Separate Search." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, California, US.

Fu, T., L. Ma, M. Li, and B. A. Johnson. 2018. "Using Convolutional Neural Network to Identify Irregular Segmentation Objects from Very High-Resolution Remote Sensing Imagery." *Journal of Applied Remote Sensing* 12 (2): 025010–025010. https://doi.org/10.1117/1.JRS.12.025010.

Gao, L., H. Liu, M. Yang, L. Chen, Y. Wan, Z. Xiao, and Y. Qian. 2021. "STransFuse: Fusing Swin Transformer and Convolutional Neural Network for Remote Sensing Image Semantic Segmentation." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:10990–11003. https://doi.org/10.1109/JSTARS.2021.3119654 .

Georganos, S., T. Grippa, S. Vanhuysse, M. Lennert, M. Shimoni, S. Kalogirou, and E. Wolff. 2018. "Less is More: Optimizing Classification Performance Through Feature Selection in a Very-High-Resolution Remote Sensing Object-Based Urban Application." *GIScience & Remote Sensing* 55 (2): 221–242. https://doi.org/10.1080/15481603.2017.1408892.

Hamedianfar, A., and H. Z. M. Shafri. 2015. "Detailed Intra-Urban Mapping Through Transferable OBIA Rule Sets Using WorldView-2 Very-High-Resolution Satellite Images." *International Journal of Remote Sensing* 36 (13): 3380–3396. https://doi.org/10.1080/01431161.2015.1060645.

Han, W., R. Feng, L. Wang, and Y. Cheng. 2018. "A Semi-Supervised Generative Framework with Deep Learning Features for High-Resolution Remote Sensing Image Scene Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 145:23–43. https://doi.org/10.1016/j.isprsjprs.2017.11.004 .

Hossain, M. D., and D. Chen. 2019. "Segmentation for Object-Based Image Analysis (OBIA): A Review of Algorithms and Challenges from Remote Sensing Perspective." *ISPRS Journal of Photogrammetry and Remote Sensing* 150:115–134. https://doi.org/10.1016/j.isprsjprs.2019.02.009 .

Huang, G., Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger. 2016. "Deep Networks with Stochastic Depth." *Proceedings of the European conference on computer vision (ECCV)*, 646–661. Amsterdam, The Netherlands. https://doi.org/10.1007/978-3-319-46493-0_39.

Janiesch, C., P. Zschech, and K. Heinrich. 2021. "Machine Learning and Deep Learning." *Electronic Markets* 31 (3): 685–695. https://doi.org/10.1007/s12525-021-00475-2.

Kipf, T. N., and M. Welling. 2017. "Semi-Supervised Classification with Graph Convolutional Networks." *Proceedings of the International Conference on Learning Representations*, Toulon, France, https://doi.org/10.48550/arXiv.1609.02907.

Kucharczyk, M., G. J. Hay, S. Ghaffarian, and C. H. Hugenholtz. 2020. "Geographic Object-Based Image Analysis: A Primer and Future Directions." *Remote Sensing* 12 (12): 2012. https://doi.org/10.3390/rs12122012.

Kuncheva, L. I., and C. J. Whitaker. 2003. "Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy." *Machine Learning* 51 (2): 181. https://doi.org/10.1023/A:1022859003006.

Lee, D. H. 2013. "Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks." *Proceedings of the Workshop on Challenges in Representation Learning, ICML*, Atlanta, Georgia, US 3:896.

Li, J., B. Sun, S. Li, and X. Kang. 2021. "Semisupervised Semantic Segmentation of Remote Sensing Images with Consistency Self-Training." *IEEE Transactions on Geoscience and Remote Sensing* 60:1–11. https://doi.org/10.1109/TGRS.2021.3134277.

Mahajan, D., R. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li, A. Bharambe, and L. Van Der Maaten. 2018. "Exploring the Limits of Weakly Supervised Pretraining." *Proceedings of the European conference on computer vision (ECCV)*, 181–196 Munich, Germany.

Momeni, R., P. Aplin, and D. S. Boyd. 2016. "Mapping Complex Urban Land Cover from Spaceborne Imagery: The Influence of Spatial Resolution, Spectral Band Set and Classification Approach." *Remote Sensing* 8 (2): 88. https://doi.org/10.3390/rs8020088.

Ozelkan, E., G. Chen, and B. B. Ustundag. 2016. "Multiscale Object-Based Drought Monitoring and Comparison in Rainfed and Irrigated Agriculture from Landsat 8 OLI Imagery." *International Journal of Applied Earth Observation and Geoinformation* 44:159–170. https://doi.org/10.1016/j.jag.2015.08.003.

Qiao, S., W. Shen, Z. Zhang, B. Wang, and A. Yuille. 2018. "Deep Co-Training for Semi-Supervised Image Recognition," *Proceedings of the European conference on computer vision (ECCV)*, 135–152. Munich, Germany.

Rousset, G., M. Despinoy, K. Schindler, and M. Mangeas. 2021. "Assessment of Deep Learning Techniques for Land Use Land Cover Classification in Southern New Caledonia." *Remote Sensing* 13 (12): 2257. https://doi.org/10.3390/rs13122257.

Sajjadi, M., M. Javanmardi, and T. Tasdizen. 2016. "Regularization with Stochastic Transformations and Perturbations for Deep Semi-Supervised Learning." In *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16)*, 1171–1179. Barcelona, Spain.

Shi, Q., M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang. 2021. "A Deeply Supervised Attention Metric-Based Network and an Open Aerial Image Dataset for Remote Sensing Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 60:1–16. https://doi.org/10.1109/TGRS.2021.3085870 .

Springenberg, J. T. 2015. "Unsupervised and Semi-Supervised Learning with Categorical Generative Adversarial Networks." *arXiv preprint arXiv*:1511.06390. https://doi.org/10.48550/arXiv.1511.06390.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *The Journal of Machine Learning Research* 15 (1): 1929–1958.

Sun, W., W. Shao, J. Peng, G. Yang, X. Meng, and Q. Du. 2020. "Multiscale Low-Rank Spatial Features for Hyperspectral Image Classification." *IEEE Geoscience and Remote Sensing Letters* 19:1–5. https://doi.org/10.1109/LGRS.2020.3034631.

Sun, W., G. Yang, K. Ren, J. Peng, C. Ge, X. Meng, and Q. Du. 2021. "A Label Similarity Probability Filter for Hyperspectral Image Postclassification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:6897–6905. https://doi.org/10.1109/JSTARS.2021.3094197.

Tang, X., Q. Ma, X. Zhang, F. Liu, J. Ma, and L. Jiao. 2021. "Attention Consistent Network for Remote Sensing Scene Classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:2030–2045. https://doi.org/10.1109/JSTARS.2021.3051569.

Tan, K., Y. Zhang, X. Wang, and Y. Chen. 2019. "Object-Based Change Detection Using Multiple Classifiers and Multi-Scale Uncertainty Analysis." *Remote Sensing* 11 (3): 359. https://doi.org/10.3390/rs11030359.

Tarvainen, A., and H. Valpola. 2017. "Mean Teachers are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning Results." In *Proceedings of the 31th International Conference on Neural Information Processing Systems (NIPS'17)*, 1195–1204. Long Beach, California, USA.

Torres-Sánchez, J., F. López-Granados, and J. M. Pena. 2015. "An Automatic Object-Based Method for Optimal Thresholding in UAV Images: Application for Vegetation Detection in Herbaceous Crops." *Computers and Electronics in Agriculture* 114:43–52. https://doi.org/10.1016/j.compag.2015.03.019.

Wang, J., S. Chen, C. H. Q. Ding, J. Tang, and B. Luo. 2022. "RanPaste: Paste Consistency and Pseudo Label for Semisupervised Remote Sensing Image Semantic Segmentation." *IEEE Transactions on Geoscience and Remote Sensing* 60:1–16. https://doi.org/10.1109/TGRS.2021.3102026.

Wang, Y., H. Chen, Q. Heng, W. Hou, Y. Fan, Z. Wu, J. Wang, et al. 2023. "Freematch: Self-Adaptive Thresholding for Semi-Supervised Learning." *Proceedings of the International Conference on Learning Representations*, Kigali, Rwanda. https://doi.org/10.48550/arXiv.2205.07246.

Wang, J., C. H. Q. Ding C, S. Chen, B. Luo, and C. He. 2020. "Semi-Supervised Remote Sensing Image Semantic Segmentation via Consistency Regularization and Average Update of Pseudo-Label." *Remote Sensing* 12 (21): 3603. https://doi.org/10.3390/rs12213603.

Wang, Y., Z. Zhang, L. Feng, Y. Ma, and Q. Du. 2021. "A New Attention-Based CNN Approach for Crop Mapping Using Time Series Sentinel-2 Images." *Computers and Electronics in Agriculture* 184:106090. https://doi.org/10.1016/j.compag.2021.106090.

Wang, W., and Z. Zhou. 2007. "Analyzing Co-Training Style Algorithms," *Proceedings of the Machine Learning: ECML: 18th European Conference on Machine Learning* , 454–465. Warsaw, Poland.

Whyte, A., K. P. Ferentinos, and G. P. Petropoulos. 2018. "A New Synergistic Approach for Monitoring Wetlands Using Sentinels-1 and 2 Data with Object-Based Machine Learning Algorithms." *Environmental Modelling & Software* 104:40–54. https://doi.org/10.1016/j.envsoft.2018.01.023.

Wu, J., L. Li, and W. Wang. 2018. "Reinforced Co-Training." *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1252–1262. New Orleans, Louisiana, US. https://doi.org/10.48550/arXiv.1804.06035.

Wu, Y., G. Mu, C. Qin, Q. Miao, W. Ma, and X. Zhang. 2020. "Semi-Supervised Hyperspectral Image Classification via Spatial-Regulated Self-Training." *Remote Sensing* 12 (1): 159. https://doi.org/10.3390/rs12010159.

Xie, Q., Z. Dai, E. Hovy, M. T. Luong, and Q. V. Le. 2020. "Unsupervised Data Augmentation for Consistency Training." In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20)*, 6256–6268.

Xie, Q., M. T. Luong, E. Hovy, and Q. V. Le. 2020. "Self-Training with Noisy Student Improves Imagenet Classification." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10687–10698. Seattle, Washington, US. https://doi.org/10.48550/arXiv.1911.04252.

Yang, Y., and S. Newsam. 2010. "Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification." *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 270–279. San Jose, California, US.

Zhai, X., A. Oliver, A. Kolesnikov, and L. Beyer. 2019. "S4l: Self-Supervised Semi-Supervised Learning." *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1476–1485. Seoul, South Korea. https://doi.org/10.1109/ICCV.2019.00156.

Zhang, R., and A. I. Rudnicky. 2006. "A New Data Selection Principle for Semi-Supervised Incremental Learning." *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)* 2:780–783. Hong Kong, China: IEEE. https://doi.org/10.1109/ICPR.2006.115.

Zhang, C., I. Sargent, X. Pan, H. Li, A. Gardiner, J. Hare, and P. M. Atkinson. 2018. "An Object-Based Convolutional Neural Network (OCNN) for Urban Land Use Classification." *Remote Sensing of Environment* 216:57–70. https://doi.org/10.1016/j.rse.2018.06.034 .

Zhao, W., S. Du, and W. J. Emery. 2017. "Object-Based Convolutional Neural Network for High-Resolution Imagery Classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10 (7): 3386–3396. https://doi.org/10.1109/JSTARS.2017.2680324 .

Zheng, Z., X. Zhang, P. Xiao, and Z. Li. 2021. "Integrating Gate and Attention Modules for High-Resolution Image Semantic Segmentation." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:4530–4546. https://doi.org/10.1109/JSTARS.2021.3071353.

Zhu, L., P. Xiao, X. Feng, L. Zhang, Y. Huang, and C. Li. 2016. "A Co-Training, Mutual Learning Approach Towards Mapping Snow Cover from Multi-Temporal High-Spatial Resolution Satellite Imagery." *Isprs Journal of Photogrammetry & Remote Sensing* 122: 179–191.

Zoph, B., and Q. V. Le. 2017. "Neural Architecture Search with Reinforcement Learning." *Proceedings of the International Conference on Learning Representations*, Toulon, France. https://doi.org/10.48550/arXiv.1611.01578.