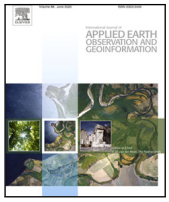
International Journal of Applied Earth Observations and Geoinformation 112 (2022) 102931

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Weakly supervised high spatial resolution land cover mapping based on self-training with weighted pseudo-labels

Wei Liu [a], Jiawei Liu [a], Zhipeng Luo [b,*], Hongbin Zhang [a], Kyle Gao [c], Jonathan Li [c]

[a] School of Software, East China Jiaotong University, Nanchang 330013, China
[b] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, 999077, Hong Kong, China
[c] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
Unsupervised domain adaptation
Land cover mapping
Self-training
Pseudo-learning
Semantic segmentation

ABSTRACT

Despite its success, deep learning in land cover mapping requires a massive amount of pixel-wise labeled images. It typically assumes that the training and test scenes are similar in data distribution. The performance of models trained on any particular dataset could degrade significantly on a new dataset due to the domain shift or domain gap across datasets, resulting in new training data requiring labor-intensive manual pixel-wise labeling. This paper proposes a land cover mapping framework combining Feature Pyramid Network (FPN) and self-training. In the FPN, we integrate ConvNeXt with a Pyramid Pooling Module (PPM). Combining the FPN and the PPM improves the segmentation performance, which benefits from the multiscale aggregation of pyramid features. To fully exploit pseudo-labels, we design an Unsupervised Domain Adaptation (UDA) land cover mapping scheme with self-training using weighted pseudo-labels of the target samples. The proposed land cover mapping framework could benefit from multiscale aggregation of pyramid features and the full use of the pseudo-labels. Comparison results on the LoveDA dataset, the latest large-scale unsupervised domain adaptation dataset for land cover mapping, empirically demonstrated that our land cover mapping approach significantly outperforms the baselines in both UDA scenarios, i.e., *Urban* → *Rural* and *Rural* → *Urban*. The models of this paper are now publicly available on GitHub.[1]

1. Introduction

The advances in remote sensing and computer techniques promote the widespread acquisition and use of timely High Spatial Resolution (HSR) remote sensing data across the globe (Han et al., 2018). The unprecedentedly fast-growing volume of timely HSR data available offers new opportunities for various applications, such as monitoring of forests, cities, natural disasters, and agriculture. HSR remote sensing technology can provide essential earth observation information to monitor geographical and ecological environments, such as climate and wetland. Specifically, land cover mapping in remote sensing aims to determine land cover types (e.g., highland, forest, and water) at every image pixel. It is considered more complex than scene classification and retrieval and is one of the most challenging remote data parsing tasks.

Due to the deep learning technology, the research and application of land cover mapping have made remarkable progress in the past decade. Deep learning in land cover mapping tends to rely heavily on large numbers of high-quality pixel-wise labeled images despite its success. Moreover, an underlying assumption of most deep learning models for land cover mapping is that the training and the test scenes are similar in data distributions. However, in the practical application of HSR land cover mapping, the distribution of training data and test data is quite different. Especially in urban and rural scenes, the representation of land cover often differs greatly in object scales, class distributions, and pixel spectrum (Wang et al., 2021b). In terms of class distribution, there are many more buildings in urban scenes as opposed to rural scenes since urban scenes have large populations. By contrast, the rural scenes have more areas of agricultural land and forests. Objects of the same class collected from different scenes typically have various scales. Since rural images usually contain large homogeneous geographical areas, such as farmland and forest, the standard variance of spectral statistics is smaller. The significant domain gap between the rural and urban scenes reduces the generalization of the land cover mapping model. The performance of models trained on one dataset could degrade significantly on a new dataset because of the distribution gap or shift across datasets (Liu et al., 2021). Relatively few studies of land cover mapping focus on model transferability between datasets with significant differences.

* Corresponding author.
E-mail address: kent-zhipeng.luo@polyu.edu.hk (Z. Luo).

[1] https://github.com/csliujw/uda-self-training

https://doi.org/10.1016/j.jag.2022.102931
Received 3 April 2022; Received in revised form 29 June 2022; Accepted 18 July 2022
Available online 9 August 2022

We focus on unsupervised domain adaptation (UDA) to avoid massive manual pixel-wise labeling on target datasets for land cover mapping in this paper. UDA aims at training a land cover mapping model in one dataset (source domain) and then applying the model to make accurate predictions in a different dataset (target domain). Self-training is a straightforward yet competitive method in UDA tasks, exploiting unlabeled target data by training with pseudo-labels generated with the model trained using the labeled source dataset. Because of the domain discrepancy and limitation of model precision, the generated labels of the target data very likely contain incorrect predictions, essentially compromising the domain adaptation (DA) process. A typical strategy (Zou et al., 2018, 2019; Wang et al., 2021a, 2020a) to mitigate the effects of noisy labels is setting a threshold to neglect pseudo labels with low-confidence scores. However, for different target domains, it is hard to determine such a threshold, which depends on various factors such as the similarity between domains, pixel location, and pixel category. Therefore, every pseudo-labels should be treated separately. It is typically not suitable to adopt a fixed threshold.

Combining a Feature Pyramid Network (FPN) and self-training, we devise a UDA scheme to bridge the gap between different domains. The proposed UDA scheme consists of two training stages, which share the same land cover mapping network. We train the land cover mapping network employing the labeled scenes from the source dataset and then generate pseudo-labels for the unlabeled target samples during the first stage. The source ground-truths and the target pseudo-labels are utilized during the second stage to fine-tune the network. In the FPN, we integrate ConvNeXt (Liu et al., 2022) with a Pyramid Pooling Module (PPM) (Zhao et al., 2017; Xiao et al., 2018).

The FPN is a generic feature extractor that exploits the inherent multi-scale, pyramidal hierarchy of deep convolutional networks to construct feature pyramids with marginal extra cost. The PPM is an effective global contextual prior for semantic segmentation, and is highly compatible with the FPN. Combining the FPN and the PPM improves the segmentation performance in the source domain, which benefits from the multiscale aggregation of pyramid features. To fully exploit pseudo-labels, we design a UDA scheme with self-training using weighted pseudo-labels of the target data. More specifically, the Jensen–Shannon divergence of the outputs of the main and the auxiliary branches of the FPN is utilized to re-weight the losses of pseudo-labels. The more inconsistent the outputs of the two branches are, the smaller the weight of the corresponding pseudo-label is. The proposed land cover mapping framework benefits from the multiscale aggregation of pyramid features and full utilization of the pseudo-labels.

The contributions of our paper are summarized as follows:

(1) We design a land cover mapping network based on an FPN using ConvNeXt with a PPM to improve the land cover mapping performance. The land cover mapping task can benefit from multiscale aggregation of pyramid features.
(2) To tackle the domain shift problem and improve the network generalization ability, we design a UDA scheme with self-training using weighted pseudo-labels. The weights are calculated by the consistency of the two outputs of the land cover mapping network for the corresponding target samples.
(3) We evaluated our land cover mapping method and the baselines on the LoveDA dataset, the latest unsupervised domain adaptation dataset for land cover mapping. Comparison results empirically indicate that our scheme is significantly superior to the baselines on both UDA scenarios: *Urban → Rural* and *Rural → Urban*.

2. Related research

2.1. Unsupervised domain adaptation

Domain adaptation (DA) is a kind of weakly-supervised learning that aims to mitigate the distribution discrepancy between different domains. This paper focuses on unsupervised domain adaptation (UDA). In the training phase of UDA for land cover mapping, the source data has labels available, while the target data has no labels available. In the last decade, a wide range of UDA methods have been devised to improve the generalization ability of models across different datasets (Wei et al., 2021; Liu et al., 2021). These UDA schemes mitigate the domain discrepancy through three different levels (Liu et al., 2021; Toldo et al., 2020), including input-level (Li et al., 2019), output-level (Vu et al., 2019; Pan et al., 2020), and feature-level (Ye et al., 2019). Specifically, a line of UDA works (Deng et al., 2019; Ye et al., 2019; Kang et al., 2020; Yan et al., 2021; Zhang et al., 2021a; Cai et al., 2022) have been proposed to tackle the domain shift problem for semantic segmentation, the task of which, like land cover mapping, is to determine the category of each pixel in an image.

Most of the recent UDA methods align source data with target data on distribution via adversarial training (Ye et al., 2019; Zhang et al., 2021a; Benjdira et al., 2019) or pseudo-label learning (Zou et al., 2018, 2019; Sohn et al., 2020; Zhang et al., 2021a,b; Zheng and Yang, 2021; Gu et al., 2022). These methods focused on aligning source and target domains to transfer shared knowledge across two significant different domains. When only a small number of labeled samples are used, the performance of UDA algorithms is still obviously behind that of the supervised or semi-supervised learning methods.

2.2. Unsupervised domain adaptation via adversarial training

The adversarial UDA methods in recent years used Generative Adversarial Networks (GANs) to approach global cross-domain alignment main through two different levels, i.e. feature-level and output-level. Ye et al. (2019) improved the adaptation capacity of the GAN by using clustering on the training set of SAR images to extract useful class information. RoadDA (Zhang et al., 2021a) aligned the features of the source scenes with the target scenes via GNAs, modeling the interclass and the intraclass discrepancies between the labeled source scenes and the unlabeled target scenes. Benjdira et al. (2019) utilized a GAN to attain image-level alignment from the source dataset to the target dataset. A cross mean teacher (CMT) UDA method (Yan et al., 2021) was developed to make full use of very pixel in the target dataset. CMT contained two teacher networks and two student networks for cross-consistency constraints. CaGAN (Xu et al., 2020) designed a class-aware GAN, which modeled the intraclass and the interclass discrepancies between the two domains, using adaptive category selection and alignment. To avoid the complexity of high-dimensional feature space adaptation, a series of works (Tsai et al., 2018; Chen et al., 2019; Luo et al., 2019) approach adversarial domain adaptation on the low-dimensional output space. In these framework, a domain discriminator is provided with prediction maps from source and target samples and it is optimized to infer which domain the samples come from.

2.3. Pseudo-label learning for domain adaptation

Pseudo-label learning for UDA can be classified as: (1) entropy minimization (Chen et al., 2019; Saito et al., 2019; Yang and Soatto, 2020), (2) curriculum learning (Bengio et al., 2009; Cascante-Bonilla et al., 2021), and (3) self-training (Choi et al., 2019b; Zheng and Yang, 2021). Entropy minimization (Vu et al., 2019; Chen et al., 2019; Saito et al., 2019; Yang and Soatto, 2020) aims to minimize the entropy of the predicted probability maps of the target data to generate the outputs for the input samples with higher confidence scores. In order to ensure the high predictive certainty of target prediction, Vu et al. (2019) proposed two entropy minimization methods: direct entropy minimization (DEM) utilizing an entropy loss and indirect entropy minimization (IEM) based on an adversarial loss. MME (Saito et al., 2019) alternately maximized the conditional entropy of target data for a domain classifier and minimized it for a feature encoding network. In order to balance the effects of the gradient of the simple samples and the gradient of the hard samples, Chen et al. (2019) proposed a max-square loss to prevent

easily transferable categories from dominating the training in the target dataset, thereby minimizing entropy. Curriculum learning (Bengio et al., 2009; Dai et al., 2020; Cascante-Bonilla et al., 2021; Shu et al., 2019) is a kind of machine learning technique that aims at gradually training models from easy to difficult in sample selection (Dai et al., 2020). To obtain some necessary properties of the target dataset, Zhang et al. (2017) designed a progressive learning approach for land cover mapping of urban scenes, which tackled easy tasks first. To alleviate the problem of false pseudo-labels for DA, PCDA (Choi et al., 2019a) combined curriculum learning with density-based clustering.

Self-training is a competitive yet straightforward approach to the UDA task and typically includes three main stages: (1) The model is trained in the labeled source dataset. (2) Pseudo-labels are generated for the target samples with the trained model. (3) The model is retrained or fine-tuned using the ground truths and the pseudo-labels. Zhang et al. (2020) proposed a layer alignment method and the feature covariance loss function to alleviate the cross-domain shift. TPLD (Shin et al., 2020) performed a easy-hard classification scheme based on confidence scores to partition the target samples into easy and hard splits. It used full pseudo-labels for the easy samples. It adopted adversarial learning to align the features of hard samples with the features of the easy ones. Similarly, RoadDA (Zhang et al., 2021a) trained the model with source scenes and labeled the target scenes utilizing the trained model. In order to partition the target data into a labeled easy split and an unlabeled hard split,it devised a classifier based on the road confidence scores. Then, it aligned the easy and hard splits features using adversarial learning to improve the segmentation performance progressively. To make full use of the target samples, Zheng and Yang (2021) modeled uncertainty of the pseudo-labels using the prediction variance and optimized the objective utilizing the uncertainty of the pseudo-labels. DistributionNet (Yu et al., 2019) modeled models each feature as a Gaussian distribution with its variance representing the uncertainty of the extracted features. TGCF-DA (Choi et al., 2019b) used student–teacher self-ensembling adaptation techniques, where an extra network provide self-supervised information for the unlabeled data. The teacher model provided pseudo-labels to share reliable knowledge with the student model by supervised training on target samples.

As mentioned above, quite a few pseudo-label learning methods have been proposed for UDA. However, there is still a significant gap between the performance of the UDA and supervised approaches. Due to the domain discrepancy and the limitation of model accuracy, the pseudo-labels are inevitably noisy. Thus, the generated labels provided to the final adapted model could significantly compromise the training process. This paper explores how to effectively improve the model's performance in the source domain and make full use of pseudo-labels to bridge the domain gap.

3. The proposed land cover mapping method

3.1. Overview

To reduce the labeling cost and fully exploit pseudo-labels, we design a UDA scheme with self-training using weighted pseudo-labels of the target images. As illustrated in Fig. 1, the proposed land cover mapping network is a Feature Pyramid Network (FPN) utilizing UPerNet (Xiao et al., 2018) with ConvNeXt (Liu et al., 2022) as the backbone. The UperNet enlarges the receptive field of deep CNNs and improves the segmentation performance, which benefits from the multiscale aggregation of pyramid features. The proposed land cover mapping network has two branches. The main takes as input the fused feature, while the auxiliary branch takes as input the output of the third downsampling layer. Both branches perform two convolution operations on their respective input features and then are resized to the image size through upsampling.

The UDA scheme consists of two training stages, sharing the same land cover mapping network. We train the land cover mapping network utilizing the labeled source images and generate pseudo-labels for the target scenes during the first stage. During the second stage, the Jensen–Shannon divergence (JSD) of the outputs of the main and the auxiliary branches of the FPN is utilized to re-weight the losses of pseudo-labels. The more inconsistent the outputs of the two branches are, the lower the confidence of the pseudo-label. The smaller the value of JSD, the greater the weight of the pseudo-label during training. The source ground truths and the target pseudo-labels are utilized to fine-tune the land cover mapping network. The proposed land cover mapping framework can benefit from the multiscale aggregation of pyramid features and the full utilization of pseudo-labels.

3.2. Land cover mapping based on feature pyramid network

As shown in Fig. 1, the proposed land cover mapping network combines ConvNeXt (Liu et al., 2022) with UPerNet. ConvNeXt is a family of pure ConvNet models, constructed entirely from standard ConvNet modules, inspired by the design of Vision Transformers. It is accurate, efficient, scalable, and very simple in design. It obtained state-of-the-art semantic segmentation scores on the large-scale benchmark ADE20K (Zhou et al., 2019), while retaining the simplicity and efficiency of standard ConvNets. We use ConvNeXt as the backbone of the UPerNet, applying a Pyramid Pooling Module (PPM) from PSPNet (Zhao et al., 2017) on the last layer of ConvNeXt before feeding the feature extracted from ConvNeXt into the upsampling stage in the FPN.

The land cover mapping network has two mapping branches: the main and auxiliary branches, which take the fused feature and the output of the third downsampling layer as inputs, respectively. Both branches perform two convolution operations on the input features and then are resized to the image size through upsampling. Before the last upsampling operation, the main and auxiliary branches are at 1/4 and 1/16 scales, respectively. Let $\mathbf{F}_m(x_i^s|\theta)$ and $\mathbf{F}_a(x_i^s|\theta)$ denote the outputs of the main and the auxiliary branches for the source sample x_i^s under the model parameter θ, respectively. For land cover mapping with the number of predefined categories C, the total supervised cross-entropy loss $\mathcal{L}_s$ for a source sample x_i^s with label one-hot encoded label y_i can be represented as the weighted sum of the main and auxiliary branches' cross-entropy losses:

$$\mathcal{L}_s(x_i^s) = -\sum_{j=1}^{C} y_i^j \log \mathbf{F}_m^j(x_i^s|\theta) - \lambda_1 \sum_{j=1}^{C} y_i^j \log \mathbf{F}_a^j(x_i^s|\theta), \tag{1}$$

where y_i^j denotes the component of y_i corresponding to class j. Likewise $\mathbf{F}_m^j(x_i^t|\theta)$ and $\mathbf{F}_a^j(x_i^t|\theta)$ denote the components of $\mathbf{F}_m(x_i^t|\theta)$ and $\mathbf{F}_a(x_i^t|\theta)$ corresponding to class j. λ_1 is the weight corresponding to the supervised cross-entropy loss of the auxiliary branch. Thus, the average cross-entropy loss for the pixel-wise labeled images in the source dataset S can be expressed as:

$$\mathcal{L}_s(S) = -\frac{1}{|S|} \sum_{x_i^s \in S} \sum_{j=1}^{C} y_i^j \log \mathbf{F}_m^j(x_i^s|\theta) - \frac{\lambda_1}{|S|} \sum_{x_i^s \in S} \sum_{j=1}^{C} y_i^j \log \mathbf{F}_a^j(x_i^s|\theta), \tag{2}$$

in which $|S|$ represents the size of the source dataset S. Using the supervised cross-entropy loss $\mathcal{L}_s(S)$, we can train the land cover mapping network utilizing the labeled source scenes, which will then be employed to generate pseudo-labels for the unlabeled target scenes.

3.3. Self-training with loss re-weighting for pseudo-labels

Standard self-training aims to produce pseudo-labels for images from target scenes with a trained source network and then retrain or finetune the network using the labeled source samples and the target samples with pseudo-labels. Because of the domain discrepancy, the generated labels on the target domain very likely contain incorrect predictions, which could compromise the training process. An intuitive way to overcome the problem is to set a fixed threshold to

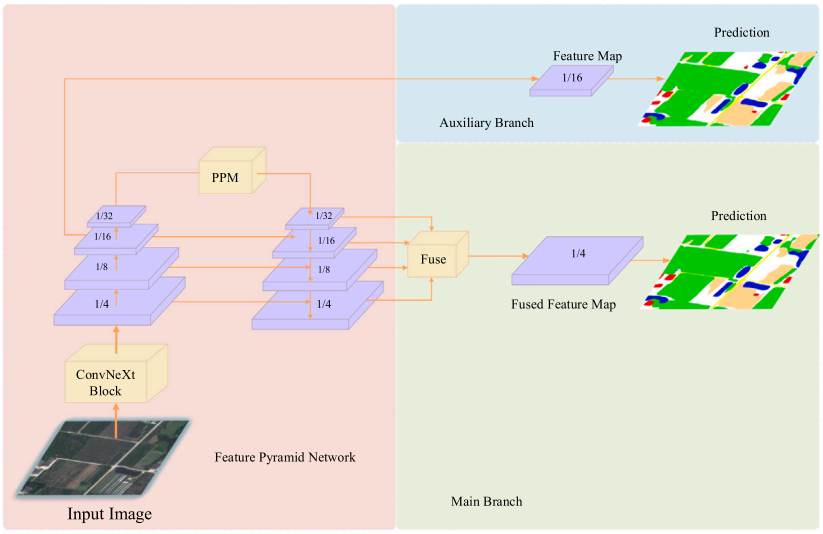


Fig. 1. The proposed land cover mapping network, which is based on a feature pyramid network using UPerNet with ConvNeXt as the backbone network. It has two branches, the main and auxiliary branches.

filter the pseudo-labels with low-confidence scores. However, it is hard to determine such a suitable threshold for different target domains, which depends on various factors such as the similarity between domains, pixel location, and pixel category. To treat every pseudo-label separately, we approach UDA using self-training with weighted pseudo-labels in this paper (see Fig. 2). The more inconsistent the outputs of the two branches are, the smaller the weight of the corresponding pseudo-label is.

We use the Jensen–Shannon divergence (JSD) to measure the mutual consistency of the outputs of the main and the auxiliary branches. The smaller the JSD of the two outputs, the more they are consistent. In the self-training stage, the JSD is utilized to re-weight losses for pseudo-labels. Formally, the JSD for the target sample x_i^t between the outputs from the main and the auxiliary branches can be represented as:

$$\begin{aligned} D_{JS}\left(x_i^t\right) &= \frac{1}{2} KL(\mathbf{F}_m(x_i^t|\theta)\|\mathbf{M}) + \frac{1}{2} KL(\mathbf{F}_a(x_i^t|\theta)\|\mathbf{M}) \\ &= -\frac{1}{2}\sum_{j=1}^{C} \mathbf{F}_m^j(x_i^t|\theta) \log \frac{\mathbf{M}^j}{\mathbf{F}_a^j(x_i^t|\theta)} - \frac{1}{2}\sum_{j=1}^{C} \mathbf{F}_a^j(x_i^t|\theta) \log \frac{\mathbf{M}^j}{\mathbf{F}_m^j(x_i^t|\theta)}, \end{aligned} \tag{3}$$

where $\mathbf{M} = \frac{\mathbf{F}_a(x_i^t|\theta)+\mathbf{F}_m(x_i^t|\theta)}{2}$. $\mathbf{M}^j$ denotes the component of $\mathbf{M}$ corresponding to class j, likewise $\mathbf{F}_m^j(x_i^t|\theta)$ and $\mathbf{F}_a^j(x_i^t|\theta)$.

The cross-entropy loss for the target sample x_i^t with a pseudo-label $\hat{\mathbf{y}}_i$ is :

$$L_{ce}\left(x_i^t\right) = -\sum_{j=1}^{C} \hat{\mathbf{y}}_i^j \log \mathbf{F}_m^j(x_i^t|\theta) - \lambda_2 \sum_{j=1}^{C} \hat{\mathbf{y}}_i^j \log \mathbf{F}_a^j(x_i^t|\theta), \tag{4}$$

in which $\hat{\mathbf{y}}_i^j$ is the component of the pseudo-label $\hat{\mathbf{y}}_i$ corresponding to class j. λ_2 is the weight corresponding to the loss of the auxiliary branch.

The more inconsistent the outputs of the two branches are, the smaller the weight of the corresponding pseudo-label is. Using the JSD, the weighted entropy-cross loss for x_i^t can be formulated as:

$$L_w\left(x_i^t\right) = \exp\{-D_{JS}\left(x_i^t\right)\} \cdot L_{ce}\left(x_i^t\right) + D_{JS}\left(x_i^t\right). \tag{5}$$

The second term $D_{JS}\left(x_i^t\right)$ in Eq. (5) is used to make the two branch's outputs for x_i^t as consistent as possible. Using the JSD of the two branch's outputs enables the self-training scheme to stress importance on high-confidence pseudo-labels. The average loss across the target dataset can be expressed as:

$$\mathcal{L}_t(T) = \frac{1}{|T|} \sum_{x_i^t \in T} L_w\left(x_i^t\right). \tag{6}$$

Substituting Eq. (5) into Eq. (6), the express of the average loss $\mathcal{L}_t(T)$ can be reformulated as:

$$\mathcal{L}_t(T) = \frac{1}{|T|} \sum_{x_i^t \in T} \exp\{-D_{JS}\left(x_i^t\right)\} \cdot L_{ce}\left(x_i^t\right) + \frac{1}{|T|} \sum_{x_i^t \in T} D_{JS}\left(x_i^t\right), \tag{7}$$

where $|T|$ is the number of images in the source dataset T.

Utilizing the labeled source scenes and the target scenes with pseudo-labels, we optimize the land cover mapping network with the following total loss:

$$\mathcal{L} = \mathcal{L}_s(S) + \lambda_3 \mathcal{L}_t(T), \tag{8}$$

where λ_3 is the weight corresponding to the average loss across the target dataset. It is important to note that the pseudo-labels are updated every fixed number of batches in the fine-tuning phase of the network.

4. Experiments

To illustrate the superiority of our method, this section details the experimental setups, comparison with recent land cover mapping UDA baselines, and the ablation studies of the proposed land cover mapping method.

4.1. Setups

4.1.1. Data sets and metrics

In order to illustrate the superiority of the proposed land cover mapping method, We compare the performance of our method with the recent UDA semantic segmentation methods on the LoveDA dataset (Wang et al., 2021b), a very recent and challenging dataset for both land cover mapping and UDA tasks. The LoveDA dataset has 5,987 high spatial resolution (HSR) samples comprising 166,768 annotated objects collected from three cities and 18 complex scenes in China. It contains two different domains, *Urban* and *Rural*.

The *Urban* areas always have more man-made objects like roads and buildings than the rural areas. In addition, the buildings in the urban scenes are densely distributed and of various shapes. The roads are wide, and water is often present in rivers or lakes adjacent to coastal urban areas. Compared with the urban scenes, the rural scenes typically

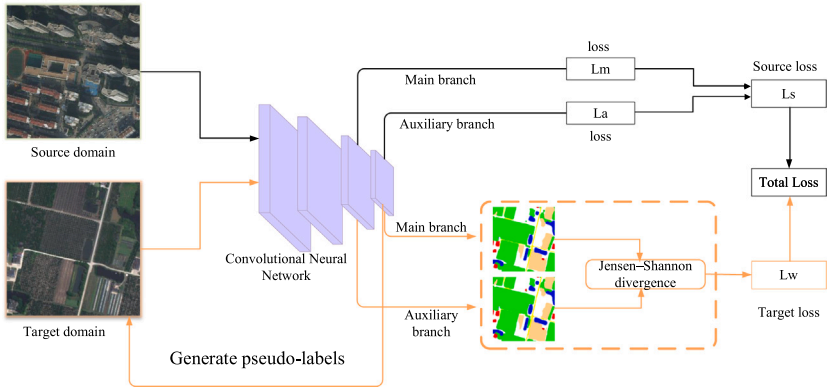


Fig. 2. The proposed self-training scheme using weighted pseudo-labels. The land cover mapping network trained utilizing images of the source scenes is employed to generate pseudo-labels for images of target scenes. The pseudo-labels are re-weighted using the Jensen–Shannon divergence between the land cover mapping network outputs during the second training stage. The more inconsistent the outputs of the two branches are, the smaller the weight of the corresponding pseudo-label is.

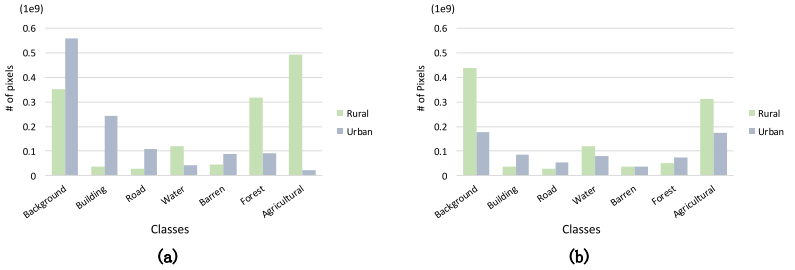


Fig. 3. Statistics for the pixels in the LoveDA dataset. (a) Number of pixels in the train datasets. (b) Number of pixels in the validation datasets.

have large homogeneous geographic areas, such as agricultural land and large water bodies, so the standard deviation of spectral statistics is lower than that of the rural scenes. As shown in Fig. 3, the LoveDA dataset has a very imbalanced class distribution. For both the *Urban* and the *Rural* scenarios, the *Road* and the *Barren* categories account for a very small proportion of the training and validation data.

There are two subtasks of cross-domain adaptation on the LoveDA dataset: (1) *Urban* → *Rural* and (2) *Rural* → *Urban*. The details of these two cross-domain subtasks can be summarized as follows.

- **Urban → Rural**: The source training set uses images collected from four cities, including Gulou,Jianghan, Qinhuai, and Qixia. The validation set used rural scene images from Huangpi and Liuhe. The test set comprises images from three cities, including Jiangning, Liyang and Xinbei. The training set, validation set and test set contain 1,156, 992 and 976 samples, respectively.
- **Rural → Urban**: The source training set uses images from the four cities, including Lishui, Pukou, Jiangxia and Gaochun areas. The validation set utilizes the images from Yuhuatai and Jintan areas. The test set uses images from three cities, including Jiangye, Wujin and Wuchang areas. The sizes of the training, validation and test sets are 1,366, 677 and 820, respectively.

The ground truths for the test sets of both scenarios are not publicly available. To obtain the test scores, we need to upload the land cover mapping results to the server associated with the LoveDA dataset.[2] To evaluate the proposed land cover mapping approach and the baselines, the server utilizes the per-class IoU and mean IoU (mIOU) for all object categories as the land cover mapping evaluation metrics. In addition to IoU, we also use other metrics to validate the effectiveness of each key component of our scheme on the validation sets of the two UDA scenarios. These metrics include producer's accuracy, user's accuracy, overall accuracy (OA) and kappa.

[2] https://codalab.lisn.upsaclay.fr/competitions/421

4.1.2. Implementation details

The proposed land cover mapping method is implemented employing the most popular open-source deep learning framework PyTorch. We train the land cover mapping network utilizing an NVIDIA Tesla V100 32 Gbs GPU. The proposed method uses the backbone ConvNeXt pre-trained on the ImageNet-22K dataset, which comprises 14.2 million samples in 22 K categories. The values of λ_1, λ_2, and λ_3 are set to 0.4, 0.5, and 0.5, respectively. The model's parameters are optimized using the AdamW optimizer (Loshchilov and Hutter, 2018) and the weight decay is set to 0.005 during the two training phases of the self-training scheme. We train the framework with the input size of 512×512, using a batch size of 8. We set the training iteration number to 10 K. For the first and second training stages, the values of the initial learning rates are set to 1e-4 and 8e-5, respectively. The predictions of the main branches are utilized as the pseudo-labels for the target scenes. The pseudo-labels are updated every 2000 batches during the second training stage. The two-stage optimization process takes around 6.5 h for either scenario. We only use the prediction from the main branch of the proposed land cover mapping network during the testing stage.

4.1.3. Baselines

Following Wang et al. (2021b), we adopt a series of advanced UDA algorithms to compare with the proposed UDA scheme for the land cover mapping task. In addition to MCD (Tzeng et al., 2014) (a domain confusion metric based approach), two types of UDA methods are evaluated: adversarial training (AdaptSeg (Tsai et al., 2018), TransNorm (Wang et al., 2019), CLAN (Luo et al., 2019), FADA (Wang et al., 2020b)) and self-training (CBST (Zou et al., 2018), PyCDA (Lian et al., 2019), IAST (Mei et al., 2020)).

4.2. Comparison results

We adopt the same experimental setting as Wang et al. (2021b) and directly report the results of these comparison algorithms from Wang et al. (2021b). Table 1 presents the comparison results on the LoveDA

Table 1
The comparison results on the test set. The abbreviations AT and ST denote adversarial training and self-training methods, respectively.

Scenario	Approach	Type	IoU (%)							mIoU (%)
			Background	Building	Road	Water	Barren	Forest	Agriculture	
Rural → Urban	Oracle	–	51.8	59.46	65.28	85.51	15.89	42.55	42.26	51.82
	Source only	–	43.3	25.36	12.7	76.22	12.52	23.34	25.14	31.27
	MCD (Tzeng et al., 2014)	–	43.6	15.37	11.98	79.07	14.13	33.08	23.47	31.53
	AdaptSeg (Tsai et al., 2018)	AT	42.35	23.73	15.61	81.95	13.62	28.7	22.05	32.68
	FADA (Wang et al., 2020b)	AT	43.89	12.62	12.76	80.37	12.7	32.76	24.79	31.41
	CLAN (Luo et al., 2019)	AT	43.41	25.42	13.75	79.25	13.71	30.44	25.8	33.11
	TransNorm (Wang et al., 2019)	AT	38.37	5.04	3.75	80.83	14.19	33.99	17.91	27.73
	PyCDA (Lian et al., 2019)	ST	38.04	35.86	**45.51**	74.87	7.71	40.39	11.39	36.25
	CBST (Zou et al., 2018)	ST	48.37	46.1	35.79	80.05	19.18	29.69	30.05	41.32
	IAST (Mei et al., 2020)	ST	48.57	31.51	28.73	86.01	**20.29**	31.77	36.5	40.48
	Ours	ST	**50.2**	**49.5**	43.86	**86.9**	15.0	**42.67**	**42.51**	**47.23**
Urban → Rural	Oracle	–	34.06	59.44	41.6	71.55	19.84	50.12	65.97	48.94
	Source only	–	24.16	37.02	32.56	49.42	14	29.34	35.65	31.74
	MCD (Tzeng et al., 2014)	–	25.61	44.27	31.28	44.78	13.74	33.83	25.98	31.36
	AdaptSeg (Tsai et al., 2018)	AT	26.89	40.53	30.65	50.09	16.97	32.51	28.25	32.27
	FADA (Wang et al., 2020b)	AT	24.39	32.97	25.61	47.59	15.34	34.35	20.29	28.65
	CLAN (Luo et al., 2019)	AT	22.93	44.78	25.99	46.81	10.54	37.21	34.45	30.39
	TransNorm (Wang et al., 2019)	AT	19.39	36.3	22.04	36.68	14	**40.62**	3.3	24.62
	PyCDA (Lian et al., 2019)	ST	12.36	38.11	20.45	57.16	**18.32**	36.71	41.9	32.14
	CBST (Zou et al., 2018)	ST	25.06	44.02	23.79	50.48	8.33	39.16	49.65	34.36
	IAST (Mei et al., 2020)	ST	29.97	49.48	28.29	64.49	2.13	33.36	**61.37**	38.44
	Ours	ST	**32.25**	**59.18**	**41.69**	**68.13**	16.23	38.69	57.99	**44.88**

test split. DeepLab (Chen et al., 2017) trained with only the labeled source scenes refers to the source-only setting. In terms of mIoU, the proposed approach outperforms the comparison algorithms by a large margin in both the *Rural → Urban* and the *Urban → Rural* experiments. In the *Rural → Urban* scenario, the proposed method obtains 47.23% mIoU. Compared with the non-domain-adaptive comparison method (i.e., the scheme only trained using the labeled source samples), the proposed method achieves a 15.96% mIoU improvement, outperforming the second-best approach by 5.91%. On the *Urban → Rural* scenario, our land cover mapping approach obtains a highest mIoU of 44.88%. Compared with the non-domain-adaptive comparison, our proposed method improves the mIoU by 13.14%, surpassing the second-best approach by 6.44%.

In terms of per-class IoU, our method also show an overall improvement over the other methods. On the *Urban → Rural* scenario, among the seven classes, we get the best performance in five classes: *Background*, *Building*, *Water*, *Forest*, and *Agriculture*. On the *Rural → Urban* scenario, among the seven classes, we get the best performance in four classes: *Background*, *Building*, *Road*, and *Water*. All algorithms perform poorly for the *Barren* category because the *Barren* category accounts for a very small proportion of the training data.

Moreover, we devise the Oracle setting to test the upper limit of our method's accuracy in a single domain. As presented in Table 1, on the *Urban → Urban* scenario, the Oracle setting outperforms the UDA methods for the *Road* and the *Building* categories by a wide margin. Specifically, the Oracle setting achieves IoUs of 59.46% and 65.25% for the *Building* and the *Forest* categories, respectively. For the *Rural* scenes, the *Road* and the *Barren* categories account for a very tiny proportion of the training data. In the *Rural → Rural* scenario, the Oracle setting obtains 50.12% IoU for the *Forest* category, outperforming the best UDA approach by 11.43%. For the *Urban* scenes, the *Forest* category accounts for a very small proportion of the training data.

We also evaluate the land cover mapping algorithms qualitatively. Since the ground truths are hidden for the test set of the LoveDA dataset, we implement the experiment on the validation split of the scene *Rural → Urban* scenario. Fig. 4 presents some qualitative examples of adapted land cover mapping for images randomly selected from the validation set. These results displayed are obtained from PyCDA, CBST, AdaptSeg, CLAN and our method. Comparing the ground truths with the predicted results, our proposed land cover mapping scheme shows significantly better results. In particular, the proposed method is more successful than the baseline methods at identifying buildings and forests.

From Table 1 and Fig. 4, we have three observations: (1) Compared with the adversarial training methods, the self-training methods achieve better performances on both scenarios. (2) Among the self-training methods, our way of exploiting pseudo-labels seems more effective. The proposed land mapping method achieves better mIoU than PyCDA, CBST, and IAST. (3) Imbalanced training data can impair the performance of UDA methods for rare categories.

4.3. Ablation study

This section presents ablation studies to validate the effectiveness of each key component of our scheme.

4.3.1. The effectiveness of multiscale structure

We investigate the effect of different multiscale structures, including DeepLabV2 (Chen et al., 2017) and UperNet. As opposed to DeepLabV2, UperNet fuses features from different stages. In both multiscale frameworks, we employ ConvNeXt (Liu et al., 2022) as the backbone and only use the source samples during the training phase. Table 2 presents the comparisons of land cover mapping on the test set of the LoveDA dataset. On the *Rural → Urban* scenario, DeepLabV2 and UperNet obtain 42.06% and 44.66% mIoU, respectively. On the *Urban → Rural* scenario, DeepLabV2 and UperNet obtain 39.79% and 43.34% mIoU, respectively. In terms of per-class IoU, UperNet also significantly outperforms DeepLabV2 across most categories for both UDA scenarios. Therefore, we conclude the proposed method based on UperNet can more efficiently extract multiscale features than methods based on DeepLabV2.

4.3.2. The effectiveness the auxiliary branch

We analyze the effect of the auxiliary branch, using only the source samples for training. As shown in Table 3, the land cover mapping models without and with the auxiliary branch obtains an mIoU of 43.66% and 44.66% on the *Rural → Urban* scenario, respectively. The auxiliary branch increase mIoU by 1%. on the *Urban → Rural* scenario, the land cover mapping model without and with the auxiliary branch achieve an mIoU of 42.81% and 43.40% on the *Rural → Urban* scenario, respectively. The auxiliary branch improves mIoU by 0.59%. The results of the two UDA scenarios in Table 3 consistently show that the auxiliary branch can effectively improve land cover mapping.

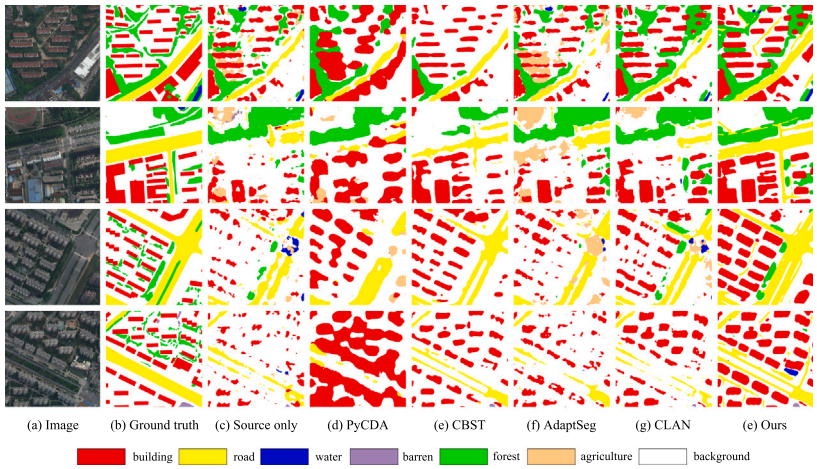


Fig. 4. Example results of adapted land cover mapping for images from the *Rural → Urban* scenario validation set. We show results obtained from Source-only, PyCDA, CBST, AdaptSeg, CLAN and our method for each image with ground truths from the validation set.

Table 2
Land cover mapping results achieved on the test split adopting various multiscale structures.

Scenario	Approach	IoU (%)							mIoU (%)
		Background	Building	Road	Water	Barren	Forest	Agriculture	
Rural → Urban	Deeplabv2	47.91	40.2	29.41	80.22	**14.79**	**42.83**	39.07	42.06
	UperNet	**48.65**	**45.03**	**40.57**	**84.31**	14.49	38.20	**41.4**	**44.66**
Urban → Rural	Deeplabv2	**29.52**	54.24	37.65	63.92	14.13	35.1	43.97	39.79
	UperNet	29.14	**60.12**	**39.69**	**66.14**	**17.26**	**39.59**	**51.84**	**43.40**

Table 3
The effectiveness of the auxiliary branch. The first and second rows of each scenario correspond to results without and with the auxiliary branches.

Scenario	Approach	IoU (%)							mIoU (%)
		Background	Building	Road	Water	Barren	Forest	Agriculture	
Rural → Urban	w/o aux	48.26	42.02	31.39	**84.64**	**17.68**	**42.41**	39.23	43.66
	w/ aux	**48.65**	**45.03**	**40.57**	84.31	14.49	38.20	**41.4**	**44.66**
Urban → Rural	w/o aux	**30.19**	**60.41**	36.29	62.72	**19.43**	37.46	**53.14**	42.81
	w/ aux	29.14	60.12	**39.69**	**66.14**	17.26	**39.59**	51.84	**43.40**

Table 4
The effectiveness of self-training. For each scenario, the first row corresponds to the results without adaptation. The second and third rows correspond to the results with adaption using the fixed threshold and the proposed weighted pseudo-labels, respectively.

Scenario	Approach	IoU (%)							mIoU (%)	Gain (%)
		Background	Building	Road	Water	Barren	Forest	Agriculture		
Rural → Urban	w/o adaptation	48.65	45.03	40.57	84.31	14.49	38.20	41.4	44.66	–
	Self-training w/ Tr	49.39	49.14	**47.14**	86.05	12.44	41.69	40.5	46.62	+1.96
	Ours	**50.2**	**49.5**	43.86	**86.9**	**15.0**	**42.67**	**42.51**	**47.23**	+2.57
Urban → Rural	w/o adaptation	29.14	**60.12**	39.69	66.14	17.26	**39.59**	51.84	43.40	–
	Self-training w/ Tr	30.28	56.99	41.32	62.86	**19.46**	39.48	51.74	43.16	−0.24
	Ours	**32.25**	59.18	**41.69**	**68.13**	16.23	38.69	**57.99**	**44.88**	+1.48

4.3.3. The effectiveness of self-training

We also analyze the usefulness of the proposed self-training using weighted pseudo-labels. The baselines of this section are the proposed land cover mapping network without domain adaptation and the proposed scheme with a fixed threshold. As indicated in Table 4, the proposed land cover mapping method with weighted pseudo-labels performs best on both scenarios. On the *Urban → Rural* scenario, the proposed land cover mapping network without domain adaptation obtains 44.66% mIoU. The proposed scheme with the fixed threshold achieves 46.62% mIoU, gaining 1.96% mIoU over the source-only model. The proposed pseudo-label learning scheme achieves 47.23% mIoU, improving the mIoU by 2.57% compared with the source-only model. On the *Rural → Urban* scenario, the proposed land cover mapping network without domain adaptation obtains 43.40% mIoU. The proposed scheme with the fixed threshold achieves 43.10% mIoU, 0.24% less than the source-only model. The proposed pseudo-label learning scheme achieves 44.88% mIoU, improving the mIoU by 1.48% compared with the source-only model. As can be seen from Table 4, pseudo-learning using a fixed threshold may compromise model performance. A fixed threshold will likely draw pseudo-labels corresponding to hard samples while filtering out noise. The proposed UDA scheme based on weighted pseudo-labels can effectively leverage pseudo-labels to bridge domain gaps.

4.3.4. Evaluation using multiple metrics

In addition to IoU, we also validate the effectiveness of each key component of our scheme on the validation sets, using other metrics including producer's accuracy, user's accuracy, overall accuracy(OA) and kappa. As presented in Table 5, in terms of OA and kappa, the auxiliary branch and the multiscale structure can effectively improve the

Table 5
Summary of the producer's accuracy, user's accuracy, overall accuracy (OA) and kappa on the validation sets. For each scenario, the first row corresponds to the results with DeeplabV2 as the backbone. The second and the third rows correspond to the results without and with the auxiliary branches, respectively. The fourth and the fifth rows corresponds to the results with adaption using the fixed threshold and the proposed weighted pseudo-labels, respectively.

Scenario	Method	User's accuracy (%)							Producer's accuracy (%)							OA (%)	Kappa (%)
		Background	Building	Road	Water	Barren	Forest	Agriculture	Background	Building	Road	Water	Barren	Forest	Agriculture		
Rural → Urban	DeeplabV2	49.84	**74.14**	**77.65**	88.88	76.18	64.51	88.63	**72.36**	69.25	61.24	79.74	50.86	71.69	62.91	68.31	60.8
	w/o aux	50.23	70.59	79.3	89.98	73.15	66.64	89.14	69.09	79.78	61.25	80.94	**53.86**	**77.24**	58.89	68.65	61.45
	w/ aux	51.17	72.98	70.25	88.6	72.01	68.05	86.66	67.19	78	68.78	82.5	50.45	74.35	61.73	68.93	61.84
	Self-training w/Tr	44.85	71.07	73.72	**91.44**	**76.84**	**75.62**	**92.96**	71.64	**82.48**	**71**	79.99	51	62.14	49.59	66.11	58.13
	Ours	**52.91**	72.81	76.9	90.01	72.29	69.19	87.29	67.89	81.11	68.61	**82.72**	51.96	74.43	**65.97**	**70.70**	**63.97**
Urban → Rural	DeeplabV2	60.04	68.72	70.75	86.28	24.81	50.23	90.05	87.37	63.05	**46.37**	64.46	20.48	40.8	49.36	66.36	49.52
	w/o aux	59.53	73.18	84.42	**93.1**	**34.73**	**70.84**	88.61	**94.78**	59.69	42.75	57.76	11.54	19.19	**51.15**	67.67	49.69
	w/ aux	60.37	69.79	78.97	86.62	23.2	59.78	92.66	88.61	69.22	45.99	70.6	24.23	35.61	47.38	67.09	50.66
	Self-training w/Tr	**63.1**	63.88	**88.66**	91.2	24.18	70.58	**95.99**	84.86	**79.79**	42.69	69.22	**71.37**	**46.91**	42.66	66.47	51.65
	Ours	60.89	**75.40**	82.94	89.82	23.87	68.27	90.66	90.06	67.98	42.94	**70.71**	22.27	35.15	49.81	**68.25**	**52.05**

performance of our method. In terms of the producer's and the user's accuracy, our method can achieve reliable results for most categories except rare ones such as the *Road* and the *Barren* categories. Therefore, how to effectively improve the performance of the UDA models for rare categories is worthy of further study.

5. Conclusion

In order to facilitate domain adaptation and reduce potential labeling costs, we promote adaptivity in two ways: (1) by designing a land cover mapping network with high performance and (2) by making full use of pseudo-labels. This paper presents a UDA framework based on an FPN and self-training for the land cover mapping task. We integrate ConvNeXt with a PPM in the FPN. Combining the FPN and PPM improves the segmentation performance, which benefits from the multiscale aggregation of pyramid features. The UDA consists of two training stages, which share the same land cover mapping network. In the first training phase, we train the network utilizing the labeled source scenes and generate pseudo labels for the target scenes. The source ground-truths and the weighted target pseudo-labels are utilized during the second stage to fine-tune the network. The proposed land cover mapping framework benefits from the multiscale aggregation of pyramid features and the full use of pseudo-labels. Comparison results on the LoveDA dataset, the latest large-scale unsupervised domain adaptation dataset for land cover mapping, empirically indicated that our land cover mapping approach is significantly superior to the baselines in both UDA scenarios. In some practical applications, a small number of samples in the target domain are available. How to make full use of the labeled samples in the target domain is a problem worthy of further studies. In the future, we will investigate semi-supervised domain adaptation algorithms for land cover mapping.

CRediT authorship contribution statement

Wei Liu: Conceptualization, Methodology, Writing – original draft. **Jiawei Liu:** Methodology, Software. **Zhipeng Luo:** Writing – review & editing. **Hongbin Zhang:** Data preparation. **Kyle Gao:** Writing – review & editing. **Jonathan Li:** Validation, Writing – review & editing.

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.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61662024, 62163016), the Natural Science Foundation of Jiangxi Province (No. 20212ACB202001), and Hong Kong Polytechnic University (Work Program: CD03).

References

Bengio, Y., Louradour, J., Collobert, R., Weston, J., 2009. Curriculum learning. In: International Conference on Machine Learning. pp. 41–48.
Benjdira, B., Bazi, Y., Koubaa, A., Ouni, K., 2019. Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images. Remote Sens. 11 (11), 1369.
Cai, Y., Yang, Y., Zheng, Q., Shen, Z., Shang, Y., Yin, J., Shi, Z., 2022. BiFDANet: Unsupervised bidirectional domain adaptation for semantic segmentation of remote sensing images. Remote Sens. 14 (1), 190.
Cascante-Bonilla, P., Tan, F., Qi, Y., Ordonez, V., 2021. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. In: AAAI, Vol. 35. (8), pp. 6912–6920.
Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848.
Chen, M., Xue, H., Cai, D., 2019. Domain adaptation for semantic segmentation with maximum squares loss. In: ICCV. pp. 2090–2099.
Choi, J., Jeong, M., Kim, T., Kim, C., 2019a. Pseudo-labeling curriculum for unsupervised domain adaptation. In: BMVC. pp. 1–13.
Choi, J., Kim, T., Kim, C., 2019b. Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation. In: ICCV. pp. 6830–6840.
Dai, D., Sakaridis, C., Hecker, S., Van Gool, L., 2020. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. Int. J. Comput. Vis. 128 (5), 1182–1204.
Deng, X., Yang, H.L., Makkar, N., Lunga, D., 2019. Large scale unsupervised domain adaptation of segmentation networks with adversarial learning. In: IGARSS. pp. 4955–4958.
Gu, X., Zhang, C., Shen, Q., Han, J., Angelov, P.P., Atkinson, P.M., 2022. A self-training hierarchical prototype-based ensemble framework for remote sensing scene classification. Inf. Fusion 80, 179–204.
Han, X., Huang, X., Li, J., Li, Y., Yang, M.Y., Gong, J., 2018. The edge-preservation multi-classifier relearning framework for the classification of high-resolution remotely sensed imagery. ISPRS J. Photogramm. Remote Sens. 138, 57–73.
Kang, G., Wei, Y., Yang, Y., Zhuang, Y., Hauptmann, A.G., 2020. Pixel-level cycle association: A new perspective for domain adaptive semantic segmentation. arXiv preprint arXiv:2011.00147.
Li, Y., Yuan, L., Vasconcelos, N., 2019. Bidirectional learning for domain adaptation of semantic segmentation. In: CVPR. pp. 6936–6945.
Lian, Q., Lv, F., Duan, L., Gong, B., 2019. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach. In: ICCV. pp. 6758–6767.
Liu, W., Luo, Z., Cai, Y., Yu, Y., Ke, Y., Junior, J.M., Gonçalves, W.N., Li, J., 2021. Adversarial unsupervised domain adaptation for 3D semantic segmentation with multi-modal learning. ISPRS J. Photogramm. Remote Sens. 176, 211–221.
Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A ConvNet for the 2020s. arXiv preprint arXiv:2201.03545.
Loshchilov, I., Hutter, F., 2018. Decoupled weight decay regularization. In: International Conference on Learning Representations.
Luo, Y., Zheng, L., Guan, T., Yu, J., Yang, Y., 2019. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In: CVPR. pp. 2507–2516.
Mei, K., Zhu, C., Zou, J., Zhang, S., 2020. Instance adaptive self-training for unsupervised domain adaptation. In: ECCV. pp. 415–430.
Pan, F., Shin, I., Rameau, F., Lee, S., Kweon, I.S., 2020. Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In: CVPR. pp. 3764–3773.
Saito, K., Kim, D., Sclaroff, S., Darrell, T., Saenko, K., 2019. Semi-supervised domain adaptation via minimax entropy. In: ICCV. pp. 8050–8058.
Shin, I., Woo, S., Pan, F., Kweon, I.S., 2020. Two-phase pseudo label densification for self-training based domain adaptation. In: ECCV. pp. 532–548.
Shu, Y., Cao, Z., Long, M., Wang, J., 2019. Transferable curriculum for weakly-supervised domain adaptation. In: AAAI, Vol. 33. (01), pp. 4951–4958.
Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L., 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. NeuraIPS.

Toldo, M., Maracani, A., Michieli, U., Zanuttigh, P., 2020. Unsupervised domain adaptation in semantic segmentation: A review. Technologies 8 (2), 35.

Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M., 2018. Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7472–7481.

Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T., 2014. Deep domain confusion: Maximizing for domain invariance. arXiv preprint arXiv:1412.3474.

Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P., 2019. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: CVPR. pp. 2517–2526.

Wang, J., HQ Ding, C., Chen, S., He, C., Luo, B., 2020a. Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label. Remote Sens. 12 (21), 3603.

Wang, X., Jin, Y., Long, M., Wang, J., Jordan, M.I., 2019. Transferable normalization: Towards improving transferability of deep neural networks. Adv. Neural Inf. Process. Syst. 32.

Wang, H., Li, H., Qian, W., Diao, W., Zhao, L., Zhang, J., Zhang, D., 2021a. Dynamic pseudo-label generation for weakly supervised object detection in remote sensing images. Remote Sens. 13 (8), 1461.

Wang, H., Shen, T., Zhang, W., Duan, L.Y., Mei, T., 2020b. Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation. In: European Conference on Computer Vision. Springer, pp. 642–659.

Wang, J., Zheng, Z., Ma, A., Lu, X., Zhong, Y., 2021b. LoveDA: A remote sensing land-cover dataset for domain adaptive semantic segmentation. In: Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2). URL https://openreview.net/forum?id=bLBIbVaGDu.

Wei, H., Ma, L., Liu, Y., Du, Q., 2021. Combining multiple classifiers for domain adaptation of remote sensing image classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 1832–1847.

Xiao, T., Liu, Y., Zhou, B., Jiang, Y., Sun, J., 2018. Unified perceptual parsing for scene understanding. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 418–434.

Xu, Q., Yuan, X., Ouyang, C., 2020. Class-aware domain adaptation for semantic segmentation of remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–17.

Yan, L., Fan, B., Xiang, S., Pan, C., 2021. CMT: Cross mean teacher unsupervised domain adaptation for VHR image semantic segmentation. IEEE Geosci. Remote Sens. Lett..

Yang, Y., Soatto, S., 2020. Fda: Fourier domain adaptation for semantic segmentation. In: CVPR. pp. 4085–4095.

Ye, F., Luo, W., Dong, M., He, H., Min, W., 2019. SAR image retrieval based on unsupervised domain adaptation and clustering. IEEE Geosci. Remote Sens. Lett. 16, 1482–1486.

Yu, T., Li, D., Yang, Y., Hospedales, T.M., Xiang, T., 2019. Robust person re-identification by modelling feature uncertainty. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 552–561.

Zhang, Y., David, P., Gong, B., 2017. Curriculum domain adaptation for semantic segmentation of urban scenes. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2020–2030.

Zhang, Z., Doi, K., Iwasaki, A., Xu, G., 2020. Unsupervised domain adaptation of high-resolution aerial images via correlation alignment and self training. IEEE Geosci. Remote Sens. Lett. 18 (4), 746–750.

Zhang, L., Lan, M., Zhang, J., Tao, D., 2021a. Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images. IEEE Trans. Geosci. Remote Sens. 1–13.

Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F., 2021b. Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In: CVPR. pp. 12414–12424.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2881–2890.

Zheng, Z., Yang, Y., 2021. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. Int. J. Comput. Vis. 129 (4), 1106–1120.

Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., Torralba, A., 2019. Semantic understanding of scenes through the ade20k dataset. Int. J. Comput. Vis. 127 (3), 302–321.

Zou, Y., Yu, Z., Kumar, B., Wang, J., 2018. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 289–305.

Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J., 2019. Confidence regularized self-training. In: ICCV. pp. 5982–5991.