Prototypical Unknown-Aware Multiview Consistency Learning for Open-Set Cross-Domain Remote Sensing Image Classification

Xiaokang Zhang, *Senior Member, IEEE,* Wanjing Wu, Mi Zhang, *Member, IEEE,* Weikang Yu, *Graduate Student Member, IEEE,* and Pedram Ghamisi, *Senior Member, IEEE*

***Abstract*—Developing a cross-domain classification model for remote sensing images has drawn significant attention in the literature. By leveraging the open-set Unsupervised Domain Adaptation (UDA) technique, the generalization performance of deep learning models has been improved with the capability to recognize unknown categories. However, it remains challenging to explore distribution patterns in the target domain using uncertain category-wise supervision from unlabeled datasets while reducing negative transfer caused by unknown samples. To develop a robust open-set UDA framework, this paper presents Prototypical Unknown-aware Multiview Consistency Learning (PUMCL) designed for remote sensing scene classification across heterogeneous domains. Specifically, it employs a consistency learning scheme with multiview and multilevel perturbations to improve feature learning from unlabeled target samples. An entropy separation strategy is utilized to facilitate open-set detection and recognition during adaptation, enabling unknown-aware feature alignment. Furthermore, the introduction of prototypical constraints optimizes pseudo-label generation through online denoising and promotes a compact category-wise feature subspace for improved class separation across domains. Experiments conducted on six cross-domain scenarios using AID, NWPU, and UCMD datasets demonstrate the method's superior performance compared to nine state-of-the-art approaches, achieving a gain of 4.5% to 21.2% in mIoU. More importantly, it shows promising class separability with clear boundaries between different classes and compact clustering of unknown samples in the feature space. The source code will be available at https://github.com/zxk688.**

***Index Terms*—Open-set, unsupervised domain adaptation, scene classification, remote sensing images, consistency learning.**

I. Introduction

REMOTE sensing image classification, as a fundamental task in Earth observation, has long been utilized for urban planning, environmental monitoring, and natural resource management [1]–[3]. The continuous advancement of sensors and platforms for Earth observation has resulted in abundant remote sensing data, presenting significant opportunities for remote sensing scene classification [4], [5]. In recent years, deep learning models have become primary tools for remote sensing image classification, owing to their exceptional ability to extract abstract features from remote sensing imagery data [6], [7]. Their excellent performances depend on the assumption that the training and test data adhere to be identically distributed.

However, in practical remote sensing applications, data shifts often arise from variations in acquisition conditions, scenes, seasons, and sensors. These differences result in performance degradation and biases when applying a trained model from the source domain to cross-domain classification tasks, as it cannot effectively characterize the representations of image data in the unseen domain [8]. Mitigating such training or methodological biases is identified as one of the key components of responsible artificial intelligence practices [9]. Fortunately, the development of transferable deep models has gained increasing attention in the literature [10]. To mitigate the impact of domain shift, unsupervised domain adaptation (UDA), a common transfer learning technique, has been used to enhance the performance of deep models on unlabeled target domains [11], [12].

The objective of UDA is to minimize distribution inconsistency across various domains by projecting the source and target domain into a shared subspace or by aligning target samples with the corresponding source class centers [13]–[17]. Specifically, as a mainstream technique in UDA, adversarial training employs a dual-branch network architecture that jointly processes both source and target domain data. By alternatively training a feature generator and a domain discriminator in an adversarial manner, it reduces the feature distribution discrepancy between the two domains [18]. Although adversarial training can effectively match the global marginal distributions, it may not ensure the separation of different classes within the target domain. In contrast, the self-training-based approach iteratively refines the model by introducing pseudo-labels, thus explicitly providing category-wise supervision information for the target domain [19]. Self-training depends on high-confidence pseudo-labels, which are typically selected based on the model's predictive uncertainty [20]. However, pseudo-labels can introduce noise due to domain discrepancies during model optimization, resulting in

This work was supported in part by the National Natural Science Foundation of China under Grant 42371374 and Grant 41801323, and was also supported by the Open Fund of Hubei Luojia Laboratory. *(Corresponding author: Weikang Yu.)*

Xiaokang Zhang and Wanjing Wu are with the School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China (e-mail: natezhangxk@gmail.com; wuwanjing@wust.edu.cn).

Mi Zhang is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China and also with Hubei Luojia Laboratory, Wuhan 430079, China (e-mail: mizhang@whu.edu.cn).

Weikang Yu is with Helmholtz Institute Freiberg for Resource Technology, Helmholtz-Zentrum Dresden-Rossendorf, 09599 Freiberg, Germany. (e-mail: w.yu@hzdr.de)

Pedram Ghamisi is with Helmholtz Institute Freiberg for Resource Technology, Helmholtz-Zentrum Dresden-Rossendorf, 09599 Freiberg, Germany, and also with Lancaster University, LA1 4YR Lancaster, U.K. (e-mail: p.ghamisi@gmail.com).

imprecise category knowledge. This issue has been identified as the primary reason why a weakly supervised approach based on pseudo-labels cannot perform on par with an Oracle (or a fully supervised model) [21].

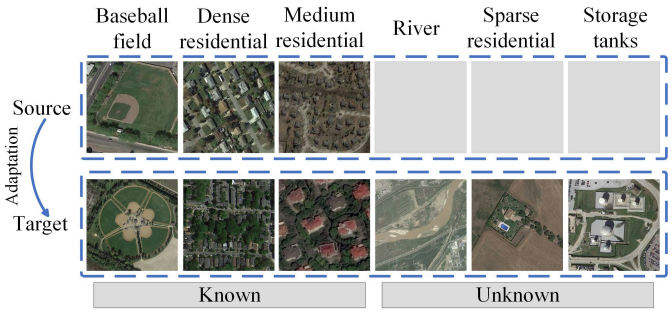


Fig. 1: Illustration of the open-set UDA problem, where the target domain contains “unknown” classes that are absent in the source domain.

Most existing works are developed assuming that the source and target domains share the same label space, i.e., in a closed-set scenario. In real-world applications, it is often impossible to guarantee that all incoming test data belong to existing classes from the source domain [22]. Therefore, developing models for open-world remote sensing is critically important [23]. Specifically, an open-set UDA setting has been introduced, which allows the target data to contain an additional ”unknown” category, covering all irrelevant classes not present in the source domain [24], [25], as shown in Fig. 1. Open-set UDA poses two key challenges. First, it is essential to accurately identify the decision boundary between known and unknown classes, even without accessible information about the unknown classes [26]. However, mitigating the influence of distribution shifts between domains is more difficult than in a closed-set scenario. Secondly, directly aligning the entire distribution of source and target domains can be risky due to the presence of unknown classes in the target domain, which may degrade the performance of the domain adaptation model—a phenomenon known as negative transfer. Therefore, open-set cross-domain classification algorithms should possess the ability to identify samples from unseen classes as ”unknown” while accurately classifying samples from known classes.

Unlike the closed-set UDA approach, unknown sample recognition strategies or modules are required in the open-set UDA, e.g., the auxiliary binary classifier to group the samples into known and unknown classes in adversarial learning [27], [28]. Alternatively, open-set recognition strategies are designed for unknown sample recognition based on prediction uncertainty, sample affinity, or distance measurement, which normally relies on predefined thresholds [29]–[31]. Since no prior knowledge is available for unknown classes, which may comprise a mixture of various land scene types, high-confidence pseudo-labels generated by the biased model cannot be considered inherently reliable in self-training [30]. Furthermore, to improve the feature representation capability of the target domain, the self-supervised-enhanced open-set UDA method has been developed by exploiting contrastive learning of unlabeled samples to mine the feature patterns [29]. While current research has yielded impressive performance within specific datasets, challenges arise when dealing with diverse scenarios. Uncertainties, including sample offsets, noisy pseudo-labels, semantic space disparities, and discrete feature distributions, collectively inhibit existing methods from providing reliable supervision information and category-wise knowledge of ground objects for model optimization.

Motivated by the abovementioned challenges, we attempt to propose a robust open-set cross-domain scene classification method, Prototypical Unknown-aware Multiview Consistency Learning (PUMCL), for remote sensing images. It employs a consistency learning strategy for both known and unknown samples with multiview and multilevel perturbations to reduce sample offsets between the two domains and minimize domain discrepancies. To address semantic space disparities caused by category inconsistency, unknown-aware feature alignment is applied in consistency learning through an entropy separation strategy with gradient reversal. This is achieved by alternately increasing the conditional entropy of known and unknown samples with respect to the classifier while minimizing discrepancies of known samples relative to the backbone. As a result, the decision boundary is shifted towards balancing the classification of both known and unknown samples. To reduce the impact of noisy pseudo-labels and discrete feature distributions on classification, prototypical constraints are constructed to reweight the generated pseudo-labels, mitigating the dispersion of feature distributions caused by domain discrepancy and shrinking category-wise subspace across domains for better class separation. The main contributions of this article are as follows:

1) A robust open-set UDA framework, PUMCL, based on consistency learning, is proposed. It fully exploits the category-wise cross-domain transferable knowledge and feature distribution patterns from the unlabeled samples without predefined thresholds. As a result, it achieves better performance than nine state-of-the-art methods in six groups of UDA experiments.
2) We construct multilevel consistency regularization strategies and incorporate entropy separation and prototypical constraints into consistency learning to enhance unknown-aware feature alignment and learn a shrinkable feature structure by introducing reliable category-wise pattern knowledge.

This article is organized as follows. In Section II, we briefly review related research on cross-domain remote sensing classification. In Section III, we detail the key components and training scheme of the proposed approach. After that, Section IV presents and analyzes the performance evaluations of our algorithm and the compared methods. Finally, Section V concludes the paper.

II. Related Work

A. Domain Alignment in Image Classification Tasks

Previous research on the cross-domain classification of remote sensing images can be divided into three categories: approaches based on divergence measurements, adversarial learning, and self-training. In the first category, feature distributions of the two domains are aligned by optimizing divergence metrics such as the maximum mean discrepancies (MMDs) [16] and conditional MMDs [32], [33], and per-class MMDs [34]. Adversarial-learning-based methods have emerged as the most developed type of UDA method in remote sensing. It generally includes a generator for representation learning and a discriminator for domain classification [35], [36]. This approach alternatively trains a feature generator and a domain discriminator in an adversarial manner, where both components engage in a game-theoretic process for feature alignment at the latent space [37], [38]. Various schemes have been developed to improve the discriminative capability of the adversarial learning models, e.g., using a weighted discriminator to perform sample instance reweighting and calibrate domain alignment according to samples' importance [39], [40]. Moreover, a fundamental theoretical framework has been established for UDA with imbalanced prediction consistency, significantly enhancing the network's discriminative ability [41]. Since a non-negligible gap exists, matching global distribution between the source and target domain cannot guarantee an improvement in class separability [42]. To address this problem, self-training-based methods have been used to improve adaptation performance by conducting category-wise feature alignment. The idea behind self-training is to use high-confidence pseudo-labels in the target domain to iteratively refine the model [43]. These pseudo-labels are typically selected using sorted output probabilities from the corresponding categories or thresholding methods based on the prediction confidence of the model. Although self-training can provide additional category-wise supervision, the generated pseudo-labels are inevitably noisy, which can mislead the classifier during backpropagation.

B. Open-set Cross-Domain Learning

Open-set adaption should exclusively focus on the known classes to avoid incorporating irrelevant information. Most methods are developed on the basis of adversarial learning and self-training in the closed-set UDA. Specifically, an adversarial learning-based approach using the backpropagation method was developed, utilizing a classifier and a feature generator to separate unknown targets from known samples [44]. [45] extended adversarial learning using multiple auxiliary classifiers with a weighting module to assess distinctive domain characteristics and indicate their likelihood of belonging to known or unknown classes. Furthermore, separate-to-adapt (STA) [46] performs adversarial domain alignment based on multi-binary classifiers and applies a coarse-to-fine weighting mechanism on the samples with an auxiliary classifier to distinguish the known samples from the unknown ones. Unknown-Aware Domain Adversarial Learning (UADAL) [27] explicitly segregates unknown features in domain adversarial learning by using entropy values as the open-set recognition indicator. In [31], an adaptive threshold segmentation method based on the Otsu algorithm is used to detect unknown samples in the testing set. In [47], an instance affinity matrix was proposed to distinguish unknown samples by increasing the affinity between similar samples across domains and reducing the affinity of irrelevant classes. Uncertainty distance measurement was proposed to identify unknown classes [48]. Self-supervised-driven open-set UDA (SSOUDA) utilized a contrastive learning strategy to effectively extract the structural knowledge of the target domain [29]. However, three trial-and-error thresholds are required to obtain reliable pseudo-labels. More recently, the multi-adversarial Open-Set Domain Adaptation Network (MAOSDAN) [28] utilized auxiliary adversarial learning to mitigate the negative transfer effect of unknown samples in network training. Another line of work to address the cross-domain classification issue is zero-shot learning [8], [49]. Although these models can recognize and even classify unknown categories, they still require textual descriptions of known-class images for assistance during both training and testing [8].

Overall, it is still challenging to retain low category-wise distribution discrepancy across different domains while achieving high interclass similarity for image classification [26]. Although the contrastive self-supervised learning-based approach can improve feature discriminative capability [50], the absence of reliable category-wise supervision could hinder the exploration of distribution patterns, leading to weak robustness.

III. Methodology

As shown in Fig. 2, the proposed PUMCL framework consists of three parts: constructing weak-to-strong views, unknown-aware consistency learning, and prototypical constraints. Specifically, the multiview consistency learning part employs consistency regularization to generate domain-invariant representations under various augmentations to mine invariant representations of the target domain. Furthermore, it leverages entropy separation to detect and recognize unknown samples for unknown-aware domain alignment, in which a gradient reverse layer (GRL) is utilized to separate the known and unknown feature spaces in an adversarial learning manner. Finally, the prototypical constraint module calculates weights based on the distance between prototypes and the sample feature within the target domain, adaptively reducing the noise in pseudo-labels used in consistency learning. Additionally, it shrinks the category subspace to enhance class separability.

A. Preliminaries

In this study, there exists a labeled source domain dataset $\mathcal{D}_s = \{x_{s,i}, y_{s,i}\}_{i=1}^{n_s}$, where $x_{s,i}$ represents a randomly sampled instance from the source domain and $y_{s,i}$ is the corresponding label, with a total of n_s images in the dataset. Also, we have an unlabeled target domain dataset $\mathcal{D}_t = \{x_{t,j}\}_{j=1}^{n_t}$, where $x_{t,j}$ represents a randomly sampled instance from the target domain without labels, with a total of n_t images. In open-set UDA scenarios, the label set of the target domain

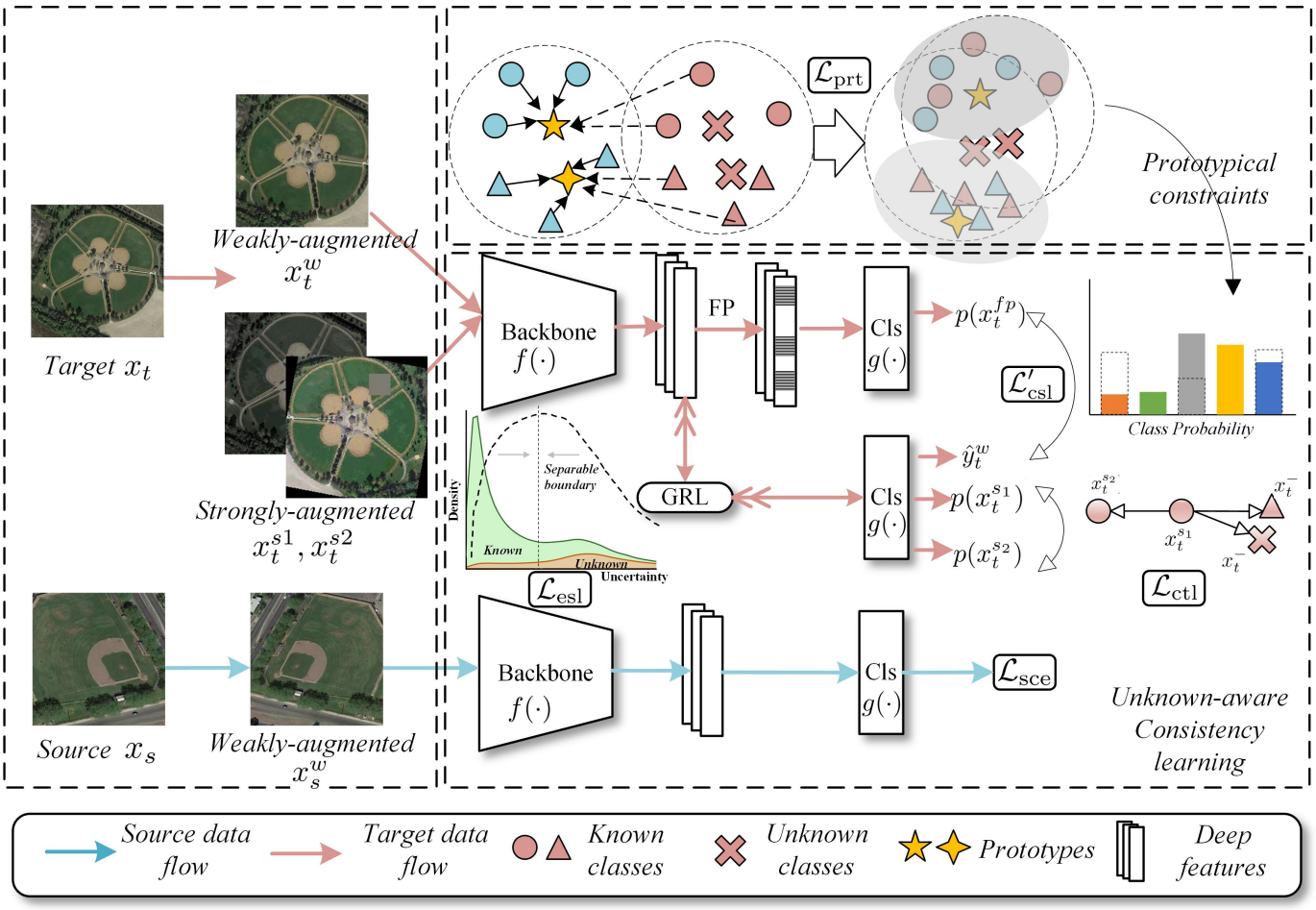


Fig. 2: Overall framework of PUMCL. Specifically, FP means feature perturbations, and GRL means gradient reverse layer.

$\mathcal{C}_t = \{1, 2, \ldots, C\}$ is larger than that of the source domain $\mathcal{C}_s = \{1, 2, \ldots, C-u\}$, i.e., $\mathcal{C}_s \subset \mathcal{C}_t$, where u is the number of the extra classes that do not exist in the source domain and are uniformly designated as unknown class C_u in the target domain.

The cross-domain classification network in the proposed framework consists of a backbone as feature extractor $f(\cdot)$ that extracts useful representations from the input data and a classifier $g(\cdot)$ that projects the representations into different classes. The objective is to guide the network in learning the category-wise knowledge of the target domain under unsupervised conditions while enabling it to accurately distinguish between known and unknown classes. As the labeled source data is the only supervised information available, the source cross-entropy loss $\mathcal{L}_{\text{sce}}$ is utilized to minimize the discrepancy between the classifier's predictions and the labels, which is defined as follows:

$$\mathcal{L}_{\text{sce}} = -\frac{1}{n_s} \sum_{i=1}^{n_s} \sum_{c \in \mathcal{C}_s} y_{s,i} \log \left(p\left(\hat{y} = c \mid x_{s,i}\right) \right) \tag{1}$$

where $y_{s,i}$ represents the true label of source domain image $x_{s,i}$, and $p\left(\hat{y} = c \mid x_{s,i}\right)$ represents the predicted conditional probability distribution for the sample $x_{s,i}$ when the predicted label $\hat{y}$ belongs to the c-th class. Notably, we utilize the weakly augmented version $x_{s,i}^w$ instead of the original image $x_{s,i}$ in the training process.

B. Multiveiw Consistency Learning

1) Consistency Regularization: Consistency regularization leverages unlabeled data by assuming that the model should produce similar predictions when presented with perturbed versions of the same image. This technique has proven effective for exploiting unlabeled data under the cluster assumption, where the decision boundary lies in low-density regions. The goal of consistency learning in UDA is to enhance the robustness of the model by learning invariant representations of the target domain samples under various augmentations. Traditional consistency regularization strategies focus on minimizing the cross-entropy between the class predictions of weakly augmented samples and strongly augmented samples. However, this approach primarily enhances discriminative features within a limited perturbation space. Inspired by the UniMatch method [51], we combined image-level and feature-level perturbations in independent streams to expand the perturbation space and obtain more robust features. Given B target domain samples in a mini-batch, each subjected to one weak augmentation $x_{t,j}^w$, based on which feature perturbation $x_{t,j}^{fp}$ is performed. Additionally, two views of strong

augmentations $x_{t,j}^{s_1}$ and $x_{t,j}^{s_2}$, generating a total of $4B$ mixed augmentation samples with each offering a unique perspective for representation learning. Specifically, weak augmentations include image flipping and random cropping, while strong augmentations are based on color transformations [52]. With respect to feature perturbation, channel dropout is adopted in this paper, which has been demonstrated sufficiently effective [51].

Instead of minimizing the divergence between the probability distributions obtained by various perturbations [51], we exploit the pseudo-labels $\hat{y}_{t,j}^{w}$ generated by the weakly augmented samples $x_{t,j}^{w}$ to explicitly guide the consistency learning. Thus, the consistency loss $\mathcal{L}_{\text{csl}}$ is expressed as:

$$\begin{aligned}\mathcal{L}_{\text{csl}} = &\frac{1}{B}\sum^{B}\mu\mathcal{H}(\hat{y}_{t,j}^{w}, p(x_{t,j}^{fp})) \\ &+\frac{\mu}{2}\left(\mathcal{H}\left(\hat{y}_{t,j}^{w}, p(x_{t,j}^{s_1})\right)+\mathcal{H}\left(\hat{y}_{t,j}^{w}, p(x_{t,j}^{s_2})\right)\right)\end{aligned} \tag{2}$$

where $\mathcal{H}(\cdot)$ represents the cross-entropy loss function between two probability distributions. It is essentially an average of three loss functions, where the weight μ is set to 0.5.

2) Contrastive Enhancement: To make the features more discriminative, we introduce contrastive enhancement in a self-supervised manner. Denoted by $f(x_{t,j}^{s_1})$ and $f(x_{t,j}^{s_2})$, the projected features of paired augmented samples $x_{t,j}^{s_1}$ and $x_{t,j}^{s_2}$ through the feature extractor $f(\cdot)$. The goal of contrastive enhancement is to maximize the similarity between the strongly augmented samples $x_{t,j}^{s_1}$ and $x_{t,j}^{s_2}$ of the same data example while minimizing the resemblance between samples from another view with different semantics which is regarded as the negative sample $x_{t,l,l\neq j}^{-}$.

Inspired by [53], [54], we adopt a simple and effective loss function to maximize the similarity between features of positive pairs while minimizing their similarity to the features of negative samples, thereby learning a more discriminative representation. Instead of using the contrast between weak and strong samples to maximize feature similarity, we construct positive and negative samples from two strong versions of the samples. As a result, the contrastive enhancement loss function $\mathcal{L}_{ctl}$ is defined as:

$$\mathcal{L}_{ctl} = -\log\frac{\exp\left(\text{sim}\left(f(x_{t,j}^{s_1}), f(x_{t,j}^{s_2})\right)/\tau\right)}{\sum_{l=1}^{2B}\mathbb{I}(l\neq j)\exp\left(\text{sim}\left(f(x_{t,j}^{s_1}), f(x_{t,l}^{-})\right)/\tau\right)} \tag{3}$$

where $\text{sim}(\cdot)$ is the cosine similarity metric function, τ is the temperature parameter, and $\mathbb{I}\in\{0,1\}$ is an indicator function to evaluate whether l is equal to j.

By exploiting consistency learning with multilevel perturbations and contrastive enhancement, the model not only learns to generate invariant representations under various augmentations but also ensures that these representations are sufficiently discriminative. This dual focus on unlabeled data promotes the extraction of more robust and generalizable features, thereby improving the model's ability to distinguish samples in the UDA context. The proposed consistency learning offers a foundational framework for unknown-aware domain alignment and open-set recognition. Notably, pattern knowledge of unknown samples can be acquired and integrated into consistency learning, as will be elaborated in Section III.D.

C. Prototypical Constraints

Although useful features of the target domain can be extracted through consistency learning, pseudo-labels $\hat{y}_{t,j}^{w}$ inevitably introduce noise in the training process and the absence of reliable supervision information hinders the class separability. Therefore, we utilize the prototypical constraints to further refine noisy pseudo-labels in the target domain and improve class separability.

1) Shrinkable Subspace Optimization: As new data emerges, the unknown class typically consists of multiple distinct classes with potentially diverse feature distributions. Therefore, it is crucial to obtain intra-class compactness and learn a clustered feature subspace. The source class prototype is a representative feature vector for each class, initialized with the source domain label:

$$\varphi_s^c = \frac{\sum_{(x_{s,i}^{w}, y_{s,i}^{w})\in\mathcal{D}_s}\sum_i \mathbb{I}\left(y_{s,i}^{w}=c\right) f\left(x_{s,i}^{w}\right)}{\sum_{y_{s,i}^{w}\in\mathcal{D}_s}\sum_i \mathbb{I}\left(y_{s,i}^{w}=c\right)}, c\in\mathcal{C}_s \tag{4}$$

where $\{\varphi_s^c\}_{c=1}^{C-u}$ represents the feature prototype set. However, calculating these prototypes is computationally intensive during training. To mitigate this, we estimate the prototypes as the moving average of cluster centroids within mini-batches, allowing us to track gradually evolving prototypes. The exponential moving average strategy throughout the training process can be expressed as follows:

$$\varphi_s^c \leftarrow \lambda_{\text{ema}}\varphi_s^c + (1-\lambda_{\text{ema}})\,\varphi_{\text{batch},s}^c \tag{5}$$

where φ_s^c denotes the prototype for class c, $\varphi_{\text{batch},s}^c$ represents the mean feature for class c calculated within the current training batch, and λ_{ema} is momentum coefficient which is set to 0.5.

We leverage the prototypical information of features to constrain each category subspace intra-domain, simultaneously constructing classifiers' decision boundaries. This process dynamically evaluates the sample-level domain discrepancy inter-domain. Since the labels for the source samples are available and can serve as reliable supervision for domain alignment, source prototypes are utilized to explicitly guide the target samples toward the source category centroids, enhancing the compactness of the feature space within the same classes. Specifically, we utilize prototypical loss, $\mathcal{L}_{prt}$, to minimize the Euclidean distances between target features and prototypes in the source domain:

$$\mathcal{L}_{prt} = \frac{1}{n_s}\sum_{i=1}^{n_s}\min_{c\in\mathcal{C}_s}\left\{\frac{\exp\left(-\left\|f\left(x_{t,i}^{w}\right)-\varphi_s^c\right\|\right)}{\sum_{c'}\exp\left(-\left\|f\left(x_{t,i}^{w}\right)-\varphi_s^{c'}\right\|\right)}\right\}. \tag{6}$$

2) Online Pseudo-Label Denoising: The commonly used pseudo-label denoising method relies on a threshold to select the high-confidence pseudo-labels while discarding those of low confidence, which lacks flexibility and global distribution patterns of the target domain. To fully exploit the knowledge from unlabeled samples, we propose an online denoising strategy for the pseudo-labels. As the model converges, the

unlabeled samples will tend to align with the category-wise feature centroids. Therefore, the target prototypes can be used as a guide for pseudo-label denoising. Meanwhile, the pseudo-labeled samples of the unknown class can be obtained and refined to enhance unknown-aware domain alignment. Specifically, the proposed approach utilizes the feature centroid of weakly augmented target samples as prototypes to incrementally refine pseudo-labels, thereby dynamically mitigating the impact of unreliable labels. The category-wise weights are determined by the distances between the target domain sample features and the target prototypes of each class $\{\varphi_t^c\}_{c=1}^{C_u}$ within the target domain, as follows:

$$\varphi_t^c = \frac{\sum_{x_{t,j}^w \in \mathcal{D}_t} \sum_j \mathbb{I}\left(\Phi(p^c(x_{t,j}^w)) = 1\right) f\left(x_{t,j}^w\right)}{\sum_{x_{t,j}^w \in \mathcal{D}_t} \sum_j \mathbb{I}\left(\Phi(p^c(x_{t,j}^w)) = 1\right)}, c \in \mathcal{C}_t \quad (7)$$

$$\omega_{t,j}^c = \frac{\exp\left(-\left\|f\left(x_{t,j}^w\right) - \varphi_t^c\right\|\right)}{\sum_{c'} \exp\left(-\left\|f\left(x_{t,j}^w\right) - \varphi_t^{c'}\right\|\right)}, c \in \mathcal{C}_t \quad (8)$$

where $\Phi(\cdot)$ denotes the conversion from the soft predictions to hard labels, $\omega_{t,j}^c$ represents the weight of sample $x_{t,j}^w$ belonging to the c-th class. It can be observed that samples close to the feature centroid can be regarded as the more trustworthy ones, while samples far away from the prototypes are likely outliers. Accordingly, we adjust the classification probabilities incrementally based on this premise. Through progressively updating the weights based on the newly acquired knowledge, the refined pseudo-labels can be obtained by:

$$\hat{y}_{t,j}^{w,c} = \Phi(\omega_{t,j}^c p^c(x_{t,j}^w)) = \begin{cases} 1, & \text{if } c = \arg\max_{c'} \omega_{t,j}^{c'} p^{c'}(x_{t,j}^w) \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where $\hat{y}_{t,j}^{w,c}$ means the one-hot label of the sample $x_{t,j}^w$ denoting whether it belongs to the c-th class.

D. Unknown-Aware Domain Alignment

1) Unknown and Known Class Separation: Normally, the "known" region exhibits a higher density of occurrences, which diminishes as predictive uncertainty increases. Conversely, the "unknown" region shows a lower density that rises with increasing uncertainty. This inverse relationship highlights the challenge of obtaining reliable information in the face of growing uncertainty, as shown in Fig. 2. Since the training process aims to fit source samples by minimizing the cross-entropy losses, the model tends to classify target samples into known classes, leading to a biased decision boundary and often failing to effectively detect unknown classes [44]. Additionally, the unknown class might consist of multiple land cover types with spectral characteristics similar to those of the known classes. As reported in [55], the classifier predicts higher entropy for unknown target samples and lower entropy for known samples because no semantics of unknown-class samples are available during the training process. Based on this observation, we attempt to increase the entropy values of the known and unknown categories conditioned on each other to find a more separable decision boundary between the two. Meanwhile, the consistency loss and prototypical constraints will incorporate category-wise target domain knowledge to guide the model in fitting both known and unknown samples, thereby enhancing its ability to distinguish between these classes.

To achieve the above purpose, we apply an entropy separation objective $\mathcal{L}_{\text{esl}}$ for the split of known and unknown feature space in an adversarial learning manner. It is implemented by alternately increasing the conditional entropy of known and unknown samples with respect to the classifier while minimizing discrepancies of known samples with respect to the feature encoder. Specifically, a GRL is utilized during backpropagation to implicitly conduct adversarial training. Finally, the entropy separation loss $\mathcal{L}_{\text{esl}}$ is designed as follows:

$$\begin{aligned} \mathcal{L}_{\text{esl}} = &- \frac{1}{n_t} \sum_{j=1}^{n_t} p\left(\hat{y} = C_u \mid x_{t,j}^w\right) \log\left(1 - p\left(\hat{y} = C_u \mid x_{t,j}^w\right)\right) \\ &- \frac{1}{n_t} \sum_{j=1}^{n_t} \left(1 - p\left(\hat{y} = C_u \mid x_{t,j}^w\right)\right) \log\left(p\left(\hat{y} = C_u \mid x_{t,j}^w\right)\right) \end{aligned} \quad (10)$$

where $p\left(\hat{y} = C_u \mid x_{t,j}^w\right)$ represents the probability of belonging to the unknown class C_u.

2) Unknown-Aware Class Recognition: Notably, the consistency learning in Equation (2) is designed for the closed-set UDA in which the source and target domains share label spaces. For the open-set UDA task, we integrate unknown-aware class recognition into consistency learning before pseudo-labeling with prototypical constraints in Equation (9) is used to refine the classifier's decision boundary, generating more reliable predictions $\hat{y}_{t,j}^w$. More specifically, an indicator function $\xi(\cdot)$ to identify whether a sample belongs to known or unknown, which is defined as:

$$\xi(\hat{y}_{t,j}^w) = \begin{cases} 1, & \text{if } \hat{y}_{t,j}^w \in \{1, 2, \cdots, C-u\} \\ 0, & \text{otherwise} \end{cases}. \quad (11)$$

Therefore, the final consistency loss, modified from Equation (2), is defined as:

$$\begin{aligned} \mathcal{L}'_{\text{csl}} = &\frac{1}{\gamma_1 B} \sum^{B} \xi(\hat{y}_{t,j}^w) \\ &\cdot \left[\mu \mathcal{H}(\hat{y}_{t,j}^w, p(x_{t,j}^{fp})) + \frac{\mu}{2} \mathcal{H}(\hat{y}_{t,j}^w, p(x_{t,j}^{s_1})) + \mathcal{H}(\hat{y}_{t,j}^w, p(x_{t,j}^{s_2}))\right] \\ &+ \frac{1}{\gamma_2 B} \sum^{B} (1 - \xi(\hat{y}_{t,j}^w)) \\ &\cdot \left[\mu \mathcal{H}(1, p(x_{t,j}^{fp})) + \frac{\mu}{2} (\mathcal{H}(1, p(x_{t,j}^{s_1})) + \mathcal{H}(1, p(x_{t,j}^{s_2})))\right] \end{aligned} \quad (12)$$

where γ_1 and γ_2 represent the proportions of known and unknown samples selected within a mini-batch.

E. Training and inference

The final objective function is to learn the optimal θ_f and θ_c of the feature extractor $f(\cdot)$ and classifier $g(\cdot)$, and the total loss function is as follows:

$$\arg\min_{\theta_f, \theta_c} \left(\mathcal{L}_{\text{sce}} + \lambda(\mathcal{L}'_{\text{csl}} + \mathcal{L}_{\text{ctl}}) + \mathcal{L}_{\text{prt}} + \arg\max_{\theta_f, \theta_c} \mathcal{L}_{\text{esl}} \right) \quad (13)$$

Table I: Details of three publicly available remote sensing datasets used in this paper.

Dataset	Classes	Images per class	Total images	Resolution (m)	Image size
UCMD	21	100	2,100	0.3	256 × 256
NWPU	45	700	31,500	0.2∼30	256 × 256
AID	31	220∼420	10,000	0.5∼8	600 × 600

where λ is the weight coefficient of $\mathcal{L}'_{\text{csl}}$ and $\mathcal{L}_{\text{ctl}}$. The objective function is optimized with respect to the parameters θ_f and θ_c. Notably, we implicitly conduct the entropy separation through maximizing $\mathcal{L}_{\text{esl}}$ using a GRL. The training process is summarized in Algorithm 1.

In the inference process, only the weakly augmented test images are input to the well-trained feature extractor $f(\cdot)$ and classifier $g(\cdot)$ to obtain the target domain classification result.

Algorithm 1 Training Process of PUMCL

Require: source domain $\mathcal{D}_s$ and target domain $\mathcal{D}_t$, backbone $f(\cdot)$ and classifier $g(\cdot)$ with parameters θ_f and θ_g , weight λ, learning rate ε, maximum iterations K.
Output: networks $f(\cdot)$ and $g(\cdot)$.

initialize prototypes: $\varphi_s^c \leftarrow y_{s,i}^w$ and $f(x_{s,i}^w)$, $\varphi_t^c \leftarrow \Phi(p(x_{t,j}^w))$ and $f(x_{t,j}^w)$
for $k \leftarrow 0$ to $K-1$ **do**
  if $k < 5$ **then**
    warm-up:
    $\theta_f \leftarrow \theta_f - \varepsilon \nabla \theta_f \left(\mathcal{L}_{\text{sce}} + \lambda \mathcal{L}_{\text{ctl}}\right)$
    $\theta_g \leftarrow \theta_g - \varepsilon \nabla \theta_g \mathcal{L}_{\text{sce}}$
    return
  else
    update source and target prototypes: $\varphi_{s/t}^c \leftarrow \lambda_{\text{ema}} \varphi_{s/t}^c + (1-\lambda_{\text{ema}})\, \varphi_{\text{batch},s/t}^c$
    online denoising: $\omega_{t,j}^c \leftarrow f(x_{t,j}^w)$ and φ_t^c, $\hat{y}_{t,j}^w \leftarrow \Phi(\omega_{t,j}^c \cdot p(x_{t,j}^w))$
    unknown identification: $\xi(\hat{y}_{t,j}^w) \leftarrow \hat{y}_{t,j}^w$
    update $f(\cdot)$ and $g(\cdot)$:
    $\theta_f \leftarrow \theta_f - \varepsilon \nabla \theta_f \left(\mathcal{L}_{\text{sce}} + \lambda(\mathcal{L}'_{\text{csl}} + \mathcal{L}_{\text{ctl}}) + \mathcal{L}_{\text{prt}} - \mathcal{L}_{\text{esl}}\right)$
    $\theta_g \leftarrow \theta_g - \varepsilon \nabla \theta_g \left(\mathcal{L}_{\text{sce}} + \lambda \mathcal{L}'_{\text{csl}} - \mathcal{L}_{\text{esl}}\right)$
  end if
end for

IV. Experiments

A. Datasets

In the experiments, we utilized three open-source remote sensing datasets to construct a set of six cross-domain tests for open-set UDA tasks: the Aerial Image Dataset (AID) [56], the Northwestern Polytechnical University (NWPU) [1] dataset, and the University of California Merced Land Use Dataset (UCMD) [57]. The details of these datasets are listed in Table I.

1) AID: The AID dataset consists of $10,000$ remote sensing images sourced from Google Earth, encompassing 30 distinct categories. Each category contains $220 \sim 420$ images, with a size of 600×600 pixels and spatial resolutions ranging approximately from 0.5 to 8 meters.

2) NWPU: The NWPU dataset comprises a total of $31,500$ images across 45 scene categories, with each category containing 700 images. All images are sized at 256×256 pixels, with spatial resolutions ranging from approximately 0.2 to 30 meters.

3) UCMD: The UCMD dataset comprises $2,100$ high-resolution remote sensing images divided into 21 scene categories, each containing 100 images. Each image is of size 256×256 pixels with a spatial resolution of 0.3 meters.

As shown in Fig. 3, the three datasets were acquired from different sensors, exhibiting varying resolutions and spectral characteristics. Following the processing in [29], we constructed the cross-domain settings using ten common classes. First, we standardized the class names across the datasets to ensure uniformity, including agricultural, baseball diamond, beach, dense residential, forest, medium residential, parking lot, river, sparse residential, and storage tanks, facilitating consistent analysis and presentation. Among these, the first seven classes are shared as known classes, while the remaining three are treated as unknown classes. To enhance the challenges, classes with similar semantics were introduced in both the known and unknown classes. For example, dense residential and medium residential are designated as known classes, whereas sparse residential is categorized as an unknown class. Finally, six cross-domain scenarios can be constructed: AID→NWPU, AID→UCMD, NWPU→AID, NWPU→UCMD, UCMD→AID and UCMD→NWPU. Here, the symbol → indicates the adaptation from the source domain to the target domain. The numbers of samples in the source and target domains for each test set are 448 and 640, respectively.

B. Baseline Approaches

The proposed approach is compared against nine state-of-the-art UDA methods and one approach without adaptation.

1) Source-Only: The model trained on the source domain is applied directly to the target domain, and no domain adaptation is done.

2) DANN [58]: A closed-set UDA method that reduces domain differences through adversarial training between the feature extractor and the discriminator. This approach involves a feature extractor learning to produce domain-invariant features while the discriminator tries to distinguish between features from the source and target domains.

3) CDAN [59]: A closed-set UDA method that leverages multilinear and entropy-conditioned reflectances to refine the domain discriminator, thereby facilitating adversarial adaptation.

4) NOUN [60]: A closed-set UDA method that normalizes output predictions to maintain consistent feature norms, achieving a simple and compact conditional domain adaptation.

5) STA [46]: An open-set UDA method that uses a coarse-to-fine weighting mechanism to gradually separate samples of known and unknown classes.

6) OSBP [44]: An open-set UDA method that introduces a domain adversarial model using a fixed threshold to distinguish between known and unknown target samples.

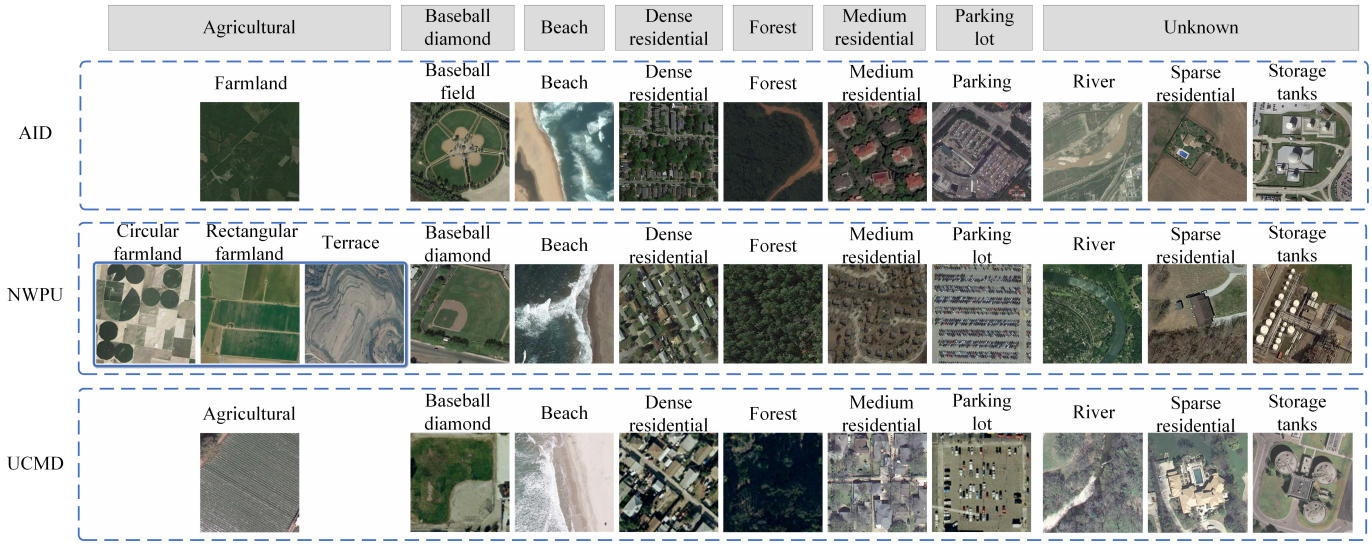


Fig. 3: Samples used in the cross-domain experiment.

7) DAMC [45]: An open-set UDA method that proposes an adversarial adaptation model with multiple auxiliary classifiers. It introduces a weighting module to evaluate different domain features, assigning weights to target samples to promote positive transfer and reduce the domain gap of shared classes.

8) UADAL [27]: An open-set UDA method that aligns the source and target-known distributions while simultaneously segregating the target-unknown distribution through adversarial learning with an additional open-set classifier.

9) SSOUDA [29]: An open-set UDA method that combines contrastive self-supervised learning with consistency self-training, in which three thresholds were predefined for pseudo-labeling.

10) MAOSDAN [28]: An open-set UDA method that employs attention-aware adaptation to distinguish unknown and known samples in the target domain with an auxiliary classifier to mitigate negative transfer effects.

C. Evaluation Metrics

To evaluate the effectiveness of our PUMCL method, we utilized five critical metrics, as follows: (1) KNO, which measures the normalized accuracy specifically for the known classes; (2) UNK, which calculates the accuracy for unknown classes; (3) Overall accuracy (OA) represents the total accuracy across all classes; (4) HOS, i.e., the harmonic mean of KNO and UNK [61]. It ensures balanced performance between known and unknown classes, addressing a common flaw of relying solely on OA metrics that could lead to artificially high scores when accuracy is high for known classes despite poor performance in unknown classes. (5) Mean Intersection over Union (mIoU). Unlike traditional accuracy metrics, which may overstate performance by misclassifying classes as unknown, mIoU provides a more comprehensive evaluation considering class variations. These metrics can be formulated as follows:

$$\mathrm{KNO} = \frac{1}{C-u}\sum_{c=1}^{C-u}\frac{\mathrm{TP}_c}{\mathrm{TP}_c+\mathrm{FP}_c} \tag{14}$$

$$\mathrm{UNK} = \frac{\mathrm{TP}_{C_u}}{\mathrm{TP}_{C_u}+\mathrm{FP}_{C_u}}, \tag{15}$$

$$\mathrm{OA} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}, \tag{16}$$

$$\mathrm{HOS} = 2\times\frac{\mathrm{KNO}\times\mathrm{UNK}}{\mathrm{KNO}+\mathrm{UNK}}, \tag{17}$$

$$\mathrm{mIoU} = \frac{1}{C}\sum_{c=1}^{C}\frac{\mathrm{TP}_c}{\mathrm{FN}_c+\mathrm{FP}_c+\mathrm{TP}_c} \tag{18}$$

where TP, FP, TN, and FN are the numbers of true positives, false positives, true negatives, and false negatives, respectively. The subscript c denotes the value of the c-th category.

These metrics collectively offer a comprehensive evaluation of our model's performance, ensuring a balanced assessment across different dimensions of class recognition.

D. Experimental Setup

In our method, we employed ResNet50 network [62] as the backbone for feature extraction, and the weights pre-trained on the ImageNet dataset are loaded. The classifier contains a pooling layer, flattened operation, and a linear projection. For optimization, Stochastic Gradient Descent (SGD) was chosen, with a momentum of 0.9. The learning rate was set to 0.003, following the annealing approach recommended by [63]. We set up our training with a batch size of 32 and planned to train for 70 epochs. Moreover, the hyperparameter λ in the total training loss function is set to 0.5. The optimal parameters of the comparison approaches were determined following the original papers. All experiments were carried out on the PyTorch platform, utilizing an NVIDIA GeForce RTX 3090 graphics card with graphics memory of 24GB.

Table II: Performance of 11 methods on six open-set domain adaptation scenarios in terms of KNO, UNK, OA, HOS and mIoU (%). The best and second-best results are highlighted in red and blue, respectively.

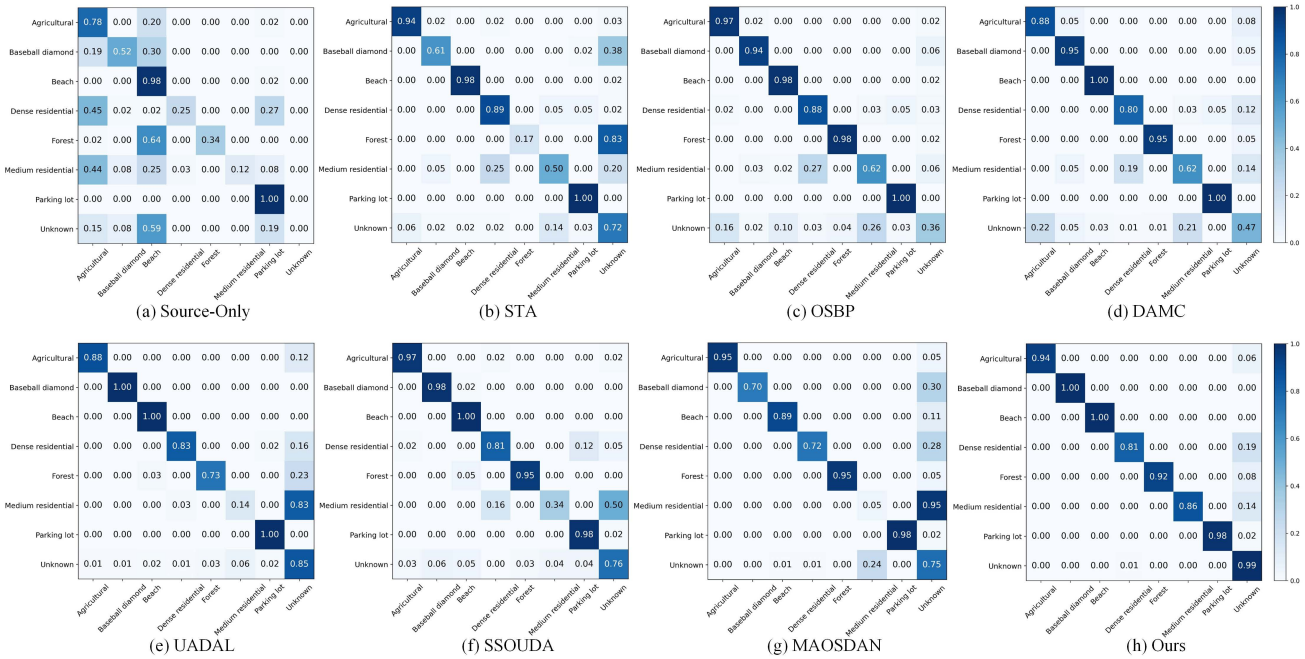
Method		Closed-set UDA				Open-set UDA						
		Source-only	DANN	CDAN	NOUN	STA	OSBP	DAMC	UADAL	SSOUDA	MAOSDAN	OURS
AID→NWPU	KNO	43.5	**90.4**	89.5	88.6	54.2	75.9	62.7	69.0	69.2	65.8	71.2
	UNK	0.0	0.0	0.0	0.0	90.6	49.0	75.0	88.0	67.7	83.3	**92.7**
	OA	30.5	63.3	62.7	62.0	65.2	67.8	66.4	74.7	68.8	71.1	**77.7**
	HOS	0.0	0.0	0.0	0.0	67.9	59.5	68.3	77.3	68.4	73.6	**80.5**
	mIoU	21.9	54.4	53.6	53.3	48.1	56.9	53.4	62.3	55.8	58.9	**66.8**
AID→UCMD	KNO	50.2	89.5	**90.2**	87.5	56.0	51.6	57.8	57.8	61.4	62.7	85.3
	UNK	0.0	0.0	0.0	0.0	76.6	57.3	78.1	85.9	68.8	**94.3**	**94.3**
	OA	35.2	62.7	63.1	61.3	62.2	53.3	63.9	66.3	63.6	72.2	**88.0**
	HOS	0.0	0.0	0.0	0.0	64.7	54.3	66.5	69.1	64.9	75.3	**89.5**
	mIoU	26.9	53.9	54.8	54.1	44.8	38.2	51.3	51.9	47.9	59.9	**81.1**
NWPU→AID	KNO	57.1	93.8	95.3	**98.2**	72.8	91.1	88.6	79.7	86.4	75.0	93.1
	UNK	0.0	0.0	0.0	0.0	71.9	35.9	46.9	84.9	75.5	75.0	**99.5**
	OA	40.0	65.6	66.7	68.8	72.5	74.5	76.1	81.2	83.1	75.0	**95.0**
	HOS	0.0	0.0	0.0	0.0	72.3	51.5	61.3	82.2	80.6	75.0	**96.2**
	mIoU	27.0	56.8	59.7	62.1	60.2	66.6	69.2	73.2	73.1	71.0	**92.0**
NWPU→UCMD	KNO	46.7	90.0	**94.6**	94.4	73.2	88.4	82.6	57.4	85.0	71.2	90.8
	UNK	0.0	0.0	0.0	0.0	73.4	24.0	58.3	72.9	53.6	92.2	**93.8**
	OA	32.7	63.0	66.3	66.1	73.3	69.1	75.3	62.0	75.6	77.5	**91.7**
	HOS	0.0	0.0	0.0	0.0	73.3	37.7	68.4	64.2	65.8	80.3	**92.3**
	mIoU	21.5	56.2	59.0	59.2	59.4	60.8	67.1	48.0	67.0	67.0	**86.9**
UCMD→AID	KNO	52.7	**92.6**	83.7	90.0	55.1	76.6	75.7	69.2	62.9	65.4	75.9
	UNK	0.0	0.0	0.0	0.0	81.3	61.5	64.1	75.0	79.2	76.6	**86.5**
	OA	36.9	64.8	58.6	63.0	63.0	72.0	72.2	70.9	67.8	68.8	**79.1**
	HOS	0.0	0.0	0.0	0.0	65.7	68.2	69.4	72.0	70.1	70.5	**80.8**
	mIoU	23.4	54.8	48.3	53.4	48.3	65.9	61.7	59.6	52.5	58.2	**70.2**
UCMD→NWPU	KNO	51.6	83.9	85.5	**85.7**	43.5	68.1	56.7	62.9	67.6	65.2	68.5
	UNK	0.0	0.0	0.0	0.0	86.5	55.7	66.1	86.5	77.6	74.5	**87.5**
	OA	36.1	58.7	59.8	60.0	56.4	64.4	59.5	70.0	70.6	68.0	**74.2**
	HOS	0.0	0.0	0.0	0.0	57.9	61.3	61.1	72.9	72.3	69.5	**76.9**
	mIoU	24.4	47.2	50.3	50.4	39.1	51.7	45.6	56.0	56.2	53.5	**61.3**

E. Quantitative Analysis

Table II presents the evaluation results of all methods across six open-set cross-domain scenarios. It is evident from the table that the non-domain adaptive method, source-only, performs the worst due to domain shift effects. This method fails to accurately identify both known and unknown classes.

In the open-world cross-domain classification task, the goal is to identify unknown samples while accurately classifying the known samples, effectively balancing performance for both. Closed-set UDA methods generally achieve higher accuracy for known samples, as evidenced by higher KNO values compared to open-set UDA methods. However, these closed-set methods struggle to correctly identify unknown classes. They often misclassify unknown samples in the target domain as known classes because they directly align the two domains without the capability to recognize unknown samples, due to the lack of prior knowledge about these unknown classes. Consequently, this results in lower OA and mIoU values.

On the other hand, open-set UDA methods may experience declines in the KNO metric due to feature overlaps caused by unknown classes, which could shift the decision boundary towards balancing the classification of both known and unknown samples. However, they significantly improve the capability to recognize unknown classes. As a result, all open-set UDA methods demonstrate better performance in comprehensive indices such as OA, HOS, and mIoU. As shown in Table II, in the AID→UCMD, NWPU→AID, and NWPU→UCMD scenarios, the KNO values in the proposed approach are comparable to those of closed-set UDA methods. However, in other scenarios, the drop in KNO is more noticeable due to the high spectral similarity between certain land use categories. Among the six open-set methods, the proposed PUMCL still achieves the highest KNO values in the four scenarios: AID→UCMD, NWPU→AID, NWPU→UCMD, and UCMD→NWPU. This advantage is attributed to the integration of prototypical constraints and multiview consistency learning, which can provide reliable supervised signals for the model training. In the UCMD→AID and AID→NWPU scenarios, although our PUMCL method ranks second to OSBP in the KNO metric (by 0.7% and 4.7%, respectively), OSBP's UNK accuracy is significantly inferior (by 43.7% and 25.0%, respectively), due to negative transfer effect of open-set UDA.



(a) Source-Only (b) STA (c) OSBP (d) DAMC

(e) UADAL (f) SSOUDA (g) MAOSDAN (h) Ours

Fig. 4: Confusion matrix on the NWPU→AID scenario.

With respect to the UNK metric, our PUMCL method outperforms all other methods, showing improvements ranging from 1.0% to 14.6%. While MAOSDAN achieves comparable UNK accuracy to our method in the AID→UCMD and NWPU→UCMD scenarios, it exhibits performance degradation on other adaptation datasets. In comparison, our methods consistently achieve stable and promising performance across all datasets. This improvement is credited to the entropy separation strategy and unknown-aware consistency learning, which provide reliable supervision information and establish clear distribution patterns for detecting and recognizing unknown classes.

For the three overall evaluation metrics, OA, HOS, and mIoU, our method achieves optimal performance in all experiments. Specifically, our PUMCL method achieves significant gains of 3.0% to 15.8% in OA, 3.2% to 14.2% in HOS, and 4.5% to 21.2% in the mIoU. These results demonstrate that our PUMCL method outperforms other methods in overall performance while achieving the balance between known and unknown class identification.

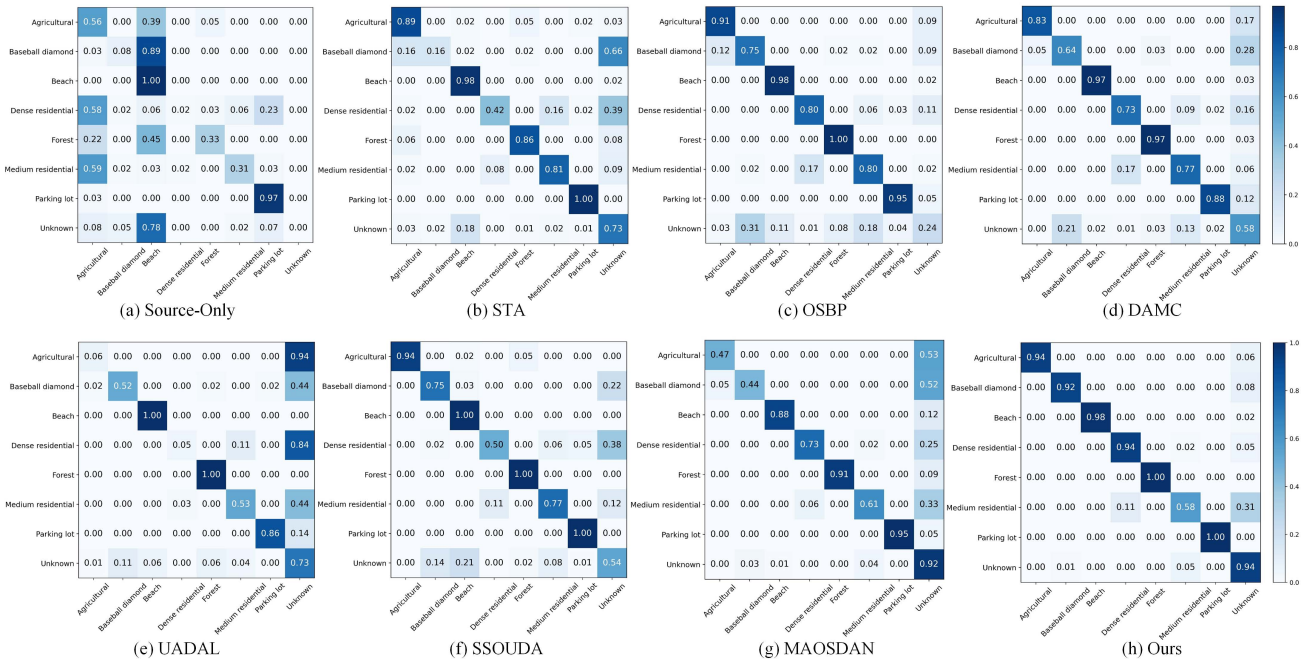
F. Confusion Matrix

To further analyze the classification performance of each class, we present the confusion matrix in the NWPU→AID scenario shown in Fig. 4, and NWPU→UCMD scenario shown in Fig. 5. In the confusion matrix, each item corresponds to the number of samples where a predicted class in a column matches an actual class in a row. After normalization along the row axis, the item in the diagonal represents the percentage of instances that are correctly predicted, corresponding to the recall of each class.

Fig. 4 illustrates a significant problem where the STA, UADAL, SSOUDA, and MAOSDAN methods frequently misclassify known classes as unknown. For example, as shown in Fig. 4(e) and (g), misclassification occurs in the "medium residential" class because the unknown classes include "sparse residential", which shares similar characteristics with "medium residential". Similarly, the "forest" class was wrongly classified into the unknown classes by STA, as shown in Fig. 4(b). This finding suggests that these four methods struggle to differentiate classes with similar features, often leading to known classes being misclassified as unknown. Conversely, OSBP and DAMC exhibit poor recognition accuracy for unknown classes and fail to effectively distinguish unknown classes from known classes. This weakness is evident in their tendency to classify unknown samples as known classes, such as "agricultural" and "medium residential", as displayed in Fig. 4(c) and (d). Fig. 5 presents another clear example of the model experiencing significant interference from the unknown class. Except for "medium residential" and "dense residential," some other known classes have been misclassified by the compared methods, e.g., "agricultural" shown in Fig. 5(d) and (e), and "baseball diamond" shown in (d), (e) and (g). As a result, serious confusion exists between many unknown samples and these categories, further exacerbating the misclassification issue. In contrast, our method can effectively distinguish confusing classes and accurately classify classes with significant cross-domain feature discrepancies. This is evidenced by the overall highest values along the diagonal in the confusion matrices, as well as by the five metrics shown in Table II, all of which are calculated based on the confusion matrices.

G. Feature Visualization

To visually compare the classification performance of different open-set methods, we utilized the t-SNE technique to



(a) Source-Only (b) STA (c) OSBP (d) DAMC

(e) UADAL (f) SSOUDA (g) MAOSDAN (h) Ours

Fig. 5: Confusion matrix on the NWPU→UCMD scenario.

visualize features extracted from the final convolutional layer in the backbone network. Fig. 6 displays the visualization results for seven open-set UDA methods and the source-only mode. We focus on the open-set UDA methods because the closed-set UDA methods fail to identify unknown classes, leading to poor feature representation. In the t-SNE plot, each point represents a feature vector of an image projected into a two-dimensional space, preserving the local relationships in the original high-dimensional feature space. This visualization allows us to observe the clustering and separation of features learned by each adaptation method. Ideally, a well-performing method should exhibit clear and distinct clusters for known classes while effectively isolating or delineating unknown classes.

It can be observed that the source-only method exhibits poor clustering and classification performance due to domain discrepancies, resulting in significant overlap between classes. DAMC, UADAL, and SSOUDA alleviate domain shifts with the generated representations that tend to be aggregated as groups. Similarly, OSBP and STA also show reasonable clustering for known classes. However, all these methods struggle to find obvious decision boundaries that accurately classify all classes, resulting in significant overlaps between them. MAOSDAN shows a promising clustering effect for known classes. However, its unknown samples show scattered distribution overlaps with the known classes, resulting in misclassification. This effect arises because MAOSDAN relies heavily on the probability of unknown predictions in the adaptation process, pushing predictions toward unknown classes.

Compared with other open-set UDA methods, our method shows better class separability, with more distinct clustering of the same categories and clearer boundaries between different classes in the t-SNE plot. Furthermore, considering that the unknown class includes three distinct categories (“river”, “sparse residential”, and “storage tank”), our PUMCL method generates representations with three distinct unknown groups in cross-domain scenarios, especially shown in AID→NWPU, NWPU→AID, UCMD→AID, and UCMD→NWPU, which indicates the effectiveness in internally detecting and distinguishing unknown classes. Overall, the t-SNE visualizations provide insights into the feature representations learned by each method and offer a qualitative assessment of their classification performance in open-set UDA scenarios. However, as shown in the plots, there are still some overlaps between the known and unknown classes in the proposed PUMCL method. Additionally, our model underperforms on known classes compared to the closed-set models, highlighting an area for improvement in future studies.

To quantify the classification separation, we employ the widely used silhouette score to evaluate the quality of feature clustering [64]. The silhouette score ranges from -1 to 1, where a score close to +1 indicates that a data point is well-matched to its own cluster and poorly matched to neighboring clusters. Table III shows a significant improvement in silhouette scores for our method compared to the other open-set methods in all UDA tests, verifying the promising separation capability of our approach.

H. Ablation Study

Considering that the NWPU dataset contains more complex samples and that the distribution of certain classes in the UCMD dataset (such as ”agricultural” and ”baseball diamond”) differs significantly from the other two datasets, we selected the UCMD→NWPU scenario for the ablation study, as shown in Table IV. The results demonstrate that the highest accuracy, with an OA of 74.2%, is achieved when all modules

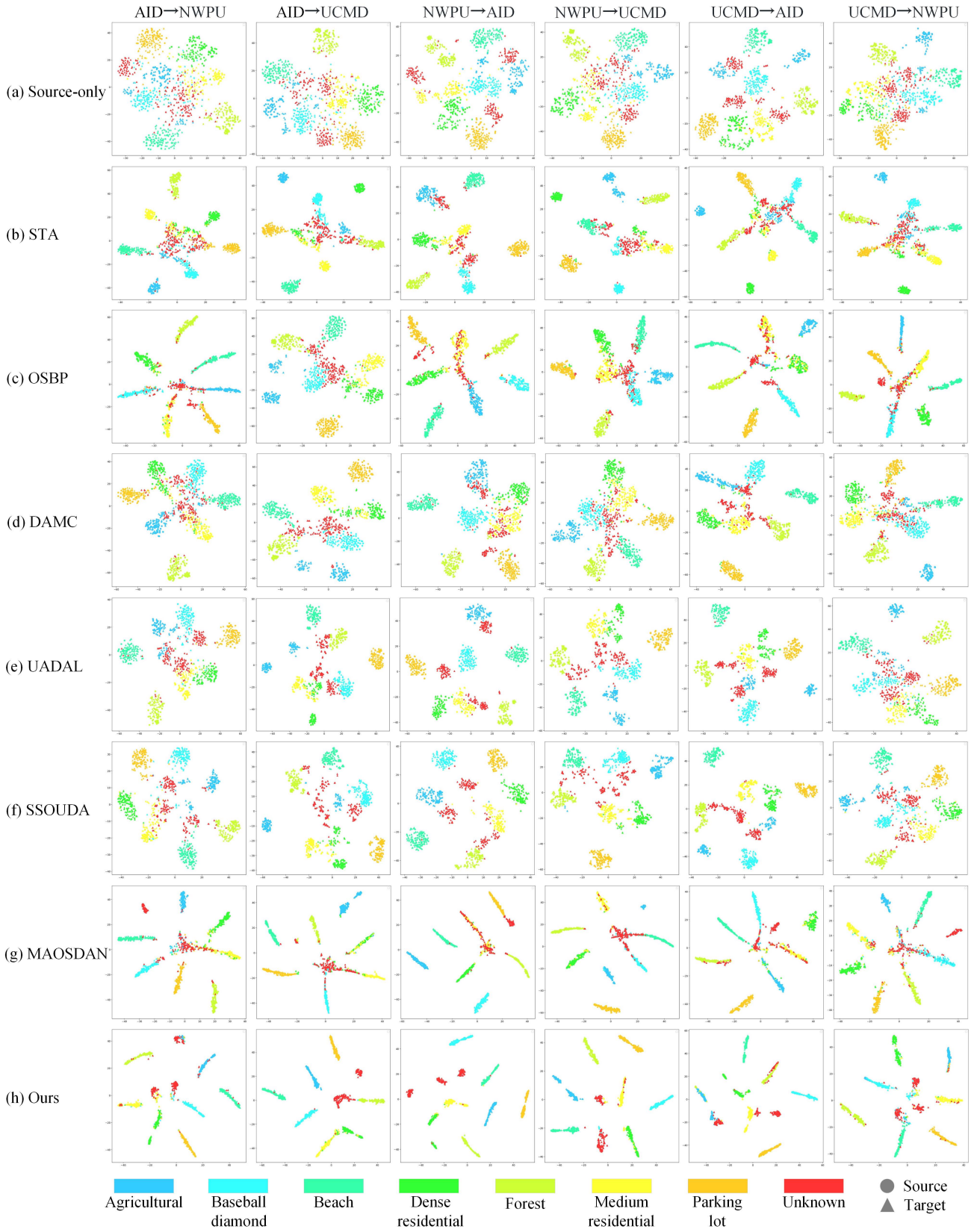


Fig. 6: The t-SNE visualization of features for the seven open-set DA along with Source-only on six open-set DA scenarios.

Table III: The silhouette scores (↑) of the seven open-set UDA methods and source-only on six sets of tests. The best results are marked in bold.

Methods	AID→ NWPU	AID→ UCMD	NWPU→ AID	NWPU→ UCMD	UCMD→ AID	UCMD→ NWPU
Source-only	0.47	0.51	0.44	0.47	0.51	0.53
STA	0.50	0.54	0.52	0.49	0.55	0.54
OSBP	0.61	0.59	0.61	0.60	0.64	0.63
DAMC	0.54	0.58	0.55	0.51	0.59	0.58
UADAL	0.60	0.60	0.62	0.54	0.61	0.55
SSOUDA	0.63	0.61	0.63	0.60	0.66	0.61
MAOSDAN	0.66	0.64	0.68	0.65	0.63	0.64
OURS	**0.70**	**0.71**	**0.73**	**0.73**	**0.72**	**0.66**

Table IV: Ablation study of each loss component for UCMD→NWPU. The best results are marked in bold.

$\mathcal{L}'_{\mathrm{csl}}$	$\mathcal{L}_{\mathrm{ctl}}$	$\mathcal{L}_{\mathrm{prt}}$	$\mathcal{L}_{\mathrm{esl}}$	KNO	UNK	OA	HOS	mIoU
✓	✓	✓	✓	68.5	87.5	**74.2**	**76.9**	**61.3**
✗	✓	✓	✓	68.3	73.4	69.8	70.8	57.4
✓	✗	✓	✓	57.1	84.4	65.3	68.1	50.8
✓	✓	✗	✓	63.6	86.5	70.5	73.3	56.7
✓	✓	✓	✗	**82.4**	0.0	57.7	0.0	49.5
✗	✗	✓	✓	30.6	83.9	46.6	44.8	24.7
✗	✓	✗	✓	66.1	69.3	67.0	67.6	53.8
✗	✓	✓	✗	75.9	0.0	53.1	0.0	41.1
✓	✗	✗	✓	54.2	**88.0**	64.4	67.1	48.8
✓	✗	✓	✗	78.3	0.0	54.8	0.0	46.9
✓	✓	✗	✗	80.4	0.0	56.2	0.0	47.5
✗	✗	✗	✓	27.9	79.2	43.3	41.3	21.7
✗	✗	✓	✗	53.3	0.0	37.3	0.0	25.9
✗	✓	✗	✗	73.2	0.0	51.2	0.0	38.3
✓	✗	✗	✗	74.8	0.0	52.3	0.0	43.5
✗	✗	✗	✗	51.6	0.0	36.1	0.0	24.4

are combined. Specifically, eliminating $\mathcal{L}'_{\mathrm{csl}}$, $\mathcal{L}_{\mathrm{ctl}}$, $\mathcal{L}_{\mathrm{prt}}$, and $\mathcal{L}_{\mathrm{esl}}$ results in an OA reduction of 4.4%, 8.9%, 3.7%, and 16.5%, respectively. This clearly highlights the individual contributions of each module to the overall performance, with their combination providing the most robust solution for open-set UDA. It can also be observed that significant performance changes occur when any two of $\mathcal{L}_{\mathrm{ctl}}$, $\mathcal{L}_{\mathrm{prt}}$ and $\mathcal{L}_{\mathrm{els}}$ are combined. Specifically, there are notable performance degradations in mIoU when $\mathcal{L}_{\mathrm{ctl}}$ and $\mathcal{L}_{\mathrm{esl}}$ are removed.

Furthermore, the proposed consistency loss $\mathcal{L}'_{\mathrm{csl}}$ serves as the foundation for domain alignment. Implementing consistency learning alone achieves an mIoU of 43.5%, which is the highest when each loss function is applied individually. When combined with the contrastive enhancement module $\mathcal{L}_{\mathrm{ctl}}$, the discriminative capability is further improved, resulting in a 5.3% improvement in mIoU. The $\mathcal{L}_{\mathrm{ctl}}$, which operates in a self-learning manner, acts as an auxiliary module to enhance the discriminative power for target samples. Particularly, $\mathcal{L}_{\mathrm{esl}}$ enhances the recognition of unknown classes. As a result, if this loss function is eliminated, the UNK values drop to 0.0%. When combined with other modules, it significantly improves the unknown recognition capability despite some sacrifice in the accuracy of known samples. Additionally, adding prototypical constraints $\mathcal{L}_{\mathrm{prt}}$ to consistency learning $\mathcal{L}'_{\mathrm{csl}}$ increases the overall classification accuracy, with a gain of 3.4% in mIoU. This improvement occurs because $\mathcal{L}_{\mathrm{prt}}$ refines the alignment between domains by providing more reliable pseudo-labels and creating more compact feature subspaces.

While the integration of all modules yields the best OA, HOS, and mIoU values, it does not guarantee the highest KNO and UNK values simultaneously. Nevertheless, it proves beneficial for improving overall classification performance while maintaining a balance between known and unknown categories.

Table V: Computational efficiency comparison of different methods. The best results are marked in bold.

Methods	FLOPs (G) ↓	Params (M) ↓	Testing Time ↓ (s/batch)	Training Time ↓ (s/batch)
DANN	5.3970	**24.30**	**0.0206**	0.27
CDAN	5.3970	26.13	0.0216	0.28
NOUN	5.3970	24.83	0.0209	0.26
STA	5.3986	30.29	0.0303	0.50
OSBP	5.3996	26.57	0.0223	**0.10**
DAMC	5.3986	27.65	0.0237	0.13
UADAL	**5.3965**	27.37	0.0225	0.30
SSOUDA	5.3970	26.08	0.0210	0.35
MAOSDAN	5.3971	27.47	0.0229	0.30
OURS	5.3971	26.42	0.0219	0.62

I. Efficiency and Complexity

To assess computational efficiency, we report the floating point operations (FLOPs) and the number of trainable parameters (Params) to determine computational and model complexity, respectively. We also evaluate the training and testing time per batch with the same batch size, as shown in Table V. The FLOPs for all methods are nearly identical, ranging from 5.3965 to 5.3996 GFLOPs, indicating similar computational demands. The parameter count varies, with STA having the highest complexity at 30.29 M and DANN the lowest at 24.30 M. The differences are that STA, DAMC, and UADAL use adversarial learning with an additional discriminator, and MAOSDAN specifically incorporates an auxiliary classifier, while our PUMCL does not require extra networks. Testing time shows minor variations, with DANN being the fastest (0.0206 s/batch) and STA the slowest (0.0303 s/batch). Notably, while our method incurs a slightly higher training time, it maintains a competitive testing time (0.0219 s/batch) and balances model complexity with 26.42 million parameters, positioning it as an effective trade-off between computational efficiency and potential performance gains in UDA tasks.

V. Conclusion

In this study, we propose a robust framework for open-set Unsupervised Domain Adaptation (UDA) in cross-domain remote sensing scene classification. Our approach effectively utilizes category-specific transferable knowledge and feature distribution patterns from unlabeled target domain samples. A key contribution is the introduction of a Prototypical Unknown-aware Multiview Consistency Learning (PUMCL) scheme, which integrates prototypical constraints to refine pseudo-labels and enhance class separation. Additionally, our method adopts an open-set adaptation strategy that adjusts conditional entropy for known and unknown samples while aligning domains through unknown-aware consistency learning.

These advancements collectively improve the adaptability and reliability of remote sensing classifiers, representing significant progress in open-set UDA for remote sensing applications. We conducted comparative experiments against three closed-set methods and six open-set methods to comprehensively evaluate the proposed method. Utilizing three public datasets, i.e., AID, NWPU, and UCMD, we conducted six groups of cross-domain experiments. Our approach achieved the highest scores in the metrics OA, HOS, and mIoU across these six scenarios, surpassing the second-best results by 3.0% to 15.8% in OA,3.2% to 14.2% in HOS, and 4.5% to 21.2% in mIoU.

We have the following findings from the experiments: Through the confusion matrix analysis, we found that unknown classes caused confusion between known and unknown classes due to their similar spectral characteristics and spatial structure. Our approach can significantly reduce class confusion by learning more discriminative features and shrinkable category-wise feature space. This has been verified by the statistical and visual analysis of feature distribution generated by different methods. Compared with other methods, the proposed approach generates clearer boundaries between different classes, especially more compact clustering of unknown classes with fewer overlaps with the known classes. Furthermore, the performance degradation observed in the ablated modules confirms the effectiveness of the designed loss functions and optimization strategies in enhancing overall accuracy.

In our future research, more computationally efficient methods will be investigated to handle large-scale datasets and reduce the training time without compromising accuracy. Moreover, we will extend the framework to adapt modalities beyond optical imagery, such as Synthetic Aperture Radar (SAR) or hyperspectral images.

ACKNOWLEDGMENT

The authors would like to thank the editors and reviewers for their instructive comments, which helped to improve this article.

REFERENCES

[1] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.

[2] K. Chen, B. Chen, C. Liu, W. Li, Z. Zou, and Z. Shi, "Rsmamba: Remote sensing image classification with state space model," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.

[3] R. Dong, L. Mou, M. Chen, W. Li, X.-Y. Tong, S. Yuan, L. Zhang, J. Zheng, X. Zhu, and H. Fu, "Large-scale land cover mapping with fine-grained classes via class-aware semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16783–16793, 2023.

[4] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and remote sensing magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[5] Y. Yang, X. Tang, Y.-M. Cheung, X. Zhang, and L. Jiao, "Sagn: Semantic-aware graph network for remote sensing scene classification," *IEEE Transactions on Image Processing*, vol. 32, pp. 1011–1025, 2023.

[6] J. Wang, W. Li, M. Zhang, R. Tao, and J. Chanussot, "Remote-sensing scene classification via multistage self-guided separation network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.

[7] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5966–5978, 2020.

[8] X. Tan, B. Xi, J. Li, T. Zheng, Y. Li, C. Xue, and J. Chanussot, "Review of zero-shot remote sensing image scene classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 11274–11289, 2024.

[9] P. Ghamisi, W. Yu, A. Marinoni, C. M. Gevaert, C. Persello, S. Selvakumaran, M. Girotto, B. P. Horton, P. Rufin, P. Hostert, F. Pacifici, and P. M. Atkinson, "Responsible ai for earth observation," 2024.

[10] D. Hong, B. Zhang, H. Li, Y. Li, J. Yao, C. Li, M. Werner, J. Chanussot, A. Zipf, and X. X. Zhu, "Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks," *Remote Sensing of Environment*, vol. 299, p. 113856, 2023.

[11] J. Peng, Y. Huang, W. Sun, N. Chen, Y. Ning, and Q. Du, "Domain adaptation in remote sensing image classification: A survey," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9842–9859, 2022.

[12] Y. Xu, P. Ghamisi, and Y. Avrithis, "Multi-target unsupervised domain adaptation for semantic segmentation without external data," *arXiv preprint arXiv:2405.06502*, 2024.

[13] Y. Liu, B. Du, Y. Chen, and L. Zhang, "Robust multiple subspaces transfer for heterogeneous domain adaptation," *Pattern Recognition*, vol. 152, p. 110473, 2024.

[14] W. Huang, Y. Shi, Z. Xiong, Q. Wang, and X. X. Zhu, "Semi-supervised bidirectional alignment for remote sensing cross-domain scene classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 195, pp. 192–203, 2023.

[15] M. Cai, B. Xi, J. Li, S. Feng, Y. Li, Z. Li, and J. Chanussot, "Mind the gap: Multilevel unsupervised domain adaptation for cross-scene hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[16] C. Yang, Y. Dong, B. Du, and L. Zhang, "Attention-based dynamic alignment and dynamic distribution adaptation for remote sensing cross-domain scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[17] J. Zhang, J. Liu, B. Pan, and Z. Shi, "Domain adaptation based on correlation subspace dynamic distribution alignment for remote sensing image scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7920–7930, 2020.

[18] Y. Huang, J. Peng, W. Sun, N. Chen, Q. Du, Y. Ning, and H. Su, "Two-branch attention adversarial domain adaptation network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[19] L. Zhang, M. Lan, J. Zhang, and D. Tao, "Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.

[20] X. Zhang, W. Yu, M.-O. Pun, and W. Shi, "Cross-domain landslide mapping from large-scale remote sensing images using prototype-guided domain-aware progressive representation learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 197, pp. 1–17, 2023.

[21] Y. Xu and P. Ghamisi, "Consistency-regularized region-growing network for semantic segmentation of urban scenes with point-level annotations," *IEEE Transactions on Image Processing*, vol. 31, pp. 5038–5051, 2022.

[22] Z. Yang, J. Yue, P. Ghamisi, S. Zhang, J. Ma, and L. Fang, "Open set recognition in real world," *International Journal of Computer Vision*, pp. 1–24, 2024.

[23] L. Fang, Z. Yang, T. Ma, J. Yue, W. Xie, P. Ghamisi, and J. Li, "Open-world recognition in remote sensing: Concepts, challenges, and opportunities," *IEEE Geoscience and Remote Sensing Magazine*, vol. 12, no. 2, pp. 8–31, 2024.

[24] J. Guo, Y. Lai, J. Zhang, J. Zheng, H. Fu, L. Gan, L. Hu, G. Xu, and X. Che, "C³da: A universal domain adaptation method for scene classification from remote sensing imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.

[25] Q. Li, Y. Wen, J. Zheng, Y. Zhang, and H. Fu, "Hyunida: Breaking label set constraints for universal domain adaptation in cross-scene hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.

[26] J. Zhang, J. Liu, B. Pan, Z. Chen, X. Xu, and Z. Shi, "An open set domain adaptation algorithm via exploring transferability and discriminability for remote sensing image scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2021.

[27] J. Jang, B. Na, D. H. Shin, M. Ji, K. Song, and I.-C. Moon, "Unknown-aware domain adversarial learning for open-set domain adaptation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 16755–16767, 2022.

[28] J. Zheng, Y. Wen, M. Chen, S. Yuan, W. Li, Y. Zhao, W. Wu, L. Zhang, R. Dong, and H. Fu, "Open-set domain adaptation for scene classification using multi-adversarial learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 208, pp. 245–260, 2024.

[29] S. Wang, D. Hou, and H. Xing, "A self-supervised-driven open-set unsupervised domain adaptation method for optical remote sensing image scene classification and retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[30] S. Zhao, Y. Zhang, Y. Luo, Y. Kang, and H. Wang, "Dynamically self-training open set domain adaptation classification method for heterogeneous sar image," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.

[31] Z. Xie, P. Duan, W. Liu, X. Kang, X. Wei, and S. Li, "Feature consistency-based prototype network for open-set hyperspectral image classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[32] S. Zhu, B. Du, L. Zhang, and X. Li, "Attention-based multiscale residual adaptation network for cross-scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.

[33] J. Qu, W. Dong, Y. Yang, T. Zhang, Y. Li, and Q. Du, "Cycle-refined multidecision joint alignment network for unsupervised domain adaptive hyperspectral change detection," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[34] L. Ma, C. Luo, J. Peng, and Q. Du, "Unsupervised manifold alignment for cross-domain classification of remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 10, pp. 1650–1654, 2019.

[35] J. Zheng, W. Wu, S. Yuan, Y. Zhao, W. Li, L. Zhang, R. Dong, and H. Fu, "A two-stage adaptation network (tsan) for remote sensing scene classification in single-source-mixed-multiple-target domain adaptation (s²m²t da) scenarios," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[36] J. Zheng, H. Fu, W. Li, W. Wu, Y. Zhao, R. Dong, and L. Yu, "Cross-regional oil palm tree counting and detection via a multi-level attention domain adaptation network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 167, pp. 154–177, 2020.

[37] S. Zhu, C. Wu, B. Du, and L. Zhang, "Adversarial pair-wise distribution matching for remote sensing image cross-scene classification," *Neural Networks*, vol. 174, p. 106241, 2024.

[38] X. Zhang, X. Yao, X. Feng, G. Cheng, and J. Han, "Dfenet for domain adaptation-based remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[39] Y. Huang, J. Peng, N. Chen, W. Sun, Q. Du, K. Ren, and K. Huang, "Cross-scene wetland mapping on hyperspectral remote sensing images using adversarial domain adaptation network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 203, pp. 37–54, 2023.

[40] Y. Huang, J. Peng, G. Zhang, W. Sun, N. Chen, and Q. Du, "Adversarial domain adaptation network with calibrated prototype and dynamic instance convolution for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[41] B. Pan, Z. Xu, T. Shi, T. Li, and Z. Shi, "An imbalanced discriminant alignment approach for domain adaptive sar ship detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–11, 2023.

[42] T. Yang, S. Xiao, J. Qu, W. Dong, Q. Du, and Y. Li, "Graph embedding interclass relation-aware adaptive network for cross-scene classification of multisource remote sensing data," *IEEE Transactions on Image Processing*, vol. 33, pp. 4459–4474, 2024.

[43] D. Hou, S. Wang, X. Tian, and H. Xing, "Pcluda: A pseudo-label consistency learning-based unsupervised domain adaptation method for cross-domain optical remote sensing image retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2022.

[44] K. Saito, S. Yamamoto, Y. Ushiku, and T. Harada, "Open set domain adaptation by backpropagation," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 153–168, 2018.

[45] T. Shermin, G. Lu, S. W. Teng, M. Murshed, and F. Sohel, "Adversarial network with multiple classifiers for open set domain adaptation," *IEEE Transactions on Multimedia*, vol. 23, pp. 2732–2744, 2020.

[46] H. Liu, Z. Cao, M. Long, J. Wang, and Q. Yang, "Separate to adapt: Open set domain adaptation via progressive separation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2927–2936, 2019.

[47] B. Niu, Z. Pan, K. Chen, Y. Hu, and B. Lei, "Open set domain adaptation via instance affinity metric and fine-grained alignment for remote sensing scene classification," *IEEE Geoscience and Remote Sensing Letters*, 2023.

[48] W. Liu, X. Nie, B. Zhang, and X. Sun, "Incremental learning with open-set recognition for remote sensing image scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[49] X. Li, C. Wen, Y. Hu, and N. Zhou, "Rs-clip: Zero shot remote sensing scene classification via contrastive vision-language supervision," *International Journal of Applied Earth Observation and Geoinformation*, vol. 124, p. 103497, 2023.

[50] Y. Ning, J. Peng, Q. Liu, Y. Huang, W. Sun, and Q. Du, "Contrastive learning based on category matching for domain adaptation in hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[51] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7236–7246, 2023.

[52] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 702–703, 2020.

[53] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," *Advances in neural information processing systems*, vol. 33, pp. 18661–18673, 2020.

[54] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, pp. 1597–1607, PMLR, 2020.

[55] Y. Xu, L. Chen, L. Duan, I. W. Tsang, and J. Luo, "Open set domain adaptation with soft unknown-class rejection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 3, pp. 1601–1612, 2023.

[56] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[57] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pp. 270–279, 2010.

[58] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International conference on machine learning*, pp. 1180–1189, PMLR, 2015.

[59] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," *Advances in neural information processing systems*, vol. 31, 2018.

[60] D. Hu, J. Liang, Q. Hou, H. Yan, and Y. Chen, "Adversarial domain adaptation with prototype-based normalized output conditioner," *IEEE Transactions on Image Processing*, vol. 30, pp. 9359–9371, 2021.

[61] S. Bucci, M. R. Loghmani, and T. Tommasi, "On the effectiveness of image rotation for open set domain adaptation," in *European conference on computer vision*, pp. 422–438, Springer, 2020.

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

[63] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of machine learning research*, vol. 17, no. 59, pp. 1–35, 2016.

[64] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)*, pp. 747–748, IEEE, 2020.

Xiaokang Zhang (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2018.

From 2019 to 2022, he was a Postdoctoral Research Associate with The Hong Kong Polytechnic University, Hong Kong, and The Chinese University of Hong Kong, Shenzhen, Shenzhen, China. He is currently a specially-appointed Professor with the School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan. He has authored or coauthored more than 40 scientific publications in international journals and conferences. His research interests include remote sensing image analysis, computer vision and machine learning.

Dr. Zhang is currently a Reviewer for more than 30 renowned international journals, such as Remote Sensing of Environment, ISPRS Journal of Photogrammetry and Remote Sensing, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON IMAGE PROCESSING and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Wanjing Wu is currently pursuing her bachelor's degree in Electronic Information Engineering at Wuhan University of Science and Technology, Wuhan, China. Her research interests include open-set domain adaptation, remote sensing and machine learning.

Mi Zhang (Member, IEEE) is an associate researcher at the School of Remote Sensing and Information Engineering, Wuhan University. He serves as the chief artificial intelligence scientist in Handleray Corporation and the director of WHU-LuoJiaAI Group. He is also a reviewer of CVPR/ICCV/ECCV, a program committee member of AAAI, and a reviewer of more than ten journals. His research interests mainly include computer vision and machine learning, with particular interest in semantic object segmentation and the construction of deep learning frameworks.

Weikang Yu (Graduate Student Member, IEEE) received the B.Eng. degree from Beihang University, Beijing, China, in 2020, and the M.Phil. degree from The Chinese University of Hong Kong, Shenzhen, in 2022. He is pursuing a PhD with the Machine Learning Group (FWGE-L), Helmholtz Institute Freiberg for Resource Technology, Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Freiberg, Germany. He also works as a beyond research fellow and a PhD student with the Chair of Data Science in Earth Observation at Technical University of Munich (TUM), Munich, Germany.

His research interests include remote sensing, image processing, and machine learning.

Pedram Ghamisi (Senior Member, IEEE) obtained his Ph.D. in electrical and computer engineering from the University of Iceland in 2015. He currently serves as (1) the head of the machine learning group at a Helmholtz institute (HZDR-HIF), Germany, and (2) visiting professor at Lancaster University, UK. He has previously held positions as Senior PI, research professor, and group leader of AI4RS at the Institute of Advanced Research in Artificial Intelligence (IARAI), Austria. He has achieved more than 10 international awards and recognitions, including the IEEE-GRSS Highest Impact Paper Awards for 2020 and 2024, Winner of the Data Fusion Contest 2017 organized by IEEE GRSS, the IEEE Mikio Takagi Prize for 2013, and the 2017 Best Reviewer Prize of IEEE GRSS. Additionally, he was named one of the top 1% of most cited researchers, as published by Clarivate since 2021. His research interests primarily revolve around deep learning, particularly in the domain of remote sensing applications. Prof. Dr. Ghamisi has also been the recipient of multiple scholarships and grants, such as the Alexander von Humboldt Fellowship in 2015, the High Potential Program (HPP) Team Leadership in 2018, the AI for Remote Sensing grant in 2020, and the 3D-ABC of the Helmholtz Foundation Models Initiative in 2024. For more information, please visit http://www.ai4rs.com.