PTDA: Progressive Pseudo-Label Learning for Cross-Domain Cloud Detection in High-Resolution Remote Sensing

Jin Kuang, Xianjun Gao, *Member, IEEE*, Yuanwei Yang, *Member, IEEE*, Siyuan Dong, Ji Dong, Yuan Kou, Meilin Tan, and Zhiwei Wang

***Abstract*— The global cloud detection of high-resolution remote sensing images (HRSIs) is crucial for acquiring high-quality imagery and optimizing data utilization. Traditional cloud detection models, which rely on limited samples and fully supervised learning, struggle to adapt to cross-temporal and cross-spatial domains. While current unsupervised domain adaptation (UDA) methods improve performance in cross-domain cloud detection to some extent, generating high-quality, reliable pseudo-labels remains a significant challenge for global cloud detection. Therefore, this article proposes a progressive pseudo-label learning for cross-domain cloud detection in high-resolution remote sensing (PTDA). First, we propose an online domain-invariant feature-guided pseudo-label generation (OPLG) strategy and learning intradomain unaligned features (LIUFs), which effectively integrate domain-invariant features and intradomain semantics to generate high-quality pseudo-labels at the feature level. LIUF then refines the pseudo-label quality at the pixel level. Second, during the model training, pseudo-label-constrained intradomain feature mining loss (PCIF Loss) is designed to suppress noisy semantic information within the domain, the hole effect of thick/thin clouds, and the noise interference of the contour boundary. Four cloud detection datasets, including MS Cloud (MS), HRC_WHU Cloud (WHU), 95 Cloud (95), and WHUS2-CD + (S2), are grouped into three cross-domain tests, MS2WHU, MS2S2, and WHU295. Our approach achieved the best performance with mean intersection over union (mIoU) of 63.99%, 58.14%, and 58.82%, and overall accuracy (OA) of 80.03%, 79.36%, and 80.49%, respectively. The experimental results show that the proposed method outperforms seven state-of-the-art (SOTA) cross-domain comparison methods. Thus, our method has important application value for cross-domain cloud detection. The available code can be downloaded from https://github.com/gasking/PTDA.**

***Index Terms*— Cloud detection, cross-domain, high-resolution remote sensing image (HRSI), pseudo-label, unsupervised domain adaptation (UDA).**

I. Introduction

IN RECENT years, with the rapid development of remote sensing technology, high-resolution remote sensing image (HRSI) has been widely applied in urban management, hydrological monitoring, land use classification, and geological exploration [1], [2]. However, atmospheric effects on the Earth's surface cause optical sensors to be easily interfered with clouds, resulting in significant cloud coverage that obscures surface information in RSIs. With the explosive growth in global ground-monitoring tasks and the number of satellites, developing cross-domain cloud detection methods with strong generalization ability is crucial for improving remote sensing data utilization and reducing sensor imaging time.

Currently, cloud detection in HRSI mainly relies on semantic segmentation methods. These methods first extract image features using convolutional neural networks (CNNs) and then perform pixelwise classification by estimating probability distributions over predefined classes. Representative fully supervised semantic segmentation methods include Unet [3] (which aggregates multiscale features through skip connections), PSPNet [4] (which uses a multiscale pyramid for feature extraction), and the DeepLab family [5] (which increases the receptive field size to improve feature extraction capabilities).

Additionally, a fully supervised segmentation model based on Transformer [6], such as TransUnet [7], uses a U-shaped encoder–decoder network combined with CNNs (local features), achieving fine-grained medical lesion segmentation and satisfactory results in radar sound data segmentation for RSIs [8]. Marsformer [9] further optimizes the encoder–decoder design with a feature enhancement module (FEM) and window transformer block (WTB), enhancing multiscale feature consistency and perception of local and contextual features. It also achieves good results in the segmentation of fine surface rocks in Mars scene tasks.

However, the aforementioned fully supervised methods not only rely on a large amount of data samples for model fitting,

Received 25 November 2024; revised 16 March 2025; accepted 29 April 2025. Date of publication 9 May 2025; date of current version 22 May 2025. This work was supported in part by the Key Project of the Scientific Research Plan of Hubei Provincial Department of Education under Grant D20231304, in part by Tianjin Science and Technology Plan Project under Grant 23YFYSHZ00190 and Grant 23YFZCSN00280, in part by the Open Fund of National Engineering Laboratory for Digital Construction and Evaluation Technology of Urban Rail Transit under Grant 2023ZH01, in part by Hunan Provincial Natural Science Foundation Project Department Union Fund under Grant 2024JJ8327, in part by Jiangxi Provincial Natural Science Foundation under Grant 20232ACB204032, in part by the Tibet Autonomous Region Science and Technology Major Project under Grant XZ202402ZD0001, and in part by the Deep Earth Probe and Mineral Resources Exploration-National Science and Technology Major Project under Grant 2024ZD1001003. *(Corresponding author: Xianjun Gao.)*

Jin Kuang, Xianjun Gao, Yuanwei Yang, Siyuan Dong, and Ji Dong are with the School of Geosciences, Yangtze University, Wuhan 430100, China (e-mail: gasquue@gmail.com; junxgao@yangtzeu.edu.cn; yyw_08@yangtzeu.edu.cn; dongsy2022@163.com; 15367065470@163.com).

Yuan Kou is with the First Surveying and Mapping Institute of Hunan Province, Changsha 421001, China (e-mail: kouyuan@xcyy.net.cn).

Meilin Tan and Zhiwei Wang are with Inner Mongolia Autonomous Region Surveying and Mapping Geographic Information Center, Hohhot 010050, China (e-mail: tanmeilin@whu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2025.3567159

but also require high-quality labeled annotations. In practical tasks such as RSI cross-domain cloud detection [10], cross-domain building detection [11], and cross-domain Mars terrain segmentation, the training dataset (source domain) is often scarce and small in scale, and the cost of manual labeling is high. Moreover, due to the influence of physical characteristics such as imaging sensor differences and spatiotemporal dynamics in remote sensing data, training a network with good cross-domain performance becomes more challenging. As a result, their generalization ability is limited, and when applied to RSIs with large spatiotemporal differences, the accuracy of cloud detection drops significantly (domain shift).

Therefore, addressing the domain shift between training and testing datasets, and improving cross-domain cloud detection performance, has become a key challenge in the field. Recently, unsupervised domain adaptation (UDA) [12] based on adversarial training (AT) have been widely applied to cross-domain cloud detection tasks in RSIs. AT introduces a domain discriminator that distinguishes between features from the source and target domains, guiding the generator to minimize the distribution gap between them. This enables the model to learn feature transfer from the source domain to the target domain without the need for target-domain labels. For instance, Mateo-García et al. [13] designed a cycle-consistent generative adversarial domain adaptation (CyCADA) framework [12] to improve the cloud detection performance by minimizing the statistical difference between satellite image sensors. Guo et al. [14] proposed group feature alignment (GFA) and entropy minimization (EM) to extract domain-invariant feature information, enhancing the performance of cross-domain cloud detection. TDAIF [10] enhances the cross-domain adaptation performance by generating a pseudo-target domain to reduce the discrepancy between the target and source domains and optimizing the decision boundary at the feature level. In sum, the UDA method based on AT improves cross-domain cloud detection performance. However, these methods still face significant limitations when applied to more complex scenarios [15], where terrain interference and uneven distribution of data add further challenges. In particular, training domain discriminators becomes difficult when there is a substantial discrepancy between the source- and target-domain distributions. Moreover, the parameters of the domain discriminator require careful tuning across different cross-domain cloud detection tasks. This is compounded by the tight coupling between the segmentation model and the domain discriminator, where fluctuations in one can adversely affect the performance of the other. These interactions may lead to significant errors in target-domain feature extraction during training, further exacerbating domain shift.

Therefore, to address these challenges and further improve the ability of domain adaptation to better fit the data distribution within the domain, while also enhancing its generalization performance, existing approaches attempt to tackle these issues through various strategies. Specifically, IntraDA [16] utilizes entropy sorting to divide target-domain samples, facilitating the generation of high-quality pseudo-labels, which improves domain invariance and semantic adaptation. IterDANet [17] adopts an iterative domain adaptation framework, combining the generator selection strategy (GSS) and entropy sorting to perform multi-subdomain unsupervised training in the target domain, effectively mitigating the impact of noise in the target domain on domain adaptation. MIC [18] enhances the robustness of the model by learning contextual semantic information from the target domain. DAFormer [19] integrates a Transformer encoder with multilevel context-aware feature fusion to enhance feature extraction, while also using pseudo-label joint training to reduce domain shifts. UDAFormer [20] leverages the output-guided biased sampling (OGBS) module and modified augmentation regularization (MAR) to boost sensitivity to low-perception category features and reduce the impact of ineffective augmentations on domain adaptation.

ProCA [21] generates high-quality pseudo-labels through class prototypes, enhancing the exploration of interclass relationships and providing effective supervision during model training. Li et al. [22] use multiple weakly supervised loss functions to alleviate domain shift caused by large differences in data distribution.

Despite these methods attempting to address the domain feature shift and domain adaptation problem in different ways, they still face the following challenges.

1) Lack of a progressive mechanism for feature interaction, failing to fully explore how features between domains can be efficiently fused.
2) How to appropriately adapt domain-invariant features and intradomain semantic information to generate high-quality pseudo-labels remains an unresolved issue.
3) Pseudo-labels in self-supervised training (ST) may introduce noise [23], and the accumulation of this noise could lead to negative optimization and model degeneration.

Therefore, to address the two major challenges in cross-domain cloud detection for HRSIs: 1) the inability to effectively integrate interdomain feature information with intradomain semantic information and 2) the difficulty of generating high-quality pseudo-labels to mitigate domain shift. We propose a progressive pseudo-label learn for cross-domain cloud detection in high-resolution remote sensing. On the one hand, the method significantly improves the model's ability to fit the data distribution within the domain through high-quality pseudo-labels. On the other hand, it enhances the adaptability of semantic features in the target domain by efficiently leveraging domain-invariant feature information, thereby alleviating the negative optimization effects caused by low-quality pseudo-labels. Finally, by incorporating high-quality pseudo-labels and pseudo-label-constrained intradomain feature mining loss (PCIF Loss), the method achieves a significant improvement in cross-domain cloud detection performance.

Specifically, we first apply domain AT to allow the model to align the data distributions of both the source and target domains, enabling the learning of domain-invariant feature information. To resolve potential conflicts between domain-invariant features and intradomain semantic information, we design the online domain-invariant feature-guided pseudo-label generation (OPLG) strategy. This strategy enables ST using a domain segmentation model as domain-invariant features (feature level), while integrating rich intradomain

semantic information from the current segmentation network iteration. Additionally, the learning intradomain unaligned feature (LIUF) strategy operates at the pixel level, guiding the generation of reliable pseudo-labels and suppressing noise interference during the self-training phase.

Finally, recognizing the challenge of directly learning target-domain semantic information from domain-invariant features, we introduce the PCIF Loss, which progressively encourages the clustering of positive samples and the separation of negative samples in the semantic feature space, thereby improving segmentation accuracy. PTDA achieves strong performance in cross-domain cloud detection tasks by synergistically combining the above strategies, making it highly valuable for practical applications in production scenarios.

In summary, our main contributions are as follows.

1) We proposes a progressive pseudo-label learning for cross-domain cloud detection in high-resolution remote sensing (PTDA). By progressively integrating domain-invariant features with intradomain semantic information, the method enhances the ability to mine cross-domain features, thereby improving the accuracy of cross-domain cloud detection.
2) OPLG and LIUF make the model fully cooperate domain-invariant feature information and domain semantic information in the training stage, so as to adaptively generate pseudo-labels to guide the model for ST, which fully improves the generalization ability of the model.
3) In the model training phase, PCIF Loss is introduced to progressively cluster positive and negative samples in the feature space. By fully utilizing pseudo-labels to mine intradomain semantic information, it facilitates the gradual integration of domain-invariant features with intradomain semantic features.
4) Finally, the effectiveness of our method is demonstrated by three HRSI cross-domain cloud detection tasks. Especially, we achieved 63.99%, 58.14%, and 58.82% mean intersection over union (mIoU) on MS Cloud → HRC_WHU Cloud (MS2WHU), MS Cloud → WHUS2-CD + (MS2S2), and HRC_WHU → 95 Cloud (WHU295), respectively, outperforming existing methods.

The organization of this article is as follows. Section II briefly describes some related works. Section III elaborates on the PTDA domain adaptation method. In Section IV, MS Cloud, HRC_WHU Cloud, WHUCD + S2 Cloud, 95 Cloud datasets, experimental configurations, and experimental analysis on the performance of the proposed method are introduced, and finally, the experimental results are discussed in depth. Section VI is the conclusion of this article.

II. Related Work

A. Cloud Detection of RSI

The task of cloud detection in HRSI is crucial for obtaining high-resolution, clear satellite images. Cloud detection methods can be divided into two categories: traditional artificial feature-based methods and deep learning methods.

Traditional artificial feature-based methods [24] distinguish clouds by thresholds or machine learning classifiers [25] according to their physical, spatial, and spectral features, such as brightness, texture, and shape. For example, Zhu et al. [26] used temperature, spectral change rate, and brightness as a joint conditional distribution and used thresholds for cloud detection. However, as HRSI contains more refined object-level information, only manually designed features are insufficient for production requirements. Therefore, machine learning methods have been gradually introduced to construct the feature decision trees or use support vector machine (SVM) and random forest methods to fit manually defined features, enhancing model applicability [27], [28]. However, tree-based methods struggle to capture local feature details in HRSI, providing only coarsely global features. This limitation leads to significant feature loss, especially creating internal "hole effects," which hinders their effectiveness in complex scenes with ground object interference.

Recently, with the rapid development of deep learning, CNNs have been widely used in cloud detection, especially in complex scenes, because of their powerful feature extraction [29] capabilities. For example, Yang et al. [30] used class correlation and knowledge distillation to address thick and thin clouds in complex scenes, effectively reducing the cloud hole effects. Zhang et al. [31] propose a channel attention module with multiscale feature extraction to enhance global feature consistency. CDNetv2 [32], combining multiple attention mechanisms and adaptive features, decodes semantic features into cloud location data, promoting effective feature interaction and reducing noise interference from "artifact" features. Zhang et al. [33] proposed a multitask driven reconfigurable network (MTDR-Net), which promoted RSI global information interaction through multiprojection and multiscale feature fusion. However, these models designed for specific cloud detection tasks lack generalization. When applied to cross-domain cloud detection, their effectiveness often falls short due to cloud diversity and ground object complexity.

B. Unsupervised Semantic Segmentation of RSIs

Due to significant spatial and temporal data distribution differences in global HRSI, the fully supervised models may achieve good performance in the source domain but struggle in the target domain. UDA can transfer learned feature knowledge from a labeled source domain to an unlabeled target domain and has been widely used in image classification [34] and semantic segmentation tasks [35]. UDA methods are divided into AT and ST.

The core idea of AT is to align the global or local feature distribution, allowing the model to learn invariant semantics across source and target domains. Meanwhile, it enhances the mining of the target-domain features, optimizing the decision boundary. For example, AdvEnt [36] used pixel entropy loss to improve segmentation. Tsai et al. [37] align global features via multiscale AT. Wang et al. [38] introduced category-based fine-grained discrimination, and CLAN [39] adjusted the gradient based on domain error weights to alleviate parameter sensitivity in domain discrimination. Guo et al. [14] proposed

a domain-invariant approach based on GFA and EM, which effectively improved the performance of cross-domain cloud detection. Mateo-García et al. [40] proposed a simple transfer method using Landsat-8 and Proba-V sensors, achieving high accuracy by accounting for physical properties. However, these methods optimize the data distribution difference between the source domain and the target domain. When the target-domain variance is too large, these networks cannot learn the semantic consistency in the target domain, and segmentation performance decreases significantly.

The ST-based methods generate pseudo-labels for the target domain, using them to guide iterative segmentation training. These methods enhance generalization and robustness by mining potential features from the target domain during the unsupervised training [41]. For example, PyCDA [42] generates high-quality pseudo-labels by constructing diverse attributes in the target domain, promoting segmentation accuracy during the domain adaptation process. Additionally, some researchers seek to explore more supervisory information from the target domain to further boost UDA segmentation performance. Liu et al. [43] introduced semantic boundary features to better adapt the segmentation decision boundary. Refign [44] uses an uncertainty-aware dense matching network to align objects in both easy and challenging images, combining an adaptive label correction strategy to reduce mispredictions. However, these methods are limited by not fully utilizing domain-invariant and rich semantic features for efficient interaction, which increases the influence of feature uncertainty on pseudo-label quality during training. Therefore, esl [45] proposed an entropy-based ST method, which uses entropy distribution to improve pseudo-label quality and alleviate interference from incorrect class feature information. Beyond CNN approaches, DAFormer [19] employs a more advanced Transformer feature extractor [46]. This enhances semantic features with a stronger feature extraction network and samples rare categories frequently. Consequently, the model can better learn the data distribution with a low proportion, improving the pseudo-label quality to a certain extent.

Inspired by the above work, we recognized that generating high-quality pseudo-labels can further acquire intradomain semantic information to improve the UDA segmentation performance. However, the current ST methods do not efficiently integrate domain-invariant features with semantic information, resulting in low-quality pseudo-labels. Therefore, how to generate high-quality reliable pseudo-labels remains a key challenge.

C. Self-Supervised Learning via Entropy Uncertainty

Uncertainty entropy distribution [36] can aid in generating higher quality pseudo-labels. Recent UDA methods apply EM to capture domain-invariant and intradomain semantic features from the target domain, thereby enhancing pseudo-label quality. For example, IntraDA [16] introduced entropy-guided self-supervised learning, using entropy to separate target-domain samples, ensuring simpler samples receive more reliable pseudo-labels. Wang et al. [47] used the model output entropy as a confidence measure in cross-domain inference to identify high-confidence samples from unlabeled datasets for feature learning. Zou et al. [48] proposed a class-balanced self-training (CNST) to mitigate class proportion effects on pseudo-label quality. Wang et al. [49] proposed U2PL, which separates reliable and unreliable pixels by predicting the entropy threshold, classifies reliable pixels as negative samples, and dynamically adjusts the threshold through training. Inspired by the above work, we find that "entropy" not only provides reliable pseudo-labels for ST, but also promotes effective interaction between domain-invariant features and intradomain semantic information.

In this article, we first use the joint condition of entropy and IoU to control the generation of reliable coarse labels. We then design OPLG and LIUF: OPLG focuses on learning domain-invariant features and intradomain semantic information at the feature level, while LIUF refines intradomain semantic information at the pixel level. These steps effectively enhance the interaction between domain-invariant features and intradomain semantic information, ensuring the generation of reliable pseudo-labels.

III. Methodology

In this section, we first give a brief overview of PTDA and illustrate the overall idea in Section III-A. The details of our approach and the problems solved at each stage are described in Sections III-B and III-C. The definition of PCIF Loss will be described in detail in Section III-D.

A. Overview of the PTDA

As shown in Fig. 1, PTDA for cross-domain cloud detection includes three parts: initial training includes training the pretrained segmentation model and coarse-reliable pseudo-label generation, progressive intradomain feature mining includes an OPLG strategy and LIUF, and PCIF Loss.

B. Initial Training

1) Pretrained Segmentation Model: Through AT, the segment model f_p effectively learns domain-invariant features between the source and target domains. First, f_p is used to train the source dataset $\mathbb{R}_s \in \{(\mathcal{X}_n^s, \mathcal{Y}_n^s)\}_{n=1}^N$ and simultaneously train interdomain discrimination model d_θ^{inter} (continuously stacked convolutional layers: the convolution kernel is 4×4, the step size is 2, and the downsampling is 32 times) is used to distinguish the source-domain data and the target-domain data $\mathbb{R}_t \in \{(\mathcal{X}_n^t)\}_{n=1}^N$. We use cross-entropy loss and binary cross-entropy loss to optimize the f_p and the domain classification model d_θ^{inter}, respectively, whose equations are defined as follows:

$$\mathcal{L}_{\text{seg}}^{\text{ce}} = -\sum_{q}^{H\times W}\sum_{k}^{C} y_{qk}^{(c)} \log\left(p_{qk}^{(c)}\right) \tag{1}$$

$$\mathcal{L}_{\text{Adv}} = \sum_{q}^{H\times W}\sum_{k}^{C}\left(y_q^{(c)} \log\left(\mathcal{D}_q^{(c)}\right) + \left(1 - y_q^{(c)} \log\left(1 - \mathcal{D}_q^{(c)}\right)\right)\right) \tag{2}$$

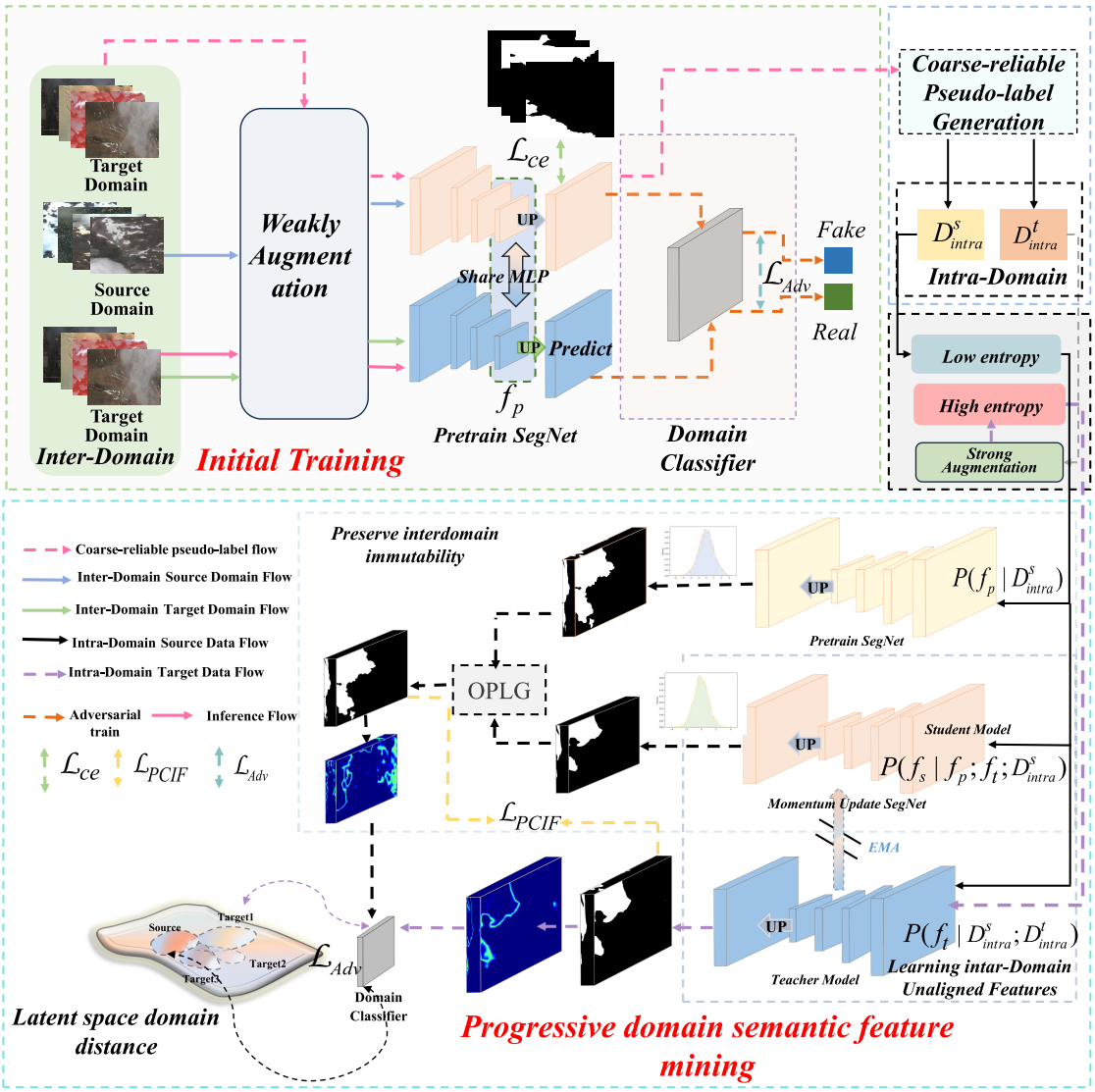

Fig. 1. Overview of PTDA framework (where the green box is the first stage of domain-invariant feature learning, the dashed box is the sample division of the target domain through entropy and IoU, the dashed box is the data enhancement means of different domains, and the cyan box is the progressive pseudo-label learning stage).

where $H \times W$ is the size of the feature layer, C is the number of classifications, k is the class sequence number, p_{qk} represents the probability distribution of the Softmax segmented feature map, and $\mathcal{D}$ represents the probability distribution of the domain feature map.

Finally, the $\mathcal{L}_{\text{pretrain}}$ of the pretrained segmentation model is defined as follows:

$$\mathcal{L}_{\text{pretrain}} = \arg\min_{\theta=1}\left(\mathcal{L}_{\text{seg}}^{\text{ce}} + \arg\max_{\theta=1}(\mathcal{L}_{\text{Adv}})\right). \tag{3}$$

2) Coarse-Reliable Pseudo-Label Generation: Subsequently, we perform inference on f_p using $\mathbb{R}_t$ to obtain intradomain high-confidence samples $\mathbb{D}_{\text{intra}}^{s} \in \{(\mathcal{T}_n^s, \mathcal{T}_n^{\text{pse}})\}_{n=1}^{N}$ (pse: pseudo_label) and intradomain low-confidence samples $\mathbb{D}_{\text{intra}}^{t} \in \{(\mathcal{T}_n^t)\}_{n=1}^{N}$.

Due to the large differences in cloud data across domains, relying solely on entropy distribution [50] to count samples can lead to misclassification. Specifically, entropy distribution often assigns high entropy distribution to boundary areas of thin clouds and internal areas of thick clouds, resulting in the misclassification of these regions as low-quality samples. If these highly uncertain pseudo-labels are used in training, they can introduce significant noise into the network, thereby disrupting domain invariance. To address this, we propose a simple method during the target-domain sample partitioning phase, which combines IoU; the label comes from the target domain with limited annotations and entropy distribution information to assess the quality of pseudo-labels. The coarse-reliable pseudo-label generation is shown in Fig. 2. Additionally, to fully utilize the domain-invariant properties of f_p, we employ class-level divergence Shannon entropy to assess the entropy uncertainty of each target sample, as defined in the following:

$$E^{(h,w)} = -\frac{1}{\log(c)} \sum_{c} p^{(h,w,c)} \log\left(p^{(h,w,c)}\right) \tag{4}$$

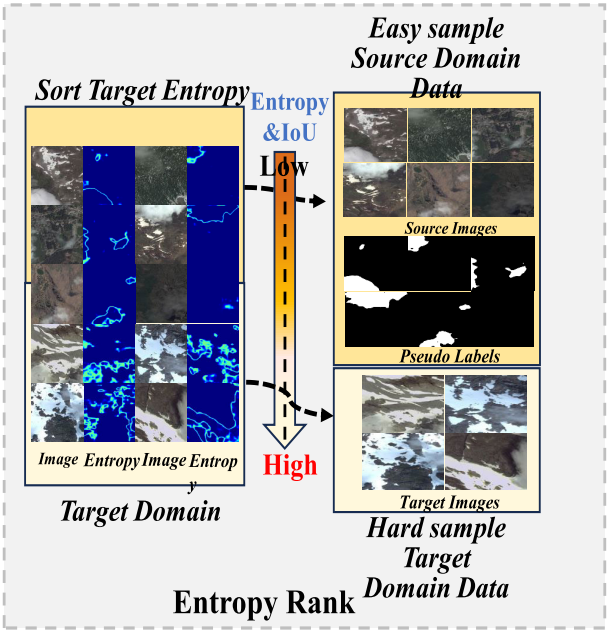

Fig. 2. Overview of coarse-reliable pseudo-label generation.

where c is the number of classes, $p^{(h,w,c)}$ is the entropy probability distribution, and h and w are the dimensions of the feature map.

Finally, the pseudocode for coarse-reliable pseudo-label generation is shown in Algorithm 1.

C. Progressive Intradomain Feature Mining

A key challenge in UDA is efficiently integrating domain-invariant features with intradomain semantic information, while mitigating the noise interference from pseudo-labels. To address this, we first design OPLG to learn domain-invariant features and intradomain semantic information at the feature level, ensuring the generation of reliable pseudo-labels. Building on this, we then introduce LIUF, which focuses on refining intradomain semantic information at the pixel level to further enhance the target-domain features.

1) Learning Intradomain Unaligned Features: The previous student–teacher model [51] in the ST of UDA is usually only used for model smoothing or for consistency feature alignment at the feature level, and rarely pays attention to sensitive detailed features and semantic features from the pixel level to guide the generation of pseudo-labels. Therefore, this article designs LIUF as shown in Fig. 3. First, f_t is used to explore the semantic feature information in the target domain, and then, EMA [52] is used to update the momentum of f_s. At this time, f_s can not only fully express the domain-invariant feature information, but also effectively extract the complex semantic information in the target domain, so as to progressively optimize the quality of the pseudo-label from the feature level to the pixel level to ensure the performance of the segmentation model in the domain. We describe this process using the following equation:

$$f_s^{(T)} = \alpha f_t^{(T-1)} + (1-\alpha) f_s^{(T-1)} \tag{5}$$

Algorithm 1 Coarse-Reliable Pseudo-Label Generation

Input: Target domain data $\mathbb{R}_t \in \{(\mathcal{X}_n^t)\}_{n=1}^N$, IoU threshold $\zeta \in (0.5, 0.85]$, Splitting weight coefficients α, segmentation model f_p, and label (*comes from the target domain with limited annotations*)

Output: $\mathbb{D}_{intra}^s$, $\mathbb{D}_{intra}^t$

Initialize: $\mathbb{D}_{intra}^s \leftarrow []$, $\mathbb{D}_{intra}^t \leftarrow []$, $\ell \leftarrow []$
for $\mathbb{X}_{cur}$ in $\mathbb{R}_t$ **do**
  Get probabilities: $P_{cur} = f_p(\mathbb{X}_{cur}; \theta)$
  Get Entropy: P_{cur} based on Eq. (4)
  Get Pseudo Label: $Pseudo_{cur} \leftarrow \arg\max(P_{cur})$
  Get Foreground IoU: $IoU_{cur} \leftarrow$ Compute_IoU$(Pseudo_{cur}, label_{cur})$
  Get $\mathbb{X}_{name} \leftarrow$ Get_name$(\mathbb{X}_{cur})$
  $\ell \leftarrow$ [Entropy, $Pseudo_{cur}$, IoU_{cur}, $\mathbb{X}_{name}$]
end for
Sort ℓ in ascending order based on entropy values
Obtain the indices for sample partitioning: $\mathcal{W} \leftarrow$ Split_sample(ℓ, α)
Partition: $\ell_{intra}^s \leftarrow \ell[:\mathcal{W}]$, $\ell_{intra}^t \leftarrow \ell[\mathcal{W}:]$
for $\ell_{intra}^{s\_cur}$ in ℓ_{intra}^s **do**
  if $\zeta_{\min} \le IoU_{cur} \le \zeta_{\max}$ **then**
    $\mathbb{D}_{intra}^s.append([\text{Read\_image}(\mathbb{X}_{name}^{cur}), Pseudo_{cur}])$
    Save pseudo label
  end if
end for
for $\ell_{intra}^{t\_cur}$ in ℓ_{intra}^t **do**
  $\mathbb{D}_{intra}^t.append([Read\_image(\mathbb{X}_{name}^{cur})])$
end for
return $\mathbb{D}_{intra}^s$, $\mathbb{D}_{intra}^t$

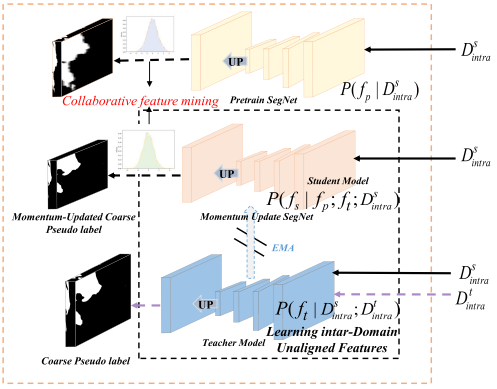

Fig. 3. Learning intra-domain unaligned features (LIUF).

where α is a momentum update factor, and the usual value range is $\alpha \in (0.995, 0.999)$. $f_\cdot^{(T)}$ denotes parameters of the segmentation model at time step T. In this article, $\alpha = 0.995$. The parameters of f_s are initialized randomly, and f_t is an intradomain segmentation model.

2) Online Domain-Invariant Feature-Guided Pseudo-Label Generation: Due to the large domain gap in the RSI domain, in order to avoid the failure of cross-domain cloud detection task caused by over-fitting wrong pseudo-labels in the ST process, this article designs the OPLG, as shown in Fig. 4. The characteristic that f_p has domain-invariant feature information

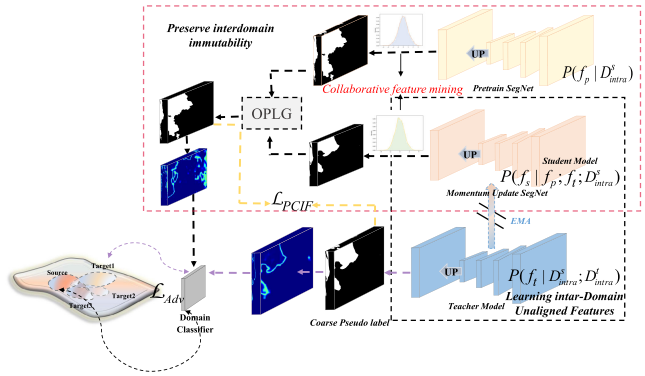

Fig. 4. Overview of the OPLG.

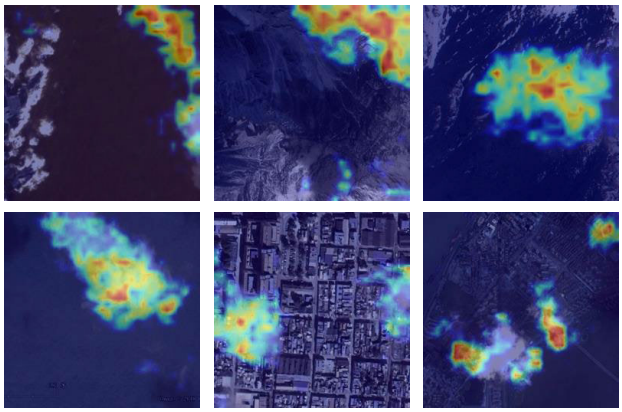

Fig. 5. PCIF Loss (Grad-CAM visualization [56]).

is used as the benchmark feature information to ensure that the model is always in the direction of positive optimization parameters, and then combined with the momentum update segment f_s [based on (5)] in the domain to generate pseudo-labels. We can describe the above process using the following equation:

$$P\left(f_p^{(c)}|D^s_{\text{intra}}\right) = -\sum_c \log \frac{f_p^{(c)}\left(D^s_{\text{intra}}\right)}{\sum_c f_p^{(c)}\left(D^s_{\text{intra}}\right)} \tag{6}$$

$$P\left(f_s^{(c)}|f_p^{(c)}; f_t^{(c)}; D^s_{\text{intra}}\right) = -\sum_c \log \frac{f_t^{(c)}\left(D^s_{\text{intra}}\right)}{\sum_c f_t^{(c)}\left(D^s_{\text{intra}}\right)} \tag{7}$$

where $P(f_p^{(c)}|D^s_{\text{intra}})$, $P(f_s^{(c)}|f_p^{(c)}; f_t^{(c)}; D^s_{\text{intra}})$ is the probability of D^s_{intra} being split by f_p and f_s. They have the ability to express domain-invariant feature information and semantic information in the target domain, where c is the classes. To extract high-quality pseudo-labels from them, we do the following:

$$P\left(f_p^{(c)}|D^s_{\text{intra}}\right) = \begin{cases} P\left(f_p^{(c)}|D^s_{\text{intra}}\right), & \text{if } P\left(f_p^{(c)}|D^s_{\text{intra}}\right) > \tau \\ 0, & \text{other} \end{cases} \tag{8}$$

$$P\left(f_s^{(c)}|f_p^{(c)}; f_t^{(c)}; D^s_{\text{intra}}\right) = \begin{cases} 0, & \text{if } P\left(f_s^{(c)}|f_p^{(c)}; f_t^{(c)}; D^s_{\text{intra}}\right) < \tau \\ P\left(f_s^{(c)}|f_p^{(c)}; f_t^{(c)}; D^s_{\text{intra}}\right), & \text{other} \end{cases} \tag{9}$$

where $P(f_p^{(c)}|D^s_{\text{intra}})$ represents the domain-invariant class probability of $f_p^{(c)}$ under the inference of D^s_{intra}, c is the number of classes, and $P(f_s^{(c)}|f_p^{(c)}; f_t^{(c)}; D^s_{\text{intra}})$ denotes the class probability of domain-invariant semantic information under the joint model parameters of τ that can prevent the error class information generated by the error class probability distribution of complex samples

$$y^{\text{invar}} = \arg\max_c \left(P\left(f_p^{(c)}|D^s_{\text{intra}}\right)\right) \tag{10}$$

$$y^{\text{var}} = \arg\max_c \left(P\left(f_s^{(c)}|f_p^{(c)}; f_t^{(c)}; D^s_{\text{intra}}\right)\right) \tag{11}$$

where y^{invar} pseudo-label contains the semantic information of domain-invariant features, y^{var} pseudo-label contains intradomain semantic information, and c represents the class information. In order to effectively combine domain feature invariance and domain segmentation features, we perform the following calculation for each category c to obtain high-quality pseudo-label:

$$\tilde{y}^{(c)} = \mathbb{Z}_{\mathcal{D}\in\mathcal{D}^s_{\text{intra}}}\left\{y^{(c)}_{\text{invar}}|y^{(c)}_{\text{var}}\right\}. \tag{12}$$

D. Pseudo-Label-Constrained Intradomain Feature Mining Loss

In the early stages of intradomain segmentation training, it is extremely difficult to extract target-domain features from domain-invariant ones. Specifically, a large number of negative samples (high entropy) and positive samples (low entropy) are clustered in the low-entropy "thick cloud" region and high-entropy "thin cloud contour" region, respectively. The cross-entropy loss does not effectively promote the clustering of positive samples and the separation of negative samples in the feature space.

Therefore, this article proposes PCIF Loss (Fig. 5 highlights how our loss function encourages the network to enhance its response to thin cloud contours and features within thick cloud regions. This promotes the clustering of positive sample features while pushing negative sample features apart in the feature space, demonstrating the strong cross-domain performance of our method in complex cloud detection scenarios) that well corresponds to meaningful patterns in the class, which verifies that our method has better localization ability in the target domain. The negative samples in the low entropy region are constantly far away from the feature classification that reaches the feature level. Then, the semantic consistency constraint is added to the f_t to clarify the multiscale feature semantic consistency to further collaboratively increase the classification semantic gradient information, so that the segmentation parameters of the model are constantly developing toward positive optimization. Finally, in order to further enhance the robustness and generalization ability of the model, the adversarial gradient is added to the ST to improve the feature extraction ability of the f_t segmentation model.

First, we cluster the intradomain semantic samples using pseudo-label $\tilde{y}^{(c)}$, as follows:

$$f^{(c)}_{\text{pos}} = f_t^{(c)} \odot \tilde{y}^{(c)} \tag{13}$$

$$f^{(c)}_{\text{neg}} = f_t^{(c)} \odot \left(1 - \tilde{y}^{(c)}\right) \tag{14}$$

where $\odot$ represents the elementwise multiplication; and $f_{\text{pos}}^{(c)}$ and $f_{\text{neg}}^{(c)}$ denote the distribution of positive and negative samples of the pseudo-label $\tilde{y}^{(c)}$ in the latent space, respectively.

Second, we are inspired by the self-attention mechanism [53]; we find that (Q, K, and V) can make full use of the spatial feature relationship of the feature map and estimate the probability of similar features without increasing the amount of computation, thus enhancing the high-quality feature expression. Hence, we use a self-attention mechanism to collaboratively mine intradomain semantic features with domain-invariant semantic information

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK.T}{\sqrt{\dim_{\text{head}}}}\right)V.T \tag{15}$$

$$\tilde{f}_{\text{pos}}^{(c)} = \text{Attention}\left(f_{\text{pos}}^{(c)}, \tilde{y}^{(c)}, \tilde{y}^{(c)}\right) \tag{16}$$

$$\tilde{f}_{\text{neg}}^{(c)} = \text{Attention}\left(f_{\text{neg}}^{(c)}, \tilde{y}^{(c)}, \tilde{y}^{(c)}\right) \tag{17}$$

where $\tilde{f}_{\text{pos}}^{(c)}$ represents the feature vector most similar to the positive samples in $f_{\text{pos}}^{(c)}$ under the guidance of pseudo-label $\tilde{y}^{(c)}$, $\tilde{f}_{\text{pos}}^{(c)}$ represents the feature vector most similar to the negative samples, and c denotes the class. Under attention-based feature aggregation, domain-invariant features and intradomain semantic information interact through iterative coupling, adaptively enhancing the model's feature discrimination ability. This allows the segmentation network f_s to efficiently capture intradomain semantic information and aggregate it with domain-invariant features from f_t, thereby improving the effective utilization of features and enhancing the cross-domain cloud detection performance of the segmentation model on target-domain HRSI images.

In addition, to eliminate the problem of ambiguous expression, multi-scale feature semantic consistency is employed for gradient guidance, which is defined as follows:

$$\mathcal{L}_{\text{consist}} = \frac{1}{h \times w}\sum_{i}^{h\times w}\sum_{c}^{C}\left(f_t^{(0)} - f_t^{(1)}\right)^2 \tag{18}$$

where $h \times w$ represents the dimensions of the feature map, i is the feature sequence, and c is the number of classes. In $f_t^{(i)}$, i represents the feature vectors at different scales. Guided by semantic consistency, this effectively alleviates the ambiguity issue of multiscale features, enhancing the model's feature extraction performance on HRSI images.

Finally, to further improve the model's robustness, inspired by GAN [54], [55], adversarial gradients are introduced during the training process. The collaborative use of adversarial gradients enhances the segmentation model's ability to extract intradomain features

$$\mathcal{L}_{\text{Adv}}^{\text{intra}} = \min_{\theta}\left(\frac{1}{|x^s|}\sum_{x}\mathcal{L}_{\text{Adv}}\left(d_\theta^{\text{intra}}\left(x^s, 0\right)\right) + \frac{1}{|x^t|}\sum_{x}\mathcal{L}_{\text{Adv}}\left(d_\theta^{\text{intra}}\left(x^t, 1\right)\right)\right) \tag{19}$$

where j in x^j represents the domain from which the sample data originate, and $\mathcal{L}_{\text{Adv}}$ is the AT loss, as referenced in (2). d_θ^{intra} is the intradomain adversarial loss (with the same network structure as d_θ^{inter}), where 0 represents the source domain and 1 represents the target domain. Finally, the efficient guidance of domain-invariant features to intradomain semantic features enables effective aggregation of class-specific domain features, improving the cross-domain cloud detection segmentation performance. The PCIF Loss is defined as follows:

$$\mathcal{L}_{\text{pos}} = -\sum_{i}^{h\times w}\sum_{c}^{C}\tilde{y}^{(c)}\log\left(f_{\text{pos}}^{(c)}\right) \tag{20}$$

$$\mathcal{L}_{\text{neg}} = -\sum_{i}^{h\times w}\sum_{c}^{C}\tilde{y}^{(c)}\log\left(f_{\text{neg}}^{(c)}\right) \tag{21}$$

$$\mathcal{L}_{\text{PCIF}} = \arg\min_{\theta}\left(\mathcal{L}_{\text{pos}} + \mathcal{L}_{\text{neg}} + \mathcal{L}_{\text{Adv}}^{\text{intra}} + \mathcal{L}_{\text{consist}}\right) \tag{22}$$

where $\mathcal{L}_{\text{pos}}$ represents the cross-entropy loss for positive samples, $\tilde{y}^{(c)}$ represents the pseudo-labels, $\tilde{f}_{\text{pos}}^{(c)}$ represents the feature vector for positive samples, $\mathcal{L}_{\text{neg}}$ represents the cross-entropy loss for negative samples, $\tilde{f}_{(\text{neg})}^{(c)}$ represents the feature vector for negative samples, $\mathcal{L}_{\text{Adv}}^{\text{intra}}$ represents the domain-specific adversarial loss, and $\mathcal{L}_{\text{consist}}$ represents the semantic consistency.

IV. Experiments and Analysis

The experiment uses a self-built MS Cloud dataset and three public data HRC_WHU Cloud, 95 Cloud, and WHUS2-CD + Cloud detection datasets for cross-domain experimental analysis. In order to evaluate our proposed method and current state-of-the-art (SOTA) methods, three sets of domain adaptation experiments are designed, one of the groups is MS Cloud → HRC_WHU Cloud (MS2WHU), one group is HRC_WHU Cloud dataset → 95 Cloud (WHU295) dataset, and MS Cloud → WHUS2-CD + (MS2S2) domain adaptation.

A. Dataset Description and Metrics

1) Dataset Description:

a) MS Cloud (MS): The dataset includes 2328 WorldView-02 RSIs (0.46-m resolution) from 2014 to 2020, covering four seasons and various landscapes such as wasteland, grassland, paddy fields, and snow-capped mountains. Both thin and back clouds are labeled as clouds, ensuring diversity in surface vegetation types and complex cross-seasonal imagery.

b) HRC_WHU Cloud (WHU): The dataset consists of 150 high-resolution images (RGB, 0.5–15-m resolution) collected from Google Earth, covering five land cover types: water, vegetation, urban, snow/ice, and barren. Thin clouds, visually detectable, are labeled as clouds. The original images are split into 256 × 256 patches, resulting in a total of 5504 images. Cloud and clear sky pixels are assigned mask values of 255 and 0, respectively.

c) 95 Cloud dataset extends the 38 Cloud dataset: Training images are from 75 Landsat 8 Collection 1 Level-1 scenes, primarily in North America. Each image has four spectral bands: red (band 4), green (band 3), black (band 2), and near-infrared (band 5). Thin clouds and fog are labeled as clouds, and natural scenes are synthesized using 432 combined bands for training.

TABLE I
TRAINING SET AND TEST SET DATA

Dataset	Train Image	Test Image	Image Size
MS Cloud	2095	233	256×256
WHU Cloud	4953	551	
WHUS2-CD+	5167	575	
95 Cloud	23670	2631	

TABLE II
THREE CROSS-DOMAIN CLOUD DETECTION TASKS

Task	Source Domain	Target Domain
MS2WHU	MS Cloud	WHU Cloud
WHU295	WHU Cloud	95 Cloud
MS2S2	MS Cloud	WHUS2-CD+

d) WHUS2-CD + (S2): The WHUS2-CD+ dataset was collected and annotated by researchers from Wuhan University, including 36 scenes from Chinese Mainland, including forests, snow, cities, deserts, and other land types.

2) Metrics: IoU, $F1$-Score, Recall, as well as class pixel accuracy (CPA) and mean CPA (mCPA) are used to evaluate the background and cloud classes, while overall accuracy (OA) and mIoU serve as comprehensive performance metrics. The quality of pseudo-labels is assessed qualitatively using the Entropy map [57], and t-SNE [58] is used for visual evaluation of segmentation performance.

B. Experimental Detail

1) Detailed Parameter Setting: The hardware setup includes an Intel Xeon Gold 6348 CPU, Nvidia A30, and Nvidia RTX 3090 GPUs. The software environment runs on Windows with Python 3.7 and PyTorch 1.13.1 + cu117. The Adam optimizer is used with cosine decay for the learning rate (min: $1e^{-6}$ and max: $2.5e^{-4}$) and a weight decay of $1e^{-3}$. Images are randomly sampled and resized to 256×256 with a batch size of 32. The maximum number of iterations for the domain transfer task is 3.5k. During pretraining, only normalization is applied. For fair comparison, all methods, including the proposed one, use the same baseline parameters, are trained from scratch, and use ResNet50 for feature extraction and DeepLabv2 for feature decoding.

2) Cross-Domain Cloud Detection Task: The experiment splits the training and test sets for four datasets (Table I), maintains the original image aspect ratio, and avoids scaling during training. Images with a length or width smaller than 256 are padded with black edges. The three cross-domain cloud detection tasks (when evaluating the accuracy on the target domain, the training and test sets are combined) are shown in Table II, and Fig. 6 illustrates the distribution of cloud content levels in the cross-domain cloud detection task.

C. Cross-Domain Cloud Detection Approaches for Comparison

To evaluate the effectiveness of the proposed cross-domain cloud detection method, we compare it with mainstream UDA methods, which fall into four categories.

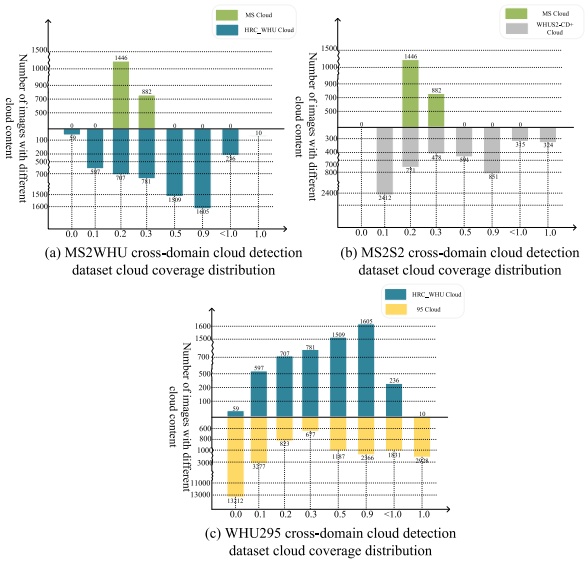

Fig. 6. Data distribution between the source domain and the target domain in the cross-domain cloud detection task (X-axis: cloud coverage and Y-axis: the number of images).

1) *Adversarial UDA Methods:* These methods, like AdvEnt, use entropy AT to minimize entropy as the optimization objective, improving cross-domain performance on specific datasets.
2) *Self-Training-Based UDA Methods:* Methods like esl use entropy as a confidence index for pseudo-label generation, guiding self-supervised learning through EM. IntraDA also uses entropy to sort target-domain samples, generating reliable pseudo-labels for adversarial ST. PDA [59] takes this further with output-level domain adaptation network (OL-DAN) and domain fusion domain adaptation network (DF-DAN) to iteratively align source- and target-domain distributions, extracting domain-invariant features.
3) *Advanced Feature Extractors:* DAFormer uses a Transformer-based architecture to enhance pseudo-label quality by incorporating multilevel context-aware feature fusion during self-training.
4) *Prototype Learning-Based UDA Methods:* ProCA utilizes prototype learning to embed category information into prototypes, aligning the target domain through class-centric distributions to improve domain-invariant feature learning. For our baseline, the feature extraction network is ResNet50, and the decoder is DeepLabv2, trained only on the source domain (Only_Source).

D. Comparative Experimental Results and Analysis

1) Cross-Domain Cloud Detection Results in MS2WHU: To fully evaluate the effectiveness of the proposed method, a comparison with mainstream UDA methods was conducted. As summarized in Table III, our method outperforms other approaches in key metrics, including IoU (Background), Recall (Background, Cloud), $F1$ (Background, Cloud), mCPA, OA, and mIoU. Notably, OA improves by 6.1% and IoU (Cloud) by 14.55% over the baseline (Only_Source).

Fig. 7. MS2WHU cross-domain visualization result (green: background and white: cloud).

1) IoU (Cloud) of our method is 54.07%, the highest among all compared methods, surpassing AdvEnt (13.72%) and esl (3.24%), both of which use EM. Our approach benefits from high-quality pseudo-labels, which integrate domain-invariant and semantic adaptation features. These labels help progressively transfer semantic information from the feature to the pixel level, ensuring positive optimization during ST. Moreover, the PCIF Loss effectively reconciles domain-invariant and semantic features, optimizing model performance.
2) Our method achieves IoU (Cloud) 10.83% higher than IntraDA, demonstrating that relying solely on entropy for intradomain sample division is insufficient for improving cross-domain cloud detection performance. In cases where cloud boundaries or interior features have high entropy, reliable pseudo-labels can be lost, making the model less effective. By incorporating IoU in the evaluation, our method improves the generation of reliable pseudo-labels.
3) DAFormer, ProCA, and PDA achieve only marginal improvements in IoU (Cloud) (1.92%, 4.38%, and 0.56%, respectively) compared to Only_Source. DAFormer struggles with negative optimization due to relying solely on the segmentation model. ProCA prevents this issue by using prototype domain-invariant features, while PDA reduces the domain gap with the DF-DAN. These observations confirm that effective adaptation of domain-invariant features and semantic information is key to improving performance. Hence, we propose the OPLG strategy to enhance model segmentation by generating pseudo-labels that combine both feature types.

Visual analysis of 11 randomly selected images (spanning rural, urban, desert, snow, and water bodies) further supports these findings. As shown in Fig. 7, except for our method, other methods struggle to accurately capture thick cloud contours and fine thin cloud details. In contrast, our approach successfully identifies cloud boundaries, even in challenging conditions like snow-cloud coexistence. Notably, ProCA and PDA fail to match our method's discrimination ability in these scenarios. Additionally, esl and IntraDA exhibit significant false and missed detections, relying solely on entropy uncertainty that leads to suboptimal performance. This reinforces the importance of integrating domain-invariant and semantic features for improved segmentation in cross-domain cloud detection tasks.

2) Cross-Domain Cloud Detection Results in MS2S2: To evaluate the cross-domain performance of our proposed method in complex scenes, we conduct experiments using the WHUS2-CD+ dataset, which covers a large area with significant ground object changes. From the results in Table IV, we draw the following conclusions.

1) Our method outperforms others in the three comprehensive indicators mCPA, OA, and mIoU, achieving 69.86%, 79.36%, and 58.14%, respectively. The mIoU surpasses the SOTA method PDA by 9.68% and ProCA by 8.54%. This demonstrates that our method not only generates high-quality pseudo-labels, but also successfully transfers domain-invariant features to domain-semantic features, enabling effective feature adaptation during training.
2) The mIoU of methods using EM, such as advEnt, esl, and IntraDA, are 54.86%, 51.22%, and 52.75%, respectively. Although EM boosts overall performance, it lacks optimization for specific categories. IntraDA focuses more on foreground cloud features, while our method builds on this by designing the OPLG strategy to improve pseudo-label quality and achieve superior mIoU.

In the qualitative analysis of MS2S2 cross-domain cloud detection, we select seven RSIs with diverse backgrounds (desert, snow, and water) and varying cloud characteristics (thin clouds with continuous or discrete distributions).

1) When the source domain lacks similar target-domain data, relying solely on the ResNet50 backbone and DeepLabv2 segmentation network cannot effectively extract domain-invariant features. The lack of feature mining effectiveness is a critical issue. Our method addresses this by focusing on pixel-level details and using semantic-guided pseudo-label generation, enhancing the model's feature learning ability. As shown in the first three columns of Fig. 8, other methods exhibit significant misjudgments, particularly in distinguishing bright clouds.
2) By LIUF, our method effectively reduce the entropy at cloud–ground boundaries using the PCIF Loss. This results in higher quaility pseudo-labels and improves the robustness of model. As shown in the sixth column of Fig. 8, DAFormer and PDA can capture the general cloud outline but struggle with fine detail, leading to a significant number of false positives. These false positives undermine the reliability of the segmentation, which is particularly problematic in cross-domain cloud detection tasks where precision is critical.

These results highlight the importance of effectively integrating domain-invariant and domain-semantic features to

TABLE III

QUANTITATIVE ANALYSIS RESULTS OF VARIOUS METHODS IN MS2WHU CROSS-DOMAIN DATASET. THE RESULTS ARE DISPLAYED IN %; THE LAST IS LISTED AS THE AVERAGE MIOU. (BOLD VALUES INDICATE THE BEST RESULTS, AND UNDERLINE VALUES INDICATE THE SUBOPTIMAL RESULT)

Method	CPA		IoU		Recall		F1-Score		mCPA	OA	mIoU
	BackGround	Cloud	Background	Cloud	Background	Cloud	Background	Cloud			
Only Source	95.95	41.84	68.43	39.52	70.46	87.73	81.26	56.66	68.89	73.93	53.98
advEnt	**96.07**	42.64	68.79	40.35	70.78	88.25	81.51	57.50	69.36	74.23	54.57
esl	87.16	**60.27**	68.37	50.83	76.03	76.44	81.21	67.40	73.71	76.16	59.60
intraDA	95.10	46.31	69.35	43.24	71.92	86.72	81.90	60.38	70.70	75.15	56.30
DAFormer	94.25	44.89	68.24	41.44	71.21	84.38	81.12	58.60	69.57	74.07	54.84
ProCA	95.97	46.46	70.05	43.90	72.17	88.85	82.39	61.01	71.21	75.74	56.98
PDA	95.40	43.04	70.39	40.08	72.86	85.39	82.62	57.23	69.22	75.28	55.24
Ours	95.59	57.52	**73.90**	**54.07**	**76.51**	**90.02**	**84.99**	**90.02**	**76.56**	**80.03**	**63.99**

TABLE IV

QUANTITATIVE ANALYSIS RESULTS OF VARIOUS METHODS IN MS2S2 CROSS-DOMAIN DATASET. THE RESULTS ARE DISPLAYED IN %; THE LAST IS LISTED AS THE AVERAGE MIOU. (BOLD VALUES INDICATE THE BEST RESULTS, AND UNDERLINE VALUES INDICATE THE SUBOPTIMAL RESULT)

Method	CPA		IoU		Recall		F1-Score		mCPA	OA	mIoU
	BackGround	Cloud	Background	Cloud	Background	Cloud	Background	Cloud			
Only Source	97.54	22.61	71.24	21.50	72.54	81.45	83.20	35.39	60.07	73.34	46.37
advEnt	97.38	37.56	70.01	35.71	75.51	87.88	85.07	52.63	67.47	77.29	54.86
esl	**98.61**	29.41	73.88	28.57	74.66	**90.91**	84.98	44.43	64.01	76.35	51.22
intraDA	93.88	38.11	71.50	34.00	75.00	33.99	83.38	50.75	66.00	75.15	52.75
DAFormer	95.82	25.61	70.72	23.55	72.97	75.52	82.85	38.12	60.72	73.14	47.13
ProCA	97.19	28.26	72.52	26.68	74.08	82.65	84.07	42.12	62.72	75.02	49.60
PDA	96.48	27.19	71.60	25.33	73.53	78.65	83.45	40.41	61.84	74.10	48.46
Ours	96.49	**43.24**	**76.03**	**40.26**	**78.19**	85.37	**86.38**	**57.40**	**69.86**	**79.36**	**58.14**

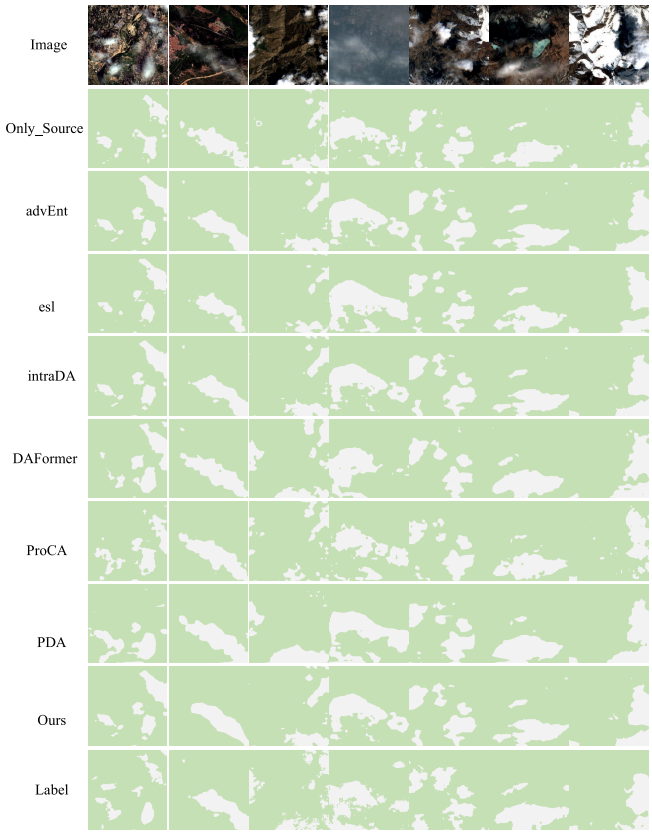

Fig. 8. MS2S2 cross-domain visualization result (green: background and white: cloud).

improve segmentation accuracy in complex cross-domain scenarios.

3) Cross-Domain Cloud Detection Results in WHU295: The analysis of Fig. 6 indicates that the 95 Cloud dataset contains a considerable proportion of images that are either entirely black or fully obscured by clouds, with the latter category being predominant. In contrast, the WHU Cloud dataset mainly contains images with partial cloud coverage, which presents challenges for cross-domain cloud detection. The large discrepancy in data distribution between the source and target domains makes it difficult to align feature distributions. Therefore, in cross-domain cloud detection tasks, the key challenge lies in leveraging domain-invariant features from the source domain to extract rare class feature information from the target domain.

As shown in Table V, our method outperforms others in mCPA, OA, and mIoU, with scores of 70.16%, 80.49%, and 58.82%, respectively. These results indicate that our method successfully utilizes the source-domain features to adapt the target-domain's semantic information, generating high-quality pseudo-labels that enhance the model's ability to extract pixel-level features during ST. In comparison, AT in IntraDA provides positive feedback to the segmentation model, with an mIoU of 56.58%, surpassing Only_Source by 14.93% and esl by 11.87%. This suggests that pseudo-labels generated purely by entropy are unreliable when domain distributions differ significantly. Inspired by this, we propose the OPLG strategy, which effectively uses domain-invariant feature information to ensure that the generated pseudo-labels positively impact the training process and enhance the model's ability to mine features from rare categories.

In the qualitative analysis of the WHU295 cross-domain cloud detection task, we randomly selected eight images with varying cloud types from the 95 Cloud dataset, including thin and thick clouds as well as images with large cloud coverage. In Fig. 9 (columns 2, 4, and 6), where the cloud distribution is sparse and there are obvious gaps between the clouds, our method is the only one that effectively fits the intermittent

TABLE V

QUANTITATIVE ANALYSIS RESULTS OF VARIOUS METHODS IN WHU295 CROSS-DOMAIN DATASET. THE RESULTS ARE DISPLAYED IN %; THE LAST IS LISTED AS THE AVERAGE MIOU. (BOLD VALUES INDICATE THE BEST RESULTS, AND UNDERLINE VALUES INDICATE THE SUBOPTIMAL RESULT)

Method	CPA		IoU		Recall		F1-Score		mCPA	OA	mIoU
	BackGround	Cloud	Background	Cloud	Background	Cloud	Background	Cloud			
Only Source	**99.57**	9.64	73.77	9.53	74.00	<u>89.66</u>	84.91	17.41	54.60	74.48	41.65
advEnt	95.77	24.36	74.08	21.96	76.59	69.01	85.11	36.01	60.06	75.84	48.02
esl	<u>99.10</u>	15.99	73.78	15.65	72.48	**89.91**	84.91	27.06	57.55	75.00	44.71
intraDA	93.33	43.03	<u>76.46</u>	<u>36.70</u>	80.89	71.38	<u>86.66</u>	<u>53.70</u>	<u>68.18</u>	<u>79.29</u>	<u>56.58</u>
DAFormer	97.38	23.29	75.09	21.82	76.64	77.46	85.77	35.82	60.34	76.70	48.45
ProCA	96.67	25.08	74.94	23.09	76.93	74.44	85.67	37.52	60.87	76.69	49.01
PDA	70.91	**57.59**	60.9	32.88	<u>81.20</u>	43.38	75.71	49.48	64.25	67.19	46.89
Ours	93.54	<u>46.77</u>	**77.56**	**40.08**	**81.95**	73.71	**87.36**	**57.23**	**70.16**	**80.49**	**58.82**

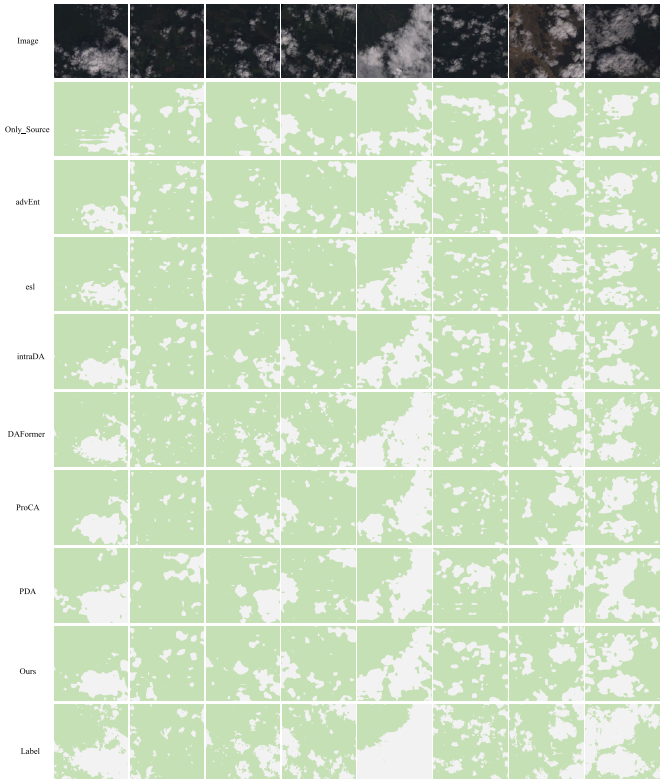

Fig. 9. WHU295 cross-domain visualization result (green: background and white: cloud).

cloud layers, while the other methods suffer from blockiness and missed detections. Furthermore, in the image with a high proportion of thick clouds (column 5), only DAFormer and our method reduce false positives while ensuring the accurate fitting of cloud layers, whereas other methods fail to do so. This clearly demonstrates that, even under highly imbalanced cross-domain category distributions, our method successfully integrates domain-invariant features with intradomain semantic information, promoting superior segmentation performance.

E. Ablation Study

1) Effectiveness of the OPLG Strategy: In this section, we evaluate different pseudo-label generation strategies. The parameter settings for the ablation experiment align with those in Section IV-B1, and the experiment is conducted on the MS2WHU cross-domain dataset.

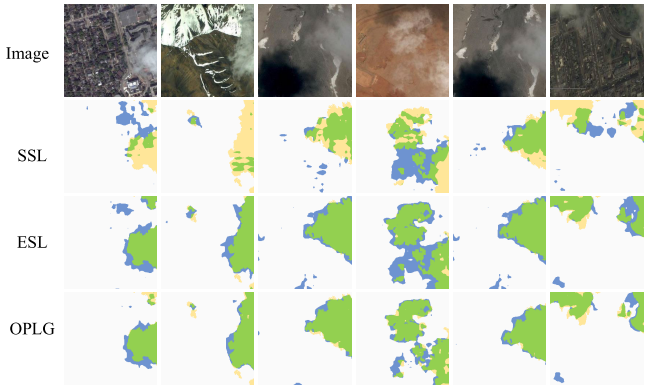

Fig. 10. Visualized result for various pseudo-label generation strategies (green: TP; yellow: FN; black: FP; and white: TN).

The quantitative results in Table VI show that pseudo-labels generated solely from the softmax class probability distribution (SSL) yield low quality, with OA and mIoU scores of 74.87% and 55.95%, respectively. Qualitative analysis in Fig. 10 further reveals that the SSL strategy performs better only in images with significant background-to-cloud foreground differences. The entropy distribution (ESL), which utilizes entropy distribution for pseudo-label generation, achieves better performance, with OA and mIoU improving by 3.26% and 6.86% over SSL. This improvement is due to the entropy distribution's ability to help the model distinguish between foreground and background features, enhancing pixel-level discrimination during ST. Inspired by entropy distribution, the proposed OPLG strategy incorporates it as baseline feature information to guide the model toward positive optimization. By leveraging the student network to assist in pseudo-label generation, OPLG achieves OA and mIoU scores of 80.03% and 63.99%, respectively. This demonstrates that effectively combining domain-invariant features with intradomain semantic information is crucial for improving ST in cross-domain cloud detection tasks. As shown in Fig. 10, OPLG not only accurately captures cloud contours, but also significantly reduces misclassifications.

2) Different Components for Impact of Network Performance: In this section, we incrementally add components to our method to analyze their impact on network performance. The parameter settings for the ablation experiment are consistent with those in Section IV-B1, and the experiments are conducted on the MS2WHU cross-domain dataset. As shown

TABLE VI

RESULT OF THE PSEUDO-LABEL ABLATION EXPERIMENT FOR OPLG. THE RESULTS ARE DISPLAYED IN %; THE LAST IS LISTED AS THE AVERAGE MIOU. (BOLD VALUES INDICATE THE BEST RESULTS)

Method	CPA		IoU		Recall		F1-Score		mCPA	OA	mIoU
	BackGround	Cloud	Background	Cloud	Background	Cloud	Background	Cloud			
SSL	94.76	46.10	69.04	42.85	71.78	85.87	81.69	60.00	70.43	74.87	55.95
ESL	85.35	67.68	69.77	55.84	**79.26**	76.15	82.19	71.67	76.52	78.13	62.81
OPLG	**95.59**	**57.52**	**73.90**	**54.07**	76.51	**90.02**	**84.99**	**90.02**	**76.56**	**80.03**	**63.99**

TABLE VII

DIFFERENT COMPONENTS FOR INFLUENCE OF NETWORK PERFORMANCE. (BOLD VALUES INDICATE THE BEST RESULTS)

Method	CPA		IoU		Recall		F1-Score		mCPA	OA	mIoU
	BackGround	Cloud	Background	Cloud	Background	Cloud	Background	Cloud			
L1	96.87	24.58	63.69	23.52	65.02	84.48	77.82	38.09	60.73	67.34	43.60
L2	94.21	50.56	70.22	46.65	73.39	85.79	82.50	63.62	72.39	76.37	58.44
L3	**95.59**	**57.52**	**73.90**	**54.07**	**76.51**	**90.02**	**84.99**	**90.02**	**76.56**	**80.03**	**63.99**

TABLE VIII

ANALYSIS OF SENSITIVITY OF CONFIDENCE TO PSEUDO-LABELS. (BOLD VALUES INDICATE THE BEST RESULTS)

p	CPA		IoU		Recall		F1-Score		mCPA	OA	mIoU
	BackGround	Cloud	Background	Cloud	Background	Cloud	Background	Cloud			
0.6	94.24	50.35	70.16	46.47	73.31	85.79	82.47	63.46	72.29	76.30	58.32
0.7	94.36	49.07	69.79	45.36	72.83	85.73	82.21	62.42	71.71	75.85	57.58
0.8	94.47	51.19	70.64	47.39	73.69	86.48	82.79	64.31	72.83	76.78	59.02
0.9	94.87	51.14	70.92	47.60	73.75	87.32	82.99	87.32	73.00	77.00	59.26
Ours(0.55)	**95.59**	**57.52**	**73.90**	**54.07**	**76.51**	**90.02**	**84.99**	**90.02**	**76.56**	**80.03**	**63.99**

in Table VII, when ST is conducted without entropy-based sample splitting, the mIoU for L1 is only 43.60%. This result is due to the use of unreliable labels during training, which causes some negative optimizations in the model. However, after incorporating sample splitting and the OPLG strategy, driven by consistent semantic alignment, the mIoU for L2 increases to 58.44%, and OA rises to 76.37%. These improvements, 14.84% and 9.03% higher than L1, respectively, stem from two factors: 1) sample splitting ensures reliable pseudo-label generation and 2) progressive semantic feature adaptation enhances model performance.

Furthermore, to strengthen the synergy between domain-invariant features and domain-specific semantic information, pixel-level supervision is achieved through a hybrid loss. As illustrated in Fig. 11, the red box highlights regions of entropy uncertainty. For example, in columns 3 and 5, entropy uncertainty is concentrated in the interior of thick clouds and areas where thin clouds are connected. From L1 to L3, this uncertainty progressively decreases, indicating that OPLG and hybrid loss effectively guide the model to aggregate positive samples in high-entropy regions. Additionally, negative samples in low-entropy areas, particularly in thin cloud regions (columns 1 and 2), are increasingly separated as the model progresses from L1 to L3.

3) Hyperparameter Sensitivity Experiments: We analyze the impact of the confidence threshold p for pseudo-labels in the ST stage, as introduced in Section III-B1. The confidence threshold plays a critical role in pseudo-label generation. If set too low, it can introduce excessive incorrect class information; if set too high, it may lead to a loss of valuable information. Therefore, selecting an optimal confidence threshold is essential for effective ST. To investigate this, we conduct sensitivity experiments with varying values of p on the ResNet50 backbone, and the results are presented in Table VIII.

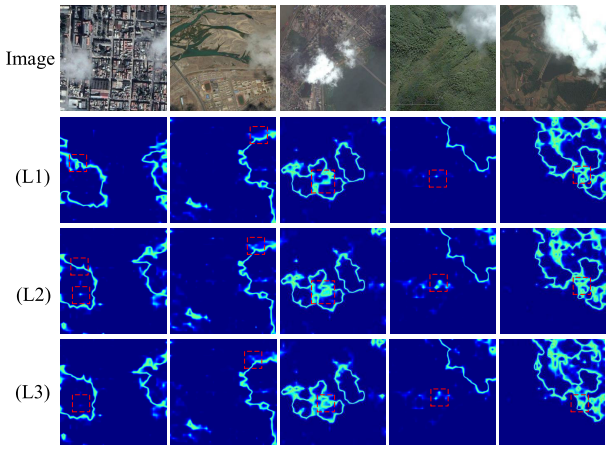

Fig. 11. Entropy visualizes cloud boundaries and interiors (as shown in the red box, the greater the entanglement, the higher the entropy).

V. DISCUSSION

A. Cross-Domain Cloud Detection Performance of Comparative Model

As shown in Fig. 12, t-SNE visualization is used to evaluate segmentation performance on the MS2WHU cross-domain cloud detection task. Three random images from the WHU Cloud dataset were selected for analysis. The Only_Source method failed to segment effectively, showing significant confusion due to poor model performance. AdvEnt improves segmentation by applying entropy AT, which helps cluster similar features more effectively. esl also benefits from EM, improving segmentation by better distinguishing foreground and background features. However, t-SNE for IntraDA shows divergence when highlighting clouds, indicating a conflict between domain-invariant and semantic information, causing the domain discriminator to struggle in distinguishing clouds from the highly altered ground features. DAFormer, with

TABLE IX
DIFFERENT FEATURE EXTRACTORS ON CROSS-DOMAIN CLOUD DETECTION PERFORMANCE. (BOLD VALUES INDICATE THE BEST RESULTS)

BackBone	Decoder	CPA		IoU		Recall		F1-Score		mCPA	OA	mIoU
		BackGround	Cloud	Background	Cloud	Background	Cloud	Background	Cloud			
ResNet50	DeepLabV2	**95.59**	57.52	73.90	54.07	76.51	**90.02**	84.99	**90.02**	76.56	80.03	63.99
MiT-B5		92.95	**65.09**	**74.88**	**59.07**	**79.39**	86.44	**85.63**	74.27	**79.02**	81.56	**66.97**

TABLE X
MODEL PARAMETERS AND COMPUTATIONAL COMPLEXITY ARE ANALYZED. (BOLD VALUES INDICATE THE BEST RESULTS)

Method	Backbone	FLOPs(G)	Params(M)	DF	FLOPs(G)	Params(M)	Resolutions	mIoU
Only_Source	ResNet50	26.664	23.729	×	×	×	256×256	53.98
advEnt				✓	1.649	2.764		54.57
esl				✓				59.60
intraDA				✓				56.30
ProCA				×	×	×		59.98
PDA		26.683	23.738	✓	1.649	2.764		55.24
Ours(ResNet50)		26.664	23.729	✓				**63.99**
DAFormer	MiT-B5	68.564	93.940	×	×	×		54.84
Ours(MiT-B5)				✓	1.649	2.764		**66.97**

its Transformer feature extractor, shows improved feature discrimination. ProCA and PDA narrow the domain gap gradually through domain-invariant features but still struggle with effective clustering due to the lack of semantic information from the source domain in the target domain. In contrast, our method fully utilizes both domain-invariant and semantic information, achieving better clustering, reducing class overlap, and obtaining clearer segmentation decision boundaries.

B. Cross-Domain Influence of Feature Extractors on Cloud Detection Performance

To assess the impact of the backbone feature extractor on our method, we replace ResNet50 with MiT-B5 while keeping DeepLabv2 in the decoder. The ablation experiment was conducted on the MS2WHU cross-domain cloud detection dataset without pretrained weights, training from scratch. As shown in Table IX, using the more powerful MiT-B5 feature extractor improves performance. With MiT-B5, our method achieves an mIoU of 66.97%, 2.98% higher than with ResNet50, and an OA of 81.56%. The Transformer-based architecture enhances feature extraction by enabling better interaction between global and local information. Additionally, multiscale feature fusion improves recognition ability. As shown in Fig. 13, MiT-B5 captures cloud contours more effectively than ResNet50, thanks to its ability to leverage global features to compensate for local feature confusion.

C. Cross-Domain Complexity of Cloud Detection Models

In this section, we analyze the number of parameters and computational complexity of the methods used in the experiments. The results are shown in Table X. For the segmentation model with ResNet50 as the backbone and DeepLabv2 as the encoder, the computational cost is 26.664G floating-point operations (FLOPs) with 23.729M parameters. The domain discriminator model (DF) requires 1.649G FLOPs and 2.764M parameters. Similarly, PDA has a computational cost of 26.683G FLOPs and 23.738M parameters. When using MiT-B5 as the backbone, the segmentation model's computational cost increases to 68.564G FLOPs with 93.940M parameters. Despite these differences, our method achieves optimal performance with an mIoU of 63.99% for ResNet50 and 66.97% for MiT-B5.

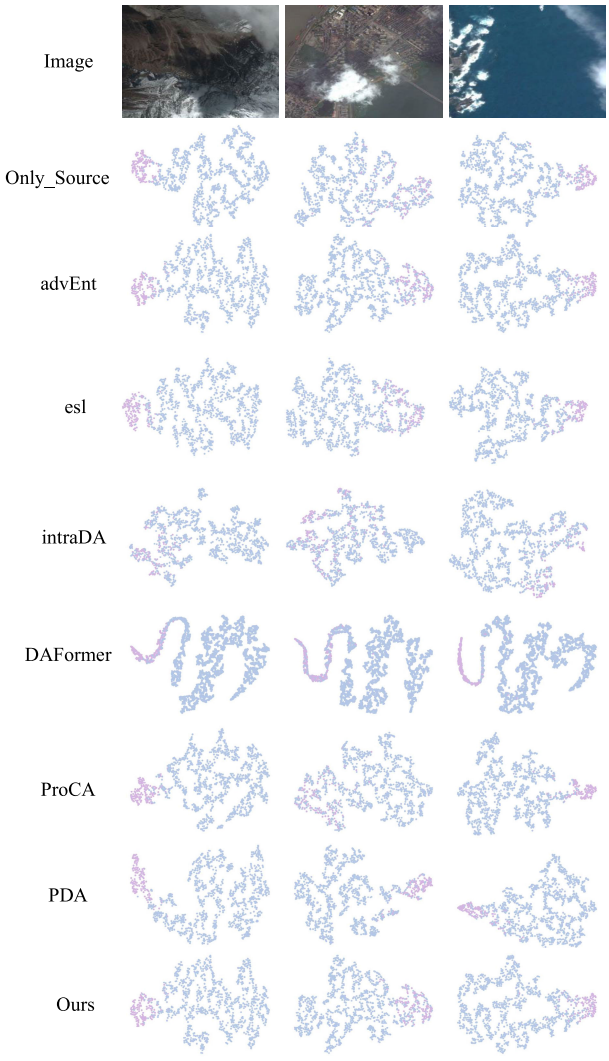

Fig. 12. t-SNE visualizes the segmentation decision boundary. (BLUE: BackGround; Purple: Cloud).

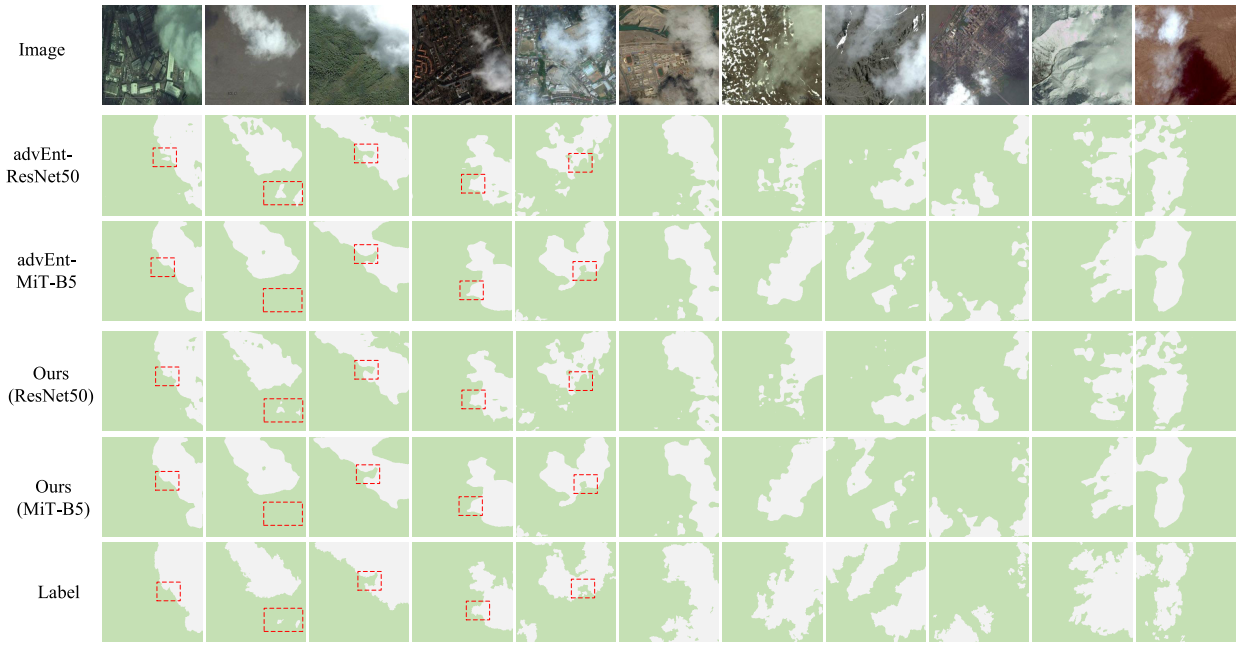

Fig. 13. Visualization of the influence of different feature extractors on cross-domain cloud detection performance.

VI. Conclusion

In this article, we propose a two-stage progressive cross-domain cloud detection framework. First, we generate high-quality, reliable pseudo-labels using the OPLG strategy. The effective collaboration between domain-invariant feature information and intradomain semantic information significantly enhances segmentation performance. Second, to address semantic ambiguity at the feature level, we design a PCIF Loss that resolves the distribution problem of positive and negative samples, ensuring more detailed feature representation from the feature to the pixel level. These steps gradually narrow the domain gap, improving the model's understanding of target-domain data distribution and further promoting target-domain feature mining, laying a solid foundation for pseudo-label generation in ST.

Our experiments show significant performance improvements across three cross-domain cloud detection tasks when compared to methods using the same feature extraction network. However, extensive experiments also highlight that extracting domain-invariant feature information solely through interdomain segmentation models remains limited by the performance of the feature extractor. Therefore, further research is needed to explore better handling of negative samples in pseudo-labels during training. In future work, we plan to use a feature queue to store negative samples, calculate the cosine distance between current pseudo-label negative samples and those in the queue, and dequeue features when the distance falls below a threshold. This approach aims to better mine effective feature information from negative samples.

References

[1] J. Zhang et al., "Monitoring plant diseases and pests through remote sensing technology: A review," *Comput. Electron. Agricult.*, vol. 165, Oct. 2019, Art. no. 104943.

[2] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, Jun. 2019.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[4] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

[5] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.

[6] H. Yang, X. Xiao, M. Yao, Y. Xiong, H. Cui, and Y. Fu, "PVSPE: A pyramid vision multitask transformer network for spacecraft pose estimation," *Adv. Space Res.*, vol. 74, no. 3, pp. 1327–1342, Aug. 2024.

[7] J. Chen et al., "TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers," *Med. Image Anal.*, vol. 97, Oct. 2024, Art. no. 103280.

[8] R. Ghosh and F. Bovolo, "TransSounder: A hybrid TransUNet-TransFuse architectural framework for semantic segmentation of radar sounder data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4510013.

[9] Y. Xiong, X. Xiao, M. Yao, H. Liu, H. Yang, and Y. Fu, "MarsFormer: Martian rock semantic segmentation with transformer," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4600612.

[10] X. Gao et al., "Two-stage domain adaptation based on image and feature levels for cloud detection in cross-spatiotemporal domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5610517.

[11] K. Gao, A. Yu, X. You, C. Qiu, and B. Liu, "Prototype and context-enhanced learning for unsupervised domain adaptation semantic segmentation of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5608316.

[12] J. Hoffman et al., "CyCADA: Cycle-consistent adversarial domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1989–1998.

[13] G. Mateo-García, V. Laparra, D. López-Puigdollers, and L. Gómez-Chova, "Cross-sensor adversarial domain adaptation of Landsat-8 and Proba-V images for cloud detection," 2020, *arXiv:2006.05923*.

[14] J. Guo, J. Yang, H. Yue, and K. Li, "Unsupervised domain adaptation for cloud detection based on grouped features alignment and entropy minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5603413.

[15] Y. Xiong, X. Xiao, M. Yao, H. Cui, and Y. Fu, "Light4Mars: A lightweight transformer model for semantic segmentation on unstructured environment like Mars," *ISPRS J. Photogramm. Remote Sens.*, vol. 214, pp. 167–178, Aug. 2024.

[16] F. Pan, I. Shin, F. Rameau, S. Lee, and I. S. Kweon, "Unsupervised intra-domain adaptation for semantic segmentation through self-supervision," in *Proc. CVPR*, Jun. 2020, pp. 3764–3773.

[17] Y. Cai, Y. Yang, Y. Shang, Z. Chen, Z. Shen, and J. Yin, "IterDANet: Iterative intra-domain adaptation for semantic segmentation of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5629517.

[18] L. Hoyer, D. Dai, H. Wang, and L. Van Gool, "MIC: Masked image consistency for context-enhanced domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 11721–11732.

[19] L. Hoyer, D. Dai, and L. Van Gool, "DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 9924–9935.

[20] H. Liu, M. Yao, X. Xiao, B. Zheng, and H. Cui, "MarsScapes and UDAFormer: A panorama dataset and a transformer-based unsupervised domain adaptation framework for Martian terrain segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4600117.

[21] Z. Jiang et al., "Prototypical contrast adaptation for domain adaptive semantic segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2022, pp. 36–54.

[22] J. Li et al., "A hybrid generative adversarial network for weakly-supervised cloud detection in multispectral images," *Remote Sens. Environ.*, vol. 280, Oct. 2022, Art. no. 113197.

[23] R. Gong, Q. Wang, M. Danelljan, D. Dai, and L. Van Gool, "Continuous pseudo-label rectified domain adaptive semantic segmentation with implicit neural representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7225–7235.

[24] M. Reguiegue and F. Chouireb, "Automatic day time cloud detection over land and sea from MSG SEVIRI images using three features and two artificial intelligence approaches," *Signal, Image Video Process.*, vol. 12, no. 1, pp. 189–196, Jan. 2018.

[25] C. Liu et al., "A machine learning-based cloud detection algorithm for the Himawari-8 spectral image," *Adv. Atmos. Sci.*, vol. 39, no. 12, pp. 1994–2007, Dec. 2022.

[26] Z. Zhu, S. Wang, and C. E. Woodcock, "Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images," *Remote Sens. Environ.*, vol. 159, pp. 269–277, Mar. 2015.

[27] P. Li, L. Dong, H. Xiao, and M. Xu, "A cloud image detection method based on SVM vector machine," *Neurocomputing*, vol. 169, pp. 34–42, Dec. 2015.

[28] J. Wei et al., "Cloud detection for Landsat imagery by combining the random forest and superpixels extracted via energy-driven sampling segmentation approaches," *Remote Sens. Environ.*, vol. 248, Oct. 2020, Art. no. 112005.

[29] H. Liu, M. Yao, X. Xiao, and Y. Xiong, "RockFormer: A U-shaped transformer network for Martian rock segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4600116.

[30] Z. Yang, Z. Yan, X. Sun, W. Diao, Y. Yang, and X. Li, "Category correlation and adaptive knowledge distillation for compact cloud detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5623318.

[31] J. Zhang, H. Wang, Y. Wang, Q. Zhou, and Y. Li, "Deep network based on up and down blocks using wavelet transform and successive multi-scale spatial attention for cloud detection," *Remote Sens. Environ.*, vol. 261, Aug. 2021, Art. no. 112483.

[32] J. Guo, J. Yang, H. Yue, H. Tan, C. Hou, and K. Li, "CDnetV2: CNN-based cloud detection for remote sensing imagery with cloud-snow coexistence," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 700–713, Jan. 2021.

[33] G. Zhang et al., "A multi-task driven and reconfigurable network for cloud detection in cloud-snow coexistence regions from very-high-resolution remote sensing images," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 114, Nov. 2022, Art. no. 103070.

[34] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 6028–6039.

[35] P. P. Busto and J. Gall, "Open set domain adaptation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 754–763.

[36] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, "ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2517–2526.

[37] Y. Tsai, W. Hung, S. Schulter, K. Sohn, M. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7472–7481.

[38] H. Wang, T. Shen, W. Zhang, L. Duan, and T. Mei, "Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Aug. 2020, pp. 642–659.

[39] Y. Luo, L. Zheng, T. Guan, J. Yu, and Y. Yang, "Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2507–2516.

[40] G. Mateo-García, V. Laparra, D. López-Puigdollers, and L. Gómez-Chova, "Transferring deep learning models for cloud detection between Landsat-8 and Proba-V," *ISPRS J. Photogramm. Remote Sens.*, vol. 160, pp. 1–17, Feb. 2020.

[41] S. Paul, Y. Tsai, S. Schulter, A. K. Roy-Chowdhury, and M. Chandraker, "Domain adaptive semantic segmentation using weak labels," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, Jan. 2020, pp. 571–587.

[42] Q. Lian, F. Lv, L. Duan, and B. Gong, "Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 6758–6767.

[43] Y. Liu et al., "Leveraging physical rules for weakly supervised cloud detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4103818.

[44] D. Brüggemann, C. Sakaridis, P. Truong, and L. Van Gool, "Refign: Align and refine for adaptation of semantic segmentation to adverse conditions," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 3173–3183.

[45] A. Saporta, T.-H. Vu, M. Cord, and P. Pérez, "ESL: Entropy-guided self-supervised learning for domain adaptation in semantic segmentation," 2020, *arXiv:2006.08658*.

[46] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Jun. 2017, pp. 5998–6008.

[47] K. Wang, D. Zhang, Y. Li, R. Zhang, and L. Lin, "Cost-effective active learning for deep image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 12, pp. 2591–2600, Dec. 2017.

[48] Y. Zou, Z. Yu, B. Vijaya Kumar, and J. Wang, "Unsupervised domain adaptation for semantic segmentation via class-balanced self-training," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 289–305.

[49] Y. Wang et al., "Semi-supervised semantic segmentation using unreliable pseudo-labels," 2022, *arXiv:2203.03884*.

[50] J. Zhu, Y. Guo, G. Sun, L. Yang, M. Deng, and J. Chen, "Unsupervised domain adaptation semantic segmentation of high-resolution remote sensing imagery with invariant domain-level prototype memory," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5603518.

[51] F. Shen, A. Gurram, Z. Liu, H. Wang, and A. Knoll, "DiGA: Distil to generalize and then adapt for domain adaptive semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 15866–15877.

[52] T. Karras, M. Aittala, J. Lehtinen, J. Hellsten, T. Aila, and S. Laine, "Analyzing and improving the training dynamics of diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 24174–24184.

[53] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[54] I. J. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[55] P. Tian et al., "3-D semantic terrain reconstruction of monocular close-up images of Martian terrains," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4600716.

[56] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[57] A. Pardyl, G. Rypeść, G. Kurzejamski, B. Zieliński, and T. Trzciński, "Active visual exploration based on attention-map entropy," 2023, *arXiv:2303.06457*.

[58] L. Van der Maaten and G. E. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, Jan. 2008.

[59] M. Liao, S. Tian, Y. Zhang, G. Hua, W. Zou, and X. Li, "PDA: Progressive domain adaptation for semantic segmentation," *Knowl.-Based Syst.*, vol. 284, Jan. 2024, Art. no. 111179.