Toward Cross-Regional Land Cover Mapping Using Regional Consistency and Certainty-Based Active Domain Adaptation With Limited Annotations

Jiaxu Guo, *Member, IEEE*, Juepeng Zheng, *Member, IEEE*, Yu Liang, Yibin Wen, Baozhu Yu, Juncheng Hu, *Member, IEEE*, Jianxi Huang, *Senior Member, IEEE*, and Haohuan Fu, *Fellow, IEEE*

***Abstract*—Semantic segmentation can automatically classify remote sensing (RS) images by pixel level into different types of land cover. However, in cross-domain semantic segmentation, how to more accurately identify the boundaries between types of surface features remains a critical challenge to be addressed. In this article, we introduce a region-aware active selection strategy incorporating region consistency and certainty for RS segmentation. The method, cross-regional active land cover mapping, enhances feature discriminability by maintaining local prediction consistency between pixels and their nearest neighbors in the source image. More accurate identification of boundaries between two types of ground objects positively contributes to improving segmentation accuracy. We conducted four sets of transfer annotation tasks on two representative RS datasets and validated our proposed method through these tasks, and our proposed method has achieved a significant improvement of 1.03%–38.23% compared to existing methods in the aforementioned transfer tasks and has narrowed the performance gap with fully supervised learning to a minimum of just 0.63%. Our method achieves outstanding performance with minimal annotation effort, offering practical and valuable information for advancing the field of RS.**

***Index Terms*—Active learning (AL), domain adaptation (DA), land cover mapping, semantic segmentation.**

Received 27 August 2025; revised 26 November 2025; accepted 21 January 2026. Date of publication 28 January 2026; date of current version 9 February 2026. This work was supported in part by the Guangdong Science and Technology Program under Grant 2024B0101040005, in part by the National Natural Science Foundation of China under Grant T2125006 and Grant 42401415, in part by Shenzhen Science and Technology Program under Grant KCXFZ20240903093759004 and Grant KJZD20230923115106012, in part by the Fundamental Research Funds for the Universities of Liaoning Province under Grant LJ212510143020, and in part by the Fundamental Research Funds for the Central Universities, Sun Yat-sen University, under Project 24xkjc002. *(Corresponding author: Juepeng Zheng.)*

Jiaxu Guo is with the School of Computer Science, Shenyang Aerospace University, Shenyang 110136, China (e-mail: jiaxu@sau.edu.cn).

Juepeng Zheng, Yu Liang, and Yibin Wen are with the School of Artificial Intelligence, Sun Yat-Sen University, Zhuhai 519082, China (e-mail: zhengjp8@mail.sysu.edu.cn; 1534484126@qq.com; wenyb5@mail2.sysu.edu.cn).

Baozhu Yu is with the College of Information Science and Engineering, Shenyang Ligong University, Shenyang 110159, China (e-mail: yubz22@sylu.edu.cn).

Juncheng Hu is with the College of Computer Science and Technology, Jilin University, Changchun 130012, China (e-mail: jchu@jlu.edu.cn).

Jianxi Huang is with the Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 611756, China, also with the College of Land Science and Technology, China Agricultural University, Beijing 100083, China, and also with the Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture and Rural Affairs, Beijing 100125, China (e-mail: jxhuang@cau.edu.cn).

Haohuan Fu is with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China, also with the Department of Earth System Science, Tsinghua University, Beijing 518055, China, and also with National Supercomputing Center in Shenzhen, Shenzhen 518055, China (e-mail: haohuan@tsinghua.edu.cn).

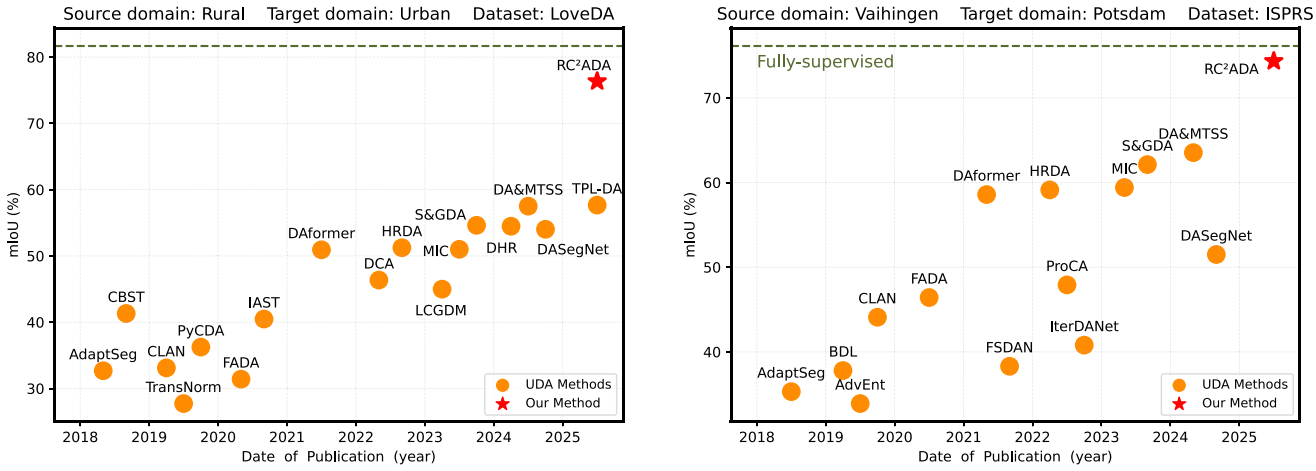
Digital Object Identifier 10.1109/JSTARS.2026.3658553

I. Introduction

LAND cover mapping involves identifying and classifying the physical material on the Earth's surface, such as vegetation, urban areas, water bodies, and bare soil. Understanding land cover is essential for various environmental and urban studies because it directly impacts ecological processes, climate modeling, natural resource management, and urban planning.

Moreover, land cover data aids biodiversity conservation by identifying key habitats and monitoring ecosystem dynamics, and is vital for carbon cycle studies given varying carbon storage capacities of land cover types. Thus, its analysis is fundamental to tackling global environmental challenges and supporting informed environmental management and policy decisions. Deep learning has become prominent in land cover classification/analysis, especially in remote sensing (RS), for its ability to automatically extract features from complex datasets. Unlike traditional methods relying on time-consuming, limited manual feature extraction, deep learning techniques can learn hierarchical data representations, enabling better land-cover classification. Furthermore, learning-based models can be trained to integrate multisource data, including optical, radar, and LiDAR data, which enhances the robustness of land cover classification [1]. This capability is particularly important in areas with diverse and complex land cover types or in regions where cloud cover frequently obscures optical imagery. In RS image analysis, cross-domain challenges occur due to discrepancies between source and target domains—differing in sensor type, resolution, spectral traits, or region. For instance, a model trained on satellite images from one region may perform poorly on those from another region or via a different sensor, as source-learned features may not generalize to the target, reducing performance. With growing diversity in RS data (captured under varying conditions and via different sensors), addressing such discrepancies is key for accurate, robust image analysis.

To address these challenges, domain adaptation (DA) has emerged as an effective approach. Through knowledge transfer between labeled source domains and unlabeled target domains,



(a) (b)

Fig. 1. SOTA DA methods and our method. (a) Performance on the DA task from rural to urban on the LoveDA semantic dataset. (b) Performance on the DA task from Vaihingen to Potsdam on the ISPRS semantic dataset.

DA mitigates domain shift to improve model efficacy in the target environment [2], [3]. This is particularly useful in RS, where acquiring labeled data for every possible scenario is often impractical or expensive [4], [5]. Active learning (AL) has significant applications in the interpretation of RS images, where acquiring labeled data are often challenging and resource intensive. In RS, AL is used to optimize the process of labeling large volumes of imagery by selectively choosing the most informative samples for human experts to label, thereby improving model performance with fewer labeled examples. Semantic segmentation plays an emerging role RS by enabling the pixelwise classification of imagery into different land cover or land use categories. In RS, semantic segmentation is used to interpret high-resolution satellite or aerial images, where each pixel is assigned a specific class label [6]. It supports the development of more advanced and accurate models for understanding complex environmental and urban systems, contributing to sustainable development and effective management of natural resources.

We have also reviewed the existing work on DA in RS imagery. As shown in Fig. 1, the performance of current methods is compared on two typical datasets, including LoveDA [7] and ISPRS [8]. It is evident that there is still a significant gap between existing methods and fully supervised learning, indicating that there is substantial room for improvement in this area. To further address the aforementioned issues, this article proposes region consistency and certainty-based method for semantic segmentation in RS for cross-region land-cover mapping. In conclusion, the principal innovative contributions of this research are described as follows.

1) We incorporate AL to strategically label image regions that are likely to mark feature boundaries. This targeted approach focuses annotation efforts where they are most needed.
2) To select the most valuable points for annotation, we designed a region identification method alongside a comprehensive evaluation metric, the regional selection index (RSI). The RSI, based on regional consistency and certainty, guides the selection process to enhance boundary delineation accuracy while optimizing computational efficiency.
3) We refine the loss function by integrating a contrastive learning-based term. Building on existing work, this addition helps mitigate the impact of inevitable noise in pseudolabels, enabling the model to achieve more precise predictive performance.

The rest of this article is organized as follows. Section II examines related scholarly work. Section III introduces the core methodology proposed herein. Section IV presents the benchmark datasets used in this research. Section V describes the experiments conducted. Section VI provides a further analysis of the aforementioned results of the experiments. Finally, Section VII concludes this article.

II. Related Works

A. Semantic Segmentation in RS

The goal of semantic segmentation, a computer vision technique, is to classify every pixel in an image according to a predefined set of categories. Within the realm of RS image data processing, it has found extensive applications [9]. At its core, the approach assigns pixel-level labels to RS images for land cover categorization, distinguishing classes such as forest, cropland, water, infrastructure, and barren ground. This plays a crucial role in land resource management, urban expansion monitoring, and environmental protection [10]. By segmenting RS images, it is possible to identify the extent and types of disasters, such as floods, wildfires, or surface damage caused by earthquakes [11]. Application scenarios include disaster area delineation, postdisaster reconstruction planning, and emergency response support [12]. Furthermore, semantic segmentation can be used to precisely extract the locations and shapes of urban infrastructure, such as buildings [13], roads, and bridges, which is valuable for urban spatial analysis, transportation network optimization, and building distribution statistics [14]. Semantic

segmentation can also identify ecological elements in RS images, such as forests, water bodies, and wetlands, to evaluate changes and health conditions in ecosystems [15]. In addition, it can segment crop cultivation areas to monitor crop growth, estimate yields, or detect pests and diseases [16]. Semantic segmentation provides detailed pixel-level classification, making it suitable for high-resolution RS imagery. Traditional methods often require manual annotation, but semantic segmentation models significantly enhance processing efficiency [17]. By leveraging deep learning, semantic segmentation is well-suited for handling large-scale data and can adapt to various RS platforms, including satellites and drones [18]. Hypergraph for HSI classification is reliably constructed via DFS/BFS balance and kernel mapping, addressing variability/similarity to outperform SOTA [19]. Besides, uncertainty-guided model was proposed to mitigate RGB-T sensors' sample-dependent information quality variation on semantic segmentation fusion [20].

Although semantic segmentation technology has made significant progress, there are still areas that require further improvement. Existing methods often produce models whose segmentation results are blurry or inaccurate when handling small objects or complex boundaries. At the same time, the generalization and robustness of these models remain insufficient. These challenges are key issues that researchers need to address.

B. DA in RS

The concept of DA comes from transfer learning, and was first introduced and studied in computer vision [21]. In recent years, it has been increasingly applied to classification tasks and has been continuously refined and integrated with innovative approaches. Weakly supervised learning, in particular, is employed to address scenarios involving low-quality data in the source domain [22]. There is also improved DA framework integrating global and local transfer perspectives, which enables data from both the source and target domains to be projected into separate subspaces [23]. Researchers have been working to improve the network structure [24], or pay attention to multiobjective DA to make the models have more excellent performance. Huang et al. [25] proposed discriminative radial DA so that the two domains can be better related. Shannon entropy constraint was incorporated into the UDA method to mitigate spectral shift in RS 3-D cross-scene classification [26]. Generative adversarial networks (GANs) have also inspired research in the field of DA. Accordingly, as a key method in transfer learning, DA has also attracted the attention of researchers in RS [27], [28]. Transfer learning methods can help researchers better use satellite RS images for a variety of analyses [29]. For example, unsupervised method was combined for multisource, multitarget, and life-long DA scenarios [30], [31]. In addition, DA also plays a relatively significant role in the classification of hyperspectral images [32], [33], [34].

An easy-to-hard adaptive architecture was also proposed as a solution for multitarget DA (MTDA) scenarios. MTDA can also be used to perform landslide cross-domain mapping from large-scale RS images, so as to apply it to risk assessment and disaster prevention [35]. It is of great practical significance to classify RS images correctly and reasonably, which can help people to develop and utilize land resources better [36].

C. AL and Active DA

AL's core innovation lies in its sample selection mechanism: identifying maximally informative points for annotation achieves dual objectives drastically decreasing labeling resources while accelerating model performance gains [37]. By improving data efficiency from the sample perspective, AL provides a practical solution for scenarios where obtaining data labels is challenging or costly. Settles [38] introduced classic AL methods, which have since evolved rapidly due to their ability to efficiently select high-value samples from unlabeled data.

Over the past decade, various sample selection strategies have been proposed, including uncertainty-based, diversity-based, query-by-committee-based, and expected model change-based approaches. In addition, hybrid query strategies [39] have been explored to balance uncertainty and diversity, enabling wider application in realistic scenarios. These strategies have been successfully applied to computer vision tasks such as image classification, object detection, and image segmentation [40], as well as practical applications such as autonomous driving [41], intelligent medical diagnostics [42], and RS [43].

Current state-of-the-art DA methods, while aiming to reduce annotation costs and enhance cross-region knowledge transfer, have not fully overcome the inherent limitations of DA. As a result, these methods have not been widely or efficiently applied in RS, with their performance still lagging behind fully supervised models. Introducing AL to DA problems could mitigate the distortion of the target-domain distribution, as AL entails minimal annotation cost and can identify the most representative samples for annotation. Our proposed method integrates ADA ideas and takes a significant step toward optimizing DA in RS. In the future, DA in RS may require the integration of advanced machine learning methods, such as AL, to balance low-cost annotation and high accuracy, effectively narrowing the gap with fully supervised models.

III. Methodology

A. Preliminaries

For an ADA problem applicable to RS image scenarios, we usually have a dataset on the source domain, ($D_S = \{(\mathbf{X}_i^S, Y_i^S)\}_{i=1}^{n_S}$), where X^S is the original image dataset and Y is the label dataset. For each Y_i, its individual pixel points are individually corresponding to X_i, and each pixel point records the corresponding class of the point. At the same time, there is also a dataset on the target domain, called $D^T = X^T$, where X^T is the image dataset in the target domain, which has not been labeled at the beginning. For images in the target domain, there will be a corresponding pseudolabel $\hat{Y}^T$. The training process involves determining the corresponding model from X to Y, ensuring that it performs well not only on the source domain D^S but also achieves the highest possible accuracy on the target domain D^T.

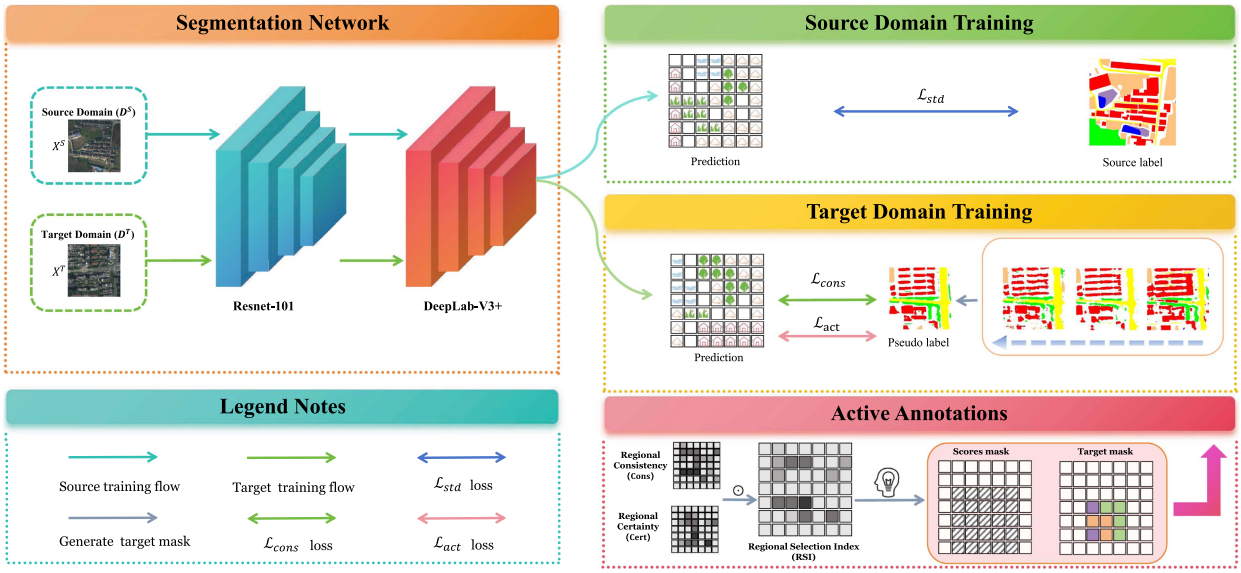


Fig. 2. Workflow framework of the proposed method RC^{2}ADA. The feature extraction part is based on the ResNet-101 backbone network, and the classifier part is based on DeepLabv3+ to perform semantic segmentation.

Therefore, for the DA scenario in semantic segmentation of cross-region land-cover mapping in RS, we propose the method termed regional-aware active DA with consistency and certainty (RC^{2}ADA) to achieve better performance. We adopt an improved region partitioning method that divides regions based on the k-pixel neighborhood around each pixel in RS images, subdividing each possible class into corresponding subregions. For these partitioned regions, we propose discriminative criteria of region consistency and region certainty: region consistency assesses the statistical information of pixels from different categories within a region, characterizing its purity by the number of categories, while region certainty represents the average certainty of pixels. Based on these criteria, AL indexes are calculated for each pixel to guide sample selection. In addition, we introduce cross-entropy loss to measure the difference between predicted results and labels, discriminative consistency loss to prevent overfitting, and negative learning loss constructed through negative pseudolabels. These losses collectively form the final training objective, enabling the trained model to achieve superior performance.

B. Active Region Definition and Selection

The overall architecture of this framework is illustrated in Fig. 2. In the field of RS imagery, traditional AL methods often focus on simple segmentation of images, such as dividing them into several equal-sized areas. However, this approach often ignores the inhomogeneity of various elements within the image. For example, in an RS image, the size of the ground features contained is often not equal. At the same scale, the area of contiguous farmland is usually much larger than that of a single building, and the area of water bodies is often different. Therefore, it is necessary to divide the areas flexibly.

To achieve superior performance in semantic segmentation tasks, it is crucial to establish well-defined neighborhood relationships for each pixel during feature extraction. The semantic contextual information derived from adjacent and surrounding pixels plays a pivotal role in determining accurate semantic class boundaries and preserving fine-grained semantic structures. Our active DA framework samples critical regions through sliding-window queries. Each query unit constitutes a fixed-size rectangular window centered at individual pixels; it can be denoted as

$$\mathcal{R}_k(i,j) = (x,y) || x-i| \le k_w, |y-j| \le k_h \tag{1}$$

where i and j represent a point in the image, and k represents the distance extending outward from this point. That is to say, $\mathcal{R}_k$ represents a square centered on the pixel (i,j) with side lengths of k_w in the x-direction and k_h in the y-direction. For each pixel in the image, there is a corresponding region. If

$$k_w = k_h = k \tag{2}$$

then the region is a square. This will be further investigated in the experimental section of this article. In semantic segmentation scenarios, an image often contains multiple categories. As for the aforementioned region $\mathcal{R}_k$, we can further divide it into several subregions that may belong to different categories. For example, the subregion belonging to class c can be described as

$$\mathcal{R}_k^c(i,j) = \{(x,y) \in N_k(i,j) | \hat{Y}_t^{(x,y)} = c\}. \tag{3}$$

After delineating the aforementioned region, we can employ statistical methods to infer the potential number of categories within this region. Next, we will introduce two important criteria used in this article: region consistency and region certainty.

C. Regional Consistency and Certainty Criteria

For the region defined in (1), when k is a positive number, the number of pixels it contains is undoubtedly greater than 1, even if the pixel is located at the edge or corner of the image. If a region contains a diverse set of pixels, assuming labeling aids in training the network, we introduce the concept of regional

consistency. It can be expressed as

$$\text{Cons}_{(i,j)} = \sum_{c=1}^{C} \frac{|\mathcal{R}_k^c(i,j)|}{|\mathcal{R}_k(i,j)|} \log \frac{|\mathcal{R}_k^c(i,j)|}{|\mathcal{R}_k(i,j)|} \tag{4}$$

larger this value, the higher the pixel homogeneity within the neighborhood centered around (i, j). It indicates that the regional consistency, and carries semantic relational information. To better evaluate the classification accuracy exhibited by the current pixel (and its surrounding region).

From the aforementioned process, we can see that the predicted result $\mathbf{P}$, carrying a considerable amount of semantic information, often exhibits a certain degree of uncertainty. Therefore, to ensure the accuracy of predictions as much as possible, it is necessary to measure its certainty. Entropy plays a crucial role in determining certainty, particularly within information theory and decision making. In the context of certainty, entropy serves as a metric for the quantity of information and degree of disorder in a system. It offers a means to quantify the uncertainty inherent in various outcomes.

$$\mathcal{H}(i,j) = -\sum_{c=1}^{C} \mathbf{P}_{t(i,j)} \log \mathbf{P}_{t(i,j)}. \tag{5}$$

Next, building upon this foundation, in conjunction with (1) and (5), we will introduce the concept of regional certainty to characterize it, whose specific definition is as follows:

$$\text{Cert}_{(i,j)} = \frac{\sum_{(x,y)\in\mathcal{R}_k(i,j)} \mathcal{H}(x,y)}{|\mathcal{R}_k(i,j)|}. \tag{6}$$

A higher value of this metric typically indicates that the trained model is more confident in the predictions—implying it deems its forecasts more reliable and accurate—while also reflecting the model's strong generalization ability.

Based on the regional consistency and regional certainty, we have obtained previously, we can assign corresponding RSI to each pixel point (i, j), as

$$\text{RSI}_{(i,j)} = \text{Cons}_{(i,j)} \odot \text{Cert}_{(i,j)}. \tag{7}$$

A higher RSI indicates better consistency and certainty at a given location. This metric assists in identifying regions with greater class diversity and more pronounced object boundaries within the image. Simultaneously, model uncertainty surfaces are identified to guide region selection, prioritizing high-entropy and heterogeneous zones for annotation. The resultant iterative calibration continuously reduces prediction divergence from reference labels.

D. Training Objectives

After selecting the regions that need annotation, the unique information and features in the target domain can be trained. During the training process, we mainly use three loss functions. First, we optimize the network by utilizing all labeled data from both the source and target domains through optimizing the standard supervised loss, which is defined as follows:

$$\mathcal{L}_{\text{std}} = \mathcal{L}_{\text{ce}}(\mathbf{X}^S, Y^S) + \mathcal{L}_{\text{ce}}(\mathbf{X}^T, \hat{Y}^T) \tag{8}$$

where L_{ce} represents the cross-entropy loss

$$\mathcal{L}_{\text{ce}} = -\frac{\sum_{(i,j)\in\mathbf{X}} \sum_{c=1}^{C} \mathcal{Y}_{(i,j,c)} \log \mathbf{P}_{(i,j,c)}}{|\mathbf{X}|} \tag{9}$$

where Y represents the label of pixel (i, j). To enforce spatial coherence, this method aligns pixel-level predictions with local context in the source domain. This dual mechanism enhances prediction consistency while mitigating source-domain overfitting. The consistency regularization term is formulated mathematically as

$$\mathcal{L}_{\text{cons}}^s = \frac{\sum_{(i,j)\in\mathbf{X}^S} \left\|\mathbf{P}_{(i,j)} - \overline{\mathbf{P}}_{(i,j)}\right\|}{|\mathbf{X}^S|} \tag{10}$$

where $\mathbf{P}_{(i,j)}$ refers to the average prediction for the neighborhood space centered at (i, j), which is the R_k mentioned earlier in the text. The calculation for this average value is as follows:

$$\overline{\mathbf{P}}_{(i,j)} = \frac{\sum_{(x,y)\in\mathcal{R}_1(i,j)} \mathbf{R}_{(x,y)}}{|\mathcal{R}_1(i,j)|} \tag{11}$$

where $\mathcal{R}_1(i, j)$ refers to the neighborhood region obtained when expanding 1 pixel in all directions with (i, j) as the center. In the inference phase, pseudolabeling generates labels for unlabeled data using predictions from a model trained on labeled data. However, in multiclass scenarios, a single pixel may receive similar probability scores for several related classes. To address this, we can initially exclude the least probable class assignments. Negative pseudolabeling specifically denotes the confident assignment of negative class labels to instances predicted with high certainty as negative. Building upon this framework, we propose an adaptive weighted loss function for AL, which yields the active loss through the following formulation:

$$\mathcal{L}_{\text{act}}^t = -\sum_{c\in\mathcal{N}(i,j)} w_c \cdot \log\left(1 - \mathbf{P}_t^{(i,j,c)} + \epsilon\right) \tag{12}$$

where $\mathcal{N}(i, j)$ denotes the set of negative labels in the pixel region (i, j), ϵ is a numerical stability term with the value of 1e-8 (IEEE floating-point compatible), and the weight w_c is dynamically calculated based on the threshold τ as follows:

$$w_c = \begin{cases} \frac{\mathbf{P}_t^{(i,j,c)} - \tau}{1-\tau}, & \text{if } \mathbf{P}_t^{(i,j,c)} > \tau \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

In other words, when the probability of a pixel being classified as class c falls below τ, it is considered unlikely to belong to that particular class. Finally, the total loss function can be formulated as follows:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{std}} + \alpha\mathcal{L}_{\text{cons}}(D_S) + \beta\mathcal{L}_{\text{act}}(D_T) \tag{14}$$

where α and β are the weighting coefficients corresponding to the respective losses. The values of these parameters will be tested and optimized in the subsequent ablation studies.

IV. Datasets

Experimental validation utilizes two RS image datasets to assess the method's labeling accuracy and generalization capacity. Each dataset comprises two domains, enabling the construction

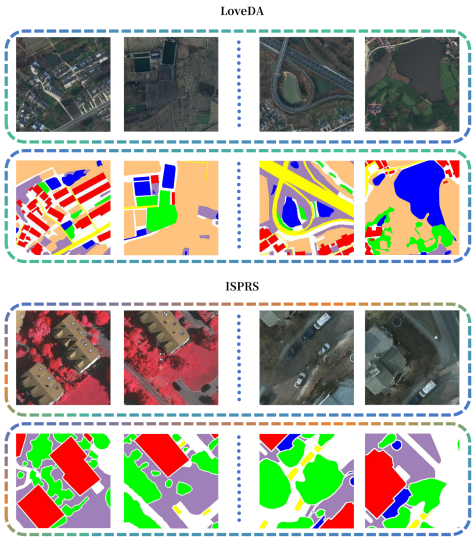


Fig. 3. Two domains contained in each of the two datasets, along with an example sample from each domain and its label. For LoveDA, the left side of the dashed line is rural, and the right side is urban; for ISPRS, the left side of the dashed line is Vaihingen, and the right side is Potsdam.

of four transfer tasks accordingly. The LoveDA and ISPRS datasets are two of the most representative RS image datasets, which are widely used in RS fields. They are especially of great significance in cross-region land-cover mapping tasks since they contain a wide range of RS images from different sensors and cover multiscale, multitype, and variably distributed landscapes. The aforementioned datasets and the typical sample examples from their different domains are illustrated in Fig. 3.

V. Experiments

A. Implementation Details

Without loss of generality, we adopt DeepLab-V3+ [48] framework, which is composed of the most versatile ResNet-101 [49] as the backbone. ResNet-101 incorporates the innovative concept of residual learning, within which the residual blocks enable the model to learn hierarchical features effectively, capturing intricate patterns in the input data. Such design helps alleviate the challenges associated with gradient vanishing or exploding during training, facilitating the training of deep networks .

As for the hyperparameters, in our experiments, we set the $\tau = 0.08$, $\alpha = 1.0$, and $\beta = 1.0$. Furthermore, considering the high resolution of the *ISPRS* dataset, we have segmented its image size for training purposes, and the size of each subimage after segmentation is 512 $\times$ 512pixels. Similarly, due to the input-friendly size of the imagery data in the *LoveDA* dataset, we employ the original images given in the dataset as inputs with 1024 $\times$ 1024pixels.

B. Evaluation Metric

The accuracy of the methods was evaluated using intersection over union (IoU) and mean intersection-over-union (mIoU). For each category, the labeling accuracy was assessed using IoU, defined as follows:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}. \tag{15}$$

For multiclass classification tasks, we employ IoU [see (15)] to evaluate class-level classification results. We employ mIoU as primary metric to quantify segmentation performance across all semantic categories. The mIoU is defined as follows:

$$\text{mIoU} = \frac{1}{C}\sum_{i=1}^{C}\frac{\text{TP}_C}{\text{TP}_C + \text{FP}_C + \text{FN}_C} \tag{16}$$

where C is the total number of classes.

C. Experimental Results

To comprehensively evaluate the effectiveness and robustness of our proposed method, we perform extensive experiments on two well-recognized benchmark datasets: *LoveDA* and *ISPRS*. Furthermore, to ensure a comprehensive analysis, we incorporate two widely adopted query strategies commonly used in mainstream AL frameworks, as discussed in Section II: random query strategies and certainty-based sampling strategies. These strategies serve as baselines to provide a holistic comparison and highlight the advantages of our proposed method. Tables I–IV show the performance of our method on four transfer tasks conducted on these datasets, respectively. The visualization results are shown in Figs. 4 and 5.

Experiments on the *LoveDA* dataset confirm RC^{2}ADA's significant improvements over competing methods. As shown in Tables I and II, RC^{2}ADA surpasses SOTA methods by 23.73% mIoU on R $\rightarrow$ U tasks and 23.15% on U $\rightarrow$ R tasks. It also outperforms classic active selection methods by 10.21% mIoU on R $\rightarrow$ U tasks and 8.99% on U $\rightarrow$ R tasks.

To objectively and comprehensively evaluate the performance of the model, we compared RC^{2}ADA with the baseline obtained from DeepLab-V3+ and the fully supervised method in terms of mIoU. On the R $\rightarrow$ U tasks shown in Table I, RC^{2}ADA outperforms a conspicuous enhancement of 23.86%, and narrows the gap between the fully supervised method to 5.33% compared to SOTA 15.54%. On the U $\rightarrow$ R tasks shown in Table II, RC^{2}ADA outperforms an improvement of 21.22%, and narrows the gap between the fully supervised method to inspiring 0.63% compared to SOTA 9.62%. Furthermore, throughout almost all categories, RC^{2}ADA also demonstrates a considerable improvement to the baseline while significantly shrinking the gap between the fully supervised method.

As exhibited in Tables III and IV, RC^{2}ADA outperforms the SOTA AADA methods by 1.03% mIoU and 4.1% mIoU on the P $\rightarrow$ V tasks and the V $\rightarrow$ P tasks. On the two transfer tasks, compared to the classic active selection methods applied in DA tasks, RC^{2}ADA also exhibits a superior performance by 2.78% mIoU and 1.55%, respectively.

VI. Discussion

A. Ablation Study on Criteria

To isolate individual component contributions, rigorous ablation analyses were conducted across multiple operational

TABLE I
RURAL → URBAN TRANSFER TASKS IN LOVEDA DATASET

Method	bkgd.	build.	road	water	barren	forest	agri.	mIoU
Baseline	38.14	54.63	52.87	71.51	46.88	45.61	57.92	52.51
Random	53.63	63.68	67.41	77.7	66.41	59.64	74.41	66.13
Entropy	64.88	49.99	43.5	71.63	33.33	19.35	66.9	49.94
Confidence	64.53	53.78	44.65	68.83	28.85	34.34	66.96	51.71
DA&MTSS [44]	40.66	59.32	60.96	79.38	48.31	54.41	59.50	57.51
DASegNet [45]	42.31	58.25	53.12	68.39	42.43	50.44	63.21	54.02
MADAv2 [43]	25.23	45.83	45.19	58.62	35.77	38.04	46.84	42.22
D^2ADA [46]	40.77	61.05	59.77	68.18	28.67	50.64	59.21	52.61
TPL-DA [47]	48.58	60.49	57.15	73.84	46.02	51.87	65.77	57.67
RC^2ADA(Ours)	**65.25**	**70.43**	**78.75**	**85.29**	**79.63**	**71.78**	**83.25**	**76.34**
Fully supervised	72.32	77.07	82.99	88.54	86.77	75.24	88.75	81.67

The bold values are the best-performing methods.

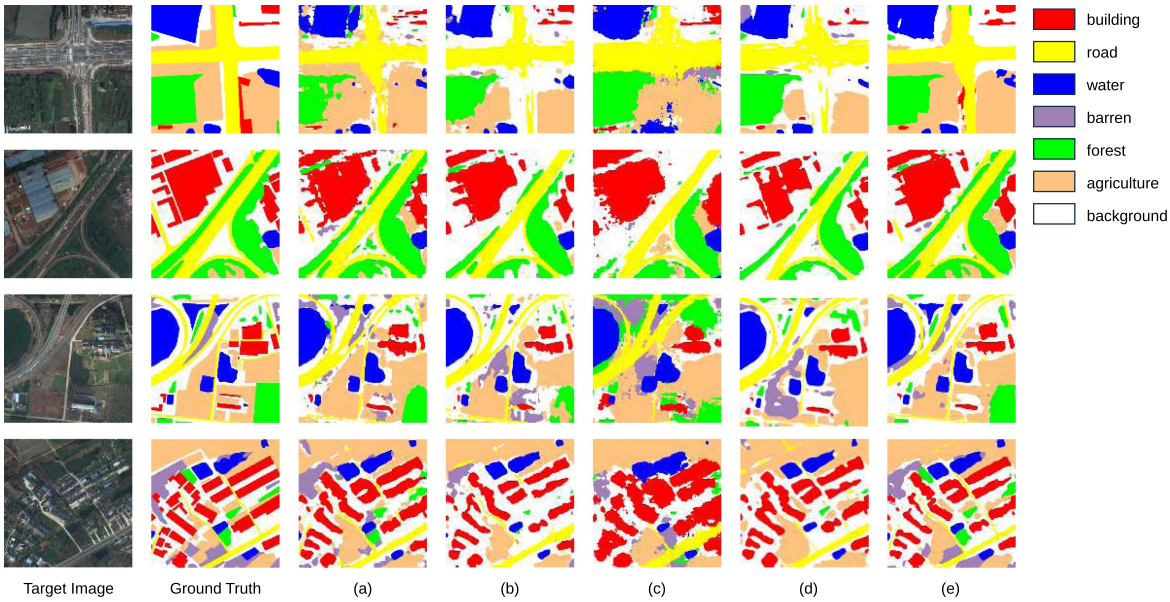


Fig. 4. Visualization of results on LoveDA semantic dataset. (a) Random. (b) Entropy. (c) MADA. (d) D^2ADA. (e) RC^2ADA method proposed in this article.

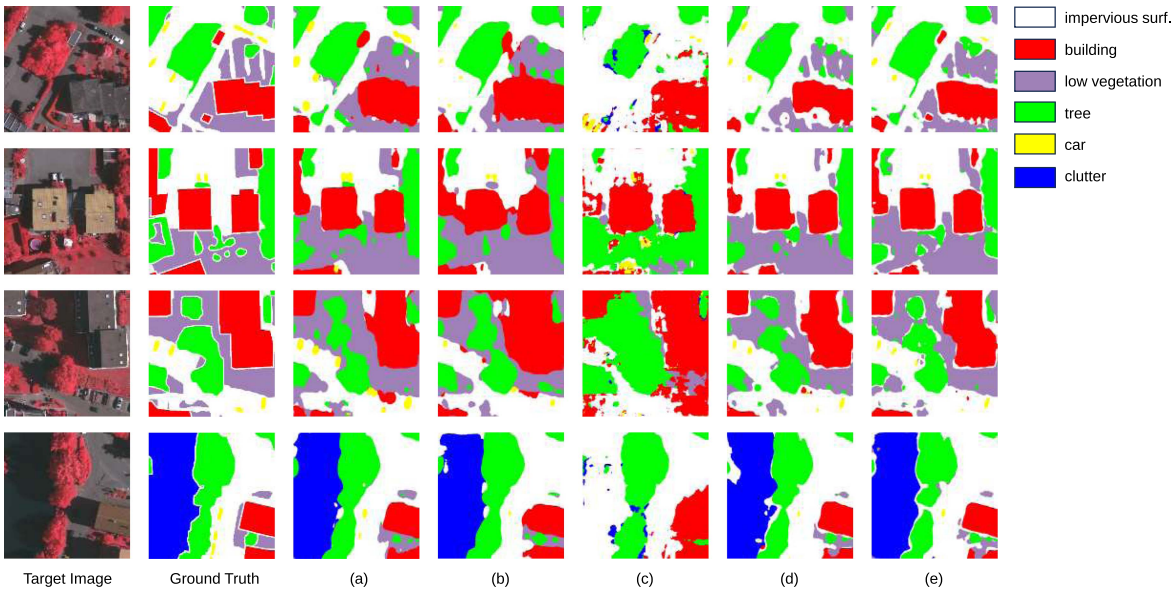


Fig. 5. Visualization of results on ISPRS semantic dataset. (a) Random. (b) Entropy. (c) MADA. (d) D^2ADA. (e) RC^2ADA method proposed in this article.

TABLE II
URBAN → RURAL TRANSFER TASKS IN LOVEDA DATASET

Method	bkgd.	build.	road	water	barren	forest	agri.	mIoU
Baseline	39.91	53.99	51.57	71.76	48.2	49.32	53.51	52.61
Random	72.17	65.81	55.04	78.57	54.7	49.49	78.11	64.84
Entropy	45.57	54.28	59.15	72.8	48.97	49.84	63.14	56.25
Confidence	44.28	57.5	56.32	72.8	47.38	47.32	61.84	55.35
DA&MTSS [44]	58.81	61.68	47.37	73.93	16.02	30.86	66.86	50.79
DASegNet [45]	58.43	43.21	34.25	63.28	42.43	4.33	50.22	42.31
MADAv2 [43]	56.71	31.66	29.52	50.65	24.51	8.97	48.79	35.83
D^2ADA [46]	59.33	**70.03**	61.8	61.32	18.07	27.32	56.67	50.65
TPL-DA [47]	61.76	56.68	33.38	64.40	35.10	5.13	56.42	44.69
RC^2ADA(Ours)	**79.48**	69.67	**67.19**	**85.47**	**72.12**	**57.66**	**85.24**	**73.83**
Fully supervised	77.73	73.02	68.15	82.61	73.88	60.29	85.51	74.46

The bold values are the best-performing methods.

TABLE III
POTSDAM → VAIHINGEN TRANSFER TASKS IN ISPRS SEMANTIC LABELING DATASET

Method	surf.	build.	vege	tree	car	bkgd.	mIoU
Baseline	35.55	53.88	1.04	1.45	22.82	1.17	19.32
Random	79.25	84.02	63.98	74.97	52.76	15.47	61.74
Entropy	72.95	73.1	55.51	72.74	27.54	9.66	51.92
Confidence	76.95	81.2	58.34	72.48	32.61	14.72	56.05
DA&MTSS [44]	66.70	83.25	49.35	52.27	50.36	35.17	56.18
DASegNet [45]	69.31	76.25	50.13	53.24	43.71	24.11	52.79
MADAv2 [43]	46.97	51.49	9.48	50.0	9.44	25.4	32.13
D^2ADA [46]	78.92	81.31	**65.82**	75.42	**61.88**	17.56	63.49
TPL-DA [47]	68.90	78.99	53.84	70.61	0.59	**47.63**	53.43
RC^2ADA(Ours)	**81.92**	**85.41**	65.42	**76.25**	60.65	17.48	**64.52**
Fully supervised	82.63	87.5	69.02	78.95	67.11	29.4	69.1

The bold values are the best-performing methods.

TABLE IV
VAIHINGEN → POTSDAM TRANSFER TASKS IN ISPRS SEMANTIC LABELING DATASET

Method	surf.	build.	vege	tree	car	bkgd.	mIoU
Baseline	42.23	43.5	3.62	1.32	29.39	0.78	20.14
Random	82.12	88.18	72.54	75.47	86.35	32.0	72.78
Entropy	78.56	84.08	69.53	71.17	85.24	27.92	69.42
Confidence	77.63	85.13	67.04	69.28	76.23	26.71	67.0
DA&MTSS [44]	74.01	74.39	63.18	72.66	69.28	27.64	63.53
DASegNet [45]	67.12	75.32	47.43	55.13	43.90	20.11	51.50
MADAv2 [43]	55.39	54.16	19.92	45.48	39.2	2.43	36.1
D^2ADA [46]	80.75	83.89	71.55	72.98	73.68	**38.55**	70.23
RC^2ADA(Ours)	**82.77**	**88.34**	**73.95**	**76.26**	**89.5**	35.18	**74.33**
Fully supervised	84.59	90.79	74.79	77.84	90.7	38.02	76.12

The bold values are the best-performing methods.

configurations. Table V shows the accuracy of ablation studies. Overall, we show that our full method achieves optimal results, underscoring the importance and complementarity inherent in the proposed region-aware factors and losses. We divided the application of the criteria into six cases, labeled (A)–(F). The specific meanings are detailed in the footnotes of Table V.

To further investigate the efficacy of the negative learning loss $\mathcal{L}^t_{\text{act}}$ and the local prediction consistency $\mathcal{L}^s_{\text{cons}}$ in active DA training on the RS data, we ablate them, respectively. As shown in the bottom half of Table V, methods (D) and (E), which incorporate the two types of loss with suitable hyperparameters, achieve a clear yet modest improvement over traditional approaches (A), (B), and (C) that lack these losses. However, the full method (F) (i.e., our RC^2ADA), which takes a weighted consideration of $\mathcal{L}^t_{\text{act}}$ and $\mathcal{L}^s_{\text{cons}}$ except for the joint utilize of the region-aware factors, showcases a better performance (D) and (E) among all transfer tasks than as expected. All in all, by contrast, the result of (F) achieves a remarkable improvement compared to method $\mathcal{L}^s_{\text{cons}}$—Only method (D) and $\mathcal{L}^t_{\text{act}}$—Only method (E). Such a result indicates that an integrating weighted utilization of $\mathcal{L}^t_{\text{act}}$ and $\mathcal{L}^s_{\text{cons}}$ with suitable hyperparameters, which is employed by our method RC^2ADA, has certain benefit to avoid overfitting to the source data and further to increase the accuracy of cross-region tasks in RS.

B. Sensitive Analysis of the Hyperparameters

To identify the optimal values for the hyperparameters, we performed three ablation studies targeting the aforementioned

TABLE V
ABLATION STUDY OF THE METHOD FACTORS

Selection			Training		R→U	U→R	**LoveDA**	P→V	V→P	**ISPRS**
Method	Consistency	Certainty	$\mathcal{L}^s_{\text{cons}}$	$\mathcal{L}^t_{\text{act}}$	mIoU	mIoU	mIoU	mIoU	mIoU	mIoU
Random: randomly selecting regions					66.13	64.84	65.49	61.74	72.78	67.26
(A)	✓				72.71	68.61	70.66	64.31	74.06	69.19
(B)		✓			69.77	67.18	68.48	62.52	73.12	67.82
(C)	✓	✓			72.67	69.44	71.06	62.61	73.59	68.1
(D)	✓	✓	✓		73.24	70.49	71.87	62.78	73.77	68.28
(E)	✓	✓		✓	72.53	70.11	71.32	64.51	73.58	69.05
(F)	✓	✓	✓	✓	**76.34**	**73.83**	**75.45**	**64.52**	**74.33**	**69.43**
Fully Supervised: all labeled source and target					81.67	74.46	78.07	69.1	76.12	72.61

Method Interpretation: (A) use region consistency only; (B) use prediction certainty only; (C) combine consistency and certainty; (D) train with $\mathcal{L}^s_{\text{cons}}$ on source samples; (E) train with $\mathcal{L}^t_{\text{act}}$ on target samples; and (F) fully RC^{2}ADA for region-based active annotating;
Average accuracy: Column *LoveDA* and *ISPRS* indicate the average accuracy calculated from the respective two transfer tasks included in the LoveDA dataset and the ISPRS semantic labeling dataset.

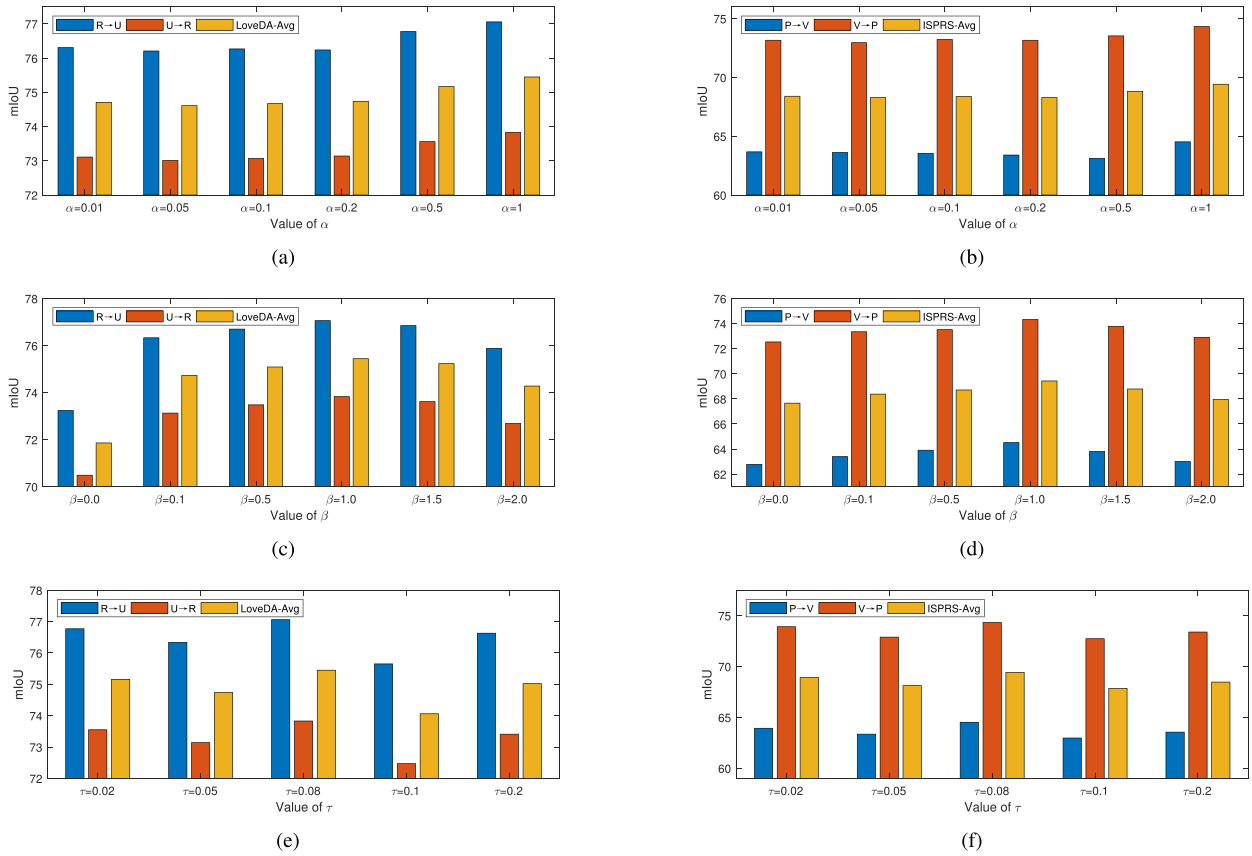


Fig. 6. Sensitivity experiments for (a) and (b) α, (c) and (d) β, and (e) and (f) τ in LoveDA and ISPRS datasets.

hyperparameters, i.e., α, β, and τ. In this part of the experiment, the value of α ranges from 0.01 to 1, β ranges from 0.0 to 2.0, and τ ranges from 0.02 to 0.2. Experiments were conducted on four transfer tasks accordingly. Fig. 6 presents the results of the aforementioned hyperparameter experiments.

We noticed that $\alpha = 1.0$ outperforms other values with a noticeable improvement. We averaged the results of transfer tasks across the two datasets, and it can be seen that the mIoU response to hyperparameter variations exhibits convergence across different transfer tasks. Therefore, in a nutshell, we set the hyperparameters $\alpha = 1.0$, $\beta = 1.0$, and $\tau = 0.08$ in our experiments, respectively.

C. Impact of Region Size and Region Shape

In the experiment, it is necessary to define the size of the region, which is described by the side length k of the square region. Here, we investigate how different values of k affect the

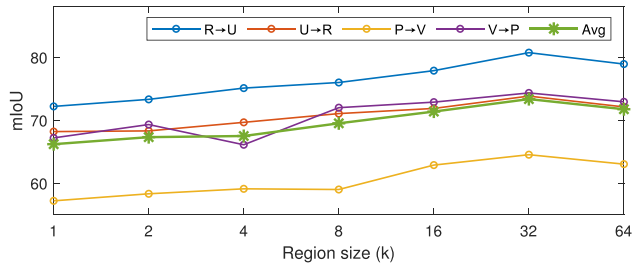


Fig. 7. Impact of changes in the value of k on mIoU.

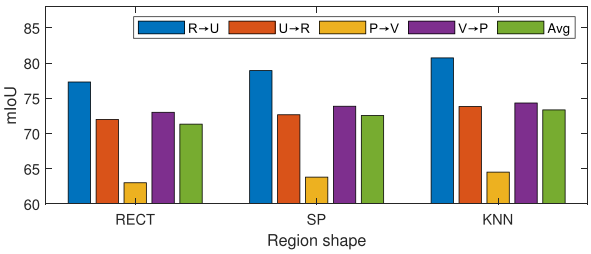


Fig. 8. Impact of region shape on mIoU.

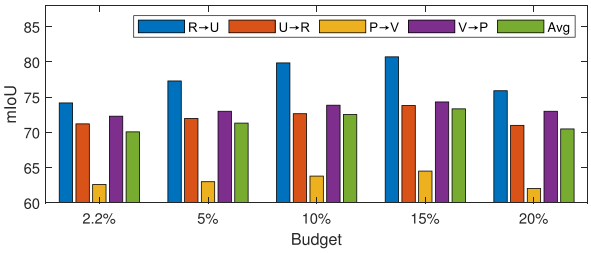


Fig. 9. Impact of active budget on mIoU.

segmentation results. Taking the rural-to-urban experiment as an example, the values of k are taken as powers of 2, ranging from 1 to 64. It can be observed in Fig. 7 that our method performs best when k is set to 32. However, when k is increased to 64, the performance decreases, which may be due to the inclusion of pixels from other categories after expansion.

Besides the region size, the influence of region shape on experimental results is also worth exploring. Therefore, for the four aforementioned transfer tasks, we conducted experiments with three region shapes: fixed-size rectangles (RECT), superpixels (SP), and the k-nearest neighbor (k-NN) neighborhood adopted in the proposed method. Fig. 8 presents the results of this part of the experiments, from which it can be observed that the k-NN neighborhood selection method in this article achieves better results both for each of the four transfer tasks and their average values.

D. Impact of Active Budget

In AL, budget usually denotes the proportion of the dataset selected for labeling/querying in each iteration, controlling the new information added to the training set, and thus, affecting the framework's efficiency and effectiveness. Taking the rural-to-urban transfer experiment as an example, this study also explored the budget's impact on results. It can be seen in Fig. 9 that the model achieves optimal performance when the budget is set to 15%.

VII. Conclusion

In this article, we propose a region-aware active DA approach incorporating an adversarial learning strategy. Our method creatively introduces AL to DA in RS imagery. Furthermore, our RC^{2}ADA comprehensively considers the consistency and certainty of the region to formulate an active adversarial label acquisition strategy, making the features more discriminative in DA tasks in RS. The efficacy of certain factors we consider in the proposed method is verified by ablation studies. We believe that our method provides a realistic and efficient way to address the DA problems in RS for relevant scenarios such as cross-region land-cover mapping. In response to the increasingly expanding demand for active DA in geoscience and RS, we will continue to explore the effectiveness of our method and propose further adaptive optimization for more practical scenarios and applications in the future.

References

[1] M. M. Al Rahhal, Y. Bazi, H. Al-Hwiti, H. Alhichri, and N. Alajlan, "Adversarial learning for knowledge adaptation from multiple remote sensing sources," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 8, pp. 1451–1455, Aug. 2021.

[2] K. Saito, D. Kim, S. Sclaroff, and K. Saenko, "Universal domain adaptation through self supervision," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 16282–16292.

[3] J. Zheng, S. Yuan, W. Li, H. Fu, L. Yu, and J. Huang, "A review of individual tree crown detection and delineation from optical remote sensing images: Current progress and future," *IEEE Geosci. Remote Sens. Mag.*, vol. 13, no. 1, pp. 209–236, Mar. 2025.

[4] J. Peng, Y. Huang, W. Sun, N. Chen, Y. Ning, and Q. Du, "Domain adaptation in remote sensing image classification: A survey," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9842–9859, 2022.

[5] J. Zheng et al., "A two-stage adaptation network (TSAN) for remote sensing scene classification in single-source-mixed-multiple-target domain adaptation (S ^{2} M^{2}T DA) scenarios," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5609213.

[6] 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.

[7] J. Wang, Z. Zheng, A. Ma, X. Lu, and Y. Zhong, "LoveDA: A remote sensing land-cover dataset for domain adaptive semantic segmentation," in *Proc. Neural Inf. Process. Syst. Track Datasets Benchmarks*, 2021.

[8] X. Yang et al., "An attention-fused network for semantic segmentation of very-high-resolution remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 177, pp. 238–262, 2021.

[9] L. Tian, X. Zhong, and M. Chen, "Semantic segmentation of remote sensing image based on GAN and FCN network model," *Sci. Programm.*, vol. 2021, 2021, Art. no. 9491376.

[10] Z. Lin, F. Zhu, Y. Kong, Q. Wang, and J. Wang, "SRSG AND S2SG: A model and a dataset for scene graph generation of remote sensing images from segmentation results," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4707411.

[11] M. Luo, T. Zhang, S. Wei, and S. Ji, "SAM-RSIS: Progressively adapting SAM with box prompting to remote sensing image instance segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4413814.

[12] Y. Zhu, L. Long, J. Wang, J. Yan, and X. Wang, "Road segmentation from high-fidelity remote sensing images using a context information capture network," *Cogn. Comput.*, vol. 14, no. 2, pp. 780–793, Mar. 2022.

[13] M. Luo, "Cross-spatiotemporal land-cover classification from VHR remote sensing images with deep learning based domain adaptation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 191, pp. 105–128, 2022.

[14] Y. Li, Y. Zhang, and Z. Zhu, "Error-tolerant deep learning for remote sensing image scene classification," *IEEE Trans. Cybern.*, vol. 51, no. 4, pp. 1756–1768, Apr. 2021.

[15] Y. Hamrouni, E. Paillassa, V. Chéret, C. Monteil, and D. Sheeren, "From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2," *ISPRS J. Photogrammetry Remote Sens.*, vol. 171, pp. 76–100, 2021.

[16] J.-X. Wang, S.-B. Chen, C. H. Q. Ding, J. Tang, and B. Luo, "RanPaste: Paste consistency and pseudo label for semisupervised remote sensing image semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 2002916.

[17] Y. Ma, Y. Wang, X. Liu, and H. Wang, "SWINT-RESNet: An improved remote sensing image segmentation model based on transformer," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 8003005.

[18] Z. Liu et al., "RSPS-SAM: A remote sensing image panoptic segmentation method based on SAM," *Remote Sens.*, vol. 16, no. 21, Nov. 2024, Art. no. 4002.

[19] Q. Wang, J. Huang, S. Wang, Z. Zhang, T. Shen, and Y. Gu, "Community structure guided network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 4404115.

[20] Q. Wang, C. Yin, H. Song, T. Shen, and Y. Gu, "UTFNet: Uncertainty-guided trustworthy fusion network for RGB-Thermal semantic segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 7001205.

[21] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Computer Vision—ECCV 2010*, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Berlin, Germany: Springer, 2010, pp. 213–226.

[22] R. Xie, H. Wei, L. Feng, and B. An, "GearNet: Stepwise dual learning for weakly supervised domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, no. 8, pp. 8717–8725.

[23] H. Yang, H. He, T. Li, Y. Bai, and W. Zhang, "Multi-metric domain adaptation for unsupervised transfer learning," *IET Image Process.*, vol. 14, pp. 2780–2790, 2020.

[24] C. Zhang, Q. Zhao, and Y. Wang, "Transferable attention networks for adversarial domain adaptation," *Inf. Sci.*, vol. 539, pp. 422–433, Oct. 2020.

[25] Z. Huang, J. Wen, S. Chen, L. Zhu, and N. Zheng, "Discriminative radial domain adaptation," *IEEE Trans. Image Process.*, vol. 32, pp. 1419–1431, 2023.

[26] Q. Wang, M. Wang, J. Huang, T. Liu, T. Shen, and Y. Gu, "Unsupervised domain adaptation for cross-scene multispectral point cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5705115.

[27] 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 Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7920–7930, Nov. 2020.

[28] J. Zheng, Y. Zhao, W. Wu, M. Chen, W. Li, and H. Fu, "Partial domain adaptation for scene classification from remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5601317.

[29] H. Chen, H. Zhang, G. Yang, S. Li, and L. Zhang, "A mutual information domain adaptation network for remotely sensed semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5537316.

[30] O. Tasar, A. Giros, Y. Tarabalka, P. Alliez, and S. Clerc, "DAugNet: Unsupervised, multisource, multitarget, and life-long domain adaptation for semantic segmentation of satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1067–1081, Feb. 2021.

[31] Y. Liang, S. Cao, J. Zheng, X. Zhang, J. Huang, and H. Fu, "Low saturation confidence distribution-based test-time adaptation for cross-domain remote sensing image classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 139, 2025, Art. no. 104463.

[32] 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 Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5518415.

[33] Q. Li, Y. Zhang, J. Zheng, Y. Zhang, J. Huang, and H. Fu, "Boosting universal domain adaptation in remote sensing with dual-classifiers consistency discrimination and cross-domain feature mixup," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5515015.

[34] W. Chen, Y. Wen, J. Zheng, J. Huang, and H. Fu, "BAN: A universal paradigm for cross-scene classification under noisy annotations from RGB and hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5610213.

[35] J. Zheng et al., "Open-set domain adaptation for scene classification using multi-adversarial learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 208, pp. 245–260, 2024.

[36] J. Guo et al., "C ^{3} DA: A universal domain adaptation method for scene classification from remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 6006705.

[37] B. Xie, L. Yuan, S. Li, C. H. Liu, and X. Cheng, "Towards fewer annotations: Active learning via region impurity and prediction uncertainty for domain adaptive semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, New Orleans, LA, USA, Jun. 2022, pp. 8058–8068.

[38] B. Settles, "Active learning literature survey," University of Wisconsin-Madison, Computer Sciences Technical Report 1648, 2009.

[39] C. Shui, F. Zhou, C. Gagné, and B. Wang, "Deep active learning: Unified and principled method for query and training," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2020, pp. 1308–1318.

[40] M. Ning et al., "Multi-anchor active domain adaptation for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 9092–9102.

[41] F. Peng, C. Wang, J. Liu, and Z. Yang, "Active learning for lane detection: A knowledge distillation approach," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 15132–15141.

[42] V. Nath, D. Yang, B. A. Landman, D. Xu, and H. R. Roth, "Diminishing uncertainty within the training pool: Active learning for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2534–2547, Oct. 2021.

[43] M. Ning et al., "MADAv2: Advanced multi-anchor based active domain adaptation segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13553–13566, Nov. 2023.

[44] D. Chen, Y. Wang, and L. Zhang, "DA&MTSS: An end-to-end remote sensing image domain adaptive semantic segmentation framework combining data augmentation and mobile threshold self-supervision," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4706216.

[45] Q. Zhao, S. Lyu, H. Zhao, B. Liu, L. Chen, and G. Cheng, "Self-training guided disentangled adaptation for cross-domain remote sensing image semantic segmentation," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 127, 2024, Art. no. 103646.

[46] T.-H. Wu et al., "D2ADA: Dynamic density-aware active domain adaptation for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, Jul. 2022, pp. 449–467.

[47] Y. Ren, J. Long, X. Gao, M. Zhang, G. Liu, and N. Su, "TPL-DA: A novel threshold-free pseudolabel learning framework for domain adaptive semantic segmentation of high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 1926–1945, 2025.

[48] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.