

## Article

# Multi-Domain Incremental Learning for Semantic Segmentation via Visual Domain Prompt in Remote Sensing Data

Junxi Li <sup>1,2</sup>, Zhiyuan Yan <sup>1,2,\*</sup>, Wenhui Diao <sup>1,2,3</sup>, Yidan Zhang <sup>1,2</sup>, Zicong Zhu <sup>4</sup>, Yichen Tian <sup>1,2</sup> and Xian Sun <sup>1,2,3</sup>

<sup>1</sup> Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China; lijunxi@aircas.ac.cn (J.L.)

<sup>2</sup> Key Laboratory of Target Cognition and Application Technology (TCAT), Beijing 100190, China

<sup>3</sup> School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100190, China

<sup>4</sup> School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China

\* Correspondence: yanzzy@aircas.ac.cn

## Highlights

### What are the main findings?

- RS-VDP utilized a visual domain prompt based on dynamic location to mitigate the distribution discrepancy between source and target images.
- A novel evaluation metric is proposed to address the issue that traditional evaluation metrics are biased by single-task accuracy.

### What are the implications of the main findings?

- RS-VDP leveraged visual domain prompts to modulate data distribution, which provides a novel paradigm for the domain incremental learning task.
- Comprehensive experiments have demonstrated the effectiveness of the proposed method and significantly reduced the degree of catastrophic forgetting.

## Abstract

Domain incremental learning for semantic segmentation has gained lots of attention due to its importance for many fields including urban planning and autonomous driving. The catastrophic forgetting problem caused by domain shift has been alleviated by structure expansion of the model or data rehearsal. However, these methods ignore similar contextual knowledge between the new and the old data domain and assume that new knowledge and old knowledge are completely mutually exclusive, which cause the model to be trained in a suboptimal direction. Motivated by the prompt learning, we proposed a new domain incremental learning framework named RS-VDP. The key innovation of RS-VDP is to utilize a visual domain prompt to change the optimization direction from input data space and feature space. First, we designed a domain prompt based on a dynamic location module, which applied a visual domain prompt according to a local entropy map to update the distribution of the input images. Second, in order to filter the feature vectors with high confidence, a representation feature alignment based on an entropy map module is proposed. This module ensures the accuracy and stability of the feature vectors involved in the regularization loss, alleviating the problem of semantic drift. Finally, we introduced a new evaluation metric to measure the overall performance of the incremental learning models, solving the problem that the traditional evaluation metric is affected by the single-task accuracy. Comprehensive experiments demonstrated the effectiveness of the proposed method by significantly reducing the degree of catastrophic forgetting.



Academic Editor: Shuying Li

Received: 29 December 2025

Revised: 23 January 2026

Accepted: 29 January 2026

Published: 1 February 2026

**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

**Keywords:** incremental learning; visual domain prompt; representative features; semantic segmentation; remote sensing

## 1. Introduction

The rapid advancement of remote sensing technology has facilitated the acquisition of high-resolution remote sensing images [1–5], which have been extensively leveraged for advancing semantic segmentation methodologies [6,7]. These technological developments underpin critical geospatial intelligence-driven applications spanning urban planning [8] and autonomous driving [9]. An important assumption so that the deep learning models perform well on the test images is that the training set and test set belong to the same data domain. This means that the spatial resolutions, spectral bands, and pixel distributions of the images in the training dataset are consistent with those in the test dataset. As depicted in Figure 1, when the model is fine-tuned on the new training images of novel domains, the performance on old domain images significantly deteriorates. The reason for the above phenomenon is called the catastrophic forgetting problem [10]. The domain incremental learning methods [11] aim at solving this problem caused by data from different domains, which eliminates the need for repeated training on the same dataset and enables the model to progressively enhance its performance across diverse data domains while preventing the misuse of private data. Therefore, domain incremental learning methods have significant research value and deserve to be widely studied.

The key problem to be solved in domain incremental learning for semantic segmentation in aerial images is the domain shift problem. Recently, the solution has been divided into rehearsal-based methods, regularization-based methods, and parameter isolation-based methods. Rehearsal-based methods [12–14] replay images or features to retain old knowledge and learn new knowledge, mainly considering the ratio of the knowledge of the old task and new task. Parameter isolation-based methods [15,16] update fewer parameters that are important to the old tasks and add new architecture for new tasks. For example, MDIL [15] was proposed to exploit domain-special parameters and domain-share parameters to study the knowledge of different domains, achieving good prediction performance. Regularization-based methods apply knowledge distillation loss or add regularization loss to increase the penalty of changing the important parameters. GSMF [17] transforms the traditional feature space to a graph feature space, and seeks to utilize the multifeature constraints loss to alleviate the forgetting.

Although the above methods have made contributions to solving the problem of domain shift, a gap remains compared to the single-task performance of the model on these datasets. On the one hand, almost all approaches assume that the model trained on the old task cannot make good predictions on the whole images in the different domains of the new tasks [18,19]. However, it is important to note that the model is trained on the old dataset, which means the knowledge of contextual information is learned. For example, the model learns that the vehicles should be on the impervious surface and not on the roof of the building. This means that the knowledge of the new data and the old data are not completely mutually exclusive. Therefore, we argue that the model should locate the locations where the domain gap between the old task and the new task is large, and adopt strategies to change the distribution of the above locations. On the other hand, the class prototypes calculated by feature vectors have the problem of semantic drift [20,21]. The class prototypes are calculated by taking the arithmetic average of the feature vectors corresponding to the labels belonging to the category. When the discriminability of the feature is not obvious, the prediction made by the model of this position is uncertain. The model utilizes

these features to participate in prototype calculation and perform parameter regularization loss calculation, which leads the optimization to a suboptimal direction. Therefore, we argue that the representative feature vector should also be determined by discriminative high-dimensional features.

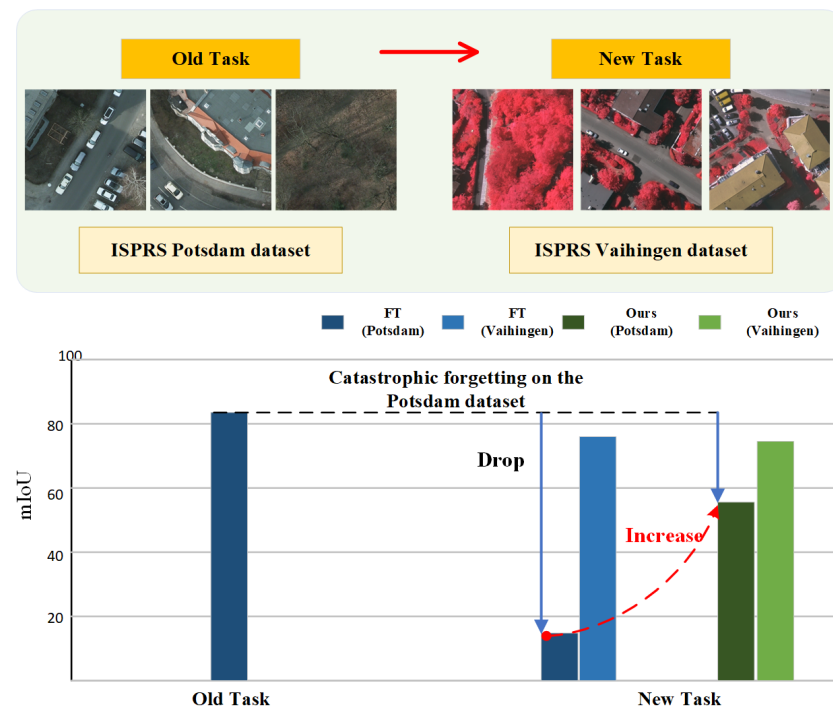


**Figure 1.** Example of the catastrophic forgetting problem in the domain incremental learning. The datasets of the old task and new task are the Potsdam and Vaihingen datasets. When training directly utilizing the new remote sensing images, the segmentation performance of the old task will drop significantly. When the proposed method was applied, the model retained the old knowledge and simultaneously learned the new knowledge.

Motivated by prompt learning technology [22,23], we propose a new domain incremental learning framework for remote sensing based on a visual domain prompt named RS-VDP, as shown in Figure 2. RS-VDP alleviates the domain shift problem by changing the learning strategy of the model from an input data space and feature space. Specifically, according to previous literature [24], the value of the entropy is related to the classification discernibility of the model. Therefore, we designed an entropy-based method to dynamically locate the locations with a large domain gap. At the same time, we use a visual domain prompt to directly change the data distribution of locations with large domain differences, thereby improving the classification accuracy of the model for locations that are difficult to distinguish. For the second problem, in order to obtain the class prototypes with high confidence, the representation feature alignment based on an entropy map module adopts a feature filtering method based on the entropy value to calculate the class prototypes for the located important feature vectors. We propose a regularization loss to optimize the parameters that play an important role in the prediction performance of the old task knowledge to the new model. This module not only increase the model's classification accuracy for old knowledge, but also solves the memory occupying problem of class prototypes. In addition, we proved that the traditional evaluation metric has a bias through formula derivation and analysis. When increasing the segmentation accuracy that has a smaller single-task accuracy, the traditional forgetting degree produces better results. Therefore, an incremental evaluation based on weighting that balances stability and plasticity to measure the performance of the algorithm more fairly is proposed. We

conducted experiments on datasets with domain differences commonly used [17,25,26] in the remote sensing field and proved the effectiveness of the proposed method.

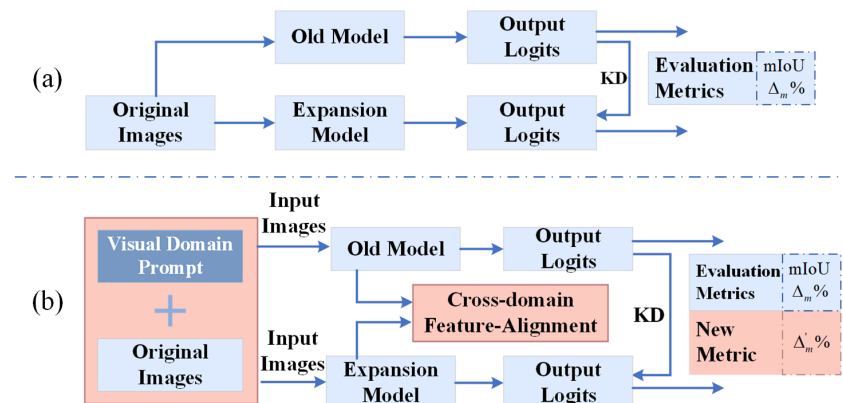


**Figure 2.** Schematic diagram of MDIL and ours. (a) Pipeline of MDIL; (b) Pipeline of ours.

In summary, our main contributions are as follows:

- We propose a novel domain incremental learning framework for the semantic segmentation task in remote sensing named RS-VDP. This module effectively reduces the degree of domain shift by adjusting both the input data space and the feature space.
- To locate the position with a large domain gap between the remote sensing images of different tasks, we design the domain prompt based on the dynamic location module. The module changes the data distribution of pixel locations with large local entropy values in the images, narrows the domain difference, and improves the classification accuracy of the incremental model.
- Considering the insufficient reliability of class prototypes and the memory usage problem, we propose representation feature alignment based on the entropy map module. Leveraging class prototypes as cross-task category cues, this module constrains the update of key parameters and aligns semantic feature distributions.
- We propose a new evaluation metric to fairly measure the plasticity and stability of incremental models. This metric solves the problem that the traditional evaluation metric is affected by the segmentation accuracy of the single-task method. We prove the scientificity of the proposed evaluation metric through formulas.

## 2. Related Work

### 2.1. Incremental Learning for Classification

Incremental learning methods [27–29] were designed to mitigate the forgetting problem, enabling the model to reserve the knowledge of old tasks and simultaneously acquire new knowledge without accessing the old dataset. The advantages of these approaches include avoidance of repeated use of the privacy and sensitive data and reducing the model's dependence on historical data. Depending on particular scenarios, incremental learning methods can be classified into class-incremental learning (CIL) [30–32], task-incremental learning (TIL) [33,34], and multi-domain incremental learning (MIL) [35,36]. CIL focuses on expanding the model's category space, where the data distribution remains consistent across incremental tasks while new categories are sequentially introduced. TIL divides the learning process into multiple independent subtasks, each with a dedicated category set and explicit task-ID. During testing, the model is required to recognize the target category under the specified task label. In contrast to CIL and TIL, DIL maintains a fixed category space throughout the incremental process, while the data domain distribution of tasks changes significantly. Further, incremental learning methods can be divided into replay-based

methods [27,37–40], parameter isolation-based methods [41,42], and regularization-based methods [43–46]. Replay-based methods retain representative samples or features from old tasks through dynamic memory storage to review the old knowledge. This kind of method repeatedly utilizes data of the old task, which inevitably leads to privacy and security issues. Regularization-based methods apply the loss functions to the model parameters or feature vectors related to old classes. Parameter isolation-based methods alleviate the forgetting problem by freezing important parameters or allowing the addition of new parameters. Nevertheless, the model fixed parameters modify the parameters that are crucial to old tasks, causing the model to optimize towards a suboptimal solution. To reconcile plasticity–stability trade-offs in domain continual learning scenarios, we utilize parameter isolation-based mechanisms.

## 2.2. Incremental Learning for Semantic Segmentation

The semantic shift of background classes is the key question to be solved for the incremental semantic segmentation task. Most methods [47–53] apply knowledge distillation losses to alleviate the forgetting. This kind of method has the advantage of not requiring the amount of incremental parameters and reducing the memory occupation. For example, SATS [53] proposes knowledge distillation by transferring the relationships between elements within each image. LSKD [50] involves a layer-specific knowledge distillation method, which can assign appropriate knowledge schemes and weights. The incremental learning methods based on replay-data reduce the degree of forgetting by replaying images [54], features [55], or auxiliary data [56] related to the old task. RECALL [56] applies webly-supervised learning to obtained the images about old tasks. ER-PASS [57] adopts a performance-aware submodular sampling strategy that considers feature diversity and task-specific performance to select representative samples, improving learning stability in domain-incremental learning. SSUL involves a class-balanced sampling strategy to store training images. Other articles [21,55,58–61] suggest preserving old knowledge using prototypes that can represent specific semantic categories. MISSNet [55] utilizes a local semantic distillation strategy and class-specific regularization loss using preserve prototypes. SIL-LAND [59] and LAG [61] employ distinct methods for calculating prototypes to distill old knowledge. The former methods update the prototypes by incorporating the old knowledge present in the images on the new task. The latter methods preserve the semantic-invariant knowledge as class prototypes. The dynamic architecture incremental learning methods [62,63] extend the architecture of the model. EACSS [63] explores the different architectures to verify the impact of the different layers of the model. RCIL [62] proposes exploiting a structural re-parameterization mechanism in order to decouple the representation knowledge of the old task and the new task. These kinds of methods can increase the amount of tasks indefinitely, so our method involves application of a dynamic architecture strategy.

## 2.3. Multi-Domain Incremental Learning

Multi-domain incremental learning models can learn knowledge from different data domains on multi-tasks without access to the knowledge of the old dataset, so as to retain the old knowledge while simultaneously acquiring new knowledge. These methods were first applied in the field of image classification and object detection. Some methods [13,18] mitigate the forgetting by storing sample data of old tasks. DISC [13] and MDMT [18] rehearse training samples or statistical parameters to replay the old knowledge. LMVD [16] proposes a tunable network architecture, which can be dynamically guided to different visual domains via the adapter residual module. Recently, multi-domain incremental learning has begun to be applied to the semantic segmentation task. MDIL [15] proposes

utilizing domain-invariant parameters and domain-specific parameters to capture homogeneous semantic features and the statistics of each domain. DILRS [19] proposes a unique parameter learning structure and new knowledge distillation loss to mitigate the impact of domain shift. GSMF [17] involves freezing feature layers and utilizing a graph space transformation and attention upsampling strategy. SimCS [12] is a parameter-free method using simulation data. Multi-domain incremental learning has been applied to many fields, including change detection [11], smart healthcare [14], and depth estimation [64], etc.

The multi-domain incremental learning methods and the domain adaptation methods have a certain degree of similarity. The differences are that the models for the incremental learning methods cannot access the old images in the subsequent tasks, but the models of the domain adaptation methods cannot exploit the label of the target domain. Recent papers [22,23] have proposed learning image-level visual domain prompts to extract domain-specific knowledge. However, these methods have not yet been decoupled and pay particular attention to the common context knowledge and the knowledge with a large domain gap. Inspired by prompt learning, we also utilize a visual domain prompt to mitigate domain shift issues.

### 3. Methodology

In this section, we first describe the notation definition and the network architecture. Then, a domain prompt module based on the dynamic location and a representative feature module based on an entropy map is designed. After that, a new evaluation metric is proposed to fairly measure the degree of forgetting.

#### 3.1. Notation Definition and Network Architecture

The objective of multi-domain incremental learning for semantic segmentation is to enable models to progressively acquire knowledge from emerging domains while maintaining performance on previously encountered domains, under the constraint that historical data are unavailable. Images from new and old domains differ in domain characteristics, meaning they are acquired by various sensors, or have different resolutions or pixel distributions.

Formally, different tasks are denoted as  $T = (1, 2, 3, \dots, t, \dots)$ . The train set and the test set of the  $t$  task are denoted as  $D_{train}^t = (X_{train}^t, Y_{train}^t)$  and  $D_{test}^t = (X_{test}^t, Y_{test}^t)$ . Similarly, the train set and the test set of the  $t-1$  task are denoted as  $D_{train}^{t-1} = (X_{train}^{t-1}, Y_{train}^{t-1})$  and  $D_{test}^{t-1} = (X_{test}^{t-1}, Y_{test}^{t-1})$ . We denote the original images of the  $t-1$  task as  $X_{train}^{t-1} = \{x_i\}_{i=1}^{K_{t-1}}$  and the corresponding ground truth as  $Y_{train}^{t-1} = \{y_i\}_{i=1}^{K_{t-1}}$ . The size of the original images is  $H \times W \times 3$  and the size of the ground truth is  $H \times W$ . It is noted that  $D_{test}^{t-1}$  and  $D_{test}^t$  are from different domains.

The overview framework of our proposed multi-domain incremental learning is shown in Figure 3. The model of the  $t$  task aims at achieving the map from the original data to enable dense prediction, namely,  $f_\theta^t : X_{train}^t \mapsto \mathbb{R}^{H \times W \times N_t}$ , where  $N_t$  denotes the total numbers of all tasks. During the training process, the model uses the data of the current task only. The old images and corresponding ground truth are not needed. The original data distribution is transformed by a visual prompt based on the dynamic location module to obtain the input data to be fed into the model. The input data are denoted  $\tilde{D}_{train}^t = (\tilde{X}_{train}^t, Y_{train}^t)$ . The input data are fed into the copy version of old model  $f_\theta^{t-1} = E^{t-1} \circ De^{t-1}$  trained on the old data, and the prediction logits  $P_{old}^{t-1} \in \mathbb{R}^{B \times N_{t-1} \times H \times W}$  are obtained. The copied model is frozen.  $E^{t-1}$  and  $De^{t-1}$  represent the encoder and decoder, respectively.  $B$  and  $N_{t-1}$  represent the batch size and the number of channels. Meanwhile, the input data are fed into the current model  $f_\theta^t = E^t \circ De^t$ , and the model outputs the prediction logits  $P_{new}^t \in \mathbb{R}^{B \times N_{t-1} \times H \times W}$  and  $P_{old}^t \in \mathbb{R}^{B \times N_{t-1} \times H \times W}$ . The former is the result of forward propagation of data through the branch of the current task, and the latter is the result of forward

propagation prediction of data through the branch of the old task. The feature map of the old model's encoder is denoted  $F_{t-1} = E_{\theta}^{t-1}(\tilde{x})$ . The feature map of the old task branch of the new model's encoder is represented as  $F_t = E_{\theta}^t(\tilde{x})$ . The domain-specific branches in the new model related to the old tasks and the decoder of the old tasks do not update the parameters, which will be described in detail in subsequent sections.

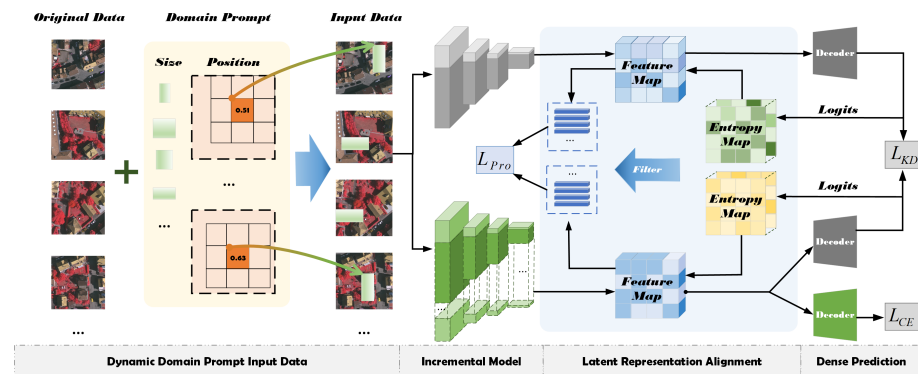


**Figure 3.** Overview framework of our proposed method. The input data come from original data and the visual domain prompt. The old models are frozen, and the parameters of the new model can be optimized. After the input data are fed into the incremental model, the representative features are calculated. We propose aligning the representative feature based on an entropy map. Then, the dense predictions are made by the model.

### 3.2. Domain Prompt Based on the Dynamic Location Module

The domain discrepancy between the old and new data distributions leads to suboptimal cross-domain generalization. Our approach circumvents this limitation by implementing parameter-efficient input adaptation via applying the visual domain prompt in the input data space. Previous research [24,65,66] has shown that a larger entropy value indicates that the probability logits of different channels are closer, which means that the classification discrimination is lower, and a smaller entropy value indicates that the probability logits are concentrated, which means that the classification discrimination is higher. Therefore, we utilize an entropy map as guidance to decide the position of the visual domain prompt. In particular, we propose a domain prompt based on the dynamic location module, whose purpose is to locate the visual domain prompt according to the local entropy weighted probability.

The module is shown as Figure 4. In contrast to locating the visual domain prompt in the area with the location of the largest entropy value corresponding to a pixel in the image, the proposed module covers the prompt to the location with a higher local entropy value composed of multiple pixels. Our proposed module locates the local areas with low discrimination by performing local probability weighting on the entropy map calculated based on the output probability. Then, the visual domain prompts are placed at the above locations. The module changes the data distribution by adding the prompt to the original data. The input data to be fed into the model can be denoted as follows:

We mainly focus on studying the size and location of the visual domain prompt. Previous methods applied a random visual prompt [22] or uncertainty domain prompt [23] to guide the location of the visual prompt. Since they do not cover the local area with a large domain gap effectively, the two approaches tend to result in the optimization of the model in an unexpected direction, which makes it difficult for the model to find the optimal solution. Therefore, we designed the domain prompt based on the dynamic location module according to the characteristics of the images, which can update the data distribution by

changing the pixels value at the area of the maximum local entropy. The entropy map is calculated based on the output logits of the model and is denoted as follows:

$$E_{\tilde{x}} = \frac{-1}{\log(N^t)} \sum_{c=1}^{N^t} P_{old}^{t-1} \log P_{old}^{t-1} \quad (1)$$

where  $E_{\tilde{x}}$  denotes the entropy map when the input image is  $\tilde{x}$ .

$$\tilde{x}_i = x_i + \text{Pad}(p) \quad (2)$$

where  $x_i$  denotes the original images of the new dataset.  $\text{Pad}(p) \in \mathbb{R}^{B \times 3 \times H \times W}$  represents the operation of padding with zero, which aims at making the size the same as the original images.

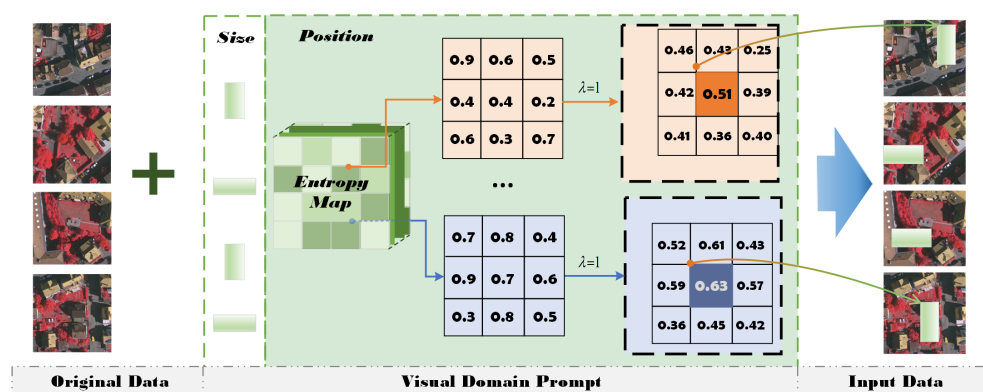


**Figure 4.** Illustration of domain prompt based on dynamic location module. Illustration of domain prompt based on dynamic location module. The input data come from the original data and domain prompt. The position of the visual domain prompt is decided by the maximum value of the locally weighted entropy map.

The value of the entropy effectively indicates the model's ability to discriminate the output of the pixels in the images. Therefore, the visual domain prompt based on the dynamic location module locates the region of maximum entropy in the images according to the entropy map. The module uses the convolution operation with the entropy map to obtain the result weighted by local probabilities. The local entropy weighting operation is denoted as:

$$\text{EWM}(h, w) = \frac{1}{(2l + 1)^2} \sum_{m=-l}^l \sum_{n=-l}^l E(h + m, w + n) \quad (3)$$

where  $\text{EWM}(h, w)$  denotes the summation of the entropy value in the local area  $(h, w)$  and  $2l + 1$  denotes the size of the kernel.

The domain prompt based on the dynamic location module applies the local entropy weighting operation before the original data are sent to the model. Since the local area with the largest local entropy value in each remote sensing images is not fixed, the visual prompt information is dynamic. After applying the above operation, the position of the pixel with the largest entropy value is used as the upper left corner of the visual domain prompt information. The input data can be obtained by the summation operation between the origin data and the domain prompt information after zero filling. The upper-left point of the visual domain information is represented by:

$$(\text{Position}_h, \text{Position}_w) = (\text{A}[\text{Max}(\text{EWM}(x))]_R, \text{A}[\text{Max}(\text{EWM}(x))]_C) \quad (4)$$

where  $\text{A}(\cdot)$  and  $\text{Max}(\cdot)$  represent the operation of argmax and the maximum, respectively.

$A[\text{Max}(\text{EWM}(x))]_R$  and  $A[\text{Max}(\text{EWM}(x))]_C$  denote the abscissa position and the ordinate position of the largest entropy value in the local area.

The domain prompt based on the dynamic location module places the visual domain prompt information to the original images by locating the most difficult to recognize area of the image, which changes the distribution of the new images and improves the attention where the domain gap is large. The updated data are beneficial for the model to learn the new knowledge, narrowing the domain gap of the old and new images. Meanwhile, we also leveraged the distillation loss to retain the prediction accuracy of the old knowledge.

### 3.3. Representation Feature Alignment Based on Entropy Map

Prototype-based regularization methods have demonstrated effectiveness in mitigating the domain shift by aligning class-specific characteristics across domains. This technique operates by computing centroid representations or prototypes for each category in both old and new domains, then enforcing alignment constraints through specially designed loss functions. The prerequisite for prototype calculation necessitates full label accessibility in both domains, requiring complete annotation of both source and target datasets.

Constrained by data privacy and prohibitive training costs, incremental learning methods under a strict data usage strategy must operate without access to historical datasets. The model aims to learn the new knowledge and maintain the old knowledge using the new images only. Considering that the prototype has the representation of semantic categories, using the prototype to align the representative feature vectors of the same category in different tasks not only makes the model retain more old knowledge, but also aligns the representation of different representative feature vectors to the same semantic category. Therefore, we utilized class prototypes and proposed the representation feature alignment module based on the entropy map. The schematic diagram is shown in the Figure 5. Moreover, since the old model still yields low prediction confidence for some pixels, we propose filtering based on entropy maps to maximize the reliability of the prototypes. In the training process, the images of the current batch are fed into the old model and the new model for forward propagation to obtain the feature map. It should be noted that the old model is frozen and only the parameters of the new model can be updated. Then, the regularization loss is calculated according to the class prototypes with high confidence to achieve the consistency of the distribution of aligned semantic features.

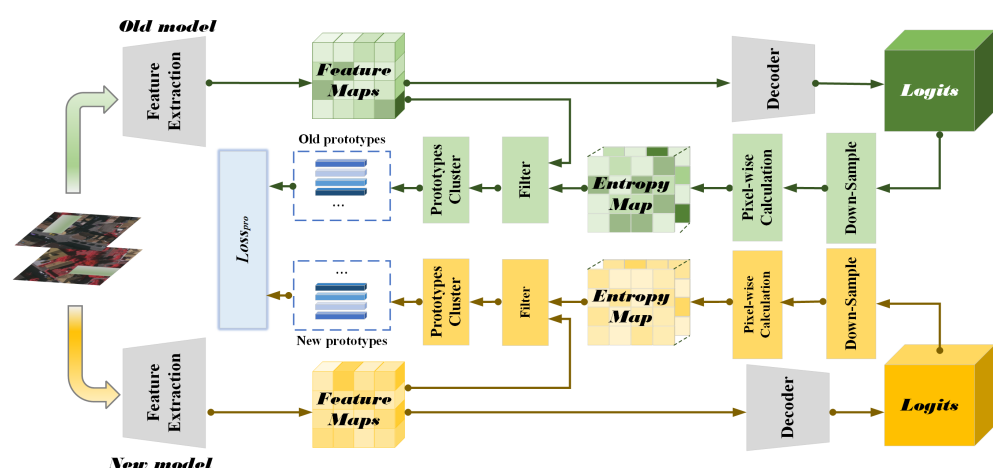


**Figure 5.** Illustration of representative feature alignment-based entropy map module.

Specifically, the entropy maps are calculated by summing the output probabilities along the channel dimension. In order to obtain the prototypes with high credibility, we filter the features of different classes based on their correspondence to the entropy value.

The class prototypes filtered according to the entropy map calculated by the new model are represented as:

$$Z_c^{new} = \frac{1}{|U_c^t|} \mathbb{I}_{E_\theta^t(r,z) < \tau} F_t \quad (5)$$

where  $Z_c^{new}$  is denoted as the prototype of the  $c_{th}$  semantic categories.  $U_c^t$  represents the pixels of feature  $F_t$  that belong to the  $c_{th}$  semantic categories after being filtered.  $|U_c^t|$  represents the total number of  $U_c^t$ .  $\mathbb{I}$  represents the index function, which can be denoted as:

$$\mathbb{I}_{E_\theta^t(r,z) < \tau} = \begin{cases} 1 & \text{if } E_\theta^t(r,z) < \tau \\ 0 & \text{if } E_\theta^t(r,z) > \tau \end{cases} \quad (6)$$

where  $\tau$  is a hyper parameter. When the entropy value of the pixel position  $(r, z)$  is less than  $\tau$ , the feature vector at this position is used to compute the class prototype. The prototypes filtered according to the entropy value of the old model can be denoted as:

$$Z_c^{old} = \frac{1}{|U_c^{t-1}|} \mathbb{I}_{E_\theta^t(r,z) < \tau} F_{t-1} \quad (7)$$

$$L_{pro} = \sum_{c=1}^{N^{t-1}} \|Z_c^{old} - Z_c^{new}\|_2 \quad (8)$$

### 3.4. The Incremental Evaluation Metric Based on Weighting

It is of great importance to impartially measure the performance degradation of different algorithms on both new and old tasks in the domain incremental learning tasks. We propose a new evaluation metric based on weighting. This evaluation metric will make the accuracy of baseline segmentation no longer affect the final forgetting performance, which can be expressed as:

$$\Delta'_m \% = \sum_{t=1}^T \omega_t \frac{\text{mIoU}_{m,t} - \text{mIoU}_{b,t}}{\text{mIoU}_{b,t}} \quad (9)$$

where  $\text{mIoU}_{m,t}$  represents the prediction accuracy of model  $m$  on the test set of task  $t$ .  $\omega_t$  represents weight, which can be denoted as:

$$\omega_t = \frac{\text{mIoU}_{b,t}}{\sum_{t=1}^T \text{mIoU}_{b,t}} \quad (10)$$

where  $\text{mIoU}_{b,t}$  refers to the performance of a single task and  $\sum_{t=1}^T \text{mIoU}_{b,t}$  represents the sum of the accuracy of  $T$  single tasks. This indicates that the weights are fixed values that do not introduce additional hyperparameters.

In accordance with the evaluation metrics used in papers [15,41,67], we also utilize  $\Delta_m \%$  to measure the overall incremental performance of the model.  $\Delta_m \%$  represents the performance degradation compared to the single task baseline  $b$ , which can be formulated as:

$$\Delta_m \% = \frac{1}{T} \sum_{t=1}^T \frac{\text{mIoU}_{m,t} - \text{mIoU}_{b,t}}{\text{mIoU}_{b,t}} \quad (11)$$

where  $\text{mIoU}_{m,t}$  denotes the prediction segmentation accuracy of the model  $m$  on the task  $t$  and  $T$  denotes the number of tasks. If  $\Delta_m \% < 0$ , it means the performance of the incremental learning model is worse than that of the single-task model.

Suppose that the prediction segmentation accuracy of the model  $m$  on the test set of the  $t$  task is increased by  $\delta \text{mIoU}$ , while the accuracy of the other tasks remains unchanged. The prediction accuracy of the model  $m$  on the test set of the  $t$  task is expressed as  $\text{mIoU}_{m,t} + \delta$ . At this time, the  $\Delta_m\%$  calculated from the new  $\text{mIoU}$  can be expressed as  $\Delta_m^\delta\%$ :

$$\begin{aligned}\Delta_m^\delta\% &= \frac{1}{T} \left[ \left( \frac{\text{mIoU}_{m,1} - \text{mIoU}_{b,1}}{\text{mIoU}_{b,1}} \right) + \left( \frac{\text{mIoU}_{m,2} - \text{mIoU}_{b,2}}{\text{mIoU}_{b,2}} \right) + \dots + \right. \\ &\quad \left. \left( \frac{\text{mIoU}_{m,t} + \delta - \text{mIoU}_{b,t}}{\text{mIoU}_{b,t}} \right) \right] \\ &= \frac{1}{T} \left[ \left( \frac{\text{mIoU}_{m,1} - \text{mIoU}_{b,1}}{\text{mIoU}_{b,1}} \right) + \left( \frac{\text{mIoU}_{m,2} - \text{mIoU}_{b,2}}{\text{mIoU}_{b,2}} \right) + \dots + \right. \\ &\quad \left. \left( \frac{\text{mIoU}_{m,t} - \text{mIoU}_{b,t}}{\text{mIoU}_{b,t}} \right) + \frac{\delta}{\text{mIoU}_{b,t}} \right] \\ &= \Delta_m\% + \frac{1}{T} \frac{\delta}{\text{mIoU}_{b,t}}\end{aligned}\quad (12)$$

It can be seen that when the accuracy is increased by  $\delta\%$ , the decrease degree of  $\Delta_m^\delta\%$  is related to  $\text{mIoU}_{b,t}$ . If  $\text{mIoU}_{b,t}$  is smaller, then  $\Delta_m^\delta\%$  is larger, that is, the evaluation metric of incremental learning is related to the accuracy of the single-task baseline method. This means that when the segmentation performance of datasets with lower prediction performance is improved as much as possible, better incremental segmentation results will be obtained. If we appropriately sacrifice the segmentation accuracy of datasets with high single-task prediction performance, in exchange for improved prediction accuracy of datasets with low single-task segmentation accuracy, the final evaluation metric of  $\Delta_m\%$  will be greatly improved.

In order to solve this problem, we propose a new evaluation metric based on weighting. When  $\Delta'_m\%$  is used as the evaluation metric, it is assumed that  $\delta$  represents the increase in the prediction accuracy of the model  $m$  to the dataset of the  $t$  task, and the accuracy of other tasks remains unchanged. Then,  $\Delta'_{m,\delta}\%$  is represented as:

$$\begin{aligned}\Delta'_{m,\delta}\% &= \omega_1 \left( \frac{\text{mIoU}_{m,1} - \text{mIoU}_{b,1}}{\text{mIoU}_{b,1}} \right) + \omega_2 \left( \frac{\text{mIoU}_{m,2} - \text{mIoU}_{b,2}}{\text{mIoU}_{b,2}} \right) + \dots + \\ &\quad \omega_t \left( \frac{\text{mIoU}_{m,t} + \delta - \text{mIoU}_{b,t}}{\text{mIoU}_{b,t}} \right) \\ &= \frac{\text{mIoU}_{b,1}}{\sum_{t=1}^T \text{mIoU}_{b,t}} \left( \frac{\text{mIoU}_{m,1} - \text{mIoU}_{b,1}}{\text{mIoU}_{b,1}} \right) + \frac{\text{mIoU}_{b,2}}{\sum_{t=1}^T \text{mIoU}_{b,t}} \left( \frac{\text{mIoU}_{m,2} - \text{mIoU}_{b,2}}{\text{mIoU}_{b,2}} \right) + \dots + \\ &\quad \frac{\text{mIoU}_{b,t}}{\sum_{t=1}^T \text{mIoU}_{b,t}} \left( \frac{\text{mIoU}_{m,t} - \text{mIoU}_{b,t}}{\text{mIoU}_{b,t}} \right) + \frac{\text{mIoU}_{b,t}}{\sum_{t=1}^T \text{mIoU}_{b,t}} \frac{\delta}{\text{mIoU}_{b,t}} \\ &= \Delta'_m\% + \frac{\delta}{\sum_{t=1}^T \text{mIoU}_{m,b}}\end{aligned}\quad (13)$$

Since both  $\delta$  and  $\sum_{t=1}^T \text{mIoU}_{m,b}$  are fixed values, the  $\Delta'_{m,\delta}$  cannot be changed by the accuracy of the single-task. Therefore, the evaluation metric  $\Delta'_{m,\delta}$  is an unbiased stability–plasticity equilibrium metric that intrinsically quantifies catastrophic forgetting in incremental segmentation paradigms. The proposed metric solves the problem that the traditional evaluation metric is affected by the accuracy of the single-task.

### 3.5. Total Loss

The domain incremental semantic segmentation method should not only maintain the prediction accuracy of the model on the old dataset, but also ensure the model has

good prediction accuracy on the new datasets with domain differences. By using the domain prompt based on the dynamic location module and the representation feature alignment module based on the entropy map, the difference between the data distribution and the semantic feature distribution is narrowed. In order to maintain consistency in the label distribution, we also use traditional knowledge distillation strategies, which can be formulated as:

$$\mathcal{L}_{KD} = \sum_{t=1}^T \sum_x l_{KL}(P_{new}^t, P_{old}^{t-1}) \quad (14)$$

where  $l_{KL}(p, q)$  denotes the loss of the Kullback–Leibler divergence.

Furthermore, when the model learns the knowledge of new datasets with domain difference, the new data and the correspondence to labels are available. Therefore, the segmentation loss function of the  $t$  task is represented as:

$$\mathcal{L}_{CE} = \frac{1}{K_t} \sum_{(x,y) \in D_{train}^t} l_{ce}(\theta; x, y) \quad (15)$$

where  $l_{ce}$  denotes the cross-entropy loss.

The total loss function of the model on the task  $t$  can be denoted as:

$$\mathcal{L}_{tot} = \mathcal{L}_{CE} + \lambda_{kd} \mathcal{L}_{KD} + \lambda_{pro} \mathcal{L}_{Pro} \quad (16)$$

## 4. Experiments and Results

We perform comprehensive empirical validation on the ISPRS Vaihingen and Potsdam datasets to rigorously assess the proposed framework's capabilities. Our experiments consist of six subsections: (1) experiment datasets, (2) implementation details, (3) evaluation metrics, (4) the ablation experiments, (5) comparative experiments, and (6) qualitative analysis.

### 4.1. Experiment Datasets

The images of the ISPRS Vaihingen [68] dataset are collected from a village in Germany. The dataset has 33 tiles of different sizes. It contains five foreground categories including tree, road, building, low vegetation, car, as well as one background class named others. To fairly compare with other methods, we also use 16 images for training and 17 images for testing. The images of the ISPRS Potsdam [68] dataset are from a city in Germany. There are 38 images of the same size. The categories of the Potsdam dataset are the same as the categories of the Vaihingen dataset. Twenty four of the images are used for training and the other 14 images are used for testing, which is the same as for the other methods.

The above two datasets belong to different domains, which can be shown in three ways. Firstly, the characteristics of the objects are different. The images in the Vaihingen dataset mainly contain independent buildings and multi-storey buildings. The objects of the images in the Potsdam dataset are mainly large buildings and narrow roads. Second, the distribution of pixels is different. The number of pixels of buildings is the largest in the Vaihingen dataset, but the number of pixels of road is the largest in the Potsdam dataset. Furthermore, the bands of the images are different. The bands of the images in the Vaihingen dataset are IRGR, but the bands of the images in the Potsdam dataset are RGB.

According to the above analysis and the experiment reported in previous papers [15,17,25,26], the Vaihingen dataset and the Potsdam dataset can be used to verify the model's ability with regard to domain incremental learning.

#### 4.2. Implementation Details

For a fair comparison with existing methods, we employ ERFNet as the backbone of our model, aligning with the architecture used in MDIL. The MDIL method serves as the baseline in our experiments. When optimizing the model, we apply the Adam optimizer. The learning rate is set to 0.001 on the first task. After that, the learning rate of the shared parameters is set to  $5 \times 10^{-7}$  and that of the special parameter is set to  $5 \times 10^{-6}$ . The parameter of the old task is not updated. We applied a poly strategy as the learning rate update scheduler, in which the power is set to 0.9. The size of the original images is  $512 \times 512$  pixels. Then, the images are resized to  $512 \times 1024$  pixels and a random flip operation is applied before feeding into the model. The data augmentation operations and the hyper-parameters are the same as for the baseline model and the other compared models, which means  $\lambda_{kd}$  is set to 0.1 and  $\lambda_{pro}$  is set to 10. All the experiments of this paper use an NVIDIA Tesla V100-32G GPU. The version of CUDA is 12.2 and the CPU is 6 vCPU Intel(R)Xeon(R) Gold6130. The version the operation system is 18.04.5.

#### 4.3. Evaluation Metrics

In order to evaluate the performance of the model, we use the  $\Delta_m\%$ ,  $\Delta'_m\%$ , and the mean intersection over union (mIoU) as the evaluation metrics. The functions of  $\Delta_m\%$  and  $\Delta'_m\%$  are described in Section 3.4. The mIoU can be denoted as follows:

$$\text{mIoU} = \frac{1}{C} \frac{\sum_{i=1}^C P_{ii}}{\sum_{i=1}^C \sum_{j=1}^C P_{ij} - \sum_{j=1}^C P_{ji} - P_{ii}} \quad (17)$$

where  $C$  represents the number of object categories, and  $P_{ij}$  denotes the pixel that belongs to class  $i$  but are predicted as class  $j$ .

#### 4.4. Ablation Experiments

So as to verify the effectiveness of our proposed method, we conduct a series of ablation experiments on the Vaihingen and Potsdam datasets. On the first task, the model is trained using the images of the Potsdam dataset. And on the next task, the Vaihingen dataset is used to train the incremental model. After that, we derive the prediction of the incremental model based on the test images of the Vaihingen and Potsdam datasets.

##### 4.4.1. Effectiveness of Each Component

The proposed method is composed of three important components, namely, domain-specific unit (DAU), dynamic visual domain prompt (DDP), and representative feature alignment (RFA). The domain-specific unit is proposed in MDIL [15]. DDP and RFA are proposed in Sections 3.2 and 3.3. The proposed evaluation metrics  $\Delta'_m\%$  are leveraged to measure the overall performance of the incremental learning model. The experimental results are depicted in Table 1.

**Table 1.** Ablation study for each component.

| Methods     | DAU | DDP | RFA | Pots  | Vai   | $\Delta_m\% \uparrow$ | $\Delta'_m\% \uparrow$ |
|-------------|-----|-----|-----|-------|-------|-----------------------|------------------------|
| Single-task | ×   | ×   | ×   | 81.26 | 77.12 | -                     | -                      |
| MDIL        | ✓   | ×   | ×   | 67.70 | 45.96 | -28.55                | -28.24                 |
| +RFA        | ✓   | ×   | ✓   | 52.15 | 69.56 | -22.81                | -23.15                 |
| +DDP        | ✓   | ✓   | ×   | 56.20 | 71.72 | -18.92                | -19.23                 |
| <b>Ours</b> | ✓   | ✓   | ✓   | 56.91 | 73.03 | <b>-17.63</b>         | <b>-17.96</b>          |

✓: The model adopts this module for training and testing; ×: The model doesn't adopt this module for training and testing; ↑: Larger evaluation metric value indicates better model performance. **Bold font** highlights the proposed method (Ours) achieves the optimal results.

The single-task learning belongs to the offline paradigm, necessitating independent model optimization for each domain-specific dataset. This method retains the independent identical distribution for each dataset. The prediction results are the theoretical maximum that can be achieved by the model with the same architecture and detail settings, with 77.12% mIoU and 81.26% mIoU achieved on the Vaihingen and Potsdam datasets. So as to increase the accuracy of the incremental learning model, the second row named MDIL is designed as DAU in the following task and adopts the cross-entropy and knowledge distillation losses to optimize the shared parameter and the domain specific parameter. It can be observed that MDIL achieves 67.70% mIoU and 45.96% mIoU on the Potsdam and Vaihingen datasets, respectively. The performance degradation is  $-28.55\%$  and  $-28.24\%$ . When adding the representation feature alignment based on the entropy map module and the domain prompt based on the location module to the baseline model MDIL, the performance degradation  $\Delta'_m\%$  of the incremental model increases by  $5.09\%$  ( $-28.24\% \rightarrow -23.15\%$ ) and  $9.01\%$  ( $-28.24\% \rightarrow -19.23\%$ ), respectively. When applying the above two new modules at the same time, our proposed method can achieve  $10.28\%$  ( $-28.24\% \rightarrow -17.96\%$ ) gains in  $\Delta'_m\%$ , with  $-18.92\%$   $\Delta_m\%$  and  $-19.23\%$   $\Delta'_m\%$ .

#### 4.4.2. Ablation Experiments for Prompt Strategy

In this section, we present ablation experiments to evaluate the effectiveness of the prompt strategies, as depicted in Table 2. According to the size and position of the prompt, there are five types of different strategies, which are fixed size and fixed position, fixed size and dynamic position, dynamic size and fixed position, dynamic size and dynamic position, and random prompt. The five type of strategies are denoted  $S_1 - S_5$ . The dynamic position means the model applies the domain prompt based on the dynamic location module. The dynamic size means the size of prompt is different in every step.  $S_5$  denotes the strategy of randomly placing fixed-size prompt information with the same size as those of the strategy  $S_2$ . The size of prompt depends on the entropy value and the threshold value, that is, the visual domain prompt is placed at the position where the entropy value is greater than 0.4.

**Table 2.** Ablation study for different prompt strategy.

| Strategy | Pots  | Vai   | $\Delta_m\% \uparrow$      | $\Delta'_m\% \uparrow$     |
|----------|-------|-------|----------------------------|----------------------------|
| $S_1$    | 54.31 | 70.48 | $-20.89$                   | $-21.21$                   |
| $S_2$    | 56.20 | 71.72 | <b><math>-18.92</math></b> | <b><math>-19.23</math></b> |
| $S_3$    | 55.61 | 68.39 | $-21.44$                   | $-21.71$                   |
| $S_4$    | 50.27 | 65.70 | $-26.47$                   | $-26.78$                   |
| $S_5$    | 52.81 | 67.12 | $-23.99$                   | $-24.28$                   |

$\uparrow$  indicates that a larger value of the evaluation metric corresponds to better model performance. **Bold font** highlights the optimal results.

The experimental results demonstrate that the proposed method with fixed size and dynamic position outperforms the others. When applying the strategy  $S_1$ , the visual domain prompt is located at the left and top of the images. The model achieves 54.31% mIoU on the Potsdam dataset and 70.48% mIoU on the Vaihingen dataset, respectively. We argue that changing the data distribution at a fixed location makes it difficult to make the domain prompt change the overall distribution of the dataset according to the characteristics of each image. Similarly, the strategy  $S_1$  also places the prompt in the middle position and the lower right corner. There is no doubt that the performance is not as good as  $S_2$ . This phenomenon may be caused by discontinuous prompts, which means the model cannot perform accuracy prediction on the context. When utilizing the visual domain prompt with dynamic size, the effects of  $S_3$  and  $S_4$  are worse than fixed size of  $S_1$  and  $S_2$ . When the

model applies the strategy  $S_5$ , the performance degradation reaches  $-24.28 \Delta'_m \%$ . The random prompting strategy cannot ensure that prompts are placed in regions requiring high attention. The model achieves optimal performance when trained with strategy  $S_2$ , yielding a performance degradation of  $-18.92 \Delta_m \%$  and  $-19.23 \Delta'_m \%$ . Therefore, the subsequent experiments are carried out based on the strategy  $S_1$ , and the data distribution difference is reduced by using the finite-size visual domain prompt information, which reduces the optimization difficulty of the model and improves the prediction accuracy.

#### 4.4.3. Ablation Experiments for Different Size of Prompt

The domain prompt based on the dynamic location module aims to change the distribution of data to enable better prediction. When designing the visual domain prompt, we conduct ablation experiments utilizing the strategy  $S_2$  with different sizes. Table 3 reports the mIoU and forgetting degradation results obtained on the Potsdam and Vaihingen datasets. The size of the visual domain prompt is designed to be 1:1 and 1:2 in the aspect ratios. The sizes are designed as  $50 \times 50$ ,  $50 \times 100$ ,  $100 \times 200$ , and  $512 \times 1024$ . As shown in Table 3, the prompt size of  $50 \times 100$  achieves the best balance across different datasets. Specifically, it yields the highest performance on the Potsdam dataset (56.20% mIoU) and the Vaihingen dataset (71.72% mIoU), while also providing the most favorable  $-18.92 \Delta_m \%$  and  $-19.23 \Delta'_m \%$ . This indicates that a  $50 \times 100$  prompt captures sufficient contextual information without introducing excessive computational overhead or redundant background noise. Smaller prompts may lack enough context, leading to suboptimal feature alignment, while larger prompts tend to include irrelevant background pixels, which degrade the model's ability to focus on target objects. When the dynamic visual domain prompt module is not applied, the model achieves 52.15% mIoU on the Potsdam dataset and 69.56% mIoU on the Vaihingen dataset, with the performance degradation reaching  $-22.81 \Delta_m \%$  and  $-23.15 \Delta'_m \%$ . In contrast, when the dynamic visual domain prompt module is employed, the prompts of different sizes all enhance the model's prediction accuracy on cross-domain data and reduce its performance degradation.

**Table 3.** Ablative analysis for different size of prompt.

| Size              | Pots  | Vai   | $\Delta_m \%$ ↑ | $\Delta'_m \%$ ↑ |
|-------------------|-------|-------|-----------------|------------------|
| $50 \times 50$    | 55.95 | 70.69 | −19.74          | −20.04           |
| $50 \times 100$   | 56.20 | 71.72 | <b>−18.92</b>   | <b>−19.23</b>    |
| $100 \times 200$  | 56.03 | 71.36 | −19.26          | −19.57           |
| $512 \times 1024$ | 55.88 | 70.62 | −19.83          | −20.13           |

↑ indicates that a larger value of the evaluation metric corresponds to better model performance. **Bold font** highlights the optimal results.

#### 4.4.4. Ablation Experiments for Prototypes

To verify the alignment degree of feature vectors after performing incremental learning tasks across different domains, we conducted ablation experiments that calculate the cosine similarity using the prototypes of the same categories across different tasks in the feature space. Specifically, we fed the same image into both the FT method and our model. The class centers are calculated in the feature space based on the feature vectors of each category. The experimental results are presented in Table 4. Each row denotes the cosine similarity between class prototypes, which are derived from the high-dimensional feature vectors of the old and new models. The FT method yields consistently low intra-category similarities across different tasks. This is because the model completely ignores the distribution of old-task categories in the high-dimensional space when learning new-task knowledge. In contrast, our method imposes constraints on the prototype vectors with high confidence scores from both the new and old models, aiming to align the feature distributions in the

high-dimensional space. As shown by the experimental results, compared with the FT method, our approach increases the average cosine similarity of class prototypes across different tasks by 0.46. These quantitative experimental results intuitively demonstrate that our method maintains favorable consistency in the high-dimensional space, which proves that the proposed approach facilitates feature alignment throughout the multi-domain incremental learning process.

**Table 4.** The average cosine similarity between class prototypes of old and new models.

| Method | Tree | Low Veg | Building | Road | Car  |
|--------|------|---------|----------|------|------|
| FT     | 0.35 | 0.28    | 0.31     | 0.45 | 0.20 |
| Ours   | 0.74 | 0.69    | 0.89     | 0.76 | 0.81 |

#### 4.4.5. Ablation Experiments for Hyper-Parameters

To ensure that the prototypes derived from features are representative, we propose representative feature alignment based on the entropy map. The old feature threshold and the new feature threshold, respectively, control the confidence of the feature vectors involved in the prototype calculation. The experiments in this section add the domain prompt based on the dynamic location module and representation feature alignment based on the entropy map module to the baseline model. The experimental results are shown as Tables 5 and 6. In the table,  $\tau$  and  $\tau'$  are denoted as the threshold of the new features and old features, separately. It can be observed that when the threshold of the old features is fixed, the feature vectors involved in the calculation of each category prototype gradually decrease as the new feature threshold decreases. But these feature vectors have higher confidence, indicating that the location feature is more likely to be predicted as the correct categories. When the new feature threshold is set to 0.1, the forgetting degradation achieves  $-17.96\%$ .

The second experiment's setting is that the threshold of the new features is fixed and the threshold of the old features is adjusted. The forgetting degree of the model goes through the process of first decreasing and then increasing. When the threshold of the old feature is 0.2, the optimal effect is achieved. The experimental results show that when the threshold of the old feature is 0.2 and the threshold of the new feature is 0.1, the model achieves excellent performance when learning the knowledge of the new task while maintaining the knowledge of the old task at the same time. When the prototype is calculated using the feature vector with credibility, the prediction effect is better. This shows that filtering representative features based on the entropy value can help to reduce the distance between prototypes of the same class between different domains, which proves the effectiveness of the method proposed in this paper.

**Table 5.** Ablative analysis for threshold of  $\tau$ .

| $\tau'$ | $\tau$ | Pots  | Vai   | $\Delta_m\% \uparrow$ | $\Delta'_m\% \uparrow$ |
|---------|--------|-------|-------|-----------------------|------------------------|
| 0.2     | 0.8    | 55.84 | 72.24 | $-18.81$              | $-19.14$               |
| 0.2     | 0.6    | 57.50 | 70.75 | $-18.75$              | $-19.02$               |
| 0.2     | 0.4    | 57.39 | 71.22 | $-18.51$              | $-18.79$               |
| 0.2     | 0.2    | 56.59 | 72.96 | $-17.87$              | $-18.21$               |
| 0.2     | 0.1    | 56.91 | 73.03 | $-17.63$              | $-17.96$               |

$\uparrow$  indicates that a larger value of the evaluation metric corresponds to better model performance.

**Table 6.** Ablative analysis for threshold of  $\tau'$ .

| $\tau'$ | $\tau$ | Pots  | Vai   | $\Delta_m\% \uparrow$ | $\Delta'_m\% \uparrow$ |
|---------|--------|-------|-------|-----------------------|------------------------|
| 0.1     | 0.1    | 58.21 | 71.09 | −18.09                | −18.36                 |
| 0.2     | 0.1    | 56.91 | 73.03 | −17.63                | −17.96                 |
| 0.4     | 0.1    | 57.21 | 71.23 | −18.61                | −18.90                 |
| 0.6     | 0.1    | 56.26 | 72.60 | −18.31                | −18.64                 |
| 0.8     | 0.1    | 56.74 | 72.06 | −18.37                | −18.68                 |

↑ indicates that a larger value of the evaluation metric corresponds to better model performance.

#### 4.5. Comparative Experiments on the SOTA Methods

##### 4.5.1. Quantitative Analysis

Our experimental protocol implements the incremental learning framework on the ISPRS Vaihingen and Potsdam aerial imagery benchmarks. The evaluation paradigm encompasses two scenarios: (1) Vaihingen  $\rightarrow$  Potsdam; (2) Potsdam  $\rightarrow$  Vaihingen. During sequential training phases, the model strictly adheres to stringent data isolation constraints, with no access to prior domain samples or annotations. Final performance metrics are derived through comprehensive assessment on test sets from both geographical domains.

Consistent with the comparison methods used in the baseline method, the methods adopted for quantitative analysis include incremental learning methods and non-incremental learning methods. Among them, the semantic segmentation method based on multi-domain incremental learning includes the method based on knowledge distillation and the method based on parameter extension. Non-incremental learning methods include single-task methods and multi-task methods. The single-task method is used to train the model on the separate dataset. The model of the multi-task method utilizes the images and labels of multiple datasets to train and test at the same time, which means the model can access all the data.

In comparison with the multi-task method, the incremental learning methods occupied less memory and avoid the use of private data. The training methods of fine tuning (FT), fixed representation (FR), learning without forgetting (LWF [43]), and incremental learning techniques (ILT [69]) are the same as the single task method on the first task. But the training method is different on the subsequent tasks. The FT method directly updates the parameters on the model trained by the first task, and then leverages the data and labels of the second task only to perform forward propagation and gradient backpropagation with the cross-entropy loss function. The learning rate applies a relatively small value. The FR approach maintains the encoder parameters as frozen while exclusively optimizing the decoder components with novel task data. In contrast, LWF implements a dual-phase knowledge preservation mechanism: (1) output probability distribution alignment through knowledge distillation strategy with Kullback–Leibler (KL) divergence loss, transferring the knowledge by the trained model in the previous task to the current model, and (2) novel knowledge learning via the cross-entropy loss function. The ILT method uses knowledge distillation loss on the output logits layer as well as on the feature layer. The baseline method MDIL uses a domain-specific parallel residual adaptation layer for the first task, which is a domain-invariant parameter for a specific domain. In the second task, the domain-specific parameters of the old task remain unchanged, while the shared parameters are updated with a smaller learning rate. The new domain-specific parameters are updated with a relatively large learning rate. The GSMF [17] method freezes feature layers and utilizes multi-feature joint loss optimization to retain prior knowledge and acquire new information. The DCL [70] applies feature-level and prediction-level consistency learning to solve the problem of forgetting. The domain feature consistency learning and class awareness consistency learning enable the target domain model to extract generalized knowledge.

To systematically validate the methodological efficacy, we first implemented incremental learning evaluation applying the Potsdam and Vaihingen datasets incrementally. Table 7 presents the performance evolution of the above-mentioned approaches in the domain incremental learning setting. The single task method yields the 77.12% mIoU of the Vaihingen dataset and 81.26% mIoU of the Potsdam dataset. The excellent results are acquired by optimizing the model parameters with a single dataset. The multi-task model optimizes all the parameters using all the dataset, with  $\Delta_m$  of  $-4.69\%$  and  $\Delta'_m$  of  $-4.78\%$ . Due to sharing the parameters of multiple datasets from different domains, the performance of the multi-task method is worse than that of the single-task method. The above two non-incremental learning methods have relatively higher requirements for data.

**Table 7.** Quantitative results for incremental learning from the Potsdam to Vaihingen dataset.

| Methods     | Step1: Potsdam |                        | Step2: Potsdam $\rightarrow$ Vaihingen |       |                            |                            |
|-------------|----------------|------------------------|----------------------------------------|-------|----------------------------|----------------------------|
|             | Pots           | $\Delta'_m\% \uparrow$ | Pots                                   | Vai   | $\Delta_m\% \uparrow$      | $\Delta'_m\% \uparrow$     |
| Single-Task | 81.26          | -                      | 81.26                                  | 77.12 | -                          | -                          |
| Multi-Task  | 81.26          | -                      | 74.59                                  | 76.22 | $-4.69$                    | $-4.78$                    |
| FT          | 81.26          | -                      | 16.43                                  | 76.61 | $-40.22$                   | $-41.26$                   |
| FR          | 81.26          | -                      | 81.26                                  | 21.56 | $-36.02$                   | $-35.08$                   |
| LwF         | 81.26          | -                      | 53.91                                  | 56.19 | $-30.40$                   | $-30.48$                   |
| ILT         | 81.26          | -                      | 55.73                                  | 55.94 | $-29.44$                   | $-29.49$                   |
| MDIL        | 79.80          | $-1.80$                | 67.7                                   | 45.96 | $-28.55$                   | $-28.24$                   |
| GSMF        | 74.87          | $-7.86$                | 61.91                                  | 65.81 | $-19.24$                   | $-19.36$                   |
| DCL         | 81.26          | -                      | 63.04                                  | 65.75 | $-18.58$                   | $-18.68$                   |
| <b>Ours</b> | 79.80          | $-1.80$                | 56.91                                  | 73.03 | <b><math>-17.63</math></b> | <b><math>-17.96</math></b> |

$\uparrow$  indicates that a larger value of the evaluation metric corresponds to better model performance. **Bold font** highlights the optimal results.

The FT method almost forgets all knowledge of the old data since the model updated all parameters trained by the old data. Nevertheless, it achieves optimal performance on the new task, with an mIoU of 76.61% and  $-41.26\%$   $\Delta'_m$ . In contrast, the FR method achieves the optimal performance on the old dataset, yielding 81.26% mIoU, but exhibits poor adaptability to the new task, with  $\Delta'_m$  of  $-35.08\%$ . The LwF method utilizes the knowledge distillation loss function on logits to retain the knowledge of the old task, improving the overall performance of the model. On the basis of LWF, the ILT method distills the knowledge on both the feature and logits layers, yielding better accuracy and less forgetting than LWF. The baseline method MDIL expands the model by adding a small number of domain-specific parameters, and the overall performance degradation  $\Delta'_m$  reaches  $-28.24\%$ . By utilizing the graph space framework with multifeature constraints, the GSMF method outperforms the MDIL method. By adopting the proposed consistency learning strategy, the DCL method achieves mean intersection over union scores of 63.04% and 65.75%, respectively, with performance degradation of  $-18.58\%$   $\Delta_m$  and  $-18.68\%$   $\Delta'_m$ . As a suboptimal method, DCL lags behind the proposed method by 0.95  $\Delta_m\%$  and 0.72  $\Delta'_m\%$ .

Quantitative evaluations substantiate that our framework achieves superior performance in the old and new domains. Compared with the baseline method MDIL, we improve the mIoU of the new task by 27.07%, with 56.91% mIoU of the old task and 73.03% mIoU of the new task. The performance degradation  $\Delta'_m$  increases to 10.28%. The excellent performance demonstrates that the proposed method is effective in mitigating the catastrophic forgetting problem in the domain incremental learning task.

Furthermore, a comprehensive comparison between our and other superior methods optimized on the Vaihingen and Potsdam datasets incrementally are shown in Table 8. Our approach establishes new performance benchmarks across all evaluated methodologies,

demonstrating superior prediction performance over existing incremental learning methods such as MDIL, ILT, LwF, MDIL<sub>r</sub>, and DCL. The advantages of the proposed method can be further elaborated from three distinct perspectives. First, the prediction performance on the original task is improved by 24.91% mIoU compared with the FT method. Second, for the novel task, our approach outperforms the FR method by 49.34% mIoU, and exceeds the DCL method by 13.18% mIoU (65.88%  $\rightarrow$  79.06%). This substantial gain further validates the superiority of our method in handling new task data. In addition, our framework improves by 4.10  $\Delta'_m\%$  and 3.78  $\Delta'_m\%$  over the baseline method and reaches 79.06% mIoU on the new dataset. Furthermore, our method also achieves competitive performance in preserving previously learned knowledge, ultimately outperforming the DCL method by 1.46  $\Delta_m\%$  and 1.84  $\Delta'_m\%$ . These consistent quantitative advancements validate the framework's optimization efficacy.

**Table 8.** Quantitative results for incremental learning from the Vaihingen to Potsdam dataset.

| Methods     | Step1: Vaihingen |                        | Step2: Vaihingen $\rightarrow$ Potsdam |       |                       |                        |
|-------------|------------------|------------------------|----------------------------------------|-------|-----------------------|------------------------|
|             | Vai              | $\Delta'_m\% \uparrow$ | Vai                                    | Pots  | $\Delta_m\% \uparrow$ | $\Delta'_m\% \uparrow$ |
| Single-task | 77.12            | -                      | 77.12                                  | 81.26 | -                     | -                      |
| Multi-task  | 77.12            | -                      | 76.22                                  | 74.59 | -4.69                 | -4.78                  |
| FT          | 77.12            | -                      | 19.19                                  | 82.63 | -36.71                | -35.71                 |
| FR          | 77.12            | -                      | 77.12                                  | 29.72 | -31.71                | -32.54                 |
| LwF         | 77.12            | -                      | 41.93                                  | 66.86 | -31.68                | -31.31                 |
| ILT         | 77.12            | -                      | 44.39                                  | 66.02 | -30.60                | -30.29                 |
| MDIL        | 76.71            | -0.53                  | 50.77                                  | 65.90 | -26.54                | -26.34                 |
| DCL         | 77.12            | -                      | 54.36                                  | 65.88 | -24.22                | -24.08                 |
| <b>Ours</b> | 76.71            | -0.53                  | 44.10                                  | 79.06 | <b>-22.76</b>         | <b>-22.24</b>          |

$\uparrow$  indicates that a larger value of the evaluation metric corresponds to better model performance. **Bold font** highlights the optimal results.

In order to clearly compare the prediction performance of every semantic class, Figure 6 presents the IoU of each class and the performance degradation when incrementally learning the knowledge of the Vaihingen and Potsdam datasets. Figure 6a shows the IoU of each class. The abscissa describes different approaches, such as FT, MDIL and ours. For each method, the left side shows the prediction IoU of the old task, and the right side is the prediction IoU of the new task. The experimental results substantiate that our framework effectively preserves old knowledge while achieving robust adaptation to novel data distributions, with the performance on emerging tasks reaching competitive levels comparable to the FT method, particularly evident in road, building, and vehicle class segmentation accuracy. In terms of overall measures, the presented method has the lowest level of performance degradation among all comparison methods. Figure 7 shows the results IoU for each category and performance degradation of every method when the Potsdam and Vaihingen datasets are used for incremental learning, respectively. The proposed method gains better prediction results for tree categories on the historical task, and the results on certain categories are still competitive. The prediction results for each category of the novel task exceed that of other advanced methods. The forgetting degree has been reduced to the lowest level compared to other methods, which alleviates the problem of catastrophic forgetting. The outstanding quantitative results substantiate the efficacy of the proposed approach.

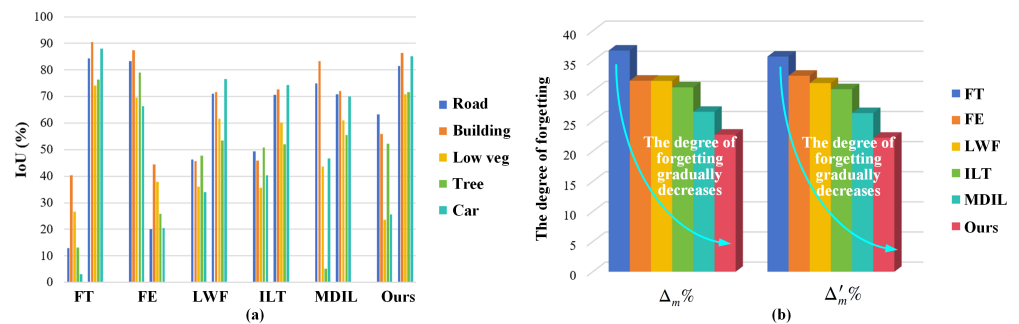


**Figure 6.** The IoU of all semantic classes and all methods for incremental learning from the Vaihingen to Potsdam dataset. (a) The Iou of every class. On the left is the prediction IoU of the old task, and on the right is the IoU of the new task; (b) The  $\Delta_m$  and  $\Delta'_m$  of every method.

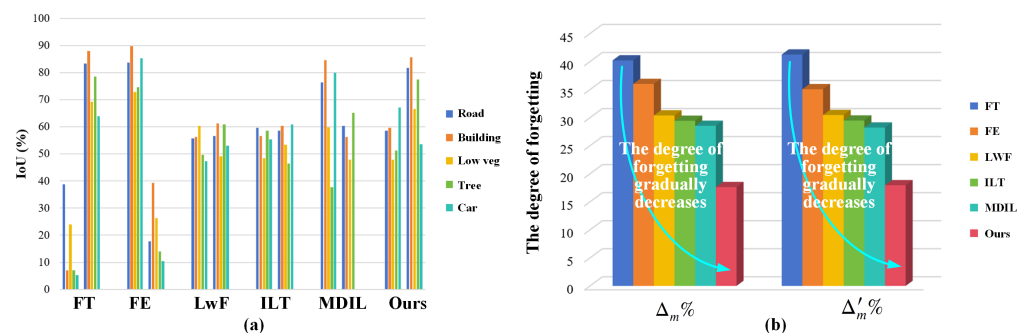


**Figure 7.** The IoU of all semantic classes and all methods for incremental learning from the Potsdam to Vaihingen dataset. (a) The Iou of every class. On the left is the prediction IoU of the old task, and on the right is the IoU of the new task; (b) The  $\Delta_m$  and  $\Delta'_m$  of every method.

In addition, memory consumption is also an important aspect to assess to verify the effectiveness of the proposed method. Classic sample-storage methods [54,57,71] reserve an average of 20 samples per category. The ISPRS Vaihingen and ISPRS Potsdam datasets each contain 5 categories. Therefore, an average of 100 samples needs to be stored in the memory buffer. We counted the memory usage per image in the Vaihingen and Potsdam datasets, respectively. In the Vaihingen training set, the smallest image occupies 316 KB of memory and the largest one occupies 596 KB, with an average memory footprint of 449 KB per image. In the Potsdam dataset, the smallest image takes up 278 KB of memory and the largest one takes up 592 KB, with an average memory consumption of 425 KB per image. Based on the average memory usage of the Vaihingen and Potsdam datasets, the total memory required for sample storage amounts to 43,700 KB. Furthermore, we measured the memory usage of our model and the baseline MDIL model, which were 31,279 KB and 30,223 KB, respectively. Compared with MDIL, our method only stores domain-specific parameters that take up 56 KB of memory, resulting in a memory saving of 43,644 KB. These results further validate that our method is a practical and scalable approach to continual learning in remote sensing.

#### 4.5.2. Qualitative Analysis

The visual results can intuitively highlight the predictive discrepancies between our framework and the baseline method. Figure 8 shows the visualization results of the incremental learning model on the Vaihingen dataset and the Potsdam dataset. The model architectures used by the two methods in different tasks are consistent. The figure shows the final prediction visualization results after all the incremental learning tasks are completed. The first two columns represent the images and corresponding labels of the Vaihingen dataset, respectively. Columns 3–4 correspond to the baseline framework and our novel

framework's prediction outputs, respectively. Qualitative analysis reveals the baseline erroneously conflates tree and low-vegetation categories, while our approach is not precise enough in predicting the category of building. In terms of the performance of the novel task, the presented approach demonstrates superior performance. For the building category, the prediction results of the baseline method are not continuous and do not reflect the integrity of the building. When only low vegetation and trees are included in the image, the prediction of the baseline model has certain noise spots, and the low vegetation is incorrectly predicted as the class of tree. For the image of the parking lot scene, the prediction results of the baseline method also contain discontinuous noise pixels, and the confidence in the roads and buildings near the car is not high enough. Moreover, the baseline method contains more false predictions than our method. The proposed approach gains outstanding segmentation results in different semantic categories and the overall prediction of most objects. The good visualization results further prove the effectiveness of the proposed approach.

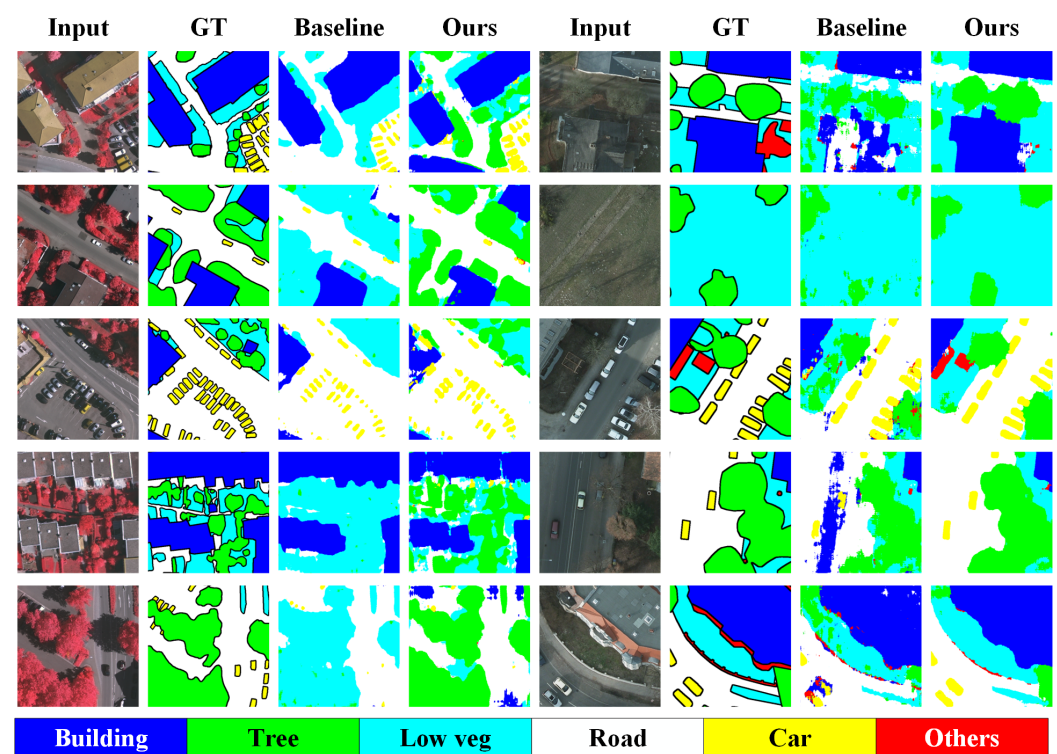


**Figure 8.** Qualitative results for incremental learning from the Vaihingen dataset to the Potsdam dataset.

Figure 9 shows the visualization results of the incremental Potsdam dataset and the Vaihingen dataset, respectively. It can be observed that the prediction results of the proposed approach are better than that of the baseline method. The fifth and sixth columns represent the original images and corresponding labels of the Vaihingen dataset, respectively. The last two columns represent the prediction performance of the baseline and the method proposed in this paper, respectively. In the first row, the baseline method incorrectly predicts the object on the roof of the building as the category of road, and cannot accurately predict the vehicle class. In the second row, there is a lot of noise in the visualization image of the baseline method. Visual results of the third row reveal the baseline framework's failure to segment vehicular instances within parking facilities coupled with imprecise architectural delineation. By comparison, the fourth row exhibits the baseline's critical misclassification of shadow-obscured residential structures with light-toned roofing as

road. In general, the baseline method does not provide good predictions for vehicles on the roads or in parking lots, and it is very difficult to accurately classify objects in the shadow or objects of similar color. Our framework achieves better visualization results on multiple semantic categories, and the prediction effect of each category exceeds that of the baseline method when visualized on a new task. The visualization results correspond to the quantitative analysis results, corroborating the efficacy of the proposed framework.

For comprehensive comparative analysis, we fed the images of the old task into the historical model and novel model of the FT method and proposed method. We visualize the t-distributed Stochastic Neighbor Embedding(t-SNE) dimensionality reduction of the features, as shown in Figure 10, where circular and triangular markers denote the feature reduction results from the novel and history model, respectively. For the FT method, the circles and triangles makers of the same color cannot coincide, which means that the updated model cannot have a uniform decision boundary for the same category in both the novel data and the old data. At this time, the decision boundary of the FT method indicates that this method has a poor quantitative segmentation effect on the old dataset, and a relatively good quantitative segmentation effect on the images of the novel task. When applying the proposed method, the models of different tasks achieve good overlap for the features of the same class after dimensionality reduction, corresponding to the circular dashed line in the figure, which proves that adjusting the data distribution using the visual domain prompt is helpful to accurately segment the objects of multiple different domains at the same time. This figure proves that the proposed method can substantiate the effectiveness of our proposed multi-domain incremental learning method.

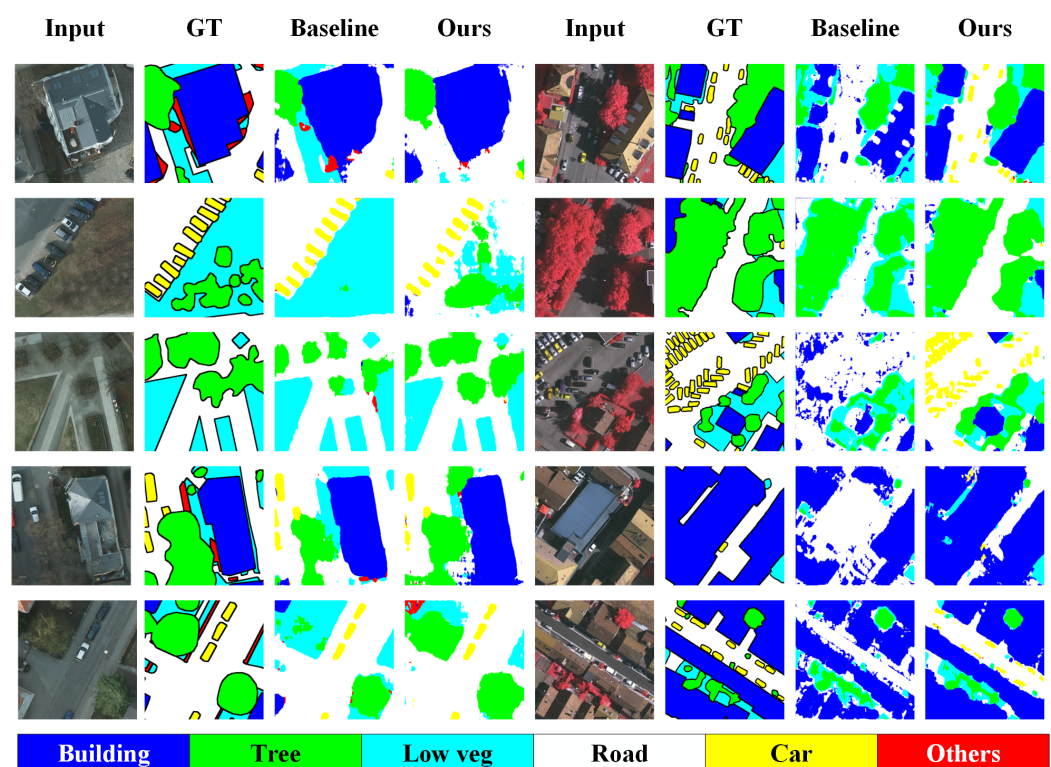


**Figure 9.** Qualitative results for incremental learning from the Potsdam dataset to the Vaihingen dataset.

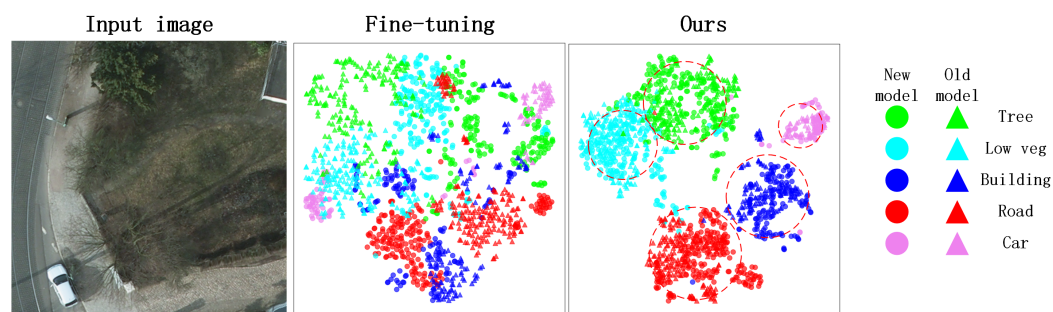


**Figure 10.** The tSNE visualisation results of fine-tuning and ours.

## 5. Discussion

In the previous section, a series of ablation experiments verify the effectiveness of the proposed modules, including DDP and RFA modules, the size and strategy of the domain prompt, the similarity of class prototypes, and the setting of hyperparameters. In addition, we perform quantitative comparisons with both non-incremental and incremental learning methods, and intuitively present the prediction visualization results. The experimental results demonstrate that our method achieves state-of-the-art (SOTA) performance with the lowest degree of forgetting. Additionally, we compare the memory consumption with replay-based methods, demonstrating that our approach reduces memory usage by over 90%. This further validates the effectiveness of the proposed method in practical applications.

Our research still has some limitations. Constrained by the limitations of existing data resources, verifying the effectiveness of domain-incremental learning methods on data from different seasons, diverse sensors, and cross-climatic zones is a highly promising research direction. In the future, we will construct new datasets that include data from different satellites, multi-seasonal scenarios, and distinct climatic conditions, to verify the robustness of the models on data from different domains. Few-shot incremental learning is also worthy of research, and we will consider conducting studies on few-shot incremental semantic segmentation for remote sensing images.

## 6. Conclusions

In this paper, we propose a new domain incremental learning framework for aerial images via a visual domain prompt named RS-VDP to alleviate the problem of catastrophic forgetting caused by domain shift. We address the issue from the perspective of changing the distribution of input images and features. RS-VDP introduces a domain prompt based on a dynamic location module to locate the visual domain prompt according to the local entropy weighted probability. This module places visual domain prompt tokens on the original images and dynamic shifts them to adapt to the new domain. The representation feature alignment based on the entropy map module uses the feature vectors with higher confidence in each batch to calculate the class prototype. We use the regularization loss to reduce the possibility of changes in important parameters that affect the prediction accuracy of old tasks, while avoiding the problem of class prototype information occupying memory. Moreover, a new evaluation metric for incremental learning models is proposed to fairly measure plasticity and stability, which solves the problem that the traditional evaluation metric is more sensitive to datasets with poor segmentation accuracy of a single-task. Extensive experiments on multi-domain datasets demonstrate that our method achieves SOTA performance and reduces the forgetting degree of the incremental learning model.

**Author Contributions:** Conceptualization, J.L. and Z.Y.; methodology, J.L. and Z.Y.; validation, J.L. and Y.Z.; formal analysis, J.L. and Z.Z.; investigation, J.L. and Y.T.; resources, J.L. and X.S.; data curation, J.L. and W.D.; writing—original draft preparation, J.L.; writing—review and editing, J.L.,

Z.Y., and W.D.; visualization, J.L.; supervision, Z.Y.; project administration, Y.T.; funding acquisition, X.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China under Grant 62501574, the National Natural Science Foundation of China under Grant 62425115, the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation under Grant Number GZC20250213, and the Open Fund of the Key Laboratory of National Geographic Census and Monitoring, Ministry of Natural Resources (2025NGCM05).

**Data Availability Statement:** All data that support the findings of this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Li, Y.; Li, X.; Li, W.; Hou, Q.; Liu, L.; Cheng, M.M.; Yang, J. Sardet-100k: Towards open-source benchmark and toolkit for large-scale sar object detection. *arXiv* **2024**, arXiv:2403.06534.
2. Sun, X.; Wang, P.; Yan, Z.; Xu, F.; Wang, R.; Diao, W.; Chen, J.; Li, J.; Feng, Y.; Xu, T.; et al. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 116–130. [\[CrossRef\]](#)
3. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 18–20 June 2018; pp. 3974–3983.
4. Xiao, C.; An, W.; Zhang, Y.; Su, Z.; Li, M.; Sheng, W.; Pietikäinen, M.; Liu, L. Highly efficient and unsupervised framework for moving object detection in satellite videos. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 11532–11539. [\[CrossRef\]](#)
5. Zhou, J.; Liu, Y.; Peng, B.; Liu, L.; Li, X. MaDiNet: Mamba Diffusion Network for SAR Target Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 10787–10800. [\[CrossRef\]](#)
6. Cheng, J.; Deng, C.; Su, Y.; An, Z.; Wang, Q. Methods and datasets on semantic segmentation for Unmanned Aerial Vehicle remote sensing images: A review. *ISPRS J. Photogramm. Remote Sens.* **2024**, *211*, 1–34. [\[CrossRef\]](#)
7. Ding, H.; Xia, B.; Liu, W.; Zhang, Z.; Zhang, J.; Wang, X.; Xu, S. A novel mamba architecture with a semantic transformer for efficient real-time remote sensing semantic segmentation. *Remote Sens.* **2024**, *16*, 2620. [\[CrossRef\]](#)
8. Jia, P.; Chen, C.; Zhang, D.; Sang, Y.; Zhang, L. Semantic segmentation of deep learning remote sensing images based on band combination principle: Application in urban planning and land use. *Comput. Commun.* **2024**, *217*, 97–106. [\[CrossRef\]](#)
9. Klietk, T.; Musa, H.; Machova, V.; Rice, L. Remote sensing data fusion techniques, autonomous vehicle driving perception algorithms, and mobility simulation tools in smart transportation systems. *Contemp. Readings Law Soc. Justice* **2022**, *14*, 137–152. [\[CrossRef\]](#)
10. McCloskey, M.; Cohen, N.J. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of Learning and Motivation*; Elsevier: Amsterdam, The Netherlands, 1989; Volume 24, pp. 109–165.
11. Weng, L.; Yang, W.; Hu, B.; Han, P.; Xue, S.; Zhang, Y.; Li, H.; Jin, J.; Bu, S. MDINet: Multi-Domain Incremental Network for Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4402315. [\[CrossRef\]](#)
12. Alfara, M.; Cai, Z.; Bibi, A.; Ghanem, B.; Müller, M. SimCS: Simulation for Domain Incremental Online Continual Segmentation. *Proc. AAAI Conf. Artif. Intell.* **2024**, *38*, 10795–10803. [\[CrossRef\]](#)
13. Mirza, M.J.; Masana, M.; Possegger, H.; Bischof, H. An efficient domain-incremental learning approach to drive in all weather conditions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2022; pp. 3001–3011.
14. Li, C.H.; Jha, N.K. PAGE: Domain-Incremental Adaptation with Past-Agnostic Generative Replay for Smart Healthcare. *arXiv* **2024**, arXiv:2403.08197. [\[CrossRef\]](#)
15. Garg, P.; Saluja, R.; Balasubramanian, V.N.; Arora, C.; Subramanian, A.; Jawahar, C. Multi-domain incremental learning for semantic segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 761–771.
16. Rebuffi, S.A.; Bilen, H.; Vedaldi, A. Learning multiple visual domains with residual adapters. In *Advances in Neural Information Processing Systems*; NeurIPS Proceedings: San Diego, CA, USA, 2017; Volume 30.
17. Huang, W.; Ding, M.; Deng, F. Domain Incremental Learning for Remote Sensing Semantic Segmentation with Multi-Feature Constraints in Graph Space. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5645215. [\[CrossRef\]](#)
18. Lyu, F.; Wang, S.; Feng, W.; Ye, Z.; Hu, F.; Wang, S. Multi-domain multi-task rehearsal for lifelong learning. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 8819–8827. [\[CrossRef\]](#)

19. Rui, X.; Li, Z.; Cao, Y.; Li, Z.; Song, W. DILRS: Domain-incremental learning for semantic segmentation in multi-source remote sensing data. *Remote Sens.* **2023**, *15*, 2541. [\[CrossRef\]](#)
20. Yu, L.; Twardowski, B.; Liu, X.; Herranz, L.; Wang, K.; Cheng, Y.; Jui, S.; van de Weijer, J. Semantic drift compensation for class-incremental learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2020; pp. 6982–6991.
21. Michieli, U.; Zanuttigh, P. Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 11–15 June 2021; pp. 1114–1124.
22. Gan, Y.; Bai, Y.; Lou, Y.; Ma, X.; Zhang, R.; Shi, N.; Luo, L. Decorate the newcomers: Visual domain prompt for continual test time adaptation. *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 7595–7603. [\[CrossRef\]](#)
23. Yang, S.; Wu, J.; Liu, J.; Li, X.; Zhang, Q.; Pan, M.; Zhang, S. Exploring sparse visual prompt for cross-domain semantic segmentation. *arXiv* **2023**, arXiv:2303.09792.
24. Vu, T.H.; Jain, H.; Bucher, M.; Cord, M.; Pérez, P. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2019; pp. 2517–2526.
25. Li, W.; Gao, H.; Su, Y.; Momanyi, B.M. Unsupervised domain adaptation for remote sensing semantic segmentation with transformer. *Remote Sens.* **2022**, *14*, 4942. [\[CrossRef\]](#)
26. Yang, Z.; Yan, Z.; Diao, W.; Ma, Y.; Li, X.; Sun, X. Active Bidirectional Self-Training Network for Cross-Domain Segmentation in Remote-Sensing Images. *Remote Sens.* **2024**, *16*, 2507. [\[CrossRef\]](#)
27. Chaudhry, A.; Rohrbach, M.; Elhoseiny, M.; Ajanthan, T.; Dokania, P.; Torr, P.; Ranzato, M. Continual learning with tiny episodic memories. In Proceedings of the Workshop on Multi-Task and Lifelong Reinforcement Learning, Long Beach, CA, USA, 15 June 2019.
28. Van de Ven, G.M.; Tuytelaars, T.; Tolias, A.S. Three types of incremental learning. *Nat. Mach. Intell.* **2022**, *4*, 1185–1197. [\[CrossRef\]](#)
29. Wang, L.; Zhang, X.; Su, H.; Zhu, J. A comprehensive survey of continual learning: Theory, method and application. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 5362–5383. [\[CrossRef\]](#)
30. Douillard, A.; Chen, Y.; Dapogny, A.; Cord, M. Plop: Learning without forgetting for continual semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2021; pp. 4040–4050.
31. Zhou, D.W.; Cai, Z.W.; Ye, H.J.; Zhan, D.C.; Liu, Z. Revisiting class-incremental learning with pre-trained models: Generalizability and adaptivity are all you need. *Int. J. Comput. Vis.* **2024**, *133*, 1012–1032. [\[CrossRef\]](#)
32. Petit, G.; Soumm, M.; Feillet, E.; Popescu, A.; Delezoide, B.; Picard, D.; Hudelot, C. An analysis of initial training strategies for exemplar-free class-incremental learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2024; pp. 1837–1847.
33. Wallingford, M.; Li, H.; Achille, A.; Ravichandran, A.; Fowlkes, C.; Bhotika, R.; Soatto, S. Task adaptive parameter sharing for multi-task learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2022; pp. 7561–7570.
34. Roy, K.; Simon, C.; Moghadam, P.; Harandi, M. Subspace distillation for continual learning. *Neural Netw.* **2023**, *167*, 65–79. [\[CrossRef\]](#)
35. Li, Y.; Xu, W.; Wang, H.; Qi, Y.; Guo, J.; Li, R. Personalized federated domain-incremental learning based on adaptive knowledge matching. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 127–144.
36. Chen, T.; Guo, Y.; Hao, S.; Hong, R. Leaving None Behind: Data-Free Domain Incremental Learning for Major Depressive Disorder Detection. *IEEE Trans. Affect. Comput.* **2024**, *16*, 758–770. [\[CrossRef\]](#)
37. De Lange, M.; Aljundi, R.; Masana, M.; Parisot, S.; Jia, X.; Leonardis, A.; Slabaugh, G.; Tuytelaars, T. A continual learning survey: Defying forgetting in classification tasks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3366–3385. [\[CrossRef\]](#)
38. Zhu, Z.; Wang, P.; Diao, W.; Yang, J.; Wang, H.; Sun, X. Few-shot incremental learning with continual prototype calibration for remote sensing image fine-grained classification. *ISPRS J. Photogramm. Remote Sens.* **2023**, *196*, 210–227. [\[CrossRef\]](#)
39. Liu, W.; Nie, X.; Zhang, B.; Sun, X. Incremental learning with open-set recognition for remote sensing image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5622916. [\[CrossRef\]](#)
40. Shao, Y.; Li, Y.; Han, X. Contrastive Dual-Pool Feature Adaption for Domain Incremental Remote Sensing Scene Classification. *Remote Sens.* **2025**, *17*, 308. [\[CrossRef\]](#)
41. Lu, X.; Sun, X.; Diao, W.; Feng, Y.; Wang, P.; Fu, K. LIL: Lightweight incremental learning approach through feature transfer for remote sensing image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5611320. [\[CrossRef\]](#)
42. Ye, D.; Peng, J.; Li, H.; Bruzzone, L. Better Memorization, Better Recall: A Lifelong Learning Framework for Remote Sensing Image Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5626814. [\[CrossRef\]](#)
43. Li, Z.; Hoiem, D. Learning without forgetting. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 2935–2947. [\[CrossRef\]](#)

44. Aljundi, R.; Babiloni, F.; Elhoseiny, M.; Rohrbach, M.; Tuytelaars, T. Memory aware synapses: Learning what (not) to forget. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 139–154.
45. Zhao, W.; Peng, R.; Wang, Q.; Cheng, C.; Emery, W.J.; Zhang, L. Life-long learning with continual spectral-spatial feature distillation for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5544214. [[CrossRef](#)]
46. Yu, C.; Zhao, X.; Gong, B.; Hu, Y.; Song, M.; Yu, H.; Chang, C.I. Distillation-constrained prototype representation network for hyperspectral image incremental classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5507414. [[CrossRef](#)]
47. Tasar, O.; Tarabalka, Y.; Alliez, P. Incremental learning for semantic segmentation of large-scale remote sensing data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3524–3537. [[CrossRef](#)]
48. Arnaudo, E.; Cermelli, F.; Tavera, A.; Rossi, C.; Caputo, B. A contrastive distillation approach for incremental semantic segmentation in aerial images. In *Proceedings of the International Conference on Image Analysis and Processing*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 742–754.
49. Klingner, M.; Bär, A.; Donn, P.; Fingscheidt, T. Class-incremental learning for semantic segmentation re-using neither old data nor old labels. In Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 20–23 September 2020; pp. 1–8.
50. Wang, Q.; Wu, Y.; Yang, L.; Zuo, W.; Hu, Q. Layer-Specific Knowledge Distillation for Class Incremental Semantic Segmentation. *IEEE Trans. Image Process.* **2024**, *33*, 1977–1989. [[CrossRef](#)] [[PubMed](#)]
51. Cermelli, F.; Mancini, M.; Bulò, S.R.; Ricci, E.; Caputo, B. Modeling the background for incremental learning in semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2020; pp. 9233–9242.
52. Li, J.; Sun, X.; Diao, W.; Wang, P.; Feng, Y.; Lu, X.; Xu, G. Class-incremental learning network for small objects enhancing of semantic segmentation in aerial imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5612920. [[CrossRef](#)]
53. Qiu, Y.; Shen, Y.; Sun, Z.; Zheng, Y.; Chang, X.; Zheng, W.; Wang, R. SATS: Self-attention transfer for continual semantic segmentation. *Pattern Recognit.* **2023**, *138*, 109383. [[CrossRef](#)]
54. Cha, S.; Kim, B.; Yoo, Y.; Moon, T. SSUL: Semantic segmentation with unknown label for exemplar-based class-incremental learning. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 10919–10930.
55. Xie, J.; Pan, B.; Xu, X.; Shi, Z. MiSSNet: Memory-inspired Semantic Segmentation Augmentation Network for Class-Incremental Learning in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5607913. [[CrossRef](#)]
56. Maracani, A.; Michieli, U.; Toldo, M.; Zanuttigh, P. Recall: Replay-based continual learning in semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 7026–7035.
57. Lee, Y.; Lee, D.; Kwak, T.; Kim, Y. ER-PASS: Experience Replay with Performance-Aware Submodular Sampling for Domain-Incremental Learning in Remote Sensing. *Remote Sens.* **2025**, *17*, 3233. [[CrossRef](#)]
58. Rebuffi, S.A.; Kolesnikov, A.; Sperl, G.; Lampert, C.H. icarl: Incremental classifier and representation learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2001–2010.
59. Li, J.; Diao, W.; Lu, X.; Wang, P.; Zhang, Y.; Yang, Z.; Xu, G.; Sun, X. Sil-land: Segmentation incremental learning in aerial imagery via label number distribution consistency. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5628820. [[CrossRef](#)]
60. Zhang, Z.; Gao, G.; Jiao, J.; Liu, C.H.; Wei, Y. CoinSeg: Contrast inter- and intra- class representations for incremental segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 843–853.
61. Yuan, B.; Zhao, D.; Shi, Z. Learning at a Glance: Towards Interpretable Data-Limited Continual Semantic Segmentation via Semantic-Invariance Modelling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 7909–7923. [[CrossRef](#)]
62. Zhang, C.B.; Xiao, J.W.; Liu, X.; Chen, Y.C.; Cheng, M.M. Representation compensation networks for continual semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 7053–7064.
63. Kalb, T.; Ahuja, N.; Zhou, J.; Beyerer, J. Effects of architectures on continual semantic segmentation. In Proceedings of the 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, AK, USA, 4–7 June 2023; pp. 1–8.
64. Yang, Z.; Li, L.; Zhang, J.; Wang, T.; Sun, Y.; Yan, C. Domain Shared and Specific Prompt Learning for Incremental Monocular Depth Estimation. In Proceedings of the 32nd ACM International Conference on Multimedia, Melbourne, Australia, 28 October–1 November 2024; pp. 8306–8315.
65. Feng, W.; Ju, L.; Wang, L.; Song, K.; Zhao, X.; Ge, Z. Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment. *Proc. Aaai Conf. Artif. Intell.* **2023**, *37*, 623–631. [[CrossRef](#)]
66. Wang, Z.; Luo, Y.; Huang, D.; Ge, N.; Lu, J. Category-adaptive domain adaptation for semantic segmentation. In Proceedings of the ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 3773–3777.

67. Kanakis, M.; Bruggemann, D.; Saha, S.; Georgoulis, S.; Obukhov, A.; Van Gool, L. Reparameterizing convolutions for incremental multi-task learning without task interference. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 689–707.
68. Rottensteiner, F.; Sohn, G.; Jung, J.; Gerke, M.; Baillard, C.; Benitez, S.; Bretkopf, U. The ISPRS benchmark on urban object classification and 3D building reconstruction. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *1*, 293–298. [[CrossRef](#)]
69. Michieli, U.; Zanuttigh, P. Incremental learning techniques for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, Montreal, BC, Canada, 11–17 October 2021.
70. Ye, Y.; Wei, W.; Zhang, L.; Ding, C.; Zhang, Y. Domain consistency learning for continual test-time adaptation in image semantic segmentation. *Pattern Recognit.* **2025**, *165*, 111585. [[CrossRef](#)]
71. Jodelet, Q.; Liu, X.; Phua, Y.J.; Murata, T. Class-incremental learning using diffusion model for distillation and replay. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Paris, France, 2–6 October 2023; pp. 3425–3433.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.