

Article

Cross-Domain Landslide Mapping in Remote Sensing Images Based on Unsupervised Domain Adaptation Framework

Jing Yang^{1,2}, Mingtao Ding^{1,3,*}, Wubiao Huang⁴, Qiang Xue^{5,6}, Ying Dong^{5,6}, Bo Chen^{1,3}, Lulu Peng^{1,3}, Fuling Zhang⁷ and Zhenhong Li^{1,3}

¹ College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China; 2020126052@chd.edu.cn (J.Y.); bo.chen@chd.edu.cn (B.C.); 2023126045@chd.edu.cn (L.P.); zhenhong.li@chd.edu.cn (Z.L.)

² Engineer School, Qinghai Institute of Technology, Xining 810016, China

³ Big Data Center for Geosciences and Satellites, Xi'an 710054, China

⁴ School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China; huangwubiao@whu.edu.cn

⁵ Xi'an Center of China Geological Survey, Xi'an 710119, China; xueqiang_79@163.com (Q.X.); dongy329@163.com (Y.D.)

⁶ Observation and Research Station of Geo-Hazards in Loess Plateau, Ministry of Natural Resources, Xi'an 710119, China

⁷ Qinghai Remote Sensing Center for Natural Resources, Xining 810000, China; linger0110@126.com

* Correspondence: mingtaoding@chd.edu.cn

Highlights

What are the main findings?

- The proposed LandsDANet demonstrates outstanding performance in cross-domain landslides mapping based on unsupervised domain adaptation.
- The implementation of Rare Class Sampling, Wallis filter and Contrastive loss collectively enhance the model's performance in cross-domain feature extraction and learning.

What are the implications of the main findings?

- The proposed method plays a critical role in rapid cross-domain landslide identification and disaster emergency response.
- The proposed method reduces the model's reliance on pixel-level annotations for training, improving the efficiency of landslide mapping.

Abstract

Rapid and accurate acquisition of landslide inventories is essential for effective disaster relief. Deep learning-based pixel-wise semantic segmentation of remote sensing imagery has greatly advanced in landslide mapping. However, the heavy dependance on extensive annotated labels and sensitivity to domain shifts severely constrain the model performance in unseen domains, leading to poor generalization. To address these limitations, we propose LandsDANet, an innovative unsupervised domain adaptation framework for cross-domain landslide identification. Firstly, adversarial learning is employed to reduce the data distribution discrepancies between the source and target domains, thereby achieving output space alignment. The improved SegFormer serves as the segmentation network, incorporating hierarchical Transformer blocks and an attention mechanism to enhance feature representation capabilities. Secondly, to alleviate inter-domain radiometric discrepancies and attain image-level alignment, a Wallis filter is utilized to perform image style transformation. Considering the class imbalance present in the landslide dataset, a Rare Class Sampling strategy is introduced to mitigate bias towards common classes and strengthen



Academic Editor: Deodato Tapete

Received: 12 December 2025

Revised: 11 January 2026

Accepted: 12 January 2026

Published: 15 January 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](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

the learning of the rare landslide class. Finally, a contrastive loss is adopted to further optimize and enhance the model's ability to delineate fine-grained class boundaries. The proposed model is validated on the Potsdam and Vaihingen benchmark datasets, followed by validation in two landslide scenarios induced by earthquakes and rainfall to evaluate its adaptability across different disaster domains. Compared to the source-only model, LandsDANet achieved improvements in IoU of 27.04% and 35.73% in two cross-domain landslide disaster recognition tasks, respectively. This performance not only showcases its outstanding capabilities but also underscores its robust potential to meet the demands for rapid response.

Keywords: landslide inventory mapping; optical remote sensing imagery; cross-domain; unsupervised domain adaptation; deep learning; semantic segmentation; transformer

1. Introduction

Large-scale seismic events and intense precipitation frequently induce cascading geohazards, particularly regional-scale landslides, which cause substantial casualties and considerable economic losses [1]. Consequently, the rapid and accurate identification of regions impacted by landslides is essential for risk assessment and emergency response [2]. However, the effectiveness of traditional landslide mapping methods is frequently hindered by a lack of prior knowledge regarding the specific affected areas, as well as insufficient time to acquire high-quality training samples necessary for constructing advanced identification models. This situation presents a significant challenge to achieving landslide inventory maps promptly. Developing a robust identification model that minimizes dependence on extensive prior knowledge and high-quality training samples is of paramount importance.

Benefiting from the powerful feature extraction and representation capabilities, deep learning techniques, especially Deep Convolutional Neural Networks (DCNN) and Vision Transformers (ViT) [3] have achieved great success in remote sensing (RS) image interpretation. Various semantic segmentation models, including ResU-Net [4], SegFormer [5], and Swin Transformer [6] have been widely utilized for the identification of newly affected landslide areas. For instance, Dong et al. [7] introduced an improved MobileNetV3 as encoder for detecting co-seismic landslides. Lv et al. [8] proposed ShapeFormer, which introduces pyramid vision transformer to capture both spectral and spatial features of landslides. Wu et al. [9] proposed a hybrid model known as SCDUNet++, combining the Transformer with CNN, and specifically designed multi-channel datasets for co-seismic landslide detection. In the process of landslide feature extraction, DCNNs excel at capturing local features within images, and ViTs possess advantages in capturing long-range contextual information through their robust sequence-to-sequence modeling capabilities, thereby improving the performance of landslide feature detection across varying scales. generalize to the different appearance and context of a new target domain.

Nevertheless, several significant challenges arise when applying DCNNs and ViTs, such as MobileNetV3 and SCDUNet++, to cross-domain landslide mapping. Firstly, deep learning-based approaches heavily depend on comprehensive pixel-level annotations for training, which are both labor-intensive and require considerable time to acquire. This requirement creates bottlenecks for practical applications, particularly in areas that have been recently affected. Secondly, deep learning models exhibit sensitivity to the data distribution between the training set (referred to as source domain data) and test set (known as target domain data). It is essential for the source and target domains to satisfy the hypothesis of being independently and equivalently distributed. However, significant

discrepancies in data distribution often arise due to variations in imaging conditions, such as observational illumination, viewing geometry, spatial or spectral resolution, and sensor characteristics. More importantly in the context of landslide mapping, fundamental differences in landslide characteristics are influenced by distinct triggering mechanisms (i.e., earthquakes or rainfall), as well as geographical environment and terrain factors. As illustrated in Figure 1, these factors lead to high intra-class variance in landslides, characterized by diverse appearances within the landslide class, and low inter-class variance, where landslides visually resemble bare lands, roads or bedrocks but belong to different semantic classes. Furthermore, there are significant variations in object scales [10]. Consequently, unlike the well-defined edges and consistent textures of objects like buildings or cars, landslides present a uniquely challenging target for semantic segmentation due to their amorphous shapes, spectral ambiguity with background elements, and terrain dependency. The inherent complexity of landslides, along with domain shifts, presents a great obstacle to robust and rapid landslide identification across diverse geographical environments.

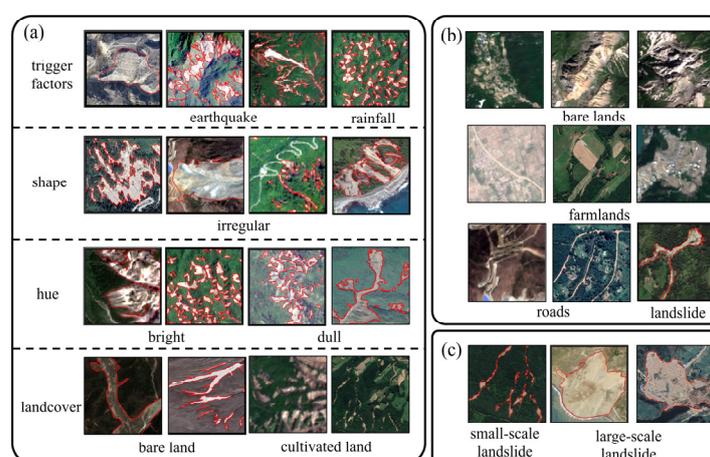


Figure 1. Challenges of cross-domain landslide identification using RS imagery. (a) significant variance within classes. (b) minimal variance between classes. (c) considerable differences in the scales of objects. The boundary of the landslide is indicated by the red outline.

To improve the generalization capability of deep learning models in newly landslide-affected areas, previous research has integrated training data from multiple study areas. This approach increases the diversity of training samples, thereby improving model performance in unseen domain to some extent. For instance, Fang et al. [11] proposed a co-seismic landslide dataset that integrates multi-source RS images, representing diverse geological backgrounds from nine events worldwide, to facilitate the mapping of newly affected landslide areas. Yang et al. [4] proposed multidomain models that combine annotated landslide samples from two different domains to enhance the model's generalization ability. However, these studies adopt a strategy that attempts to rigidly memorize the landslide features of different domains during the training process rather than truly learning to eliminate domain-specific disparities. Consequently, when the trained model encounters an entirely new region with data characteristics absent from the existing mixed training set, its generalization performance tends to decline sharply. Furthermore, this method suffers from high costs and limited scalability, as it requires retraining the model from scratch whenever new study areas are introduced.

Unsupervised domain adaptation (UDA), a specific instance of transfer learning, is designed to tackle the challenges posed by domain shifts between source and target domains. It enables models to learn domain-invariant knowledge, thereby enhancing performance on unlabeled target domains. Recently, there has been significant interest in adversarial

learning [12] within the field of computer vision, which has proven effective for UDA in semantic segmentation. This approach provides a powerful framework for learning deep representations without relying on extensive annotated datasets. To address the challenges arising from domain discrepancies and the limited availability of annotated samples in cross-domain landslide mapping, we aim to construct a model for domain adaptation that utilizes adversarial learning to identify features invariant across different domains.

Therefore, we propose LandsDANet, a comprehensive framework for domain adaptation that utilizes a generative adversarial network (GAN) to identify landslides across different domains in RS images. Firstly, adversarial learning is employed to reduce the disparities in data distribution between the source and target domains in the output space. Based on GAN, the proposed model features an enhanced transformer-based segmentation model as the generator to predict outputs, along with a discriminator that identifies whether the outputs are derived from the source or the target domains. By leveraging adversarial loss, the segmentation model seeks to deceive the discriminator, aiming to produce distributions that closely resemble each other between the predictions of the source and target in the output space. Secondly, a Wallis filter is utilized to convert source domain images into those resembling the target, effectively aligning the source and target domains at the image level. Considering the class imbalance in the landslide dataset, a Rare Class Sampling (RCS) strategy is also adopted to reinforce the learning of rare classes. By increasing the sampling frequency of images that contain rare classes, the network is able to learn them more robustly and mitigate confirmation bias. Thirdly, a contrastive loss is adopted to enhance the representations of features. The International Society for Photogrammetry and Remote Sensing (ISPRS) in Potsdam and Vaihingen, covering urban, suburban and natural scenes [13] serves as a benchmark for the evaluation of the general domain adaptation capability of proposed LandsDANet framework. Subsequently, the proposed model was applied to cross-domain landslide identification task, including the Meizhou rainfall-triggered landslide dataset and the Taiwan co-seismic landslide dataset. The workflow of this study is shown in Figure 2. The main contributions of this study can be summarized as follows:

- (1) We propose LandsDANet, a novel UDA framework specifically designed for the rapid identification of landslides across diverse domains. This network employs adversarial learning to align data distributions at the output level and supports end-to-end training, facilitated by flexible implementation schemes.
- (2) The Wallis filter is introduced for image style transfer, effectively minimizing low-level domain discrepancies and achieving image-level alignment. A sampling strategy specifically focusing on landslide rare category, is implemented to prevent the model from exhibiting excessive bias towards more common classes during the adaptation process.
- (3) The contrastive loss function is introduced to reduce the representation distance among features belonging to the same category while maximizing the distance between features from different categories, enhancing the model's robustness in category discrimination and improving the cross-domain segmentation results.
- (4) To evaluate the model's cross-domain adaptability, experiments were performed using the ISPRS benchmark dataset alongside landslide datasets created by authors. Our proposed model exhibits exceptional performance relative to leading UDA methods, and comprehensive ablation studies corroborate the significance of each module's contribution. This framework substantially enhances the accuracy of cross-domain landslide identification while effectively mitigating domain gaps caused by distinct triggering mechanisms and imaging conditions.

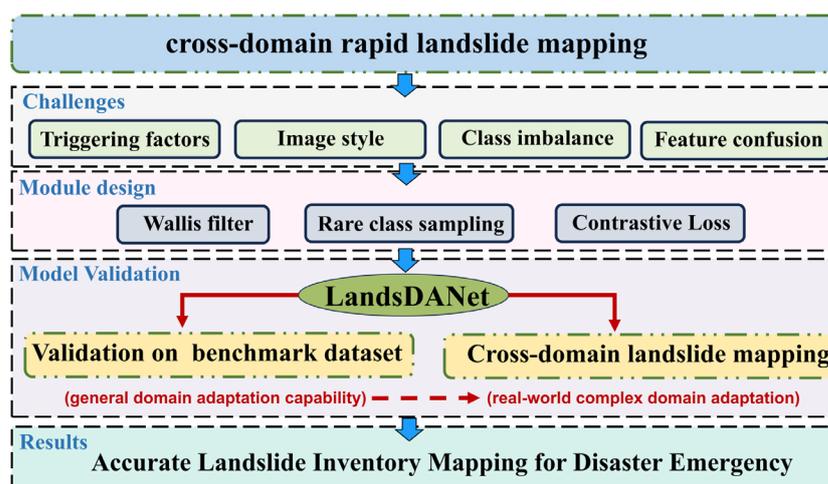


Figure 2. Workflow of this study.

2. Related Work

2.1. UDA Semantic Segmentation in the Computer Vision Field

UDA seeks to transfer knowledge acquired from a labeled source domain to target domains that are heterogeneous, unlabeled, and not previously encountered, thus effectively mitigating the domain gaps [14]. Traditional UDA primarily minimizes distribution discrepancy between source and target domains by employing discrepancy measures such as Maximum Mean Discrepancy (MMD) [15]. Recently, a variety of GAN-based UDA techniques have been specifically tailored for tasks involving semantic segmentation. These methods emphasize the alignment of distributions between domains at image, feature, and output levels. Image-level alignment UDA methods integrate image-to-image translation techniques to transform source images into target-like images. Examples of typical image-to-image translation models include CyCADA [16], CycleGAN [17] and ColorMapGAN [18]. However, these models always generate low-quality style-transferred images, especially in scenarios where the domain shift is complex. Feature-level alignment is implemented to reduce data distributions in a high-dimensional latent space [19]. For instance, Hoffman et al. [20] employ full convolutional domain adversarial learning for both global and class-specific adaptation in semantic segmentation task. However, high-dimensional features frequently contain noise, which can interfere with the training of the discriminator. Tsai et al. [21] noted that the output space contains sufficient spatial and local information, proposing a domain adaptation approach for the output space. Building on the work of [21], Vu et al. [22] introduced entropy maps, with losses grounded in the entropy of segmentation predictions. Their model achieved strong performance on two challenging datasets and performed favorably compared to leading methods in the field. Subsequently, several new variants based on category information [23] and fine-grained alignment [24] have been proposed. Wang et al. [24] utilized a domain discriminator with fine-grained capabilities for class-level alignment, aiming to preserve the internal semantic structure across different domains. However, class-level alignment is insufficient to handle intra-class variance and lacks contextual consistency constraints.

As another choice of UDA, self-training (ST) [25] utilizes a pretrained model to generate pseudo-labels tailored to target domains. Following this, it engages in additional training rounds using the labeled data from the source domain, aiming to progressively align the feature spaces between the source and target domains. However, the domain shift complicates the generation of pseudo-labels in the target domain, potentially accumulating errors that adversely affect model convergence. To mitigate the negative impact of pseudo-labels, several studies have utilized confidence thresholds or reweighting tech-

niques. Zhang and Yang [26] proposed to dynamically adjust confidence thresholds based on prediction variance to rectify the pseudo-labels for UDA. Additionally, Zhang et al. [27] developed ProDA, which utilized prototypical pseudo-label denoising and target structure learning to enhance the reliability of pseudo-labels and boost model performance. However, for categories exhibiting diversity, their feature distributions may present multiple characteristics, leading to inaccuracies in prototypes.

The aforementioned methods are typically assessed using both synthetic datasets and real-world datasets, such as GTA5 and Cityscapes. However, these datasets exhibit substantial discrepancies when compared to remote sensing datasets. These discrepancies underscore the critical necessity of developing UDA models specifically tailored to remote sensing data, particularly in UDA scenarios involving class-imbalanced landslide datasets.

2.2. UDA Semantic Segmentation in the Remote Sensing Field

With the widespread application of UDA methods in natural image semantic segmentation tasks, an increasing number of studies have begun to concentrate on the domain shift issues associated with RS images. Typical applications of UDA in RS include cross-domain land cover classification, road and building extraction, and post-disaster damage assessment, among others. In the field of land cover classification, the Potsdam and Vaihingen datasets, provided by the ISPRS, have been extensively adopted as standard benchmarks. Zhu et al. [28] proposed MemoryAdaptNet which incorporates an output space adversarial learning strategy alongside a memory module designed to retain invariant domain-level information. This model is utilized for land cover classification using the Potsdam and Vaihingen datasets. Zhang et al. [29] developed a stagewise domain adaptation network for road extraction, which incorporates a GAN-based method for both interdomain and intradomain feature alignment. Peng et al. [30] proposed a network aimed at cross-domain building extraction, which integrates adversarial learning with a ST training strategy. Chen et al. [31] created a method for cross-domain building extraction by decoupling style-semantic features, utilizing contrastive learning to ensure cross-domain semantic consistency while employing a memory mechanism to store style features as well as domain-invariant semantic features. Overall, these studies effectively solve cross-scene domain shift problems and lay a foundation for domain adaptation in RS field.

While UDA has demonstrated promising results in land cover classification and urban extraction, its application to cross-domain landslide identification, characterized by heterogeneous datasets and complex contextual dependencies, remains significantly under explored. Xu et al. [10] introduced ADANet, which employs a multi-level output space adaptation approach for cross-domain landslide recognition, with prediction accuracy depends on a fine-tuning module. Li et al. [32] and Yu et al. [33] proposed image-to-image translation and style transfer for unsupervised cross-domain landslide identification, limited by low-quality images generation. Zhang et al. [34] proposed a prototype-guided and domain-aware approach for progressive representation learning in multi-target landslide domain adaptation mapping, facing challenges in initial prototype determining and requiring numerous iterations. Li et al. [35] proposed PluTsa, which is a method for unsupervised multitemporal landslide detection based on progressive label upgradation and cross-temporal style adaptation. Wei et al. [36] introduced a universal adapter module, that can achieve efficient and highly accurate transferable landslide mapping by fine-tuning only a limited number of parameters. Wang et al. [37] proposed a weakly supervised landslide extraction method that guides the Segment Anything Model in generating precise landslide segmentation masks by automatically producing point and box prompts from Class Activation Maps. This approach achieves high-accuracy boundary segmentation using only image-level labels. While these methods represent significant advancements in

cross-domain landslide mapping, they exhibit inherent constraints, including dependence on specific data modalities, algorithmic complexity, and sensitivity to hyperparameters, particularly when rapid, single-date, cross-domain adaptation is required. To address these gaps, we propose LandsDANet, a novel UDA framework specifically designed for cross-disaster landslide identification from single-data imagery. It integrates a lightweight Wallis filter for robust image-level alignment, a RCS strategy to handle extreme imbalance, and an adversarial learning scheme with contrastive loss for feature discrimination and domain-invariant representation learning. The design aims to overcome the limitations of prior methods while ensuring practicality for rapid disaster mapping.

3. Methodology

3.1. LandsDANet Architecture

The proposed LandsDANet framework is designed to address the specific cross-domain landslide identification challenges. Specifically, the RCS strategy tackles the extreme class imbalance; the Wallis filter mitigates low-level spectral ambiguity; and the combination of adversarial learning and contrastive loss targets the high-level semantic shift and boundary ambiguity arising from irregular shapes and terrain dependency. The architecture of LandsDANet is presented in Figure 3. This network comprises a segmentation network, denoted as G , and a discriminator, referred to as D . The dataset for the source domain D_S , consists of source images x_s and ground truth y_s . In contrast, the dataset for the target domain D_T , consists solely target images x_T . D_S and D_T can be landslide datasets from various locations, imaging conditions and different times. The objective is to train G to achieve accurate predictions on x_T by transferring knowledge learned from D_S to D_T and learning domain-invariant features. As presented in Figure 3, considering the low proportion of landslide pixels within the landslide sample, RCS is introduced to increase the sampling frequency of images containing rare classes from the source domain. Then, to alleviate the radiometric discrepancy of D_S and D_T , a Wallis filter is employed to transform x_s into target-stylized images x_{ST} . Both x_{ST} and x_T are then fed into G , a semantic segmentation network based on an encoder–decoder structure, known as ISegFormer. At the decoder, segmentation results p_{ST} and p_T are generated, respectively. On the one hand, segmentation losses \mathcal{L}_{CE} and \mathcal{L}_{contr} were calculated with p_{ST} and y_s for the purpose of optimizing G . On the other hand, to align the distribution of p_{ST} and p_T in the output space, these two predictions were input into D , facilitating the gradient propagation from D to G via adversarial loss \mathcal{L}_{adv} . This process promotes G to produce segmentation distributions in the target domain that closely resemble those of the source prediction. Through continuous iterations and learning, G progressively learns the domain-invariant features. During the testing phase, G can be directly utilized to produce prediction maps for the target domain. The implementation of the proposed LandsDANet will be made publicly available at <https://github.com/HuangWBill/LandsDANet> (accessed on 10 January 2026).

3.2. ISegFormer Semantic Segmentation Network and Discriminator Network

Recently, researchers have introduced models based on ViT for the task of semantic segmentation demonstrating superior performance compared to DCNNs. In this work, the segmentation backbone G is an improved version of the SegFormer architecture [5], which we refer to as ISegFormer. Specifically, we preserve its hierarchical Transformer encoder while integrate a Convolutional Block Attention Module (CBAM) [38] after each stage to enhance feature representation, as shown in Figure 4a. This modification enables the model to concentrate more effectively on discriminative features crucial for landslide identification. In the encoder part, multi-scale features are directly generated through overlapping Patch Embedding and transformer blocks [9]. The Overlap Patch Embedding firstly converts the

original image into multiple overlapping patches using convolution operations, thereby preserving spatial continuity between adjacent regions. These patches are then flattened into an embedding sequence and processed by the transformer block, which primarily comprises Self-Attention [3], a Mix Feedforward Network (Mix-FFN), and Overlap Patch Merging [9]. The Self-Attention calculates the attention weights of each patch relative to all other patches in the sequence, enabling the model to capture dependencies over long distances and understand global contextual relationships across the entire image. It was defined as:

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d_{head}}}\right)V \tag{1}$$

where the query, key, and value matrices, designated as Q , K and V , possess identical dimensions $H \times W \times C$. $\text{SoftMax}(\cdot)$ is the activation function and d_{head} is the dimension of each attention head.

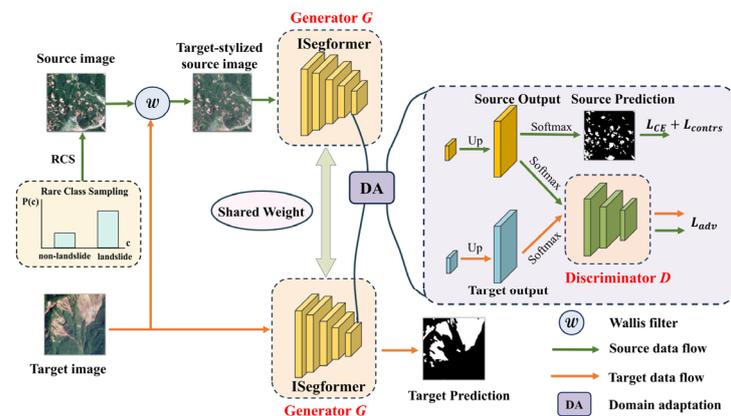


Figure 3. The proposed LandsDANet network. Here, RCS denotes Rare Class Sampling, \mathcal{L}_{CE} , \mathcal{L}_{contr} and \mathcal{L}_{adv} denote cross-entropy loss (CE), contrastive loss, and adversarial loss, respectively.

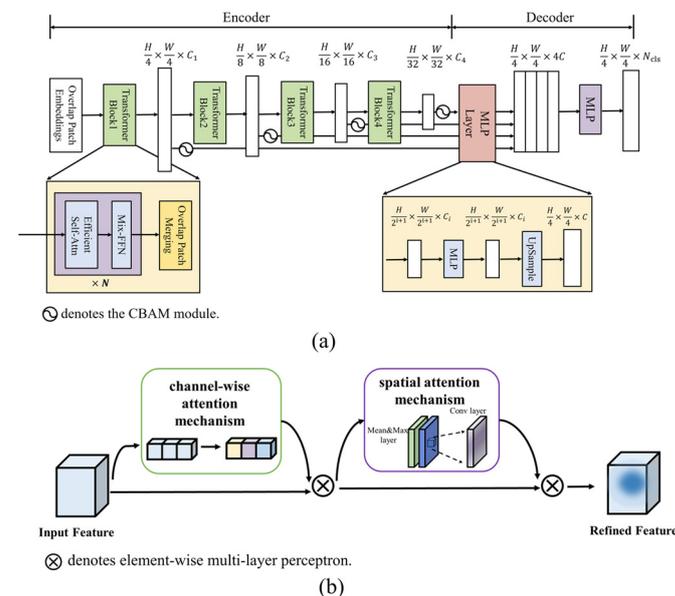


Figure 4. (a) The architecture of ISegFormer. (b) The architecture of the CBAM module.

Mix-FFN encodes positional information via depth-wise convolutional layers, thus obviating the need for positional encodings. Additionally, it incorporates nonlinear transformations to capture complex relationships in input data. It can be defined as:

$$x_{out} = \text{MLP}(\text{GELU}(\text{Conv}_{3 \times 3}(\text{MLP}(x_{in})))) + x_{in} \tag{2}$$

where x_{in} denotes the feature derived from Self-Attention, x_{out} represents output feature, $MLP(\cdot)$ denotes the multilayer perceptron, $Conv_{3 \times 3}$ indicates convolution operation with a kernel size of 3×3 and $GELU(\cdot)$ is Gaussian Error Linear Unit activation function.

Overlap patch merging is responsible for aggregating information from overlapping patches and producing a coarser representation. Specifically, when given an input image of dimensions $H \times W \times 3$, four successive Transformer blocks yield multi-scale feature maps X_i with a size of $\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C_i$, where $i = 1, 2, 3, 4$.

The ALL-MLP decoder consists of four primarily stages: (1) unifying the channel dimensions of the feature maps produced by the encoder, (2) performing upsampling, (3) concatenating and merging feature maps, and (4) making segmentation predictions. This architecture integrates both local and global multi-level attention features while preserving high representational capabilities and reducing computational complexity.

The CBAM is integrated after the output of each stage in the Transformer block to capture the complex, high-level semantic differences in landslide morphologies influenced by varying geological conditions. As presented in Figure 4b, CBAM comprises both channel and spatial attention mechanisms. For a feature map $F \in \mathbb{R}^{H \times W \times C}$, the attention mechanism can be outlined as follows:

$$F' = M_C(F) \otimes F \quad (3)$$

$$F'' = M_S(F') \otimes F' \quad (4)$$

where \otimes denotes element-wise multi-layer perceptron (MLP) and $F'' \in \mathbb{R}^{H \times W \times C}$ is the final output. $M_C(F)$ is a channel-wise attention mechanism that utilizes average-pooling and max-pooling to focus on which channels are significant for landslide detection. $M_S(F)$ is a spatial attention mechanism that focuses on where the informative parts for landslide detection are located.

The discriminator adopts fully convolutional layers. To generate high-dimensional feature representations, four successive convolutional layers are implemented, each featuring a kernel size of 4×4 and a stride of 2, followed by a leaky *Relu*. Subsequently, a convolutional layer along with an up-sampling operation is added to facilitate channel dimension reduction and spatial size recovery. Lastly, a sigmoid layer is employed to limit the output values within the range of 0 to 1.

3.3. Wallis Filter and Rare Class Sampling

Significant color discrepancies between source and target RS images arise from variations in acquisition time, geographic location and illumination conditions. While GAN-based methods have found success in the style transfer of natural images, such as CycleGAN [17], their application to RS images still presents challenges. The complex semantic information and high intra-class heterogeneity inherent in RS images often lead to training instability or even model collapse in adversarial frameworks. The Wallis filter is a deterministic and parameter-light radiometric normalization algorithm. It performs statistical adjustments for each pixel based on mean and variance value, achieving stable inter-image transformation while preserving the structural and semantic integrity of objects. Additionally, it effectively reduces model complexity and enables rapid image-level alignment. It is computed that for a 512×512 image patches with a batch size of 2, the Wallis filter inference time is approximately 0.4 milliseconds. Thus, for landslides triggered by earthquakes and rainfall, we adopted a Wallis filter-based transformation to convert x_S into x_{ST} , thereby

minimizing the discrepancies in low-level radiation caused by different surface materials, such as bare rock and wet soil. The formula is written as follows:

$$x_{ST} = \omega \left[\sigma(x_T) \left(\frac{x_S - \mu(x_S)}{\sigma(x_S)} \right) + \mu(x_T) \right] + (1 - \omega)x_S \quad (5)$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ represent the mean and variance operations, respectively, and ω controls the extent of the transformation between images. In our experiments, we set $\omega = 1$ to apply the full transformation.

Landslide detection datasets frequently exhibit class imbalance, also known as long-tail distribution, which could lead to confirmation bias towards common classes during the training process. This means that the model predominantly relies on dominant categories for its predictions, making it challenging to effectively learn rare classes due to a lack of sufficient samples. Experimental observations indicate that the starting time and final performance of rare categories can vary significantly depending on the random seed. If relevant samples arise in the later phases of training, the model may struggle to learn these categories effectively due to pre-existing established biases that have already been established. To address this issue, we introduced RCS to adjust the sampling strategy of the source domain data, giving priority to sampling images that contains rare categories. It is particularly effective for increasing the detection rate of small landslides. The frequency f_c for each class c within the source domain dataset is determined by counting the number of pixels assigned to class c :

$$f_c = \frac{\sum_{i=1}^{N_S} \sum_{j=1}^{H \times W} [y_S^{i,j,c}]}{N_S \cdot H \cdot W} \quad (6)$$

The sampling probability $P(c)$ for class c is established in accordance with f_c :

$$P(c) = \frac{e^{1-f_c/T}}{\sum_{c'=1}^C e^{1-f_{c'}/T}} \quad (7)$$

where temperature T serve as a paramant that influences the smoothness of the distribution. The objective of T is to maximize the resampling count for the pixel categories with the lowest frequency. As a result, samples with rare categories will have a higher sampling probability through the RCS.

3.4. Loss Functions

Given a $x_{ST} \in X_S$ and its corresponding label $y_S \in Y_S$ from D_S , and a $x_T \in X_T$ from D_T , we forward both x_{ST} and x_T to G (i.e., ISegFormer) and acquire the predicted result p_{ST} and p_T . Firstly, p_{ST} and y_S are utilized to compute the segmentation loss in a supervised manner, with the objective of reducing the discrepancy between the predicted results and the actual ground truth for the source domain landslide mapping task. To facilitate the model training, we utilized Cross-Entropy (CE) loss function as the segmentation loss:

$$\mathcal{L}_{CE}(X_S, Y_S) = -E_{(x_S, y_S) \sim (X_S, Y_S)} \sum_{i=1}^{C_K} y_S^{(i)} \log G(x_S)^{(i)} \quad (8)$$

However, the CE loss modifies pixel-wise predictions independently, ignoring the inter-pixel relationships and failing to provide direct supervision for the learned representations [39]. Therefore, we introduced a contrastive loss to regularize the structural embedding space and optimize hard sample sampling, which enhances the discriminative

ability between landslide features and spectrally similar non-landslide features of G . The formula is:

$$\mathcal{L}_{Contrs} = \frac{1}{|P_i|} \sum_{i^+ \in P_i} -\log \frac{\exp(i \cdot i^+ / \tau)}{\exp(i \cdot i^+ / \tau) + \sum_{i^- \in N_i} \exp(i \cdot i^- / \tau)} \quad (9)$$

where P_i and N_i refer to the collections of pixels embedding for the positive and negative samples corresponding to pixel i , respectively. τ is a hyperparameter and we set $\tau = 0.1$ in this experiment through cross-validation. i^+ is a positive sample feature vector and i^- is a negative sample feature vector. As shown in Equation (9), the primary objective of contrast loss is to learn a structured embedding space by bringing similar class pixel samples closer together while distancing pixel samples from different classes.

Finally, the supervised segmentation loss is defined as:

$$\mathcal{L}_{seg} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{Con} \quad (10)$$

where λ denotes the weight of pixel-wise contrastive loss and is set as 0.5.

In addition, we transmit p_T to D to produce an adversarial loss aimed at optimizing G . The definition of the adversarial loss is as follows:

$$\mathcal{L}_{adv}(X_T) = -E_{x_t \sim P_T(x)} [\log D(G(x_T))] \quad (11)$$

where $\mathcal{L}_{adv}(X_T)$ represents the adversarial loss and $P_T(x)$ denotes the distribution of target domains.

To generate domain-invariant features, p_{S_T} and p_T are fed into D , yielding a scalar output. This scalar reflects the probability of the input derived from the source domain instead of the target domain. In this context, label 0 is assigned to the target training example, while label 1 is assigned to the source training example. This singular scalar is utilized to then employed to compute cross-entropy loss \mathcal{L}_D and can be expressed as:

$$\mathcal{L}_D(X_S, X_T) = -E_{x \sim P_S(x)} [\log D(G(X_S))] - E_{x \sim P_T(x)} [\log(1 - D(G(X_T)))] \quad (12)$$

4. Datasets

To evaluate the proposed LandsDANet framework, we employ a validation approach involving both general benchmark datasets and real-world landslide datasets. This design is motivated by the need to first assess the general domain adaptation capability of the model core components using well-annotated, controlled datasets, followed by an assessing of their performance on the more complex and application-specific task of cross-disaster landslide identification.

4.1. General Remote Sensing Datasets

The Potsdam and Vaihingen benchmark datasets are two publicly accessible semantic segmentation datasets in RS. The Potsdam dataset consists of 38 images with a spatial resolution of 5 cm. The Vaihingen dataset comprises 33 images with a resolution of 9 cm. Figure 5 shows the percentage of pixels associated with each class in relation to the total pixel count within the standard datasets. Our analysis reveals that the car and clutter categories exhibit lower percentages in both datasets compared to other classes, which could be classified as rare classes. Figure 6 presents examples from the benchmark datasets. We observed huge domain discrepancies arising from diverse geolocation, imaging sensors, and illumination conditions. This highlights the necessity of utilizing UDA for the semantic segmentation of RS images.

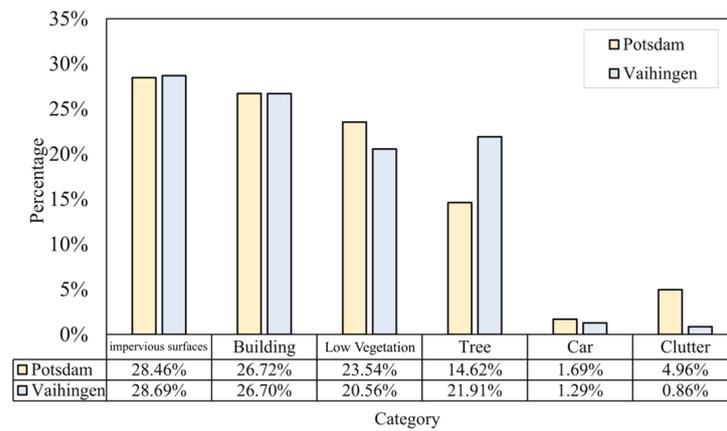


Figure 5. Pixel percentage of each class to the total number of pixels.

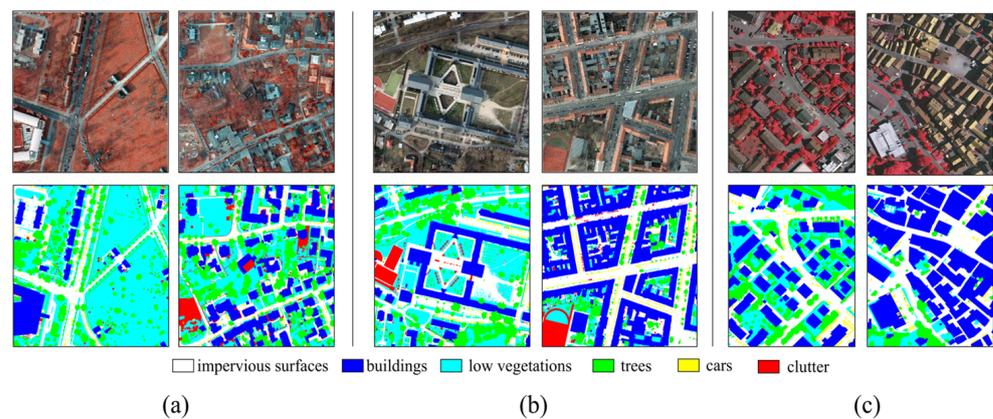


Figure 6. Examples of images and ground truths of the ISPRS benchmark datasets. (a) Potsdam dataset with [IR, R, G] channels. (b) Potsdam dataset with [R, G, B] channels. (c) Vaihingen dataset with [IR, R, G] channels.

4.2. Landslide Inventory

Two scenarios of landslide occurrence, including co-seismic and rainfall-triggered landslides are selected, and the locations of the study areas are presented in Figure 7a. The details are illustrated as follows.

4.2.1. 2024 Hualien Earthquake

On 3 April 2024, a Mw 7.4 earthquake struck Hualien County in Taiwan, China [40]. The study area is located within an active plate collision zone, characterized predominantly by hard, coherent metamorphic and sedimentary rocks. The strong ground shaking triggered rock-dominated failures, including rock avalanches and slides. These processes expose extensive areas of fresh bedrock, resulting in landslides that appear light-colored, exhibit high spectral reflectance, maintain relatively uniform shapes, and are distributed along steep topography. It is estimated that the earthquake triggered 1243 landslides, covering a total area approximately 21.5 km², resulting in 18 fatalities and 1555 injuries [41]. The post-event optical images, obtained from 17 April to 29 April 2024, was shown in Figure 7c, which with a spatial resolution of 3 m and from PlanetScope satellites. Through the interpretation of RS images, we found that most of the landslides were small and distributed on both sides of the river.

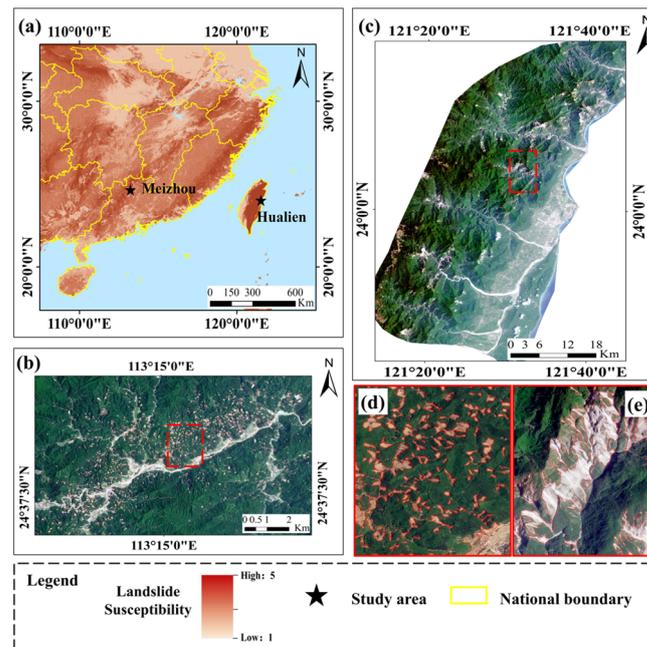


Figure 7. (a) Locations of the Meizhou and Hualien areas. (b,c) Post-event images of the Meizhou and Hualien areas captured by PlanetScope. (d,e) Enlarged localized imagery of landslides occurring in Meizhou and Hualien regions. (d) Following the landslide depicted in the diagram, bedrock was exposed. (e) Following the landslide depicted in the diagram, loose soil was revealed. The red-bordered area in (d,e) is an enlarged view of the red dashed area in (b,c).

4.2.2. 2024 Meizhou Rainfall

From April to June 2024, Guangdong Province in Eastern China experienced multiple rounds of rainfall, particularly affecting regions such as Meizhou, Shantou, and Zhaoqing, which were hit by torrential to exceptionally heavy rain. These cities are situated in the subtropical hilly region of South China, characterized by deeply weathered granite and covered by thick residual soils. The prolonged intense rainfall triggered numerous shallow soil slides and debris flows. These landslides, which develop within the thick weathered mantle, incorporate moist soil and vegetation debris, leading to darker spectral tones, more diffuse boundaries, and often flow-like morphologies. It is reported that the torrential rain in Meizhou City led to 5 fatalities, 15 individuals reported missing, and 13 people stranded due to mountain floods and landslides. To evaluate the landslide damage caused by the heavy rainfall, we obtained PlanetScope satellite images captured on 14 May 2024, with the study area shown in Figure 7b.

As shown in Figure 7d,e and Figure 8, significant differences in landslide morphology, scale, hue, and geographic environment were observed between the two datasets, attributed to varying triggering factors. Earthquake-induced landslides typically exhibit steeper and more concentrated distributions, with greater exposure of bedrock. In contrast, rainfall-induced landslides are generally more dispersed and occur in soil-covered regions, frequently accompanied by debris flows. These distinctions constitute the fundamental source of domain shifts, underscoring the critical importance of conducting landslide UDA experiments. The proposed LandsDANet framework, with its integrated design addressing both low-level radiometric alignment and high-level semantic adaptation, is specifically engineered to bridge such geophysically based domain gaps.

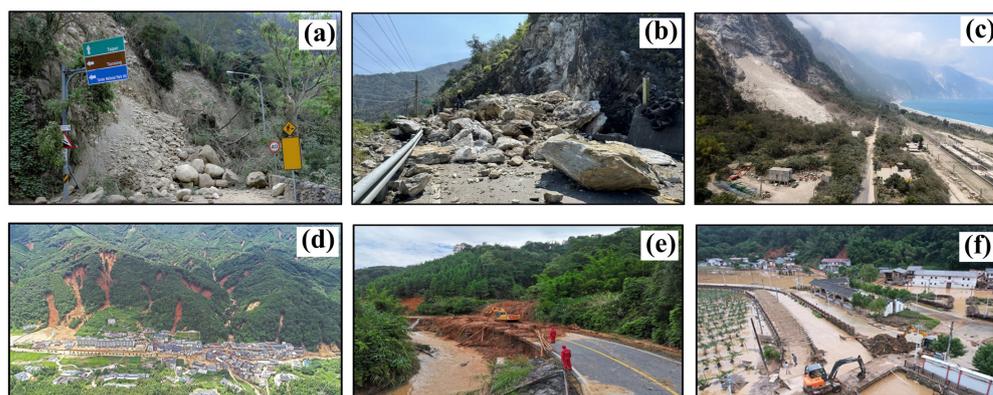


Figure 8. Field photographs of landslides from the two study cases. (a–c) Landslides triggered by the 2024 Hualien, Taiwan earthquake. (d–f) Landslides triggered by the 2024 Meizhou, Guangdong torrential rainfall. Image sources: (a) Reuters ¹; (b) Sohu ²; (c) CNN ³; (d) Guangzhou Daily ⁴; (e) BTime ⁵; (f) Hubei Daily ⁶. ¹ <https://www.reuters.com/world/asia-pacific/taiwan-searches-18-still-missing-after-earthquake-2024-04-05/> (accessed on 5 April 2024); ² <https://www.sohu.com/picture/769114627> (accessed on 4 April 2024); ³ https://edition.cnn.com/asia/live-news/taiwan-earthquake-hualien-04-04-23-hnk-intl#h_9cf5276cdbc5a30e473fb7ae79f0de52 (accessed on 6 April 2024); ⁴ https://gzdaily.dayoo.com/pc/html/2024-06/21/content_870_861439.htm (accessed on 21 June 2024); ⁵ <https://item.btime.com/454lhjtaoof9iuonpaiuqor1cj9> (accessed on 19 June 2024); ⁶ http://v.cnhubei.com/content/2024-06/22/content_18078674.html (accessed on 23 June 2024). Note: These images are used for academic illustrative purposes only; copyright resides with the original publishers.

The satellite RS images utilized in this study were taken from the PlanetScope constellation of Planet Labs Company (Planet Labs Federal, Inc., San Francisco, CA, USA). The images used in this research are provided by reference [11], utilizing red, green and blue spectra bands. These acquired images have undergone atmospheric and geometric corrections. The dataset preprocessing phases primarily involve the generation of binary masks and the cropping of sample batches. Firstly, the binary landslide masks were generated through visual interpretation of the post-event PlanetScope imagery by experts. Then, the interpreted landslide polygons were rasterized into a binary mask containing geospatial information using ArcMap (v10.2), where ‘1’ represents landslide and ‘0’ represents non-landslide. Subsequently, the imagery and the binary mask were batch clipped into 512×512 pixels for model training. In the Meizhou dataset, we collected 710 samples for training and 304 samples for testing. In the Taiwan dataset, we obtained 698 samples for training and 196 samples for testing.

The generated inventories are interpreted from post-event imagery alone, without direct comparison to pre-event images. While interpreters prioritized fresh scars likely associated with the triggering event, the inventories may include pre-existing landslides or other spectrally similar features that were visible after the event. Thus, the study mainly focuses on rapid mapping of post-event landslides from two different triggering events.

4.3. Validation Metric and Network Training

To assess the performance of the model, evaluation metrics such as F1-score (F1), intersection over union (IoU), mean IoU, precision and recall are employed. The formulae for each metric are presented below:

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (16)$$

$$Precision = \frac{TP}{TP + FP} \quad (17)$$

where TP represents true positives, FP represents false positives, TN denotes true negatives, and FN indicates false negatives.

The LandsDANet is implemented on a system with an NVIDIA GeForce RTX 4090 Laptop GPU (16 GB VRAM). A batch size of 2 is selected and 512×512 patches are processed for 60,000 iterations. Throughout the warm-up phase, which lasts until iteration t_{warm} , the learning rate at iteration t is calculated as $l_t = l_{init} \cdot t / t_{warm}$, where t_{warm} is 1500, and l_{init} is the starting learning rate. The training procedures for the proposed LandsDANet are detailed in Algorithm 1. We conducted bidirectional UDA experiments utilizing both the benchmark datasets and landslide datasets, respectively. Additionally, data augmentation is employed to enhance data diversity through random horizontal and vertical flips.

Algorithm 1 Training Process of the Proposed Method

Input: Source samples $(x_s, y_s) \in D_S$, target samples $x_T \in D_T$, Training epochs E , Steps per epoch T

Output: The predicted label p_{ST} of source image, the predicted label p_T of target image, optimized segmentation model parameters θ_S and optimized discriminator parameters θ_D

for epoch = 1 to E **do**

for step = 1 to T **do**

 Sample (x_s, y_s) from D_S , sample x_T from D_T

 Calculate x_{ST} with x_s and x_T based on Equation (5)

 Freeze θ_D

 Calculate \mathcal{L}_{seg} with (x_{ST}, y_s) based on Equation (10)

 Calculate \mathcal{L}_{adv} with x_T based on Equation (11)

 Update weights θ_S

 Unfreeze θ_D

 Freeze θ_S

 Calculate \mathcal{L}_D with p_{ST} and p_T respectively based on Equation (12)

 Update weights θ_D

 Unfreeze θ_S

end for

end for

5. Result and Accuracy Assessment

To ensure that LandsDANet has a reliable foundation for handling complex, data-scarce cross-domain landslide detection tasks, we first validate its core design on publicly available benchmark datasets. These datasets exhibit distinct domain shifts, including variations in geography, sensors, and resolution, and containing rare class. This makes them particularly suitable for specifically evaluating the effectiveness of modules such as Wallis filter, rare class sampling, and contrastive loss in addressing common challenges, including radiometric differences, class imbalance, and feature discriminability. Secondly, we compared LandsDANet with the current state-of-the-art (SOTA) UDA methods designed for common object semantic segmentation. This comparison reflects LandsDANet's

competitiveness in addressing the fundamental issues of domain alignment. Moreover, the subsequent superior performance of LandsDANet in cross-disaster landslide recognition tasks can be more confidently attributed to our landslide-specific components, which tackle challenges beyond general domain shifts, such as extreme class imbalance and spectral ambiguity of bare ground. Consequently, the experiments in Section 5.1 serve as a crucial preliminary validation, aimed at ensuring the robustness of the design before applying it to the more challenging landslide detection tasks in Section 5.2.

5.1. Validation on General Domain Shifts

We performed comparative experiments with a deep model that does not incorporate DA (ISegFormer) and four SOTA UDA methods: AdaptSegNet [21], AdvEnt [22], ADANet [10] and MemoryAdaptNet [28]. Meanwhile, we established three UDA tasks, designated as P2V_S, V2P_S and P2V_D. P and V refer to the initial letters of the Potsdam and Vaihingen datasets, respectively. S and D indicate whether the spectral composition of the source and target domains is identical or distinct. The spectral composition of Potsdam dataset is [IR, R, G] and [R, G, B], and the Vaihingen dataset is [IR, R, G].

5.1.1. P2V_S Task: Adaptation Under Geographic and Illumination Variations

The P2V_S task primarily aims to assess the effectiveness of LandsDANet in mitigating domain discrepancies caused by geolocation, spatial resolution, and illumination. As shown in Table 1, the proposed LandsDANet achieved the highest IoU and F1-score in each category, with mIoU and mean F1-score values of 56.90% and 70.62%, respectively. In contrast, the source-only model suffers from domain shift, resulting in the lowest mIoU and mean F1-score values of 43.55% and 57.82%, respectively. The AdaptSegNet and AdvEnt yield slightly better results than the source-only model, with mean F1-scores of 59.26% and 59.99%, and mIoU of 44.01% and 45.66%, respectively. ADANet attained the second-highest IoU value of 23.61% and F1-score of 38.20% in the clutter category. It is noteworthy that the recognition accuracy for this category is the lowest among all domain-adaptive methods, primarily due to the low pixel proportions of the clutter category. Compared to MemoryAdaptNet, which utilizes invariant domain-level prototype memory, LandsDANet presents an improvement of 9.55% and 7.95% in both mIoU and F1-score. The IoU for the rare class “car” presents a marked improvement to 49.94%, significantly higher than all compared methods. This can be attributed to the frequent sampling of the RCS technique. Furthermore, this result establishes a foundation for future approaches aimed at tackling the recognition challenges associated with extremely rare categories, such as “landslide”. Figure 9 presents representative results from LandsDANet and other methods in the P2V_S task, showing that the source-only model produces confusing outputs with considerable noise. Although the performance of the four SOTA methods improves following domain adaptation, shortcomings remain in terms of unclear boundaries and semantic confusion. There is still some noise at the boundary in the output from AdaptSegNet. The ADANet misclassifies low vegetations and trees (as presented in the seventh row of Figure 9), while the MemoryAdaptNet exhibits classification errors among clutter, buildings, and impervious surfaces (as presented in the seventh row of Figure 9). In contrast, our LandsDANet demonstrates superior cross-domain adaptation capabilities and enhanced performance due to its unique structural design. In particular, it maintains a clear distinction between semantically similar classes, such as low vegetation and trees.

Table 1. Domain adaptation results for all methods in the P2V_S task.

Method	Impervious Surfaces		Building		Low Vegetation		Tree		Car		Clutter		mIoU	mF1
	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1		
Source only	55.39	71.29	66.42	79.83	45.10	62.17	56.92	72.55	27.94	43.68	9.53	17.41	43.55	57.82
AdaptSegNet [21]	58.70	73.98	66.49	79.87	39.84	56.98	49.01	65.78	32.44	48.98	17.61	29.94	44.01	59.26
AdvEnt [22]	66.94	80.2	69.08	81.72	44.39	61.49	49.65	66.36	31.09	47.44	12.82	22.73	45.66	59.99
ADANet [10]	67.23	80.4	65.77	79.35	42.15	59.31	47.22	64.15	35.8	52.73	23.61	38.2	46.96	62.36
MemoryAdaptNet [28]	64.95	78.75	67.17	80.36	46.71	63.68	46.18	63.19	37.54	54.59	21.52	35.42	47.35	62.67
LandsDANet (ours)	75.36	85.95	80.26	89.05	55.43	71.32	56.51	72.21	49.94	66.62	23.9	38.57	56.9	70.62

Bold values represent the best values.

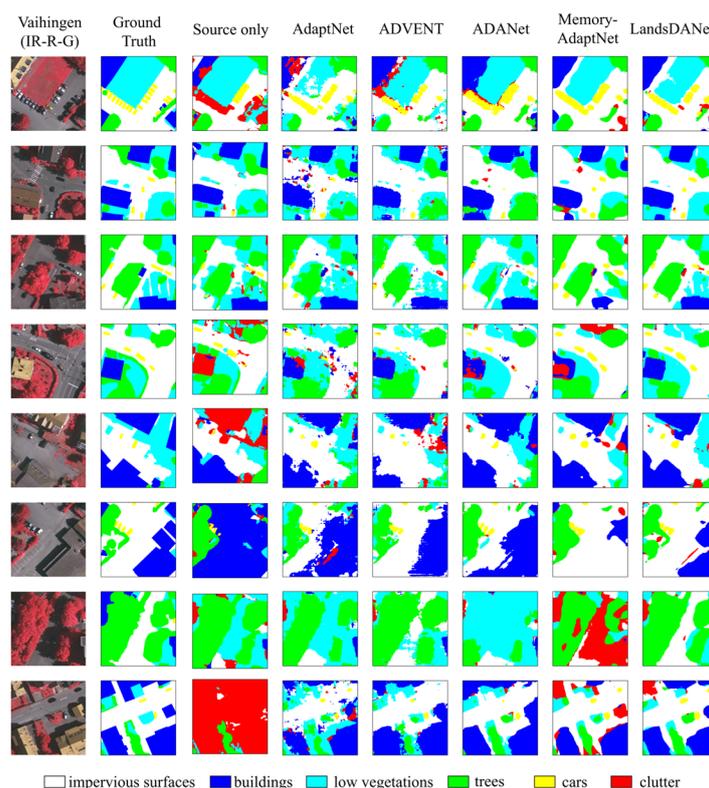


Figure 9. Examples of visualized results for all methods in the P2V_S task.

5.1.2. V2P_S Task: Adaptation Under Spatial Resolution Variation

The V2P_S task adapts the source domain Vaihingen to the target domain Potsdam with the same channel composition. We found that the lower spatial resolution and fewer training samples in the Vaihingen dataset further increases the challenges of UDA. Notably, under these conditions, LandsDANet demonstrates significant superiority, achieving a mIoU of 53.87% and a mean F1-score of 66.69% as shown in Table 2. This underscores its capability to perform effective domain adaptation even with limited source data. The source-only model and ADANet demonstrate comparable performance, both achieving mIoU and F1-score values of 43.67% and 56.37% for the source-only model, and 43.48% and 58.47% for ADANet, respectively. AdaptSegNet shows the lowest performance, underperforming the source-only model by 5.32% and 3.02% in mIoU and F1-score, respectively. Among all methods evaluated, the prediction accuracy for the clutter category is highest when using ADANet, with an IoU of 15.08% and F1-score of 26.20%. This may be attributed to the fact that the heterogeneity of the “clutter” within the class increases the difficulty of image-level alignment. Meanwhile, ADANet’s fine-tuning approach may facilitate a more flexible adaptation to the “clutter” class. As shown in Figure 10, LandsDANet generates

the highest-quality semantic segmentation results, characterized by greater accuracy in semantic details and smoother boundaries. Specifically, the categories of low vegetation and buildings are often confused by the SOTA methods, primarily due to the limited quantity of training samples available in the Vaihingen dataset, which offers less guidance regarding segmentation loss. This limitation results in a poor alignment across domains. In other words, this further demonstrates the capability of our approach to achieve cross-domain adaptation even with fewer samples.

Table 2. Domain adaptation results for all methods in the V2P_S task.

Method	Impervious Surfaces		Building		Low Vegetation		Tree		Car		Clutter		mIoU	mF1
	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1		
Source only	64.93	78.73	60.83	75.64	44.08	61.18	20.41	33.91	67.53	80.62	4.23	8.12	43.67	56.37
AdaptSegNet [21]	51.77	68.22	53.92	70.07	42.02	59.18	32.38	48.92	42.34	59.48	7.67	14.24	38.35	53.35
AdvEnt [22]	58.41	73.74	58.11	73.51	39.18	56.31	31.66	48.09	47.34	64.26	5.45	10.34	40.03	54.37
ADANet [10]	61.68	76.30	57.53	73.04	32.47	49.02	35.3	52.18	58.8	74.06	15.08	26.20	43.48	58.47
MemoryAdaptNet [28]	60.11	75.08	59.64	74.72	29.14	45.14	43.44	60.57	49.71	66.41	8.63	15.89	41.78	56.30
LandsDANet (ours)	66.77	80.07	76.71	86.82	49.88	66.56	49.25	65.99	70.83	82.93	9.76	17.78	53.87	66.69

Bold values represent the best values.

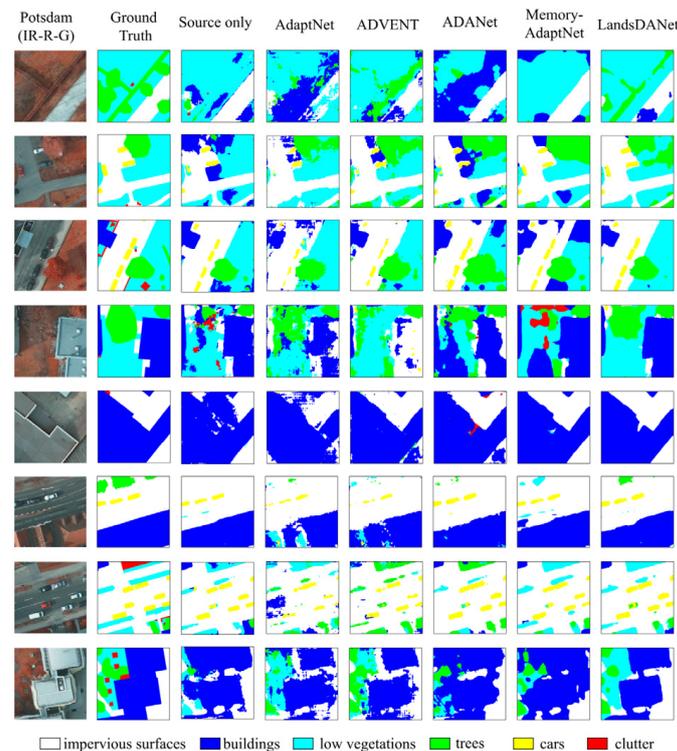


Figure 10. Examples of visualized results for all methods in the V2P_S task.

5.1.3. P2V_D Task: Adaptation Under Channel Composition Variation

This scenario utilizes the Potsdam dataset as the source domain and the Vaihingen dataset as the target domain to evaluate the domain adaptation under discrepancies in spectral band composition, spatial resolution, and geographical location. Table 3 presents the quantitative evaluation results, where the mIoU and F1-score of the source-only model are 29.21% and 40.67%, respectively. The best-performing model, LandsDANet, achieves mIoU and F1-score values of 48.5% and 63.02%, respectively, reflecting improvements of 19.29% and 22.35% compared to the source-only model. AdvEnt and ADANet demonstrate superior predictive performance in the low vegetation and clutter categories, respectively.

The results obtained from MemoryAdaptNet are unsatisfactory, possibly attributing to the memory module failing to effectively capture historical pseudo-invariant features in the presence of a significant domain disparity. The mIoU and mean F1-score of MemoryAdaptNet are 40.66% and 55.64%, respectively, indicating decreases of 7.84% and 7.38% compared to the proposed LandsDANet model. Figure 11 visualized the comparative results. The source-only model suffers from serious misidentification issues, with buildings, low vegetation and trees being mistakenly identified as clutter. The performance of segmentation is enhanced to different extents following domain adaptation. Nevertheless, AdaptSegNet and AdvEnt still exhibit some noisy segmentation, and the boundaries remain unsmooth. MemoryAdaptNet shows considerable misclassification between clutter and impervious surfaces. In contrast, our LandsDANet achieves the best segmentation performance, with an improvement in boundary smoothness.

Table 3. Domain adaptation results for all methods in the P2V_D task.

Method	Impervious Surfaces		Building		Low vegetation		Tree		Car		Clutter		mIoU	mF1	
	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1			
Source only	57.62	73.11	60.91	75.71	11.11	19.99	15.49	26.82	27.34	42.94	2.81	5.46	29.21	40.67	
State-of-the-art methods	AdaptSegNet [21]	60.04	75.03	68.16	81.07	32.29	48.82	32.98	49.6	42.34	59.48	14.17	24.83	43.05	57.76
	AdvEnt [22]	63.92	77.99	74.06	85.10	35.42	52.32	42.02	59.17	47.34	64.26	14.06	24.65	45.94	60.40
	ADANet [10]	62.33	76.79	67.53	80.62	34.48	51.28	40.64	57.79	58.8	74.06	25.4	40.51	45.25	60.88
	MemoryAdaptNet [28]	60.4	75.31	63.53	77.7	28.21	44.01	39.66	56.80	49.71	66.41	14.50	25.32	40.66	55.64
LandsDANet (ours)	60.89	75.70	82.12	90.18	27.61	43.27	45.68	62.71	70.83	82.93	24.63	39.53	48.50	63.02	

Bold values represent the best values.

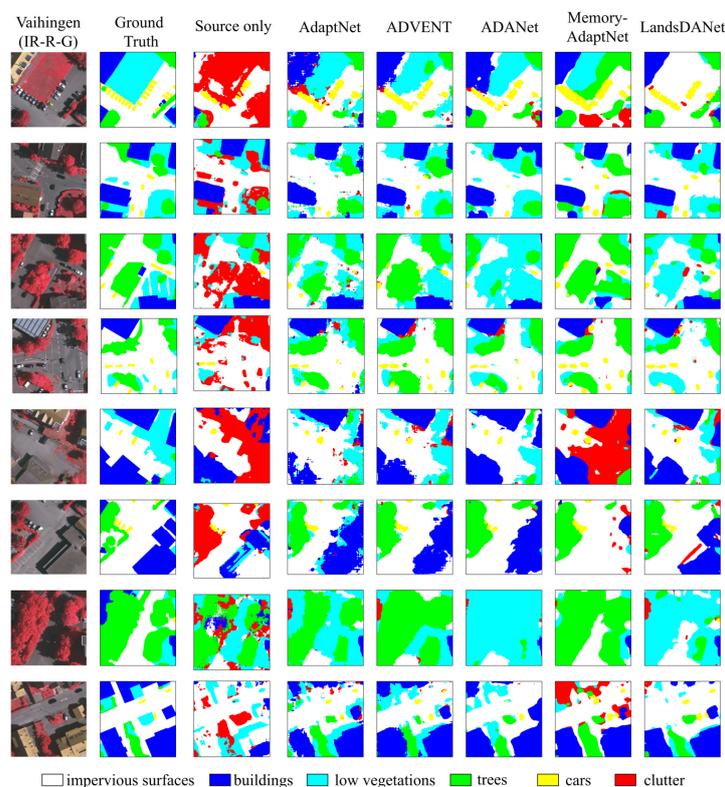


Figure 11. Examples of visualized results for all methods in the P2V_D task.

The above experiments demonstrate that the proposed LandsDANet framework achieves competitive performance on standard benchmarks, effectively handling domain shifts induced by variations in geography, sensor types, spatial resolution, and illuminations. Its core modules effectively address fundamental challenges in UDA, including style

transfer, class imbalance, and feature distribution matching. Building upon this validated foundation, we further applied the framework to the more challenging task of cross-domain landslide detection. Unlike the controlled settings of benchmark datasets, real-world landslide mapping involves additional complexities stemming from variations in triggering mechanisms, geological environments, and terrain characteristics. These factors result in deeper and more semantically diverse domain shifts. The subsequent experiments are designed to evaluate whether this method can effectively address geophysical challenges, thereby assessing its practical utility for rapid disaster response.

5.2. Cross-Domain Landslide Detection Experiments

Having established the domain adaptation capability of LandsDANet on general datasets, we now focus on its primary application of cross-domain landslide identification under real-world, geophysical complex conditions. We designed two UDA experiments utilizing the Meizhou (M) dataset, which represents rainfall-induced shallow soil landslides, and the Taiwan (T) dataset, which represents earthquake-induced rock collapses. The experiments involve two transfer directions: from Meizhou to Taiwan (M2T) and from Taiwan to Meizhou (T2M).

5.2.1. M2T Task

The validation results related to various cross-domain landslide mapping methods are summarized in Table 4. Compared to other models, LandsDANet exhibited superior and robust performance, demonstrating precision, recall, IoU, and F1-score values of 68.47%, 69.85%, 52.85% and 69.15%, respectively. Applying the source-only training model directly to the target domain leads to inaccurate predictions due to the model's inadequate generalization capability, causing the IoU, recall, precision and F1-score only reach 17.12%, 19.43%, 59.01% and 29.23%, respectively. The performance was improved to a certain extent after adaptation. AdaptSegNet and AdvEnt obtain the near-accuracies through global adversarial learning, but the results are unsatisfactory, with IoU for the landslide category only reaching 22.24% and 25.98%, respectively. MemoryAdaptNet outperforms other SOTA methods, but is still overshadowed by the LandsDANet, with IoU, recall, precision and F1-score being 9.41%, 11.66%, 5.31% and 8.58% lower than LandsDANet, respectively. These results reflected that UDA methods could reduce the domain gap to some extent, and our LandsDANet is the most suitable for cross-domain landslide identification. Figure 12 shows the landslide identification results for the Taiwan area based on LandsDANet. From the four sub-regions (marked as a–d) at Figure 12, we found that most landslides were accurately identified. However, some of the cultivated land and bare soil in the coastal area were mistakenly identified as landslides due to its similar spectral characteristics to landslides. Figure 13 shows the qualitative examples of landslide detection results obtained by the source-only model, SOTA methods and LandsDANet for the M2T case, respectively. Similarly to the analysis of Table 4, based on the source-only model, the majority of landslides were missed and the model had almost no generalization ability. The implementation of domain adaptation methods led to a steady reduction in both missed and falsely identified instances (as shown in Figure 13d–h). Our proposed method generates prediction maps that are both more accurate and comprehensive than those of comparative methods.

Table 4. Quantitative evaluation results of the M2T task.

Method	Landslide		mIoU	mF1	Recall	Precision	
	IoU	F1					
Source only	17.12	29.23	56.35	63.50	19.43	59.01	
State-of-the-art methods	AdaptSegNet [21]	22.24	36.38	59.02	67.12	25.77	61.86
	AdvEnt [22]	25.98	41.24	60.91	69.56	31.13	61.08
	ADANet [10]	37.49	54.53	66.41	76.07	59.25	50.51
	MemoryAdaptNet [28] (2023)	43.44	60.57	69.93	79.38	58.19	63.16
LandsDANet (ours)	52.85	69.15	74.95	83.83	69.85	68.47	

Bold values represent the best values.

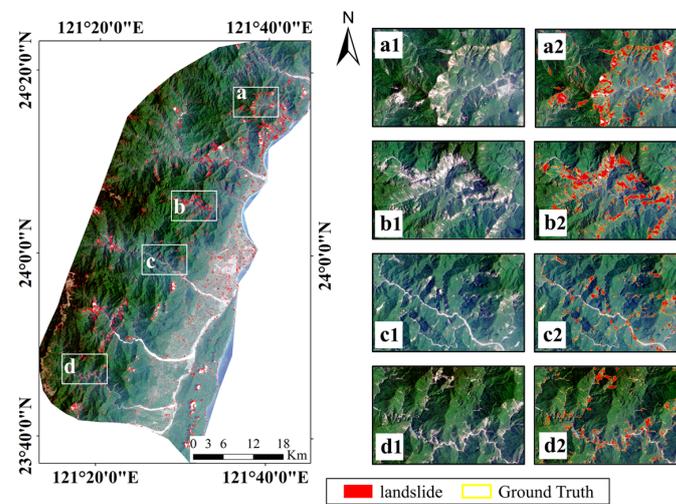


Figure 12. Landslide identification result based on LandsDANet in the four sub-regions (a–d) of Taiwan area. (a1) post-event optical image; (a2) predicted results; (b1) post-event optical image; (b2) predicted results; (c1) post-event optical image; (c2) predicted results; (d1) post-event optical image; (d2) predicted results. The boundaries of the ground truth are depicted in the subgraph with yellow outlines.

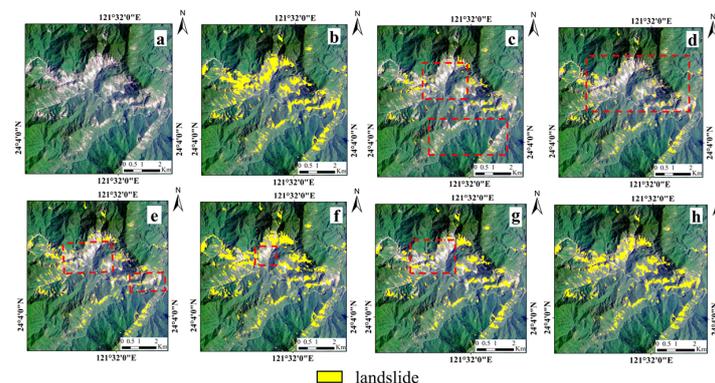


Figure 13. Examples of landslide identification maps produced by various methods for the M2T case: (a) post-event image, (b) ground truth, (c) source only, (d) AdaptSegNet, (e) AdvEnt, (f) ADANet, (g) MemoryAdaptNet, and (h) LandsDANet. Red boxes show differences in results.

5.2.2. T2M Task

In this scenario, the Taiwan area, where the landslides were triggered by earthquake, serves as the source domain, and the Meizhou region, where the landslides were induced

by rainfall, serves as the target domain. Table 5 and Figure 14. show LandsDANet achieves the highest performance with recall, precision, IoU, and F1-score values of 65.34%, 60.40%, 45.75% and 62.77%, respectively. The source-only model exhibited the poorest performance, with IoU, recall, precision and F1-score values of 18.71%, 20.23%, 71.39% and 31.54%. As illustrated in Figure 15c, most landslides were not identified. Compared with AdaptSegNet which with the output space alignment, LandsDANet shows the improvement of 19.19% on IoU, 25.56% on recall, 12.57% on precision and 20.80% on F1-score. Compared with MemoryAdaptNet which has the domain-level prototype memory, LandsDANet presents the improvement of 7.53%, 6.25%, and 8.83% on IoU, recall and precision, respectively. ADANet, as the best-performing model among the four SOTA methods, yields result that are slightly inferior to those of LandsDANet we proposed. Figure 13d–g show that some landslides were not identified, and lakes were misidentified as landslides. These performance differences confirm that LandsDANet has better domain adaptation ability in the field of the landslide detection task.

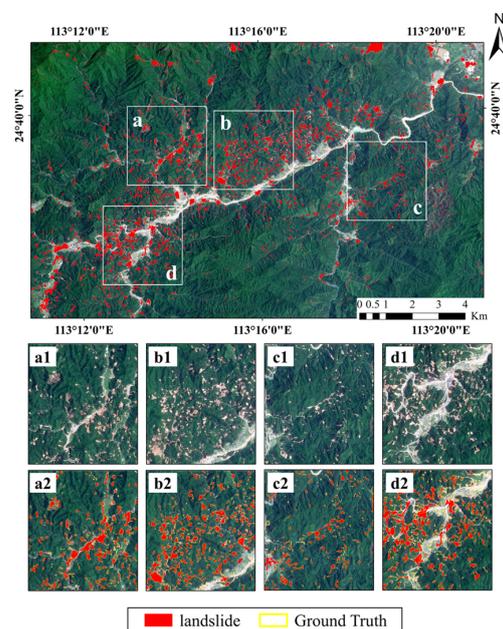


Figure 14. Landslide identification result based on LandsDANet in the four sub-regions (a–d) of Meizhou area. (a1) post-event optical image; (a2) predicted results; (b1) post-event optical image; (b2) predicted results; (c1) post-event optical image; (c2) predicted results; (d1) post-event optical image; (d2) predicted results. The boundaries of the ground truth are depicted in the subgraph with yellow outlines.

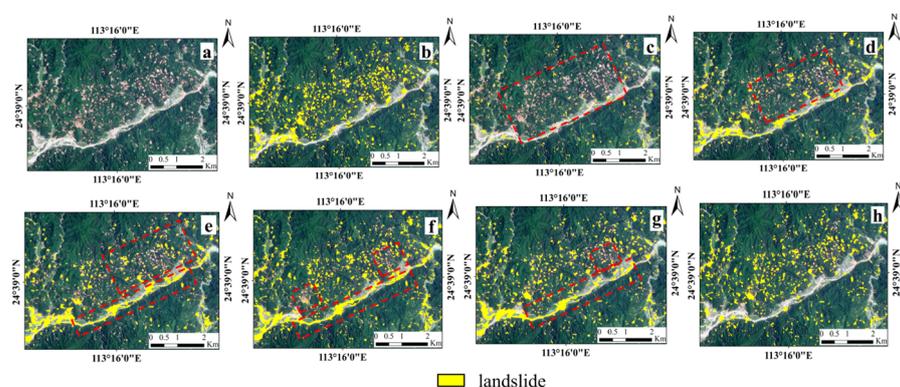


Figure 15. Examples of landslide extraction maps produced by various methods for the T2M case: (a) post-event image, (b) ground truth, (c) source only, (d) AdaptSegNet, (e) AdvEnt, (f) ADANet, (g) MemoryAdaptNet, and (h) LandsDANet. Red boxes show differences in results.

Table 5. Quantitative evaluation results of the T2M task.

Method	Landslide		mIoU	mF1	Recall	Precision	
	IoU	F1					
Source only	18.71	31.54	57.72	64.94	20.23	71.39	
State-of-the-art methods	AdaptSegNet [21]	26.56	41.97	61.48	70.07	34.84	52.77
	AdvEnt [22]	30.53	46.78	63.45	72.47	42.50	52.01
	ADANet [10]	41.95	59.11	69.22	78.66	67.12	52.81
	MemoryAdaptNet [28] (2023)	38.22	55.30	67.46	76.81	54.15	56.51
LandsDANet (ours)	45.75	62.77	71.52	80.70	60.40	65.34	

Bold values represent the best values.

6. Discussion

6.1. Ablation Study

The proposed LandsDANet benefits from the RCS, Wallis filter and contrastive loss modules. To assess how each module contributes to the model’s performance, we conducted four different ablation experiments. We established the most basic output-level alignment as a baseline and progressively incorporated the RCS, Wallis filter and contrastive loss. The experiments were executed on both the P2V_S and T2M domain adaptation tasks, and the evaluation metrics were calculated accordingly. The validation results are presented in Table 6 and illustrated in Figures 16 and 17.

Table 6. Ablation study results on P2V_S and T2M task.

Method	P2V_S Task		T2M Task	
	mF1	mIoU	mF1	mIoU
No adaptation	57.82	43.55	64.94	57.72
Baseline	62.27	48.02	706	61.6
+RCS and Wallis filter	66.52	52.45	78.14	68.94
+Contrastive Loss	69.59	55.46	78.29	68.69
+RCS+Wallisfilter+ContrastiveLoss	70.62	56.90	80.70	71.52

Bold values represent the best values.

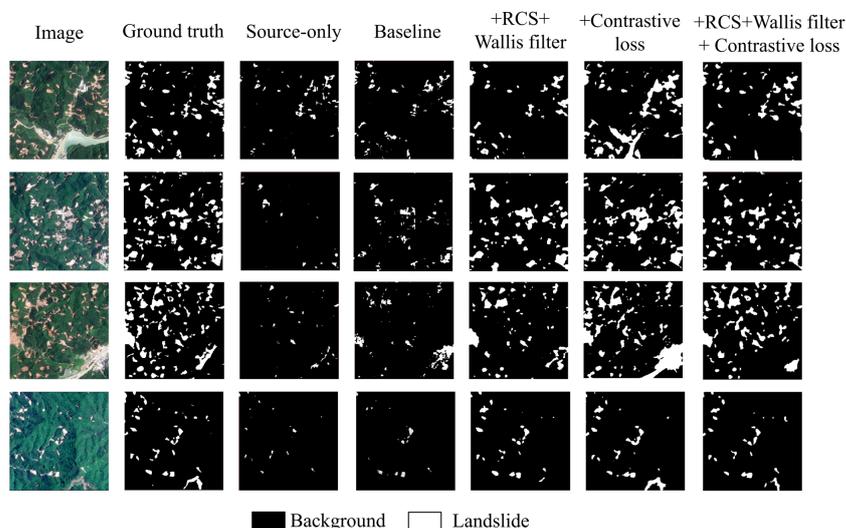


Figure 16. Results of ablation study visualized for the T2M task.

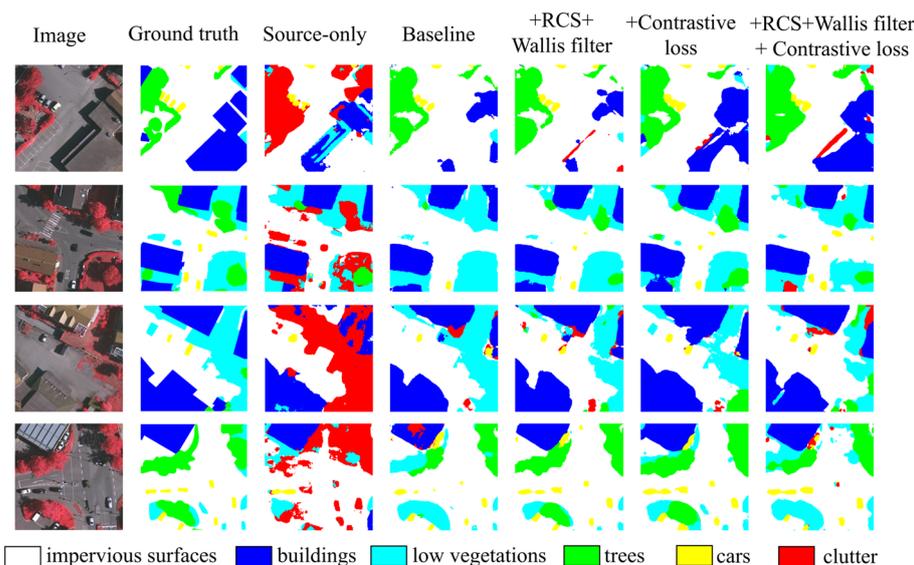


Figure 17. Results of ablation study visualized for the P2V_S task.

6.1.1. Ablation of Wallis Filter and RCS Module

By incorporating the RCS and Wallis filter, as demonstrated in Table 6, the performance of cross-domain semantic segmentation showed significant improvement. Compared to the baseline model, mIoU values increased by 4.43% and 7.34% in the P2V_S and T2M tasks, respectively, and F1-score improved by 4.25% and 8.08%, respectively. Notably, in the T2M task, recall increased by 15.93% to 46.40%, and precision increased by 9.67% to 75.90%. The improvement in recall, particularly for small landslides, can be largely attributed to the RCS strategy, which facilitates their frequent sampling during the training process. As presented in Figure 16, most landslides were accurately identified compared to the baseline and source-only models. The Wallis filter and RCS obviously alleviated the misclassification problem of rare category induced by domain shift.

6.1.2. Ablation of Contrastive Loss Module

As shown in Equation (9), we introduced contrastive loss to further optimize the training of G . In the embedding space, contrastive loss enforced the clustering of pixel features belong to the same class while pushing the dissimilar classes farther apart, thereby explicitly addressing intra-class compactness and inter-class dispersion. Through incorporating contrastive loss, the mIoU and mean F1-score in the P2V_S task increased by 7.44% and 7.32%, respectively, while those in the T2M task, also increased by 7.09% and 8.23%, respectively. Notably, in the T2M task, recall improved from 47.96% to 78.43%, whereas precision decreased from 19.3% to 46.93% compared to the baseline model. The introduction of contrastive loss significantly reduced omission errors, with the majority of landslides being accurately identified (Figure 16). However, this improvement was accompanied by an increase in false positives, as muddy rivers and exposed soil areas were incorrectly classified as landslides. This trade-off underscores the role of contrastive loss as a powerful feature discriminator that prioritizes comprehensive detection, aligning with the imperative of minimizing risk in rapid emergency assessments.

At last, the collaboration of Wallis filter, RCS and contrastive loss significantly enhances the domain adaptation capability. Table 6 demonstrates the LandsDANet's superior performance, attaining 70.62% in mean F1-score and 56.90% in mIoU for the P2V_S task, respectively. Additionally, it shows results of 80.70%, 71.52%, 60.39% and 65.34% in mean F1-score, mIoU, recall, and precision, respectively, for the T2M task. This indicates that these three modules can mutually promote each other to enhance performance in the target

domain. Furthermore, as presented in Table 6 and Figure 16, the balance between recall and precision in the T2M task, combined with the optimal F1-score, further demonstrates the complementarity of RCS, Wallis filter and contrastive loss, ultimately achieving the best prediction results.

6.2. Computational Complexity Analysis

We compared the performance and computational efficiency of various UDA methods on the P2V_S and T2M tasks. As presented in Table 7, the proposed LandsDANet achieved the lowest FOLPs, and relatively high parameter and inference times among all compared methods while maintaining competitive performance, with the highest mean F1-score value of 70.62% and 80.70% in the P2V_S and T2M tasks, respectively. Compared to AdaptSegNet, which has the lowest number of parameters, LandsDANet improves the mean F1-score by 11.36% and 10.63% in the P2V_S and T2M tasks, respectively, with an additional 42.21 million parameters. Notably, while the number of parameters in ADANet is lower than of our model, its FLOGs are 4.6 times those of LandsDANet, possibly attributing to the adoption of a dense feature pyramid in the decoder parts, whereas our model strategically employs a more efficient lightweight ALL-MLP decoder. Although MemoryAdaptNet has the lowest inference time, it possesses the highest number of parameters, and its performance is inferior to that of LandsDANet. These comprehensive comparisons demonstrate that LandsDANet achieves an optimal balance between accuracy and efficiency, providing a robust and effective UDA framework that is particularly valuable for automated cross-domain landslide inventory mapping and disaster response.

Table 7. Complexity Analysis of Different Methods on P2V_S and T2M Task.

Method	FLOPs (G)	Params (M)	Inference Time (ms)	mF1 (%)	
				P2V_S Task	T2M Task
AdaptSegNet [21]	184.72	42.83	42	59.26	70.07
AdvEnt [22]	186.12	43.16	48.61	59.99	72.47
ADANet [10]	485.68	71.36	61.23	62.36	78.66
MemoryAdaptNet [28]	143.51	134.01	39.60	62.67	76.81
LandsDANet	105.42	85.04	61.65	70.62	80.70

Bold values represent the best values.

6.3. Effects of Decoder

In the proposed LandsDANet, we utilized the ISegFormer as the segmentation network G , which consists of a transformer encoder and a simple yet efficient MLP decoder. The CBAM module was integrated to connect the encoder and decoder, adaptively determining the importance of each low-level feature channel and position information. Recently, numerous studies have proposed hybrid deep neural networks that combine transformer encoders with DCNN decoders for semantic segmentation [42]. To further enhance the performance of G , inspired by [42], we also introduced a CNN-based decoder to replace MLP. Figure 18 shows the architecture of the proposed decoder, which employs a hierarchical feature fusion strategy to progressively recover spatial resolution. Its key components include Atrous Spatial Pyramid Pooling (ASPP) for capturing multi-scale contextual information and a squeeze-and-excitation (SE) channel attention module to enhance feature representation. First, the deepest feature F_4 was enhanced through ASPP to capture contextual information at multiple scales, whereas the remaining features (namely F_3 , F_2 , and F_1) were converted to 512 channels using 1×1 convolution. Each feature connects to two branches: one branch links to the fusion block and the other is first upsampled before being merged with corresponding-resolution encoder features via elementwise addition.

In the feature fusion block, four feature representations are upsampled to $\frac{1}{4}$ of the input resolution and concatenated. To enhance feature representation capabilities, SE channel attention module is integrated to dynamically adjust the importance of each channel. Finally, two consecutive 3×3 convolutions and one 1×1 convolution are applied to obtain the predicted result. To ensure a fair comparison, we conducted tests under the same training conditions as LandsDANet. We implemented the hybrid transformer-CNN network for the P2V_S UDA task, achieving a mIoU of 52.36% and mean F1-score of 66.60%, respectively. While these metrics are slightly lower than those of LandsDANet, our analysis reveals that the model incurs significantly higher computational demands: it requires 238.49 G FLOPs, contains 108 million parameters, and exhibits an inference time of 66.74 ms per sample. These substantially higher resource requirements, particularly the 238.49 G FLOPs nearly double that of LandsDANet, directly impacting computational efficiency and deployment feasibility. Considering the requirement for a lightweight model in landslide emergency scenarios, we employ ISegFormer as the segmentation network G . In future work, we will explore more lightweight hybrid transformer-CNN networks to address the conflicting demands between representation learning and deployment inference.

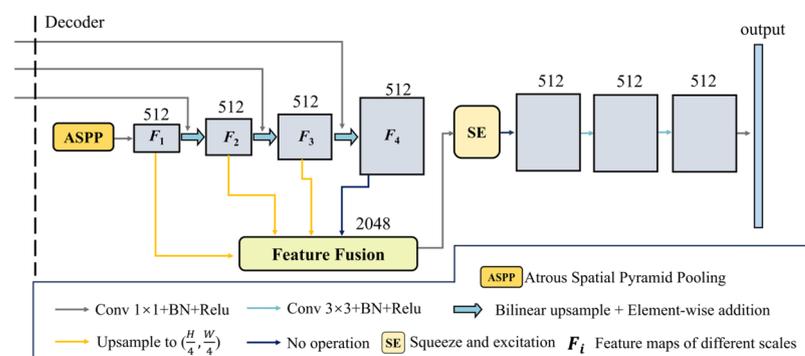


Figure 18. The architecture of the CNN-based decoder.

6.4. Advantages of LandsDANet for Cross-Domain Landslide Identification

The proposed LandsDANet framework demonstrates significant advantages in cross-domain landslide detection. Firstly, LandsDANet achieves remarkable cross-domain generalization capabilities, as evidenced by its superior performance compared to source-only models and four SOTA UDA methods. Specifically, the F1-score values increased by 39.92% and 31.23%, while the IoU values increased by 35.73% and 27.04% for the M2T and T2M tasks, respectively, when compared to the source-only method. The F1-Score serves as a metric that balances the risk of omission with the cost of commission. The higher F1-score achieved by LandsDANet across various tasks suggests that it delivers optimal results, with the generated inventory maps being comprehensive enough to guide broad rescue efforts while maintaining sufficient accuracy for effective field assessments. The significant improvement in IoU indicates not only accurate identification but also precise spatial delineation of landslide boundaries. This capability facilitates more reliable estimations of affected areas and volumes, which is critical for planning the scale of response. Essentially, the advantages of LandsDANet extend beyond superior benchmark metrics; for time-critical disaster response, this represents a significant advancement. Secondly, as discussed in Section 6.2, the architecture maintains competitive computational efficiency, facilitating its deployment for processing large-scale satellite imagery in practical scenarios. Importantly, the framework mitigates the reliance on labor- and time-intensive pixel-level annotations of target domain samples, which represent a major bottleneck in large-scale landslide rapid mapping. This advantage is crucial for enabling rapid, near-real-time landslide assessment in emergency response situations. In addition, we analyzed the model's

performance across different landslide scales to evaluate its capability to detect small objects. As presented in Appendix A Tables A1 and A2, the F1-score exhibited a decline for smaller landslides (i.e., areas less than 900 m²). Although our framework effectively aligns domain-invariant features for typical landslides, the detection of very small landslides remains constrained by the fundamental limits of pixel-wise segmentation at moderate resolutions. Future enhancements could involve the implementation of explicit multi-scale learning or loss functions specifically designed for small objects.

We employed the t-distributed stochastic neighbor embedding (t-SNE) [43] method to visualize the generated representation embeddings of the target domain images into a two-dimensional space. Considering the benchmark datasets which contain six categories would yield more pronounced results via t-SNE; thus, we selected the P2V_S task for experimentation. As shown in Figure 19, the representations before the application of UDA are convoluted, exhibiting numerous overlaps. Through UDA, LandsDANet alleviated domain shifts and ensured that pixels of the same category in the target image were clustered together, particularly for the building and tree categories. This outcome confirms that our method can generate more category-discriminative representations, thereby improving the performance of UDA.

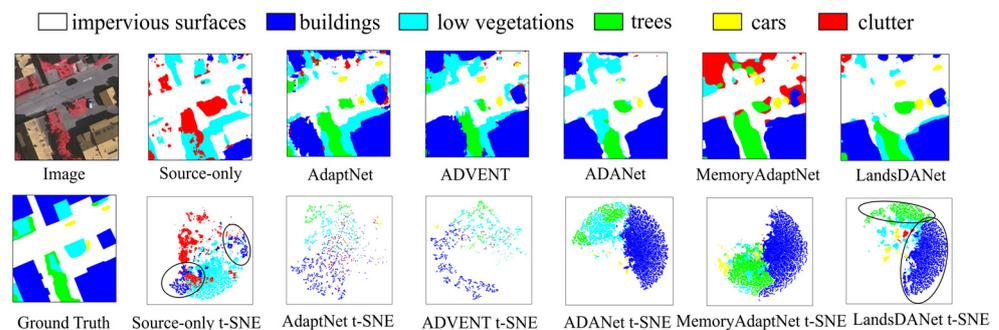


Figure 19. t-SNE embedding results of P2V_S task. The **top** row represents the prediction result of each model, while the **bottom** row displays the corresponding t-SNE results. The black circle represents the clustering of features.

6.5. Limitations and Future Study

The proposed method employs a pixel-wise UDA semantic segmentation framework for cross-domain landslide detection. Firstly, adversarial domain adaptation has demonstrated effectiveness in mitigating domain shift, the sensitivity and instability associated with GAN-based training remain a challenge. The tuning of hyperparameters, such as the learning rate and τ in contrastive loss, is critical for achieving convergence; however, this process can be time-consuming. Secondly, this study exclusively utilized post-event optical images containing red, blue and green bands obtained from PlanetScope, while commonly used terrain data and spectral indices were disregarded [44]. This resulted in spectral ambiguity between landslides and visually similar features, such as bare agricultural fields, exposed riverbank sediments, and dry sparse vegetation areas in the M2T task. Thirdly, while our experiments validate the framework across two major triggers, i.e., earthquakes and rainfall, its performance on landslides induced by other triggering factors, across various lithologies, or under diverse phenological conditions remains unexplored. Lastly, our model shows robustness to moderate GSD variations, its performance under extreme resolution shifts (e.g., from sub-meter to multi-meter) combined with semantic shifts warrants further investigation. Future research will focus on integrating more stable and efficient adversarial learning frameworks to mitigate training instability and reduce hyperparameter sensitivity. Additionally, we will explore few-shot learning to facilitate rapid adaptation to novel target domains with minimal labeled data [45]. Furthermore,

we will create a comprehensive large-scale landslide dataset that incorporates pre- and post-event optical images, terrain data, Synthetic Aperture Radar images, and spectral indices to identify a more diverse, and multi-temporal landslide inventories.

7. Conclusions

This study presents LandsDANet, an innovative semantic segmentation framework for UDA, aiming at tackling the discrepancies in data distributions for cross-domain landslide identification. We employ the Wallis filter to transform source images into target-stylized images for image-level alignment instead of complicated image-to-image translation models. Additionally, we utilized RCS to frequently sample images containing the rare landslide category, addressing the issue of class imbalance. Adversarial learning is introduced to align the source and target domains at the output level, further reducing domain discrepancies. We utilized ISegFormer as a semantic segmentation network which integrates a CBAM attention mechanism to generate accurate predictions. Moreover, the inclusion of contrastive loss enhances the discriminative ability and robustness of LandsDANet as an auxiliary loss function. The efficiency of LandsDANet is first assessed using the Potsdam and Vaihingen benchmark datasets, which differ in geographic locations, spectral band composition and spatial resolution. Subsequently, we validate it in two landslide scenarios induced by earthquakes and rainfall to assess its cross-disaster domain adaptability. The landslides in these two scenarios exhibit significant differences in morphology, scale, hue, and geographic environment. Compared to the source-only model, performance improved across all tasks. The F1-score achieved gains of 31.23% on the T2M task and 39.92% on M2T task. Meanwhile, the mIoU improved by 13.35% for P2V_S task, 10.2% for V2P_S task, and 19.29% for P2V_D task. These results confirming that the proposed LandsDANet is highly effective in mitigating domain shifts. Future studies will focus on developing more lightweight UDA models that can be deployed on resource-constrained edge devices or cloud platforms, facilitating rapid disaster response and continuous monitoring.

Author Contributions: Conceptualization, J.Y. and W.H.; methodology, J.Y.; software, J.Y. and W.H.; validation, Q.X. and J.Y.; formal analysis, J.Y., L.P. and B.C.; investigation, L.P. and B.C.; resources, M.D.; data curation, W.H. and F.Z.; writing—original draft preparation, J.Y.; writing—review and editing, J.Y., M.D. and B.C.; visualization, F.Z.; supervision, Z.L.; project administration, Z.L.; funding acquisition, M.D. and Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 42374027), Geological Survey Projects of China Geological Survey, (grant Nos. DD20230436 and the 2022 Qinghai Province “Kunlun Talent” Program for High-End Innovation and Entrepreneurship.

Data Availability Statement: The PlanetScope imagery used in this study is provided by Fang et al. (2024) [11], and is publicly available for research purposes under the CC-BY 4.0 license. The landslide inventory (ground truth masks) generated and analyzed in this study will be made available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to acknowledge the providers of the optical images used in this study. The Planet optical image data utilized in this research available from: <https://doi.org/10.5194/essd-16-4817-2024-supplement> (accessed on 10 November 2025).

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

Appendix A

Table A1. Performance Assessment of LandsDANet under Different Landslide Scales in Taiwan area.

Area/m ²	Detection Rate	Precision	Recall	F1
<90	30.5	34.3	26.9	30.2
99–900	46.6	37.6	31.0	34.0
909–9000	88.1	54.9	62.6	58.5
>9009	98.2	56.6	67.2	61.4

Table A2. Performance Assessment of LandsDANet under Different Landslide Scales in Meizhou area.

Area/m ²	Detection Rate	Precision	Recall	F1
<90	19.7	36.6	18.8	24.9
99–900	45	49.8	32.4	39.2
909–9000	86.6	65.7	62.6	64.1
>9009	96	65.9	63.1	64.5

References

1. Keefer, D.K. Landslides caused by earthquakes. *Geol. Soc. Am. Bull.* **1984**, *95*, 406–421. [[CrossRef](#)]
2. Fan, X.; Liu, B.; Luo, J.; Pan, S.; Han, S.; Zhou, Z. Comparison of earthquake-induced shallow landslide susceptibility assessment based on two-category LR and KDE-MLR. *Sci. Rep.* **2023**, *13*, 833. [[CrossRef](#)]
3. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, A.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
4. Yang, J.; Ding, M.; Huang, W.; Li, Z.; Zhang, Z.; Wu, J.; Peng, J. A Generalized Deep Learning-Based Method for Rapid Co-Seismic Landslide Mapping. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2024**, *17*, 16970–16983. [[CrossRef](#)]
5. Xie, E.; Wang, W.; Yu, A.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–14 December 2021; pp. 12077–12090.
6. Gao, M.; Chen, F.; Wang, L.; Zhao, H.; Yu, B. Swin Transformer-Based Multiscale Attention Model for Landslide Extraction From Large-Scale Area. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4415314. [[CrossRef](#)]
7. Dong, A.; Dou, J.; Li, C.; Chen, Z.; Ji, J.; Xing, K.; Zhang, J.; Daud, H. Accelerating Cross-Scene Co-Seismic Landslide Detection Through Progressive Transfer Learning and Lightweight Deep Learning Strategies. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4410213. [[CrossRef](#)]
8. Lv, P.; Ma, L.; Li, Q.; Du, F. ShapeFormer: A Shape-Enhanced Vision Transformer Model for Optical Remote Sensing Image Landslide Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2681–2689. [[CrossRef](#)]
9. Wu, L.; Liu, R.; Ju, N.; Zhang, A.; Gou, J.; He, G.; Lei, Y. Landslide mapping based on a hybrid CNN-transformer network and deep transfer learning using remote sensing images with topographic and spectral features. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *126*, 103612. [[CrossRef](#)]
10. Xu, Q.; Ouyang, C.; Jiang, T.; Yuan, X.; Fan, X.; Cheng, D. MFFENet and ADANet: A robust deep transfer learning method and its application in high precision and fast cross-scene recognition of earthquake induced landslides. *Landslides* **2022**, *19*, 1617–1647. [[CrossRef](#)]
11. Fang, C.; Fan, X.; Wang, X.; Nava, L.; Zhong, H.; Dong, X.; Qi, J.; Catani, F. A globally distributed dataset of coseismic landslide mapping via multi-source high-resolution remote sensing images. *Earth Syst. Sci. Data* **2024**, *16*, 4817–4842. [[CrossRef](#)]
12. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
13. Rottensteiner, F.; Sohn, G.; Jung, J.; Gerke, M.; Breikopf, U. The ISPRS benchmark on urban object classification and 3D building reconstruction. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *I-3*, 293–298. [[CrossRef](#)]
14. Liu, X.; Xing, F.; You, J.; Lu, J.; Kuo, C.; Fakhri, G.E.; Woo, J. Subtype-Aware Dynamic Unsupervised Domain Adaptation. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 2820–2834. [[CrossRef](#)] [[PubMed](#)]
15. Ouyang, L.; Key, A. Maximum Mean Discrepancy for Generalization in the Presence of Distribution and Missingness Shift. *arXiv* **2021**, arXiv:2111.10344.

16. Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.-Y.; Isola, P.; Saenko, K.; Efros, A.A.; Darrell, T. CyCADA: Cycle-consistent adversarial domain adaptation. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1989–1998.
17. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2242–2251.
18. Tasar, O.; Happy, S.L.; Tarabalka, Y.; Alliez, P. ColorMapGAN: Unsupervised Domain Adaptation for Semantic Segmentation Using Color Mapping Generative Adversarial Networks. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7178–7193. [[CrossRef](#)]
19. Chen, Y.-H.; Chen, W.-Y.; Chen, Y.-T.; Tsai, B.-C.; Wang, Y.-C.-F.; Sun, M. No more discrimination: Cross city adaptation of road scene segmenters. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1992–2001.
20. Hoffman, J.; Wang, D.; Yu, F.; Darrell, T. FCNs in the wild: Pixel-level adversarial and constraint-based adaptation. *arXiv* **2016**, arXiv:1612.02649.
21. Tsai, Y.-H.; Hung, W.-C.; Schuster, S.; Sohn, K.; Yang, M.-H.; Chandraker, M. Learning to adapt structured output space for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7472–7481.
22. 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, Long Beach, CA, USA, 16–20 June 2019; pp. 2517–2526.
23. Luo, Y.; Zheng, L.; Guan, T.; Yu, J.; Yang, Y. Taking a Closer Look at Domain Shift: Category-Level Adversaries for Semantics Consistent Domain Adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 2502–2511.
24. Wang, H.; Hen, T.; Zhang, S.W.; Duan, L.; Mei, T. Classes Matter: A Fine-Grained Adversarial Approach to Cross-Domain Semantic Segmentation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 642–659.
25. French, G.; Mackiewicz, M.; Fisher, M. Self-ensembling for visual domain adaptation. *arXiv* **2017**, arXiv:1706.05208.
26. Zheng, Z.; Yang, Y. Rectifying Pseudo Label Learning via Uncertainty Estimation for Domain Adaptive Semantic Segmentation. *Int. J. Comput. Vis.* **2020**, *129*, 1106–1120. [[CrossRef](#)]
27. Zhang, P.; Zhang, B.; Zhang, T.; Chen, D.; Wang, Y.; Wen, F. Prototypical Pseudo Label Denoising and Target Structure Learning for Domain Adaptive Semantic Segmentation. *arXiv* **2021**, arXiv:2101.10979. [[CrossRef](#)]
28. Zhu, J.; Guo, Y.; Sun, G.; Yang, L.; Deng, M.; Chen, J. Unsupervised domain adaptation semantic segmentation of high-resolution remote sensing imagery with invariant domain-level prototype memory. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5603518. [[CrossRef](#)]
29. Zhang, L.; Lan, M.; Zhang, J.; Tao, D. Stagewise Unsupervised Domain Adaptation With Adversarial Self-Training for Road Segmentation of Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5609413. [[CrossRef](#)]
30. Peng, D.; Guan, H.; Zang, Y.; Bruzzone, L. Full-level domain adaptation for building extraction in very-high-resolution optical remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5607317. [[CrossRef](#)]
31. Chen, J.; Zhu, J.; He, P.; Guo, Y.; Hong, L.; Yang, Y. Unsupervised Domain Adaptation for Building Extraction of High-Resolution Remote Sensing Imagery Based on Decoupling Style and Semantic Features. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4406917. [[CrossRef](#)]
32. Li, P.; Wang, Y.; Si, T.; Ullah, K.; Han, W.; Wang, L. DSFA: Cross-scene domain style and feature adaptation for landslide detection from high spatial resolution images. *Int. J. Digit. Earth* **2023**, *16*, 2426–2447. [[CrossRef](#)]
33. Yu, B.; Chen, F.; Chen, W.; Shi, G.; Xu, C.; Wang, N.; Wang, L. Cross-domain landslide mapping by harmonizing heterogeneous remote sensing datasets. *GISci. Remote Sens.* **2025**, *62*, 2559457. [[CrossRef](#)]
34. Zhang, X.; Yu, W.; Pun, M.-O.; Shi, W. Cross-domain landslide mapping from large-scale remote sensing images using prototype-guided domain-aware progressive representation learning. *ISPRS J. Photogramm. Remote Sens.* **2023**, *197*, 1–17. [[CrossRef](#)]
35. Li, P.; Wang, G.; Liu, G.; Fang, Z.; Ullah, K. Unsupervised Landslide Detection From Multitemporal High-Resolution Images Based on Progressive Label Upgradation and Cross-Temporal Style Adaption. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4410715. [[CrossRef](#)]
36. Wei, R.; Li, Y.; Li, Y.; Zhang, B.; Wang, J.; Wu, C.; Yao, S.; Ye, C. A universal adapter in segmentation models for transferable landslide mapping. *ISPRS J. Photogramm. Remote Sens.* **2024**, *218*, 446–465. [[CrossRef](#)]
37. Wang, J.; Zhang, X.; Ma, X.; Yu, W.; Ghamisi, P. Auto-Prompting SAM for Weakly Supervised Landslide Extraction. *IEEE Geosci. Remote Sens. Lett.* **2025**, *22*, 6008705. [[CrossRef](#)]
38. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Convolutional block attention module (CBAM). In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.

39. Zhao, S.; Wang, Y.; Yang, Z.; Cai, D. Region mutual information loss for semantic segmentation. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 11117–11127.
40. Chang, J.M.; Chao, W.A.; Yang, C.M.; Huang, M.W. Coseismic and subsequent landslides of the 2024 Hualien earthquake (M7.2) on April 3 in Taiwan. *Landslides* **2024**, *21*, 2591–2595. [[CrossRef](#)]
41. Chen, Y.; Song, C.; Li, Z.; Chen, B.; Yu, C.; Hu, J.-C.; Cai, X.; Zhu, S.; Wang, Q.; Ma, Y.; et al. Preliminary analysis of landslides induced by the 3 April 2024 Mw 7.4 Hualien, Taiwan earthquake. *Landslides* **2025**, *22*, 1551–1562. [[CrossRef](#)]
42. Zhang, C.; Jiang, W.; Zhang, Y.; Wang, W.; Zhao, Q.; Wang, C. Transformer and CNN Hybrid Deep Neural Network for Semantic Segmentation of Very-High-Resolution Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4408820. [[CrossRef](#)]
43. Laurens, V.D.M.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
44. Liu, X.; Peng, Y.; Lu, Z.; Li, W.; Yu, J.; Ge, D.; Xiang, W. Feature-Fusion Segmentation Network for Landslide Detection Using High-Resolution Remote Sensing Images and Digital Elevation Model Data. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4500314. [[CrossRef](#)]
45. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2016**, arXiv:1511.06434. [[CrossRef](#)]

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.