applied sciences

MDPI

*Article*

# An Anomaly Detection-Based Domain Adaptation Framework for Cross-Domain Building Extraction from Remote Sensing Images

**Shaoxuan Zhao, Xiaoguang Zhou and Dongyang Hou ***

School of Geosciences and Info-Physics, Central South University, Changsha 410083, China
* Correspondence: houdongyang1986@csu.edu.cn

**Abstract:** Deep learning-based building extraction methods have achieved a high accuracy in closed remote sensing datasets. In fact, the distribution bias between the source and target domains can lead to a dramatic decrease in their building extraction effect in the target domain. However, the mainstream domain adaptation methods that specifically address this domain bias problem require the reselection of many unlabeled samples and retraining in other target domains. This is time-consuming and laborious and even impossible at small regions. To address this problem, a novel domain adaptation framework for cross-domain building extraction is proposed from a perspective of anomaly detection. First, the initial extraction results of images in the target domain are obtained by a source domain-based pre-trained model, and then these results are classified into building mixed and non-building layers according to the predicted probability. Second, anomalous objects in the building layer are detected using the isolation forest method. Subsequently, the remaining objects in the building layer and the objects in the non-building layer are used as positive and negative samples, respectively, to reclassify the mixed layer using the random forest classifier. The newly extracted objects are fused with the remaining objects in the building layer as the final result. Four different experiments are performed on different semantic segmentation models and target domains. Some experimental results indicate that our framework can improve cross-domain building extraction compared to the pre-trained model, with an 8.7% improvement in the F1 metric when migrating from the Inria Aerial Image Labeling dataset to the Wuhan University dataset. Furthermore, experimental results show that our framework can be applied to multiple target domains without retraining and can achieve similar results to domain adaptation models based on adversarial learning.

**Keywords:** building extraction; remote sensing image; domain adaptation; semantic segmentation; anomaly detection

check for updates

**Citation:** Zhao, S.; Zhou, X.; Hou, D. An Anomaly Detection-Based Domain Adaptation Framework for Cross-Domain Building Extraction from Remote Sensing Images. *Appl. Sci.* **2023**, *13*, 1674. https://doi.org/10.3390/app13031674

Academic Editor: Sungho Kim

Received: 3 January 2023
Revised: 20 January 2023
Accepted: 27 January 2023
Published: 28 January 2023

**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

As one of the important elements of urban geographic information, the accurate extraction of buildings plays an important role in the analysis of the human living environment urban planning layout, disaster damage assessment, and illegal building detection [1–3]. Therefore, building extraction from remote sensing images has been a hot research topic in the field of remote sensing [4].

In recent years, with the development of artificial intelligence technology, building extraction from remote sensing images is gradually dominated by deep semantic segmentation models represented by UNet [5], PSPNet (pyramid scene parsing network) [6], and so on [7–9]. These deep learning-based building extraction methods have achieved F1 values above 90% on closed datasets thanks to the powerful feature extraction capability of deep learning models. However, these deep semantic segmentation models suffer from the problems of requiring a large number of samples and a weak generalization ability. That is, when the distribution of the training dataset (source domain) differs from that of the actual

dataset (target domain), it is necessary to reselect the training samples and carry out the training again, which is often time-consuming and laborious [10,11].

Currently, the mainstream approach to solve the above-mentioned model migration problem is domain adaptation [12–14]. The domain adaptation methods aim to transfer the knowledge learned from a source domain to a new target domain [15,16]. For example, Tasar et al. [17] presented a DAugNet for multisource, multitarget, and life-long domain adaptation problems by using only one encoder, one decoder, and one discriminator. Shi et al. [18] proposed a joint pixel- and representation-level network based on a cycle generative-adversarial network and adversarial learning. These domain adaptation methods mainly use the idea of adversarial learning and do not require labeled data in the target domain [19]. This indeed improves the performance of cross-domain building extraction while reducing the labeling workload. However, these methods still require a large number of unlabeled samples in the target domain to complete the training process with the labeled samples in the source domain and, even worse, they need to be retrained whenever the target domain changes. This condition is difficult to be satisfied for a small scope of cross-domain building extraction scenarios. Moreover, it should be noted that the research on domain adaptation is still in its initial stage.

To address the abovementioned issues, we attempt to investigate the domain adaptation method from a novel perspective of anomaly detection and propose an anomaly detection-based domain adaptation framework (AD-DAF) for cross-domain building extraction from remote sensing images in this paper. The framework is designed with anomaly detection and reclassification processes based on a conventional pre-trained model. The anomaly detection process in our framework is a bridge between the knowledge of object types in the source domain and the knowledge of target features in the target domain, which is used to refine the accuracy of the pre-trained model for cross-domain building extraction from the target domain. The effect of this process is equivalent to the training process of the mainstream domain adaptation methods. The reclassification process is designed to refine again the effect of cross-domain building extraction from the predicted non-building data. It should be noted that the framework can be adapted to a variety of mainstream segmentation models. In other words, our framework can be used as a post-processing step for mainstream semantic segmentation models. The main contributions of this paper can be summarized as follows.

(1) A novel anomaly detection-based domain adaptation framework is proposed to extract buildings from cross-domain remote sensing images. Compared with the current mainstream domain adaptation methods, our framework does not require a large number of unlabeled samples in the target domain and it retrains in new target domain.

(2) Numerous experiments are conducted on three widely used deep learning models, large area data with different region types, and publicly available datasets. Our experimental results indicate that our proposed framework can improve the performance of cross-domain building extraction and can achieve the role of becoming the state-of-the-art similarly to other recent domain adaptation methods.

The remainder of this paper is organized as follows. Section 2 provides the related work, including building extraction in the closed datasets, domain adaptation, and its applications in remote sensing. Section 3 describes our anomaly detection-based domain adaptation framework, including source domain-based pre-training, anomaly detection for the target domain, and local region reclassification. Section 4 reports four experiments and their analysis, followed by the conclusions and future work in Section 5.

## 2. Related Work

### 2.1. Building Extraction in Closed Datasets

Building extraction in closed datasets means that the training set (source domain) and test set (target domain) come from the same dataset (domain). In terms of the image features, building extraction methods can be divided into handcrafted feature-based and deep-level feature-based methods. The handcrafted-based extraction methods mainly rely

on features of samples such as the texture, structure, and spectrum [20]. These features are then typically used as the input to traditional machine learning methods (i.e., support vector machine and random forest) to train classification models. This type of method is more effective for a small-scale building extraction. However, its accuracy in large-scale building extraction has much room for improvement due to the diversity of the features.

Unlike the handcrafted-based extraction methods, the features of samples are automatically learned by various deep semantic segmentation models in the deep-level feature-based methods. For example, Hui et al. [21] modified the deep semantic segmentation model UNet with the Xception module and multitask learning to extract the buildings. Liu et al. [22] proposed an efficient segmentation network by incorporating factorized residual block and the dilated convolutions for building extraction from high-resolution aerial images. In addition, to further improve the extraction accuracy, attention mechanisms, which can concentrate on significant target features [23], are incorporated into deep semantic segmentation models. For example, Chen et al. [24] incorporated self-attention and reconstruction-bias modules into the UNet architecture for building extraction. Guo et al. [25] added an attention block and multiple losses to the UNet architecture for enhancing the character of the building. These methods indeed have a higher accuracy of building extraction than the handcrafted-based extraction methods. However, these methods require a much higher number of labeled samples than the handcrafted-based methods. In addition, the accuracy of these methods drops dramatically when the target domain does not coincide with the source domain source (i.e., domain bias). It should be noted that both of these two types of methods require labeled samples and assume that the samples are correct. They usually use cross-validation or batch training to reduce the impact of incorrect samples, which belongs to partial formal verification methods [26].

### 2.2. Domain Adaptation in Remote Sensing

In fact, the domain bias phenomenon is very common in the field of remote sensing due to the imaging conditions and sensors, etc. Therefore, domain adaptation methods, which are used to reduce the difference between the target and source domain distributions, have been widely applied in remote sensing fields, such as remote sensing image retrieval [19], scene classification [27], object detection [28], semantic segmentation [18,29], change detection [30], and so on. For example, Hou et al. [19] used pseudo-label consistency learning to reduce domain bias in retrieval of different optical remote sensing images. Zhang et al. [31] employed a correlation subspace dynamic distribution alignment method with convolutional neural networks to reduce the domain difference in remote sensing image scene classification. Ji et al. [32] used a generative adversarial network to align the source and target domains in a land cover classification.

Recently, many scholars have also used domain adaptation methods for building extraction as well. For instance, Peng et al. [10] utilized a simple Wallis filter method and adversarial learning to reduce the differences between different domains for cross-domain building extraction. Na et al. [29] proposed a segmentation network based on a domain adaptation transmission attack for cross-domain building extraction. Dias et al. [16] used two adversarial learning-based methods to explore the impact of domain adaptation on building extraction. These domain adaptation methods mainly rely on adversarial learning, self-learning, and differential alignment. They have a substantial improvement in the accuracy in the target domain compared to the methods for closed datasets. This is because these methods migrate the source domain knowledge to the target domain and reduce the effect of domain bias. However, these methods require not only labeled source domain data but also a large amount of unlabeled target domain data. In small-scale scenarios, large amounts of unlabeled data are difficult to obtain. In addition, these methods require a retraining of the model when the target domain or study area is replaced. Unlike these methods, this paper focuses on correlating the knowledge of source and target domains from the perspective of anomaly detection to reduce the domain bias.

## 3. Methodology

### *3.1. Overall Architecture*

The current mainstream domain adaptation frameworks often require a retraining of the deep learning models with large amounts of unlabeled data in the target domain and large amounts of labeled data in the source domain. That is, retraining with large amounts of unlabeled and labeled data is required for each change in the target domain. This is difficult to apply for target domains with an insufficient amount of data. To tackle this problem, our AD-DAF framework fuses a deep semantic segmentation model with an anomaly detection idea, emphasizing the combination of knowledge in both the source and target domain. The introduction of the anomaly detection idea allows the AD-DAF framework to be applied to multiple target domains simultaneously with only one pre-training. This drastically reduces the need for retraining with large amounts of data. The overall architecture of our anomaly detection-based domain adaptation framework is shown in Figure 1.

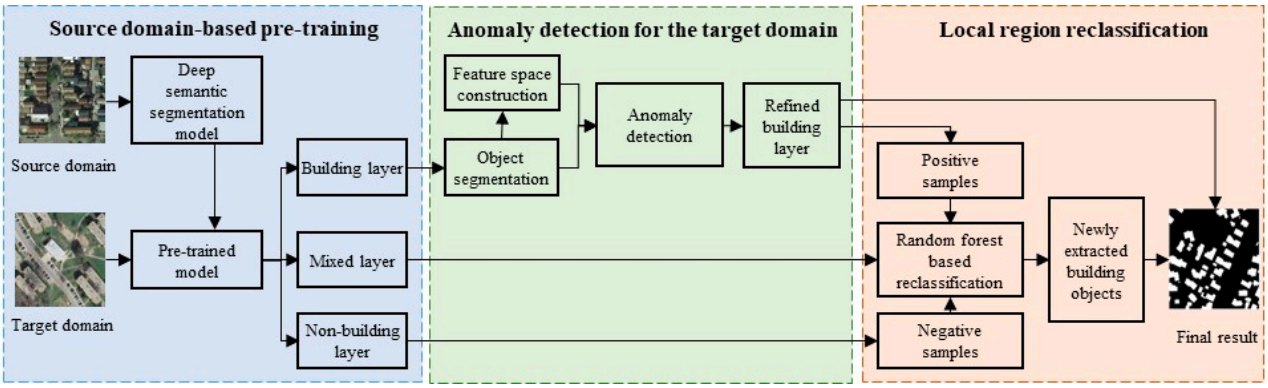



**Figure 1.** Overall architecture of our anomaly detection-based domain adaptation framework.

It can be seen from Figure 1 that our AD-DAF framework mainly consists of three processes: source domain-based pre-training, anomaly detection for the target domain, and local region reclassification. The source domain-based pre-training process is responsible for learning the building extraction model from the source domain in order to obtain the initial results of the building extraction for the target domain. The process of the anomaly detection for the target domain is responsible for detecting misclassification from the initial results of the building extraction. The local region reclassification process is designed to re-extract buildings from the above remaining results. The details are given in the following Section.

### *3.2. Source Domain-Based Pre-Training*

The source domain-based pre-training process focuses on creating a pre-trained model for the building extraction by using a deep semantic segmentation model and an open access source domain dataset, such as the Inria Aerial Image Labeling [33] and the WHU dataset (Wuhan University dataset) [34]. The AD-DAF framework is not limited to a particular deep semantic segmentation model. Any of the deep semantic segmentation models such as fully convolutional networks, UNet, PSPNet, and LinkNet [35], and their improvements can be applied to this framework. This means that our framework can be used as a complement to the current approaches. In other words, this also means that our framework can be used as a post-processing method for the current mainstream building extraction methods.

The pre-trained model can output the classification probability that each pixel is classified as a building. Based on the output probability, the input image in the target domain can be divided into building, mixed, and non-building layers through the given thresholds. Regions with a pixel classification probability greater than 0.6 are classified as the building layer, regions with a pixel classification probability less than 0.2 are classified

as the non-building layer, and regions with a pixel classification probability between 0.2 and 0.6 are classified as the mixed layer. Obviously, the above given thresholds and the pre-trained model cannot ensure that the building layer is completely correct; it can only ensure that most of the pixels in the building layer are buildings. This is mainly because the pre-trained model is obtained based on the labeled images in the source domain, which holds bias towards the unlabeled images in the target domain.

To reduce the possibility of bias between the images in the source and target domains, a test time augmentation strategy [36] is applied in this paper. Specifically, each input image in the target domain is rotated by 0°, 90°, 180°, 270°, flipped up and down, and flipped horizontally to obtain six new images. The final classification probability of the input image is the average of the classification probabilities of the corresponding pixels in these six augmented images. This will subsequently improve the accuracy of the classification probability.

In addition, it is important to note that to reduce the disturbance of roads on the building extraction, roads are roughly extracted by using the same deep semantic segmentation model and DeepGlobe road extraction dataset [37], and then these extracted roads are removed from the output of the pre-trained model.

### 3.3. Anomaly Detection for the Target Domain

The above pre-trained model is less effective in segmenting the target domain due to the data distribution bias between the source and the target domains. Therefore, the process of anomaly detection for the target domain is introduced into our AD-DAF framework to further remove outliers (i.e., non-buildings) in the building layer. In this paper, the isolation forest approach [38] is selected to perform anomaly detection for the target domain. It is an unsupervised anomaly detection approach with less execution time, memory requirements, and a high precision [39]. The isolation forest approach isolates each anomaly data by constructing binary trees. In general, the binary trees are constructed from the random properties of the dataset and anomaly data are those nodes that have the shortest average path lengths on the binary trees [40].

In this paper, the color feature, texture feature, morphological building index, and morphological shadow index are selected to construct the feature space as the random properties for anomaly detection [41]. The specific processes are shown below. First, the mean-shift algorithm [42] is used to segment the building layer into different objects. Then, the feature vectors of each segmented object are computed using the above features. Next, binary trees are constructed from the feature vectors of the segmented objects and non-building objects are detected based on the shortest average path lengths on the binary trees. Finally, a refined building layer is obtained for the local region's reclassification.

In practice, the above anomaly detection approach can also be used in the non-building layer in order to reduce the possibility of buildings in the non-building layer and to add anomaly-free objects to the mixed layer.

### 3.4. Local Region Reclassification

The above anomaly detection only refines the extraction accuracy of the building layer, but in fact the mixed layer still contains a large number of building objects due to the weak generalization ability of the pre-training model. Therefore, we design the local reclassification process for the mixed layer in order to extract the building objects from it.

The random forest method is chosen as the core method for this local reclassification process. This is because it has a strong noise immunity and high classification accuracy [43]. The random forest method is a supervised classification algorithm. In this paper, the refined building layer is selected as the positive samples and the non-building layer is selected as the negative samples for the random forest method. Specifically, the sample set is randomly selected from the refined building layer and the non-building layer through an unordered sampling with a replacement strategy. This sample set is then divided into a training set and a validation set in a 2:1 ratio. Then, a corresponding decision tree model is set up for each training sample set and continues to split until all the training samples at that node

are of the same type. Furthermore, the generated multiple decision trees are formed into a random forest and the best classification of the mixed layer is decided based on the voting probability. Finally, the combined random forest is used to extract building objects from the mixed layer and the newly extracted building objects are fused with the objects in the refined building layer as the final result.

## 4. Experiments and Analysis

To verify the effectiveness of our AD-DAF framework, four experiments are designed in this paper. The first experiment is to verify the applicability of different deep semantic segmentation models in the proposed framework. The second experiment is to verify the effectiveness of the framework for cross-domain building extraction from remote sensing images in large areas. The third experiment is to test the performance of the framework on a benchmark dataset. The fourth experiment is to compare the differences between the framework and the main domain adaptation methods.

These experiments are implemented via Python on a desktop server with an Intel (R) Core (TM) i9-10900K processor, 32GB RAM (random access memory). The four experiments are conducted by the Keras library with TensorFlow 2.1 backend. In the pre-training process, the learning rate is $1 \times 10^{-4}$ and the number of epochs is set to 20.

The Precision, Recall, F1, and IoU (intersection over union) score are selected to quantitatively evaluate the accuracy of the cross-domain building extraction. Specifically, Recall expresses the proportion of correctly extracted buildings to the total buildings in the image [44]. Precision represents the proportion of correctly extracted buildings in the extracted result [44]. F1 is the weighted average of Recall and Precision. IoU is the percentage of the intersection of the predicted and true results in their concurrent sets. The larger their values are, the more effective the method is. They can be calculated by the following equations [3].

$$\text{Recall} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{4}$$

where $TP$ denotes the number of correctly predicted building pixels in the extracted result, $FP$ denotes the number of incorrect building pixels in the extracted result, and $FN$ denotes the number of missed building pixels.

### 4.1. Experiment I: Semantic Segmentation Model Applicability

In this experiment, three dominant models, UNet, PSPNet, and LinkNet, are chosen to test the applicability of our AD-DAF framework. The Inria Aerial Image Labeling Dataset (IAILD) is selected as the source domain for the pre-training model. It is an open access dataset for urban building detection with building and non-building semantic classes. The dataset consists of 180 images with $5000 \times 5000$ pixels and a pixel spatial resolution of 0.3 m. It mainly covers different forms of urban settlements in the United States and Australia [33], which makes it often used to test the generalization capabilities of deep semantic segmentation models. In the 180 images, 150 images are selected for training and the remaining 30 images for validation. Specifically, each training image with $5000 \times 5000$ pixels is split into $384 \times 384$ pixel patches under the condition of a 20% overlap, and finally 40,836 training samples for deep semantic segmentation models are obtained.

A dense residential image in the urban area and a sparse residential image in the suburban area from different regions are selected as the target domains to verify the domain adaptation effect of our AD-DAF framework. Figure 2 shows the two images from

different regions. The first image is selected from Google Images at Shenzhen, China, with 500 × 500 pixels and a spatial resolution of 1 m. The second image is selected from the ISPRS (International Society for Photogrammetry and Remote Sensing) semantic segmentation dataset [45] at Postdam, Germany, with 6000 × 6000 pixels and a spatial resolution of 0.05 m. The building labels of the first image are obtained by manual vectorization, and the building labels of the third image are directly from the ISPRS dataset. Compared with the source domain of IAILD, the above two images are from different regions with different resolutions; thus, these two images can be considered as two different target domains in the experiments.

(**a**) (**b**)

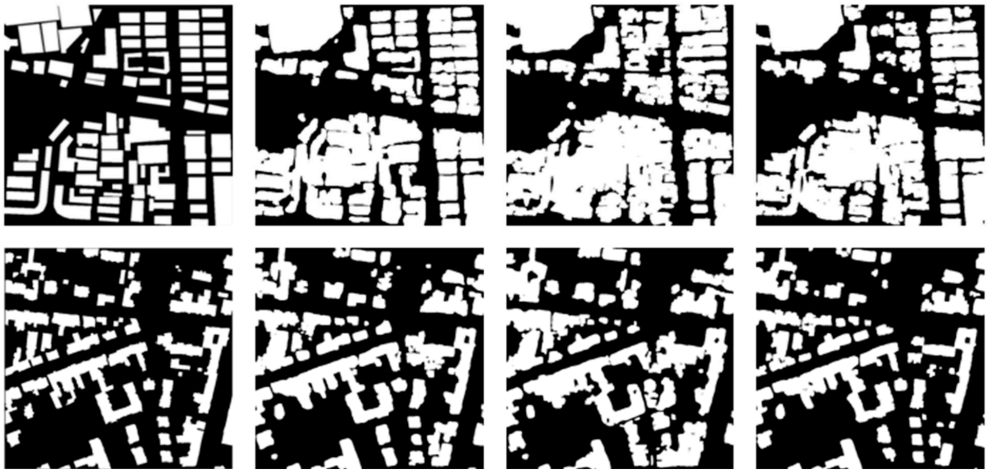
**Figure 2.** Two high-resolution remote sensing images for cross-domain testing. (**a**) The first image from Google Images at Shenzhen, China, with 500 × 500 pixels and a spatial resolution of 1 m for urban area. (**b**) The third image from ISPRS semantic segmentation dataset at Postdam, Germany, with 6000 × 6000 pixels and a spatial resolution of 0.05 m for suburban area.

Table 1 represents the Recall, Precision, and F1 of our AD-DAF framework on the three semantic segmentation models in two different target domains. It can be seen from this table that three models behave differently in different target domains. Specifically, the LinkNet model in our AD-DAF framework shows the largest increase in F1 metrics in the two target domains compared to the other two models, while the UNet model in our AD-DAF framework has the largest average F1 value of 87.05%. This indicates that our AD-DAF framework with the UNet model has the most stable effect for building extraction. Therefore, we choose the UNet model as our pre-trained model in the remaining three experiments. For the Precision metrics, our AD-DAF framework with the three models is mostly lower than their only pre-trained models. This is mainly because our method introduces some error classes while substantially increasing the recall. In addition, it also can be seen that the recall and F1 metrics using only the pre-trained model of the three semantic segmentation models are all lower than those of our proposed AD-DAF framework, with the value of F1 improving by at least 7.89% and at most 16.96%. These experimental results indicate that our AD-DAF framework is able to improve the migration ability of different deep segmentation models and is also applicable to different target domains.

**Table 1.** The results on the three semantic segmentation models in two different target domains.

| Images | Methods | UNet | | | PSPNet | | | LinkNet | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **Recall** | **Precision** | **F1** | **Recall** | **Precision** | **F1** | **Recall** | **Precision** | **F1** |
| (a) | Only Pre-trained | 81.63% | 75.31% | 78.34% | 62.27% | 79.82% | 69.96% | 53.42% | 86.01% | 65.91% |
| | AD-DAF | 92.16% | 81.65% | 86.59% | 93.49% | 72.97% | 81.96% | 90.86% | 76.17% | 82.87% |
| (b) | Only Pre-trained | 64.05% | 91.66% | 75.41% | 71.35% | 81.19% | 75.95% | 64.32% | 94.91% | 76.68% |
| | AD-DAF | 90.75% | 84.48% | 87.51% | 88.39% | 79.73% | 83.84% | 90.81% | 86.87% | 88.79% |

Figure 3 shows the building extraction effects of our AD-DAF framework with three different models in two different target domains. In this Figure, the first and second rows correspond to the results of the first image and the second image, respectively. The white area indicates the building footprint. As shown in Figure 3, our AD-DAF framework can extract some building objects under all the three models. Compared with the other two models, the building outlines extracted by our AD-DAF framework with the UNet model are clearer. This is consistent with the results of the quantitative metric F1 in Table 1.



(**a**) (**b**) (**c**) (**d**)

**Figure 3.** Semantic segmentation results of our framework with three different models in two different target domains. The first and second rows correspond to the results of the first image and the second image, respectively. The white area indicates the building footprint. (**a**) Ground truth. (**b**) Results of our framework based on UNet. (**c**) Results of our framework based on PSPNet. (**d**) Results of our framework based on LinkNet.

### *4.2. Experiment II: Large Area Applicability*

In this experiment, two images with relatively larger areas (about 1 $km^2$, 5201 × 5201 pixels) and with a spatial resolution of 0.2 m are used as two different target domains. The first image represents an urban area and is obtained from Google Images at Jilin, China, where the building types include residential areas, educational sites, industrial sites, and so on. The second image represents a suburban area and is also obtained from Google Images at Yunnan, China, in which the buildings are mainly low single bungalows. Their details are shown in Figure 4. The UNet-based pre-trained model by the IAILD dataset in experiment I is also used for the pre-trained model in this experiment.

(**a**) (**b**)

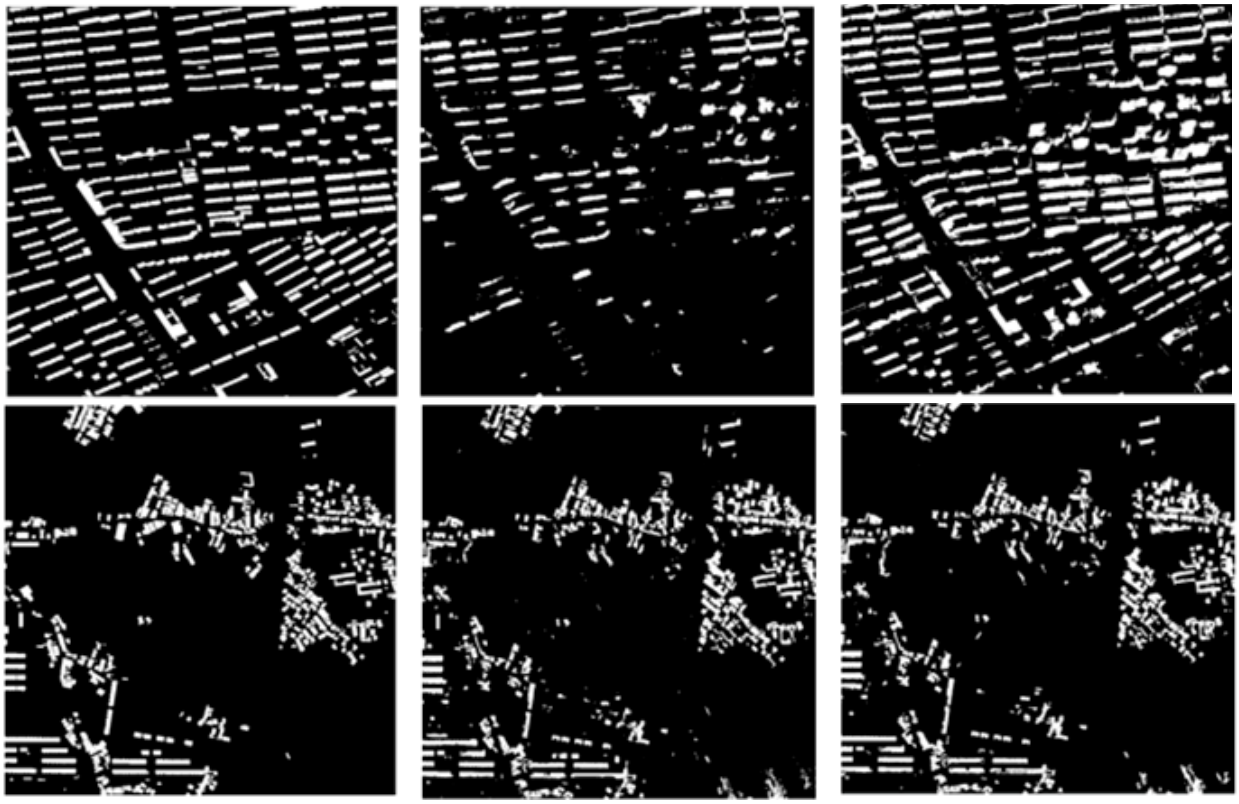
**Figure 4.** Two remote sensing images with larger area as target domains. (**a**) The first image for urban area. (**b**) The second image for suburban area.

Table 2 provides the Recall, Precision, and F1 of our AD-DAF framework with the UNet model on the two relatively larger areas. As can be seen from this table, our AD-DAF framework improves by 41.92% in the Recall metric and 26.93% in the F1 metric on the first image compared with the only pre-trained model. This is because the first image is slightly distorted and its angle is completely different from the angle of the image in the source domain of the IAILD dataset. This causes the pre-trained model based on the IAILD dataset to be less effective in extracting buildings in this image. In contrast, our method uses anomaly detection and reclassification to further improve the building extraction. On the second image, our method only improves by 9.42% in the Recall metric and 8.23% in the F1 metric. The above experimental results show that our AD-DAF framework is also effective under relatively large areas.

**Table 2.** The results of our framework with the UNet model on the two large areas.

| Images | Methods | Recall | Precision | F1 |
|---|---|---|---|---|
| (a) | Only Pre-trained | 42.97% | 72.94% | 54.08% |
| | AD-DAF | 84.89% | 77.46% | 81.01% |
| (b) | Only Pre-trained | 72.12% | 72.98% | 72.55% |
| | AD-DAF | 81.54% | 80.03% | 80.78% |

Figure 5 shows the building extraction results of the two methods on the two target domains. It is apparent from Figure 5 that the completeness of the segmentation results by our AD-DAF framework is significantly better than those of the only pre-trained model. This again indicates the advantage of our AD-DAF framework in the cross-domain building extraction. However, it should be noted that there is still a gap between the building outlines of our AD-DAF framework and the ground truth.



(**a**) (**b**) (**c**)

**Figure 5.** Semantic segmentation results on the larger areas. The white area indicates the building footprint. (**a**) Ground truth. (**b**) Results by using the only pre-trained model based on the UNet model. (**c**) Results of our framework with the UNet model.

### *4.3. Experiment III: Performance on the Benchmark Dataset*

In this experiment, the WHU satellite building dataset I [34] is selected as the target domain to verify the effectiveness of our AD-DAF framework. This dataset contains a total of 204 images which are collected from cities over the world. These images are from various remote sensing satellites such as QuickBird, IKONOS, and ZY-3, with the spatial resolution of 0.3 m to 2.5 m. Some example images are shown in Figure 6. Due to the diversity of sensors and imaging conditions in this dataset, it is challenging to extract buildings from this dataset, especially when this dataset is used as the target domain. These 204 images are all used as a test set for the target domain, which is consistent with the number of images used for the experiments in the literature. The pre-trained model still uses the UNet-based model with the source domain of the IAILD dataset already trained in Experiment I.

**Figure 6.** Some examples of WHU Satellite Dataset I. The lime green color indicates the building outline.

Table 3 presents the Recall, Precision, F1, and IoU of our AD-DAF framework with the UNet model on the WHU benchmark dataset. As can be seen from the table, the performance of the UNet model decreases substantially when migrating from the IAILD dataset to the WHU dataset and improves substantially when adding anomaly detection and reclassification to the pre-trained UNet model, with the Recall improving by 18.37%, the F1 improving by 8.7%, and the IOU improving by 9.94%. Furthermore, it can be seen from this table that the results of our AD-DAF framework are close to the testing results without the cross-domain of the WHU dataset, with only a 3.78% reduction in the F1 and 4.44% reduction in the IoU. This indicates that our added anomaly detection and reclassification process can indeed increase the effectiveness of the cross-domain building extraction. Figure 7 shows some building extraction results on the WHU Satellite Dataset I. Visually, the extraction results of our method do not differ much from the ground truth, and the differences are mainly in the building outlines.

**Table 3.** The results on the WHU satellite building dataset I.

| Images | Methods | Recall | Precision | F1 | IoU |
|---|---|---|---|---|---|
| WHU->WHU | UNet [22] | 73.30% | 73.10% | 73.20% | 57.70% |
| IAILD->WHU | Only pre-trained | 55.39% | 67.19% | 60.72% | 43.32% |
| IAILD->WHU | AD-DAF | 73.76% | 65.56% | 69.42% | 53.26% |

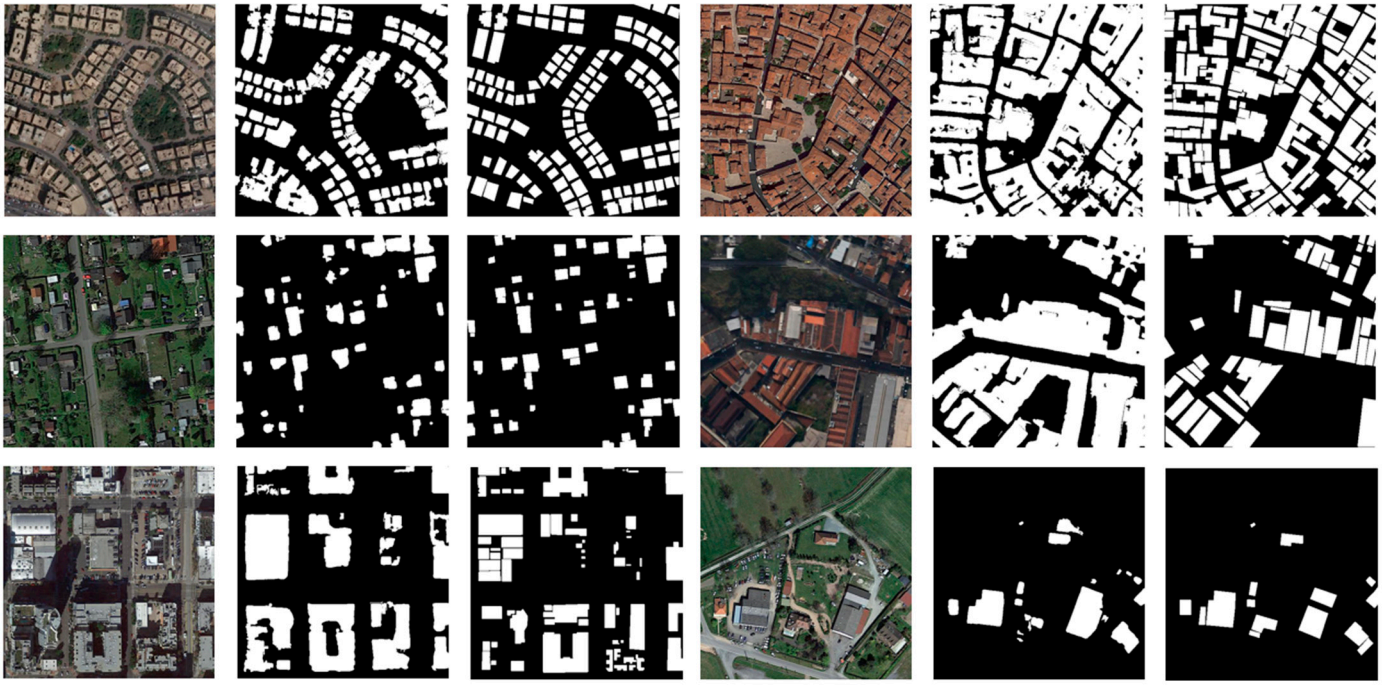



**Figure 7.** Some building extraction results on WHU Satellite Dataset I. The white area indicates the building footprint. (**a**,**d**) are the remote sensing image. (**b**,**e**) are the extraction results of our framework with the UNet model. (**c**,**f**) are the ground truth.

It should be noted that the pre-trained model based on the IAILD dataset has remained unchanged, while the target domain has changed in Experiments I, II, and III. More importantly, our method has been effective on different target domains yet with the same source domain. However, our AD-DAF framework is not retrained when the target domain changes. In contrast, the current mainstream domain adaptation methods need to be retrained with a large amount of unlabeled data in the new target domain. This further indicates that our AD-DAF framework can be applied to multiple target domains with only one pre-trained model.

### *4.4. Experiment IV: Comparisons with Other Domain Adaptation Methods*

To further evaluate the performance of our newly proposed framework, two other open access datasets of ISPRS Postdam semantic segmentation dataset (abbreviated as Postdam) [45] and the other WHU aerial building dataset are chosen as the source and target domains, respectively. The Postdam dataset consists of aerial orthorectified color imagery with a spatial resolution of 0.3 m. The other WHU aerial building dataset includes aerial images with a spatial resolution of 0.075 m covering in Christchurch, New Zealand. Four recently proposed adversarial learning-based domain adaptation methods of the DAugNet, DATA (Domain Adaptive Transfer Attack) [29], JPRNet (Joint Pixel and Representation level Network) [18], and FDANet (full-level domain adaptation network) [10] are selected for the comparisons. In the experiment, we divide the Postdam dataset into the training and validation sets in a ratio of 8:2 and use them to train the UNet pre-trained model in our AD-DAF framework. A total of 50 images are randomly selected from the WHU aerial dataset for cross-domain building extraction. In addition, F1 and IoU are chosen as the evaluation metrics.

Table 4 provides the comparison results from the source domain of the Postdam datasets to the target domain of the WHU aerial building dataset by different methods. As can be seen from this table, our F1 values are lower than the FDANet and DATA methods and higher than the DAUgNet and JPRNet methods, while our IoU values are only lower

than the FDANet method and better than the other three methods. The experimental results show that our AD-DF framework is comparable to some of the recent domain adaptation methods based on adversarial learning, but they also have some shortcomings with individual methods. However, the adversarial learning-based domain adaptation methods require a large amount of unlabeled data in the target domain to be retrained in concert with the labeled data in the source domain, which makes it necessary to retrain every time when the target domain changes. In contrast, our AD-DF framework only needs to be trained once based on the labeled data in the source domain to satisfy the requirements of multiple target domains and does not require a large amount of unlabeled data in the target domain. This again demonstrates the effectiveness of our anomaly detection and reclassification process.

**Table 4.** The comparison results from the source domain of the Postdam datasets to the target domain of the WHU aerial building dataset.

| Methods | Potsdam->WHU | |
|---|---|---|
| | F1 | IoU |
| DAugNet [10] | 0.7363 | 0.5827 |
| DATA [10] | 0.7951 | 0.6599 |
| JPRNet [10] | 0.7713 | 0.6277 |
| FDANet [10] | 0.8337 | 0.7148 |
| Our AD-DAF | 0.7862 | 0.6713 |

## 5. Conclusions

In this paper, we propose an anomaly detection-based domain adaptation framework for cross-domain building extraction called AD-DAF. Our proposed framework is composed of source domain-based pre-training, anomaly detection for the target domain, and local region reclassification. Our proposed framework leverages the anomaly detection idea to make full use of the feature knowledge of the target domain objects with the pre-classification knowledge of the source domain. Our framework has the advantage that a single pre-training can be used for multiple target domains at the same time. This is completely different from the mainstream domain adaptation methods that require retraining when the target domain is changed. The first three experimental results demonstrate that our framework can significantly improve the cross-domain building extraction of the pre-trained model, with the F1 improving by 8.7%, and the intersection over union improving by 9.94% when migrating from the Inria Aerial Image Labeling Dataset to the Wuhan University dataset. The last experimental result shows that our framework is comparable to the current mainstream domain adaptation methods.

Despite the advantages of the proposed framework, there is still much room for improvement in the accuracy and applicability of the method. First, the accuracy of our method is slightly lower than that of some adversarial learning-based domain adaptation methods. In fact, our newly proposed framework is not intended to replace the current domain adaptation methods but to complement them. On the one hand, our framework can provide initial pseudo-label samples for the mainstream supervised domain adaptation methods. On the other hand, our framework can be used as a post-processing step for mainstream deep learning semantic segmentation models and domain adaptation methods to further improve the effectiveness of cross-domain building extraction. Therefore, one of our future works will be to explore the possibility of integrating our framework with adversarial learning-based domain adaptation methods for a higher accuracy. Second, although we have performed method validation on different scenarios and benchmark datasets, these validations do not include all remote sensing data types because of the richness and diversity of remote sensing data types. Therefore, we will also validate the method on more benchmark datasets in the future. Finally, our method has more anomaly detection and reclassification steps than the mainstream deep semantic segmentation or

domain adaptation methods. To simplify the usage process, we intend to develop a building extraction web system using dynamic service computing technology in the future [46,47].

**Author Contributions:** Conceptualization, S.Z., X.Z. and D.H.; methodology, S.Z., X.Z. and D.H.; validation, S.Z.; writing—original draft preparation, S.Z. and D.H.; funding acquisition, X.Z. and D.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded in part by the National Natural Science Foundation of China under Grant 41971360 and Grant 42171457, in part by the Hunan Provincial Natural Science Foundation of China under Grant 2021JJ40721.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank ISPRS (https://www.isprs.org/education/benchmarks/UrbanSemLab/Default.aspx (accessed on 2 January 2023)), Inria Sophia Antipolis (https://project.inria.fr/aerialimagelabeling/ (accessed on 2 January 2023)) and Wuhan University (http://gpcv.whu.edu.cn/data/ (accessed on 2 January 2023)) for making the semantic segmentation datasets publicly available. The authors would also like to thank the anonymous reviewers for their comments to improve this paper.

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

## References

1. Yu, Y.; Ren, Y.; Guan, H.; Li, D.; Yu, C.; Jin, S.; Wang, L. Capsule Feature Pyramid Network for Building Footprint Extraction From High-Resolution Aerial Imagery. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 895–899. [CrossRef]
2. Liu, T.; Yao, L.; Qin, J.; Lu, N.; Jiang, H.; Zhang, F.; Zhou, C. Multi-scale attention integrated hierarchical networks for high-resolution building footprint extraction. *Int. J. Appl. Earth Obs.* **2022**, *109*, 102768. [CrossRef]
3. Zhao, H.; Zhang, H.; Zheng, X. A multiscale attention-guided UNet++ with edge constraint for building extraction from high spatial resolution imagery. *Appl. Sci.* **2022**, *12*, 5960. [CrossRef]
4. Huang, H.; Chen, Y.; Wang, R. A lightweight network for building extraction from remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]
5. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
6. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
7. Yu, B.; Yang, A.; Chen, F.; Wang, N.; Wang, L. SNNFD, spiking neural segmentation network in frequency domain using high spatial resolution images for building extraction. *Int. J. Appl. Earth Obs.* **2022**, *112*, 102930. [CrossRef]
8. Feng, D.; Xie, Y.; Xiong, S.; Hu, J.; Hu, M.; Li, Q.; Zhu, J. Regularized Building Boundary Extraction From Remote Sensing Imagery Based on Augment Feature Pyramid Network and Morphological Constraint. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 12212–12223. [CrossRef]
9. Wang, Y.; Zeng, X.; Liao, X.; Zhuang, D. B-FGC-Net: A Building Extraction Network from High Resolution Remote Sensing Imagery. *Remote Sens.* **2022**, *14*, 269. [CrossRef]
10. 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.* **2022**, *60*, 1–17. [CrossRef]
11. Yao, X.; Wang, Y.; Wu, Y.; Liang, Z. Weakly-supervised domain adaptation with adversarial entropy for building segmentation in cross-domain aerial imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8407–8418. [CrossRef]
12. Liu, W.; Su, F. Unsupervised adversarial domain adaptation network for semantic segmentation. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1978–1982. [CrossRef]
13. Wang, J.; Ma, A.; Zhong, Y.; Zheng, Z.; Zhang, L. Cross-sensor domain adaptation for high spatial resolution urban land-cover mapping: From airborne to spaceborne imagery. *Remote Sens. Environ.* **2022**, *277*, 113058. [CrossRef]
14. Chen, J.; Zhu, J.; Guo, Y.; Sun, G.; Zhang, Y.; Deng, M. Unsupervised Domain Adaptation for Semantic Segmentation of High-Resolution Remote Sensing Imagery Driven by Category-Certainty Attention. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [CrossRef]
15. Teng, W.; Wang, N.; Shi, H.; Liu, Y.; Wang, J. Classifier-constrained deep adversarial domain adaptation for cross-domain semisupervised classification in remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 789–793. [CrossRef]

16. Dias, P.; Tian, Y.; Newsam, S.; Tsaris, A.; Hinkle, J.; Lunga, D. Model Assumptions and Data Characteristics: Impacts on Domain Adaptation in Building Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [CrossRef]
17. Tasar, O.; Giros, A.; Tarabalka, Y.; Alliez, P.; Clerc, S. DAugNet: Unsupervised, multisource, multitarget, and life-long domain adaptation for semantic segmentation of satellite images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1067–1081. [CrossRef]
18. Shi, L.; Wang, Z.; Pan, B.; Shi, Z. An end-to-end network for remote sensing imagery semantic segmentation via joint pixel-and representation-level domain adaptation. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1896–1900. [CrossRef]
19. Hou, D.; Wang, S.; Tian, X.; Xing, H. PCLUDA: A Pseudo-label Consistency Learning-Based Unsupervised Domain Adaptation Method for Cross-domain Optical Remote Sensing Image Retrieval. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–14. [CrossRef]
20. Sun, G.; Huang, H.; Zhang, A.; Li, F.; Zhao, H.; Fu, H. Fusion of multiscale convolutional neural networks for building extraction in very high-resolution images. *Remote Sens.* **2019**, *11*, 227. [CrossRef]
21. Hui, J.; Du, M.; Ye, X.; Qin, Q.; Sui, J. Effective building extraction from high-resolution remote sensing images with multitask driven deep neural network. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 786–790. [CrossRef]
22. Liu, Y.; Zhou, J.; Qi, W.; Li, X.; Gross, L.; Shao, Q.; Zhao, Z.; Ni, L.; Fan, X.; Li, Z. ARC-Net: An efficient network for building extraction from high-resolution aerial images. *IEEE Access* **2020**, *8*, 154997–155010. [CrossRef]
23. Hou, D.; Wang, S.; Tian, X.; Xing, H. An Attention-Enhanced End-to-End Discriminative Network With Multiscale Feature Learning for Remote Sensing Image Retrieval. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8245–8255. [CrossRef]
24. Chen, Z.; Li, D.; Fan, W.; Guan, H.; Wang, C.; Li, J. Self-attention in reconstruction bias U-Net for semantic segmentation of building rooftops in optical remote sensing images. *Remote Sens.* **2021**, *13*, 2524. [CrossRef]
25. Guo, M.; Liu, H.; Xu, Y.; Huang, Y. Building extraction based on U-Net with an attention block and multiple losses. *Remote Sens.* **2020**, *12*, 1400. [CrossRef]
26. Krichen, M.; Mihoub, A.; Alzahrani, M.Y.; Adoni, W.Y.H.; Nahhal, T. Are Formal Methods Applicable To Machine Learning And Artificial Intelligence? In Proceedings of the 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia, 22–24 May 2022; pp. 48–53.
27. Song, S.; Yu, H.; Miao, Z.; Zhang, Q.; Lin, Y.; Wang, S. Domain adaptation for convolutional neural networks-based remote sensing scene classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1324–1328. [CrossRef]
28. Zhao, S.; Zhang, Z.; Guo, W.; Luo, Y. An Automatic Ship Detection Method Adapting to Different Satellites SAR Images with Feature Alignment and Compensation Loss. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [CrossRef]
29. Na, Y.; Kim, J.H.; Lee, K.; Park, J.; Hwang, J.Y.; Choi, J.P. Domain adaptive transfer attack-based segmentation networks for building extraction from aerial images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5171–5182. [CrossRef]
30. Liu, J.; Xuan, W.; Gan, Y.; Zhan, Y.; Liu, J.; Du, B. An End-to-end Supervised Domain Adaptation Framework for Cross-Domain Change Detection. *Pattern Recognit.* **2022**, *132*, 108960. [CrossRef]
31. Zhang, J.; Liu, J.; Pan, B.; Shi, Z. Domain adaptation based on correlation subspace dynamic distribution alignment for remote sensing image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7920–7930. [CrossRef]
32. Ji, S.; Wang, D.; Luo, M. Generative adversarial network-based full-space domain adaptation for land cover classification from multiple-source remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 3816–3828. [CrossRef]
33. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Worth, TX, USA, 23–28 July 2017; pp. 3226–3229.
34. Ji, S.; Wei, S.; Lu, M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 574–586. [CrossRef]
35. Chaurasia, A.; Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
36. Kimura, M. Understanding Test-Time Augmentation. In Proceedings of the International Conference on Neural Information Processing, Sanur, Indonesia, 8–12 December 2021; pp. 558–569.
37. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. Deepglobe 2018: A challenge to parse the earth through satellite images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 172–181.
38. Chabchoub, Y.; Togbe, M.U.; Boly, A.; Chiky, R. An in-depth study and improvement of Isolation Forest. *IEEE Access* **2022**, *10*, 10219–10237.
39. Cheng, Z.; Zou, C.; Dong, J. Outlier detection using isolation forest and local outlier factor. In Proceedings of the Conference on Research in Adaptive and Convergent Systems, Chongqing, China, 24–27 September 2019; pp. 161–168.
40. Lesouple, J.; Baudoin, C.; Spigai, M.; Tourneret, J.-Y. Generalized isolation forest for anomaly detection. *Pattern Recogn. Lett.* **2021**, *149*, 109–119. [CrossRef]
41. Wei, D.; Hou, D.; Zhou, X.; Chen, J. Change Detection Using a Texture Feature Space Outlier Index from Mono-Temporal Remote Sensing Images and Vector Data. *Remote Sens.* **2021**, *13*, 3857.
42. Tao, W.; Jin, H.; Zhang, Y. Color image segmentation based on mean shift and normalized cuts. *IEEE Trans. Syst. Man Cybern. Part B (Cybernetics)* **2007**, *37*, 1382–1389. [CrossRef] [PubMed]

43. Johnson, B.A.; Iizuka, K.; Bragais, M.A.; Endo, I.; Magcale-Macandog, D.B. Employing crowdsourced geographic data and multi-temporal/multi-sensor satellite imagery to monitor land cover change: A case study in an urbanizing region of the Philippines. *Comput. Environ. Urban Syst.* **2017**, *64*, 184–193.
44. Yu, M.; Zhang, W.; Chen, X.; Liu, Y.; Niu, J. An End-to-End Atrous Spatial Pyramid Pooling and Skip-Connections Generative Adversarial Segmentation Network for Building Extraction from High-Resolution Aerial Images. *Appl. Sci.* **2022**, *12*, 5151. [CrossRef]
45. Markus Gerke, I. *Use of the Stair Vision Library within the ISPRS 2D Semantic Labeling Benchmark (Vaihingen)*; ResearcheGateP: Brlin, Germany, 2014.
46. Xing, H.; Liu, C.; Li, R.; Wang, H.; Zhang, J.; Wu, H. Domain Constraints-Driven Automatic Service Composition for Online Land Cover Geoprocessing. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 629. [CrossRef]
47. Xing, H.; Chen, J.; Wu, H.; Zhang, J.; Li, S.; Liu, B. A service relation model for web-based land cover change detection. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 20–32.


**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.