

Article

Not peer-reviewed version

A Semi-Supervised Remote Sensing Image Change Detection Network Based on Consistent Regularization for Multi-Scale feature Fusion and Multiplexing

[Yongjun Qi](#) , [Zhengzheng Jia](#) , [Yuanmeng Song](#) , [Shaohua Yang](#) , [Jie Zhu](#) ^{*} , [Xin Liu](#) , [Hongxing Zheng](#) ^{*}

Posted Date: 24 October 2024

doi: 10.20944/preprints202410.1886.v1

Keywords: Remote sensing; Change detection; Semi-supervised learning; Multi-scale feature fusion and multiplexing; Consistent regularization



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

A Semi-Supervised Remote Sensing Image Change Detection Network Based on Consistent Regularization for Multi-Scale Feature Fusion and Multiplexing

Yongjun Qi ¹, Zhengzheng Jia ¹, Yuanmeng Song ¹, Shaohua Yang ², Jie Zhu ^{1,*}, Xin Liu ³
and Hongxing Zheng ^{1,*}

¹ School of Computer Science and Engineering of North China Institute of Aerospace Engineering, Langfang 065000, China; qyjun@189.cn

² School of Aeronautics and Astronautics of North China Institute of Aerospace Engineering, Langfang 065000, China; ysh0130@stumail.nciae.edu.cn

³ School of Optoelectronic Engineering, Xidian University, Xi'an 710071, China; xinliu@mail.xidian.edu.cn

* Correspondence: zhujie0424@126.com (J.Z.); hxzheng@hebut.edu.cn (H.Z.); Tel.: 15613628683 (J.Z.)

Abstract: Remote sensing image change detection is a key application technology in remote sensing. In recent years, the development of deep learning has brought new opportunities for remote sensing image change detection. However, to achieve high change detection accuracy, large-scale, good-quality labeled data is required to train the model, which is often difficult to obtain and costly. This paper proposes a semi-supervised remote sensing image change detection network based on consistent regularization for multi-scale feature fusion and reuse (SCMFM-CDNet), which can enhance change detection performance with limited labeled data. Specifically, the SCMFM-CDNet consists of two parts, the supervised stage multi-scale feature fusion and reuse network (MFM-CDNet) and the consistency regularization method of semi-supervised learning. The MFM-CDNet network uses channel attention mechanism and spatial attention mechanism to enhance the expression ability of features, and gradually completes the extraction and detection of change features through multi-scale feature fusion and reuse strategy. The consistency regularization method of semi-supervised learning adopts the teacher model and student model for alternate training, updating model parameters by minimizing the differences in their predictions, thus achieving semi-supervised learning. This paper conducted extensive experiments on three public datasets, LEVIR-CD, WHU-CD, and GoogleGZ-CD. Under the three labeling data ratios of 5%, 10%, and 20%, the F1 scores of the SCMFM-CDNet network on the LEVIR-CD dataset were 88.62, 90.78, and 90.63, respectively. On the WHU-CD dataset, the F1 scores were 88.40, 90.17, and 90.83, respectively. On the GoogleGZ-CD dataset, the F1 scores were 81.44, 82.36, and 85.40, respectively. The results indicate that the proposed network outperforms existing semi-supervised change detection methods in terms of comprehensive performance. In addition, this paper also conducted ablation experiments to evaluate the contribution of the global context module (GCM) and feature fusion reuse module (FFM) in the MFM-CDNet network. The experimental results show that these two modules are crucial for improving network performance.

Keywords: remote sensing; change detection; semi-supervised learning; multi-scale feature fusion and multiplexing; consistent regularization

1. Introduction

Remote sensing image change detection is a key application of remote sensing technology, which identifies changes in the surface or landscape by analyzing remote sensing images of the same

geographical area acquired at different time points. This technology can provide up-to-date geographic information and is widely used in various fields such as urban planning [1,2], environmental monitoring [3], agricultural monitoring [4,5], and natural disaster management [6,7]. In recent years, deep learning has been widely applied in various fields of remote sensing, such as image fusion [8], semantic segmentation [9], and change detection.

Remote sensing image change detection methods are mainly divided into three categories: supervised, unsupervised, and semi-supervised. Supervised methods rely on a large amount of labeled data for training and can perform well when there is sufficient labeled data, accurately identifying the areas of change in the images. However, the acquisition of labeled data is costly, especially in high-resolution remote sensing images, where manual annotation is complex and time-consuming. Unsupervised methods do not require labeled data and rely on differences in image features to detect changes. Although they reduce the dependence on labeled data, due to the lack of supervisory signals, the models struggle to capture complex change patterns and are prone to noise interference, leading to lower detection accuracy, which is not sufficient to meet the needs of high-precision applications. Semi-supervised methods combine the advantages of both, enhancing the model's detection capability by utilizing the potential information of unlabeled data when labeled data is insufficient. It not only reduces the cost of labeling but also effectively addresses the issues of unclear change edge detection and the omission of extremely small targets through strategies such as regularization and multi-task learning. Especially in dealing with complex change detection tasks in high-resolution remote sensing images, semi-supervised methods can improve the model's generalization ability, reduce the risk of overfitting, and significantly enhance detection accuracy.

Therefore, in order to fully exploit the limited labeled data and a large amount of unlabeled data, to improve detection performance with minimal model runtime and inference time, and to further address the issues of unclear detection of change edges and the omission of extremely small targets, this paper proposes a Semi-supervised Remote Sensing Image Change Detection Network Based on Consistent Regularization for Multi-scale feature Fusion and Multiplexing (SCMFM-CDNet). The network mainly consists of two parts: the supervised stage change detection network (MFM-CDNet) and the consistency regularization method used to achieve semi-supervised learning. The main contributions of this paper are:

(1) This paper proposes a semi-supervised remote sensing image change detection network SCMFM-CDNet, which includes the supervised MFM-CDNet and the semi-supervised consistency regularization method. The supervised part enhances the recognition ability of the change areas through the effective fusion and reuse of multi-scale features, especially in the detection of small targets and edge details. At the same time, the semi-supervised part uses consistency regularization to strengthen the feature learning of unlabeled data, improving the model's detection performance under limited labeled data and enhancing the model's generalization ability.

(2) In the supervised part of SCMFM-CDNet, the GCM module and the FFM module have been added. The GCM module enhances the detection capability for extremely small targets by integrating features from different levels. The FFM module improves the localization accuracy of edge change areas and reduces the problem of unclear detection of change edges by fusing, refining, and reusing features from different levels.

(3) Conducted extensive experiments on the LEVIR-CD, WHU-CD, and GoogleGZ-CD datasets, including three label ratios of 5%, 10%, and 20%, the proposed SCMFM-CDNet network has demonstrated significant performance in both training and inference, proving the effectiveness and advantages of the proposed method.

The structure of the rest of this paper is as follows. In Section 2, a review of the related literature on existing change detection methods is presented. The proposed method is introduced in Section 3. The experimental results and analysis are provided in Section 4. Finally, the conclusions are summarized in Section 5.

2. Related Work

2.1. Supervised Change Detection Methods

Supervised change detection methods rely on annotated training data, among which the application of convolutional neural networks (CNN) and U-Net and other architectures has achieved significant results. For example, the VGG network [10] proposed by Simonyan and Zisserman and the ResNet [11] proposed by He et al. have performed well in image classification and have become fundamental models in change detection.

The FCN model [12] proposed by Long et al. successfully achieved pixel-level semantic segmentation by replacing fully connected layers with convolutional layers, providing strong support for change detection. The U-Net structure, with its symmetric encoder-decoder design, can effectively capture detailed information in images, making it widely popular in medical image segmentation and remote sensing change detection [13].

The dynamic Siamese network [14] proposed by Guo et al. provides a new perspective for visual object tracking and can be extended to change detection tasks. This method processes two images by sharing weights, using similar feature extraction and similarity measurement to generate change maps, significantly improving the accuracy of change detection.

In addition, the combination of feature selection and deep learning has also promoted the progress of change detection. By introducing a multi-scale fusion mechanism into the Siamese network, the multi-scale fusion Siamese network (MSF-Siamese) proposed by Li et al. effectively uses the different scale features of images to enhance change detection performance [15]. On this basis, the AMFNet proposed by Zhan et al. uses the attention mechanism to guide multi-scale feature fusion, further improving the accuracy of change detection in remote sensing images [16].

To overcome the problem of data imbalance, improved loss functions are often used in supervised methods, such as weighted cross-entropy loss and focal loss [17]. These methods adjust the weights of different categories during the training process to improve the detection capability of change areas. In addition, for the sparsely populated change areas of pixels, the multi-scale method using spectral and texture features by Fonseca et al. has also become an effective solution, promoting the accuracy and stability of change detection [18].

2.2. Unsupervised Change Detection Methods

Unsupervised change detection methods do not rely on annotated data. Autoencoders [19] and Generative Adversarial Networks (GANs) [20,21] are among the most promising techniques currently used in unsupervised change detection. Autoencoders perform change detection by learning the latent features of images, while GANs, through the adversarial training of generators and discriminators, can better capture the change information in images.

The application of autoencoders in unsupervised learning allows models to obtain deep feature representations by learning to reconstruct the input images. The SNUNet-CD proposed by Fang et al., which uses a densely connected Siamese network structure, can achieve change detection of high-resolution remote sensing images in an unsupervised setting and has shown good performance [22]. In addition, the fully convolutional Siamese network proposed by Daudet et al. is also suitable for change detection tasks and has demonstrated its potential in remote sensing image analysis [23].

The advantage of unsupervised change detection is its low dependence on annotated data, which enables effective change detection even in the absence of annotations. The downside is that the detection accuracy is relatively low. Researchers are actively exploring the potential of unsupervised learning, especially in the analysis of high-resolution remote sensing images.

2.3. Semi-Supervised Change Detection Methods

In semi-supervised learning, feature fusion techniques are widely applied, and the combination of multi-scale features can extract richer information, thereby enhancing the detection capability of change areas in complex scenes. The Siamese network based on multi-scale attention proposed by Li et al. demonstrated the potential of using different feature scales for change detection in a semi-

supervised setting [24]. Wang et al. proposed a change detection method based on existing vector polygons and the latest imagery, which introduces an attention mechanism to improve detection accuracy [25].

Adversarial learning methods have also shown good results in semi-supervised change detection. The AdvNet method proposed by Hung et al. utilizes a semi-supervised semantic segmentation method based on adversarial networks, where the discriminator enhances the accuracy of semantic segmentation by combining adversarial loss and standard cross-entropy loss, and is appropriately modified for change detection tasks [26]. This method was first proposed in 2019. The SemiCD method proposed by Wele and Patel adopts a generative adversarial network-based framework, including a segmentation network and two discriminator networks, for change detection in high-resolution remote sensing images [27].

Additionally, multi-task learning strategies have shown good results in semi-supervised change detection. By training multiple related tasks simultaneously, the model can obtain stronger feature representation capabilities. The dual-branch network proposed by Maa et al. effectively detects changes in remote sensing images through a dual-branch structure, fully exploring the potential information between labeled and unlabeled data [28]. The TCNet method proposed by Shu et al. combines convolutional neural networks (CNN) and visual transformers for processing change detection in high-resolution remote sensing images [29]. Wang et al. designed a change region-based contrast loss function (RCL method) to enhance the model's feature extraction capability for change objects and uses the uncertainty of unlabeled data to select reliable pseudo-labels for model training [30].

Recently, Yang et al. proposed the UniMatch-DeepLabv3+ method at CVPR 2023, which is a deep learning model combining semi-supervised learning and semantic segmentation [31]. The core innovation lies in unifying semi-supervised learning methods, significantly improving the performance of semantic segmentation by combining labeled and unlabeled data. This method has shown good flexibility and accuracy in practical applications, especially in data-scarce situations.

With in-depth research, semi-supervised methods show a promising application prospect in remote sensing change detection. Combined with the latest deep learning technologies, these methods can more effectively utilize limited annotated data and improve overall performance when dealing with change detection tasks in high-resolution remote sensing images.

3. Methodology

The overall architecture of the SCMFM-CDNet model is shown in Figure 1. It mainly consists of two parts: the supervised multi-scale feature fusion and reuse network MFM-CDNet, and the semi-supervised learning consistency regularization method based on the teacher-student model for updating parameters. The working principle of semi-supervised consistency regularization is to use the teacher model to generate pseudo-labels for unlabeled data, and to add consistency loss and regularization terms when training the student model. This ensures that the model's predictions for the data and its perturbed versions remain consistent, thereby forcing the model to learn more generalized feature representations and iteratively updating model parameters after performance improvement. Initially, MFM-CDNet is used for supervised learning on labeled data to generate initial parameters. Then, for unlabeled data, the teacher-student model is used for iterative training for semi-supervised learning. The main idea is to use the teacher model to produce pseudo-labels to guide the student model to learn the features of unlabeled data, thus achieving the effect of semi-supervised learning. The quality of pseudo-labels plays a key role in the performance of the final model.

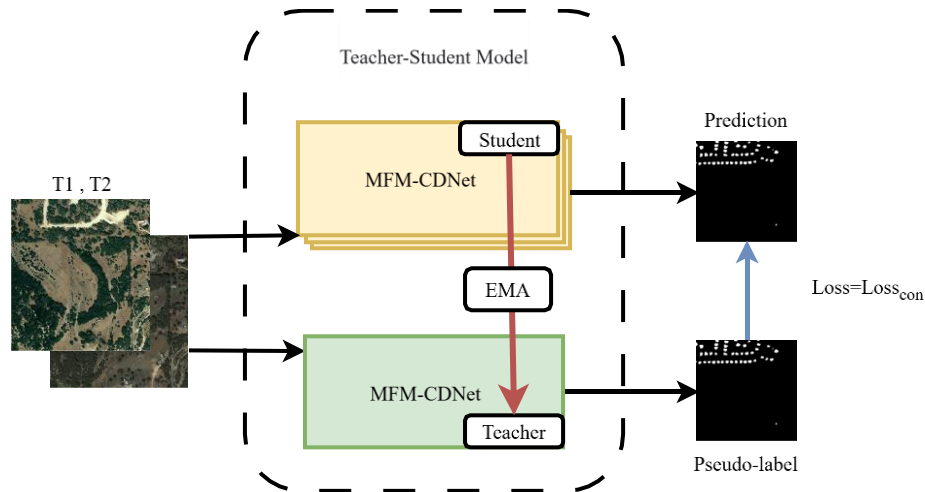


Figure 1. Overall architecture of the SCMFM-CDNet model.

3.1. MFM-CDNet Structure

MFM-CDNet consists of an encoder and a decoder parts. The former uses the VGG-16 network as the backbone, employing 5 VGG16_BN blocks, which correspond to the operations of VGG16_BN at layers 0-5, 5-12, 12-22, 22-32, and 32-42. Features are extracted at different stages to achieve feature fusion from coarse to fine. In the decoder part, the Feature Enhancement Module (FEM), Global Context Module (GCM), and Feature Fusion and Reuse Module (FFM) are used in combination. The overall structure of the MFM-CDNet model is shown in Figure 2.

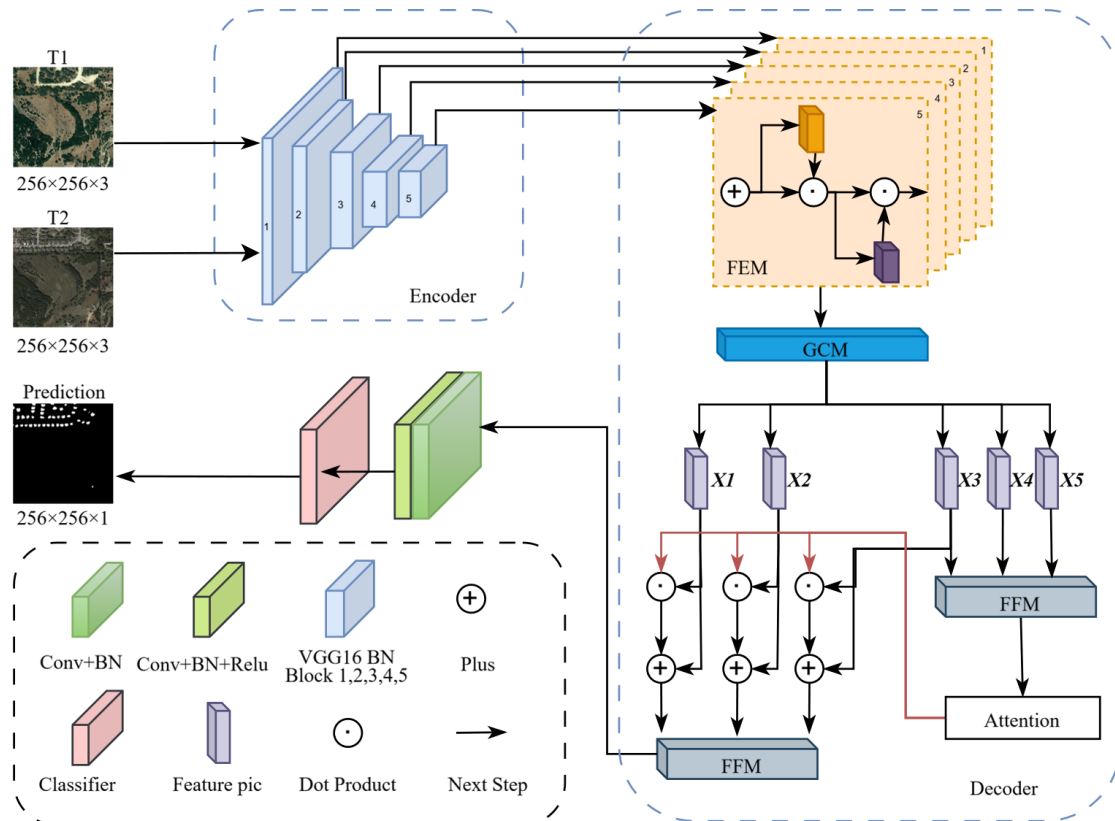


Figure 2. Overall architecture of the MFM-CDNet model.

(1) Feature Enhancement Module (FEM)

The FEM primarily guides the attention by sequentially passing the feature maps through a channel attention module and a spatial attention module, enabling the model to accurately locate important areas in the image and significantly reduce computational complexity.

The channel attention module can highlight the most important feature channels in the image, enabling the model to obtain more refined and distinctive feature representations, and reducing the computational load on unimportant features. The channel attention mechanism mainly addresses the question of ‘what the network should focus on’. The spatial attention module allows the network to focus on key areas within the image, primarily addressing the question of ‘where the network should concentrate’.

(2) Global Context Module (GCM)

The main function of the GCM is to integrate multi-scale contextual information of images to enhance the expressive power of features. In this way, the model can not only capture local details but also integrate features from different levels, which can improve the detection ability of extremely small targets.

The GCM consists of four parts. The first part is a 2D convolutional unit, while the second to fourth parts each contain four 2D convolutional units. Each 2D convolutional unit is composed of a convolutional layer and a batch normalization layer, but they have different kernel sizes and dilation factors. The outputs of these four units are horizontally concatenated, and then placed together with the output of the feature map obtained directly through the 2D convolution operation into the ReLU activation function, ultimately resulting in a feature-enhanced map that contains multi-scale contextual information. The detailed architecture of the GCM module is shown in Figure 3.

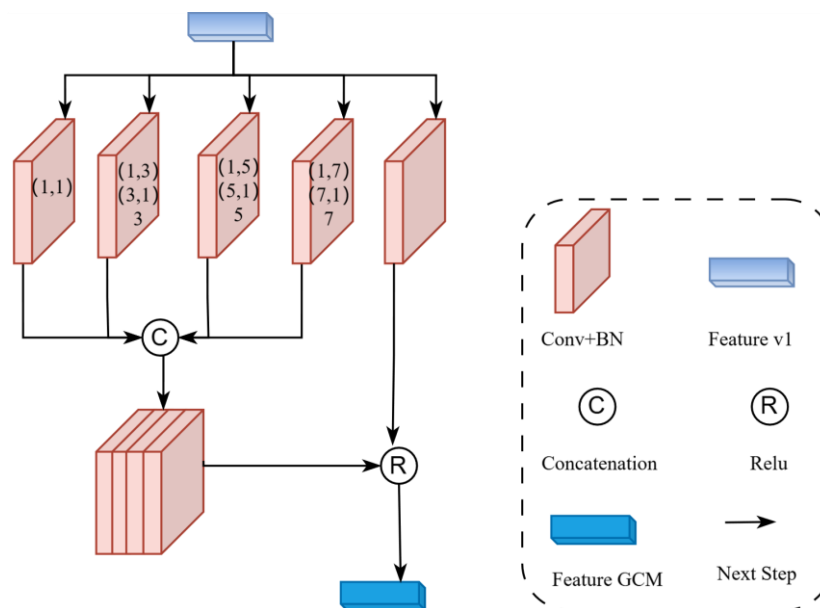


Figure 3. Architecture of the GCM module.

(3) Feature Fusion and Multiplexing Module (FFM)

The FFM primarily fuses, refines, and reuses features. Its main function is to integrate and optimize feature information from different levels. Through a series of operations, it enhances the model’s ability to recognize positive samples, improves the localization accuracy of edge change areas, and reduces the occurrence of missed detections. This module mainly correlates the features with the attention maps processed by the FFM, which primarily refines the decoding of features from three stages to enhance the accuracy of change information. The detailed operation involves bilinear upsampling of the input feature $X1$, multiplying the upsampled feature with another layer of features $X2$, bilinearly upsampling the input feature $X2$ and multiplying the upsampled feature with another layer of features $X3$, bilinearly upsampling the input feature $X1$ twice and multiplying the upsampled feature with another layer of features $X3$, horizontally concatenating the results of the three

operations, and then passing them through two 3x2D two-dimensional convolutional modules to produce the final decoded value of FFM. This results in an aggregated and refined feature map with higher spatial resolution and stronger feature expression capabilities. The detailed architecture of the FFM is shown in Figure 4.

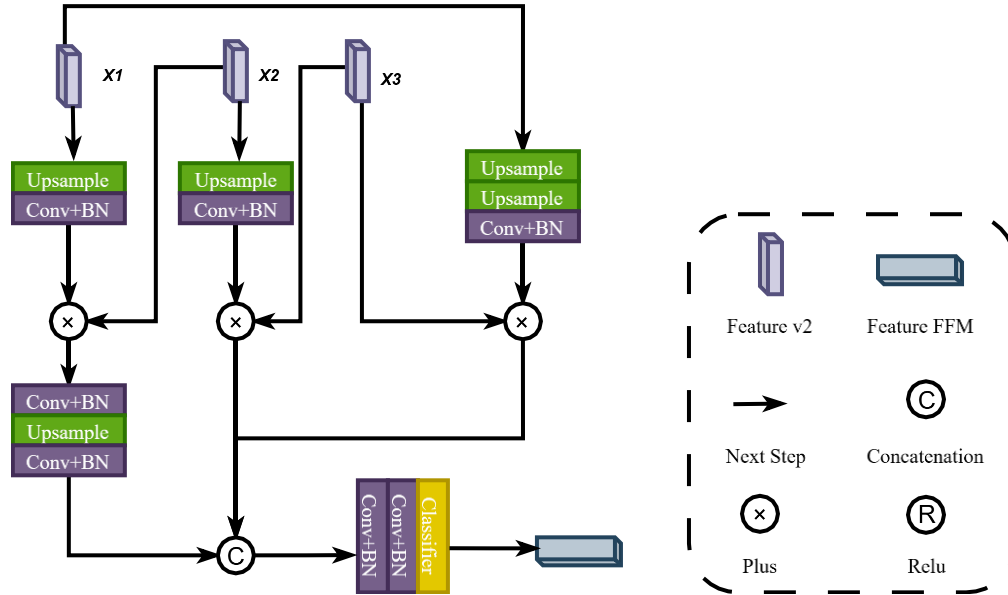


Figure 4. Architecture of the FFM module.

3.2. Teacher-Student Module

In the semi-supervised learning task, for unlabeled data, this paper employs the mean teacher method to learn features. The main idea is to switch between the roles of the student model and the teacher model. When using the backbone network as the teacher model, it primarily generates pseudo-labels for the student model. When using the backbone network as the student model, it learns from the pseudo-labels generated by the teacher model. The parameters of the teacher model are obtained by taking a weighted average of the parameters from previously iterated training of the student model. The target labels for the unlabeled data come from the predictions of the teacher model, i.e., the pseudo-labels. Since the pseudo-labels generated by the teacher model can serve as prior information in later training, they guide the student model in learning the features of a large amount of unlabeled data, thereby achieving the effect of semi-supervised learning.

The parameters of the teacher model are updated through EMA of the parameters from the student model, with the policy as follows.

$$\theta'_t = \alpha \theta'_{t-1} + (1 - \alpha) \theta_s, \quad (1)$$

where, θ'_t represents the parameters of the teacher model at the current step, θ'_{t-1} represents the parameters of the teacher model at the previous step, θ_s represents the parameters of the student model, and α is a smoothing factor set to 0.9, which controls the smoothness of the changes in the teacher model's parameters.

3.3. Loss Function

Change detection tasks are considered a special type of binary classification problem. In this paper, a binary cross-entropy loss function is used in conjunction with the Sigmoid function and BCELoss. The supervised loss function is:

$$Loss_{sup} = \{l_1, \dots, l_n\}, \quad (2)$$

$$l_n = -[y_n \cdot \log(\alpha(X_n)) + (1 - y_n) \cdot \log(1 - \alpha(X_n))], \quad (3)$$

$$\alpha(X) = \frac{1}{1 + \exp(-x)}, \quad (4)$$

$\alpha(X)$ is a Sigmoid function that maps x to the interval (0,1).

In the semi-supervised part, the input is the result of supervised training, and the semi-supervised loss function is as follows:

$$Loss_{con} = -[\hat{y}_t \cdot \log(\alpha(X_n)) + (1 - \hat{y}_t) \cdot \log(1 - \alpha(X_n))], \quad (5)$$

$$\hat{y}_t = \frac{1}{1 + \exp(-\hat{x}_t)}. \quad (6)$$

$$Loss = Loss_{sup} + \beta Loss_{con}. \quad (7)$$

β represents the difference in coefficients between the two loss functions, and in this case, β is used with a value of 0.2, at which point the impact of this coefficient on model accuracy is minimal.

3.4. Training Process

The training process of the model is shown in Table 1.

Table 1. Pseudo Code for Model Training.

Algorithm SCMFM-CDNet	
Input:	
Labelled image (X_l^1, X_l^2) , Unlabelled image (X_u^1, X_u^2) ,	
Hyperparameters : lr, B, T, α	
Initialise: MFM-CDNet model parameters θ , teacher model parameters θ_t	
Output:	
Trained SCMFM-CDNet model with parameters θ	
1: procedure Semi-Supervised Training θ, α	
2: for each=1 to T_{sup} do	
3: Shuffle (X_l^1, X_l^2)	
4: for batch B in (X_l^1, X_l^2) do	
5: Forward pass: Compute predictions P_l using (X_l^1, X_l^2) and θ	
6: Compute $Loss_{sup}$ by (Eq.(2)-(3))	
7: Backward pass: Update θ using $Loss_{sup}$ and lr	
8: end for	
9: end for	
10: for each= T_{sup} to T do	
11: for batch B in (X_u^1, X_u^2) do	
12: Forward pass: Compute predictions P_u using (X_l^1, X_l^2) and θ	
13: Compute $Loss_{con}$ using P_u and θ_t	
14: Backward pass: Update θ using $Loss_{con}$ and lr	
15: end for	
16: Update θ_t by (Eq.(1))	
17: end for	
18: end procedure	
19: return θ	

4. Experiments

In this section, this paper will validate the effectiveness of the proposed method through extensive experiments, which mainly include dataset introduction, experimental setup, comparison methods, experimental results and analysis, and ablation experiments.

4.1. Dataset Introduction

This experiment utilizes three public datasets, which are LEVIR-CD, WHU-CD, and GoogleGZ-CD.

The LEVIR-CD dataset is provided by Beihang University and is a large-scale Very High Resolution (VHR) building change detection dataset. It consists of 637 pairs of high-resolution image blocks, with a spatial resolution of 0.5 meters and an image size of 1024×1024 pixels. The dataset focuses on building-related changes, including various types of structures such as villas, high-rise apartments, small garages, and large warehouses.

The WHU-CD dataset is provided by Wuhan University and is a high-resolution remote sensing image dataset for building change detection. The dataset includes a large image pair sized at 32507×15345 pixels with a spatial resolution of 0.2 meters. It reflects significant reconstruction activities in the area following an earthquake, with change information primarily focused on buildings. The images are typically cropped into non-overlapping blocks of 256×256 pixels.

The GoogleGZ-CD dataset is provided by Wuhan University and is a large-scale high-resolution multispectral satellite image dataset specifically for change detection tasks. The dataset contains 20 pairs of seasonal high-resolution images covering the suburban areas of Guangzhou, China, with an acquisition time span from 2006 to 2019. The image resolution is 0.55 meters, and the size varies from 1006×1168 pixels to 4936×5224 pixels. It covers various types of changes, including water bodies, roads, farmland, bare land, forests, buildings, and ships, with buildings being the main changes. The image pairs are cropped into non-overlapping blocks of 256×256 pixels.

The detailed number of samples in the training set, validation set, and test set of the three datasets is shown in Table 2.

Table 2. Detailed Information of Datasets.

Title 1	Training Set Samples	Validation Set Samples	Test Set Samples
LEVIR-CD	7120 pairs	1024 pairs	2048 pairs
WHU-CD	6096 pairs	1184 pairs	1910 pairs
GoogleGZ-CD	804 pairs	330 pairs	330 pairs

4.2. Experimental Setup

The model in this paper was implemented on PyTorch and trained and tested on an NVIDIA RTX 4090. Regarding the specific parameter configuration of the model, this paper uses the AdamW optimizer to minimize the loss function with a weight decay coefficient of 0.0025 and a learning rate of 5e-4. Due to the limitations of GPU resources and through extensive experimental validation, this paper sets the batch size to 16 and the number of training epochs to 180 to ensure that the model can converge. In the semi-supervised experiment, the dataset is constructed through random sampling. Specifically, 5%, 10%, and 20% of the training data are randomly selected as labeled data, and the rest of the data is processed as unlabeled data.

Experimental comparison methods selection: Currently advanced semi-supervised detection models include AdvNet, SemiCD, RCL, TCNet, UniMatch-DeepLabv3+, and supervised models include our MFM-CDNet model.

To clearly investigate the performance of SCMFM-CDNet, this experiment uses three evaluation metrics: Precision (P), Recall (R), and F1 score. The formulas are shown as follows:

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{10}$$

In these metrics, TP stands for true positives, FP stands for false positives, and FN stands for false negatives.

4.3. Experimental Results and Analysis

In this paper, 5%, 10%, and 20% of the training data were randomly selected as labeled data, with the remaining data treated as unlabeled. Comparative experiments on the detection performance of different methods were conducted on the LEVIR-CD, WHU-CD, and GoogleGZ-CD datasets.

The quantitative accuracy results of different methods at different annotation ratios on the LEVIR-CD dataset are shown in Table 3.

Table 3. Quantitative Accuracy Results of Different Annotation Ratios on the LEVIR-CD Dataset.

Method	Labelled Ration								
	5%			10%			20%		
	P	R	F1	P	R	F1	P	R	F1
MFM-CDNet	93.98	74.06	82.84	94.21	80.11	86.59	95.31	83.53	89.03
AdvNet	89.19	74.99	81.48	90.98	80.41	85.37	91.34	82.37	86.62
SemiCD	56.34	39.86	46.68	91.85	82.19	86.75	91.30	84.07	87.53
RCL	82.26	77.87	80.01	85.79	81.86	83.78	86.60	85.87	85.14
TCNet	91.95	77.91	84.35	91.80	86.57	89.11	92.21	87.80	89.95
UniMatch-DeepLabv3+	94.08	81.08	87.10	94.51	82.42	88.05	94.62	83.05	88.46
SCMFM-CDNet	90.51	86.87	88.62	91.98	89.62	90.78	92.09	89.22	90.63

The SCMFM-CDNet network proposed in this paper has shown significant superiority on the LEVIR-CD dataset. Among the three label ratios, the SCMFM-CDNet network ranks first in recall and F1 score, indicating the network’s superior comprehensive performance. Although the accuracy of SCMFM-CDNet is slightly lower than the UniMatch-DeepLabv3+ method, the recall rate R of the SCMFM-CDNet network under the three label ratios is much higher than that of UniMatch-DeepLabv3+ and other methods. This means that the proportion of actual change areas that the network in this paper can detect is far higher than other methods, demonstrating the high performance of the SCMFM-CDNet network proposed in this paper. In general, the SCMFM-CDNet network has shown significant superiority in improving the accuracy and efficiency of change detection through its innovative semi-supervised strategy.

The quantitative accuracy results of different methods at different annotation ratios on the WHU-CD dataset are shown in Table 4.

Table 4. Quantitative Accuracy Results of Different Annotation Ratios on the WHU-CD Dataset.

Method	Labelled Ration								
	5%			10%			20%		
	P	R	F1	P	R	F1	P	R	F1
MFM-CDNet	85.74	75.47	80.28	89.35	79.06	83.89	90.32	84.11	87.10
AdvNet	78.76	64.80	71.10	79.97	76.06	77.96	78.40	86.44	82.22
SemiCD	86.22	72.49	78.76	82.51	80.49	81.49	89.88	87.63	88.74
RCL	74.63	68.73	85.36	78.24	77.15	77.69	80.52	86.95	83.61
TCNet	87.05	83.73	85.36	90.33	82.03	85.98	94.56	83.91	88.92

UniMatch-DeepLabv3+	92.56	84.51	88.35	89.29	87.88	88.58	86.27	92.47	89.26
SCMFM-CDNet	90.95	85.98	88.40	91.58	88.81	90.17	91.83	89.86	90.83

The quantitative accuracy results of various methods on the WHU-CD dataset indicate that the SCMFM-CDNet network proposed in this paper ranks first in recall and F1 score at label ratios of 5% and 10%. Additionally, at a 10% label ratio, the SCMFM-CDNet network’s accuracy also ranks first, demonstrating the network’s superior comprehensive performance. At a 20% label ratio, the SCMFM-CDNet network’s recall rate is second, lower than the UniMatch-DeepLabv3+ method, but still better than other methods. The main reason for this is that the WHU-CD dataset has relatively fewer positive samples (changed areas) and more negative samples (unchanged areas), with the changed areas being more distinct and less complex. The high performance of UniMatch-DeepLabv3+ in semantic segmentation allows it to detect changed areas with high accuracy. However, on the more complex LEVIR-CD and GoogleGZ-CD datasets, the performance of our method is significantly higher than that of UniMatch-DeepLabv3+. Particularly on the complex GoogleGZ-CD dataset, it can be seen that UniMatch-DeepLabv3+ is inferior to the SCMFM-CDNet network proposed in this paper in terms of generalization ability.

The quantitative accuracy results of different methods at different annotation ratios on the GoogleGZ-CD dataset are shown in Table 5. Compared to some other datasets, the GoogleGZ-CD dataset has a smaller scale. From a semi-supervised perspective, with only 5% labeled data, SCMFM-CDNet achieved an F1 score of 81.44%, which is higher than other semi-supervised change detection methods. This demonstrates its strong performance under conditions of limited labeled data, aligning with the original intention of semi-supervised learning to train using a small set of labeled data.

Table 5. Quantitative Accuracy Results of Different Annotation Ratios on the GoogleGZ-CD Dataset.

Method	Labelled Ration								
	5%			10%			20%		
	P	R	F1	P	R	F1	P	R	F1
MFM-CDNet	80.03	77.61	78.80	84.59	73.09	78.42	89.25	75.11	81.57
AdvNet	67.64	52.86	59.34	76.87	61.18	68.14	85.54	59.09	69.90
SemiCD	70.33	52.99	60.44	75.45	62.09	68.12	88.25	60.45	71.75
RCL	74.90	68.27	71.43	76.78	78.25	77.51	78.82	80.07	79.44
TCNet	78.97	70.95	74.75	75.79	81.79	78.68	85.46	79.62	82.43
UniMatch-DeepLabv3+	61.68	45.60	52.50	81.79	52.47	63.93	79.19	59.55	67.97
SCMFM-CDNet	81.86	81.02	81.44	81.02	83.75	82.36	82.64	88.36	85.40

The qualitative evaluation results of different comparison methods at a labeling rate of 10% on various datasets are shown in Figure 5. To make the visualization more intuitive, we use white to express the TP (true positives), black to represent the TN (true negatives), red to illustrate the FP (false positives), and light blue to depict the FN (false negatives). Black indicates unchanged areas, while white indicates changed areas. Red represents areas where the model incorrectly identifies unchanged areas as changed, i.e., false alarms. Light blue represents areas where the model fails to detect actual changes, i.e., missed detections. By observing the visualization, the following conclusions can be drawn: Although all methods still have common difficulties in missed detections in some areas, compared with other semi-supervised methods, the SCMFM-CDNet network proposed in this paper has relatively fewer blue and red parts in the visualization, that is, there are fewer false alarms and missed detections, and the problems of unclear change edge detection and missed detections of extremely small targets have been significantly improved.

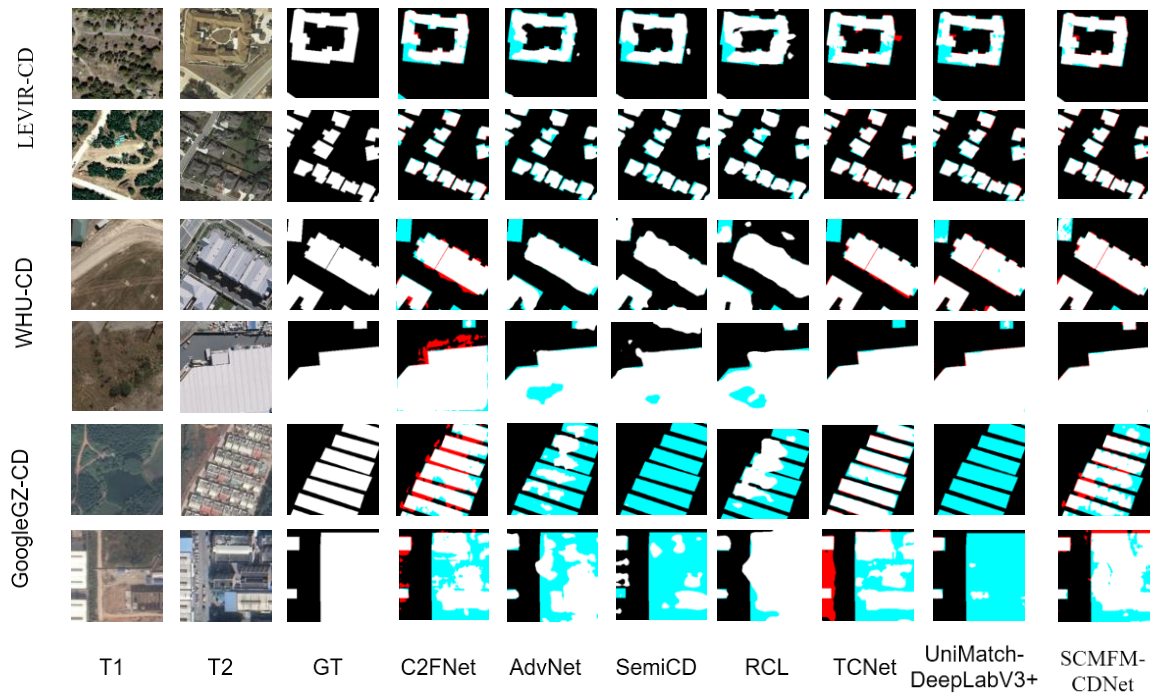


Figure 5. Qualitative evaluation results at a labeling rate of 10%.

4.4. Ablation Experiments

Ablation experiments mainly verify the effects of the Global Context Module (GCM) and the Feature Fusion and Reuse Module (FFM) in the MFM-CDNet network. This paper selects 5% and 10% label ratios in the GoogleGZ-CD dataset for validation. The quantitative accuracy results of ablation experiments at different annotation ratios on the GoogleGZ-CD dataset are shown in Table 6.

Table 6. Quantitative Accuracy Results of Ablation Experiments at Different Annotation Ratios in the GoogleGZ-CD Dataset.

Method	Labelled Ration					
	5%			10%		
	P	R	F1	P	R	F1
w/o GCM	70.28	75.31	72.71	74.22	76.04	75.12
w/o FFM	80.35	80.03	80.19	79.45	81.35	80.39
SCMFM-CDNet	81.86	81.02	81.44	81.02	83.75	82.36

From the table, it can be seen that the performance is better when both the Global Context Module (GCM) and the Feature Fusion and Reuse Module (FFM) are present, as both modules play a key role in the feature extraction process.

5. Conclusions

To address the problems of relatively difficult acquisition of annotated dual-time-phase remote sensing images, unclear detection of change edges and missed detection of very small targets, we propose a semi-supervised remote sensing image change detection network based on multi-scale feature fusion and multiplexing with consistent regularisation, including multi-scale feature fusion and multiplexing with supervised phase, and an instructor-student model used to implementing a teacher-student model for semi-supervised learning. Extensive experiments and detailed ablation experiments were done on three public datasets with three different labelling ratios of 5%, 10% and 20%, and the proposed network shown higher performance than other semi-supervised change

detection networks in both training and inference. In the future, in terms of the feature extraction in the supervised part of the MFM-CDNet and the upgrading strategy of the govern's parameters in the unsupervised part of the MFM-CDNet, good progress must be possible.

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

Funding: This work was supported by the Hebei Province Central Guiding Local Science and Technology Development Fund Project under grant number 236Z4901G and the Science and Technology Innovation Program of Hebei, China (Grant No. SJMYF2022X18). Additionally, this research was funded by three projects from the North China Institute of Aerospace Engineering: Grant Nos. YKY-2023-39, YKY-2023-40, and YKY-2024-70.

Acknowledgments: The authors would like to thank the Assistant Editor of this article and anonymous reviewers for their valuable suggestions and comments.

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

References

1. Tison C, Nicolas JM, Tupin F, Maître H. A new statistical model for Markovian classification of urban areas in high-resolution SAR images. *IEEE Transactions on Geoscience and Remote Sensing*. 2011, 42(10), 2046-2057. DOI:10.1109/TGRS.2004.834630.
2. Thompson AW, Prokopy LS. Tracking urban sprawl: Using spatial data to inform farmland preservation policy. *Land Use Policy*. 2009, 26(2), 194-202. DOI:10.1016/j.landusepol.2008.02.005.
3. Chen H, Qi Z, Shi Z. Remote Sensing Image Change Detection with Transformers. *arXiv*, 2021, submitted. DOI:10.1109/TGRS.2021.3095166.
4. Sommer S, Hill J, Mégier J. The potential of remote sensing for monitoring rural land use changes and their effects on soil conditions. *Agriculture, Ecosystems and Environment*. 1998, 67(2-3), 197-209. DOI:10.1016/S0167-8809(97)00119-9.
5. Fichera CR, Modica G, Pollino M. Land Cover classification and change-detection analysis using multi-temporal remote sensed imagery and landscape metrics. *European Journal of Remote Sensing*. 2012, 45(1), 1-18. DOI:10.5721/EuJRS20124501.
6. Gillespie TW, Chu J, Frankenberg E, Thomas D. Assessment and Prediction of Natural Hazards from Satellite Imagery. *Progress in Physical Geography*. 2007, 31(5), 459. DOI:10.1177/0309133307083296.
7. Dong L, Shan J. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *Isprs Journal of Photogrammetry and Remote Sensing*. 2013, 84(oct.), 85-99. DOI:10.1016/j.isprsjprs.2013.06.011.
8. Zhang L, Zhang L, Du B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*. 2016, 4(2), 22-40. DOI:10.1109/MGRS.2016.2540798.
9. Song A, Kim Y. Semantic Segmentation of Remote-Sensing Imagery Using Heterogeneous Big Data: International Society for Photogrammetry and Remote Sensing Potsdam and Cityscape Datasets. *International Journal of Geo-Information*. 2020, 9(10), 601. DOI:10.3390/ijgi9100601.
10. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv*, 2014, submitted. DOI:10.48550/arXiv.1409.1556.
11. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, USA, 2016, 770-778. DOI:10.1109/CVPR.2016.90.
12. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017, 39(4), 640-651. DOI:10.1109/CVPR.2015.7298965.
13. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, Cham, 2015, 234-241. DOI:10.1007/978-3-662-54345-0_3.
14. Guo Q, Feng W, Zhou C, Liu Z, Wang Y. Learning Dynamic Siamese Network for Visual Object Tracking. In *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy, 2017, 196-205. DOI:10.1109/ICCV.2017.196.

15. Li Y, Weng L, Xia M, Hu K, Lin H. Multi-Scale Fusion Siamese Network Based on Three-Branch Attention Mechanism for High-Resolution Remote Sensing Image Change Detection. *Remote Sensing*. 2024, 16(10), 1665. DOI:10.3390/rs16101665.
16. Zhan Z, Ren H, Xia M, Lin H, Wang X, Li X. AMFNet: Attention-Guided Multi-Scale Fusion Network for Bi-Temporal Change Detection in Remote Sensing Images. *Remote Sensing*. 2024, 16(10), 1765. DOI:10.3390/rs16101765.
17. Li X, He M, Li H, Shen H. A Combined Loss-Based Multiscale Fully Convolutional Network for High-Resolution Remote Sensing Image Change Detection. *IEEE Geoscience and Remote Sensing Letters*. 2022, 19, 1-5. DOI:10.1109/LGRS.2022.3141892.
18. Fonseca A, Marshall MT, Salama S. Enhanced detection of artisanal small-scale mining with spectral and textural segmentation of Landsat time series. *Remote Sens*. 2024, 16(10), 1749. DOI:10.3390/rs16101749.
19. Chakraborty D, Ghosh A. Unsupervised change detection in hyperspectral images using feature fusion deep convolutional autoencoders. *arXiv*, 2021, submitted. DOI: 10.48550/arXiv.2109.04990.
20. Wu C, Du B, Zhang L. Fully Convolutional Change Detection Framework With Generative Adversarial Network for Unsupervised, Weakly Supervised and Regional Supervised Change Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2023, 45(8), 9774-9788. DOI:10.1109/TPAMI.2023.3237896.
21. Wang JJ, Dobigeon N, Chabert M, Wang DC, Huang J, Huang TZ. CD-GAN: a robust fusion-based generative adversarial network for unsupervised change detection between heterogeneous images. *arXiv*, 2022, submitted. DOI:10.48550/arXiv.2203.00948.
22. Fang S, Li K, Shao J, Li H, Zhou Z, Zhao Z, Yang J, Wang C. SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geoscience and Remote Sensing Letters*. 2022, 19, 1-5. DOI:10.1109/LGRS.2021.3056416.
23. Daudt RC, Saux BL, Boulch A. Fully Convolutional Siamese Networks for Change Detection. 2018 25th IEEE International Conference on Image Processing (ICIP). Athens, Greece, 2018, 4063-4067. DOI:10.1109/ICIP.2018.8451652.
24. Li J, Zhu S, Gao Y, Zhang G, Xu Y. Change Detection for High-Resolution Remote Sensing Images Based on a Multi-Scale Attention Siamese Network. *Remote Sensing*. 2022, 14(14) 3464. DOI:10.3390/rs14143464.
25. Wang S, Zhu Y, Zheng N, Liu W, Zhang H, Zhao X, Liu Y. Change Detection Based on Existing Vector Polygons and Up-to-Date Images Using an Attention-Based Multi-Scale ConvTransformer Network. *Remote Sensing*. 2024, 16(10), 1736. DOI:10.3390/rs16101736.
26. Hung WC, Tsai YH, Liou YT, Lin YY, Yang MH. Adversarial Learning for Semi-Supervised Semantic Segmentation. *arXiv*, 2018, submitted. DOI:10.48550/arXiv.1802.07934.
27. Wele GCB, Patel VM. Revisiting Consistency Regularization for Semi-supervised Change Detection in Remote Sensing Images. *arXiv*, 2022, submitted. DOI:10.48550/arXiv.2204.08454.
28. Maa C, Weng L, Zhang QY. Dual-branch network for change detection of remote sensing image. *Engineering Applications of Artificial Intelligence: The International Journal of Intelligent Real-Time Automation*. 2023, 123(Pt.B), 106324-1-106324-12. DOI:10.48550/arXiv.2204.08454.
29. Shu Q, Pan J, Zhang Z, Wang M. MTCNet: Multitask consistency network with single temporal supervision for semi-supervised building change detection. *Int. J. Appl. Earth Obs. Geoinformation*. 2022, 115, 103110. DOI:10.1016/j.jag.2022.103110.
30. Wang J, Li T, Chen S, Tang J, Luo B, Wilson RC. Reliable Contrastive Learning for Semi-Supervised Change Detection in Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*. 2022, 60, 1-13. DOI:10.1109/TGRS.2022.3228016.
31. Yang L, Qi L, Feng L, Zhang W, Shi Y. Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation. *arXiv*, 2023, submitted. DOI:10.48550/arXiv.2208.09910.

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.