A Segmentation Map Difference-Based Domain Adaptive Change Detection Method

Huakang Tang, Honglei Wang, and Xiaoping Zhang

***Abstract*—Deep neural network (DNN) has been widely used in remote sensing image change detection (CD) in recent years. Due to the scarcity of training data, a large number of labeled data onto other fields become the source of DNN concept learning in remote sensing image CD. However, the distribution of features of the CD data and other data varies greatly, which prevents DNN from being better applied for one task to another. To solve this problem, a domain adaptive CD method based on segmentation map difference is proposed to this article, which includes the pretraining stage and the CD stage. In the pretraining stage, the domain adaptive UNet (Ada-UNet) is applied as the basic network of remote sensing image segmentation for network training with the purpose of learning the concepts of different features. In the CD stage, strict threshold segmentation results are used to train the channel attention network, which makes it more efficient to utilize the high-dimensional feature map. The probabilistic map generated by the three-channel attention networks is evaluated, and then it is used to accurately classify the changing pixels. In this article, experiments are carried out on datasets with different feature distributions. The results show that this method has strong domain adaptability and can greatly reduce the influence of the difference in feature distributions of the CD results.**

***Index Terms*—Change detection (CD), channel attention, domain adaptive, feature extraction, remote sensing.**

I. Introduction

REMOTE sensing image is the main form of remote sensing data at present, which directly reflects the information of land utilization and land cover. The change detection (CD) based on remote sensing image has a wide range of applications in environmental monitoring [1], military reconnaissance [2], urban planning [3], etc. In recent decades, the large increase of geodesic satellites and the enhancement of imaging capability has put forward new requirements for the CD method. In practical applications, it is necessary to quickly obtain accurate information from massive image data.

Manuscript received April 12, 2021; revised June 28, 2021 and August 5, 2021; accepted September 11, 2021. Date of publication September 20, 2021; date of current version October 6, 2021. This work was supported by the Science and Technology Cooperation Program of Guizhou Province under Grant (QKH[2016]5103) *(Corresponding author: Honglei Wang.)*

Huakang Tang is with the School of Electrical Engineering at Guizhou University, Guizhou University, Guiyang 550025, China (e-mail: 18311744452@163.com).

Honglei Wang is with the School of Electrical Engineering, Guizhou University, Guiyang 550025, China, and also with the Key Laboratory of "Internet +" Collaborative Intelligent Manufacturing in Guizhou Province, China (e-mail: gzdxhlwang@163.com).

Xiaoping Zhang is with the Science and Technology Department of Guizhou Province, Yunyan District, Guiyang 550000, China (e-mail: 2215093175@qq.com).

Digital Object Identifier 10.1109/JSTARS.2021.3113327

CD is one of the most important means in remote sensing image analysis and interpretation [4]. It aims to identify differences in remote sensing images of the same geographic location at different times. At present, the CD method can be divided into two categories. One is based on the original difference map of two images to generate the difference map through the log-ratio operator [5], mean-ratio operator [6], PCA [7], and other methods. The difference maps are divided into changed or unchanged categories through threshold-based [8], clustering-based [9], or level set [10] to generate the change map. The other is based on the classification or segmentation method. In this kind of method, images are classified by using neural network [11], and all kinds of generated feature images are fused and processed. The difference map method is a pixel-level method, which usually produces a lot of noise and cannot identify false changes such as shadows. The method based on classification requires deep neural network (DNN) to extract image features. In the process of feature extraction, the ability of DNN to extract unbalanced features will be greatly reduced because the features of ground objects contained in a large range of remote sensing images are usually not evenly distributed [12].

There have been many studies that have applied the classification-based method and the difference map-based method to the CD task [13]–[15]. The classification-based or segmented method can learn the global feature changes. However, it is not enough to reflect the image detail changes [16]. Based on the difference map processing method, the difference map of the three-channels can retain the original change information to a greater extent, but it is not enough to reflect the change of the global features of the image [17]. It is a natural method to combine the difference map of the three-channel s with the segmentation difference map. This combination leads to two problems. First, as a feature extractor, the deep learning method often has a poor effect on data with different feature distributions [18]. Second, the high-dimensional feature map processing after the combination becomes the key to the effect of CD, because the high-dimensional feature map is often interrelated [19].

Domain adaptive is an excellent method to solve the problem of unbalanced feature distribution in the CD dataset [20]. In recent years, domain adaptive transfer learning has developed rapidly in various computer vision tasks due to the lack of natural image data and the difficulty in labeling [21]. The domain adaptive method can shorten the distance of feature distribution so that the neural network can better extract features [22]. Domain adaptive method has extensive applications, such as image classification [23], target detection [24], streetscape classification

[25], etc. The channel attention module can effectively extract the correlation between channels. In recent years, many channel attention networks have been proposed and used to solve various visual problems. The channel attention models, such as SKNet [26], CCNet [27], and SENet [28], have been used for semantic segmentation and achieved good results.

In this article, we propose a new CD method based on semantic segmentation feature maps to generate change maps from high-resolution remote sensing images. This structure requires less target domain data to train the channel attention network, which can effectively improve the precision of CD. Our main contributions are threefold.

1) To solve the problem of multiscale CD, we use a lightweight UNet to extract features, and the segmented feature map and the original difference map are stacked as the input of the CD network. This structure can extract image features with fewer parameters.
2) To reduce the difference of feature distribution between different CD datasets, this article uses the adaptive batch normalization (AdaBN) to improve the UNet network. The AdaBN can narrow the difference of feature distribution between different data domains.
3) To make it more efficient to use high-dimensional feature maps, the ability of three-channel attention networks is evaluated to extract channel correlations of high-dimensional feature maps, and the balanced cross-entropy loss function is used to balance the changed and unchanged pixels in the ground truth.

The rest of this article is organized as follows. Section II discusses the related work, Section III introduces the method we proposed in detail, Section IV is the experiment, Section V further discusses the method, and Section VI concludes the article.

II. Previous Work

In this section, we first briefly review the current CD network. Then we analyze the application of the domain adaptive method in remote sensing images. Finally, we review the application of the channel attention method in the convolutional neural network.

A. Change Detection Network

The main factor limiting the improvement of pixel-level CD precision is the extraction of spatial features. To extract spatial feature information, some network structures are designed to extract high-level features of images [29], [30]. Different from conventional deep learning tasks, the CD task is to detect the difference between two images, and many existing structures need to be further processed to obtain the CD map.

The structure based on a deep belief network can extract the low-level features of images, but it is difficult to select the appropriate pixel based on the pixel method. The CNN-based method preprocesses the difference maps of various operators, then classifies them with CNN, and obtains the change binary graph by voting. This method does not directly use the feature extraction ability of CNN, and still cannot extract ground changes from the perspective of features. Siamese network [31], [32] is very popular in the field of CD. The conventional CD process directly uses the convolutional neural network to extract features and then uses the Siamese structure method to obtain the differences between two images. After further processing, most of the information can be detected. Based on the idea of the Siamese neural network, features can be extracted directly from two images.

Due to its concise structure and the small number of parameters, UNet [17] is widely used in semantic segmentation and other fields. It can integrate features of multiple scales to generate a semantic segmentation map. In this article, the traditional semantic segmentation structure UNet is used as the feature extraction network of Siamese neural network, and a Siamese neural network based on UNet is constructed.

B. Domain Adaptive

Although deep learning has powerful automatic feature extraction capability, it is a part of machine learning. Therefore, one of the most important criteria of machine learning needs to be met, that is, the distribution of data features used for training and testing needs to be consistent [33]. For tasks with inconsistent data feature distribution, if the deep learning method is directly used without considering data distribution, the tasks will only stay at the research level, rather than a specific application. Domain adaptive [34] for the study of modern deep learning is very important because the deep learning algorithm is data-driven [35], nature has all kinds of data distribution, to seek a deep learning algorithm that can adapt to all kinds of data distribution was very hard. Instead of seeking a deep learning algorithm that can adapt to various data distributions, it is better to study how to shorten the feature distribution distance in the data domain.

Many existing studies have made great contributions to reducing the difference of feature distribution among similar data sets, which are mainly divided into several levels of domain adaptive algorithms [36]–[38]. By resampling the samples in the source domain, the sample adaptive algorithm makes its distribution approach to the distribution of the target domain, finds out the samples that are closest to the target domain from the source domain, and lets them join the data learning in the target domain with high weight. The feature-level adaptive algorithm projects the source domain and the target domain into the common feature subspace so that the training knowledge from the source domain can be directly applied to the target domain. Taking into account the error of the target domain, the model level adaptation is mainly through modifying the error function of the source domain.

Generally speaking, there are obvious differences between different data domains at the data feature level. There are many adaptive algorithms at the feature level from the following aspects: data distribution [39], feature selection [40], and feature transformation [41]. Since the problem of unbalanced data distribution mainly exists among the datasets of the CD task of remote sensing images, this study considers solving the problem from the perspective of data distribution. Minimizing the probability distribution distance is designed to balance the

probability distribution of the data, common methods include TCA [42] and JDA [43].

To better combine domain adaptation with feature extraction network, this article introduces adaptive batch normalization (AdaBN [44]) improved UNet network. The ordinary BN layer can only accelerate the network calculation process and can do nothing about the difference in data distribution. AdaBN is used to improve the UNet structure to minimize the distance of feature distributions among different datasets. When the network model expands to an unknown domain, the AdaBN algorithm first calculates the mean and variance of the domain through a forward propagation. Then it updates the parameters of the BN layer of the entire network, freezing the BN parameters, and then trains the network.

C. Channel Attention

Adding the learning mechanism of acquiring spatial correlation to the network can effectively improve the network performance without additional supervision. Channel attention is a kind of spatial attention [45]–[48]. After the image passes through the convolution layer, the local spatial information can be extracted. Although nonlocal blocks can produce long range dependence on images, it is difficult to determine the weight of each data set. As a good channel attention algorithm, CCNet is introduced into semantic segmentation. Although it can obtain the context information on the cross paths, it does not pay enough attention to the single-channel as a whole. SKNet is also used for channel attention extraction, but adding SKNet to the network will add a lot of parameters. Since the CD task has further operations after feature extraction, SKNet will add a lot of parameters, making the model difficult to train.

The SE block [28] is a lightweight threshold mechanism designed to model the relationships between channels. The goal of our introduction of SE Blocks is to enhance the features with more information when the network is processing the multiscale feature map. The main reason for this is to make full use of these features later and to suppress unwanted features. SENet can capture the overall attention of a single map and can represent the relationships between channels.

SENet have three steps. First, it investigates the signals that characterize each channel, compresses the global spatial information into channel descriptors, and uses rms pooling to generate statistics for each channel. Then, the degree of dependence of each channel is investigated with the threshold mechanism of the sigmoid activation function. The final sigmoid function is to adjust the weight of each channel feature according to the input data according to the weight of each channel, which helps to better distinguish important features.

III. Proposed Method

In this section, we describe the proposed CD method in detail. The structure of this method is shown in Fig. 1.

This method consists of three steps. First, the feature extraction network is trained in the source domain. Second, the channel attention network is trained by a high-dimensional feature map combined difference map and original difference map.

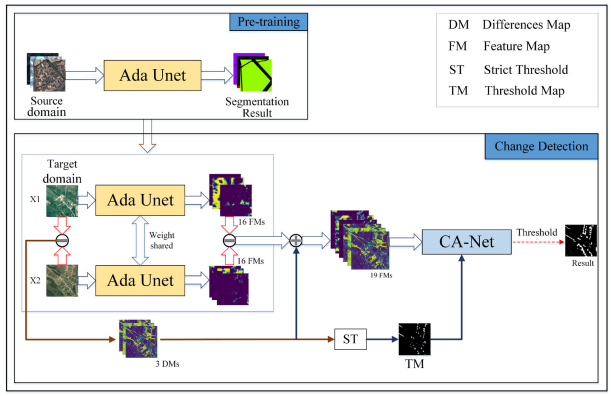

Fig. 1. Flowchart of the proposed method.

Third, using the probabilistic map generated by the network obtains the result through threshold segmentation.

Here, we set the T1 temporal image of the target domain as X1 and T2 temporal image as X2.

DM represents the difference map of two images, FM represents the feature map, TM represents the ground change map after simple threshold segmentation, and ST represents the threshold segmentation under strict conditions.

A. Training Feature Extraction Network

According to whether spatial neighborhood information is used or not, feature extraction methods are mainly divided into two categories. One type is to use one-dimensional data for feature extraction, which mostly utilizes neural networks such as deep belief networks [49] and stack autoencoder [50]. The other is CNN [51], which has achieved great success in most image tasks. CNN is undoubtedly suitable for feature extraction tasks. SegNet [52] and UNet structures are commonly used for image feature extraction. SegNet extracts image features using sampling structure from top to bottom, which is different from other sampling structures.

SegNet network can locate the pooling of the operation; therefore, points can be more accurate image information on the sampling stage. But the multiscale features of the sampling stage under the figure have not been fully used and this will lead to the loss of valuable information.

UNet structure can extract multiscale features very well. Like most CNNs, it has obvious defects for the problem of inconsistent feature distribution between training data and test data. Current CD methods almost do not consider the difference of feature distribution between data domains. However, the reason for the great success of deep learning is a basic assumption, which is the data feature distribution used in the process of training and testing is consistent [53]. To overcome this problem, an AdaBN operation that can feature the alignment of data is introduced in UNet. Adaptive batch standardization can shorten the distance between two data domains and enhance the adaptability of the network to different datasets.

In the feature extraction stage, this article uses an improved UNet model structure, as shown in Fig. 2. UNet consists of

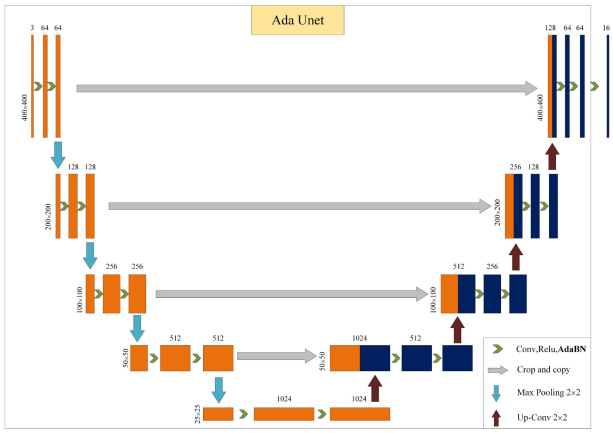

Fig. 2. Ada-UNet network. The difference with UNet is we replace BN with AdaBN, a forward propagation is required to update the parameters of AdaBN before Ada-UNet training.

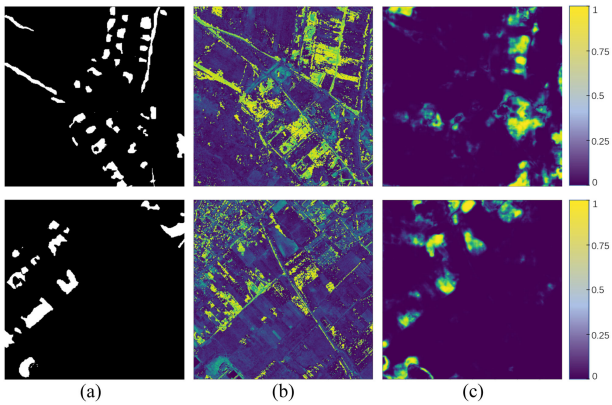

Fig. 3. Features of the same ground object at different scales. (a) is the ground truth, (b) is one of the channels of difference map, and (c) is one of the channels for segmentation of difference map.

four downsampling layers and four upsampling layers. Each downsampling layer and the convolution layer at the end of the upsampling combines ReLU and AdaBN and finally carries out semantic segmentation. Semantic segmentation training is performed on the GID dataset [54], and then network weights are shared so that the two networks output consistent feature mappings. Finally, the subtraction is carried out directly to retain as much feature information as possible.

The same object in different scales of the characteristic map is shown in Fig. 3. Fig. 3(a) is the ground truth, Fig. 3(b) is one of the original image subtraction resulting directly channel feature map, Fig. 3(c) is the transformation of the CD task of remote sensing image into the processing task of multiscale feature difference map obtained by the subtraction of a channel feature map.

B. Adaptive Segmentation Difference Map CD Network

Aiming at the problem of inconsistent feature distribution in different data sets, this article designs an adaptive attention CD network for feature maps. The network structure is composed of three parts as follows:

1) Ada-UNet Feature Extraction: In this article, a two-channel structure of UNet based on the idea of Siamese neural network is proposed to solve the problem of insufficient training samples. Multiscale feature maps were extracted from X1 and X2. The original UNet was composed of four upsampling layers and four downsampling layers. Considering that the high-dimensional feature map contains more information, we finally selected the 16-dimension output feature map to retain the maximum object-level feature extraction information. The input image X of the proposed adaptive UNet (Ada-UNet) structure is 400 × 400×3, and the output Y of the Ada-UNet substructure is 400 × 400×16.

2) Difference Map Fusion: In order to make the data distribution of the feature difference map more reasonable, Z-score standardization [55] was first performed on the data. Z-score is described as follows:

$$Y_i^* = \frac{Y_i - \mu}{\sigma}, i = 1, \ldots, N \tag{1}$$

where Y_i^* is the normalized multidimensional difference map, Y_i represents the multidimensional difference map, μ is the mean value of all feature maps, σ is the standard deviation of all feature maps, and N represents dimension.

After the standardized difference map is obtained, the DM with the most image change information is fused for stacking in the channel dimension direction. Finally, the multiscale difference map containing the most change information is obtained with a dimension of 400 × 400 × 19.

3) Multiscale Channel Attention Network: In order to explore the relationship between channel direction of feature map, this article introduces the weight between channel attention network learning channels. We introduce the strict threshold segmentation map as the reference data for training, aiming to learn the weight relationship between channels. Generally, there are a few changed pixels in the CD data set and more pixels do not changed. Data types are unevenly distributed. In order to alleviate this problem, this article introduces root mean square pooling [56], which can better reflect the dispersion degree of a feature layer.

The introduction of channel attention network can effectively learn the correlation between feature layers, but how to better use channel attention module needs experiments. In order to make reasonable use of multiscale difference map, we have designed three different levels of encoding decoding structure for experiments. The results show that high dimension map can retain the maximum reservation details, but might introduce more noise. Coding decoding structure part can restrain noise, but it may ignore details. The last layer of the network was trained with 1×1 Conv and two filters, and the network was trained with Sigmoid. A single filter was used during the test, and softmax was used to activate the probabilistic map. Network Settings is very important. Thus the three network in this article are verified through experiments.

4) Loss Function: Using the appropriate loss function can effectively improve network performance and reduce training time. Contrastive loss function, softmax loss function, and cross-entropy loss function [57] are the most commonly used

loss functions for CD. Contrastive loss is used to evaluate the similarity between two images, while softmax loss is usually used for multiclass classification. Cross-entropy loss is used to measure the similarity between two probability distributions and is more suitable for CD tasks, since the purpose of CD is to classify pixels as changed or unchanged. The cross-entropy loss function can be expressed as follows:

$$L = -\frac{1}{N}\sum_{n=1}^{N}[y_n \log \hat{y}_n + (1-y_n)\log(1-\hat{y}_n)] \quad (2)$$

where N represents the number of samples, y_n represents the true value of samples, with a value of 0 or 1, $\hat{y}$ represents the pixel that is unchanged or changed and represents the network prediction result. In the CD task, the number of unchanged pixels is generally far less than the number of changed pixels, which means that the network prediction result is more inclined to have no change, because the pixels of $y_n = 0$ are far more than the pixels of $y_n = 1$, which will lead to some changes in the original loss function and $y_n \log \hat{y}_n$ will tend to 0.

Because the training data samples are not balanced, the loss function may fail and the network tends to predict samples that are easy to learn. Therefore, for the bias sampling, the conventional cross-entropy function must be improved to make it more suitable for the sample imbalance problem. The balanced cross-entropy loss [17] is introduced, which can solve the above two problems, and it can be expressed as follows:

$$L_{fl} = -\frac{1}{N}\sum_{n=1}^{N}[\beta y_n \log \hat{y}_n + \hat{y}_n(1-\beta)(1-y_n)\log(1-\hat{y}_n)] \quad (3)$$

where N represents the sample number; y_n represents the true value of the sample, with a value of 0 or 1, represents the pixel that has not changed or changed; $\hat{y}$ represents the network prediction result; and β represents the balance factor, $\beta = \frac{p_-}{p_- + p_+}$ The addition of balance factor is mainly to reduce the influence of sample imbalance on network preference and to balance the training loss of the two kinds of samples.

C. Threshold Analysis

Through data processing, it is found that the difference graph of the two-phase images mainly presents a bimodal distribution, as shown in Fig. 4, in which the pixels with higher values tend to be the parts that actually change. That is to say, D is the pixel that actually changed. Therefore, pixels in part D are used as truth values to train the network in order to obtain the weights between feature maps. In the CD task, the pixels that are most difficult to determine whether they have changed are parts B and C, so we regard them as uncertain parts.

In order to better train channel attention CD network, this research uses strict threshold segmentation maps as the training samples of the network. We restrict the result between the two conditions by a strict condition and a relaxed condition, which obtain a more accurate result.

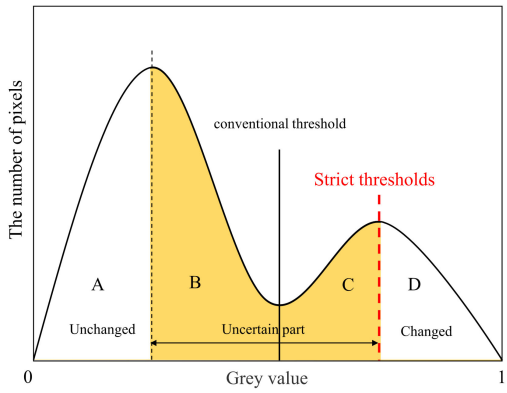

Fig. 4. Multiscale channel attention network. (a) is the encoding and decoding structure with two up-down sampling, (b) is the encoding and decoding structure with one up-down sampling, and (c) is the attention convolution layer of 1×1 channel.

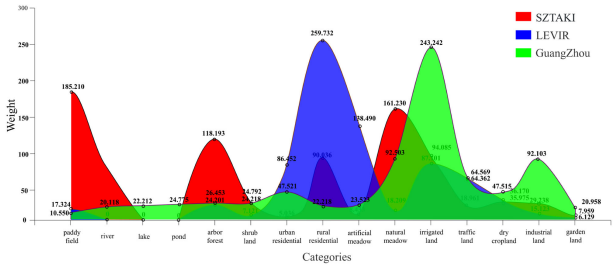

Fig. 5. Grayscale distribution of pixels.

If the result obtained by an untrained Ada-UNet network is regarded as a looser result, the stricter result is the pixel in part D of Fig. 4.

To contribute to the geoscience community, the implementation of our proposed method will be released through GitHub.[1]

IV. Experiment

First, this section provides a detailed description of the datasets used and selects reasonable evaluation indicators. Second, some basic methods used in CD are analyzed. Finally, the proposed method is used to detect the change of datasets with different feature distributions.

A. Datasets and Evaluation Metrics

1) Datasets Description: The experimental data set adopted in this article is shown in Table I, including two training sets and three test sets.

This study first completes the remote sensing image semantic segmentation in Ada-UNet network. Then, the feature is extracted with UNet network and the AdaBN, which has the ability to align the different data set features. In order to test the network's ability to solve tasks only with few samples, we choose two training datasets containing less data. The training set is the semantic segmentation data of remote sensing images,

[1][Online]. Available: https://github.com/tanghuakang/A-SMD-Based-Domain-Adaptive-Change-Detection-Method

TABLE I
DATASETS DESCRIPTION

Dataset		Region	Resolution(m)	Size(Pixels)	Number of Images
Training	GID	China	1.0	6800 × 7200	5 Training,10Testing
	Zurich Summer	Switzerland	0.62	600 × 1600	20
Test	SZTAKI	Japan	1.5	952 × 640	13 pairs × 2 stage
	GuangZhou	China	3	-	20 pairs × 2 stage
	LEVIR	China	1.0	800 × 600	128 pairs × 2 stage

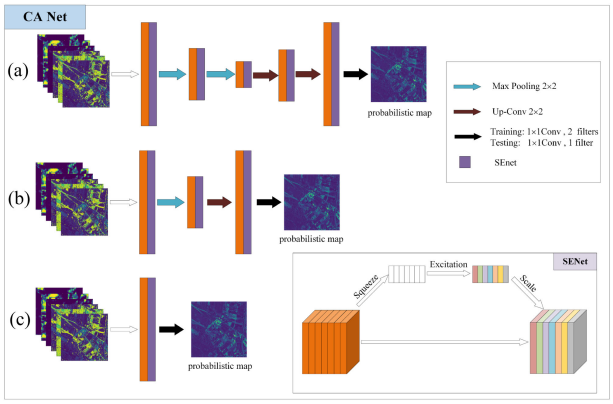

Fig. 6. Categories and weights of features in the datasets.

among which there are 15 images in Gaofen Image Dataset, including training data and test data. There are five images of training data and 10 images of test data. The Image size is 6800 × 7200 pixels and the resolution is 1 m. Zurich Summer Semantic Segmentation Dataset [58] contains a total of 20 images, each of which has a size of 600 × 1600 and a resolution of 0.62 m. The location of the images is Switzerland.

In order to test the generalization ability of this method for datasets with inconsistent feature distributions, several datasets with obvious differences in feature distributions were selected for testing. Through the analysis of these datasets, we found that the feature distribution of the dataset was significantly different, as shown in Fig. 5. The SZTAKI dataset [59] was sampled from three prefectures in Japan, with a total of 13 pairs of time-stage images. The image size was 952 × 640 and the resolution was 1.5 m. The Guangzhou dataset [60] was sampled in Guangzhou, China, with a total of 20 pairs of temporal images with a resolution of 3 m. The LEVIR dataset [61] was sampled from all over the world and verified 64 pairs of two-stage images with a resolution of 1 m and an image size of 800 × 600.

The ground object distribution of the three test datasets is shown in Fig. 5. As can be seen from the figure, the feature distribution of the three data sets is greatly different. The Guangzhou dataset has almost all the features, and the features are evenly distributed. The SZTAKI dataset lacks some features, and its distribution presents a bimodal trend. LEVIR dataset does not have a lot of features, the distribution of features is completely irregular.

2) Evaluation Metrics: In order to verify the effectiveness of the method we proposed, we selected the following evaluation indicators. The first metrics are TP and FN. Since the labeling of the CD task is all made manually, the labeling is highly subjective. Some data sets are mainly biased to the architectural change, while others are mainly biased to the change of land use type. Although the labeling tendency of the data set is different, the changes of the labeling must be the changes of the actual ground objects. Therefore, our method at least needs to detect the changes of man-made labeling. Thus, precision and recall are required to evaluate whether the labeled changes are detected. OA was used to evaluate the precision of the changed and unchanged pixels judged by the network. Taking into account both accuracy and recall rate, the F1-score was the harmonic average of the model's precision and recall rate. To sum up, four evaluation indexes were adopted: precision, recall, OA, and F1-score. As shown in the following equations:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{5}$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{7}$$

where TP, FP, TN, and FN denote the number of true positives, the number of false positives, the number of true negatives, and the number of false negatives, respectively.

B. Experimental Setup

1) Data Processing: Gaofen Image Dataset contains a total of five training data and 10 test data. We make the detailed classification of 10 test data for network training and cut 6800 × 7200 pixels image into 400 × 400 pixels image according to the size of input of network. Zurich Summer data set has a total of 20 images which are also cut into 400 × 400 pixels image. The above images are used for network training. Through the above sampling, the GID data set is cropped into 3060 images of 400 × 400 pixels, and the Zurich Summer data set is cropped into 160 images of 400 × 400 pixels. The principle of data selection in the target domain is as follows. Appropriately sample from each dataset, use a small amount of data for strict threshold segmentation and extract feature for CD. We randomly selected 20 images of 400 × 400 pixels from each data for network effect verification and calculated evaluation indexes.

Although the UNET network can show a strong feature fitting ability under the condition of a small amount of data, for the image task, 3220 images are obviously not able to meet the

need of the network training, so we adopted some necessary data enhancement means in the network pretraining stage. The data enhancement methods used in this article include random flip and random rotation.

2) Network Setup: The network structure proposed in this study is shown in Fig. 1, and the detailed description is as follows.

The Ada-UNet network for semantic segmentation consists of four downsampling layers and four upsampling layers. The downsampling layer consists of Conv + ReLu + AdaBN + Conv + ReLu + AdaBN + MaxPooling. The upsampling layer consists of Conv_Transpose + ReLu + AdaB + Conv + ReLu + AdaBN + Conv + ReLu + AdaBN, and finally consists of two sorting layers: Conv + ReLu + AdaBN + Conv + ReLu + AdaBN + Conv + softmax. Fig. 2 is the detailed parameter information.

After the semantic segmentation, the three-channel difference diagram, which is a subtraction result of the segmentation feature map and the original image channel by channel, is stacked with channel dimensions to obtain the feature map of 19 channels. Then, the channel weight is learned by the channel attention network. If there are fewer 3×3 convolution layers in the network, the network will only pay attention to the channel weight but ignore the spatial information of the feature graph. If there are more 3×3 convolution layers, the network space information will account for a large weight but ignore the channel information. Here, we design three-channel attention network schemes, as shown in Fig. 6. Fig. 6(a) and (b) consists of two upsampling and downsampling structures and one upsampling and downsampling structure, respectively. The lower sampling structure is: 3×3 Conv + ReLu + AdaBN + MaxPooling. The upsampling structure is: Conv_Transpose + ReLu + AdaBN; Fig. 6(c) is 1×1 Conv. Different from conventional SENet, rms pooling is adopted in this article, which is different from global average pooling. rms pooling can better express the contribution degree of feature information in feature graph for the final result. The channel attention network is trained with the results of strict threshold segmentation to obtain the channel weight.

This study is implemented through tensorflow. The source domain image semantic segmentation training the optimizer for stochastic gradient descent, we use learning rate is 0.01, the batch size is 32, training target domain CD, we use the optimizer is stochastic gradient descent, the vector is 0.01, the batch size is 16 strict threshold segmentation to produce training data, we use part D of Fig. 4 as the training ground truth of multiscale channel attention network, finally to threshold segmentation results using the conventional threshold.

3) Loss Function: Since the balance loss function is adopted in this study to balance the imbalance between the changed samples and the unchanged samples in the training data, the calculation method of β is as follows:

$$\beta = \frac{p_n}{p_n + p_c} \tag{8}$$

$$1 - \beta = \frac{p_c}{p_n + p_c}. \tag{9}$$

p_n represents the total number of pixels that have not changed, and p_c represents the total number of pixels that have changed.

C. Baseline Method

Because this research focuses on migration domain adaptive problems in the field of study, in order to highlight the effectiveness of this method in our study, we selected the method to compare as follows.

1) The threshold segmentation method [8].
2) PCA [62]: The traditional PCA method to construct the difference image is to carry out PCA transformation on the two stages of the image respectively, and then use the first principal component to carry out the difference calculation. In this article, the variance contribution rate is set as 80, and then the classification is carried out by k-means.
3) Fuzzy C-means [9], membership matrix index is set as 2, the maximum number of iterations is set as 1000, and the pixels in the difference map are clustered.
4) Random forest [63].
5) Siamese-KNN [64] generate differences variation graph by using CNN and K-nearest neighbors.
6) VGG16_LR uses the feature map generated by VGG16 to perform low-rank decomposition.
7) PCANet [65].

D. Result Evaluation

By training 87 epochs, the segmentation accuracy of Ada-UNet reached 94.5%. The cross-entropy loss function is used in the pretraining process. Under this precision, it can be considered that the segmentation feature map generated by our semantic segmentation network can well represent the real situation on the ground. In order to demonstrate the effect of our proposed method on different feature distribution data sets, this section demonstrates the effectiveness of this study through the following experiments. Then, the ablation study was carried out to determine the contribution of each module of the proposed method to the network. Finally, the number of parameters of the proposed method was calculated and compared.

1) Dataset With a Balanced Distribution of Features: By analyzing the distribution of ground objects in the dataset, we find that the Guangzhou data set is a dataset with a balanced distribution of ground features.

In order to verify the generality of the proposed method, we first conduct a CD experiment on the Guangzhou dataset. According to the description of the previous method, we use the strict threshold segmentation results to train the channel attention network and then use the general threshold segmentation to detect the change of gray image.

The changes of a large number of artificial targets are marked more in the Guangzhou dataset, such as building roads, etc., while the changes of ground vegetation, water bodies, and rocks are considered less. The results of a comparison between the method in this article and other methods are shown in Table II. Structure B of the method proposed in this article achieves the best results.

The comparison between the probabilistic map generated by the network proposed in this article and the ground truth is shown in Fig. 7. Figs. 8(a) and 7(b) are images of two-time stages of a certain area on the ground in the Guangzhou dataset,

TABLE II
COMPARISON OF DETECTION RESULTS OF GUANGZHOU DATASET

Method		Precision	Recall	OE	F1-Score
PCA		0.6893	0.6657	0.6761	0.6773
FCM		0.6582	0.6357	0.6405	0.6468
RF		0.7016	0.6783	0.6844	0.6898
Siam-KNN		0.8481	0.8193	0.8305	0.8335
VGG16-LR		0.8637	0.8348	0.8464	0.8490
PCANet		0.8690	0.8396	0.8498	0.8541
ours	a	0.8875	**0.8579**	0.8703	0.8724
	b	**0.8890**	0.8568	**0.8703**	**0.8726**
	c	0.8470	0.8277	0.8353	0.8310

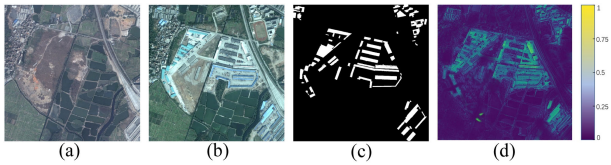

Fig. 7. Comparison of real ground change map and probabilistic map. (a) and (b) represent remote sensing images in time periods T1 and T2, (c) represents marked real ground truth, and (d) represents probabilistic map generated by this method.

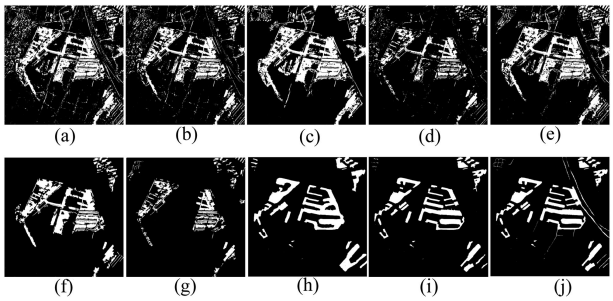

Fig. 8. Comparison of detection effect of feature distribution balanced dataset. (a) Threshold. (b) PCA. (c) FCM. (d) Random forest. (e) Siamese KNN. (f) VGG16-LR. (g) PCANET. (h), (i), and (j) are the results obtained by the three structures proposed in this article.

respectively. It can be seen that there are few buildings in the area in time T1, while great changes have taken place in time T2. It can be seen from Fig. 7(c) and (d) that the real ground change is basically consistent with the high probability area of the probabilistic map, but the result cannot be obtained by simple visual measurement. In this article, the probability map is divided by the conventional threshold value to generate the change map.

The traditional pixel-based methods are shown in Fig. 8(a), (b), (c), and (d), respectively, and the results in all aspects are still superior to the method proposed in this article.

By comparing the results of each method, we find that the threshold segmentation method can obtain the most complete land surface changes, but many of the land surface changes generated by the threshold segmentation method are pseudo changes, that is, this method cannot distinguish pseudo changes.

TABLE III
COMPARISON OF DETECTION RESULTS OF SZTAKI DATASET

Method		Precision	Recall	OE	F1-Score
PCA		0.6794	0.6532	0.6593	0.6670
FCM		0.6688	0.6428	0.6486	0.6555
RF		0.6903	0.6649	0.6712	0.6774
Siam-KNN		0.8582	0.8219	0.8631	0.8397
VGG16-LR		0.8619	0.8327	0.8444	0.8470
PCANet		0.8642	0.8403	0.8500	0.8521
ours	a	0.8794	**0.8501**	**0.8622**	0.8645
	b	**0.8994**	0.8335	0.8599	**0.8651**
	c	0.8574	0.8308	0.8412	0.8347

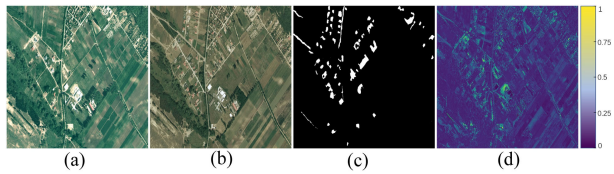

Fig. 9. Comparison of real ground change map and probabilistic map. (a) and (b) represent remote sensing images in time periods T1 and T2, (c) represents marked real ground truth, and (d) represents probabilistic map generated by this method.

Actually, PCA, random forests, and FCM also cannot eliminate pseudo changes; the cause of this phenomenon is mainly because traditional pixel-based method cannot identify advanced features, and CD data set of tag data is usually carried out for calibration, generally has an obvious tendency, is negligible for some very slight change. Fig. 8(e), (f), and (g) shows some advanced methods in the field of remote sensing image CD results, but is ahead of the results of the methods in this article.

2) Dataset With Uneven Distribution of Features: Through the analysis of the distribution of the ground object of data collection, we found that the SZTAKI dataset is a ground features distribution imbalanced dataset. SZTAKI dataset in Japan contains 13 images, mainly with the variation of buildings and roads, and methods to the spring and autumn period and the change does not take into account the ground vegetation part puts forward three types of channel attention network, we use each picture of strict threshold figure of three network training, the results such as Table III, visible, the best comprehensive performance is our structure B.

In the process of training and testing, first of all, based on datasets SZTAKI adaptive semantic segmentation, segmentation feature difference image, then to strict threshold segmentation of image difference figure, get variation, using the feature difference and variation of network training.

Then, the cross-validation method was used to carry out the experiment and the heat map of the multiscale channel attention CD network B completed by training was shown in Fig. 9(d). Fig. 9(a), (b), and (c) were the time stage T1, time stage T2, and ground truth, respectively.

In unbalanced distributed datasets, structure B proposed in this article performs best. The traditional pixel-based methods

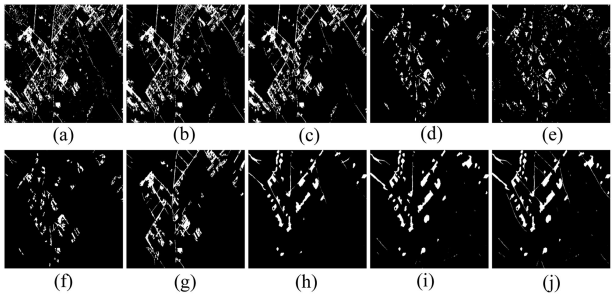

Fig. 10. Comparison of detection effects of dataset with uneven feature distribution. (a) Ordinary threshold segmentation. (b) PCA. (c) FCM. (d) Random forest. (e) Siamese KNN. (f) VGG16-LR. (g) PCANET. (h), (i), and (j) are the results obtained by the three structures proposed in this article.

are shown in Fig. 10(a), (b), (c), and (d), respectively. Compared with the pixel-based method, the method proposed in this article undoubtedly surpasses the traditional unsupervised method in every aspect, because the pixel-based method does not have the ability to learn advanced features from the image. By comparing the results of each method, we can also find that the threshold segmentation method can fully find out the changes between two images, and the selection of threshold is the key to the good or bad results.

As shown in Fig. 10(a), we choose a strict threshold, which leads to many real changes not being detected. PCA and random forest method can identify most of the changes, but cannot effectively exclude the false changes. The fuzzy C-means method eliminates a lot of pseudo changes, but the change diagram is rather disorderly because there are no learning advanced features. Fig. 10(e), (f), and (g) shows some advanced methods in the field of CD of remote sensing images. Siamese KNN and VGG16_LR methods can detect a lot of changes, but the detection of subtle changes is relatively rough. PCANet can distinguish some ground features and eliminate the interference of noise well, but the recognition of ground features using PCANet is not accurate enough.

3) Dataset With Extremely Uneven Distribution of Features: In order to obtain the detection effect of this method when the feature distribution is extremely uneven, the LEVIR dataset is used for testing. Through data analysis, we found that the data collection of all the characteristics of the rural, residential, and irrigated land is the most significant and other characteristics are less, and none of the characteristics of the data set river, lakes, and other water features, and characteristics of the distribution is very uneven.

This dataset can be used to test the ability of the proposed method to detect the data with great differences in feature distribution. In this article, the method and the results of the benchmark method for such as shown in Table IV, according to the results we can see various methods based on the DNN indicators are lower than before the results of the two datasets, for the uneven characteristic distribution difference of the data, the conventional method based on pixel almost has no effect, but the need for using the methods of extracting features of the DNN became the accuracy of a block. Fortunately, the proposed approach is still in a variety of methods to obtain the best results.

TABLE IV
COMPARISON OF DETECTION RESULTS OF LEVIR DATASET

Method		Precision	Recall	OE	F1-Score
PCA		0.6753	0.6525	0.6578	0.6637
FCM		0.6696	0.6529	0.6611	0.6568
RF		0.7012	0.6942	0.6977	0.6962
Siam-KNN		0.8018	0.7831	0.7899	0.7923
VGG16-LR		0.8196	0.7852	0.7977	0.8020
PCANet		0.8235	0.8053	0.8122	0.8143
ours	a	0.8680	0.8524	0.8588	0.8601
	b	**0.8827**	**0.8652**	**0.8726**	**0.8739**
	c	0.8130	0.8002	0.8050	0.8065

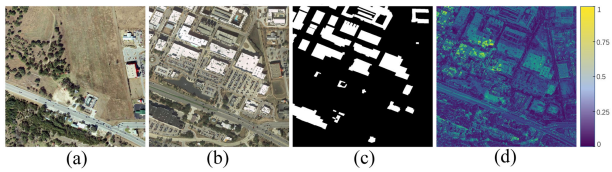

Fig. 11. Comparison of real ground change map and probabilistic map. (a) and (b) represent remote sensing images in time periods T1 and T2, (c) represents marked real ground truth, and (d) represents probabilistic map generated by this method.

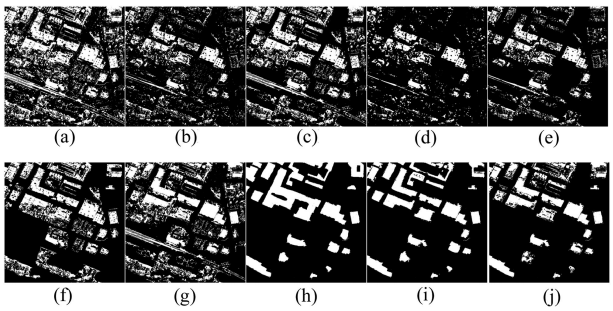

Fig. 12. Comparison of detection effects of dataset with extremely uneven feature distribution. (a) Ordinary threshold segmentation. (b) PCA. (c) FCM. (d) Random forest. (e) Siamese KNN. (f) VGG16-LR., (g) PCANET. (h), (i), and (j) are the results obtained by the three structures proposed in this article.

The comparison between the real ground change diagram and the probabilistic map is shown in Fig. 11. Fig. 11(a) and (b) shows the real pictures of two-time stages in a certain region in the LEVIR dataset, respectively.

It can be found that the former stage is mainly roads and wasteland, while the latter stage is mostly wasteland and trees turned into buildings. From the comparison of real change and probability map, this method can basically identify the ground change, and the gray level distinction between the changed area and the unchanged area is obvious.

Fig. 12 is a comparison of the effects of the various methods with those presented in this article. Fig. 12(a), (b), (c), and (d) shows the results generated by the pixel-based method. As the pixel-based method focuses more on the change of details, the overall effect of the pixel-based method is obviously worse than that of the feature-level method. Fig. 12(e), (f), and (g) shows

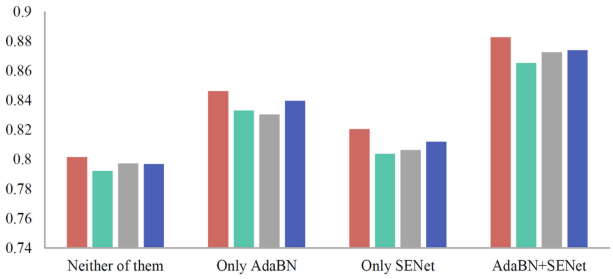

Fig. 13. Ablation study.

the performance of some advanced methods on the data with extremely uneven feature distribution. These results are obtained after the training of the whole dataset sample.

Although the effect is good, the training cost is much higher than the method proposed in this article. Fig. 12(h), (i), and (j) show an experiment of the attention CD network of three-channel s. According to the data, the convolution of upper and lower sampling of one layer is more suitable for feature extraction of multichannels. The two-layer network with up and down sampling makes the changing regions stick together. Although convolution can extract features, too many layers make the network pay more attention to features than to channels. The channel attention network without up and down sampling structure recognizes many false changes.

4) Ablation Experiments: In this section, the ablation experiment was used to verify the effectiveness of the proposed method. In order to reduce the difference in data feature distribution, AdaBN was introduced into UNet. Since there is a strong correlation between each segmentation feature difference map and the final CD result, to make efficient use of this correlation, we introduce the correlation between SENET learning channels in the generation stage of the change probability map. The following experiments will respectively discuss the absence of AdaBN and SENet in the method. Both modules are absent and both modules are present.

To make the proposed method more convincing, we conducted experiments on LEVIR, a data set with the largest difference in feature distribution. The other settings of the experiment were not changed at all. Since structure B performed best in the channel attention network, we chose structure B to test the detection effect of each module in the ablation experiment. The experimental indicators are shown in Fig. 13.

As can be seen from the figure, when the two modules are not added, the indicators of the method are low, which is similar to the effect of the traditional deep learning method. In the experiment of "without SENet," AdaBN was added separately, and it can be seen that AdaBN can effectively improve the detection effect of the method, and all indicators in the results have been significantly increased. However, in the experiment of "without AdaBN," the improvement of network detection effect by adding SENet alone was not so great. The experiment with both AdaBN and SENet achieves the highest detection precision and recall, which proves that the combination of AdaBN and SENet enables the network to detect changes in the ground more accurately.

V. Discussion

A. Adaptive Semantic Segmentation

The CD method proposed in this article shows that the difference in feature distribution among different remote sensing image data sets can be solved by improving the structure of the feature extraction stage. In this paper, UNet structure is used to extract image features. In order to extract the distribution characteristics of different data sets, the structure of UNet is improved by adding AdaBN layer. The addition of AdaBN layer can balance the feature distribution distance between different data. This simple change enables Ada-UNet to improve the semantic segmentation accuracy of source domain data to 94.5% in the case of a small number of training data sets.

In the target domain data, we use the method of weight sharing to make the semantic segmentation conditions of different stages in the same region the same. Since AdaBN needs to propagate forward once to update the parameters of batch normalization processing, we first use images of different stages in the same region as input data to propagate forward once. In this stage, this study did not evaluate the semantic segmentation effect but directly reflected the detection result in the CD graph in the next stage, which needs to be considered in future research.

B. Channel Attention Network

If the channel attention network is directly used in the 1×1 Conv, many noises may be judged as changing pixels. However, too many convolutional layers will lead to too many parameters to be trained, making it difficult for the network to be trained. It is more likely to encounter the problem of fewer training samples. Therefore, the attention CD network of three-channel proposed according to the problem with a small number of samples is shown in Fig. 6. In the experiment, we also carried out corresponding effect verification. Fig. 6(b) has the best effect among the three structures. According to the evaluation index, it can be found that the more unbalanced the data distribution is, the better the effect of Structure B in this method is compared with other methods.

For the channel attention network in this article, we find that the overall accuracy is around 90%. Although it performs well in unsupervised tasks, strict threshold segmentation of the data set is still needed in the method. By analyzing the results of strict threshold segmentation, we find that even under such strict conditions, there are still a small number of pixels that are false changes. Therefore, future research can focus on how to use reinforcement learning, continuous learning, and other means to improve the authenticity of network training data.

C. Threshold Analysis

From the analysis of the above results, it can be found that among the three kinds of channel attention networks proposed in this article, structure (b) has the best comprehensive performance.

So we only analyze the strict threshold difference diagram required by training structure (b) and the probabilistic map generated by structure (b) when analyzing the threshold value.

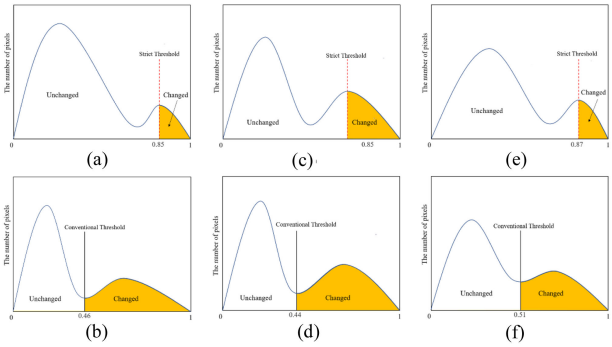

Fig. 14. Pixel distribution of difference map and probabilistic map. (a) and (b), (c) and (d), and (e) and (f) are the gray value distribution and probability value distribution of difference map in Guangzhou, SZTAKI, and LEVIR, respectively.

The difference map used in the experiment on the three datasets and the pixel distribution of the generated probabilistic map is shown in Fig. 14. Among them, the difference between the data sets a and b is Guangzhou figure and probabilistic map pixel distribution, can be seen from the diagram, the figure to take advantage of the critical threshold segmentation results training channel network, 0.85, the threshold value is greater than 0.85 pixels as a real ground probabilistic map generated by the pixel 0.46 threshold value can get Fig. 8(i). The rest of the two data sets are as shown.

By analyzing the threshold changes, we can find that the network can draw away the distance between the changed pixels and the unchanged pixels in the data set with a balanced distribution of features, making it easier to use the conventional threshold for segmentation.

But for the balance of the distribution characteristics of data sets, feature extraction has a certain influence, Fig. 14(f) said LEVIR probabilistic map of a data set that obvious than Guangzhou dataset and SZTAKI more fuzzy pixel distribution of the dataset, this is because the characteristics of the distribution of unbalanced datasets feature extraction difficulty bigger although there are very strong in this article, the structure of the domain adaptive feature extraction ability, but the distribution characteristics of balanced data still have some influence.

D. Comparison of Parameters

In order to better illustrate the superiority of the proposed method, we make a horizontal comparison in terms of the number of parameters. The deep learning methods used in the experiment of this article mainly include VGG-LR based on the VGG network and PCANet based on CNN. At the same time, we also compared the parameters before and after the addition of the SENet module and AdaBN module. The results of parameter comparison are shown in Fig. 15.

By calculating the parameters in the experiment, it can be found that the method proposed in this article has a relatively small number of parameters. This is not only because UNet is a lightweight neural network, but also because SENet and AdaBN modules almost do not add any parameters to the network. This means that adding AdaBN and SENet into the same network

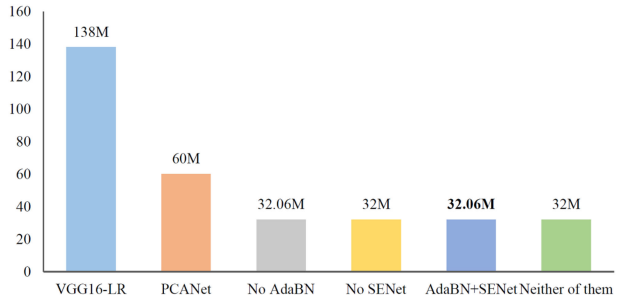

Fig. 15. Comparison of parameters.

structure can greatly improve the accuracy of the model without greatly increasing the model parameters.

SENet plays a significant role in the proposed approach and does not result in a significant increase in the number of parameters. At the same time, CCNet, SKNet, and other modules, which have similar functions with the SENet module, did not do experiments to verify the relationship between parameters and accuracy. These will be the focus of our future research.

VI. Conclusion

In this article, we propose a domain adaptive CD method based on segmentation map difference. In this method, the pretrained Ada-UNet is used for semantic segmentation of the original image in the CD datasets. The semantic segmentation feature map and the original image difference map was combined, and then strict threshold segmentation was performed on the original image difference grayscale image. The combined high-dimensional feature map and strict threshold segmentation map were used to train the channel attention network. A probabilistic map is generated by the trained channel attention network, and the CD results are obtained by dividing the probabilistic map.

Compared with the existing methods, our method has a better comprehensive performance when the experiments are carried out in datasets with different feature distributions. More importantly, it is a method that can be trained with only a small amount of data, which solves the problem that the data feature distribution for training and testing needs to be consistent, and it is more applicable to practical situations.

References

[1] A. P. Tewkesbury *et al.*, "A critical synthesis of remotely sensed optical image change detection techniques," *Remote Sens. Environ.*, vol. 160, pp. 1–14, 2015.

[2] A. Ertürk, M.-D. Iordache, and A. Plaza, "Sparse unmixing with dictionary pruning for hyperspectral change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 321–330, Jan. 2016.

[3] I. R. Hegazy and M. R. Kaloop, "Monitoring urban growth and land use change detection with GIS and remote sensing techniques in Daqahlia governorate Egypt," *Int. J. Sustain. Built Environ.*, vol. 4, no. 1, pp. 117–124, 2015.

[4] X. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 5, pp. 1566–1577, Sep. 1998.

[5] Y. Bazi, L. Bruzzone, and F. Melgani, "Automatic identification of the number and values of decision thresholds in the log-ratio image for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 349–353, Jul. 2006.

[6] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.

[7] J. S. Deng *et al.*, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, 2008.

[8] M. Hussain *et al.*, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogrammetry Remote Sens.*, vol. 80, pp. 91–106, 2013.

[9] A. Ghosh, N. S. Mishra, and S. Ghosh, "Fuzzy clustering algorithms for unsupervised change detection in remote sensing images," *Inf. Sci.*, vol. 181, no. 4, pp. 699–715, 2011.

[10] K. Zhang *et al.*, "Active contours with selective local or global segmentation: A new formulation and level set method," *Image Vis. Comput.*, vol. 28, no. 4, pp. 668–676, 2010.

[11] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2019.

[12] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, 2019, Art. no. 27.

[13] C. Wu *et al.*, "A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion," *Remote Sens. Environ.*, vol. 199, pp. 241–255, 2017.

[14] W. Chen, L. Zhang, and D. Bo, "Kernel slow feature analysis for scene change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2367–2384, Apr. 2017.

[15] L. Wan, Y. Xiang, and H. You, "An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9941–9959, Dec. 2019.

[16] F. C. Stingo *et al.*, "An integrative Bayesian modeling approach to imaging genetics," *J. Amer. Stat. Assoc.*, vol. 108, no. 503, pp. 876–891, 2013.

[17] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNet++," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1382.

[18] D. Teney and J. Piater, "Multiview feature distributions for object detection and continuous pose estimation," *Comput. Vis. Image Understanding*, vol. 125, pp. 265–282, 2014.

[19] J. Azorin-Lopez *et al.*, "Constrained self-organizing feature map to preserve feature extraction topology," *Neural Comput. Appl.*, vol. 28, pp. 439–459, 2017.

[20] D. Liu *et al.*, "PDAM: A panoptic-level feature alignment framework for unsupervised domain adaptive instance segmentation in microscopy images," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 154–165, Jan. 2021.

[21] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, 2018.

[22] M. Talo *et al.*, "Application of deep transfer learning for automated brain abnormality classification using MR images," *Cogn. Syst. Res.*, vol. 54, pp. 176–188, 2019.

[23] J. Yu, D. Tao, and M. Wang, "Adaptive hypergraph learning and its application in image classification," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3262–3272, Jul. 2012.

[24] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster r-cnn for object detection in the wild," in *Proc. Conf., IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 3339–3348.

[25] Z. Li, Q. Wu, B. Cheng, L. Cao, and H. Yang, "Remote sensing image scene classification based on object relationship reasoning CNN," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2020.3017542.

[26] X. Li, W. Wang, X. Hu *et al.*, "Selective Kernel networks," 2019, *arXiv:1903.06586*.

[27] Z. Huang *et al.*, "CCNet: Criss-cross attention for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2020.3007032.

[28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[29] X. Wang *et al.*, "Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 276.

[30] A. Song *et al.*, "Change detection in hyperspectral images using recurrent 3-D fully convolutional networks," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1827.

[31] R. C. Daudt, B. L. Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.

[32] Z. Zhang *et al.*, "Change detection between multimodal remote sensing data using Siamese CNN," 2018, *arXiv:1807.09562*.

[33] M. J. Cracknell and A. M. Reading, "Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information," *Comput. Geosciences*, vol. 63, pp. 22–33, 2014.

[34] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Subsense: A universal change detection method with local adaptive sensitivity," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 359–373, Jan. 2015.

[35] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[36] Y. Fang, Z. Chen, W. Lin, and C.-W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3888–3901, Sep. 2012.

[37] B. Bhanu, S. Lee, and J. Ming, "Adaptive image segmentation using a genetic algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 12, pp. 1543–1567, Dec. 1995.

[38] Z. Wang and C. Gong, "A multi-faceted adaptive image fusion algorithm using a multi-wavelet-based matching measure in the PCNN domain," *Appl. Soft. Comput.*, vol. 61, pp. 1113–1124, 2017.

[39] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, "Strong-weak distribution alignment for adaptive object detection," in *Proc. Conf., IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6956–6965.

[40] J. Li, J. Zhao, and K. Lu, "Joint feature selection and structure preservation for domain adaptation," in *Proc. 25th Int. Joint Conf. Artif. Intelli.*, 2016, pp. 1697–1703.

[41] L. Zhang, J. Fu, S. Wang, D. Zhang, Z. Dong, and C. L. Philip Chen, "Guide subspace learning for unsupervised domain adaptation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3374–3388, Sep. 2019.

[42] J. Xie, L. Zhang, L. Duan, and J. Wang, "On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis," in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, 2016, pp. 1–6.

[43] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2200–2207.

[44] Y. Li *et al.*, "Adaptive batch normalization for practical domain adaptation," *Pattern Recognit.*, vol. 80, pp. 109–117, 2018.

[45] X. Zhang *et al.*, "Pyramid channel-based feature attention network for image dehazing," *Comput. Vis. Image Understanding*, vol. 197, 2020, Art. no. 103003.

[46] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. European Conf. Comput. Vis.*, 2018, pp. 286–301.

[47] J. Fu *et al.*, "Dual attention network for scene segmentation," pp. 3146–3154, *arXiv:1809.02983*.

[48] T. Panboonyuen *et al.*, "Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning," *Remote Sens.*, vol. 11, no. 1, 2019, Art. no. 83.

[49] P. Hamel and D. Eck, "Learning features from music audio with deep belief networks," in *Proc. 11th Int. Soc. Music Inf. Retrieval Conf.*, 2010, pp. 339–344.

[50] C. Xing, L. Ma, and X. Yang, "Stacked denoise autoencoder based feature extraction and classification for hyperspectral images," *J. Sensors*, vol. 2016, pp. 1–10, 2016.

[51] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[52] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[53] Z.-H. Zhou, "Learnware: On the future of machine learning," *Front. Comput. Sci.*, vol. 10, no. 4, pp. 589–590, 2016.

[54] X. Tong, Q. Lu, G. Xia, and L. Zhang, "Large-scale land cover classification in Gaofen-2 satellite imagery," in *Proc. Conf., IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 3599–3602.

[55] I. B. Mohamad and D. Usman, "Standardization and its effects on K-means clustering algorithm," *Res. J. Appl. Sci., Eng. Technol.*, vol. 6, no. 17, pp. 3299–3303, 2013.
[56] C. Gulcehre *et al.*, "Learned-norm pooling for deep feedforward and recurrent neural networks," 2014, *arXiv:1311.1780*.
[57] F. Xia, T.-Y. Liu, J. Wang, W. Zhang, and H. Li, "Listwise approach to learning to rank: Theory and algorithm," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1192–1199.
[58] M. Volpi and V. Ferrari, "Structured prediction for urban scene semantic segmentation with geographic context," in *Proc. Joint Urban Remote Sens. Event (JURSE)*, 2015, pp. 1–4.
[59] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multilayer conditional mixed Markov model," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3416–3430, Oct. 2009.
[60] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021.
[61] X. Wu, Z. Shi, and Z. Zou, "A geographic information-driven method and a new large scale dataset for remote sensing cloud/snow detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 174, no. 11, pp. 87–104, 2021.
[62] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and K-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.
[63] D. Liu *et al.*, "Using local transition probability models in Markov random fields for forest change detection," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2222–2231, 2008.
[64] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep Siamese convolutional network for optical aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.
[65] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1792–1796, Dec. 2016.

Huakang Tang is currently working toward the master's degree in control engineering with the School of Electrical Engineering, Guizhou University, Guiyang, China.

His research interests include remote sensing image processing, deep learning, and change detection.

Honglei Wang received the Ph.D. degree in management science and engineering from the Chinese Academy of Sciences, Beijing, China, in 2007.

He is currently a Professor with the School of Management and the School of Electrical Engineering, Guizhou University (GZU), Guiyang, China; the Director of the Key Laboratory of "Internet+" Collaborative Intelligent Manufacturing, Guizhou Province, China; the Executive Director of the China Society of Systems Engineering; the Director of the China Society of Optimal Selection, Overall Planning, and Economic Mathematics; the Chairman of the Guizhou System Engineering Society; and the Vice President of the Guizhou Automation Society. In the past three years, 11 papers have been authored or coauthored by him in major journals and international conferences. His research interests include manufacturing big data and intelligent manufacturing, and management system engineering (optimization and decision making).

Dr. Wang was the recipient of the second and third prizes of the Science and Technology Progress in Guizhou Province. His research was funded by the National Natural Science Foundation of China.

Xiaoping Zhang, photograph and biography not available at the time of publication.