A Siamese Network Based U-Net for Change Detection in High Resolution Remote Sensing Images

Tao Chen, *Senior Member, IEEE*, Zhiyuan Lu, Yue Yang, Yuxiang Zhang, *Member, IEEE*, Bo Du, *Senior Member, IEEE*, and Antonio Plaza, *Fellow, IEEE*

***Abstract*—Remote sensing image change detection (RSICD) is a technique that explores the change of surface coverage in a certain time series by studying the difference between multiple remote sensing images (RSIs) collected over the same area. Traditional RSICD algorithms exhibit poor performance on complex change detection (CD) tasks. In recent years, deep learning (DL) techniques have achieved outstanding results in the fields of RSI segmentation and target recognition. In CD research, most of the methods treat multitemporal remote sensing data as one input and directly apply DL-based image segmentation theory on it while ignoring the spatio-temporal information in these images. In this article, a new siamese neural network is designed by combing an attention mechanism (Siamese_AUNet) with UNet to solve the problems of RSICD algorithms. SiameseNet encodes the feature extraction of RSIs by two branches in the siamese network, respectively. The weights are shared between these two branches in siamese networks. Subsequently, an attention mechanism is added to the model in order to improve its detection ability for changed objects. The models are then compared with conventional neural networks using three benchmark datasets. The results show that the Siamese_AUNet newly proposed in this article exhibits better performance than other standard methods when solving problems related to weak CD and noise suppression.**

***Index Terms*—Attention blocks, change detection (CD), remote sensing, siamese networks.**

I. Introduction

Remote sensing image change detection (RSICD) is a technique to identify interested differences in land characteristics that can obtain real-time and accurate information on surface changes. Conducting land use type change detection (CD) surveys is of great significance for better protecting ecological environments, managing natural resources, studying social development, and understanding the relationship between humans and nature [1], [2].

In the early days, the main researches were focused on medium-resolution remote sensing images (RSIs). As a result, detectable surface changes were only performed at large spatial scales, such as land surveys [3], urban studies [4], ecosystem monitoring [5], disaster monitoring and assessment [6], [7]. In recent years, with the access to more and more high-quality RSIs, a large amount of remote sensing data has provided data support for further exploration of more efficient and accurate RSICD methods. The traditional algorithms applied to medium-resolution RSIs are no longer satisfied the needs of current application. Along with the increasing spatial resolution, the feature information is more abundant, and problems, such as poor separability, high rate of missed detection and strong interference brought by the mixing of changed and unchanged features are intensified. The detection ratio is poor in practical problems. Therefore, new algorithms with strong robustness, excellent learning ability and high accuracy of results are urgently needed.

The traditional methods could be summarized into image algebraic methods and image transformation methods, focusing on medium resolution data [8], [9]. With the development of machine learning (ML), a variety of ML models have been applied to RSICD. For example, support vector machines [8], decision trees [9], random forests [10], and long short-term memory models [11] have been applied, while a variety of information-assisted classification methods combining RSIs and GIS have also been applied [12], [13]. Overall, as a newly developed field, RSICD still suffers from low accuracy, poor applicability and severely is severely limited by image noise [14].

In order to exploit high-resolution RSIs in CD while eliminating false detection noise, object-oriented methods are of particular importance. More and more researches combine ML algorithms with object-oriented data analysis methods to adopt segmentation before computation to detect changes at the object

Manuscript received November 18, 2021; revised January 21, 2022 and February 17, 2022; accepted March 2, 2022. Date of publication March 11, 2022; date of current version March 23, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62071439 and Grant 61871259, in part by the Opening Foundation of Qilian Mountain National Park Research Center (Qinghai) under Grant GKQ2019-01, in part by the Opening Foundation of Beijing Key Laboratory of Urban Spatial Information Engineering, under Grant 20210209, and in part by the Opening Foundation of Geomatics Technology and Application Key Laboratory of Qinghai Province, under Grant QHDX-2019-01. *(Corresponding author: Tao Chen.)*

Tao Chen, Yue Yang, and Yuxiang Zhang are with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China (e-mail: taochen@cug.edu.cn; lunayoung1020@cug.edu.cn; zhangyx@cug.edu.cn).

Zhiyuan Lu is with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China, and also with the Institute of Computing Technology, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China (e-mail: luzhiyuan@rails.cn).

Bo Du is with the School of Computer Science, Wuhan University, Wuhan 430079, China (e-mail: gunspace@163.com).

Antonio Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10071 Cáceres, Spain (e-mail: aplaza@unex.es).

Digital Object Identifier 10.1109/JSTARS.2022.3157648

level in RSIs. Lefebvre *et al.* [15] proposed an object-oriented analysis method combining RSIs, geometric features, and texture analysis methods for surface CD. Gamanya *et al.* [16] used a hierarchical image segmentation method to extract objects from multitemporal RSIs and applied a standardized, fuzzy logic-based, automatic object-oriented classification method to successfully obtain changes in the Harela region of Zimbabwe. Bontemps *et al.* [17] proposed an automatic probabilistic CD method and applied it to detect long time series changes in the tropical forest environment of the state of Lonja, Brazil, between 2001 and 2004. In addition, there are some other methods that consider the relationship between pixels in the neighborhood space and combine those adjacent pixels for analysis. For example, Hao *et al.* [18] carried out unsupervised CD using level sets. Bruzzone and Prieto [19] performed automatic unsupervised CD by Markov random fields. The conditional random field method used by Zhou *et al.* [20] in their CD study on high resolution RSIs. The above three methods are successfully applied to perform object-oriented CD, which further introduce spatial information into the spectral information and improves the accuracy of object-oriented CD. After early exploration, CD has achieved increasingly higher accuracy high accuracy. However, there are too many human interventions and threshold adjustments, which leads to a poor robustness of the model. Although for specific problems it can achieve good results, there are problems, such as a low degree of automation, poor noise suppression, complex features, poor accuracy.

The emergence of deep learning (DL) has brought in a new direction to the research on RS, and associated research has been carried out in numerous fields, such as ecological evaluation [21], [22], RS segmentation [23], landslide detection [24], [25], image classification [26]–[28], etc. In the field of CD, researchers exploit DL to alleviate the problems of complex feature detection, strong noise interference, poor separability, and low automation in RSICD. Various DL-based methods for RSICD have emerged in the last few years. Amin *et al.* [29] proposed a new convolutional neural network (CNN) feature CD method based on high-resolution RSIs, which uses pretrained CNN to generate change result maps directly from pair images. Varghese *et al.* [30] proposed a parallel CNN architecture for locating and identifying changes between street scene image pairs. Gao *et al.* [31] proposed a deep cascade network (DCNet) for synthetic aperture radar (SAR) image CD to extract features, and introduced residual learning to solve the blast gradient problem. Song *et al.* [32] proposed a DL model integrating a sample generator and a fully CNN on hyperspectral data. Peng *et al.* [33] proposed a new end-to-end CD method on high-resolution RSIs in order to release the error accumulation problem in the training process.

However, most of the above methods are based on traditional RSI segmentation algorithms. Multitemporal data are often simply overlaid to serve as training data for DL algorithms, which ignores the characteristics of changed ground features itself caused by spatial, temporal, and other factors. Satisfactory results are often not achieved in complex scenes, such as seasonal and/or lighting changes. In order to solve this problem, it is necessary to pay attention to the correlation properties between multi-temporal data. siamese networks can be used for image comparison by multiple inputs [34], which is suitable for RSICD. Du *et al.* [35] proposed a Deep slow feature analysis CD algorithm for multitemporal RSIs based on deep networks and slow feature analysis theory. Guo *et al.* [36] proposed a new fully convolutional siamese metric network and used a custom contrast loss method to improve the variability of changing scenes while reducing the variability of invariant scenes. Chen *et al.* [37] proposed a new dual-attentional fully convolutional siamese network (DASNet) for RSICD in high-resolution RSIs, which captured long-range dependencies, obtained more discriminative feature representations, and improved the recognition performance of the model through a dual-attention mechanism. Sakurada *et al.* [38] constructed a semisupervised scene CD network to reduce the errors caused by camera point-of-view differences during multitemporal remote sensing imaging. Chen *et al.* [39] proposed a neural network model relying on siamese network and self-attentive mechanism to solve the error effects and alignment errors in dual-temporal RSICD studies.

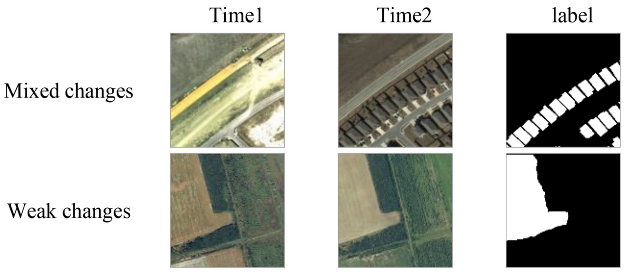


Fig. 1. Challenges in change detection.

Several of the above methods focus only on the siamese structure or the combination of siamese structure with other theories. However, relatively simple methods are only appropriate for reducing differences from features extracted from siamese structure. The models are not pure neural network models and often require additional processing. In order to further improve the CD accuracy, two important problems must be solved, which are mixed changes and weak changes. For easy understanding, these two cases are given in Fig. 1. Mixed changes indicate that there may be a variety of change scenarios interfering with the task, e.g., wasteland to house change is noticed, while wasteland to concrete surface change is ignored. Weak changes indicate that a change has occurred, but it is intuitively difficult to distinguish. An example is the reclamation of wasteland. The land may be bare before, however after reclamation, no significant change can be detected from the RSIs.

In order to address the above-mentioned problems, we propose a new RSICD framework based on siamese networks, named Siamese_AUNet. The two main contribution of the article can be summarized as follows.

1) Conventional CD models use image overlay analysis, which ignores the temporal variability between data. In this article, the siamese structure is used to extract features from images of two different periods separately, with

both spatial correlation and temporal variability between images.

2) Then atrous spatial pyramid pooling (ASPP) [40] was subsequently computed on the feature matrix to increase the multiscale feature detection capability of the model. In addition, the feature attention model (FAM) was added to increase the detection capability of the model as well as the noise suppression capability.

The rest of the article is organized as follows. Section II introduces the related theories and methods. Section III describes the parameters, accuracy evaluation indexes, the three datasets used in this article, and the experimental results. Section IV discusses the experimental contents. Section V concludes the article with and gives some plausible future research lines.

II. Proposed Siamese_AUNet

A. Siamese Networks

Siamese networks are a coupled network architecture built on two artificial neural networks [34]. It computes two input samples separately with the same network weights and outputs their representations embedded in a high-dimensional space to compare the degree of similarity of the two samples. During the training process, one sample is computationally encoded by the convolutional network to obtain a set of features, after which the weights and bias parameters of this network are maintained and the same computational encoding is performed on another sample. Finally, the similarity of these two sets of features is subsequently compared by training learning.

By processing the two sets of samples with the same network parameters, it can better reduce the distance of sample values between unchanged regions and increase the distance of samples in changed regions.

B. Feature Attention Module

In this article, we use an FAM for spatial and channel attention on the deep feature layer. The construction strategy of the FAM refers to both the nonlocal attention module (nonlocal AM) as well as the convolutional block attention module (CBAM) attention module. To introduce the FAM in more detail, the principles of nonlocal AM and CBAM are explained separately below.

1) Nonlocal AM: The Nonlocal AM is used to capture the long-range dependencies of an image directly by computing the interaction between any two positions of the image [41]. Its computational principle can be represented by (1), where x denotes the input signal and y denotes the output signal with the same sample volume as the input x. $f(x_i, x_j)$ is used to compute the combinatorial relationship between all possible positions j associated with i. $g(x_j)$ is used to compute the eigenvalues of the input signal at position j. $C(x)$ is the normalization parameter. The i represents the corresponding current position, and j is a nonlocal response obtained by weighting, which can be regarded as the global response

$$y_i = \frac{1}{C\left(x\right)} \sum\nolimits_{\forall j} f\left(x_i, x_j\right) g\left(x_j\right). \qquad (1)$$

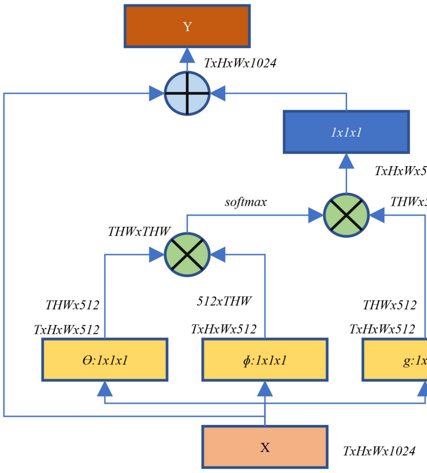


Fig. 2. Structure of Nonlocal AM.

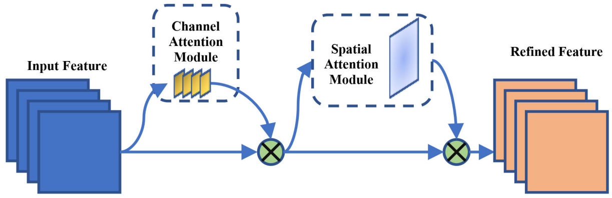


Fig. 3. Structure of CBAM.

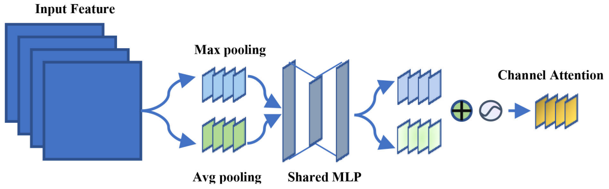


Fig. 4. Structure of CAM.

It can be seen from Fig. 2 that nonlocal AM is essentially a complex convolution operation that preserves the spatial information of the original input signal as well as extracts the dependencies between the inter-internals of the input signal. Its structure is shown in Fig. 2.

2) CBAM: It consists of two blocks, which are the channel attention module (CAM) and the spatial attention module (SAM) [42]. These two blocks are connected in sequence to form a new attention network model, and the structure is shown in Fig. 3.

The interchannel relationship and inter-spatial relationship of the input features are captured by CAM and SAM, respectively. Finally, the feature matrix after the attention process is obtained by multiple matrix operations. The CAM module focuses on learning the feature relationships between different channels and using the inter-channel relationships of the features to create a channel attention map. To compute channel attention efficiently, the module compresses the spatial dimensions of the input feature maps. The input features are processed using average pooling and maximum pooling together and computed through a weight sharing network as shown in Fig. 4.

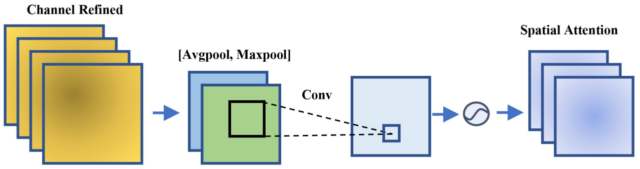


Fig. 5. Structure of SAM.

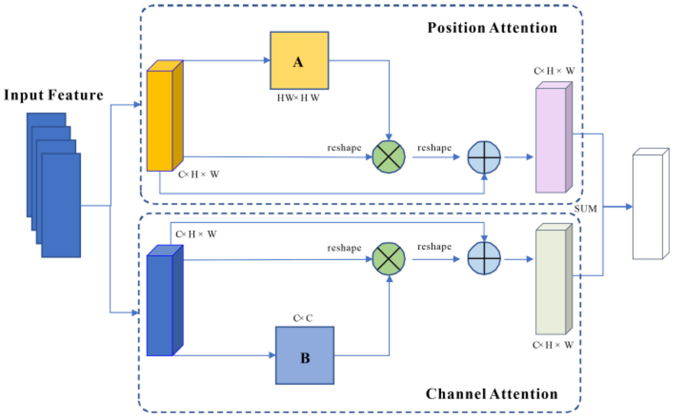


Fig. 6. Structure of FAM.

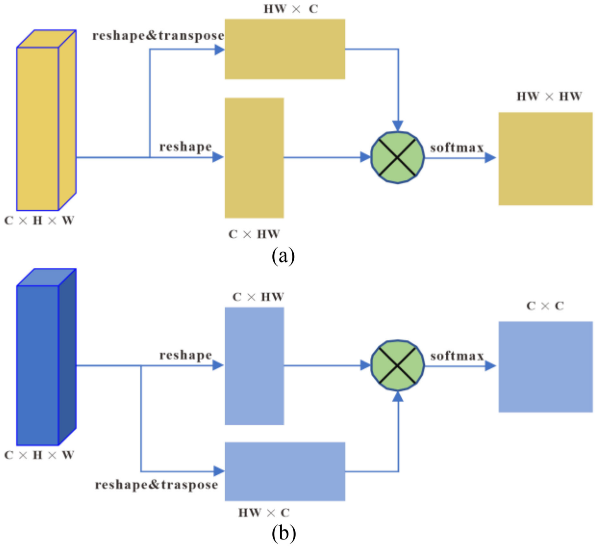


Fig. 7. Structure of A&B in FAM. (a) A block. (b) B block.

Different from CAM, SAM mainly performs global maximum pooling and average pooling operations for pixel values at the same location on different feature layers in the axial direction, and obtains two spatial attention feature layers as shown in Fig. 5.

3) Feature Attention Model: In FAM, we replace the maximum pooling and average pooling methods in CBAM with the convolution operation in nonlocal AM, so that it can process the input signal in spatial and channel dimensions, respectively. The structure of FAM is shown in Fig. 6.

In Fig. 6, A and B represent a matrix operation, respectively. Processing details of A and B are shown in Fig. 7(a) and (b). The position attention and channel attention are represented, respectively. The position attention aims to use the association between any two points features to mutually enhance the representation of the features. Channel attention, on the other hand, can highlight interdependent feature maps by mining the interdependencies between channel maps to enhance the feature representation of specific semantics. Then, the two outputs are summed and fused to obtain the final features for pixel point classification.

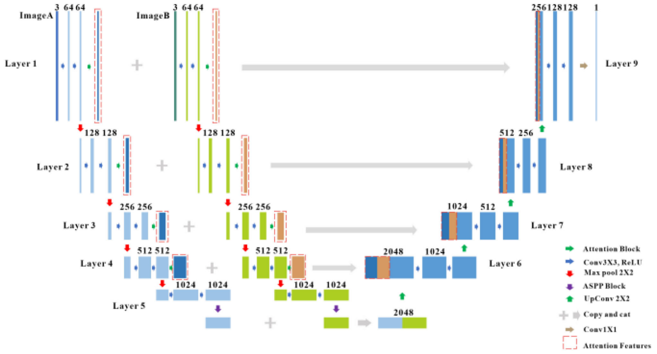


Fig. 8. Structure of Siamese_AUNet.

C. Atrous Spatial Pyramid Pooling

The ASPP structure was first proposed by Google's DeepLabv2 segmentation network [44]. Inspired by spatial pyramid pooling [40], a similar structure was designed that samples a given input sample in parallel with an atrous convolution at different sampling rates, equivalent to capturing the contextual information of the image at multiple scales.

In RSICD, multiscale feature mixing is a very common phenomenon. However, the detection capability of the model decreases when the change objects differ greatly in scale. In this article, we adopt the ASPP mechanism to enhance the multiscale feature learning capability of the model and improve the detection performance of multiscale change objects.

D. Siamese_AUNet Network Architecture and Formulation

The Siamese_AUNet is a siamese feature extraction self-organizing network based on UNet network by improving the structure on the left feature extraction end of the network. UNet is a DL network for image segmentation modified from a fully CNN [43]. The FAM is then applied to each feature layer to perform the attention operation to obtain the combination of the noted features. An ASPP module is added to the bottom end of the siamese network to enhance the multiscale feature learning capability of the network. On the right side of the model is the decoding end, and the change binary map is finally obtained after decoding process. In order to represent the network structure features more intuitively, the parameters are given in Table I, and the structure is shown in Fig. 8.

As shown in Fig. 8, Siamese_AUNet consists of two main parts: the feature extraction siamese network structure for processing multitemporal data and the decoding network for

TABLE I
PARAMETERS OF THE SIAMESE_AUNET

	Layer Type	Filters	Size of Filters
Layer 1	Conv1-1	64	3×3×64
	Conv1-2	64	3×3×64
	Pooling	1	\
	Attention	1	\
Layer 2	Conv2-1	128	3×3×128
	Conv2-2	128	3×3×128
	Pooling	1	\
	Attention	1	\
Layer 3	Conv3-1	256	3×3×256
	Conv3-2	256	3×3×256
	Pooling	1	\
	Attention	1	\
Layer 4	Conv4-1	512	3×3×512
	Conv4-2	512	3×3×512
	Pooling	1	\
	Attention	1	\
Layer 5	Conv5-1	1024	3×3×1024
	Conv5-2	1024	3×3×1024
	ASPP	1024	\
	Combine	\	\
	UpConv	1024	3×3×1024
Layer 6	Combine	\	\
	Conv6-1	1024	3×3×1024
	Conv6-2	1024	3×3×1024
Layer 7	Combine	\	\
	Conv7-1	512	3×3×512
	Conv7-2	512	3×3×512
Layer 8	Combine	\	\
	Conv8-1	256	3×3×256
	Conv8-2	256	3×3×256
Layer 9	Combine	\	\
	Conv9-1	128	3×3×128
	Conv9-2	128	3×3×128
	Conv9-3	1	1×1×1

TABLE II
TRAINING PROCEDURE OF THE PROPOSED SIAMESE_AUNET

```
Input :
    $t0$: pre event image
    $t1$: post-event image
Function Siamese AUNet($t0$, $t1$):
    $a_1, a_2, a_3, a_4, a_5$ = VGGconv($t0$)
    $b_1, b_2, b_3, b_4, b_5$ = VGGconv($t1$)
    $x$ = aspp($a5$)
    $y$ = aspp($b5$)
    $L_6$ = torch.cat(($x$, $y$), $dim$ = 1)
for $i$ in 1 : 4 do
    $L_{6-i}$ = Upsample($L_{7-i}$)
    $L_{6-i}$ = FAM($L_{6-i}$)
    $L_{6-i}$ = torch.cat(($a_{5-i}$, $b_{5-i}$,$L_{6-i}$), $dim$ = 1)
    $L_{6-i}$ = Upconv($L_{6-i}$)
end
$out$ = Conv($L_2$)
return $out$
```

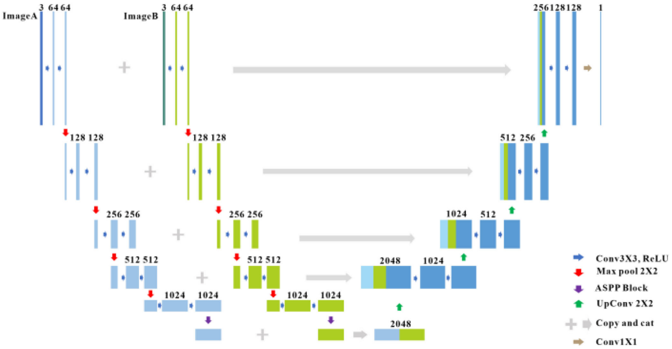


Fig. 9. Structure of SiameseNet.

analyzing feature differences, both of which are constructed from visual geometry group (VGG) networks [41]. In the feature extraction part, the data of different time phases are processed by different branches of the siamese network, and each branch shares the weight parameters. Through the siamese network, the remote sensing data of each time phase are mapped into shallow or deep features. At each layer of the siamese network there is an attention module, which pays attention to the extracted features for subsequent decoding of the network and enriching the feature information. The deep-level features extracted by the siamese network are subsequently enhanced for multi-scale feature representation by an ASPP module, which also uses shared parameters for multi-feature data processing. In the feature difference analysis network, the noted feature matrix is combined with the deep-level feature matrix to characterize the variation of the sample after multiple convolutions and upsampling layers.

In the study of the problem of RSICD, the essence is to identify the semantic information of change. In Table I, Conv4-2, Conv5-2 get the deep semantic information, which can reflect the deep features of the image. However, in the VGG network, the resolution of the original image is reduced by 16 times due to the large number of pooling downsampling operations. In order to solve this problem, we combine the shallow features of Conv1-2, Conv2-2, and Conv3-2, which retain the image details, with the deep features for semantic difference analysis after attention mechanism, to retain the detail information and obtain the deep semantic difference information of the image to a certain extent. It is worth noting that, in order to ensure that each feature layer is in the same range for variation detection, batch normalization is used in each convolutional module in the network.

In the proposed Siamese_AUNet, the siamese structure facilitates the extraction of shallow and deep features between multi-temporal RSIs. The multitemporal phase data processed by the same network structure at the same time have the same size and similar properties for each output. The whole detailed process of training and generating binary change maps for Siamese_AUNet is given in Table II.

In order to evaluate the effect of FAM and the siamese structure, this article improves the UNet and constructs a pure siamese UNet, which we named SiameseNet, and the structure is shown in Fig. 9. Subsequently, AUNet was constructed by adding the same FAM as in the Siamese_AUNet. The structure of AUNet is shown in Fig. 10.

In this section, the model-related structure, theory and algorithms are introduced. The VHR RSIs bring richer detail information, which has an impact on the amount of image information, the amount of GPU operations and the scale of model detection. This article addresses this issue and aims to build a model that is applicable to multiple resolutions and multi-scale detection capability, from the following three aspects.

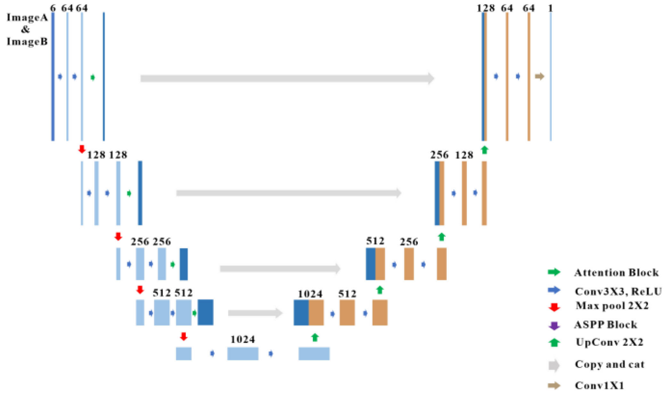


Fig. 10. Structure of AUNet.

1) The siamese network structure is used to minimize the model parameters, and the images at different times share the model parameters for feature extraction separately, which preserves the temporal variability among the data and the difference features.
2) The ASPP structure increases the multiscale feature learning capability of the model. This alleviates the problem of poor multiscale detection caused by the shrinking coverage of VHR images in the same size region and the variable objects with irregular scales.
3) When the image information increases, the uninterested interference information increases as well. Applying FAM can effectively avoid the model to focus too much on useless objects.

Based on the above three points, we expect the constructed model to achieve better results in VHR image CD.

III. Experimental Results and Analysis

In this article, a total of three public datasets of high quality are selected for the experiment. These three datasets include LEVIR dataset [39], WHU building dataset [45], and SZTAKI Air change benchmark dataset [46]. Due to the different spatial resolutions of these datasets and the different scales of the research objects involved, different cropping size are chosen for each dataset during the training process to ensure the optimal efficiency and accuracy. All datasets are 8bit RGB images, and the labels are 8bit grayscale images. The same hyperparameters: “EPOCH;” “BATCH_SIZE;” and “LEARNING_RATE” are set for each network for the same dataset, to determine the performance of different models for the same dataset. It is important to emphasize that each dataset is partially selected in advance as a test set, which is not involved in training and validation datasets. For the training and validation sets, they are randomly divided according to a fixed ratio during the training process. As an indicator, *Time* is used to record the time spent in training the model.

Meanwhile, to make the experiment results more reasonable, a SOTA model comparison experiment was added. Chen *et al.* [37] carried out better work, and their proposed DASNet is one of the best models among many CD models. The comparison experiment with them can greatly increase the reliability of Siamese_AUNet constructed in this article.

BCEWithLogitsLoss was chosen as the loss function for the experiments. This is because in the CD dataset, the ratio of positive and negative samples is very disparate, and the conventional loss function does not work well in dealing with such problems. The BCEWithLogitsLoss can amplify the positive sample weights, which is more conducive to detecting changes.

Four evaluation metrics (EMs) are used to evaluate the accuracy of the results, including Precision, Recall, F1, and Dice. The prediction label is determined by judging the difference between the pixel values of the predicted value and the true value of the same location. The calculations of each EM are as follows.

1) Precision is an indicator to evaluate the proportion of positive samples correctly predicted by the model to all positive samples predicted, and is calculated as follows:

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

2) Recall is an index to evaluate the proportion of positive samples correctly predicted by the model to all true positive (TP) samples, expressed as

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}. \tag{3}$$

3) The F1 value is the summed average of the precision and recall rates, expressed as follows:

$$F1 = \frac{2}{\frac{1}{\mathrm{precision}}+\frac{1}{\mathrm{recall}}} = \frac{2\times\mathrm{precision}\times\mathrm{recall}}{\mathrm{precision}+\mathrm{recall}}. \tag{4}$$

4) The Dice is commonly used to measure the similarity of two samples and is expressed as follows:

$$\mathrm{Dice} = \frac{2\times\mathrm{TP}}{2\times\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}. \tag{5}$$

Where TP/TN is the positive/negative samples which are correctly predicted by the model, FN means the positive samples which are incorrectly predicted as negative samples by the model, FP is the situation that the negative samples which are incorrectly predicted as positive samples by the model.

A. LEVIR Dataset

In LEVIR dataset, a total of 4923 pairs of samples are involved in model training, of which 10% are randomly chosen as validation samples, which are not involved in model training and are only used to evaluate the network training process. Each training Batch contains 12 sets of samples, and the training samples are all iterated once for one Epoch, with a total of ten iterations. The initial learning rate of training is 0.001, and the learning rate is dynamically optimized according to the training process.

In the training process, a python script was constructed to train multiple networks one by one, which realized the sequential experiments of different networks with the same parameters for this dataset. The network models involved were UNet, AUNet, SiameseNet, and Siamese_AUNet. The model was validated and evaluated after using each 640 pairs of images for training. The experimental accuracy of LEVIR dataset was given in Table III.

TABLE III
EXPERIMENTAL ACCURACY OF LEVIR DATASET

	Precision	*Recall*	*F1*	*Dice*	*Time(h)*
DASNet	0.6942	**0.8901**	0.7661	0.7604	1.50
UNet	0.8558	0.7355	0.7888	0.7911	0.97
AUNet	**0.8743**	0.7802	0.8206	0.8248	**0.90**
SiameseNet	0.8216	0.8283	0.8248	0.8329	2.04
Siamese_AUNet	0.8582	0.8702	**0.8557**	**0.8565**	2.59

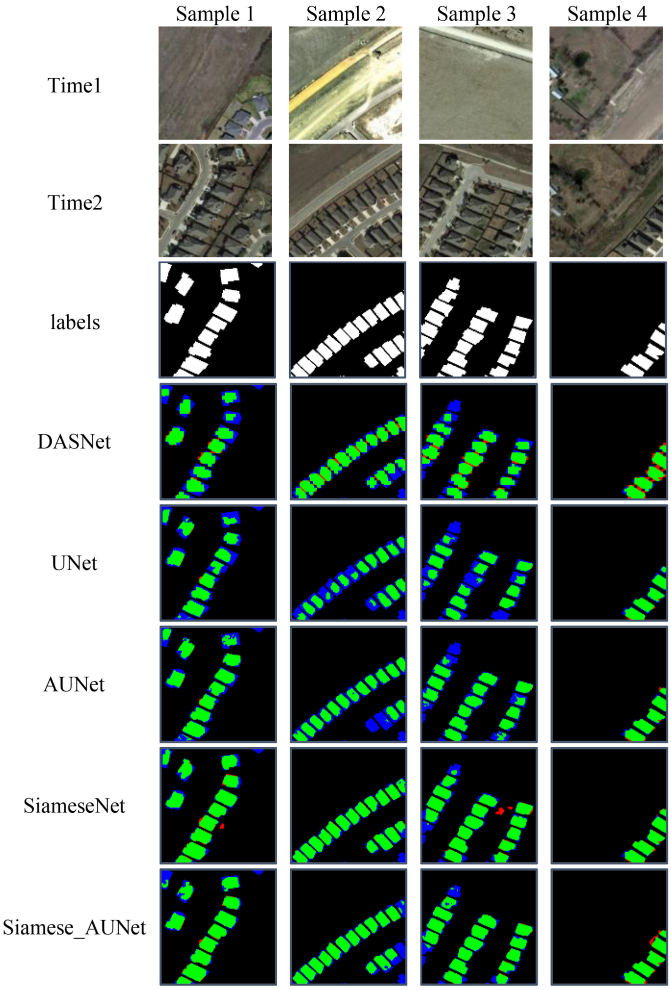


Fig. 11. Experimental results of LEVIR dataset (The rendered colors represent TPs (Green), FPs (red), and FNs (Blue)).

The validation samples result of LEVIR dataset are shown in Fig. 11 and the precision curve comparison of test samples results of LEVIR dataset is shown in Fig. 12.

In the process of accuracy evaluation of LEVIR dataset, the validation samples are not involved in training and are only used as accuracy evaluation data. The accuracy of each validation was recorded and the accuracy curve was plotted by the TensorBoard visualization tool. From Table III and Fig. 12, it can be found that the Siamese_AUNet model maintains the highest accuracy in F1, and Dice throughout the training process, while it ranks second place in Precision metrics. This is because Siamese_AUNet is more detailed in the detection of changing boundaries and has a stronger sensitivity to ground difference detection.

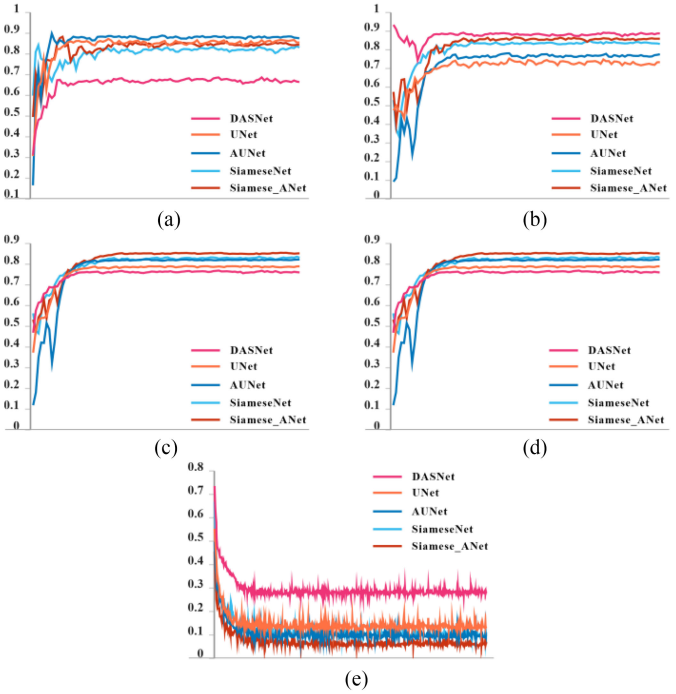


Fig. 12. Precision curve comparison of validation samples results of LEVIR dataset.

But at the same time, it increases the anti-interference ability for the complex ground feature conditions. From Fig. 11, we can also see that the Siamese_AUNet model detects more changed areas, and the boundaries of these detected changed area are more regular, even the sharp angles of changed areas can be identified, which also proves the superiority of Siamese_AUNet compared with the traditional model. From the results, we can also find that the SiameseNet performs much better than the UNet. This also shows that the siamese network is more sensitive to differences, but due to its relatively simple model structure and poor noise suppression ability, it does not perform as well as Siamese_AUNet, which incorporates the attention module on this basis.

B. WHU Building Dataset

In WHU building dataset, there are 1827 pairs of training samples, due to the large-scale variation in the dataset, the size of each image block is set to 512 × 512 pixels in the image cropping step. When the model training, 10% of the data is randomly selected as the validation sample, and each training batch contains four groups of samples. The training samples are all iterated once for one Epoch, 10 iterations in total, and the initial learning rate is 0.001. The learning rate is dynamically optimized according to the training process.

The results are given in Table IV.

The validation sample results chart is shown in Fig. 13 and the validation accuracy curve is shown in Fig. 14.

Analyzing the experimental results, it was found that the DASNet occupied three highest values among four EMs, which were recall, F1, and dice, while the highest value of recall was

TABLE IV
EXPERIMENTAL ACCURACY OF WHU BUILDING DATASET

	Precision	*Recall*	*F1*	*Dice*	*Time(h)*
DASNet	0.8205	**0.9586**	**0.8755**	**0.8754**	1.26
UNet	**0.9174**	0.8048	0.8501	0.8501	**0.58**
AUNet	0.9155	0.6627	0.7590	0.7590	0.61
SiameseNet	0.4784	0.9139	0.6280	0.6280	1.90
Siamese_AUNet	0.8202	0.8633	0.8447	0.8447	2.40

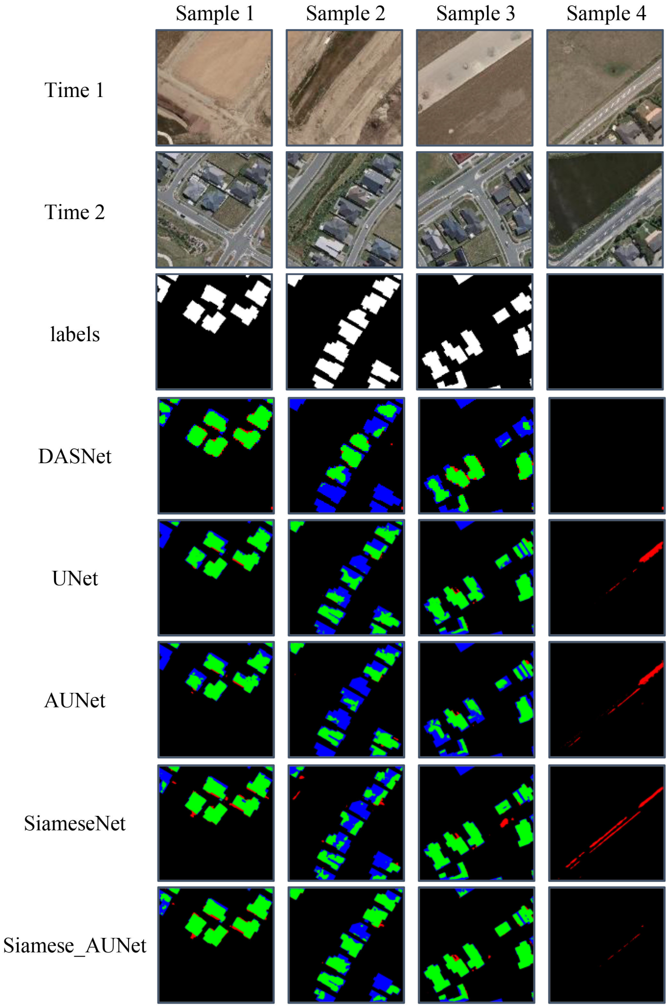


Fig. 13. Experimental results of WHU building dataset (The rendered colors represent TPs (Green), FPs (red), and FNs (Blue)).

obtained by UNet. F1 and Dice indicates the overall superiority or inferiority of the model's performance on a certain dataset. A high Precision indicates that the model has a strong ability to suppress noise during the training of this dataset and a high detection accuracy for the change region, but in the experiments, this often leads to a high omission rate, i.e., a low recall. Conversely, a high "Recall" may appear accompanied by a low "Precision," such as SiameseNet.

Both of the DASNet and Siamese_AUNet maintain a good Recall along with the high Precision metric. The reason DASNet is better than Siamese_AUNet is that the WHU dataset has a higher resolution, which is extremely important for DASNet.

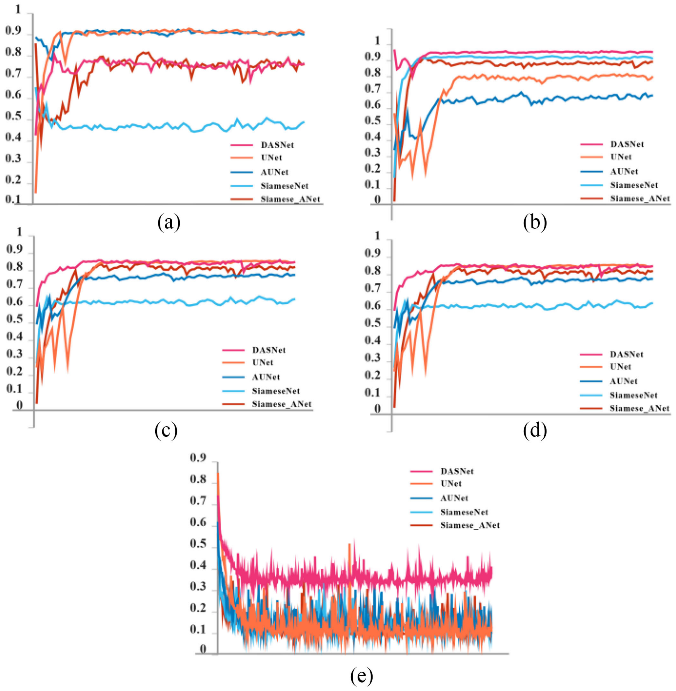
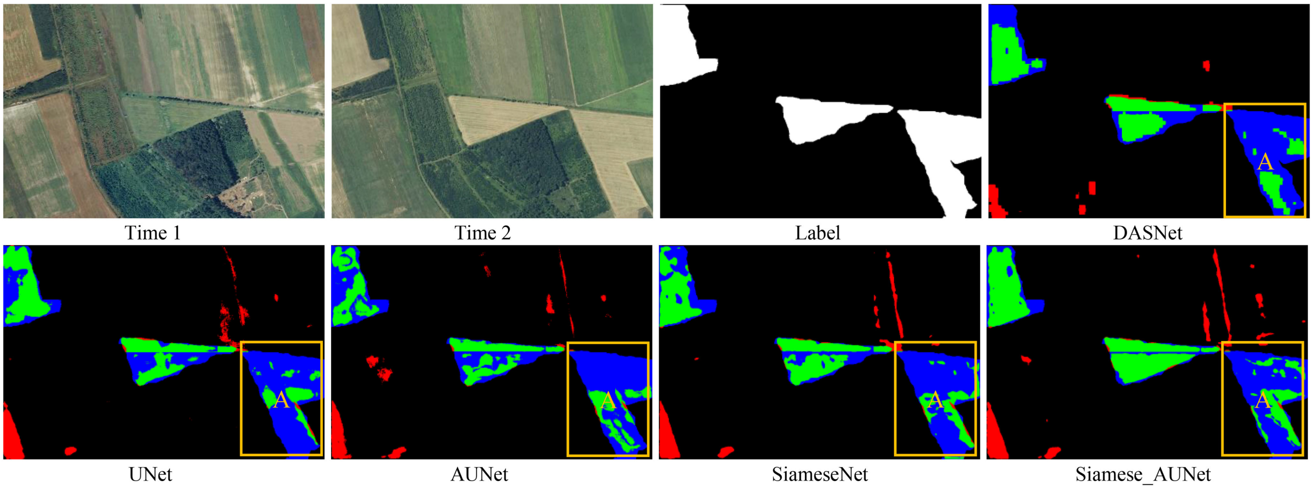


Fig. 14. Precision curve comparison of experimental results of WHU building dataset.

The higher resolution allows it to retain a certain amount of texture information when extracting deep features, and has the ability to better depict boundaries when restoring changing areas. For SiameseNet, which obtained the highest Recall, but very low Precision, indicating that the network predicted variation areas with a high commission error. In this dataset, the UNet model also achieves better results, but from the local results it can be found that the UNet model is more blurred than Siamese_AUNet in the edge regions and is not able to permute the original true change edges; there are also some gaps in the changed objects, while the SiameseNet and the Siamese_AUNet effectively reduce this phenomenon.

Compared with our other models, the siamese network not only improves the detection ability of the changed region, but also greatly increases the commission error while improving the detection ability. As a result, the FP of Siamese_AUNet is effectively suppressed by FAM. The ability that the FAM can suppress the FP is also confirmed in the comparison of the results of UNet and AUNet.

In this experiment, DASNet achieved overall better indicators than our model, although they are very close. After review and experiments, we have known that the detection accuracy is mainly disturbed by pseudochange noise. DASNet only uses deep features in the training process, which are not affected by noise, so the result has a very high recall. However, it abandons the shallow features, which is affected by noise. Its description of the change boundary will not be perfect, especially when the change involves a small number of pixels, it is easier to be ignored as noise. For WHU dataset, its resolution reaches 0.2 m, which will no longer affect DASNet. It not only achieves higher recall, but also higher Precision. For our model, it obtains



Time 1 Time 2 Label DASNet

UNet AUNet SiameseNet Siamese_AUNet

Fig. 15. Experimental results of SZTAKI Air change benchmark dataset (The rendered colors represent TPs (Green), FPs (red), and FNs (Blue)).

TABLE V
EXPERIMENTAL ACCURACY OF SZTAKI AIR CHANGE BENCHMARK DATASET

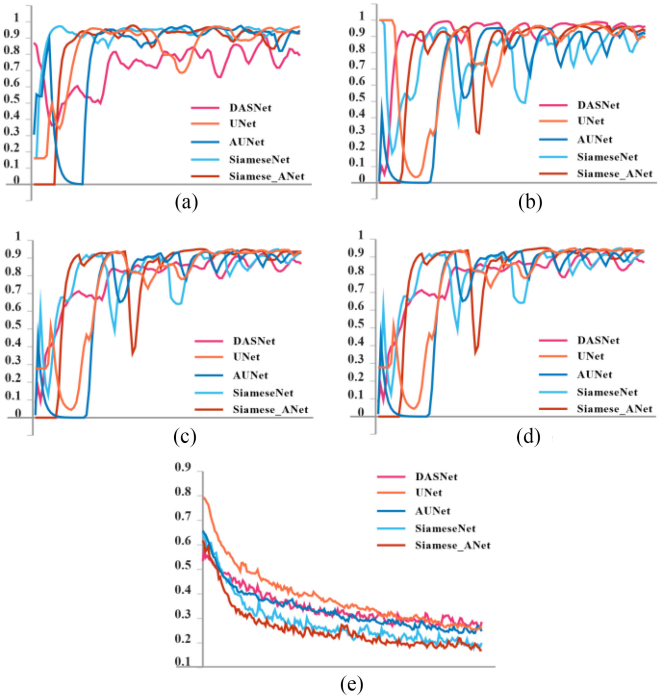
	Precision	*Recall*	*F1*	*Dice*	*Time(h)*
DASNet	0.7641	**0.9691**	0.8546	0.8545	0.06
Unet	**0.9803**	0.8801	0.9276	0.9275	**0.03**
AUNet	0.9661	0.9050	0.9302	0.9346	0.04
SiameseNet	0.9241	0.9271	0.9317	0.9317	0.06
Siamese_AUNet	0.9363	0.9611	**0.9318**	**0.9422**	0.07

semantic information from deep features and texture information with noise from shallow features. Although we have taken a variety of ways to mitigate this interference, it has not achieved the effect of complete removal. Therefore, our model is inferior to DASNet in terms of overall accuracy.

C. SZTAKI Air Change Benchmark Dataset

SZTAKI Air change benchmark dataset is an earlier dataset used in DL-based RSICD. The dataset has a slightly lower image resolution, so it does not focus on detecting changes in object levels, but instead annotates regional changes in artificial buildings, forest land, and cultivated land. There are 122 pairs of samples in this dataset, and the Tiszadob_3 image pair with the size of 512 × 786 × 3 pixel is used as the verification set. All samples are finally cut into 256 × 256 pix image patches. This dataset involves many variations and the amount of data is small, so 20 epochs were set during the training, and the validation accuracy was calculated for each epoch. In addition, the model training process can be monitored in real time by verifying the accuracy and loss curve to avoid overfitting. Each batch contains six sets of samples. Because the sample size of this model is small, the learning rate is fixed at 0.001 to ensure learning stability. Multimodel experiments were carried out, and the accuracy index results are given in Table V. The validation example graph shown in Fig. 15 and the validation accuracy curve is shown in Fig. 16.

It can be found from the results that Siamese_AUNet achieves two highest values among the four Ems, which is better than the other four models. The highest value of the Precision index is obtained by the Unet, which means that the model has a higher detection rate, and the highest value of the Recall index is obtained by the DASNet, which means that the model has a lower detect error. From Fig. 15, we can see that the missed detection area of Siamese_AUNet is concentrated in the "A" corner of the image. Meanwhile, it can be found that Siamese_AUNet achieves results that are closer to the real change situation than the labels, and the impact of label quality on the model will be discussed in detail in the following section. In addition, although the proposed models all adopt the means to preventing overfitting, due to the small sample size of this



(a) (b) (c) (d) (e)

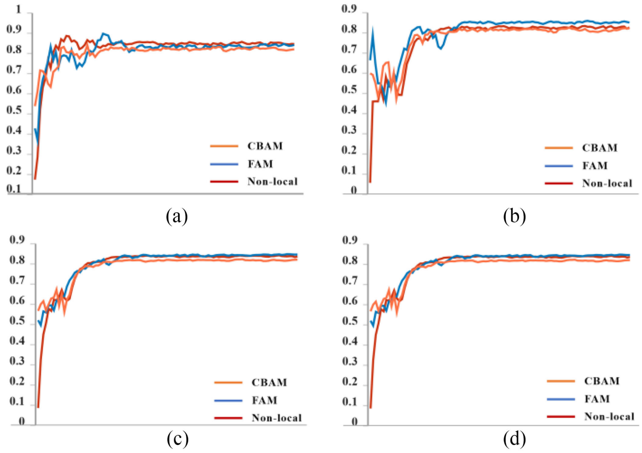
Fig. 16. Precision curve comparison of experimental results of SZTAKI Air change benchmark dataset.

TABLE VI
EXPERIMENTAL ACCURACY OF LEVIR

Attention block	*Precision*	*Recall*	*F1*	*Dice*	*Time(h)*
Non-local	0.8491	0.8327	0.8341	0.8356	**2.28**
CBAM	0.8313	0.8263	0.8205	0.8216	2.30
FAM	**0.8603**	**0.8753**	**0.8581**	**0.8572**	2.56

TABLE VII
EXPERIMENTAL ACCURACY OF WHU

Attention block	*Precision*	*Recall*	*F1*	*Dice*	*Time(h)*
Non-local	0.6938	0.7966	0.7417	0.7416	**1.71**
CBAM	0.8113	0.8257	0.8138	0.8126	1.84
FAM	**0.8261**	**0.8587**	**0.8403**	**0.8411**	2.21



(a) (b)

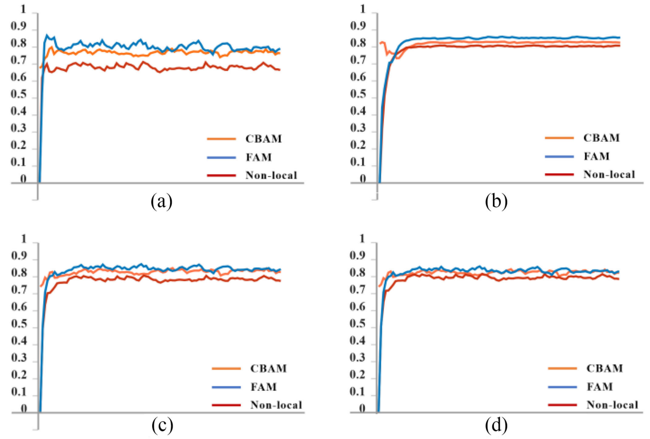
(c) (d)

Fig. 17. Ablation experiment accuracy of LEVIR.

dataset, the real-time change of loss values during the training of the model is recorded using the TensorBoard tool in order to prevent training overfitting. As seen in Fig. 16, the loss values continued to decrease and no overfitting occurred. The changes of the accuracy index of each validation were recorded during the training process, as shown in Fig. 16. Due to the small number of validation samples and the fact that some image pairs were randomly selected from the validation samples for each validation, there is a wide range of fluctuations in the validation accuracy (for example, the A marked with yellow in Fig. 15) when there are image pairs with large differences between the validation results and the real labels. Observing the accuracy curve, we can find that Siamese_AUNet maintains a smooth curve in the later stage, which also proves the stability and reliability of the model compared with other models.

D. Ablation Experiments

In order to further evaluate the difference between the impact of FAM and the other two attention blocks on Siamese_AUNet, ablation experiments were carried out for the LEVIR and WHU datasets, respectively. The same accuracy metrics were calculated and the results are given in Tables VI and VII, and the accuracy curves are shown in Figs. 17 and 18. The STAZKI dataset was not selected here because this dataset is inappropriate to differentiate the model accuracy.



(a) (b)

(c) (d)

Fig. 18 Ablation experiment accuracy of WHU.

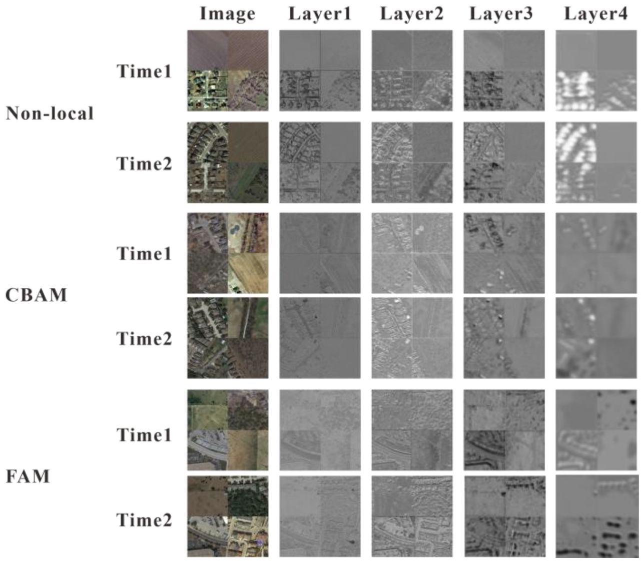


Fig. 19. Feature maps of LEVIR.

The accuracy metrics of the two datasets reflect the difference in detection accuracy of the model for using different Attention blocks. Overall, FAM has higher accuracy among the three Attention blocks. In addition, the accuracy metrics shown for the three models also confirm that the siamese structure combined with the Attention module is an effective approach.

In each layer of the model, the attention feature matrix is processed by a convolution block to obtain a grayscale attention image, and Figs. 19 and 20 show the recorded grayscale attention images. From these figures, it can be found that for the shallow features, such as Layer1 and Layer2, the main response is the texture information of the image as well as the detail information, which helps to portray the boundary of the change region. For the deeper features, such as Layer3 and Layer4, the main response is the spatial information of the key objects in the image, such as the location of the change region, while the specific detail information is lost more. This also explains why the detection results reflected by DASNet do not reconstruct the boundary well. After Attention block, the buildings in the image are well noticed, which is the effect we expect to see, and this also shows that Attention block has a certain role.

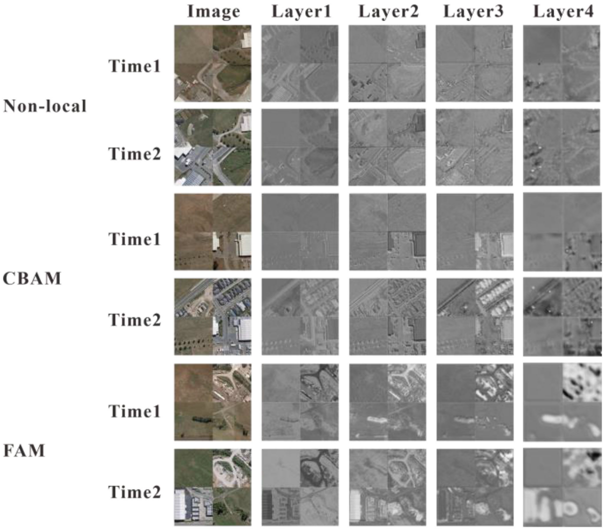


Fig. 20. Feature maps of WHU.

TABLE VIII
EXPERIMENTAL SUMMARY

	LEVIR	*WHU*	*SZTAKI*	*COUNT*
DASNet		√		1
UNet		√		1
AUNet	√		√	2
SiameseNet	√		√	2
Siamese_AUNet	√	√	√	3

E. Experimental Summary

We summarize the performance of each model in the three datasets. We considered a model successful when its F1 and Dice accuracy achieved the top three in the experiment, and both were better than 0.8. We counted the performance of each model separately and summarized its number of successes. The results are given in Table VIII.

It can be found that Siamese_AUNet shows stable metrics for high resolution images (LEVIR), VHR images (WHU), and medium resolution images (SZTAKI), indicating that our proposed Siamese_AUNet is more robust.

IV. DISCUSSION

RSICD is an important but arduous and challenging task. There has been an important trend to use siamese networks for this task, but there are still some issues that need to be addressed. In this article, a new RSICD framework based on siamese network has been proposed. Experiments are carried out on three commonly used datasets, and the results show that the proposed framework achieves a good performance. However, some key issues still need to be discussed in the further application of Siamese_AUNet.

1) The training cost is huge. In the study for high-resolution RSICD, not only the number of remote sensing data is large, but also the resolution is high, both of which bring exponential increase for the cost of training. In this article, the model is constructed and trained from the initial state, and although certain results are achieved, it consumes a lot of time and resources, which is more difficult to meet in practical applications. Therefore, exploring how to build siamese networks using pre-trained models is a very important direction.
2) Further research on AM is necessary. The FAM used in this article is improved from nonlocal AM and CBAM, which achieves better results and also brings a huge amount of computation compared to both UNet and SiameseNet. It is important to explore a lightweight and accurate AM for future research.
3) Manual data labeling often brings more errors, such as incorrect labeling and omission of labeling. Although data errors are inevitable, a more objective, scientific, and efficient annotation method to build the dataset is still a direct solution to improve the model training effect. Therefore, it is necessary to explore a semisupervised labeling scheme.
4) In the WHU experiment, our method did not achieve better results than DASNet. The details discussion we have given in the experimental section, too much focus on the spatial features is the main reason for this problem. In fact, the main role of spatial features in the study is to provide boundary texture information. We will continue to investigate how to maximize the role of spatial features while eliminating noise interference.

V. CONCLUSION

This article introduces a new siamese network architecture (which is different from the traditional convolutional and siamese networks) for RSICD problems. The proposed Siamese_AUNet incorporate both multibranch weight-share characteristics of siamese networks and have robust image segmentation ability as VGGNet. The ASPP module and the attention mechanism are combined, and experiments are conducted on three public datasets. The following conclusions are drawn based on the analysis of the experimental data.

1) The siamese structure enhances the detection ability of the network for weakly changing objects, and the FAM is adopted to suppress the appearance of interference noise for better results.
2) Applying ASPP to the SiameseNet network can significantly improve the detection ability of multiscale change units, making the description of the change boundary closer to the real situation. This proves that the ASPP module can effectively solve the multiscale feature detection problem in the CD problem.
3) Siamese_AUNet has achieved good results in the two CD challenges of suppressing noise and detecting weak changes. Experiments on three publicly available datasets also further confirmed the robustness of the newly proposed model.

REFERENCES

[1] D. Lu, P. Mausel, E. Brondizio, and E. Moran, “Change detection techniques,” *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, Jun. 2004.
[2] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, “Digital change detection methods in ecosystem monitoring: A review,” *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, May 2004.

[3] H. Li *et al.*, “Using land long-term data records to map land cover changes in China over 1981–2010,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1372–1389, Apr. 2017.

[4] M. E. Zelinski, J. Henderson, and M. Smith, “Use of Landsat 5 for change detection at 1998 Indian and Pakistani nuclear test sites,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3453–3460, Jan. 2014.

[5] J. Hu and Y. Zhang, “Seasonal change of land-use/land-cover (LULC) detection using MODIS data in rapid urbanization regions: A case study of the Pearl river delta region (China),” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 4, pp. 1913–1920, Jan. 2013.

[6] J. E. Vogelmann *et al.*, “Monitoring landscape change for LANDFIRE using multi-temporal satellite imagery and ancillary data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 2, pp. 252–264, Apr. 2010.

[7] Z. Li, W. Shi, S. W. Myint, P. Lu, and Q. Wang, “Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method,” *Remote Sens. Environ.*, vol. 175, pp. 215–230, Mar. 2016.

[8] G. Cao, Y. Li, Y. Liu, and Y. Shang, “Automatic change detection in high-resolution remote-sensing images by means of level set evolution and support vector machine classification,” *Int. J. Remote Sens.*, vol. 35, no. 16, pp. 6255–6270, Aug. 2014.

[9] X. J. Huang, Y. W. Xie, J. J. Wei, M. Fu, L. L. Lv, and L. L. Zhang, “Automatic recognition of desertification information based on the pattern of change detection-CART decision tree,” *J. Catastrophol.*, vol. 32, no. 1, pp. 36–42, Jan. 2017.

[10] D. Seo, Y. Kim, Y. Eo, M. Lee, and W. Park, “Fusion of SAR and multispectral images using random forest regression for change detection,” *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 10, Oct. 2018, Art. no. 401.

[11] H. Lyu and H. Lu, “Learning a transferable change detection method by recurrent neural network,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 5157–5160.

[12] B. C. Pijanowski, D. G. Brown, B. A. Shellito, and G. A. Manik, “Using neural networks and GIS to forecast land use changes: A land transformation model,” *Comput. Environ. Urban Syst.*, vol. 26, no. 6, pp. 553–575, May 2002.

[13] X. W. Chen, “Using remote sensing and GIS to analyse land cover change and its impacts on regional sustainable development,” *Int. J. Remote Sens.*, vol. 23, no. 1, pp. 107–124, Jan. 2002.

[14] L. P. Zhang and C. Wu, “Advance and future development of change detection for Multi-temporal remote sensing imagery,” (in Chinese), *Acta Geodaetica et Cartographica Sinica*, vol. 46, no. 10, pp. 1447–1459, Oct. 2017.

[15] A. Lefebvre, T. Corpetti, and L. Hubert-Moy, “Object-oriented approach and texture analysis for change detection in very high resolution images,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2008, pp. IV–663–IV–666.

[16] R. Gamanya, P. De Maeyer, and M. De Dapper, “Object-oriented change detection for the city of Harare, Zimbabwe,” *Expert Syst. Appl.*, vol. 36, no. 1, pp. 571–588, Jan. 2009.

[17] S. Bontemps, P. Bogaert, N. Titeux, and P. Defourny, “An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution,” *Remote Sens. Environ.*, vol. 112, no. 6, pp. 3181–3191, Jun. 2008.

[18] M. Hao, W. Shi, H. Zhang, and C. Li, “Unsupervised change detection with expectation-maximization-based level set,” *IEEE Geosci. Remote Sens.*, vol. 11, no. 1, pp. 210–214, Jan. 2014.

[19] L. Bruzzone and D. F. Prieto, “Automatic analysis of the difference image for unsupervised change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[20] L. Zhou, G. Cao, Y. Li, and Y. Shang, “Change detection based on conditional random field with region connection constraints in high-resolution remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3478–3488, Aug. 2016.

[21] D. Y. Zhu, T. Chen, Z. W. Wang, and R. Q. Niu, “Detecting ecological spatial-temporal changes by remote sensing ecological index with local adaptability,” *J. Environ. Manage.*, vol. 299, Dec. 2021, Art. no. 113655.

[22] D. Y. Zhu, T. Chen, N. Zhen, and R. Q. Niu, “Monitoring the effects of open-pit mining on the eco-environment using a moving window-based remote sensing ecological index,” *Environ. Sci. Pollut. Res.*, vol. 27, no. 13, pp. 15716–15728, Feb. 2020.

[23] X. X. Zheng and T. Chen, “High spatial resolution remote sensing image segmentation based on the multiclassification model and the binary classification model,” *Neural Comput. Appl.*, pp. 1–8 2021, doi: 10.1007/s00521-020-05561-8.

[24] X. Gao, T. Chen, R. Q. Niu, and A. Plaza, “Recognition and mapping of landslide using a fully convolutional DenseNet and influencing factors,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7881–7894, 2021, doi: 10.1109/JSTARS.2021.3101203.

[25] H. Cai, T. Chen, R. Q. Niu, and A. Plaza, “Landslide detection using densely connected convolutional networks and environmental conditions,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5235–5247, 2021, doi: 10.1109/JSTARS.2021.3079196.

[26] X. Wang, K. Tan, Q. Du, Y. Chen, and P. Du, “CVA2E: A conditional variational autoencoder with an adversarial training process for hyperspectral imagery classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5676–5692, Aug. 2020, doi: 10.1109/TGRS.2020.2968304.

[27] X. Wang, K. Tan, Q. Du, Y. Chen, and P. Du, “Caps-TripleGAN: GAN-Assisted CapsNet for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7232–7245, Sep. 2019, doi: 10.1109/TGRS.2019.2912468.

[28] N. Wambugu *et al.*, “Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, 2021, Art. no. 102603.

[29] A. M. El Amin, Q. Liu, and Y. Wang, “Convolutional neural network features based change detection in satellite images,” in *Proc. 1st Int. Workshop Pattern Recognit., Int. Soc. Opt. Photon.*, 2016, Art. no. 100110W.

[30] A. Varghese, J. Gubbi, A. Ramaswamy, and P. Balamuralidhar, “ChangeNet: A deep learning architecture for visual change detection,” in *Proc. Eur. Conf. Comput. Vis.Workshops*, 2018, vol. 11130, pp. 129–145.

[31] Y. Gao, F. Gao, J. Dong, and S. Wang, “Change detection from synthetic aperture radar images based on channel weighting-based deep cascade network,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4517–4529, Nov. 2019.

[32] A. Song, J. Choi, Y. Han, and Y. Kim, “Change detection in hyperspectral images using recurrent 3D fully convolutional networks,” *Remote Sens.*, vol. 10, no. 11, Nov. 2018, Art. no. 1827.

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

[34] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, “Signature verification using a ‘Siamese’ time delay neural network,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 7, no. 4, pp. 669–688, 1993.

[35] B. Du, L. Ru, C. Wu, and L. Zhang, “Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.

[36] E. Guo *et al.*, “Learning to measure change: Fully convolutional Siamese metric networks for scene change detection,” 2018, *arXiv:1810.09111*.

[37] J. Chen *et al.*, “DASNet: Dual attentive fully convolutional siamese networks for change detection of high resolution satellite images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1194–1206, 2021. doi: 10.1109/JSTARS.2020.3037893.

[38] K. Sakurada, M. Shibuya, and W. Wang, “Weakly supervised silhouette-based semantic scene change detection,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 6861–6867.

[39] H. Chen and Z. Shi, “A spatial-temporal attention-based method and a new dataset for remote sensing image change detection,” *Remote Sens.*, vol. 12, no. 10, May 2020, Art. no. 1662.

[40] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[41] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015, *arXiv:1409.1556*.

[42] S. Woo *et al.*, “CBAM: Convolutional block attention module,” in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[43] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Med. image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[44] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected CRFS,” in *Proc. Int. Conf. Learn. Representation*, 2015.

[45] S. P. Ji, S. Q. Wei, and M. Lu, “Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019.

[46] 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.

Tao Chen (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan China, in 2008. He is currently an Associate Professor with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China.

From 2015 to 2016, he was a Visiting Scholar with the University of New South Wales, Sydney, NSW, Australia. He has published more than 50 scientific papers including IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), *Journal of Environmental Management*, *Remote Sensing*, *Environmental Earth Sciences*, and *Environmental Science and Pollution Research* and guest edited two journal special issues. His research interests include image processing, machine learning, and geological remote sensing.

Dr. Chen regularly is a Program Committee (PC) member of the IEEE International Conference on Data Mining, the International Joint Conferences on Artificial Intelligence, and the Scientific Committee of the Conference of the Arabian Journal of Geosciences. He is also a Reviewer for more than 20 Science Citation Index (SCI) journals, including *Remote Sensing of Environment*, TGRS, JSTARS, GRSL, IEEE ACCESS, *Computers and Geosciences*, *Neural Computing and Application*, *Neurocomputing*, *Ecological Indicators*, *Remote Sensing*, *Environmental Earth Sciences*, *Environmental Science and Pollution Research*, and *Land Degradation and Development*.

Zhiyuan Lu received the M.S. degree in geological engineering from China University of Geosciences, Wuhan, China, in 2021.

He is currently a Research Intern with China Academy of Railway Sciences, Beijing, China. His research interests include remote sensing image change detection, GNSS measurements, InSAR, and remote sensing geological analysis.

Yue Yang received the B.S. degree in geoinformatics in 2020 from the China University of Geoscience, Wuhan, China, where she is currently working toward the M.S. degree in earth exploration and information technology.

Her current research interests include remote sensing change detection and remote sensing application.

Yuxiang Zhang (Member, IEEE) received the B.S. degree in the sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2011, and the Ph.D. degree in the photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2016.

She was a Visiting Scholar with the University of Sydney, Sydney, NSW, Australia, from 2019 to 2020. She is currently an Associate Professor with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China. Her current research interests include hyperspectral image processing, pattern recognition, and machine learning.

Bo Du (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2010.

He is currently a Professor with the School of Computer Science, Wuhan University, Wuhan, China, and the Institute of Artificial Intelligence, Wuhan University. He is also the Director of the National Engineering Research Center for Multimedia and Software, Wuhan University. He has published more than 80 research articles in IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CYBERNETICS (TCYB), IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL). A total of 13 of them are Essential Science Indicators hot articles or highly cited articles. His research interests include pattern recognition, hyperspectral image processing, and signal processing.

Dr. Du is currently a Senior Program Committee Member of the International Joint Conferences on Artificial Intelligence and the Association for the Advancement of Artificial Intelligence. He is also a Reviewer for 20 Science Citation Index magazines, including IEEE TPAMI, TCYB, TGRS, TIP, JSTARS, and GRSL. He was the recipient of the Highly Cited Researcher by the Web of Science Group in 2019 and 2020, received the IEEE Geoscience and Remote Sensing Society 2020 Transactions Prize Paper Award, received the IJCAI Distinguished Paper Prize, was the IEEE Data Fusion Contest Champion, and received the IEEE Workshop on Hyperspectral Image and Signal Processing Best Paper Award in 2018. He was an Area Chair for the International Conference on Pattern Recognition. He is currently an Associate Editor for *Neural Networks*, *Pattern Recognition*, and *Neurocomputing*.

Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the Department of Technology of Computers and Communications, University of Extremadura, Cáceres, Spain, in 1999 and 2002, respectively.

He is the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. He has authored more than 600 publications, including more than 200 Journal Citation Reports journal articles (more than 160 in IEEE journals), 23 book chapters, and around 300 peer-reviewed conference proceeding papers. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza was a Member of the Editorial Board of IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER, from 2011 to 2012, and IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE in 2013. He was also a member of the Steering Committee of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He was contributed to hyperspectral data processing and parallel computing of earth observation data. He was a recipient of the Recognition of Best Reviewers of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2009 and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2010, for which he was an Associate Editor, from 2007 to 2012. He was also a recipient of the Most Highly Cited Paper Award from the Journal of Parallel and Distributed Computing, from 2005 to 2010, the 2013 Best Paper Award of the JSTARS journal, the Best Column Award of the IEEE Signal Processing Magazine in 2015, Paper Awards at the IEEE International Conference on Space Technology, and the IEEE Symposium on Signal Processing and Information Technology. He has guest edited ten special issues on hyperspectral remote sensing for different journals. He was the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS), from 2011 to 2012 and the President of the Spanish Chapter of the IEEE GRSS, from 2012 to 2016. He has reviewed more than 500 manuscripts for more than 50 different journals. He was the Editor-in-Chief for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, from 2013 to 2017. He is also an Associate Editor for the the IEEE ACCESS (receiving the recognition as an outstanding Associate Editor of the journal in 2017).