applied sciences

MDPI

Article

Crop Land Change Detection with MC&N-PSPNet

Yuxin Chen [1,2], Yulin Duan [2,3], Wen Zhang [1,*], Chang Wang [1], Qiangyi Yu [2,3] and Xu Wang [4]

[1] School of Civil Engineering, University of Science and Technology Liaoning, Anshan 114000, China; cyx0402@ustl.edu.cn (Y.C.); wangchang324@163.com (C.W.)
[2] State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China; duanyulin@caas.cn (Y.D.); yuqiangyi@caas.cn (Q.Y.)
[3] Key Laboratory of Agricultural Remote Sensing, Ministry of Agriculture and Rural Affairs/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[4] School of Resources and Civil Engineering, Liaoning Institute of Science and Technology, Benxi 117000, China; wangxu1983411@126.com
* Correspondence: zhangwen_ln@163.com

Abstract: To enhance the accuracy of agricultural area classification and enable remote sensing monitoring of agricultural regions, this paper investigates classification models and their application in change detection within rural areas, proposing the MC&N-PSPNet (CBAM into MobileNetV2 and NAM into PSPNet) network model. Initially, the HRSCD (High Resolution Semantic Change Detection) dataset labels undergo binary redrawing. Subsequently, to efficiently extract image features, the original PSPNet (Pyramid Scene Parsing Network) backbone network, ResNet50 (Residual Network-50), is substituted with the MobileNetV2 (Inverted Residuals and Linear Bottlenecks) model. Furthermore, to enhance the model's training efficiency and classification accuracy, the NAM (Normalization-Based Attention Module) attention mechanism is introduced into the improved PSPNet model to obtain the categories of land cover changes in remote sensing images before and after the designated periods. Finally, the final change detection results are obtained by performing a different operation on the classification results for different periods. Through experimental analysis, this paper demonstrates the proposed method's superior capability in segmenting agricultural areas, which is crucial for effective agricultural area change detection. The model achieves commendable performance metrics, including overall accuracy, Kappa value, MIoU, and MPA values of 95.03%, 88.15%, 93.55%, and 88.90%, respectively, surpassing other models. Moreover, the model exhibits robust performance in final change detection, achieving an overall accuracy and Kappa value of 93.24% and 92.29%, respectively. The results of this study show that the MC&N-PSPNet model has significant advantages in the detection of changes in agricultural zones, which provides a scientific basis and technical support for agricultural resource management and policy formulation.

Keywords: deep learning network model; self-attention; semantic segmentation; change detection

Citation: Chen, Y.; Duan, Y.; Zhang, W.; Wang, C.; Yu, Q.; Wang, X. Crop Land Change Detection with MC&N-PSPNet. *Appl. Sci.* **2024**, *14*, 5429. https://doi.org/10.3390/app14135429

Academic Editor: Nathan J. Moore

Received: 2 April 2024
Revised: 18 June 2024
Accepted: 20 June 2024
Published: 22 June 2024

1. Introduction

Cultivated land is the material basis for human survival and development and is related to national economic and social development. However, the phenomenon of "non-agriculturalization" in agricultural areas is becoming increasingly severe. This is characterized by excessive deforestation, slash-and-burn cultivation, conversion of farmland to other uses, and escalating conflicts between humans and land. Not only is the size of agricultural areas decreasing, but the cultivation capacity of arable land is decreasing, which in turn affects food production and even destroys the global ecosystem, affecting socio-economic as well as ecologically sustainable development [1,2]. Therefore, the protection of cultivated land is of utmost urgency, and precise detection of changes in agricultural areas is becoming increasingly vital [3].

There are two main categories of methods for recognizing the type of change in features [4]: first classification and then change, as well as direct change detection. The former is to classify different simultaneous images separately, and then compare the classification knots of the before and after images pixel by pixel. In the latter case, change detection is performed directly by superimposing images of different time phases, simultaneously detecting and classifying the features of the changed part and the features of the unchanged part, so as to minimize the impact of classification errors on the change detection results. In post-classification change detection, traditional remote sensing image segmentation methods include those based on spectral, spatial, and a combination of spectral and spatial image segmentation [5–7]. Spectral-based image segmentation, or pixel-based image segmentation, primarily analyzes individual pixels. Spatial-based image segmentation attempts to incorporate spatial location information by detecting edges between regions for image segmentation. However, traditional methods often suffer from low robustness, over-segmentation, and high computational demands [8].

In recent years, with the rapid development of deep learning and neural networks, many scholars at home and abroad have begun to use deep learning for semantic segmentation based on multitemporal images to detect the change region and achieve good research results. Cao et al. [9] proposed the Res-UNet network by replacing the basic units of the ResNet [10] network with the convolutional layers of the U-Net [11] network, combining the underlying features obtained from down-sampling with the up-sampled inputs through jumping connection. The improved U-Net network achieved high classification accuracy in high-resolution remote sensing images. Wang et al. [12] proposed a DeepLabV3+ network that incorporates a class feature attention mechanism, which showed better performance in segmenting boundary areas of different features in high-resolution remote sensing images in experiments on a public dataset (GID), but the model was unable to accurately segment in some small scenes. Yuan et al. [13] proposed Pyramid-SCDFormer, a twin network model based on Transformer, which solves the problem of accurate recognition of small-scale objects and changing boundaries by using the self-attention mechanism to capture multi-scale features. Literature combined Siamese networks and U-net++ networks to design the SNUNet-CD network, which achieved independent feature extraction from different images and combined multi-level semantic information [14]. Zhan et al. [15] proposed a twin neural network change detection model SSCNN-S that combines spectral and spatial information, which effectively retains spatial information and improves change detection accuracy without loss of efficiency. Based on the Non-Local [16] mechanism, a spatiotemporal attention module was developed to combine multi-scale feature information with a pyramid pooling module [17]. Yuan Peng et al. [18] took Changzhou City as the study area and improved the U-Net network by using the residual structure and attention mechanism, proposing the RMAU-Net network model to realize the fine extraction of cropland. Although these studies have improved the accuracy of remote sensing image segmentation to some extent, there are still defects in the continuity and omission of segmentation at the edges of land cover and small targets.

Agricultural area classification and remote sensing monitoring play a crucial role in modern agricultural management. Accurate agricultural area classification can help policymakers to formulate more precise agricultural policies, improve the efficiency of agricultural production, reduce the waste of resources, and promote sustainable agricultural development. However, existing classification models often show certain limitations when facing complex agricultural environments and changes, such as insufficient classification accuracy, discontinuous smoothing of feature boundaries, and inaccurate change detection. The general aim is to better maintain high-dimensional features, ensure that the feature edge segmentation is continuous and not missed, ensure the light weight of the network, and improve the efficiency. Thus, in this paper, firstly, the backbone network ResNet50 is replaced by the more lightweight MobileNetV2 [19] model. Secondly, the CBAM (Convolutional Block Attention Module) module is added to the MobileNetV2 backbone feature extraction network to enhance the parsing ability of the agricultural area

as a way to improve the shallow feature classification accuracy and, lastly, the attention mechanism NAM is added after feature fusion to capture richer and more accurate features at different layers to capture richer and more differentiated feature information, which improves the accuracy and robustness of the model without increasing the complexity of the model. Based on extracting the results of the two periods of agricultural areas, the change detection results are obtained by the difference method. Based on the improvement of the method in this paper, excellent performance indexes are achieved in the experiments, which are of great significance for realizing efficient agricultural monitoring.

2. Materials and Methods

The network architecture in this paper primarily consists of two main components. Firstly, the replacement of the backbone feature extraction network. The lightweight MobileNetV2 backbone feature extraction network is utilized instead of the ResNet50 backbone feature extraction network to reduce model parameter count and enhance network training efficiency and accuracy. Secondly, after feature fusion, a lightweight and efficient attention mechanism (Normalization-Based Attention Module, NAM) is introduced to improve model accuracy and robustness without increasing model complexity. The workflow of the proposed agricultural area change detection method is illustrated in Figure 1.

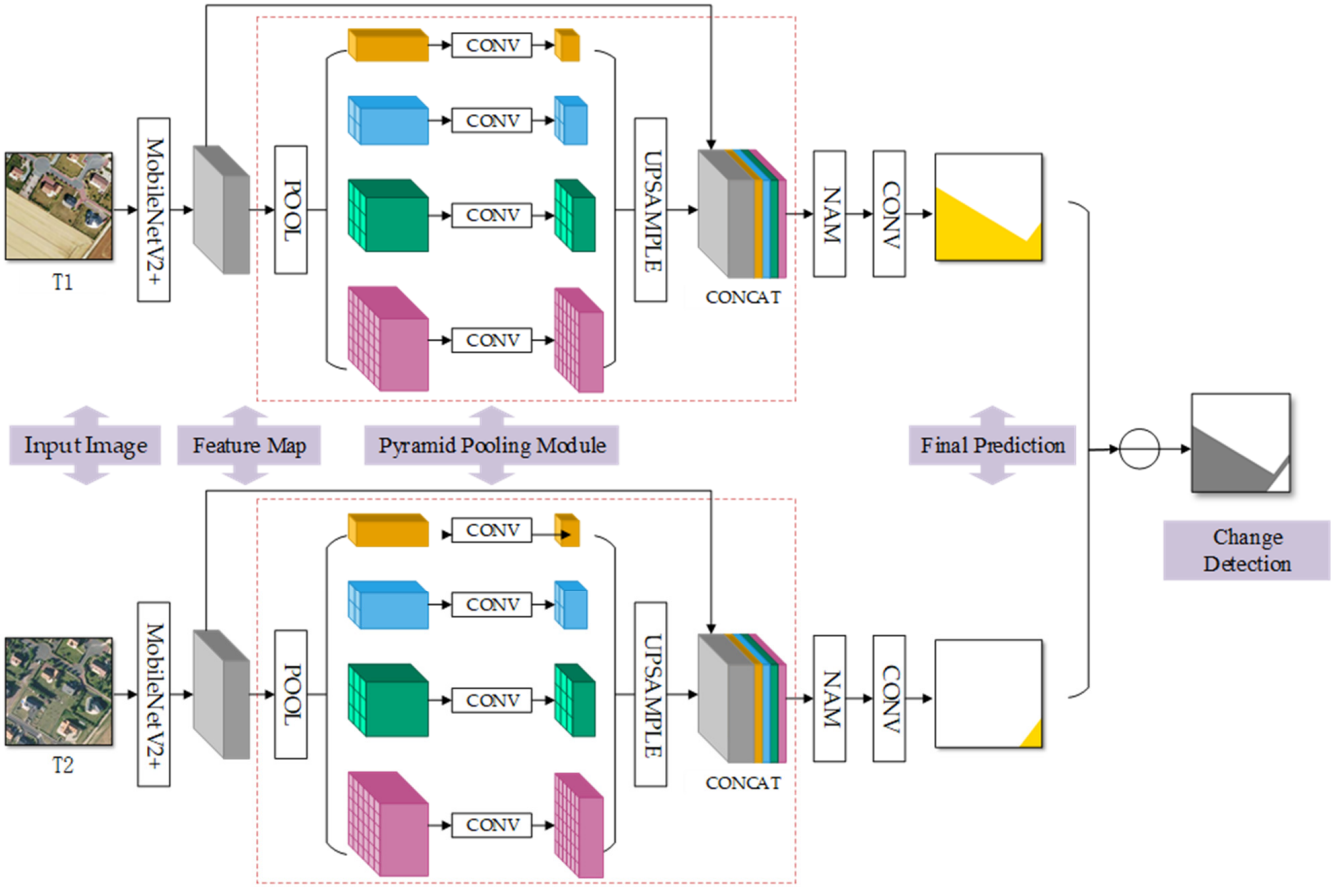


Figure 1. Flowchart of the methodology for detecting changes in agricultural areas.

2.1. Backbone Feature Extraction Network

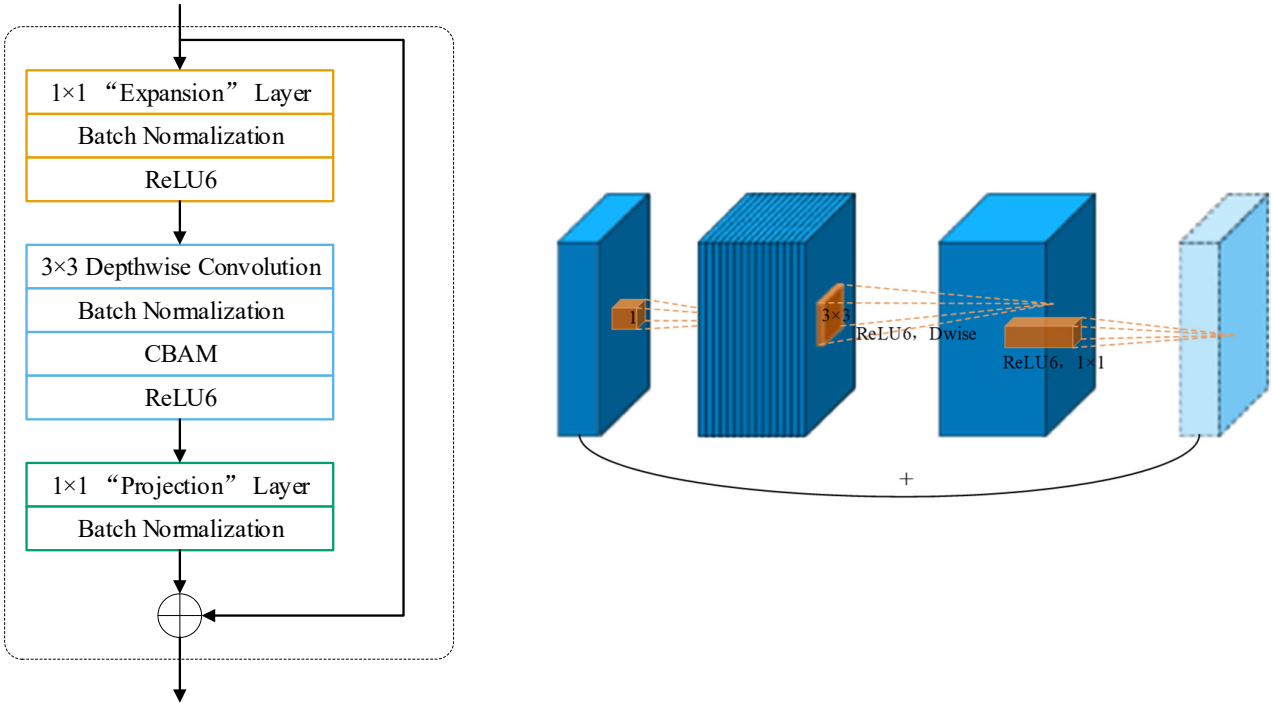
The original PSPNet [20] network utilizes ResNet50 as the backbone feature extraction network. ResNet50 is a deep residual network capable of learning more complex feature representations, thereby enhancing performance. Additionally, the deep design of ResNet50 may require more time and computational resources during training. However, in the study area the intra-class differences are relatively small, and this paper focuses on binary

classification. Hence, the advantages of deep network structures cannot be fully utilized. Moreover, deep network structures suffer from issues such as a large number of parameters and susceptibility to overfitting.

In this paper, an improved MobileNetV2 is used to replace the original network as the backbone network, which can improve the detection accuracy and stability while maintaining high efficiency without significantly expanding the model's complexity and computational volume. Firstly, MobileNetV2 adopts inverted residuals to produce the 3 × 3 depthwise separable convolution after 1 × 1 point-by-point convolution upscaling (Depthwise Convolution, DW). Then, the CBAM [21] module is embedded before the activation layer in the 3 × 3 network structure. The CBAM module can significantly improve the model's ability to detect the agricultural areas in remote sensing images and increase the weight of the main features to enhance the segmentation performance, especially in the case where the agrarian areas are similar to the non-agricultural areas. Finally, MobileNetV2 undergoes 1 × 1 point-by-point convolution to reduce dimension. Overall, this paper uses a single convolution kernel to apply a filter per each input channel and utilize the linear combination between channel features, which can obtain more information containing new features and enhance the efficiency and accuracy compared with the standard convolution. This expansion–contraction linear bottleneck structure of MobileNetV2 is capable of solving the problem of the small convolution kernel processed by DW as well as the gradient vanishing due to the ReLU6 activation function (expression as in Equation (1)) that makes the neuron output zero. The structure is shown in Figure 2.

$$\text{ReLU6}(x) = \min(\max(0, x), 6), \quad (1)$$

Therefore, to maintain good performance while reducing computational costs, this study opts to use MobileNetV2 instead of ResNet50 as the backbone feature extraction network.



(**a**) Inverse Residual Module (**b**) Network Block Diagram

Figure 2. MobileNetV2 backbone network structure: (**a**) inverse residual module, (**b**) network block diagram.

2.2. Attention Mechanism

To improve the training accuracy and speed of the model, this study introduces attention mechanisms into the PSPNet model. The essence of attention mechanisms lies in applying human perception and attention to machines, enabling them to discern important from unimportant parts of the data. In this paper, attention modules, namely CBAM and NAM, are respectively incorporated into the backbone feature extraction network and after feature fusion.

2.2.1. CBAM

Attention refers to important spatial and channel information within feature channels. It is commonly assumed that the feature channels obtained through convolutional network pooling possess equal importance. However, in reality, the importance of features in each channel varies. Compared to using channel attention mechanisms or spatial attention mechanisms separately, the CBAM (Convolutional Block Attention Module) module (illustrated in Figure 3) integrates both channel and spatial attention mapping processes. This integration allows for the preservation of more useful feature information.

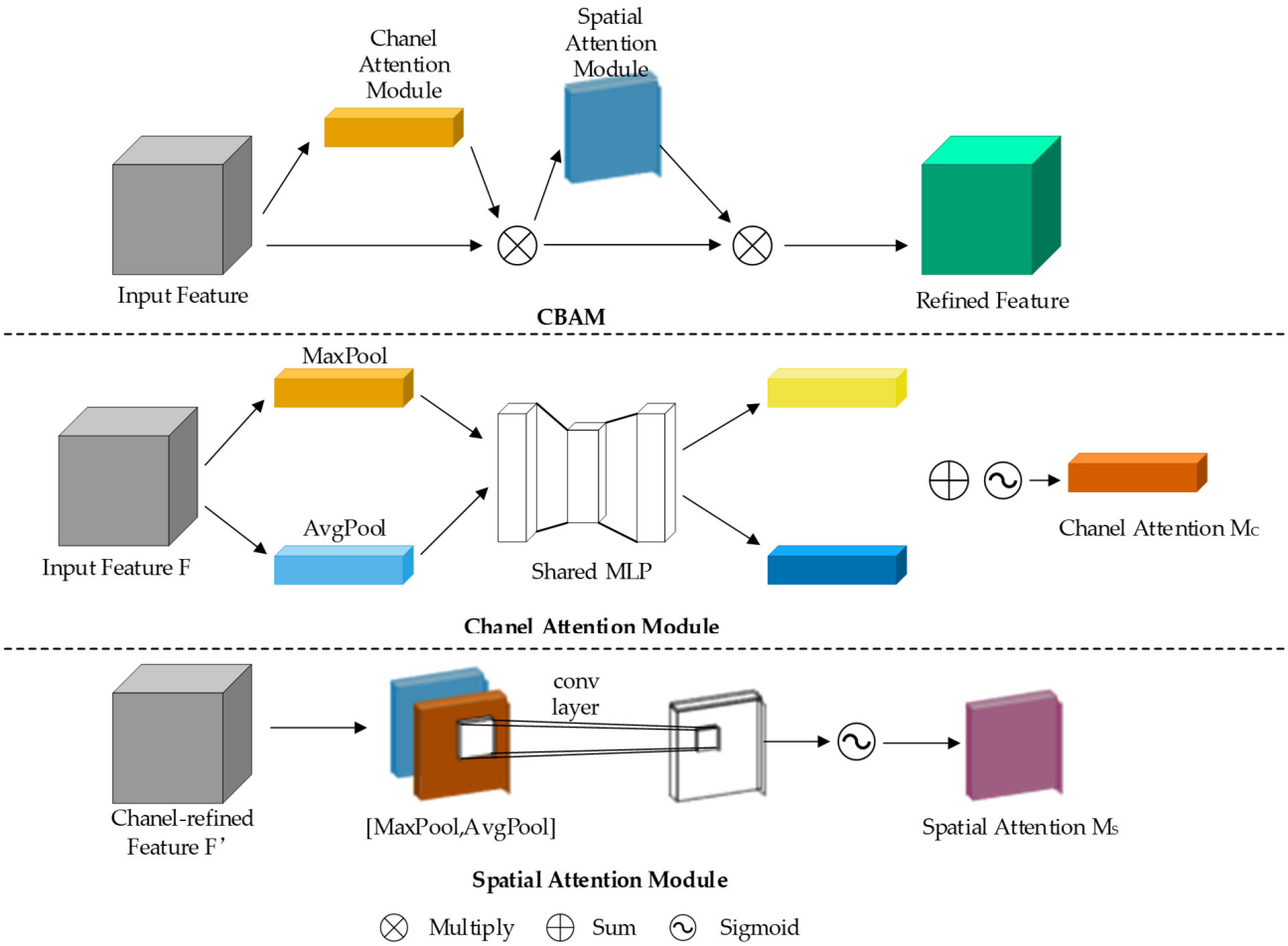


Figure 3. CBAM structure.

The network embedded in the CBAM mechanism first uses channel attention mapping to globally pool and maximally pool the feature map F generated by convolution, and then inputs the pooling results into the multilayer perceptron to perform the summation operation, generates the channel weight coefficients through the Sigmoid activation function, and then multiplies the weight coefficients by the original feature map F to get the feature

map adjusted by the channel weights. The process of the channel attention mapping is shown in Equation (2).

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (2)$$

where MLP denotes multilayer perceptron, which is the activation function. Subsequently, spatial attention mapping is performed on the weighted feature maps by serial concatenation of global maximum pooling and average pooling, using convolution to downsize into single-channel feature maps, using the Sigmoid function to activate and to generate the spatial feature matrices, and then the weight matrices and the feature maps undergo the dot-multiplication operation to get the final required spatial feature maps. The process of spatial attention mapping is shown in Equation (3).

$$M_s(F) = \sigma\left\{f^{7\times7}[AvgPool(F); MaxPool(F)]\right\} \quad (3)$$

where f denotes a convolutional layer with a 7×7 convolutional kernel which is the Sigmoid function and serial concatenation.

Finally, the feature map F of the output of the previous convolutional layer is summed with the feature map via the CBAM mechanism to obtain the input of the next convolutional layer.

2.2.2. NAM

The NAM [22] (Normalization-Based Attention Module) embedded in feature fusion in this paper is a normalization-based attention mechanism designed to reduce the weights of less significant features. This method applies sparse weight penalties to attention modules, making these weights computationally more efficient while maintaining the same level of performance. It helps the model to capture richer and more distinctive feature information at different levels, facilitating the detection of agricultural areas affected by factors such as planting patterns and crop types. Adding contribution factors for weights to the attention mechanism further suppresses insignificant features. The NAM used in this study incorporates the scaling factor of Batch Normalization to represent the importance of weights, avoiding the need for additional fully connected layers and convolutional layers as seen in SE [23], BAM [24], and CBAM.

The integration method of the NAM adopts the CBAM module, redesigning the channel attention (see Figure 4a) and spatial attention sub-modules (see Figure 4b). Thus, the NAM can be embedded at the end of each network block. For the channel attention sub-module, the scaling factor from BN (Batch Normalization) is used (i.e., the variance in BN, as shown in Equation (4)), reflecting the magnitude of changes in each channel and the importance of that channel. Equation (5) represents the final output feature obtained, where γ is the scaling factor for each channel, enabling the derivation of weights for each channel. If the same normalization method is applied to each pixel in space, spatial attention weights can be obtained, as shown in Equation (6), referred to as pixel normalization. To suppress unimportant features, a regularization term is added to the loss function, as shown in Equation (7).

$$B_{out} = BN(B_{in}) = \gamma \frac{B_{in} - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}} + \beta, \quad (4)$$

$$M_c = \text{sigmoid}(W_\gamma(BN(F_1))), \quad (5)$$

$$M_s = \text{sigmoid}(W_\lambda(BN_s(F_2))), \quad (6)$$

$$Loss = \sum_{(x,y)} l(f(x, W), y) + p\sum g(\gamma) + p\sum g(\lambda), \quad (7)$$

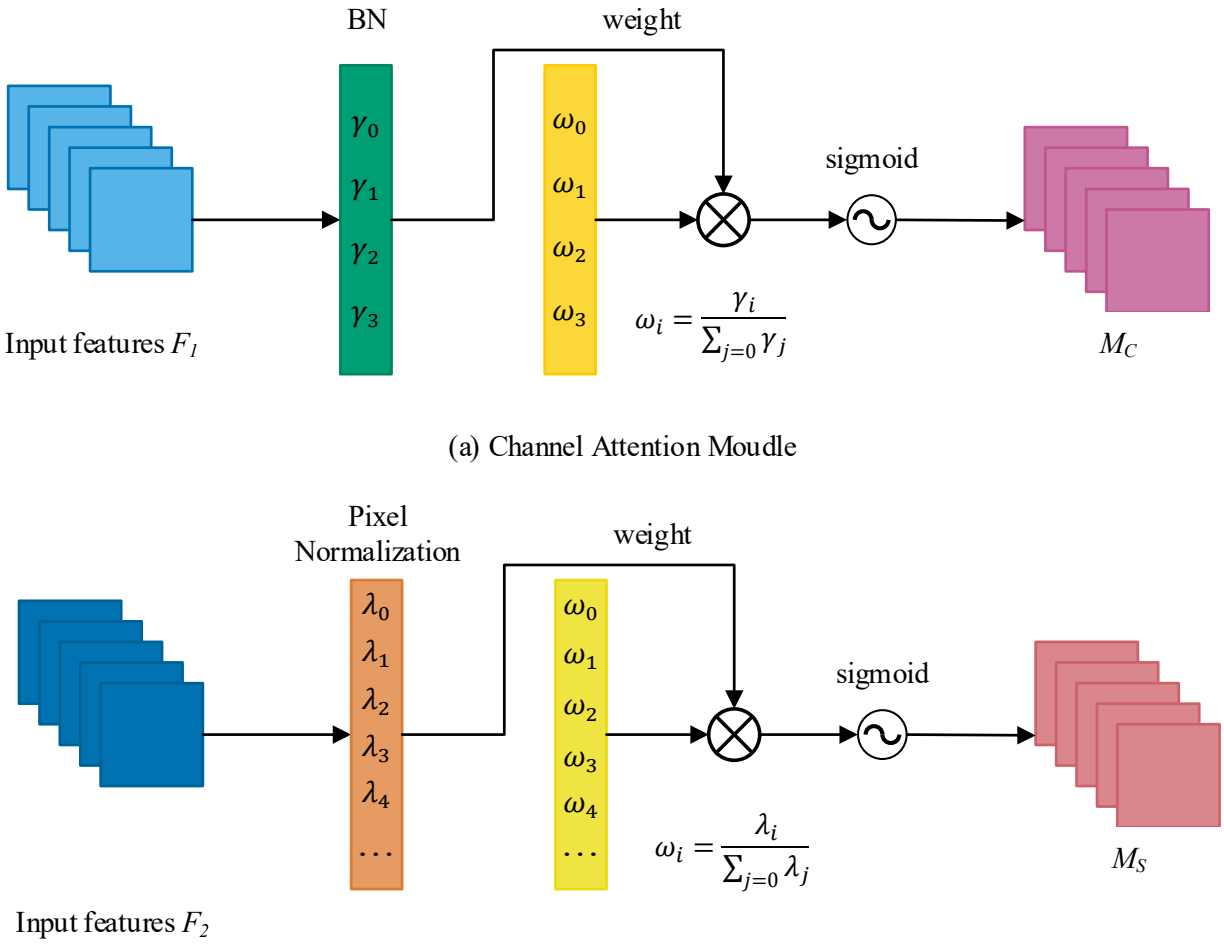


Figure 4. NAM structure: (**a**) Channel Attention Module, (**b**) Spatial Attention Module.

2.3. Data

To validate the effectiveness of the proposed method, this paper selected the High-Resolution Semantic Change Detection (HRSCD) dataset captured by the French National Institute of Geographic and Forest Information (IGN). The HRSCD dataset, sourced from an open access publication on semantic change detection, is composed of aerial images stitched together. It is suitable for multi-label tasks such as semantic segmentation. The dataset comprises 291 pairs of RGB images stored in TIFF format, each with dimensions of 10,000 × 10,000 pixels. Each image pair consists of an early image captured in either 2005 or 2006 and a second image captured in 2012. The label data is stored in a single-channel TIFF format with a resolution of 0.5 m per pixel. The imaging bands include the R, G, and B bands, covering six land cover classes: bare land, urban areas, agricultural areas, forests, wetlands, and water bodies.

To facilitate targeted detection of changes in agricultural zones, a binary classification was used in this study. Therefore, manual visual interpretation was performed using the ArcGIS 10.8 platform to categorize the images into agricultural and non-agricultural zones. Subsequently, the change detection labels were redrawn using semantic segmentation tags. Due to the large image data, the original images and labels were resized to 256 × 256 image segments before model training. In total, 3,042 image segments were obtained. A subset of the dataset is shown in Figure 5.

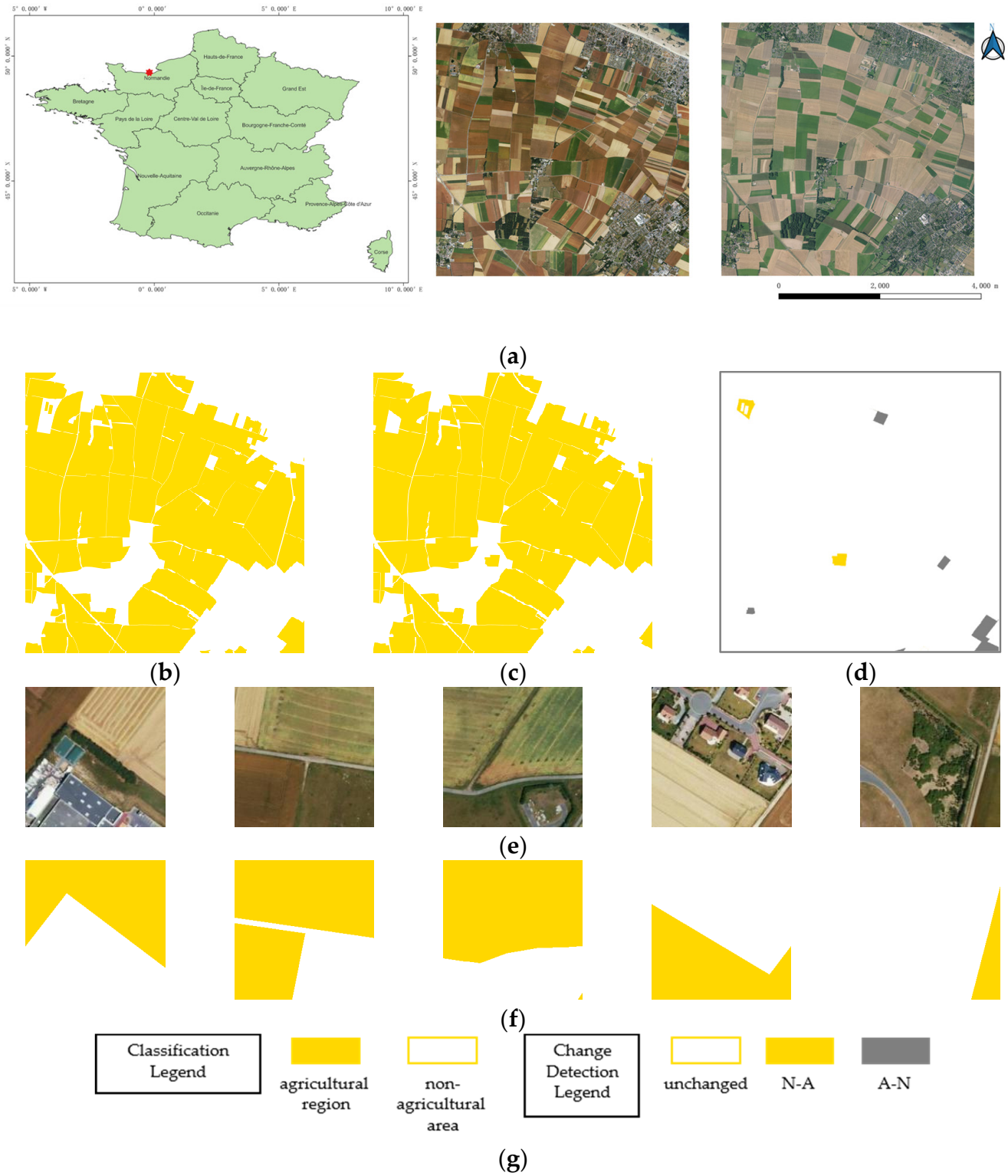


Figure 5. Original images and labels: (**a**) geographic location of the study area and original images of the two phases; (**b**) 2006 image labeling; (**c**) 2012 image labeling; (**d**) CD labels; (**e**) example of training images; (**f**) example of training labels; (**g**) legends.

2.4. Evaluation

To validate the accuracy of different improvement schemes, this paper uses the confusion matrix to categorize the validation dataset and six objective evaluation metrics, namely Mean Integration over Union (MIoU), Mean Pixel Accuracy (MPA), Overall Accuracy (OA),

Producers Accuracy (PA), Users Accuracy (UA), and Kappa Coefficient. The formulas for calculating the six objective evaluation indexes are shown in Equations (8)–(13).

$$\mathrm{MIoU} = \frac{1}{\mathrm{k}+1}\sum_{\mathrm{i}=0}^{\mathrm{k}} \frac{\mathrm{P_{ii}}}{\sum_{\mathrm{j}=0}^{\mathrm{k}} \mathrm{P_{ji}} + \sum_{\mathrm{j}=0}^{\mathrm{k}} \mathrm{P_{ij}} - \mathrm{P_{ii}}}, \tag{8}$$

$$\mathrm{MPA} = \frac{1}{\mathrm{k}+1}\left(\sum_{\mathrm{i}=0}^{\mathrm{k}} \mathrm{P_{ii}} / \sum_{\mathrm{i}=0}^{\mathrm{k}} \sum_{\mathrm{i}=0}^{\mathrm{k}} \mathrm{P_{ij}}\right), \tag{9}$$

$$\mathrm{OA} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_{ii}}}{\mathrm{N}}, \tag{10}$$

$$\mathrm{PA} = \frac{\mathrm{x_{ii}}}{\mathrm{a_i}}, \tag{11}$$

$$\mathrm{UA} = \frac{\mathrm{x_{ii}}}{\mathrm{b_i}}, \tag{12}$$

$$\mathrm{Kappa} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{a_i b_i}}{\mathrm{N}^2}, \tag{13}$$

where x_{ii} represents the number of samples in the i category that were correctly categorized. The $N = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}$ in Equations (10) and (13) represents the number of samples and n represents the number of categories. The $a_i = \sum_{j=1}^{n} x_{ji}$ in Equation (11) represents the number of samples of the i category in the real category. Equation (12) $b_i = \sum_{j=1}^{n} x_{ij}$ represents the number of samples of the i category in the prediction result.

2.5. Training and Validation

In this paper, PyTorch is used as the development framework, and the hardware environment is Intel(R) Core(TM) i9-12900HCPU and NVIDIA GeForce RTX 3060GPU, with 16GB of RAM and 8GB of video memory, and the software environment is Window11, python3.9, CUDA12.1.68, and pytorch1.13.1. Based on the experimental results, the optimal parameters are selected, and the data are randomly divided according to the ratio of 9:1; there are a total of 2736 images in the training dataset and 306 images in the validation dataset. The accuracy and efficiency of the method in this paper are affected by factors such as the down-sampling multiplier, the structure of the backbone network, the type of optimizer, and the training ratio of the dataset. The basis for the determination of the four parameters is mainly evaluated by the effect of the network performance on the dataset (overall accuracy, Kappa, training time, Loss, Val_Loss). In this paper, the backbone networks are selected as MobileNetV2 and ResNet50 for experimental comparison. The size of the down-sampling multiplier is chosen between 8 and 16. The optimizer type is chosen between SGD (Stochastic Gradient Descent), and Adam (Adaptive Gradient Algorithm); the proportion of the training part of the dataset decreases step by step, and the experimental results are shown in Figure 6.

From Figure 6a, it can be seen that the best training effect is achieved when the backbone network is MobileNetV2, the down-sampling multiplier is 8, the optimizer is Adam, and the ratio of the training set to the validation set is 1:9. From Figure 6b,c when epoch = 120, both Loss and Val_Loss are stabilized. Batch size is set to the maximum within the acceptance range of the graphics card. Therefore, in this paper, the parameters are set as follows: the backbone network structure is MobileNetV2, the down-sampling multiplier is 8, the optimizer is Adam, the batch size is 8, the number of training rounds epoch is 120, the gradient descent function is cos, and the loss function is chosen as Cross Entropy Loss (CEL). The computational formulas are as follows:

$$Loss = -\frac{1}{N}\sum_{i} \sum_{c=1}^{M} y_{ic} \log(p_{ic}), \tag{14}$$

In Equation (14), the M represents the number of total categories of classification and takes 1 when the y_{ic} predicted feature type and the labeled feature type are the same and

0 otherwise; p_{ic} represents the probability that a feature is i predicted to be c a feature; N represents the number of samples.

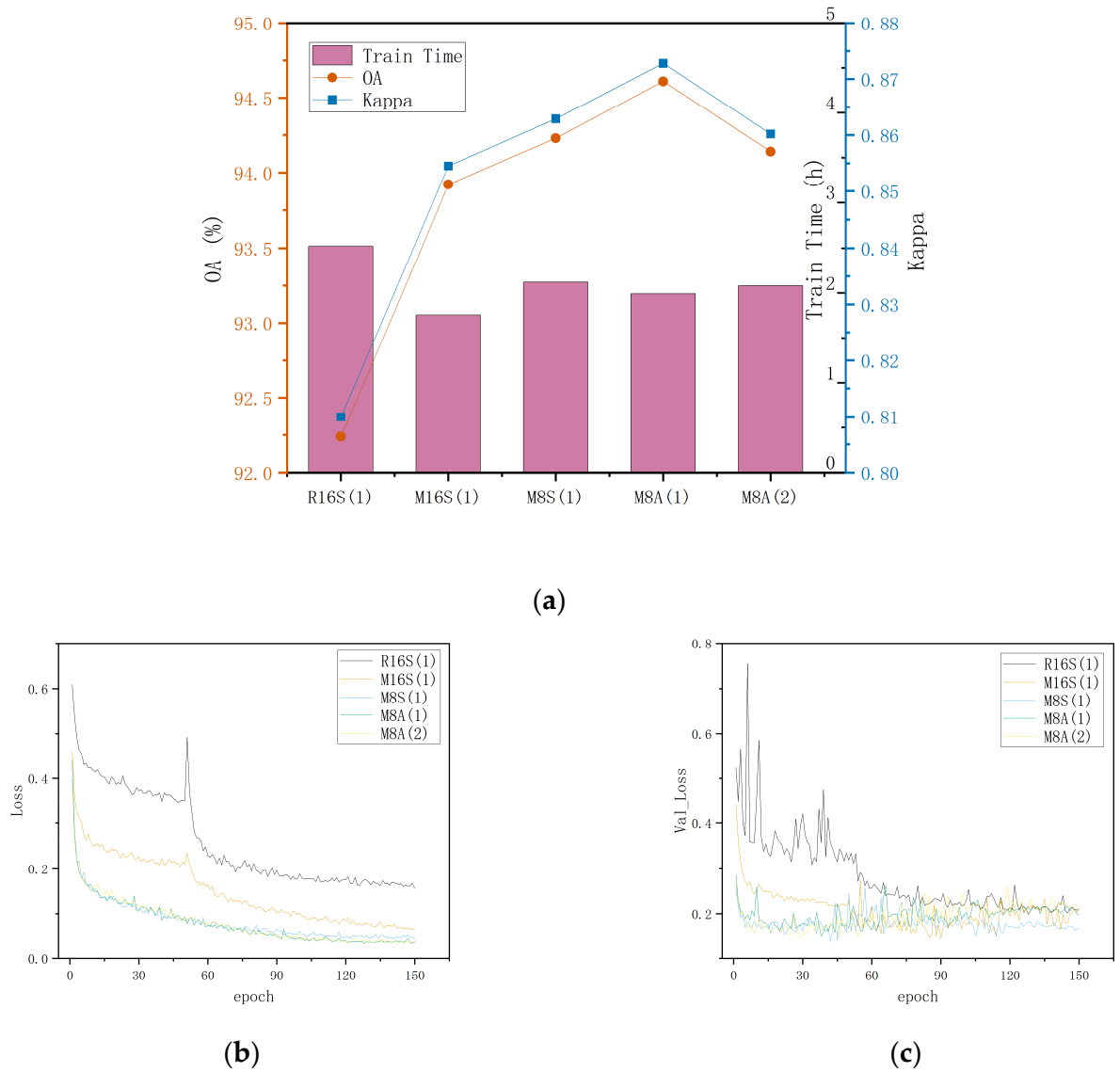


Figure 6. Impact of backbone network structure, down-sampling multiplier, optimizer type, and dataset ratio on information detection accuracy and efficiency: (**a**) effect of different parameter settings on OA, Kappa, and training time; (**b**) different parameter settings' effect on Loss; (**c**) different parameter settings' effect on Val_Loss. (Note: in the figure, M stands for MobileNetV2 as the backbone feature extraction network, R stands for ResNet50 as the backbone feature extraction network, 8 and 16 stand for different down-sampling multiplicities, S stands for the selection of SGD optimizer, A stands for the selection of Adam as the optimizer, and (1) (2) stands for the ratios of the training set to the validation set of 1:9 and 2:8, respectively).

3. Results

3.1. Experimental Result

The experimental results based on the proposed method are shown in Figure 7 below. The overall edge extraction effect has better performance with a small amount of mis-extraction, the overall accuracy and Kappa value reached 95.03 and 0.8815, respectively, and the training time was 2.109 h.

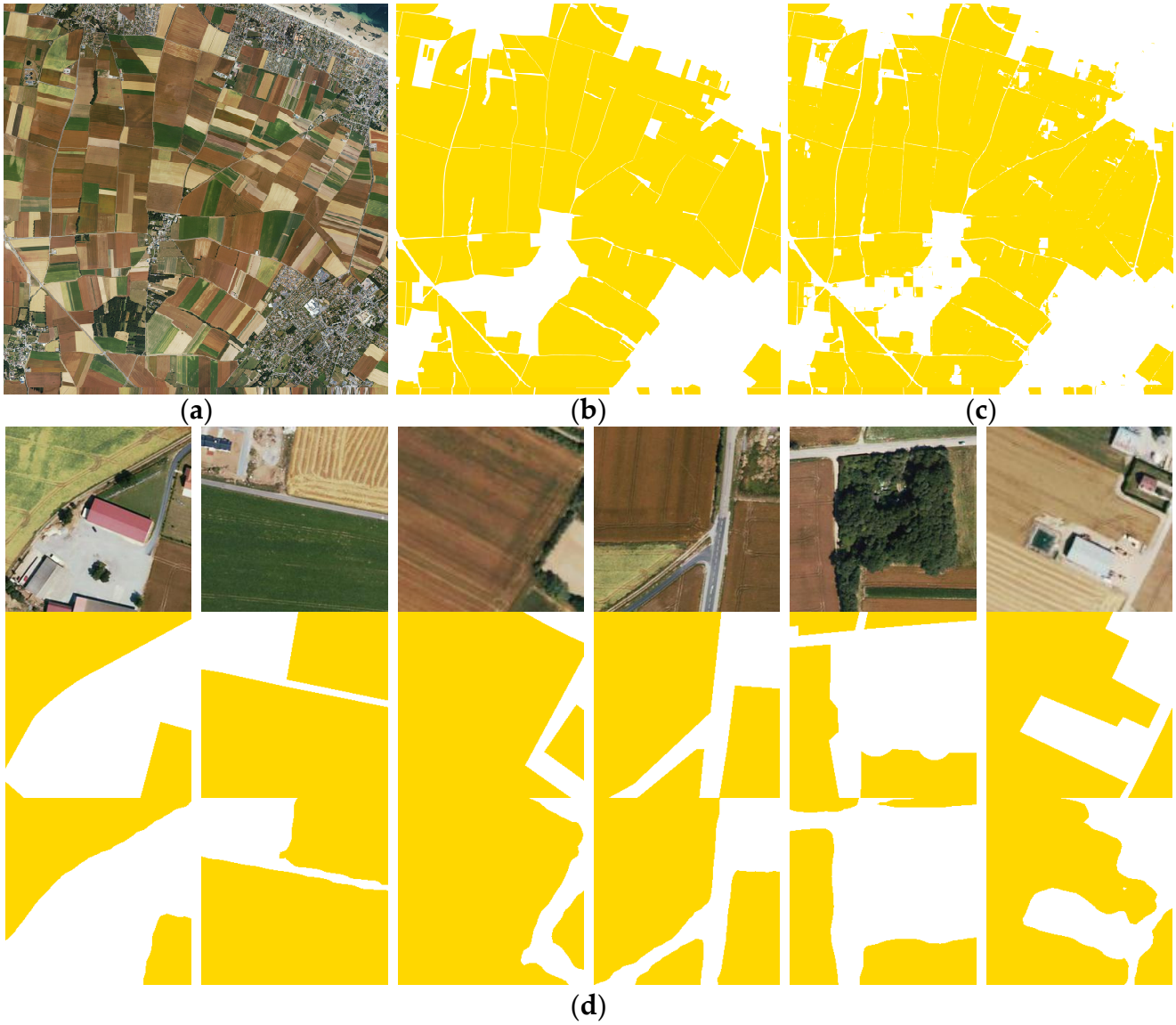


Figure 7. Extraction results of the method in this paper: (**a**) original figures; (**b**) 2006 ground truth; (**c**) 2012 ground truth; (**d**) detailed drawings (from top to bottom: original, label, result of MC&N-PSPNet).

3.2. Control Experiment

To verify the effectiveness of the semantic segmentation model proposed in this paper, it is compared with the results of U-Net, DeepLabV3+ [25], and proposed methods with those reported in previous work [26]. Some of the results are shown in Figure 8. In this paper, the performance of different network models is quantitatively evaluated using six evaluation metrics such as average crossover and merge rate, average pixel accuracy, user accuracy, producer accuracy, overall accuracy, and Kappa coefficient. The results are shown in Table 1.

Figure 8. Extraction results of different model agricultural areas: (**a**) original figures; (**b**) ground truth; (**c**) original network; (**d**) U-Net; (**e**) DeepLabV3++; (**f**) literature [26] (**g**) MC&N-PSPNet.

Table 1. Comparison of the precision of extraction results of different methods in agricultural areas.

Feature Type	Original Network		U-Net		Literature [26]		DeepLabV3++		MC&N -PSPNet.	
	UA%	PA%	UA%	PA%	UA%	PA%	UA%	PA%	UA%	PA%
Agricultural Area	97.26	91.95	96.79	91.97	97.20	95.48	**99.03**	87.01	97.36	**95.58**
Non-Agricultural Area	80.59	93.05	80.09	91.65	89.52	93.35	66.34	**96.79**	**89.74**	93.72
MIoU	82.87		81.84		88.55		75.60		**88.90**	
MPA	88.97		88.45		93.37		82.69		**93.55**	
OA%	92.24		91.70		93.86		89.06		**95.03**	
Kappa	0.8099		0.7971		0.8674		0.7170		**0.8815**	

UA and PA in the table represent user accuracy and producer accuracy for each object type, respectively. The bolded font represents the highest precision.

As can be seen in Figure 8, the agricultural area extracted by the original network is incomplete, and the boundary is not clear. The layers of the ResNet50 backbone feature extraction network used by PSPNet are deeper, the internal differences of each wetland type in the test area are small, and the advantages of the deep network structure cannot be expressed, which causes the agricultural area extraction results to have the phenomena of sticking and fragmentation, as shown in Figure 8c. The extraction of the agricultural area with simple edges is performed well by U-Net, but when faced with complex edges, the edges are blurred and inaccurate, and the training time is the longest among all the methods. U-Net has a good effect for the extraction of agricultural areas with simple edges, but when facing agricultural areas with complex edges, the edges are fuzzy and inaccurate, there is the phenomenon of missed extraction, and the training time is the

longest among all methods. As for the DeepLabV3+ classification confusion phenomenon, the extraction results have an obvious "pretzel phenomenon" that is serious, wherein a small field is part of an incorrect mention: the non-agricultural area being part of the classification for the agricultural area. In the literature [26], through the introduction of the SENet module for classification before PSPNet feature extraction, the method applied in the HRSCD dataset can only be extracted from the general outline because the boundary of the misclassification is serious and a small number of missed points will occur. This paper's method effectively reduces the misclassification phenomenon in the extraction process of agricultural areas by superimposing the results of pyramid pooling, improving the correct rate of information, and improving the accuracy and robustness of the model without increasing the complexity of the model through the attention mechanism NAM, which improves the extraction of agricultural areas. However, the edges of the extraction results of the method in this paper also have a slight amount of jaggedness. A loophole in labeling is also found when comparing the results of multiple groups of methods, such as in Detailed Figure 1: each method determines the parcel in the upper-left corner as an agricultural area, and reviewing the original figure reveals that there is indeed a problem with visual interpretation, and the subsequent semi-supervised or unsupervised approach will be used to complete the work of producing labels to eliminate the influence of subjectivity. This finding also laterally certifies the efficient accuracy of the model. As can be seen from Table 1, the production accuracy of this paper's method in the agricultural area, the user accuracy in the non-agricultural area, and the MioU, MPA, overall accuracy, and Kappa coefficient are all higher than those of other methods. The overall accuracy is 95.03% and the Kappa coefficient is 0.8815, which indicates that the results of this paper have a high degree of consistency with the actual distribution of the landforms.

In order to verify the efficiency of the attention module used in this paper, the NAM module is replaced with CBAM, SENet, ECA, and BAM modules in the same location in turn for comparison tests, and the accuracy assessment results are shown in Table 2. The accuracy assessment results show that the production accuracy and user accuracy of agricultural and non-agricultural areas, as well as the overall six evaluation accuracies of MioU, MPA, OA, and Kappa value, are higher than the other methods, which corroborates that the addition of NAM is the most suitable for the network model structure in this paper.

Table 2. Comparison of the accuracy of the extraction results for different attention modules for agricultural areas.

Feature Type	CBAM		SENet		ECA		BAM		MC&N -PSPNet.	
	UA%	PA%	UA%	PA%	UA%	PA%	UA%	PA%	UA%	PA%
Agricultural Area	97.27	92.07	94.56	92.39	96.67	92.39	95.27	92.71	**97.36**	**95.58**
Non-Agricultural Area	81.08	93.63	82.48	87.09	82.10	91.63	83.14	88.64	**89.74**	**93.72**
MIoU	83.57		80.61		82.95		81.88		**88.90**	
MPA	89.37		88.52		89.38		89.20		**93.55**	
OA%	92.56		90.84		92.18		91.53		**95.03**	
Kappa	81.86		0.7819		0.8110		0.7978		**0.8815**	

The most effective value plus black display.

The MC&N-PSPNet proposed in this paper introduces the CBAM module into the MobileNetV2 backbone feature extraction network, enhancing the model's ability to resolve agricultural areas in remote sensing images. This improvement aids in distinguishing between agricultural and non-agricultural areas that may exhibit similarities, leading to greater efficiency, albeit with some loss in effectiveness. Furthermore, the introduction of the NAM module after feature fusion helps the model to capture richer and more distinctive feature information at different levels, facilitating the detection of agricultural areas with varying features due to differences in planting patterns and crop types. The agricultural area producer accuracy, non-agricultural area user accuracy, overall accuracy

(OA), and Kappa values achieved by this method surpass those of other methods. The lower agricultural area user accuracy is attributed to significant differences in spectral information features among different crops, leading to potential misclassification. Experimental results demonstrate that the network model proposed in this paper achieves an overall accuracy of 95.03% and a Kappa coefficient of 0.8815 on the HRSCD dataset, indicating a high level of consistency with the actual distribution of land features.

3.3. Ablation Experiment

To verify the effectiveness of the backbone feature extraction network before and after the replacement, the improvement of the replaced MobileNetV2, and the addition of the NAM module on the efficiency and accuracy improvement of the model, ablation experiments are conducted in this paper, and the results are shown in Table 3.

Table 3. Ablation experiment results.

Network Model	Backbone	IBN	NAM	OA%	Kappa	Training Time
B1	ResNet50			91.24	0.7999	2.522
B2	MobileNetV2			94.61	0.8728	**1.993**
B2 + IBN	MobileNetV2	√		94.75	0.8761	2.116
B2 + NAM	MobileNetV2		√	94.86	0.8784	2.418
B2 + IBN + NAM	MobileNetV2	√	√	**95.03**	**0.8815**	2.109

Baseline uses the original PSPNet model, the original network. Values in bold font are the best, B1: Baseline 1, B2: Baseline 2, IBN: improved backbone network.

As can be seen from Table 3, after replacing the backbone feature extraction network ResNet50 with MobileNetV2, the overall accuracy is improved by 3.37%, the Kappa coefficient is increased by 0.0729, and the training time is accelerated by 0.589h; after improving the backbone network and adding the CBAM module, the overall accuracy is improved by 3.51% and the Kappa coefficient is increased by 0.0762; after adding the attention mechanism NAM, the extraction accuracy is improved by 3.62%, and the Kappa coefficient is improved by 0.0785; after improving the backbone feature extraction network and adding the NAM module after feature fusion, the overall extraction accuracy is improved by 3.79%, the Kappa coefficient is improved by 0.0816, and the training time is accelerated by 0.413 h.

3.4. Change Detection

Change detection is carried out by different methods based on the results extracted from the agricultural area in this paper. The graph of the results of the experimental detection part of the change detection is shown in Figure 9.

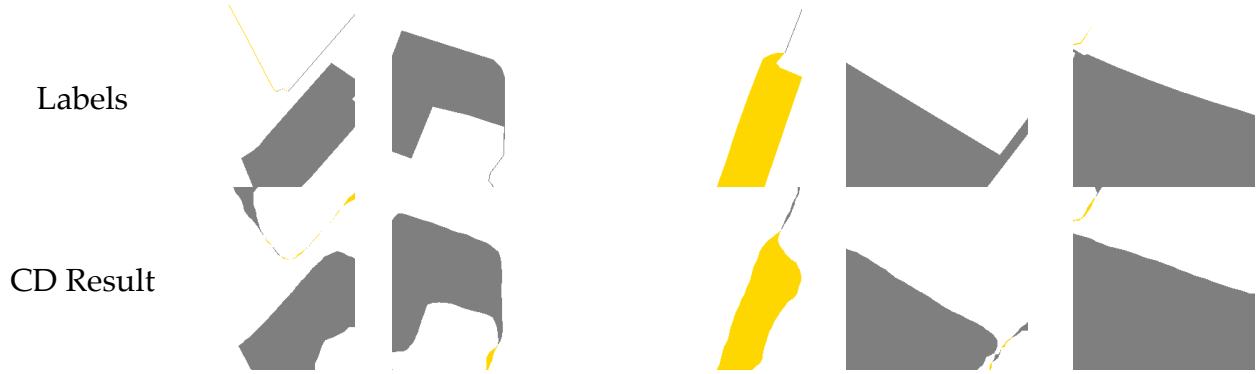


Figure 9. Chart of change detection results.

Based on the method in this paper, the OA and Kappa values for change detection reached 93.24% and 92.29%, respectively.

4. Discussion

Application Development, Directions, and Limitations

The method in this paper will ensure the accuracy and efficiency of complete extraction of the basis of all agricultural areas, but the ability to screen out the visual interpretation of label loopholes, and then optimize the label. Through the results of change detection, it can also be seen that the method in this paper can detect the direction of change, and accurately obtain the bi-directional change of agricultural and non-agricultural areas. Based on this, future research will continue to optimize the dataset and the model according to different application directions, so that it can detect the specific change in agricultural land to another land type or detect contaminated soil. By comparing the recognition results with the recent literature [26], it is found that there are omissions in the recognition of the method in literature [26], and the edge smoothing is worse than the method of this paper. As far as accuracy is concerned, the overall accuracy of this paper's method is 1.17% higher. In terms of efficiency, the total training time of this paper's method is 0.31h faster.

Regarding the application level of this paper's methodology, this has important potential benefits for agricultural practices, for example, in precision agriculture management. It accurately categorizes and monitors changes in agricultural zones, so that farmers and agricultural managers can more accurately understand changes in arable land and develop more effective planting and management strategies to improve crop yields and quality. In terms of resource optimization and allocation, the model can identify and monitor changes in agricultural zones to help agricultural managers to rationally allocate resources, such as water resources, fertilizers, and pesticides, to reduce waste and improve resource utilization efficiency. In disaster assessment and response, after natural disasters (e.g., floods, droughts), the model can quickly assess the impacts of disasters on agricultural zones, provide accurate change detection data, and support post-disaster recovery efforts and the implementation of disaster mitigation measures. In terms of environmental protection, by monitoring changes in agricultural areas, environmental degradation and soil erosion problems caused by irrational farming can be detected and prevented promptly, promoting sustainable agricultural development. In terms of policy formulation and planning, the government and relevant institutions can use the data provided by the model to formulate scientific agricultural policies and planning, promote the process of agricultural modernization, and enhance the overall competitiveness of agriculture. In summary, the MC&N-PSPNet model proposed in this paper not only technically realizes high-precision classification and change detection in agricultural areas, but also shows great potential and value in practical applications.

However, practical applications still suffer from dataset limitations, model complexity, image quality, real-time processing capabilities, and adaptability to different domains. For example, detecting changes in cropping patterns or crop types is critical for accurate monitoring of agricultural areas. Different crops exhibit unique spectral characteristics and growth cycles that may not be fully captured by current datasets and model configurations. To improve model sensitivity to different crop types, future research will integrate multi-temporal and multi-spectral data. This approach can help to distinguish between various crops and their respective growth stages. In addition, combining auxiliary data sources such as crop growth cycles, soil moisture, and meteorological data can further improve the accuracy and reliability of the model in identifying and categorizing crop types. The ability of this paper's method to generalize across different agricultural landscapes and cropping practices is also a key aspect to consider. Agricultural regions exhibit significant variability in field sizes, shapes, and crop arrangements, as influenced by regional agricultural practices and local policies. Current models may face challenges when applied to regions with different agricultural systems than those represented in the training data. Non-photorealistic rendering of remote sensing images poses additional challenges. Non-photorealistic images, often resulting from preprocessing techniques aimed at enhancing specific features, can introduce artifacts and distortions that affect the model's performance. These distortions might lead to misclassification and reduce the model's

accuracy. To mitigate this, the model should be trained and tested on both photorealistic and non-photorealistic datasets, ensuring it can handle various image processing artifacts. Additionally, developing advanced preprocessing methods that preserve essential image features while minimizing distortions can enhance the model's performance in real-world applications. In summary, while the MC&N-PSPNet model demonstrates significant potential in detecting changes between agricultural and non-agricultural areas, addressing these limitations through the integration of diverse data sources, expansion of training datasets, consideration of non-photorealistic rendering effects, and optimization of computational requirements will be crucial for enhancing its accuracy, reliability, and generalizability. These improvements will provide a more comprehensive and scientifically robust foundation for agricultural resource management and policy formulation.

5. Conclusions

Based on the PSPNet model, this paper achieves rapid and accurate extraction of agricultural areas through the replacement of the backbone feature extraction network, improvement of the backbone network, and addition of an attention mechanism. The overall accuracy and Kappa value of the final change detection reached 93.24% and 92.29%, respectively, outperforming other competitive methods in terms of performance and achieving a better trade-off between model complexity and performance. This research result also has a vital theoretical value in monitoring agricultural areas by using remote sensing technology and further improves the theoretical system of a remote sensing classification model. The main conclusions are as follows:

(1) Replacement of backbone feature extraction network. The backbone feature extraction network in the original PSPNet is replaced by ResNet50 with the improved MobileNetV2. This substitution addresses the issue of minimal intra-class differences within the study area, which hinder the expression of the advantages of deep network structures. Moreover, it enhances model accuracy while reducing parameter count and shortening training time.

(2) Addition of NAM module. The NAM module is integrated into the PSPNet network structure, enhancing model training accuracy and efficiency without increasing network complexity. This achieves the original goal of creating a more lightweight network.

(3) To sum up, detecting changes in cropland requires the application of suitable techniques for seamless monitoring. While 2D change detection has been widely used, incorporating 3D detection techniques could provide additional valuable insights. Accurate mapping of changes in cropland is essential for understanding the underlying causes of and for analyzing both the ecological and socio-economic consequences of these changes. This understanding is critical for devising effective land management and policy-making strategies. Future work will focus on the extraction of diverse features to enhance the robustness of change detection models. By incorporating multi-temporal, multi-spectral, and spatial features, we aim to capture more complex patterns of change. Additionally, we plan to explore and implement more advanced classifiers, which have shown great promise in other remote sensing applications. These efforts will contribute to improving the accuracy and reliability of cropland change detection, ultimately supporting sustainable agricultural practices and land use management.

Author Contributions: Y.C., W.Z. and C.W. devised the project, the main conceptual ideas, and the proof outline; Y.C. performed the training of the neural network and the writing of the first draft; Y.D. and Q.Y. analyzed the background literature; W.Z. and C.W. were responsible for communicating and coordinating with all co-authors in a timely manner; W.Z., Y.D. and Q.Y. critically revised the article; Y.C., Y.D., W.Z., C.W., Q.Y. and X.W. contributed to the preparation of and figure items at various stages. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for Central Non-Profit Scientific Institution of China (Y2023PT05), the Education Bureau of Liaoning Province (LJKMZ20220638), Liaoning Institute of Science and Technology (2307B29).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available at [https://doi.org/10.21227/azv7-ta17].

Acknowledgments: First of all, I would like to express my heartfelt thanks to all the people who helped me during the research toward and writing of this paper. I would like to thank Zhang and Duan for their valuable comments and revisions of the article, Wang for his assistance with the experiments, Duan and Wang for their financial support of this study, and, finally, I sincerely thank all those who spent a lot of time reading this thesis and gave me many suggestions that will benefit me in my future studies.

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

Glossary

Abridge	Annotations
PSPNet	Network model: Pyramid Scene Parsing Network
MC&N-PSPNet	Methods in this paper's network model: CBAM into MobileNetV2 and NAM into PSPNet
ResNet50	Network model: Residual Network-50
MobileNetV2	Network model: Inverted Residuals and Linear Bottlenecks
HRSCD	Dataset: High Resolution Semantic Change Detection
CBAM	Attention modules: Convolutional Block Attention Module
NAM	Attention modules: Normalization-Based Attention Module
SENet	Attention modules: Squeeze-and-Excitation Networks
ECA	Attention modules: Efficient Channel Attention
BAM	Attention modules: Bottleneck Attention Module
MIoU	Evaluation indicators: Mean Integration over Union
MPA	Evaluation indicators: Mean Pixel Accuracy
OA	Evaluation indicators: Overall Accuracy
PA	Evaluation indicators: Producers Accuracy
UA	Evaluation indicators: Users Accuracy
Kappa	A measure of classification accuracy

References

1. Zhang, L.; Cheng, J. Talking about arable land protection with changes in China's arable land in 2015. *West China Dev.* **2018**, *3*, 58–62.
2. Foley, J.A.; DeFries, R.; Asner, G.P. Global Consequences of Land Use. *Science* **2005**, *309*, 570–574. [CrossRef]
3. Wang, J.; Li, P.; Zhan, Y.; Tian, S. Research on the Protection and Enhancement of Cultivated Land Quality in China. *China Popul. Resour. Environ.* **2019**, *29*, 87–93.
4. Zhang, L.; Wu, C. Advance and Future Development of Change Detection for Multi-temporal Remote Sensing Imagery. *Acta Geod. Cartogr. Sin.* **2017**, *46*, 1447–1459. [CrossRef]
5. Thenkabail, P.S. *Remote Sensing Data Characterization, Classification, and Accuracies: Advances of the Last 50 Years and a Vision for the Future*; CRC Press: Boca Raton, FL, USA, 2015. [CrossRef]
6. Wang, Y.; Meng, Q.; Qi, Q.; Yang, J.; Liu, Y. Region merging considering within-and between-segment heterogeneity: An improved hybrid remote-sensing image segmentation method. *Remote Sens.* **2018**, *10*, 781. [CrossRef]
7. Yang, J.; He, Y.; Caspersen, J. Region merging using local spectral angle thresholds: A more accurate method for hybrid segmentation of remote sensing images. *Remote Sens. Environ.* **2017**, *190*, 137–148. [CrossRef]
8. Cheng, H.-D.; Jiang, X.H.; Sun, Y.; Wang, J. Color image segmentation: Advances and prospects. *Pattern Recognit.* **2001**, *34*, 2259–2281. [CrossRef]
9. Cao, K.; Zhang, X. An improved res-unet model for tree species classification using airborne high-resolution images. *Remote Sens.* **2020**, *12*, 1128. [CrossRef]
10. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
11. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18, 2015, pp. 234–241.

12. Wang, Z.; Wang, J.; Yang, K.; Wang, L.; Su, F.; Chen, X. Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3+. *Comput. Geosci.* **2022**, *158*, 104969. [CrossRef]
13. Yuan, P.; Zhao, Q.; Zhao, X.; Wang, X.; Long, X.; Zheng, Y. A transformer-based Siamese network and an open optical dataset for semantic change detection of remote sensing images. *Int. J. Digit. Earth* **2022**, *15*, 1506–1525. [CrossRef]
14. Fang, S.; Li, K.; Shao, J.; Li, Z. SNUNet-CD: A densely connected Siamese network for change detection of VHR images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 3056416. [CrossRef]
15. Zhan, T.; Song, B.; Xu, Y.; Wan, M.; Wang, X.; Yang, G.; Wu, Z. SSCNN-S: A Spectral-Spatial Convolution Neural Network with Siamese Architecture for Change Detection. *Remote Sens.* **2021**, *13*, 895. [CrossRef]
16. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
17. Chen, H.; Shi, Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens.* **2020**, *12*, 1662. [CrossRef]
18. Yuan, P.; Wang, K.; Xiao, J. Cropland Extraction from High-Score Imagery Based on RMAU-Net Network Modeling. *Hubei Agric. Sci.* **2023**, *62*, 182–188+196. [CrossRef]
19. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
20. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
21. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
22. Liu, Y.; Shao, Z.; Teng, Y.; Hoffmann, N. NAM: Normalization-based attention module. *arXiv* **2021**, arXiv:2111.12419. [CrossRef]
23. Jie, H.; Li, S.; Gang, S.; Albanie, S. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *42*, 2011–2023. [CrossRef]
24. Park, J.; Woo, S.; Lee, J.-Y.; Kweon, I.S. Bam: Bottleneck attention module. *arXiv* **2018**, arXiv:1807.06514. [CrossRef]
25. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
26. Cui, W.; Wu, Y.; Xue, W. Research on forest land extraction method based on improved PSPNet model for high-resolution remote sensing images. *Sci. Technol. Innov.* **2024**, *62*, 52–55.