remote sensing

MDPI

Article

LPMSNet: Location Pooling Multi-Scale Network for Cloud and Cloud Shadow Segmentation

Xin Dai [1], Kai Chen [1], Min Xia [1,*], Liguo Weng [1] and Haifeng Lin [2]

[1] Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China; 20211249020@nuist.edu.cn (X.D.); 20211249015@nuist.edu.cn (K.C.); 002311@nuist.edu.cn (L.W.)

[2] College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China; haifeng.lin@njfu.edu.cn

* Correspondence: xiamin@nuist.edu.cn

Abstract: Among the most difficult difficulties in contemporary satellite image-processing subjects is cloud and cloud shade segmentation. Due to substantial background noise interference, existing cloud and cloud shadow segmentation techniques would result in false detection and missing detection. We propose a Location Pooling Multi-Scale Network (LPMSNet) in this study. The residual network is utilised as the backbone in this method to acquire semantic info on various levels. Simultaneously, the Location Attention Multi-Scale Aggregation Module (LAMA) is introduced to obtain the image's multi-scale info. The Channel Spatial Attention Module (CSA) is introduced to boost the network's focus on segmentation goals. Finally, in view of the problem that the edge details of cloud as well as cloud shade are easily lost, this work designs the Scale Fusion Restoration Module (SFR). SFR can perform picture upsampling as well as the acquisition of edge detail information from cloud as well as cloud shade. The mean intersection over union (MIoU) accuracy of this network reached 94.36% and 81.60% on the Cloud and Cloud Shadow Dataset and the five-category dataset L8SPARCS, respectively. On the two-category HRC-WHU Dataset, the accuracy of the network on the intersection over union (IoU) reached 90.51%. In addition, in the Cloud and Cloud Shadow Dataset, our network achieves 97.17%, 96.83%, and 97.00% in precision (P), recall (R), and F1 score (F_1) in cloud segmentation tasks, respectively. In the cloud shadow segmentation task, precision (P), recall (R), and F1 score (F_1) reached 95.70%, 96.38%, and 96.04%, respectively. Therefore, this method has a significant advantage over the current cloud and cloud shade segmentation methods.

Keywords: cloud and cloud shadow detection; convolutional neural network; multi-scale extraction; channel space attention; scale fusion restoration

check for updates

Citation: Dai, X.; Chen, K.; Xia, M.; Weng, L.; Lin, H. LPMSNet: Location Pooling Multi-Scale Network for Cloud and Cloud Shadow Segmentation. *Remote Sens.* **2023**, *15*, 4005. https://doi.org/10.3390/rs15164005

Academic Editor: Mohamed Lamine Mekhalfi

Received: 4 July 2023
Revised: 3 August 2023
Accepted: 10 August 2023
Published: 12 August 2023

1. Introduction

Following the swift growth of satellite imagery in recent decades, the segmentation task of remote sensing photos is now employed flexibly in disaster warning, geological research, and other tasks [1–3]. The exact segmentation of clouds as well as cloud shading can significantly improve the meteorological bureau's monitoring and forecasting efficiency on the atmospheric environment [4–6]. As a result, cloud as well as cloud shade segmentation perform an important part in current meteorological forecasting and early warning activities [7–9].

Traditional cloud and cloud shadow segmentation techniques are classified in two types: threshold-based approaches and colour, texture, as well as shape-based methods. These techniques have better results in some simple scenarios. They are, however, ineffective in picture segmentation under complicated situations and lighting conditions and could be frequently interfered with a variety of elements, leading to false detection and missed detection.

Recently, cloud and cloud shade segmentation methods based on fully connected networks and convolutional networks have become a research hotspot. Also, there are some semantically segmented networks for satellite photos that already exist today. Zhao et al. [10] proposed the Pyramid Pooling Module. It can obtain multi-scale information of images and complete effective image classification tasks. DeepLabV3Plus proposed by Chen et al. [11] uses dilated convolutions to obtain contextual information in different regions about the image, which improves segmentation accuracy. Zhang et al. [12] proposed ACFNet. Relying on the attention category feature module, the network can adaptively segment different classification categories based on per-pixel calculations. To overcome the issue of excessive computation in an attention module and to enhance segmentation efficiency, CCNet [13] introduced a Criss-Cross Attention Module by improving Non-Local Neural Networks [14]. The emergence of visual Transformer has also promoted further development of semantic segmentation models. CvT [15] improves Transformer performance and efficiency by incorporating convolution into ViT. Swin-T [16] employs a window-shifting attention operation to restrict self-attention computation to non-overlapping local windows, while also allowing cross-window connections, leading to higher efficiency and segmentation performance. PvT [17] improves the Transformer architecture and introduces the pyramid structure into the Transformer. When downsampling the feature map, the size of the feature map can be reduced, reducing the amount of calculation, thereby improving the ability of dense level prediction. However, the general problem of the above methods is that they do not have a high performance improvement in the segmentation tasks for cloud and cloud shadow. They lack the ability to capture the edge details of cloud and cloud shadow, resulting in blurred edges of segmented images and poor segmentation results.

The aforementioned networks have become semantically segmented nets with broad applicability in recent years; however, they still have flaws when it comes to cloud as well as cloud shade segmentation tasks [18]. Xia et al. [19] recommended GAFFNet to overcome these challenges. The Arous Spatial Pyramid Pooling Module (ASPP) was improved in this network to gather multi-scale deep semantic info. However, the module's shortcoming is the fact that it solely employs the convolution operation to extract feature information, ignoring the importance of global information in semantic segmentation information, causing the loss of picture feature info. A Parallel Asymmetric Dual-Attention Network PANDA was proposed by Xia et al. [20]. Using a Dual Attention Module, the network improves its dedication to the split object's category info. The problem is that the convolution-based attention module has a limited ability to extract global information, cannot fully capture the picture's long-distance dependencies, as well as is simple to use to make the segmentation the objective's edges rough. Miao et al. [21] proposed a Multi-Scale Fusion Module (MF), which fuses local info and multi-scale info to obtain globally feature info. Although it recognises the relevance of global features, the convolution operation's capacity to extract global features is restricted, and network can still disregard the edge info of cloud as well as cloud shade. Hu et al.'s [22] cloud as well as cloud shade detection network CDUNet optimises cloud edge segmentation by processing and analysing image semantic info. Simultaneously, they employed self-attention to increase a network's ability to regulate global info, boosting the network's segmentation effect. The network's downside is that it employs an extensive amount of convolution operations, leading to an excessive amount of calculation variables for the model as well as low segmentation performance. MSPFANet proposed by Lu et al. [23] employs a Multi-Scale Pooling Feature Aggregation Module. This module use MSPFA for extracting multi-scale semantically info, so as to achieve the acquisition of deep semantic information of cloud and cloud shade. However, this method lacks the capture of shallow global information, and it is easy to lose the edge information of cloud and cloud shade. Classical semantic segmentation networks can effectively extract semantic information at different levels of feature maps [24], but they lack effective attention modules, ignoring the importance of attention operations for capturing information about segmented target objects [25]. At the same time, they perform too many down-sampling operations, which severely degrades the feature map information and causes the picture's location information to shift, which ends in missing and erroneous detection. In addition, networks dedicated to cloud and cloud shade segmentation also have shortcomings. They fail to capture enough long-range dependencies of pictures. Image's long-range relationship can assist the network in comparing and classifying the correlation between pixels at different distances [26,27]. Due to a lack of appropriate long-range dependencies, these networks may lose picture feature information.

Aiming to address the above problems, this paper designs a cloud and cloud shade segmentation network with ResNet50 [28,29] as the backbone. First, in the encoding stage, we improve the ResNet50. Simultaneously, we designed the Location Attention Multi-Scale Aggregation Module (LAMA). This module can obtain multi-scale information of feature maps through pooled convolution of different convolution kernels and can improve the semantic information of different levels obtained via the backbone network. Using the pooling convolution, the parallel Location Attention Module (LA) may extract the position information of the image's H and W directions, respectively, as well as build the attention feature map in which LPMSNet focuses on a target object. The above feature maps can complement the position information that may be lost in multi-scale extraction. Following that, this paper presents a Channel Spatial Attention Module (CSA). A parallel Channel Attention Module (CA) as well as improved Non-Local Neural Networks comprise CSA. The improved NLNN can extract long-range dependencies of feature maps. The deep CSA can further process the deep semantic information extracted using the SAMA to capture more abstract and global dependencies. While the shallow CSA can extract more detailed and local dependencies. The Channel Attention Module can use shared weights through MLP to increase the module's generalisation and efficiency. Finally, the Scale Fusion Repair Module (SFR) is created to complete an upsampling operation. This module can fuse and decode the features of different levels extracted from the encoding structure through skip connections and the upsampling information input into the SFR, effectively complementing the feature information. The stacked convolution operation can repair the edge information lost by the cloud and cloud shade in the downsampling operation. Studies reveal that the strategy suggested in this research outperforms the previously mentioned semantic segmentation network. Several innovations are made via the model and architecture given in this work:

1. The Location Attention Multi-Scale Aggregation Module (LAMA) aims to gather multi-scale info by pooling convolutions of different convolution kernels, thus enhancing LPMSNet's capacity to collect semantic info across various scales. The parallel-input Location Attention Module (LA) can extract location information along the horizontal and vertical directions of the feature map through average pooling convolution, respectively. This location info can direct the network's attention to the intended objects. Simultaneously, after horizontal and vertical position information is embedded with multi-scale information, it can complement the position code lost in multi-scale feature extraction.
2. A Channel Spatial Attention Module (CSA) is created to eliminate the detrimental impact of background noise on cloud as well as cloud shade segmentation. The enhanced Non-Local Neural Networks can extract the long-distance dependencies of feature maps across the network's shallow and deep layers. Channel Attention Module (CA) can dynamically adjust the weight of features, assisting NLNN in focusing on long-distance dependencies at different levels. The internal MLP can share weights via convolution operations, reducing the number of parameters as well as boosting the model's generalisation capability and efficiency.
3. During the downsampling process, the proposed Scale Fusion Restoration Module (SFR) could combine distinct categories of contextual info and deep semantic info. Simultaneously, SFR effectively fixes the edge info of cloud as well as cloud shade via stacked convolution operations, increasing the cloud and cloud shade segment impact.

2. Methodology

At present, most cloud as well as cloud shade segment networks are currently not optimal for cloud and cloud shade edge segmentation. We present a Location Pooling Multi-Scale Network (LPMSNet) to address the difficulties of misclassification induced by background noise interference and rough edge segmentation in cloud and cloud shade segmentation, as demonstrated through Figure 1. It could segment cloud as well as cloud shade quickly and successfully. We design a cloud and cloud shade segmentation network with a modified ResNet50 as the backbone. We start with the enhanced ResNet50 for feature extraction and then utilise the Location Attention Multi-Scale Aggregation Module (LAMA) for obtaining multi-scale feature information from a picture of any size. Simultaneously, the parallel Location Attention Module (LA) within LAMA embeds a picture's location attention feature map into multi-scale info to supplement the missing location encoding during the downsampling procedure. The Channel Spatial Attention Module (CSA) deep-processes the feature info extracted in the deep and shallow layers of LPMSNet and dynamically adjusts the network's focus on long-distance dependencies at different levels. Finally, the Scale Fusion Restoration Module (SFR) fuses the context information of each layer in the improved ResNet50 with the deep semantic information through skip connections to obtain richer and more accurate semantic segmentation results. Simultaneously, the convolution stacked inside SFR may repair the edge info of cloud as well as cloud shade, making the segmentation target's edges clearer.

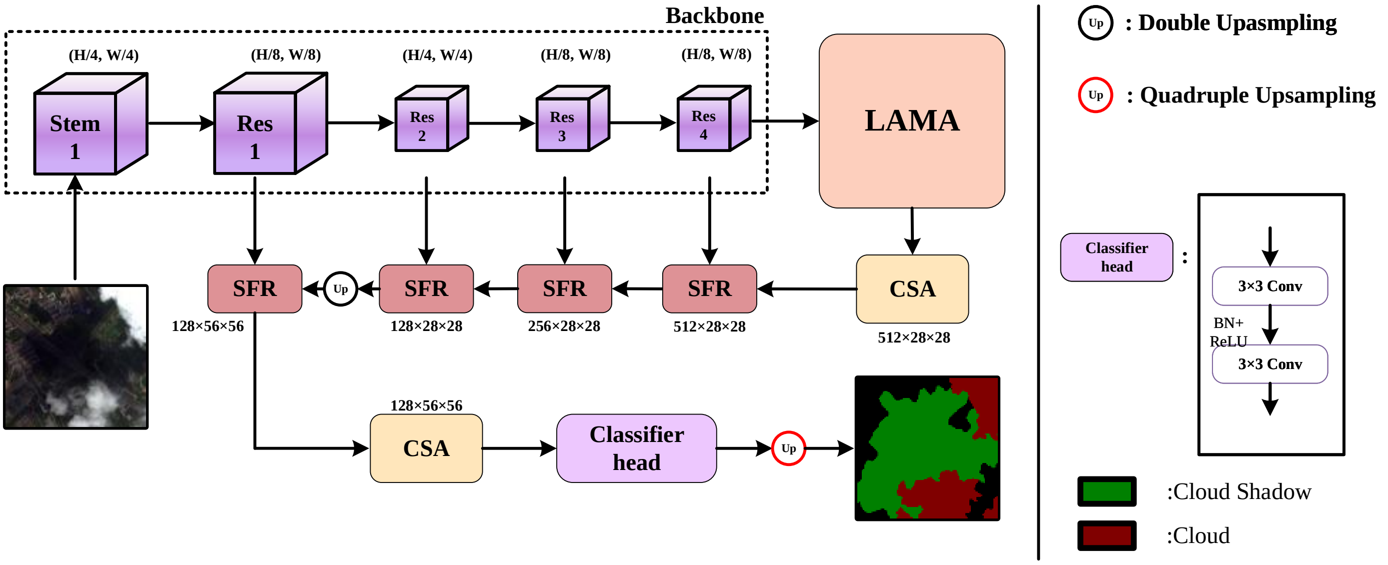

Figure 1. LPMSNet-based cloud as well as cloud shadow semantic segmentation network framework. The encoding module includes ResNet50, Location Attention Multi-Scale Aggregation Module (LAMA), Channel Spatial Attention Module (CSA), and Scale Fusion Restoration Module (SFR).

2.1. Backbone

We make improvements to the overall architecture of ResNet50. First, we remove its last fully connected layer. And, instead of using the original 32 times downsampling operation, only eight times downsampling is performed on the feature map. Instead, the dilated convolution is added to the last three or four layers. Its specific structure is shown in Table 1. When the convolutional network is excessively deep, the upgraded ResNet50 solves the usual problems of gradient explosion and gradient disappearance.

Table 1. Architecture of original and modified ResNet50.

Original			Modified	
Layer Stem	50 layer 7 × 7, stride 2	Size 1/2	Modified 50 layer	Size 1/2
L1	3 × 3, Max pool, stride 2 $\begin{pmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 3 \times 3, 64 \end{pmatrix}$	1/4		1/4
L2	$\begin{pmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 3 \times 3, 512 \end{pmatrix}$	1/8		1/8
L3	$\begin{pmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 3 \times 3, 1024 \end{pmatrix}$	1/16	Dilated convolution	1/8
L4	$\begin{pmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 3 \times 3, 2048 \end{pmatrix}$	1/32	Dilated convolution	1/8

2.2. Location Attention Multi-Scale Aggregation Module

The selected scene background in the Cloud and Cloud Shadow Dataset is more complicated, and the pixel value is comparable with the cloud and cloud shade, which strongly interferes with network segmentation, and it is simple to produce detecting errors and omissions in the network. Simultaneously, the cloud's edge feature info is highly rich, shallow clouds are difficult to detect, and the shape of cloud edges is rather rough, resulting in it being easy for the network to misplace the segmentation target's edge info. As a result, cloud and cloud shade segregation presents significant issues. To tackle the issues described above, we need not only enough semantic information to identify the broad category of cloud as well as cloud shade but also a considerable amount of geographical information to mend the segmentation edges of cloud as well as cloud shade. Inspired by the Pyramid Pooling Module (PPM), we design the Location Attention Multi-Scale Aggregation Module (LAMA), whose structure is shown in Figure 2. This module is made up of two parts: a Multi-Scale Aggregation Module (MA) and a Location Attention Module (LA). The Multi-Scale Aggregation Module (MA) extracts multi-scale features in feature maps by parallelising average pooling convolutions of different sizes. Moreover, MA can retain global information at different scales through pooling convolution operations, so that the network can accurately capture the category information of cloud and cloud shadow, which helps to reduce the occurrence of missed and false detection. Meanwhile, the Location Attention Module (LA) aggregates the input feature maps into two independent direction-aware map features by using global pooling operations to extract location feature information along the horizontal and vertical directions of the feature maps, respectively, as shown in Figure 3. This operation can fully extract the position of the pixels of cloud and cloud shadow in the image and generate the attention feature map that the network is interested in for the segmentation target. Afterwards, we embed this attention feature map into a Multi-Scale Aggregation Module. The positional encoding may be lost during the multi-scale extraction procedure owing to the downsampling operation of the feature maps at various stages [30]. At this moment, the attention feature map with position encoding can assist the multi-scale aggregation module in completing the missing position information, allowing correct capture of cloud as well as cloud shade edge info.

The Loading Attention Module's structural representation is as follows:

$$x_{i+1} = Conv\{Cat[Avg_H(x_i), Avg_W(x_i)]\} \tag{1}$$

$$out_h = \text{Sigmoid}\{Conv[Split(BS(x_i))]\} \tag{2}$$

$$out_w = \text{Sigmoid}\{Conv[Permute[Split(BS(x_i))]]\} \tag{3}$$

Here, Avg_H and Avg_W represent the feature map's global average pooling in the horizontal and vertical dimensions, respectively; Cat indicates the connection of two feature information. $Conv$ means 1×1 convolution operation; BS means BatchNorm and Sigmoid operation; $Split$ means splitting feature information in a certain dimension; $Permute$ means swapping the dimension order of a certain feature information; and Sigmoid means Sigmoid operation.

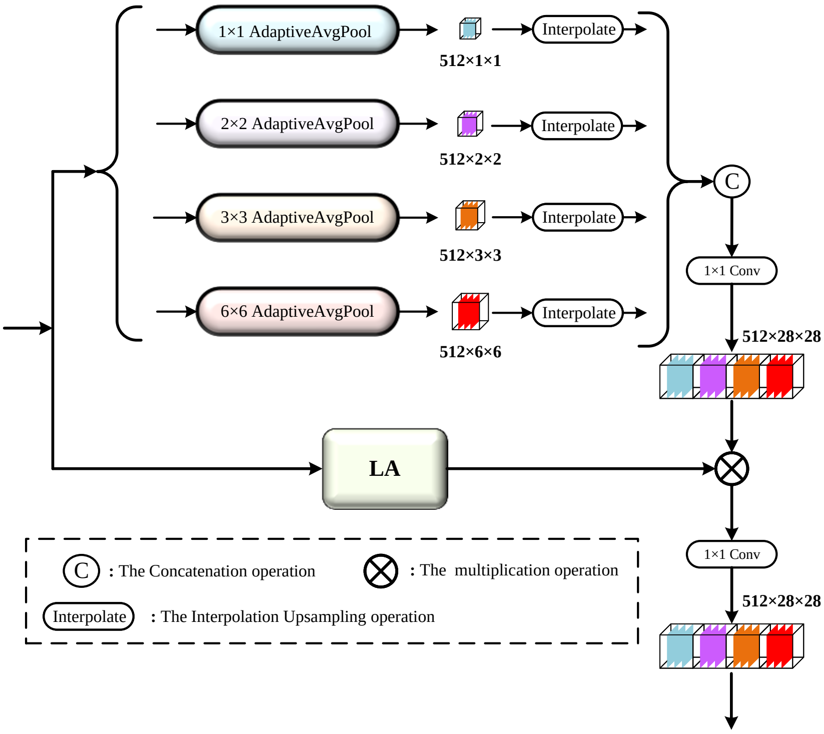

Figure 2. Location Attention Multi-Scale Aggregation Module (LAMA) structure.

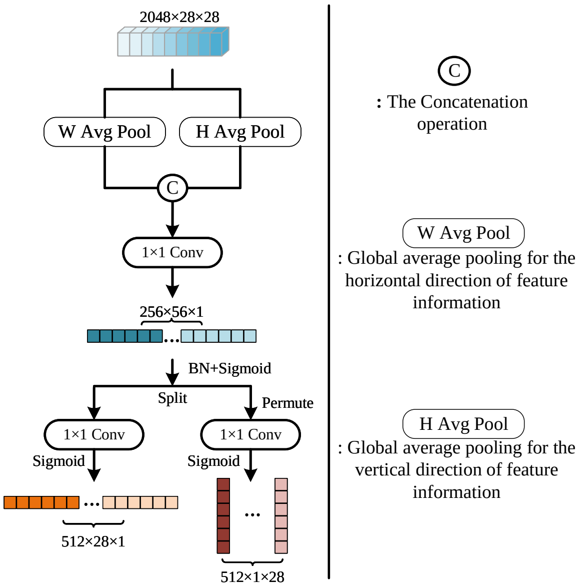
2048×28×28
W Avg Pool
H Avg Pool
C
1×1 Conv
256×56×1
BN+Sigmoid
Split
Permute
1×1 Conv
1×1 Conv
Sigmoid
Sigmoid
512×28×1
512×1×28
C : The Concatenation operation
W Avg Pool : Global average pooling for the horizontal direction of feature information
H Avg Pool : Global average pooling for the vertical direction of feature information

Figure 3. Location Attention Module (LA) structure.

2.3. Channel Spatial Attention Module

The method of attention could assist the model in concentrating on relevant characteristics and areas, reducing missing and erroneous detection of segmentation targets. As a result, they are commonly utilised in computer vision tasks [31–34]. The majority of contemporary attention modules are classified as channel attention modules, spatial attention modules, and attention modules that combine channel and space. A single channel module or spatial attention module lacks the advantages of each other and cannot achieve better attention feature extraction. Attention modules that combine channels and spaces, such as CBAM [35] and SK [36], tend to use convolution operations to extract feature information from images. These modules could successfully increase the system's focus on key characteristics and locations. Convolution, on the other hand, is incapable of capturing long-distance dependencies well, and simple convolution processes are susceptible to missing and erroneous detection. As a result, in this research, a novel Channel Spatial Attention Module is built, and its framework schematic is presented in Figure 4.

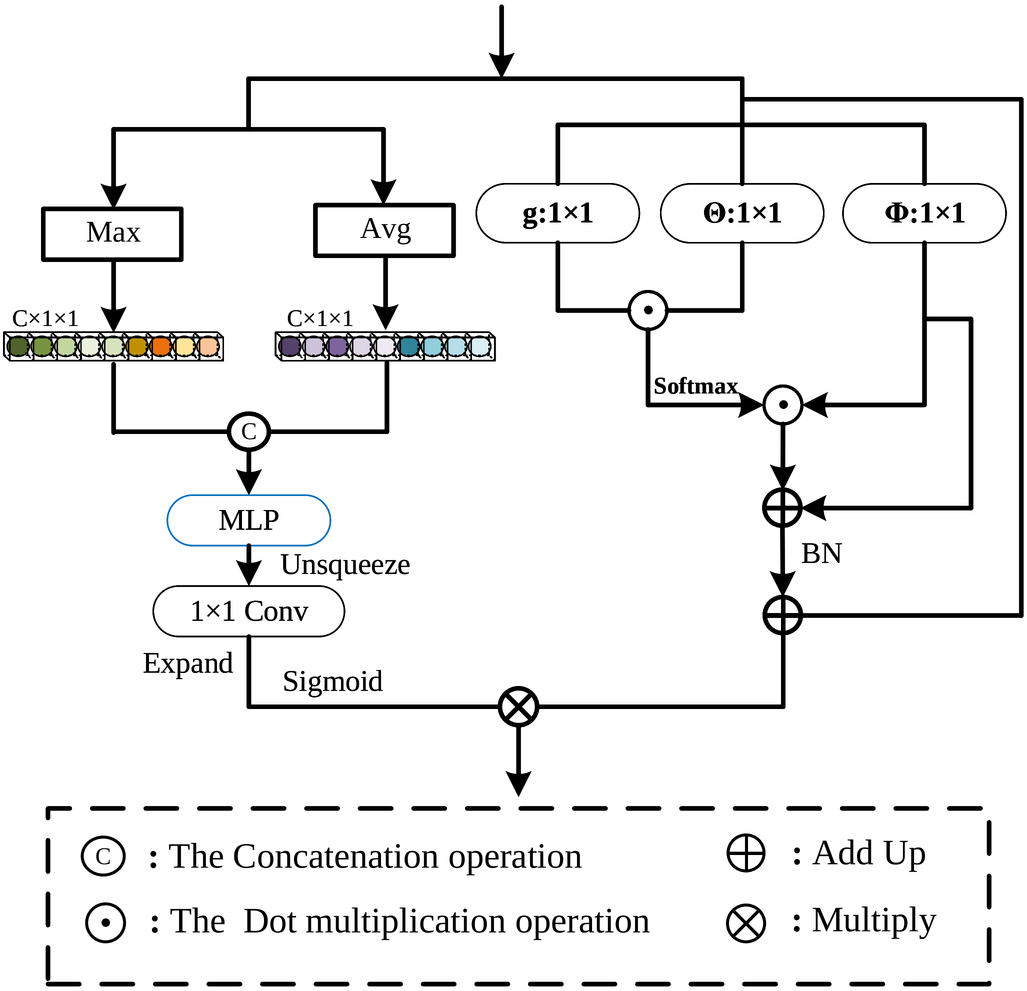

Figure 4. Channel Spatial Attention Module (CSA) structure.

This paper's Channel Spatial Attention Module is made up of two sub-modules. For obtaining high-level characteristics, the Channel Attention Module on the left employs average and maximum pooling procedures. The feature information is merged via the connection operation when it passes via two global pooling scales [37]. Then, the Multi-Layer Perceptron (MLP) further extracts the high-level abstract features in the above feature info. The MLP can effectively model the input semantic information and has a strong model expression ability, so as to accurately identify objects such as cloud and cloud shade that need to be detected. Furthermore, the MLP can share weights via the convolution operations, lowering the total amount of parameters as well as enhancing the model's generalisation ability and performance. Then, as a selector, we utilise a 1 × 1 convolution to flexibly concentrate on the feature representations of the two pooling layers as well as the MLP. Finally, we re-weight the original feature map using Sigmoid. Its structural expression is as follows:

$$\mathrm{x}_{i+1} = Cat[Avg(x_i), Max(x_i)] \tag{4}$$

$$x_{out} = ES\{Conv[UQ(MLP(x_{i+1}))]\} \tag{5}$$

Here, *Avg* and *Max* are the average and maximum pooling operations, respectively; *Cat* indicates connecting two feature information; *MLP* represents the MLP module; *UQ* means dimension compression for a certain dimension of feature information; *Conv* means 1×1 convolution operation; and *ES* means expanded function and Sigmoid function.

The right branch's improved Non-Local Neural Networks can successfully obtain the long-distance relationships required for semantic classification. Long-distance dependencies can assist the network in capturing the global consistency of cloud and cloud shading; better understanding the overall structure of objects; and avoiding backdrop noise interference, which can lead to erroneous and missed detection. Simultaneously, it allows network to more accurately record cloud as well as cloud shade border geometries. Then, the attention unit Φ and the original unprocessed feature information are fused in a way similar to the residual connection to supplement the feature information of the image. Finally, the Spatial Attention Module enhances its ability to capture long-range dependencies through weighted embedding into the improved NNLL. The Channel Attention Module can dynamically adjust the weight of features, assisting NLNN in focusing on long-distance dependencies at different levels. The Channel Attention Module can dynamically adjust the weight of features, assisting NLNN in focusing on long-distance dependencies at different levels. In the entire network structure, the deep CSA can further process the deep semantic information extracted by the SAMA to capture more abstract and global dependencies, while shallower CSA in the upsampling locations can extract more detailed and local dependencies. CSA effectively lowers the occurrence of erroneous as well as missed detection and can better restore cloud and cloud shade edge details.

2.4. Scale Fusion Restoration Module

In the decoding stage, we must effectively use the many layers of contextual info acquired in the encoding stage, along with the semantic info collected in the deep network. Therefore, a simple upsampling operation cannot achieve accurate segmentation of cloud and cloud shade. Figure 5 depicts our Scale Fusion Restoration Module (SFR). To acquire the contextual information extracted during the encoding stage, this module employs skip connections. Then, SFR integrates it with the deep network's rich semantic information to efficiently accomplish the cloud as well as cloud shade classification task. Different levels of context information can help the network to better identify the relationship between cloud and cloud shade pixels, and rich deep semantic information can help the network capture the overall category information of objects. Finally, the thinning convolution module at the conclusion with a stacked convolution kernel of 33 may successfully refine the segmentation border of segmented cloud as well as cloud shade, making it clearer. The structural expression for this module is as follows:

$$U_i = DConv[Cat(U_1, U_2)] \tag{6}$$

$$U_j = Conv(U_2) \tag{7}$$

$$U_{out} = U_i + U_j \tag{8}$$

Here, U_1 represents the feature information of the shallow layer of the network, U_2 represents the feature information of the deep layer of the network, *Cat* represents the connection of two feature information, *DConv* represents two stacked 3 × 3 convolution modules, and *Conv* represents 1 × 1 convolution.

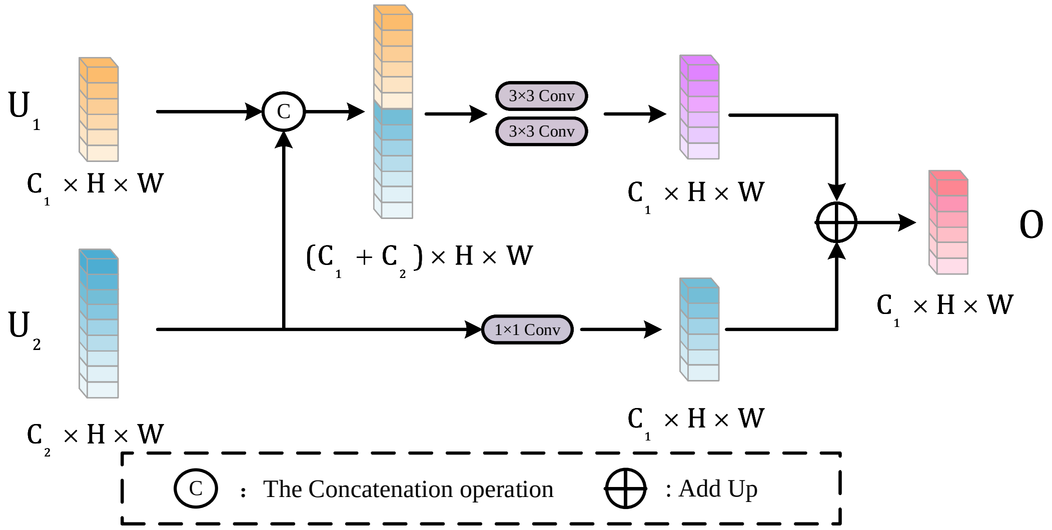

Figure 5. Scale Fusion Restoration Module (SFR) structure.

3. Experiment

3.1. Dataset Introductions

3.1.1. Cloud and Cloud Shadow Dataset

This set of images is derived from remote sensing image data collected by the Landsat-8 and Sentinel-2 satellites, as shown on Google Earth. The dataset has a large dimension span, complex background, and large noise interference, including complex areas such as cities, farmland, hills, plains, and plateaus, and has high requirements for segmentation algorithms. Due to the limitation of GPU computing power, we cut the original high-quality remote sensing dataset with a size of 4800 × 2692 into small images with a size of 224 × 224. After filtering (removing image data with only a single classification category), we obtained a total of 9217 images. We selected 80% images as the training and validation set and 20% images as the test set. The images are labeled into three types: cloud, cloud shadow, as well as backdrop. Figure 6 shows the images and corresponding labels.

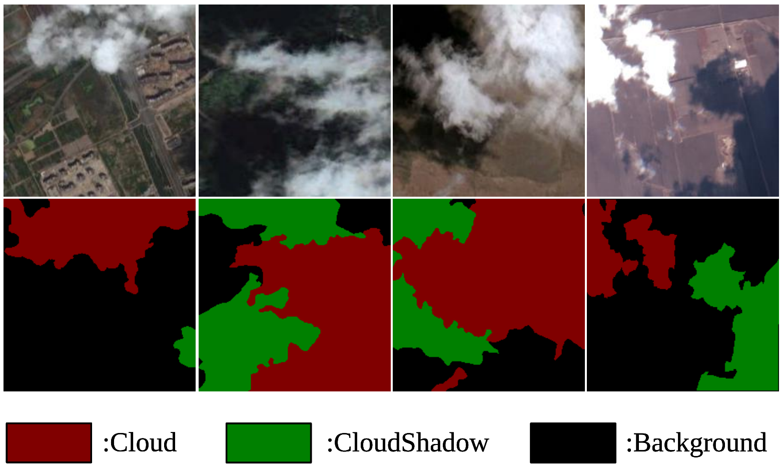

Figure 6. Cloud and Cloud Shadow Dataset.

3.1.2. HRC WHU Dataset

We employ the HRC WHU Dataset to test the proposed method's generalisation performance. The photos in this dataset were chosen from Google Earth by Wuhan University experts in the area of satellite picture comprehension, and the experts digitised the relevant reference cloud photographs [38]. The dataset is rich in context, including regions such as plants, water, snow, cities, and deserts. We cropped 150 original images of size 1280 × 720 to images of size 256 × 256 and filtered out the image data with only a single classification category. We performed data improvement techniques on the dataset to boost the model's generalisation capability as well as obtained approximately 6000 picture data. We chose 4/5 images as the training and validation set, as well as 1/5 images as the test set. These photos have two classes: cloud and background. Figure 7 shows part of the image data and the corresponding labels.

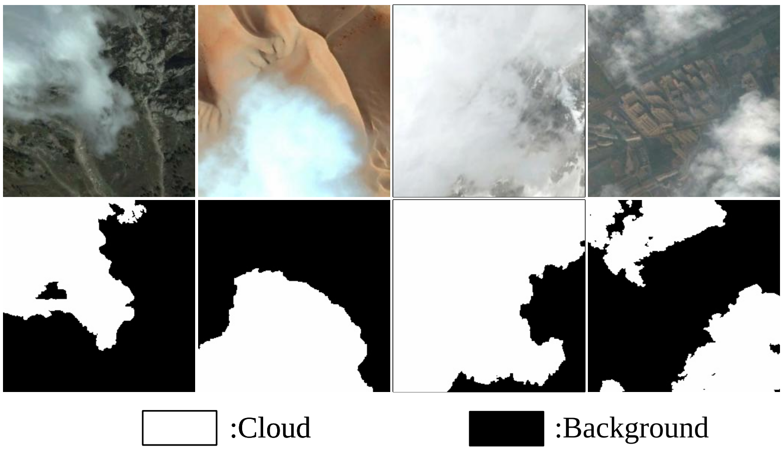

Figure 7. HRC WHU Dataset.

3.1.3. L8SPARCS Dataset

This dataset contains a 1000 × 1000 pixel subset of 80 Landsat 8 OLI/TRS scenes [39]. We split 80 scene images into 256 × 256 tiny images. Simultaneously, data augmentation operations were performed on them. Afterwards, we selected 80% images as the training and validation set and 20% images as the test set. Each picture has different labels, including categories such as cloud, cloud shadow, snow/ice, water, and backdrop. Figure 8 shows the images and corresponding labels.

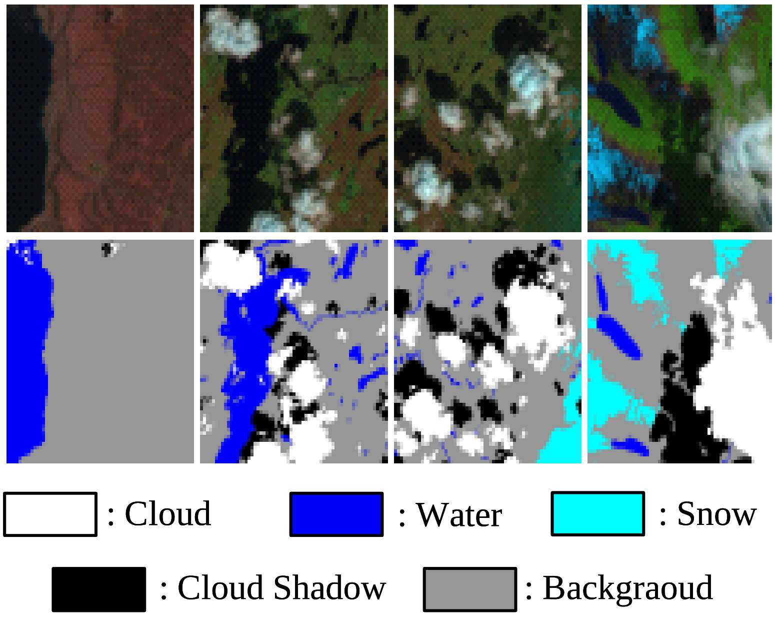

Figure 8. L8SPARCS Dataset.

3.2. Experimental Parameter Setting

All experiments were performed under the configuration environment of an Intel Core i7-11700K CPU@3.60 GHz and NVIDIA RTX3080ti. The operating system we used was Windows 10, and the framework was built on Pytorch (1.10.0). The optimiser used was adaptive matrix estimation (Adam) [40], the learning rate strategy adopted the "Poly" strategy [41], and the loss function used the Cross-Entropy Loss Function [42,43]. Their formulas are as follows:

$$lr = base\_lr \times (1 - \frac{epoch}{num\_epoch})^{power}, \tag{9}$$

where lr is the new learning rate, $base\_lr$ is the baseline learning rate, $epoch$ is the number of iterations, num_epoch is the maximum number of iterations, and $power$ controls the shape of the curve (usually, it is greater than 1). In our model, $power$ is set to 0.9 and num_epoch is set to 300. Due to the GPU card's restricted memory capacity, the batch size is limited to 16 during training.

$$Loss = -\sum_{i=1}^{n} y_i \log y_i', \tag{10}$$

where y_i is the label value and y_i' is the predicted value.

The loss function used in this article is BCEWithLogitsLoss. In this paper, precision (P), recall (R), F_1 score, pixel precision (PA), mean pixel precision (MPA), intersection over union (IoU), and average intersection over union ($MIoU$) are used as evaluation indicators [44]. The aforementioned assessment index has the following formula:

$$P = \frac{TP}{TP + FP}, \tag{11}$$

$$R = \frac{TP}{TP + FN}, \tag{12}$$

$$F_1 = 2 \times \frac{P \times R}{P + R}, \tag{13}$$

$$IoU = \frac{TP}{TP + FP + FN}, \tag{14}$$

$$PA = \frac{\sum_{i=0}^{k} \rho_{\mathrm{i},j}}{\sum_{i=0}^{k} \sum_{j=0}^{k} \rho_{i,j}}, \tag{15}$$

$$MPA = \frac{1}{k} \sum_{i=0}^{k} \frac{\rho_{i,j}}{\sum_{j=0}^{k} \rho_{i,j}}, \tag{16}$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{\rho_{i,j}}{\sum_{j=0}^{k} \rho_{i,j} + \sum_{j=0}^{k} \rho_{j,i} - \rho_{i,i}}, \tag{17}$$

where true positive (TP) represents the number of correctly predicted building (water) pixels, false positive (FP) represents the number of wrongly predicted building (water) pixels, true negative (TN) represents the number of correctly classified non-building (water) pixels, false negative (FN) represents the number of misclassified building (water) pixels, k denotes the category of object segmentation (excluding background), $\rho_{i,i}$ denotes the true class, and $\rho_{i,j}$ denotes the amount of pixels belonging to category i but predicted to be category j.

3.3. Ablation Experiments on Cloud and Cloud Shadow Dataset

This paper demonstrates the segmentation effect of each module of LPMSNet in the Cloud and Cloud Shadow Dataset hierarchically through ablation experiments. In this section, we adopt ResNet50 as the baseline model. In the benchmark model, we planned to use the improved ResNet50 for downsampling in the encoding stage, and then used the deconvolution module to upsample the downsampled output in the decoding stage to retrieve the original image. For proof of the viability of every part as well as the entire network, we sequentially add the suggested SFR, LAMA, and CSA to the model. Within this section, we evaluated using the MIoU evaluation index, and the specific data are provided in Table 2. At the same time, in order to verify the differences and advantages between the LAMA and CSA proposed in this paper and the existing modules, we added common multi-scale extraction modules and attention modules in the ablation experiments for comparison, such as ASPP, PPM, CBAM, SE, Non-Local Neural Network, and other modules. According to the table, all of the modules suggested in this work achieved the best performance.

Table 2. Ablation experiments of different modules of this model (bold numbers represent the optimal results, ↑ indicates the accuracies of the models proposed in this paper that rose after adding).

Method	MIoU (%)
ResNet50	92.23
ResNet50 + SFR	92.63 (0.40 ↑)
ResNet50 + SFR + ASPP	93.06
ResNet50 + SFR + PPM	93.21
ResNet50 + SFR + LAMA	93.55 (0.92 ↑)
Swin-T + SFR + LAMA + CSA	94.01
ResNet18 + SFR + LAMA + CSA	93.79
ResNet34 + SFR + LAMA + CSA	93.83
ResNet50 + SFR + LAMA + SE	93.79
ResNet50 + SFR + LAMA + CBAM	93.88
ResNet50 + SFR + LAMA + Non-Local	94.02
ResNet50 + SFR + LAMA + CSA	94.34 (0.79 ↑)

In order to test the performance analysis of this module under other backbone networks, we continued to add the experiment of replacing the basic backbone with Swin, ResNet18, and ResNet34 in the ablation experiment. In Table 2, we can see that the networks with Swin, ResNet18, and ResNet34 as the backbone have an accuracies of 94.01%, 93.79%, and 93.83% on the MIoU segmentation index, respectively. Obviously, the network models with the above backbones are all less accurate than those with ResNet50 as the backbone. The model based on ResNet50 is 0.33%, 0.55%, and 0.51% higher than the above models in terms of MIoU accuracy. Therefore, the network with ResNet50 as the backbone has a greater advantage in the segmentation task of Cloud and Cloud Shadow Datasets than other backbone networks.

The heatmap in Figure 9 intuitively shows the corresponding segmentation effect after adding each module in turn in the ablation experiment. The heat map visualises LPMSNet's attention to the two classification targets of cloud and cloud shade. Different colours in the heat map show the network's focus on the area. The darker the red area, the more attention the area receives from the model, and the segmentation effect will be relatively better. The yellow-green area has the next highest degree of attention, and the blue area represents the place with less attention [45–47]. In Figure 9, a total of two remote sensing pictures of cloud and cloud shade with different backgrounds are selected, and the test pictures are divided into two rows. The first line reflects the network's focus on the cloud, while the second represents the network's focus on the cloud shadow. We can intuitively see that as the modules proposed in this paper are added sequentially, the attention to cloud and cloud shade in the heat map is becoming higher and higher and the network is significantly less disturbed by background noise. As a result, each module provided in this research can improve the cloud as well as cloud shade segmentation impact. The cloud layer in this region of the test picture is relatively thin, as indicated by the oval box in the first row of Figure 9, and the network without the CSA is incomplete in capturing the cloud layer in this area, as indicated by the lighter colour of attention. This is due to the CSA's ability to dynamically extract correct long-distance dependencies of cloud and cloud shade, as well as to produce attention feature maps that are interested in cloud as well as cloud shade, hence avoiding the previously described missing detection of thin cloud, as shown in the second row of oval boxes in the second picture of Figure 9. The red area in the heat map becomes darker and darker as the parts presented in this research are added progressively, indicating that the network's attention to this area is gradually rising. The evidence presented above also demonstrates that CSA introduced in the article could assist the network in focusing on the segmentation target and effectively improving the segmentation effect. The module suggested in the article additionally provides an outstanding network segmentation impact.

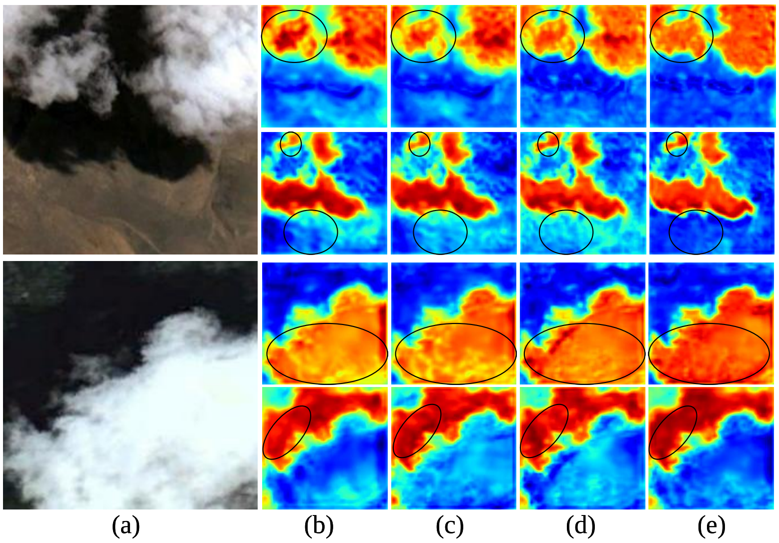

Figure 9. Heatmaps of different modules. (**a**) Test image; (**b**) method using backbone; (**c**) method using backbone and SFR; (**d**) method using backbone, SFR, and LAMA; (**e**) method using backbone, SFR, LAMA, and CSA.

In order to prove that CSA has greater advantages than other attention modules, we have added ablation experiments and heat map experiments for the Squeeze and Excitation Module (SE), the Convolutional Block Attention Module (CBA), and the Non-Local Neural Network (NLNN). From the Table 2, we can see that the models added with SE, CBAM, and NLNN have segmentation accuracies of 93.79%, 93.88%, and 94.02% on the Cloud and Cloud Shadow Dataset, respectively. It is obvious that the CSA proposed in this paper outperforms the above attention modules in segmentation metrics by 0.55%, 0.46%, and 0.32%, respectively. In addition to the advantages in segmentation accuracy, our CSA also outperforms the aforementioned attention modules on heatmap segmentation. In Figure 10, we selected two pictures. The first line of each picture is the segmentation effect of each model for cloud shadow, and the second line is the segmentation effect of cloud. We can see a small area of cloud shadow in the purple circle box in the first row of the first image. Only the heatmap of the model with CSA has the deepest red in this region. This shows that the CSA has a stronger ability to segment cloud shadow than other attention modules. In the second row of green boxes, the rest of the attention modules have lighter red in this area. This suggests that they pay less attention to the cloud. Only CSA has a darker red in this region, with a stronger focus on cloud. The above advantages are due to the fact that CSA can combine the Channel Attention Module and NLNN to adaptively adjust the focus of the network on the long-distance dependence of cloud and cloud shadows, improve the segmentation ability of the network, and avoid missed detection and false detection. Therefore, compared with the existing attention modules, CSA has a greater advantage in the segmentation task of the Cloud and Cloud Shadow Dataset.

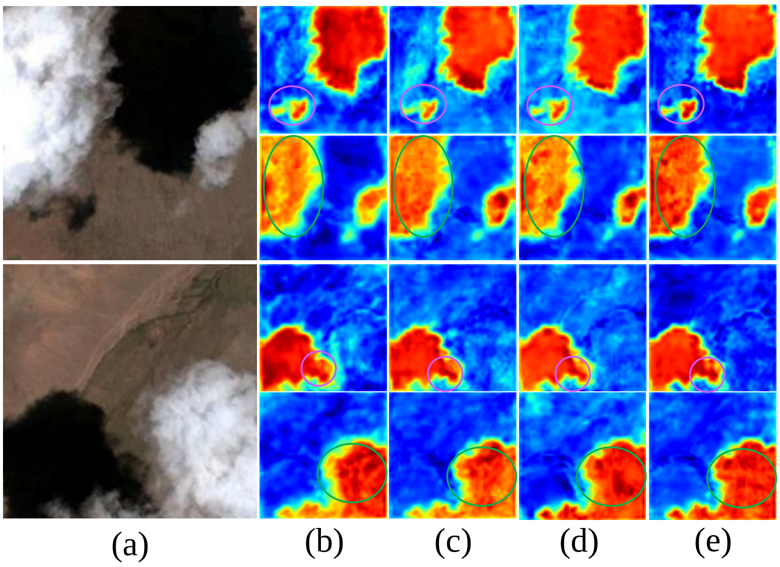

Figure 10. Heatmaps of different modules. (**a**) Test image; (**b**) method using SE; (**c**) method using CBAM; (**d**) method using Non-Local Neural Network; (**e**) method using CSA.

1. Ablation experiments of the SFR: The Scale Fusion Restoration Module (SFR) can combine the contextual information gathered during the downsampling step with the deep semantic info obtained via LAMA. This strategy allows the two types of information to guide and fuse each other, increases the network's ability to segment images, and perfects the feature info obtained with LPMSNet. Simultaneously, the end-stacked convolution could repair cloud as well as cloud shade edge details during the upsampling stage, boosting the segmentation effect. As shown in Table 2, SFR improves the model's MIoU value from 92.23% to 92.63%. This data demonstrates that the module can effectively fuse multi-scale information during the upsampling stage to boost the cloud and cloud shade segmentation impact.
2. Ablation Experiments of LAMA: The Location-Attention Multi-Scale Aggregation Module (LAMA) consists of a Location Attention Module and a Multi-Scale Aggregation Module. The Multi-Scale Aggregation Module efficiently recovers the feature map's multi-scale information by pooling convolutions of multiple scales, capturing the properties of cloud and cloud shading of different sizes and better segmenting their semantic categories. The Location Attention Module retrieves the feature map's positioning info via pooling convolution and generates attentional feature maps with the original image's horizontal and vertical positional encodings, respectively. The attention feature map focuses on the classification information of cloud and cloud shade, which can supplement the position encoding lost in multi-scale feature extraction. In addition, in order to highlight the superiority of LAMA compared with the existing multi-scale extraction modules, we made a comparison experiment with PPM and ASPP modules in Table 2. We can clearly find that the network with the LAMA module is 0.49% and 0.34% higher in evaluation indicators than the network with the ASPP and PPM modules. Therefore, compared with the existing multi-scale extraction models, the model proposed in this paper not only is stronger than them in terms of realisation functions but also has a greater advantage in terms of segmentation index accuracy than ASPP and PPM.
3. Ablation Experiment of CSA: The Channel Spatial Attention Network (CSA) can first obtain the long-distance dependence of the feature map through the improved NLNN. Design methods like residual connections can preserve the original feature information of the input. The Channel Attention Module gathers high-level characteristics via pooling operations and then feeds these high-level features into MLP to increase the model's generalisation capability and performance. CSA could make the

network dynamically focus on the long-distance dependencies between cloud and cloud shade, avoiding missed and false detection. At the same time, this module can share weights via convolution to enhance the module's robustness. Within the entire network structure, the deep CSA can further process the deep semantic information extracted with the SAMA to capture more abstract and global dependencies, while a shallower CSA in the upsampling locations can extract more detailed and local dependencies. It can be seen from Table 2 that the CSA can effectively improve the network's attention to cloud and cloud shade classification information. The overall evaluation index of the network with CSA increased by 0.79%, which also proves that CSA can significantly improve the segmentation effect of cloud and cloud shade.

3.4. Comparative Experiments on Different Datasets

3.4.1. Generalisation Experiment of Cloud and Cloud Shadow Dataset

We compared the suggested approach with the present excellent semantic segmentation model in the Experiment section to demonstrate the feasibility and benefits of the algorithm proposed in this paper. Each of the networks listed in this lab has its own benefits. FCN [48] gathers visual feature information using the full convolution network, that is a pixel-level classification network. UNet [49] uses stacked fully convolutional layers as the encoding structure and decoding structure and uses skip connections to fuse the information collected by the encoding structure to achieve effective classification. PSPNet proposes the Pyramid Pooling Module. It can obtain multi-scale information of images and complete effective image classification tasks. DeepLabV3Plus uses dilated convolutions to obtain contextual information in different regions about the image, which improves segmentation accuracy. ACFNet can rely on the attention category feature module, the network can adaptively segment different classification categories based on per-pixel calculations. To overcome the issue of excessive computation in the attention module and to enhance segmentation efficiency, CCNet introduces Criss-Cross Attention. OCRNet [50] uses object region representations to enhance pixel-wise representations of segmented objects. DFN [51] presents Channel Attention Block to integrate the characteristics of adjoining stages to accomplish successful segmentation. DDRNet [52] is made up of two depth branches connected by several bilateral fusions to achieve object segmentation. HRNet [53] designs a new cascaded pyramid structure for feature extraction to complete the semantic segmentation task. The above networks are some basic networks widely used in semantic segmentation tasks. GAFFNet, PANDA, CSDNet [54], MSPFANet, DBNet, etc. are specialised networks specially applied to cloud and cloud shade segmentation. The MSPFANet uses an improved pyramid pooling model to improve the multi-scale information extraction ability of the network and mine deep multi-scale semantic information. At the same time, the network also uses mutual fusion modules to guide information fusion at different levels. Lu proposed a Dual-Branch Network consisting of a convolutional network and a Transformer to extract the semantic and spatial details of an image, respectively. CSDNet combines a Multi-Scale Feature Fusion Module (MFF) and a Controllable Depth Supervision and Feature Fusion Structure (CDSFF) to achieve effective classification of cloud and cloud shade. CvT, PvT and Swin-T are improved Transformer networks for semantic segmentation tasks. CvT enhances the visual Transformer's efficacy and effectiveness by incorporating convolution into ViT to provide a combined effect. PvT improves the Transformer architecture and incorporates the pyramid structure. Downsampling the feature map reduces the size of the feature map, reducing the amount of calculation and thus boosting the ability of dense level prediction.

As indicated in Tables 3 and 4, we chose PA, MPA, and MIoU as overall assessment indicators for the Cloud and Cloud Shadow Dataset, and P, R, and F_1 as specific category evaluation indicators. Table 3 demonstrates that our technique performs best on PA, MPA, and MIoU scores, with 97.62%, 97.05%, and 94.38%, respectively. Table 4 displays performance of the above networks' segmentation metrics on a single class. It can be seen that the index scores of SLCANet on the Cloud and Cloud Shadow Dataset in terms of P, R and

F1 score are also better than those of existing networks. In the cloud-related segmentation tasks, the above segmentation indicators reached 97.17%, 96.83%, and 97.00%, respectively. In the cloud shade segmentation task, the above indicators reached 95.70%, 96.38%, and 96.04% respectively.

Table 3. Evaluation results of different models on Cloud and Cloud Shadow Dataset (bold numbers represent the optimal results).

Method	PA (%)	MPA (%)	MIoU (%)
UNet [49]	95.56	94.74	90.09
DeepLab V3Plus (ResNet 50) [11]	96.11	95.02	90.87
FCN-8s [48]	96.15	95.22	91.12
PSPNet (ResNet 50) [10]	96.16	95.07	91.24
HRNet [53]	96.46	95.68	91.66
DDRNet [52]	96.67	95.85	92.12
OCRNet (ResNet 101) [50]	96.54	95.82	92.14
DFN (ResNet101) [51]	96.59	95.86	92.45
CCNet (ResNet 50) [13]	96.69	95.95	92.51
ACFNet (ResNet 50) [12]	96.92	96.32	92.81
Swin-T [16]	95.69	94.51	90.08
CvT [15]	96.31	95.54	91.68
PvT [17]	96.91	96.11	92.95
GAFFNet (ResNet 18) [19]	96.11	95.07	91.04
PANDA [20]	96.15	95.37	91.32
CSDNet [54]	97.12	96.32	93.05
DBNet [1]	97.27	96.41	93.12
MSPFANet [23]	97.41	96.56	93.27
LPMSNet (ours)	97.62	97.05	94.38

Table 4. Single-class evaluation index results of different models on Cloud and Cloud Shadow Dataset (bold numbers represent the optimal results).

	Cloud			Cloud Shadow		
Method	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)
UNet [49]	96.32	92.71	94.47	91.68	94.07	92.89
DeepLab V3Plus (ResNet 50) [11]	94.68	95.45	95.03	92.92	93.63	93.31
FCN-8s [48]	95.32	95.15	95.24	93.21	93.91	93.58
PSPNet (ResNet 50) [10]	94.61	96.04	95.32	93.17	94.11	93.57
HRNet [53]	95.53	95.78	95.64	94.13	93.47	93.79
DDRNet [52]	96.33	95.45	95.88	94.51	93.91	94.21
OCRNet (ResNet 101) [50]	96.21	95.33	95.76	93.91	94.74	94.35
DFN (ResNet101) [51]	96.14	95.46	95.84	94.03	95.12	94.45
CCNet (ResNet 50) [13]	95.98	95.52	96.94	94.17	95.23	94.68
ACFNet (ResNet 50) [12]	96.21	96.08	96.16	94.34	95.32	94.84
Swin-T [16]	94.91	94.67	94.78	91.84	93.43	92.61
CvT [15]	95.74	95.56	95.63	92.45	93.87	93.94
PvT [17]	96.24	96.51	96.37	94.32	95.04	94.68
GAFFNet (ResNet 18) [19]	95.31	94.93	95.21	92.49	94.36	93.41
PANDA [20]	95.49	95.32	95.41	93.80	94.31	94.12
CSDNet [54]	96.12	96.57	96.31	94.14	95.95	95.03
DBNet [1]	96.79	96.56	96.62	94.89	94.41	94.58
MSPFANet [23]	96.87	96.63	96.56	94.21	94.51	94.35
LPMSNet (ours)	97.17	96.83	97.00	95.70	96.38	96.04

We selected 5 better-performing networks from 16 control networks and demonstrated their segmentation performance. As shown in Figures 11 and 12, we use black, red, and green to represent the labels corresponding to the background, cloud, and cloud shadow, respectively. We can illustrate the benefits of the algorithm suggested in the article by comparing the segmentation results of these five better-performing networks. Because of characteristics such as shape and thickness, the cloud layer puts the network's segmentation capacity to the test in the cloud as well as cloud shade classification task. Shallow and small-area cloud, for example, are readily disregarded by the network, leading to errors in detection. Simultaneously, as to the dataset's complicated backdrop, the network is easily disrupted by background noise, resulting in false detection. Figure 11 shows five networks with higher evaluation indicators and the segmentation effect of LPMSNet. We can clearly see that the network segmentation effect of ACFNet and PvT is poor, both have missing as well as erroneous detection, and the edge details of the cloud and cloud shade they segment are also rough. This area is characterised by relatively thin cloud, as seen in the oval boxes with the first and second rows of Figure 11, or the urban background in this area is similar to cloud shadow in terms of pixel values. It is obvious that, except for LPMSNet, a large number of false detections and missed detections have occurred in other networks. As shown in the third row of the figure, CSDNet ignores the shallow cloud layer at the edge of the region, and the segmentation effect on the cloud layer is poor. Although DBNet, ACFNet, MSPFANet and PvT have captured the cloud layer information in this area, their attention is not enough, and only a part of the cloud layer area is segmented. Only the LPMSNet proposed in this paper accurately captures the shallow cloud cover in this region and achieves effective segmentation. In addition, as shown by the blue circle in the first row of Figure 11, the edge of the cloud in this area is relatively rugged, and none of the five comparison networks can clearly identify the edge of the cloud in this area. Only LPMSNet is able to roughly segment out the rugged cloud edge details in this region. According to the segmentation impact illustrated in Figure 11, LPMSNet can effectively recognise the classification category info of cloud as well as cloud shade, as well as cleanly segment their rough edge features.

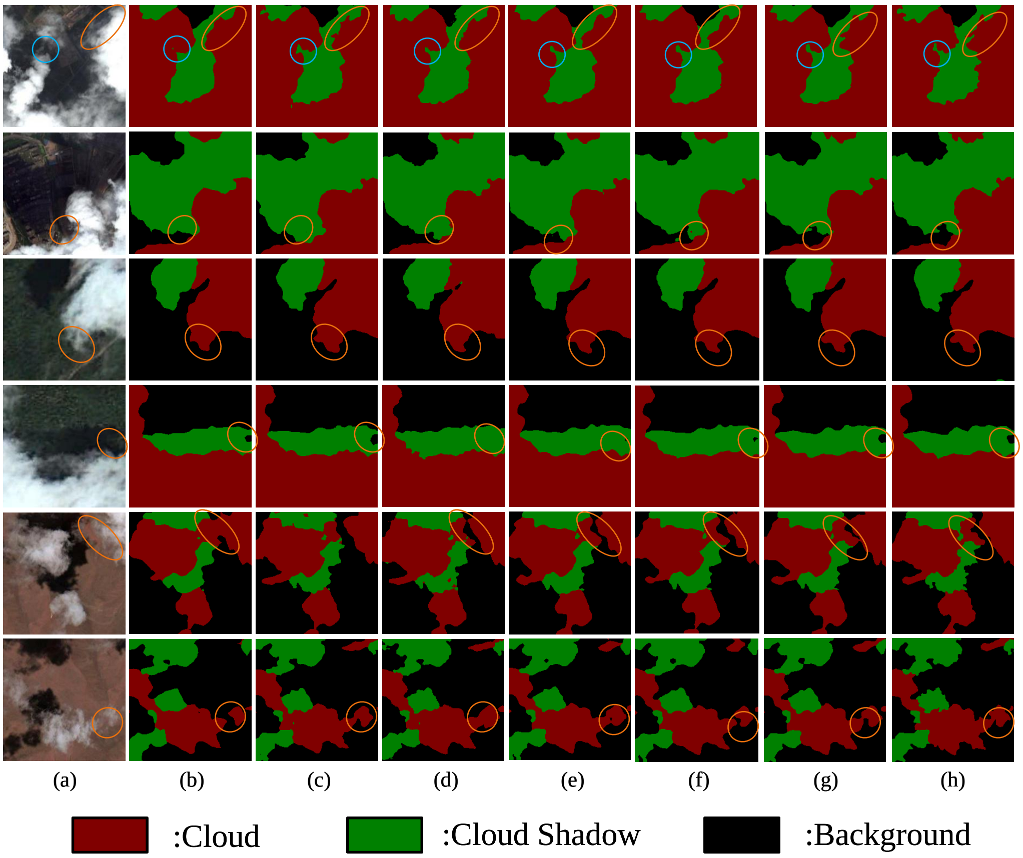

Figure 11. Prediction results for Cloud and Cloud Shadow Dataset. (**a**) Original image; (**b**) prediction map of ACFNet; (**c**) prediction map of PvT; (**d**) prediction map of CSDNet; (**e**) prediction map of DBNet; (**f**) prediction map of MSPFANet; (**g**) prediction map of LPMSNet; (**h**) corresponding labels. (The circles indicate areas that can be focused on in the image).

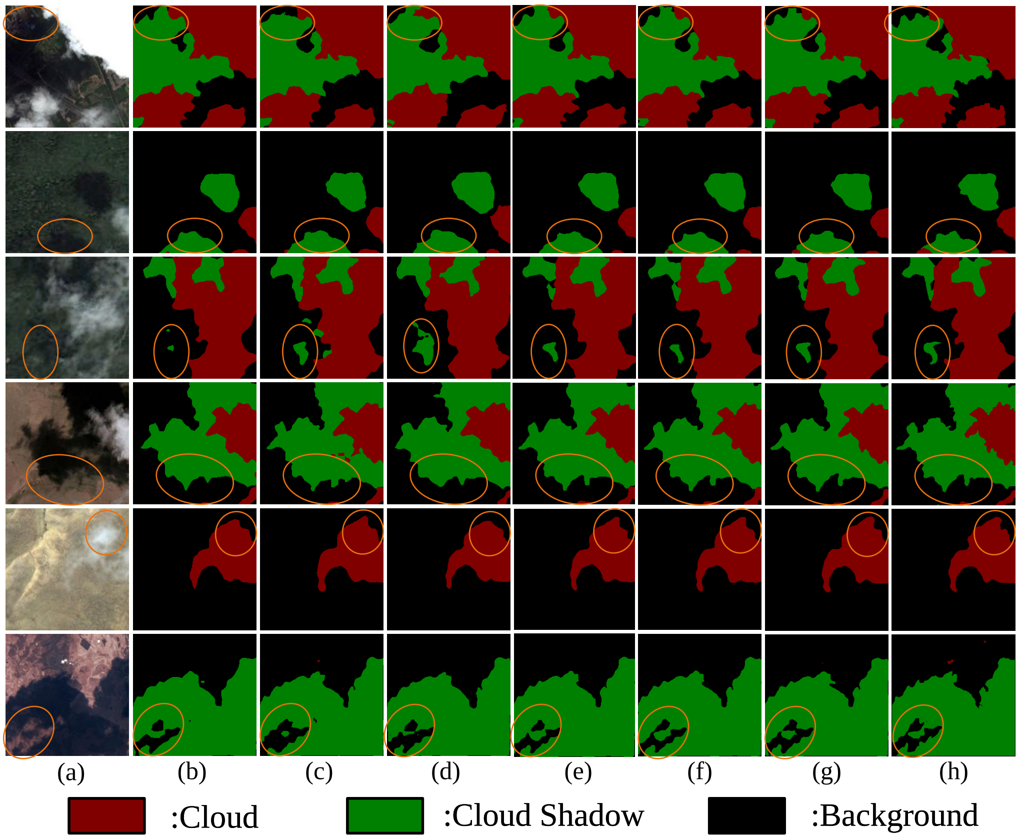

Figure 12. Prediction results for Cloud and Cloud Shadow Dataset. (**a**) Original image; (**b**) prediction map of ACFNet; (**c**) prediction map of PvT; (**d**) prediction map of CSDNet; (**e**) prediction map of DBNet; (**f**) prediction map of MSPFANet; (**g**) prediction map of LPMSNet; (**h**) corresponding labels. (The circles indicate areas that can be focused on in the image).

In order to further verify that LPMSNet still has a strong segmentation ability in complex backgrounds, we selected pictures with different backgrounds including cities, deserts, waters, and rocks to display the segmentation effects. These two groups of photos, as illustrated that the first as well as third rows of Figure 12, will cause the network to misclassify cloud shade as a background because their city and water backgrounds are similar in pixel value to cloud shade. We can clearly see from the oval box that ACFNet, PvT, CSDNet, DBNet and MSPFANet all have a large number of misclassification phenomena. They confuse cloud shade with background classification information, resulting in false detection. The edge information of cloud shading in this location is relatively complex, as seen in the second as well as sixth lines of Figure 12. ACFNet, PvT, and CSDNet all segment the edge of cloud shade in this area too simply and lack the rugged and complex characteristics of cloud shade edges. Furthermore, as illustrated within the picture's fourth row, the cloud cover in this area is relatively thin. Thin clouds are more difficult to distinguish, and the pixel values of the thin cloud and the desert rock background are quite similar in this area. These features will cause missing as well as erroneous detection in the network, putting the network's segmentation capability to the test. Obviously, only LPMSNet suggested in this research can effectively capture the category information of the cloud layer while clearly segmenting the complicated edge characteristics of the cloud layer for the segmentation impact of this region. LPMSNet first effectively collects image feature info using the improved ResNet50. LAMA feeds the above feature information into pooled convolutions of multiple scales to gather rich multi-scale cloud as well as cloud shade info. Using pooled convolutions, the location attention module within LAMA collects location-encoded information about cloud and cloud shade in both the vertical as well as horizontal axes of the image. This process may fully extract the position of the cloud and cloud shade pixels in the image to construct the attention feature map for the segmentation target that the network is interested in. Following that, the network embeds the previously mentioned attention feature map into multi-scale information to supplement the location information lost by LAMA in multi-scale sampling to prevent missing as well as erroneous detection. The improved NLNN is then used by the CSA to fully capture the long-range relationships of cloud as well as cloud shade. The reweighted feature info is embedded in the improved NLNN via the CA. The effective integration of the above two modules allows the CSA to dynamically modify the network's attention on the long-distance dependence about cloud as well as cloud shade. The deep CSA captures more abstract and global dependencies by further processing the deep semantic information extracted by the SAMA, while a shallower CSA in the upsampling locations can extract more detailed and local dependencies. Finally, SFR fuses the upper and lower information of different scales extracted in the upsampling stage with the semantic info obtained in LPMSNet's deep layer, allowing the two types of information to be fused and guided and improving the model's perception of cloud and cloud shade edges and classification info. The convolution module at the conclusion of the SFR could increase the network's edge segmentation details for cloud and cloud shade, making their border segmentation more obvious. As a result, our suggested technique outperforms the other control networks discussed above.

3.4.2. Generalisation Experiment of HRC-WHU Dataset

The HRC-WHU Dataset is a binary classification dataset. We use black and white to represent the labels corresponding to the background and cloud, respectively. In Figure 13, we selected data images containing complex backgrounds such as towns, snow, desert rocks, water, and grass. As indicated in Table 5, we used P, R, and IoU as assessment indicators to reflect the segmentation effect of the cloud dataset in this section. The table demonstrates that the model provided in this research performs well on these four indicators, with scores of 94.84%, 95.78%, 94.87%, and 90.51%, respectively.

Table 5. Evaluation results of different models on HRC-WHU Dataset (bold numbers represent the optimal results).

Method	**P (%)**	**R (%)**	**F_1 (%)**	**IoU (%)**
UNet [49]	90.86	94.88	92.88	86.68
DeepLab V3Plus (ResNet 50) [11]	92.04	94.69	93.38	87.54
FCN-8s [48]	92.51	93.48	92.96	86.97
PSPNet (ResNet 50) [10]	92.57	94.95	94.12	89.06
HRNet [53]	92.33	94.21	93.51	87.98
DDRNet [52]	93.11	95.27	94.17	88.98
OCRNet (ResNet 101) [50]	92.09	94.62	93.84	88.38
DFN (ResNet101) [51]	92.40	94.64	93.38	87.68
CCNet (ResNet 50) [13]	92.35	94.46	93.89	88.56
ACFNet (ResNet 50) [12]	93.78	94.99	93.78	89.65
Swin-T [16]	91.69	93.06	92.32	85.91
CvT [15]	91.94	94.94	93.45	87.76
PvT [17]	91.90	93.98	92.94	86.87
GAFFNet (ResNet 18) [19]	93.94	95.43	94.75	90.08
PANDA [20]	92.25	93.95	93.04	87.11
CSDNet [54]	92.14	94.67	93.38	87.65
DBNet [1]	93.79	95.08	94.71	90.02
MSPFANet [23]	93.88	95.32	94.78	90.11
LPMSNet (ours)	94.84	95.78	94.87	90.51

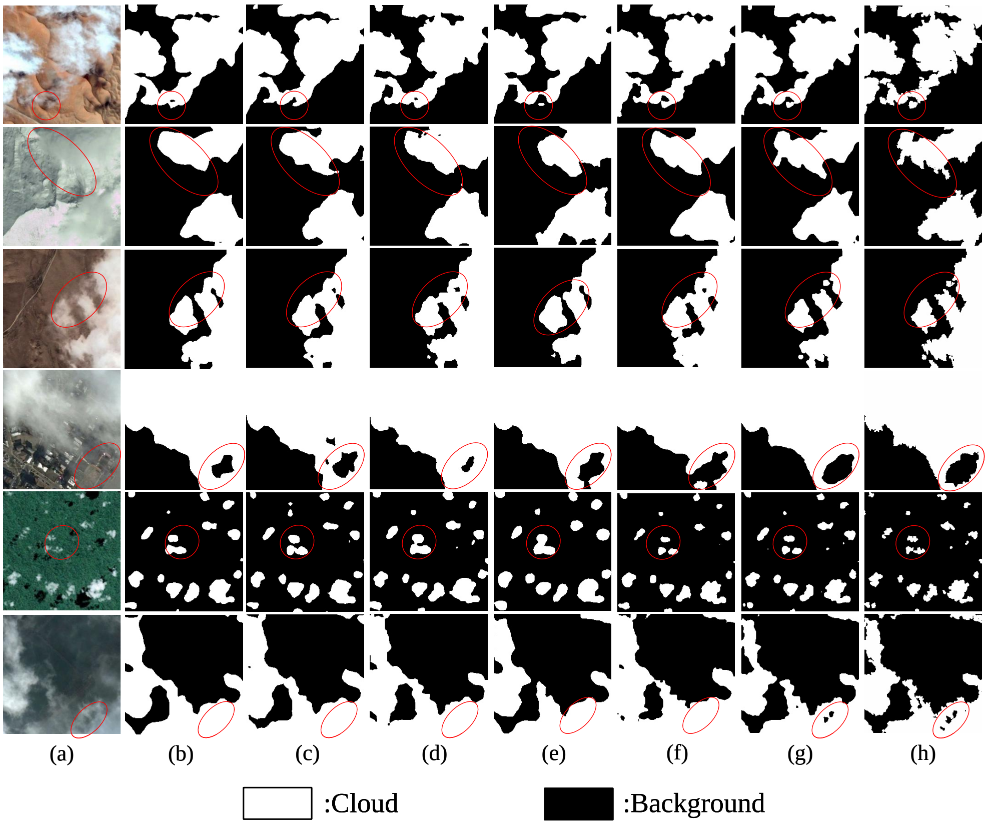

Figure 13. Prediction results for HRC-WHU Dataset. (**a**) Original image; (**b**) prediction map of PSPNet; (**c**) prediction map of ACFNet; (**d**) prediction map of GAFFNet; (**e**) prediction map of DBNet; (**f**) prediction map of MSPFANet; (**g**) prediction map of LPMSNet; (**h**) corresponding labels. (The circles indicate areas that can be focused on in the image).

Figure 13 clearly demonstrates that the segmentation impact of LPMSNet on this dataset has significant benefits over other networks. As shown in marked areas in the first and third lines of Figure 13, against the background of desert rocks, the cloud in the marked areas are relatively shallow. It can be clearly seen that only LPMSNet detects the approximate range of thin cloud in this area. Both PSPNet and ACFNet failed to identify the cloud layer in this area, and missed detection occurred. Although GAFFNet, DBNet, and MSPFANet segmented cloud, they mistakenly classified a large number of backgrounds into cloud, resulting in misclassification. This area has snow as the background, which is apparent in the second row of the picture, and the snow and cloud have comparable pixel values, making it simple to affect the network's accurate classification. In this area, LPMSNet accurately captures the category information of cloud and roughly segments the cloud edge information in this area. In contrast to other control networks, the edges of the cloud they segmented in this area are relatively rough, missing most of the information. Furthermore, as illustrated in the fourth row of the picture, GAFFNet incorrectly labels the urban background as a cloud layer, resulting in the misclassification issue. Although PSPNet, ACFNet, GAFFNet and DBNet did not have a wide range of misclassifications, they still missed some cloud information. The LAMA suggested in this paper gathers cloud as well as cloud shade feature information at multiple scales via a Multi-Scale Aggregation Module. These data can assist the network in adaptively identifying the category info of a cloud layer as well as cloud shade in various places, hence boosting the overall semantic understanding ability. The Location Attention Module can extract location information from images in the horizontal as well as vertical dimensions, build attention feature maps for cloud and cloud shade regions, and supplement the location encoding lost due to multi-scale downsampling. The CSA can adaptively modify the network's attention to cloud and cloud shade category information, as well as edge information, to improve the ability to acquire feature map long-distance relationships and to avoid missing as well as incorrect detection within the network. SFR can effectively combine contextual information from different layers in the encoding stage as well as multi-scale information extracted in deep networks. The two kinds of information guide and fuse each other to complete the decoding operation of the feature information in the upsampling. The end-stacked convolution modules can repair the edge information of cloud and cloud shade, making their edges clearer. As a result, LPMSNet suggested within the article not only has the highest accuracy in the comparison network but also has a good segmentation effect and a strong generalisation performance.

3.4.3. Generalisation Experiment of L8SPARCS Dataset

The L8SPARCS Dataset contains five classification categories, namely cloud, cloud shadow, snow/ice, water, and ground background. We use grey, black, white, dark blue, and light blue for the labels corresponding to background, cloud shadow, cloud, water, and snow/ice, respectively. We performed comparison experiments utilising the L8SPARCS Dataset to further validate the segmentation performance of this technique on multi-classification datasets, and the outcomes of the experiments are presented in Table 6. We select Class Pixel Accuracy as the single-class category segmentation index, as well as PA, MPA, and MIoU as the evaluation index of the overall segmentation effect. The network described within the article reach the maximum value in the three overall indicators of PA, MPA, and MIoU, which are 93.27%, 90.01%, and 81.60%, respectively, as demonstrated in Table 6. In terms of single-class category segmentation indicators, LPMSNet has the highest evaluation indicators for cloud, cloud shade, and background, which are 91.65%, 81.63%, and 95.87%, respectively. Although LPMSNet's scores on snow/ice and water are not the highest, the gap between it and the top indicators is not very large.

Table 6. Overall evaluation index and single classification evaluation index on L8SPARCS Dataset (bold numbers represent the optimal results).

	Class Pixel Accuracy					Overall Results		
Method	**Cloud (%)**	**Cloud Shadow (%)**	**Snow/Ice (%)**	**Water (%)**	**Land (%)**	**PA (%)**	**MPA (%)**	**MIoU (%)**
UNet [49]	88.82	75.81	93.51	93.76	93.41	91.67	89.09	78.82
DeepLab V3Plus (ResNet 50) [11]	87.24	68.94	91.85	90.07	93.26	90.41	86.25	75.81
FCN-8s [48]	89.12	71.57	91.81	94.27	93.14	91.23	87.94	76.47
PSPNet (ResNet 50) [10]	91.11	76.07	91.51	90.69	92.12	91.01	88.31	76.25
HRNet [53]	87.51	70.21	91.96	90.22	93.41	90.52	86.32	75.89
DDRNet [52]	88.80	70.03	88.73	89.84	92.21	90.13	85.92	74.58
OCRNet (ResNet 101) [50]	87.79	78.11	91.12	93.03	94.12	91.89	88.86	78.68
DFN (ResNet101) [51]	87.31	68.74	90.00	91.38	94.25	91.00	86.33	76.46
CCNet (ResNet 50) [13]	88.01	75.45	92.42	90.71	93.35	91.21	88.01	76.69
ACFNet (ResNet 50) [12]	90.47	78.49	93.46	93.25	92.93	91.81	89.69	77.78
Swin-T [16]	87.68	71.54	91.58	90.67	93.87	91.12	87.09	77.38
CvT [15]	85.67	73.85	91.58	85.21	94.55	90.96	86.15	76.91
PvT [17]	88.97	72.36	94.47	89.32	93.92	91.36	87.79	78.25
GAFFNet (ResNet 18) [19]	89.48	75.89	91.95	95.17	94.08	92.18	89.37	79.01
PANDA [20]	85.79	71.94	87.91	88.38	92.25	89.68	85.25	73.41
CSDNet [54]	89.12	76.43	91.41	90.23	95.68	92.81	88.56	80.67
DBNet [1]	90.24	78.56	93.78	91.14	95.75	93.01	89.54	81.21
MSPFANet [23]	89.07	73.18	91.23	88.91	93.56	90.89	87.34	78.02
LPMSNet (ours)	91.65	81.63	94.02	95.04	95.87	93.27	90.01	81.60

As illustrated in the third row about Figure 14, the cloud in this location are very shallow, and the network can easily misclassify thin cloud as cloud shade on the ground. OCRNet, CSDNet, and GAFFNet all have the above-mentioned missed detection phenomenon, they ignore the shallow cloud cover in the area and directly classify the area as cloud shade. LPMSNet detected the shallow cloud information well and segmented it effectively. The area of the cloud layer in this area is small, as seen in the second row about the image, and UNet, OCRNet, CSDNet, and GAFFNet all missed these small-area cloud layers, resulting in a missed detection phenomenon. Because of the LAMA, LPMSNet may simultaneously concentrate on the category info of big or small portions of cloud and cloud shade in the above-mentioned areas to achieve successful segmentation and to avoid missed detection. The edge of the cloud shade is relatively rough, as seen in the fourth row about the picture, which is a test of the network's capacity to collect the segmented target's edge information. The edges of cloud shade segmented by OCRNet, GAFFNet, and CSDNet are relatively rough, and the details of edges and corners are ignored, and even misclassification of large areas occurs. Only LPMSNet accurately detects the edge information of cloud shade in this area, while avoiding large-scale false detection. The advantage of the technique described within the research is that LAMA can mine sufficient multi-scale image information and successfully recognise the category information of cloud as well as cloud shade on multiple scales. Simultaneously, the location code generated by this module can supplement missing location information in multi-scale data in order to increase the network's picture segmentation accuracy. With an attention module, the CSA may dynamically apply weights to the network's region of interest, precisely identify the present image's long-distance dependencies, as well as better optimise the network's recognition effect on cloud and cloud shade. The SFR successfully realises mutual guidance between the above information, increases the network's decoding ability, and avoids missing as well as erroneous detection by fusing context info from multiple layers as well as deep semantic info. The stacked convolution at the end can improve the network's repair of cloud and cloud shade edge information, making their edge details clearer. The above segmentation impact demonstrates that the method we suggest can effectively eliminate the phenomenon of missing as well as erroneous detection, as well as efficiently extract the target object's edge information. LPMSNet has the best segmentation index and segmentation effect on the L8SPARCS Dataset. It performs better on multi-classification dataset and has strong generalisation performance.

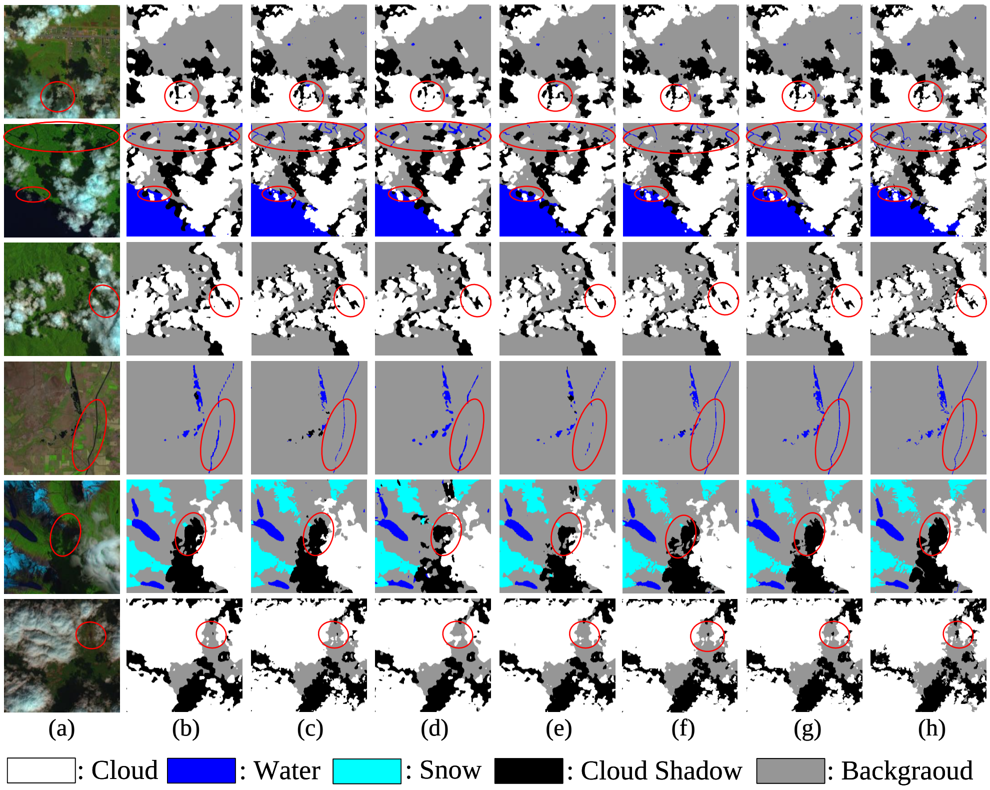

Figure 14. Prediction results for L8SPARCS Dataset. (**a**) Original image; (**b**) prediction map of OCRNet; (**c**) prediction map of UNet; (**d**) prediction map of GAFFNet; (**e**) prediction map of CSDNet; (**f**) prediction map of DBNet; (**g**) prediction map of LPMSNet; (**h**) corresponding labels. (The circles indicate areas that can be focused on in the image).

4. Conclusions

4.1. Limitations and Future Research Directions

Nevertheless, the technique requires additional refinement in cloud and cloud shade identification efforts. The amount of parameters for the LPMSNet as well as computational complexity provided in this paper could be constantly optimised in the future to minimise network training time while maintaining the precision of segmentation.

4.2. Summary

We present a Location Pooling Multi-Scale Network within this article to accomplish end-to-end cloud and cloud shade identification using multispectral remote sensing photos. The approach first collects feature info from multiple layers about the image using the upgraded ResNet50 and then uses the Location Attention Multi-Scale Aggregation Module to additionally collect multi-scale info from cloud as well as cloud shade in the network's deep layer. Simultaneously, its internal Location Attention Module embeds the obtained picture's location code into multi-scale info to supplement the lost position info of cloud as well as cloud shade. Then, the Channel Spatial Attention Module processes the multi-scale information extracted via the Location Attention Multi-Scale Aggregation Module and the semantic information obtained via the Scale Fusion Restoration Module in the upsampling stage in the deep and shallow layers of the network, respectively. It adaptively adjusts the focus of the network on the long-distance dependence of cloud as well as cloud shade, improves network segmentation capabilities, and avoids missed and false detection. At last, the Scale Fusion Restoration Module fuses the contextual info taken from the deep layer of the network with the contextual info obtained during the upsampling stage to guide the network through the decoding operation. The end-stacked convolutional modules can focus on the edge information of cloud as well as cloud shade, and refine and repair their edge details. LPMSNet can successfully realise the identification as well as segmentation tasks about cloud as well as cloud shade and can obtain a sharper segmentation effect map of cloud and cloud shade remote sensing images by performing the aforementioned procedures.

Author Contributions: Conceptualisation, X.D. and M.X.; methodology, M.X. and L.W.; software, X.D. and K.C.; validation, L.W. and H.L.; formal analysis, M.X. and K.C.; investigation, X.D.; resources, M.X. and L.W.; data curation, X.D.; writing—original draft preparation, X.D.; writing—review and editing, M.X.; visualisation, X.D.; supervision, M.X.; project administration, M.X.; funding acquisition, M.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported in part by the National Natural Science Foundation of China (42075130).

Data Availability Statement: The data and the code of this study are available from the corresponding author upon request.

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

References

1. Lu, C.; Xia, M.; Qian, M.; Chen, B. Dual-branch network for cloud and cloud shadow segmentation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5410012. [CrossRef]
2. Zhang, E.; Hu, K.; Xia, M.; Weng, L.; Lin, H. Multilevel feature context semantic fusion network for cloud and cloud shadow segmentation. *J. Appl. Remote Sens.* **2022**, *16*, 046503. [CrossRef]
3. Chen, K.; Xia, M.; Lin, H.; Qian, M. Multi-scale Attention Feature Aggregation Network for Cloud and Cloud Shadow Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5612216. [CrossRef]
4. Qu, Y.; Xia, M.; Zhang, Y. Strip pooling channel spatial attention network for the segmentation of cloud and cloud shadow. *Comput. Geosci.* **2021**, *157*, 104940. [CrossRef]
5. Hu, K.; Zhang, E.; Xia, M.; Weng, L.; Lin, H. MCANet: A Multi-Branch Network for Cloud/Snow Segmentation in High-Resolution Remote Sensing Images. *Remote Sens.* **2023**, *15*, 1055. [CrossRef]
6. Wang, D.; Weng, L.; Xia, M.; Lin, H. MBCNet: Multi-Branch Collaborative Change-Detection Network Based on Siamese Structure. *Remote Sens.* **2023**, *15*, 2237. [CrossRef]

7. Chen, B.; Xia, M.; Qian, M.; Huang, J. MANet: A multi-level aggregation network for semantic segmentation of high-resolution remote sensing images. *Int. J. Remote Sens.* **2022**, *43*, 5874–5894. [CrossRef]
8. Song, L.; Xia, M.; Weng, L.; Lin, H.; Qian, M.; Chen, B. Axial Cross Attention Meets CNN: Bibranch Fusion Network for Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 32–43. [CrossRef]
9. Gao, J.; Weng, L.; Xia, M.; Lin, H. MLNet: Multichannel feature fusion lozenge network for land segmentation. *J. Appl. Remote Sens.* **2022**, *16*, 016513. [CrossRef]
10. 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.
11. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
12. Zhang, F.; Chen, Y.; Li, Z.; Hong, Z.; Liu, J.; Ma, F.; Han, J.; Ding, E. Acfnet: Attentional class feature network for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6798–6807.
13. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. Ccnet: Criss-cross attention for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 603–612.
14. 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, 18–22 June 2018; pp. 7794–7803.
15. Wu, H.; Xiao, B.; Codella, N.; Liu, M.; Dai, X.; Yuan, L.; Zhang, L. Cvt: Introducing convolutions to vision transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 22–31.
16. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
17. Wang, W.; Xie, E.; Li, X.; Fan, D.P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 568–578.
18. Ji, H.; Xia, M.; Zhang, D.; Lin, H. Multi-Supervised Feature Fusion Attention Network for Clouds and Shadows Detection. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 247. [CrossRef]
19. Xia, M.; Wang, T.; Zhang, Y.; Liu, J.; Xu, Y. Cloud/shadow segmentation based on global attention feature fusion residual network for remote sensing imagery. *Int. J. Remote Sens.* **2021**, *42*, 2022–2045. [CrossRef]
20. Xia, M.; Qu, Y.; Lin, H. PANDA: Parallel asymmetric network with double attention for cloud and its shadow detection. *J. Appl. Remote Sens.* **2021**, *15*, 046512. [CrossRef]
21. Miao, S.; Xia, M.; Qian, M.; Zhang, Y.; Liu, J.; Lin, H. Cloud/shadow segmentation based on multi-level feature enhanced network for remote sensing imagery. *Int. J. Remote Sens.* **2022**, *43*, 5940–5960. [CrossRef]
22. Hu, K.; Zhang, D.; Xia, M. CDUNet: Cloud detection UNet for remote sensing imagery. *Remote Sens.* **2021**, *13*, 4533. [CrossRef]
23. Lu, C.; Xia, M.; Lin, H. Multi-scale strip pooling feature aggregation network for cloud and cloud shadow segmentation. *Neural Comput. Appl.* **2022**, *34*, 6149–6162. [CrossRef]
24. Ma, Z.; Xia, M.; Weng, L.; Lin, H. Local Feature Search Network for Building and Water Segmentation of Remote Sensing Image. *Sustainability* **2023**, *15*, 3034. [CrossRef]
25. Chen, J.; Xia, M.; Wang, D.; Lin, H. Double Branch Parallel Network for Segmentation of Buildings and Waters in Remote Sensing Images. *Remote Sens.* **2023**, *15*, 1536. [CrossRef]
26. Zhang, C.; Weng, L.; Ding, L.; Xia, M.; Lin, H. CRSNet: Cloud and Cloud Shadow Refinement Segmentation Networks for Remote Sensing Imagery. *Remote Sens.* **2023**, *15*, 1664. [CrossRef]
27. Chen, B.; Xia, M.; Huang, J. MFANet: A Multi-Level Feature Aggregation Network for Semantic Segmentation of Land Cover. *Remote Sens.* **2021**, *13*, 731. [CrossRef]
28. 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.
29. Weng, L.; Pang, K.; Xia, M.; Lin, H.; Qian, M.; Zhu, C. Sgformer: A Local and Global Features Coupling Network for Semantic Segmentation of Land Cover. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 6812–6824. [CrossRef]
30. Hu, K.; Wang, T.; Shen, C.; Weng, C.; Zhou, F.; Xia, M.; Weng, L. Overview of Underwater 3D Reconstruction Technology Based on Optical Images. *J. Mar. Sci. Eng.* **2023**, *11*, 949. [CrossRef]
31. Bello, I.; Zoph, B.; Vaswani, A.; Shlens, J.; Le, Q.V. Attention augmented convolutional networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3286–3295.
32. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Vedaldi, A. Gather-excite: Exploiting feature context in convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; Volume 31.
33. Dai, X.; Xia, M.; Weng, L.; Hu, K.; Lin, H.; Qian, M. Multi-Scale Location Attention Network for Building and Water Segmentation of Remote Sensing Image. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5609519. [CrossRef]
34. Li, X.; Xu, F.; Liu, F.; Lyu, X.; Tong, Y.; Xu, Z.; Zhou, J. A Synergistical Attention Model for Semantic Segmentation of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5400916. [CrossRef]

35. 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.
36. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective kernel networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 510–519.
37. Zhang, S.; Weng, L. STPGTN—A Multi-Branch Parameters Identification Method Considering Spatial Constraints and Transient Measurement Data. *Comput. Model. Eng. Sci.* **2023**, *136*, 2635–2654. [CrossRef]
38. Li, Z.; Shen, H.; Cheng, Q.; Liu, Y.; You, S.; He, Z. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 197–212. [CrossRef]
39. Hughes, M.J.; Hayes, D.J. Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. *Remote Sens.* **2014**, *6*, 4907–4926. [CrossRef]
40. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
41. Leng, Z.; Tan, M.; Liu, C.; Cubuk, E.D.; Shi, X.; Cheng, S.; Anguelov, D. Polyloss: A polynomial expansion perspective of classification loss functions. *arXiv* **2022**, arXiv:2204.12511.
42. Li, X.; Xu, F.; Lyu, X.; Gao, H.; Tong, Y.; Cai, S.; Li, S.; Liu, D. Dual attention deep fusion semantic segmentation networks of large-scale satellite remote-sensing images. *Int. J. Remote Sens.* **2021**, *42*, 3583–3610. [CrossRef]
43. Li, X.; Xu, F.; Xia, R.; Li, T.; Chen, Z.; Wang, X.; Xu, Z.; Lyu, X. Encoding contextual information by interlacing transformer and convolution for remote sensing imagery semantic segmentation. *Remote Sens.* **2022**, *14*, 4065. [CrossRef]
44. Elmezain, M.; Malki, A.; Gad, I.; Atlam, E.S. Hybrid Deep Learning Model–Based Prediction of Images Related to Cyberbullying. *Int. J. Appl. Math. Comput. Sci.* **2022**, *32*, 323–334.
45. Ma, C.; Weng, L.; Xia, M.; Lin, H.; Qian, M.; Zhang, Y. Dual-branch network for change detection of remote sensing image. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106324. [CrossRef]
46. Yin, H.; Weng, L.; Li, Y.; Xia, M.; Hu, K.; Lin, H.; Qian, M. Attention-guided siamese networks for change detection in high resolution remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *117*, 103206. [CrossRef]
47. Li, X.; Xu, F.; Liu, F.; Xia, R.; Tong, Y.; Li, L.; Xu, Z.; Lyu, X. Hybridizing Euclidean and Hyperbolic Similarities for Attentively Refining Representations in Semantic Segmentation of Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5003605. [CrossRef]
48. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
49. 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 ; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
50. Yuan, Y.; Chen, X.; Wang, J. Object-contextual representations for semantic segmentation. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020 ; Springer: Berlin/Heidelberg, Germany, 2020; pp. 173–190.
51. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Learning a discriminative feature network for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1857–1866.
52. Hong, Y.; Pan, H.; Sun, W.; Jia, Y. Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes. *arXiv* **2021**, arXiv:2101.06085.
53. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5693–5703.
54. Zhang, G.; Gao, X.; Yang, Y.; Wang, M.; Ran, S. Controllably deep supervision and multi-scale feature fusion network for cloud and snow detection based on medium-and high-resolution imagery dataset. *Remote Sens.* **2021**, *13*, 4805. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.