remote sensing

MDPI

Article

An Effective Cloud Detection Method for Gaofen-5 Images via Deep Learning

Junchuan Yu [1,*], Yichuan Li [1], Xiangxiang Zheng [1], Yufeng Zhong [2] and Peng He [1]

[1] Department of Research and Development, China Aero Geophysical Survey and Remote Sensing Center for Natural Resources, Beijing 100083, China; liyichuan@agrs.cn (Y.L.); zhengxiang@agrs.cn (X.Z.); hepeng@agrs.cn (P.H.)

[2] School of Earth Science and Surveying Engineering, University of Mining & Technology, Beijing 100101, China; zyf@cumtb.edu.cn

* Correspondence: jcyu@agrs.cn

Received: 19 June 2020; Accepted: 29 June 2020; Published: 1 July 2020

Abstract: Recent developments in hyperspectral satellites have dramatically promoted the wide application of large-scale quantitative remote sensing. As an essential part of preprocessing, cloud detection is of great significance for subsequent quantitative analysis. For Gaofen-5 (GF-5) data producers, the daily cloud detection of hundreds of scenes is a challenging task. Traditional cloud detection methods cannot meet the strict demands of large-scale data production, especially for GF-5 satellites, which have massive data volumes. Deep learning technology, however, is able to perform cloud detection efficiently for massive repositories of satellite data and can even dramatically speed up processing by utilizing thumbnails. Inspired by the outstanding learning capability of convolutional neural networks (CNNs) for feature extraction, we propose a new dual-branch CNN architecture for cloud segmentation for GF-5 preview RGB images, termed a multiscale fusion gated network (MFGNet), which introduces pyramid pooling attention and spatial attention to extract both shallow and deep information. In addition, a new gated multilevel feature fusion module is also employed to fuse features at different depths and scales to generate pixelwise cloud segmentation results. The proposed model is extensively trained on hundreds of globally distributed GF-5 satellite images and compared with current mainstream CNN-based detection networks. The experimental results indicate that our proposed method has a higher F1 score (0.94) and fewer parameters (7.83 M) than the compared methods.

Keywords: Gaofen-5; deep learning; cloud detection; big data; MFGNet; quality assessment

1. Introduction

Gaofen-5 (GF-5) is the fifth flight unit of the China High-Resolution Earth Observation System (CHEOS) [1], which was successfully launched in May 2018. It carries two land observation payloads, including a visible and shortwave infrared hyperspectral camera, a multispectral imager, and four atmospheric observation payloads, including a greenhouse gas detector, multiangle polarization detector, differential absorption spectrometer for atmospheric trace gas, and atmospheric environment infrared sensor [2]. GF-5 imagery can be widely used in environmental monitoring, geological mapping, urban heat island monitoring, thermal effluent monitoring, and other fields, by virtue of its wide spectrum range and high spatial and spectral resolution characteristics. GF-5 imagery is of great relevance for global-scale quantitative remote sensing applications [3]. However, the annual mean global cloud cover is approximately 66% [4], which brings significant challenges to the large-scale application of remote sensing data. According to our statistics, the daily peak value of the data obtained by the GF-5 land observation payloads is above 300 scenes. For data providers, it is crucial

to quickly, efficiently, and accurately detect clouds to ensure the daily quality of subsequent remote sensing products.

Over the past two decades, many cloud detection methods have been developed and they can be roughly classified into three categories: threshold-based, traditional machine learning (TML)-based, and convolutional neural network (CNN)-based methods. As a traditional and efficient cloud recognition algorithm, threshold-based algorithms are often used in cloud detection for multispectral and hyperspectral images [5–7]. The basic principle of this type of method is to use the reflectance difference between the cloud and other objects in the visible–shortwave infrared spectral range, and manually design feature extraction rules to identify and segment clouds; the method is used in the International Satellite Cloud Climatology Project (ISCCP) [8], the AVHRR Processing Scheme over Clouds, Land, and Ocean (APOLLO) [9], and Clouds from the Advanced Very High-Resolution Radiometer (CLAVR) [10]. As a well-known threshold-based algorithm, the Fmask [5] algorithm and its improved versions [11] have made significant contributions to the cloud detection of Landsat satellite imagery. At present, the series of algorithms has been updated to version 4.0 [12]. In addition, a series of threshold-based algorithms that combine new ideas such as multitemporal information [13–15] and dynamic thresholds [16,17] has been proposed and widely used in cloud detection applications for particular types of satellite images, including Landsat [10], Moderate Resolution Imaging Spectroradiometer (MODIS) [18], Sentinel [19], Han Jing-1 [20], GF-5 multispectral images [21], etc. However, this type of method still has some shortcomings. First, it is not suitable for some high-resolution satellite data with only four optical bands, such as Chinese Ziyuan-3 and Gaofen-2 data. Second, the calculation process is mainly based on pixels, which can easily lead to a "salt-and-pepper" (SAP) effect [22,23]. Third, the threshold-based method relies on expert knowledge, so it is difficult to determine a proper threshold, especially in dealing with data for complex surfaces [24,25]. Fourth, as mentioned earlier, the hand-crafted approach is often designed for specific payloads, and the applicability of the algorithm is limited.

With the rise of machine learning technologies, traditional machine learning algorithms such as artificial neural networks (ANN) [26], support vector machines (SVM) [27], and random forest (RF) [28] have been proposed for cloud classification issues [29,30]. The most significant improvement of the TML-based methods over the threshold-based methods is that they eliminate the problem of setting thresholds. Although they still rely on hand-crafted features, the choice of features is more flexible [31,32]. However, as with threshold-based algorithms, the classification-based method has challenges in overcoming the SAP effect, and there is still room for improvement in fusing spatial and spectral information.

Recent advances in deep learning, especially deep CNN, have led to a remarkable breakthrough in remote sensing applications, including image fusion, image registration, scene classification, object detection, land use and land cover classification, segmentation, and object-based image analysis [33]. Unlike TML-based methods, which first perform the extraction of hand-crafted features and then apply shallow classification techniques, CNN-based methods automatically extract important features [34]. In recent years, scholars have made many advances in research on CNN-based cloud recognition methods. Table 1 lists selected CNN-based cloud detection methods that have become prominent in recent years, which can be divided into two major categories: objectwise and pixelwise. The objectwise method needs to apply superpixel segmentation, such as simple linear iterative clustering (SLIC), to the target and then uses CNN to classify the superpixel results. Although this kind of method overcomes the SAP effect to some extent, recognition accuracy is still severely affected by the initial superpixel result [22]. To minimize the error caused by superpixel segmentation, some postprocessing steps have been developed to refine the final mask, such as conditional random fields (CRFs) [35] and Markov random fields (MRFs) [36]; these steps are very time-consuming [22,37,38].

Table 1. Summary of the main convolutional neural network (CNN)-based methods for cloud detection in recent years.

Authors	Date	Pixel-/Objectwise	Method	CNN Structure
Sorour Mohajerani et al. [39]	2020	Pixel	Cloud-Net+ [39]	U-shape + Branches
Zhiwei Li et al. [31]	2019	Pixel	MSCFF [31]	U-shape + Branches
Jingyu Yang et al. [40]	2019	Pixel	CDnet [40]	Multi-scale
Jacob Hobroe Jeppesen et al. [41]	2019	Pixel	RS-Net [41]	U-shape
Dengfeng Chai et al. [38]	2019	Pixel	Modified SegNet [38]	U-shape
Zhengfeng Shao et al. [25]	2019	Pixel	MF-CNN [25]	Multi-branch
Yongjie Zhan et al. [22]	2019	Pixel	FCN [22]	Linear stack + Branches
Johannes Dronner et al. [42]	2018	Pixel	CS-CNN [42]	U-shape
Han Liu et al. [43]	2018	Object	SLIC+HFCNN+ Deep Forest [43]	Linear stack
Giorgio Morales et al. [44]	2018	Object	ASLIC+CNN [44]	Linear stack
Lei Wang et al. [23]	2018	Object	ASLIC+CNN [23]	Dual-branch
Zhengsheng Guo et al. [45]	2018	Object	ASLIC+CNN [45]	Dual-branch
Yue Zi et al. [46]	2018	Object	SLIC+PCANet [46]	Dual-branch
Yang Chen et al. [47]	2018	Object	SLIC+MCNNNs [47]	Multi-branch
Fengying Xie et al. [48]	2017	Object	SLIC+Multi-level CNN [48]	Dual-branch

What is interesting in Table 1 is the rapid increase of cloud detection based on pixelwise CNN methods. This type of approach can perform feature extraction and classification at the same time to implement end-to-end segmentation. Compared with the methods mentioned above, pixelwise CNN methods have apparent advantages: first, they can integrate spatial information and spectral information. Second, hand-crafted features and sophisticated remote sensing preprocessing steps are not needed; third, given sufficient training samples, they have higher accuracy and stronger generalization abilities. However, the pixelwise CNN method still has room for improvement in cloud detection. According to our statistics, the U-shape [38,41,42] and the linear stack structures [22], inspired by U-Net [49], SegNet [50], and VGG [51], are the two mainstream architectures for cloud segmentation. It is well known that cloud images contain different types of representations: high-level semantic information and low-level information such as color, shape, and location information. As such, these architectures, with a single processing pipeline that relies on multistage cascaded CNNs, may lead to the loss of spatial information and may result in inaccurate boundary definitions [52–54]. Some meaningful practices relating to the fusion of features at different depths and scales to expand the receptive field of the network have also been reported [25,31,39,40]. However, further research is needed, especially on ways to reduce the loss of spatial information and how to capture and fuse the relevant and meaningful multi-scale contextual information instead of simple concatenation.

Benefiting from deep learning technology, completing a quality assessment of massive satellite data requires fewer resources and even thumbnails can be used to implement accurate cloud detection [40]. Inspired by the excellent CNN architectures of bilateral segmentation network (BiSeNet) [54], pyramid scene parsing network (PSPNet) [55], and squeeze-and-excitation network (SENet) [56], in this paper, we propose a new cloud detection method, a multiscale fusion gated network, using a dual-branch CNN architecture for cloud detection of GF-5 preview RGB images. First, we design a new lightweight backbone network combining the advantages of SENet [56] and ResNeXt [57]. Then, we introduce two attention modules: one is a spatial pyramid pooling attention (SPPA) module based on the channel attention mechanism to extract multiscale semantic features; the other is a low-level feature spatial attention (LFSA) module with a spatialwise attention mechanism for extracting beneficial shallow features. Finally, a gated multilevel feature fusion (GMFF) module is employed to deeply fuse features at different depths and scales to generate the pixelwise cloud segmentation result. The remainder of this paper is organized as follows. The proposed method is described in Section 2. The data source and experiment settings are described in Section 3. Experiments with evaluations

and comparisons are presented in Section 4, and the conclusion, along with a discussion, is presented in Section 5.

2. Methods

The linear stack structure and the U-shape structure are two classic frameworks for semantic segmentation, though there is still much room for improvement. For the linear stack structure, repeated downsampling and resizing operations lose much spatial information, and global context information is not fully exploited. U-shape structures such as U-Net [49] try to fill in the missing details by using skip-connections, but still cannot fundamentally solve the problems [54]. Although shallow features are abundant in spatial information, they are still too noisy to provide sufficient and useful information related to the target [58]. This kind of single processing pipeline, which has a limited effect in improving spatial information loss, often leads to inaccurate boundary definitions [52]. Another critical factor affecting the segmentation results is the size of the receptive field of the convolutional layer, especially for the recognition of targets with multiscale features. Recent work has focused on how to enlarge the receptive field and obtain more global context information. However, further research is needed, especially on how to collect relevant and effective global contextual information and how to fuse features from different depths and scales instead of simply concatenating them. The proposed architecture is designed to offer an improvement plan to address the issues mentioned above.

The multiscale fusion gated network (MFGNet) with dual-branch CNN architectures is mainly composed of four core modules, i.e., an SPPA module, an LFSA module, a GMFF module, and a new backbone network. As shown in Figure 1, patches of size H × W × C (256 × 356 × 3 in this case) are input to the backbone network. The features extracted by the backbone network are divided into two parts, which are input into the SPPA and LFSA modules, respectively, for multiscale semantic and shallow information extraction. In the end, the features from different depths and scales are fused by the GMFF module and output as a cloud segmentation mask with the same size as the input image.

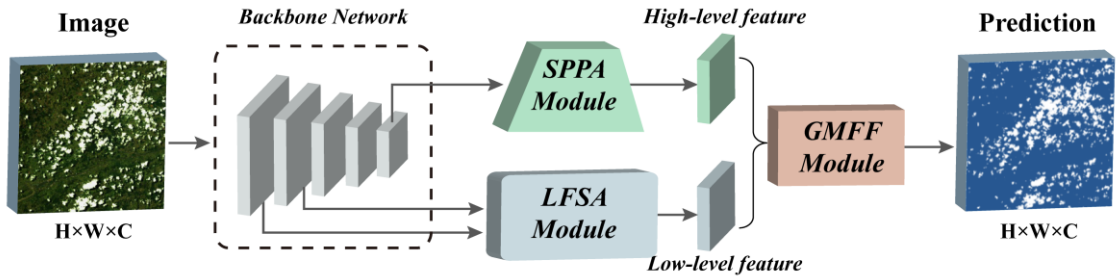

Figure 1. An overview of the multiscale fusion gated network (MFGNet) for cloud segmentation based on the dual-branch CNN architecture. Abbreviations: SPPA, spatial pyramid pooling attention; LFSA, low-level feature spatial attention; GMFF, gated multilevel feature fusion.

2.1. Backbone Network

The backbone network plays an essential role in improving the efficiency and accuracy of segmentation. Recently, lightweight skeleton networks such as the MobileNet series [59,60], ResNeXt [57], and Xception [61] have achieved state-of-the-art performance in many classification and segmentation tasks. We used ResNet as the base structure and proposed a new lightweight backbone network, combining the advantages of Xception and SENet. As shown in Table 2, the new backbone network consisted of a stack of 5 stages. In an effort to keep, as much as possible, sufficient quantities of shallow information, only stages 1, 3, and 4 were downsampled, and the output features were 1/2, 1/4, and 1/8 of the input image size, respectively. Stage 1 was composed of 3 CBR (Conv + BN + ReLU) blocks, which consisted of a convolutional layer (Conv), a batch normalization layer (BN), and a rectified linear unit (ReLU). The remaining stages, with the same topology and different hyperparameters, wre composed of several residual convolutional blocks (RCB) and an identity convolutional block

(ICB). The adjustable hyperparameters of RCB and ICB included stride, dilation rate, squeeze and excitation (SE) option, skip connection option, etc.; in stages 2–5, the number of filters of each stage was multiplied by a factor of 2. The kernel size of all convolutional layers was set to 3 × 3.

Table 2. Specification of the backbone network.

Stage	Input	Output	Operator	Filters	SE	Stride	Dilation Rate
1	256	256	CBR	32		1	1
	256	256	CBR	32		1	1
	256	128	CBR	64		2	1
2	128	128	RCB	64		1	1
	128	128	ICB	64	yes	1	1
	128	128	ICB	64	yes	1	1
3	128	64	RCB	128		2	2
	64	64	ICB	128	yes	1	2
	64	64	ICB	128	yes	1	2
	64	64	ICB	128	yes	1	2
4	64	32	RCB	256		2	4
	32	32	ICB	256	yes	1	4
	32	32	ICB	256	yes	1	4
	32	32	ICB	256	yes	1	4
	32	32	ICB	256	yes	1	4
	32	32	ICB	256	yes	1	4
5	32	32	RCB	512		1	5
	32	32	ICB	512	yes	1	5
	32	32	ICB	512	yes	1	5

Figure 2 shows two types of convolutional layers with different structures used in the backbone network. Both RCB and ICB were composed of depthwise separable convolutional layers (SepConv), BN, and ReLU. The main difference between RCB and ICB is that the former used skip connections, and the output feature was 1/2 of the input size. In addition, there was an SE unit in the ICB block, which was used to adjust the channel weight of the output layer.

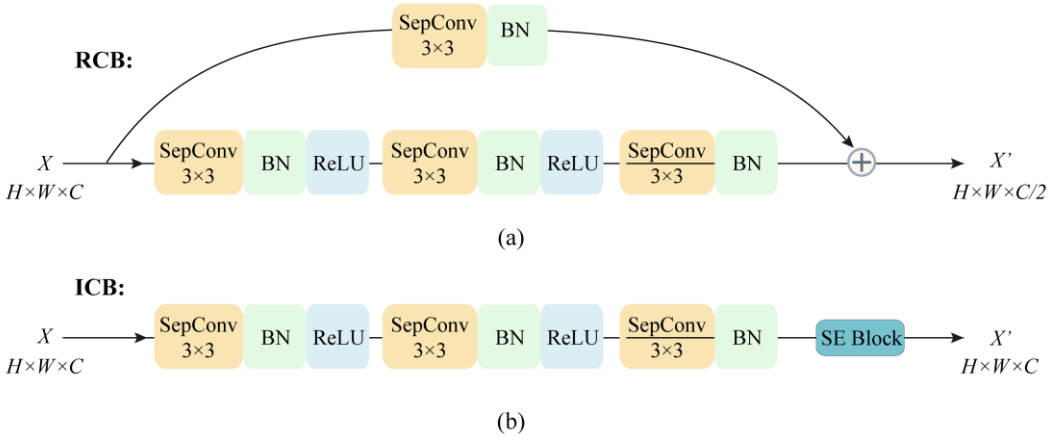

Figure 2. Two types of convolutional block in the backbone network. (**a**) Residual convolutional block with skip connections (Stride = 2). (**b**) Identity convolutional block (Stride = 1).

2.2. Spatial Pyramid Pooling Attention Module

Enlarging receptive fields and extracting multiscale features can help obtain more global context information. Recently, many practices have tried these two aspects and some useful network structure solutions have been proposed. For example, global convolution network (GCNet) [58] adopts a "large

kernel" to enlarge the receptive field, PSPNet utilizes a spatial pyramid pooling (SPP) module to obtain a multiscale pooling feature [55], and DeepLab [62] proposes atrous spatial pyramid pooling to capture the context information of different receptive fields. Inspired by PSPNet, we introduced an SPPA module that combines the advantages of SPP and channelwise attention. As shown in Figure 3, the features of the final stage output of the backbone network with a size of 1/8 of the original image were connected to a CBR block, and the number of channel dimensions was reduced to 256. A spatial pyramid pooling (SPP) submodule was applied to capture the context information from different scales. The SPP consisted of five average pooling layers, with kernel and stride sizes of 1×1, 2×2, 4×4, 8×8, and 16×16 respectively. Then, we directly upsampled all five pyramid level layers to 1/8 of the original image, and then concatenated them. Before being fed into the next channelwise attention block, the feature combination was reduced by a 1×1 Conv. In the attention block, the spatial information from all the channels was squeezed by average pooling and output as a one-dimensional vector of size $1 \times 1 \times C$. Followed by two 1×1 Conv and one active layer, the computed weight vector was able to reweight the feature and control feature selection. After upsampling by a factor of 4, the output of the SPPA module was a feature map with 1/2 the size of the input image.

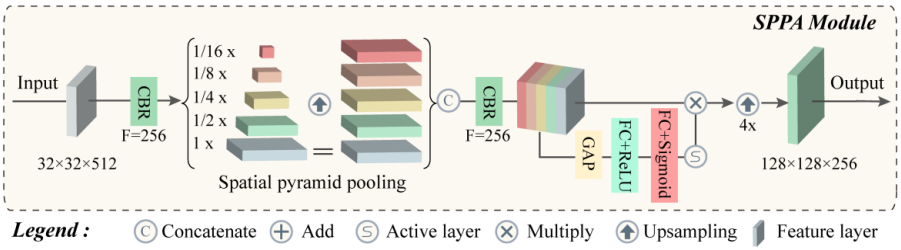

Figure 3. Illustration of the spatial pyramid pooling attention module. CBR = convolutional layer (Conv) + batch normalization layer (BN) + rectified linear unit (ReLU), GAP = global average pooling, F = filters.

2.3. Low-Level Feature Spatial Attention Module

The LFSA module (Figure 4) is mainly used to extract and fuse spatial features at different scales. The core step is that the first two levels of features generated by stages 1 and 2 of the backbone network are further refined by a spatialwise attention block. First, the maximum value of each pixel in all channels was calculated at the spatial scale and concatenated with the original features to enhance the weight of the cloud targets. We reduced the channel dimension of the features by a 1×1 Conv and utilized a squeeze factor to adjust the dimension reduction ratio. Then, a $1 \times 1 \times 1$ pointwise Conv and an active sigmoid layer were used to generate a spatial attention map. By multiplying by the input feature combination, the spatial attention map was able to reweight the feature and emphasize meaningful features on the spatial dimension. After being refined by the spatial attention (SA) block, the low-level features, holding 1/2 the size of the input image from different stages, were concatenated as the final output of LFSA.

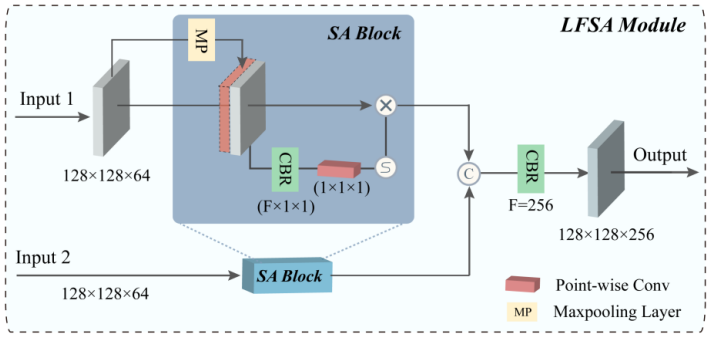

Figure 4. Illustration of the low-level feature spatial attention (LFSA) module. SA = spatial attention.

2.4. Gated Multilevel Feature Fusion Module

As a common technique in segmentation tasks, the traditional method of fusing shallow spatial information with semantic information is to simply concatenate or sum these features and then apply postprocessing operations such as CRF to refine the results. In a variation from previous practice, the GMLFF module (Figure 5), which is based on the attention mechanism, pays more attention to further extracting the useful information in the feature combination. In this case, we first combined the shallow and deep features and reduced the dimensions of the channels by a 1×1 Conv. Then, we pooled the concatenated features to a vector and computed a weight vector. Through multiplying by the concatenated features, the weight vector was able to adjust the weight of the useful information. In a variation from the attention methods used in the SPPA module, a skip connection was utilized to bring more abundant information. Followed by the upsampling layer, the feature map was fed into a 1×1 convolutional layer to get the final cloud segmentation mask with the same size as the original input image.

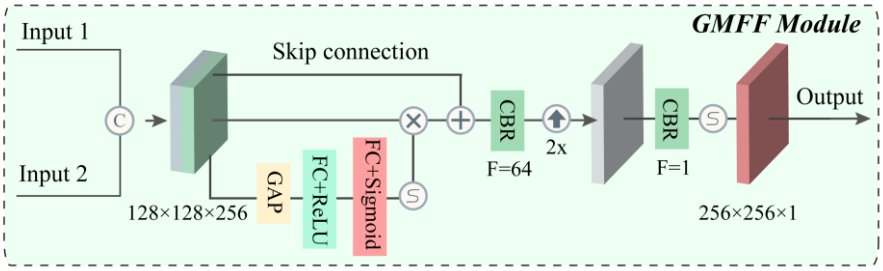

Figure 5. Illustration of the gated multilevel feature fusion (GMFF) module.

3. Experiments

3.1. Experimental Data

3.1.1. Dataset

GF-5 data providers need to complete precise cloud detection tasks with hundreds of scenes every day. Just a few years ago, cloud estimation of most satellite data relied mainly on the manual monitoring of RGB images. As mentioned before, traditional methods can achieve accurate cloud recognition of hyperspectral or multispectral data; however, they are not suitable for large-scale productions, especially for GF-5 hyperspectral images with massive data volumes. Moreover, as the first step of data processing, the cloud detection process is required to not consume too much time in data preprocessing, such as performing decompressions and atmospheric corrections. It is not difficult to find that, in most cases, there is sufficient information to make a clear judgment on the cloud through its color, shape, texture, shadow, spatial relationship, and many other features from RGB images. This is the main reason we chose the GF-5 preview RGB image (i.e., thumbnails) as the training dataset for cloud detection.

More than 1600 scenes of GF-5 RGB images with a size of 2008×2083 were collected on a global scale, covering the period from January to March 2019. Some scenes with invalid data were eliminated, and the rest of the images were further selected according to the land-cover types and cloud coverage to ensure that the model could be applied to common scenarios. A collection of images containing typical scenarios was chosen to evaluate the visual performance of the prediction results, and the remaining 717 scenes (Figure 6) were used for model training and quantitative evaluation. The scenes contained multiple collections such as cloudy, sunny, snow, cloud and snow coexisting, etc., covering common scenarios such as cities, mountains, forests, farmland, etc. All selected images were manually labeled with reference cloud masks (RCMs). In the first stage, the threshold-based method with careful thresholding was used to label the cloud, and in the second stage, the mask results were visually

checked and corrected, especially for complex samples with snow or ice. The RGB images and the RCMs together formed a four-band dataset.

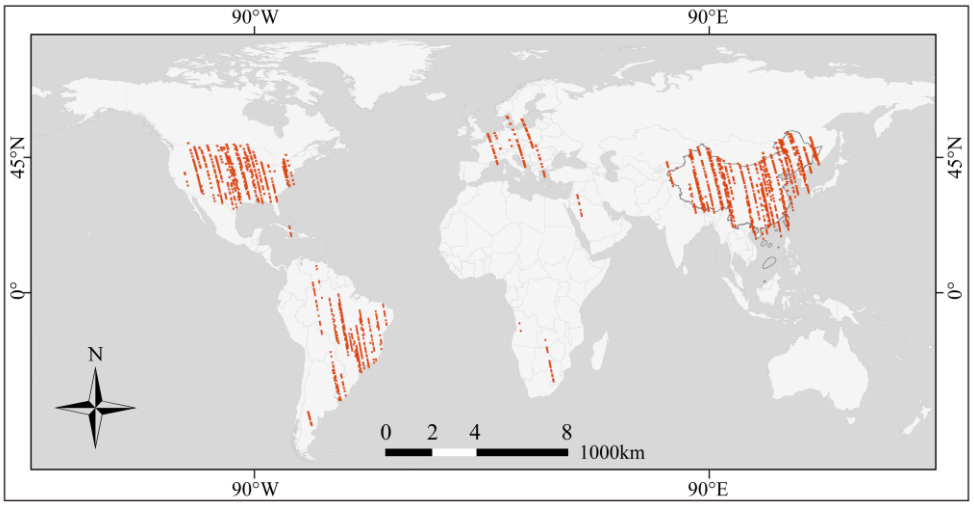

Figure 6. Distribution of Gaofen-5 (GF-5) satellite imagery.

3.1.2. Data Processing

In this process, some necessary preprocessing was performed on the entire GF-5 image to meet the constraints of the algorithm and the hardware, such as graphic processing unit (GPU) memory. As shown in Figure 7, we used a fixed-size window to crop the data into patches of size 256 × 256. The random cropping strategy was achieved by randomly setting the starting point coordinates and the rotation angle of the window. Further, during the cropping process, we set a threshold to keep more positive samples in the patches, to balance the positive and negative sample balance of the dataset.

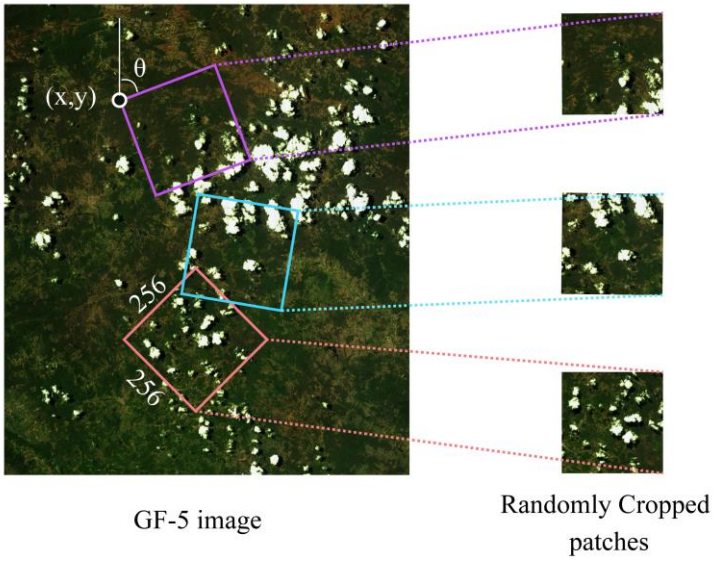

Figure 7. Random cropping strategy for the dataset.

Remote sensing images from the same place acquired at different times will show some radiation differences due to temporal issues. Considering that this phenomenon is mainly manifested as differences in brightness, contrast, etc., it is reasonable to perform color transformation on the data to enrich the diversity of the data. In addition, the spatial transformation of the data also helps the algorithm better identify the cloud target in the background.

To improve the generalization ability of the model and make it adapt to images acquired at different times and scenes, we adopted a random expansion strategy for the batch data before training. As shown in Figure 8, the data augmentation strategy proposed in this task included color-based methods such as saturation, brightness, contrast, and sharpness, and geometry-based methods such as rotate, flip, shift, zoom in, and zoom out. Except for rotate and flip, the amplitudes of the other transformations were set to ±20%.

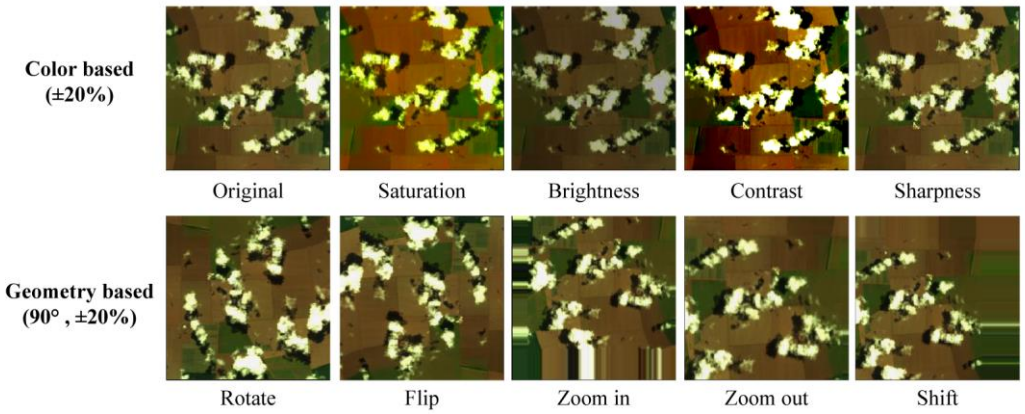

Figure 8. Data augmentation-based color and geometry transformation.

Random samples of 12k patches were used in this task. These data were split into two groups, i.e., 80% for training, and 20% for validation. All the input data were normalized to values between 0 and 1.

3.2. Experiment Settings

3.2.1. Model Training and Prediction

In the training stage, the patches of size 256 × 256 were input to the backbone network with five processing stages. The output of the backbone network's first two stages was input to the LFSA module. During this time, low-level features of different depths were selectively extracted through the SA block and output as a feature of size 128 × 128 × 256. At the same time, the features from stage 5 were input to the SPPA module. After pyramid pooling, the high-level features of different scales were concatenated and selectively extracted through a channelwise attention mechanism and output at the size of 128 × 128 × 256. Finally, the low-level spatial information and high-level semantic information were deeply fused by the GMFF module, and a cloud mask with a size of 256 × 256 × 1 was output as the final result.

The training procedure was performed in a TensorFlow (1.13.1) framework on an NVidia GeForce GTX 1080Ti GPU and optimized by the adaptive moment estimation (Adam) algorithm [63] (initial learning rate as 0.001) with "Binary_crossentropy" loss. One hundred epochs were used for training, and the batch size was 20. The convolution weights were initialized by "Glorot_uniform," and were drawn randomly from a uniform distribution within [-limit, limit] with the limit being defined in [64]. The biases in the convolutional layers were initialized with a constant of 0. To prevent overfitting, a dropout layer with dropout rate of 0.2 was added on the top of the backbone network.

The size of a GF-5 image is 2008 × 2830, which means that it needs to be divided into multiple patches for prediction. An overlap-tile strategy [49], which retains only the intermediate prediction results of each patch (Figure 9), was used to ensure the seamless segmentation of large images.

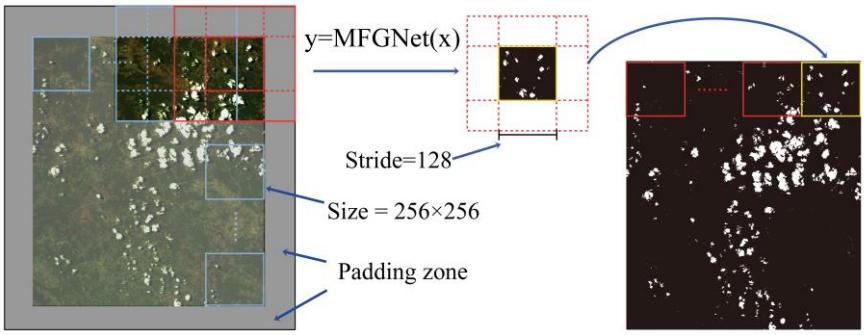

Figure 9. Overlap-tile strategy for seamless segmentation of GF-5 images.

In our experiments, BiSeNet, PSPNet, SegNet, and FCN8 (fully convolutional network with 8× upsamping) were also evaluated as reference methods on the same dataset with the same training parameter settings as the MFGNet. An ablation experiment was also conducted to test the performance of the main modules in the MFGNet, which is described in detail in Section 4.

3.2.2. Evaluation Metrics

The performance of the proposed model was quantitatively measured by the agreements and differences between predicted results and RCMs. The most common metrics Equations (1)–(5), the overall accuracy, recall, F1 score, precision, and intersection over union (IoU), were deployed as the evaluation index to evaluate the compared methods. For reference, a general analysis of accuracy metrics for classification tasks can be found in [65]. These metrics are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{PN}} \tag{2}$$

$$\text{Overall Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3}$$

$$\text{F1 score} = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}} \tag{4}$$

$$\text{IoU} = \frac{\text{TP}}{\text{FP} + \text{FN} + \text{TP}} \tag{5}$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively.

4. Results

4.1. Evaluation of the MFGNet

Observing changes in loss and accuracy is a simple and effective way to evaluate the quality of a model during training. The loss and accuracy of the training and validation set of each epoch were computed and are displayed in Figure 10a. As depicted in the figure, the accuracy curve of the training set rises rapidly; meanwhile, the loss curve drops rapidly and reaches stability after a few epochs. Although the loss of the validation set shows periodic oscillations, a stabilized curve is achieved after 60 epochs. The validation loss reaches the lowest point in the 80–100 epoch, and the curve trends of the training and validation sets during this period are the same, which indicates that the model is not overfitting.

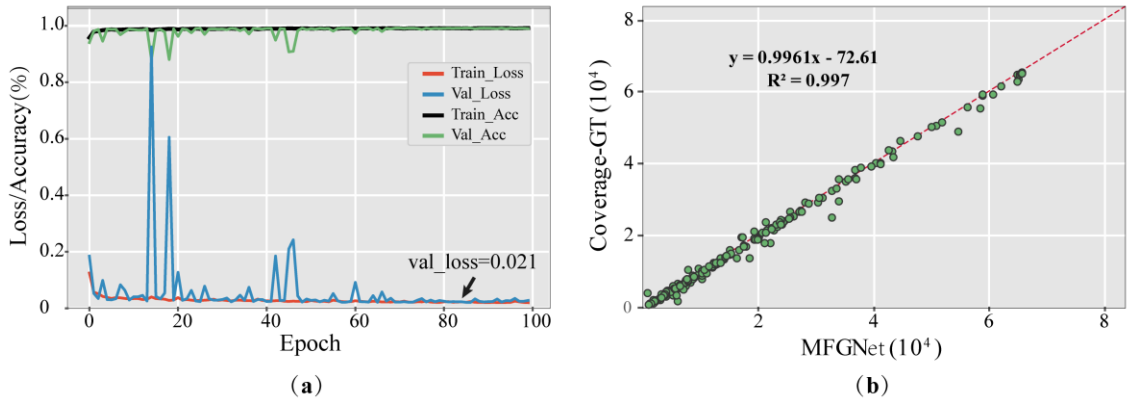

Figure 10. Diagram for evaluation model training on the validation set: (**a**) The loss and accuracy curve for the MFGNet; (**b**) The RCM plotted against the predicted cloud coverage for the MFGNet.

Figure 10b shows that a scatter plot of the RCMs and predicted cloud coverage was employed to investigate the performance of the MFGNet models further. The proposed model showed significant performance, and the predictions were highly consistent with the RCM with an R^2 of 0.99. It should be noted that the validation set contained cloud coverage data from different scenarios, and no patches with full cloud coverage or without any cloud coverage were evaluated, thereby resulting in a more reliable evaluation of the model.

4.2. Comparison Results

To quantitatively evaluate the performance of each model in the cloud detection task, we adopted overall accuracy, recall, F1 score, precision, and IoU as the evaluation metrics. From the results in Table 3, clearly the proposed MFGNet consistently outperformed all reference methods in terms of all metrics. In general, all CNN-based methods are effective for cloud detection, with both accuracy and precision reaching 95% or more, which not only exceeds the performance of traditional methods but also approaches the accuracy of manual labeling. It is worth mentioning that the recall value of the MFGNet is significantly better than other models, which indicates that this model has a lower false-negative rate. F1 score, which combines the evaluation results of precision and recall, can better represent the overall performance of the model, while IoU is used to judge the degree of coverage of the segmentation result on the target, which is more convincing for the segmentation task. Both of the above two comprehensive indicators of the MFGNet reached 0.9, which was significantly better than for the other methods. In general, the results show that the proposed models have better performance than other methods, and it also implies that they can perform more robustly on the validation set, which contains many kinds of cloud coverage data obtained from different scenarios.

Table 3. Performance of different methods for cloud detection.

Model	Accuracy	Precision	Recall	F1 Score	IoU
FCN8	0.96	0.96	0.80	0.83	0.73
SegNet	0.97	0.97	0.87	0.88	0.80
PSPNet	0.96	0.96	0.81	0.84	0.74
BiSeNet	0.97	0.97	0.88	0.89	0.81
MFGNet	**0.99**	**0.99**	**0.93**	**0.94**	**0.90**

4.3. Example Scene and Performance

Cloud segmentation examples for whole scene GF-5 imagery are shown in Figures 11 and 12. Four types of cases, including cloud-only, ice and snow coexisting, snow-only, and cloud and snow coexisting cases, are shown as a comparison. Figure 11a–d shows the recognition results for images with different cloud coverage. At first glance, most algorithms work quite well in the cloud segmentation

task, and apart from the apparent errors of FCN8 and SegNet, there is not much difference between the others. However, through careful comparison of the details, it is not difficult to find that the visual performance of the MFGNet is much better than the comparison method (discussed later). Figure 11 also reveals that, with the exception of individual cases, all methods performed well on ice recognition, indicating that the CNN-based method has fully learned the differences between ice and cloud features.

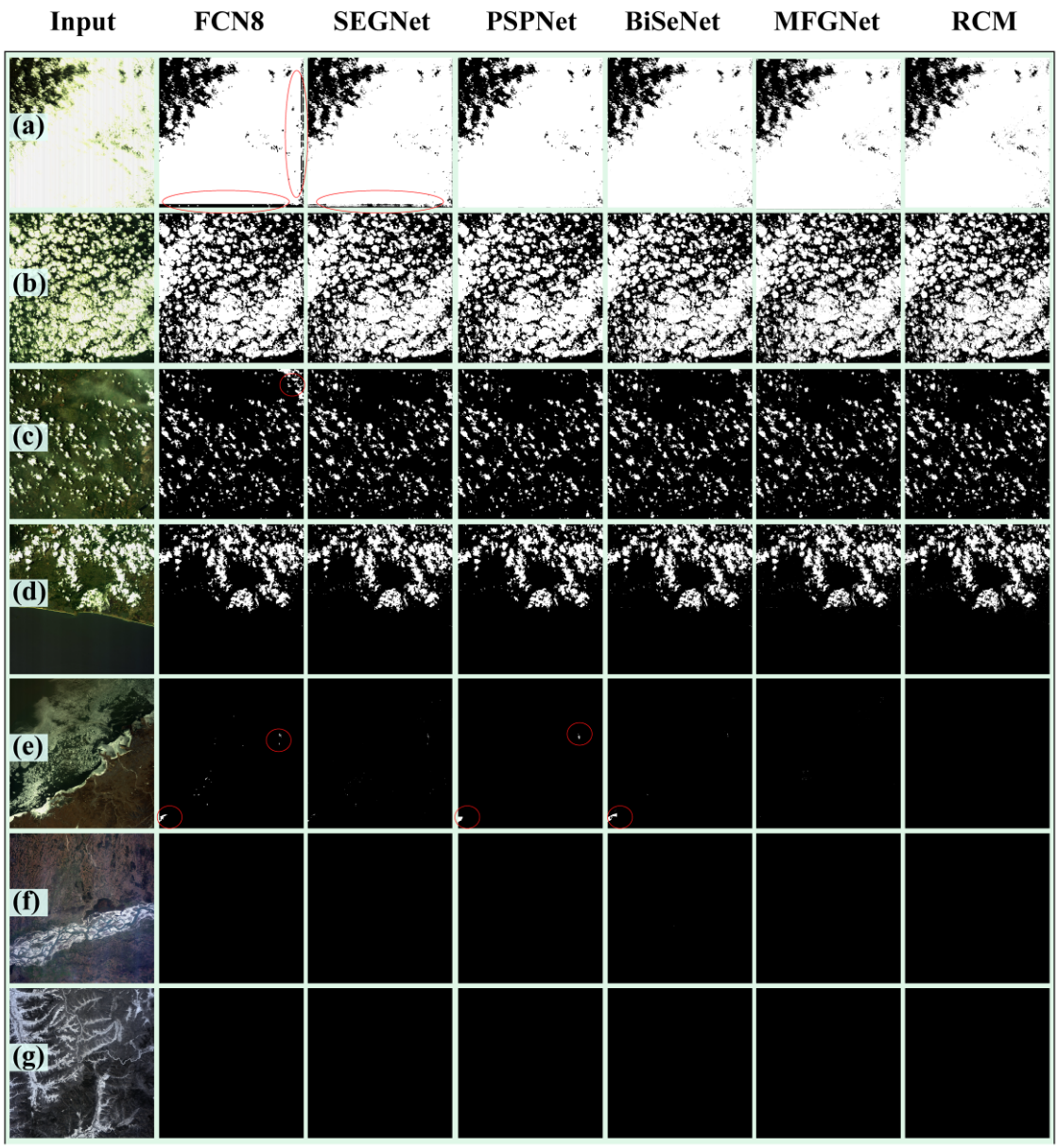

Figure 11. Comparison of cloud segmentation of CNN-based methods in the presence of cloud and ice. (**a**–**d**) are for cloud-only cases, and (**e**–**g**) are for ice and snow coexisting cases. All sizes of RGB images are 2008 × 2083. Abbreviations: RCM, reference cloud mask.

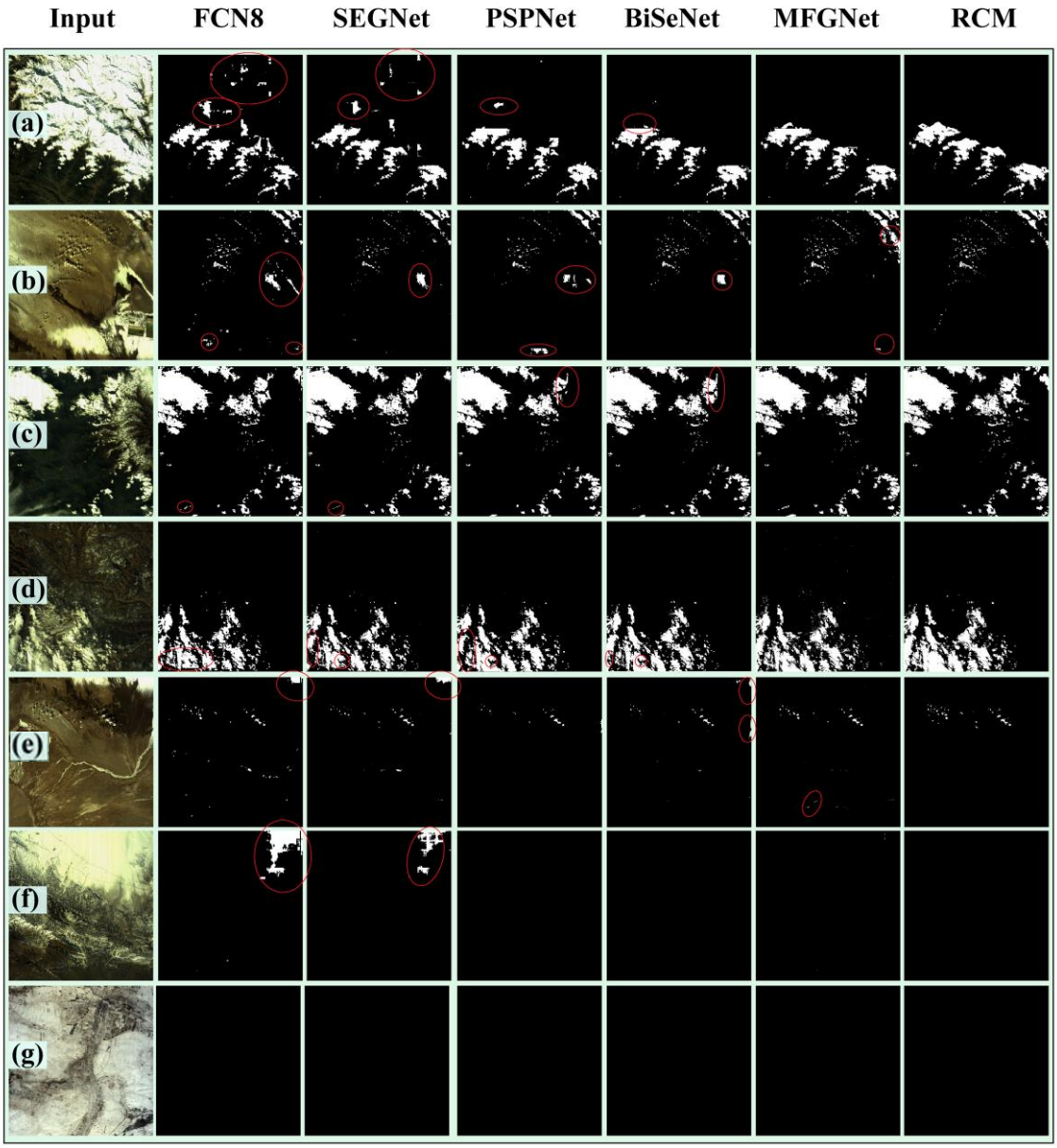

Figure 12. Comparison of cloud segmentation of CNN-based methods in the presence of cloud and snow. (**a**–**e**) are for cloud and snow coexisting cases, and (**f**,**g**) are for snow-only cases. All sizes of RGB images are 2008 × 2083.

The most serious issue for cloud recognition is the elimination of the misidentification of snow. In terms of statistics, the values of cloud and snow are very close. The low distinguishability in the values of cloud and snow makes for challenges in their identification. However, it is not difficult to find that in most cases, there is a distinguishable difference between snow and clouds, since the distribution of snow is closely related to the terrain. Surprisingly, most CNN-based methods showed the potential to distinguish between clouds and snow, indicating that the characteristics of clouds and snow can be learned through neural networks, even based on RGB imagery with only three bands. As shown in Figure 12, benefiting from the dual-branch CNN architecture, the MFGNet still achieved the best visual performance in all the experimental results. The SPPA module provided sufficient receptive fields while acquiring the multiscale features of the cloud. The attention mechanism adopted by the MFGNet ensures that the model can focus on the relevant and effective features to distinguish between clouds and snow. Furthermore, the fusion of features from different depths and scales improves the accuracy of the prediction. However, there is still room for improvement. By observing the detail

of the tough cases, which were inaccurate and marked with a red circle, we divided the inaccurate predictions into two types; the first was the interference of thin clouds, and the other was 100% snow. The misidentification of these two cases may be related to the inaccurate labeling of samples and the limitations of the model's capabilities, which will be discussed in the next section.

4.4. Efficiency Evaluation

As mentioned earlier, cloud detection is the first step in a data quality assessment. The detection process needs to be accurate and efficient. We calculated the MFLOPs (millions of floating point operations per second), #Params (number of network parameters), model size, and time cost of each method in the experiment to illustrate the efficiency performance. As we can see from Table 4, the model size of the MFGNet was much smaller than that of the other methods, which shows that the proposed methods achieve the highest accuracy with the fewest parameters. In addition, the MFLOPs of the MFGNet was only 15.72, which was not only the smallest in the comparison experiments but also reached the current level of the mainstream lightweight network. This indicates that the model has a higher tolerance for hardware devices. All models performed similarly in time cost and could complete the prediction of a scene of 2k × 2k images within 10 s. This means that CNN-based methods can complete cloud detections of more than 300 scenes in less than an hour, which is much more efficient than traditional methods. In short, the efficiency evaluation of the proposed model means that it shows great promise for practical applications.

Table 4. Efficiency comparison of the different methods.

Model	MFLOPs	#Params (10^6)	Model Size (mb)	Time Cost (Seconds/Scene)
FCN8	67.18	33.60	384	3.59
SegNet	35.74	10.19	116	3.75
PSPNet	44.01	21.97	251	9.90
BiSeNet	52.53	26.19	299	7.76
MFGNet	**15.72**	**7.83**	**90.5**	9.87

5. Discussion

5.1. Method Advantage Analysis

The previous experimental results show that the MFGNet outperforms reference methods of cloud segmentation. We believe that this mainly depends on the architecture of the CNN-based methods. It is reasonable that FCN8's segmentation results were not satisfactory. As an early proposed network, the depth of FCN8 is limited, and a lot of useful information is lost during the repeated upsampling process, which leads to misidentification and omission. Benefitting from the SPP module, PSPNet can distill more deep semantic information. Although its segmentation accuracy performance is slightly better than that of FCN8, the problem of loss of spatial information (LSI) still exists. These architectural defects in the models have led to a decrease in evaluation metrics, and they have limited capabilities in small target recognition. SegNet proposed a new upsampling strategy and added more shallow information to the decoder; to some extent, it improved the problem of LSI but also led to inaccurate boundary definitions. BiSeNet employs a dual-branch CNN structure to solve the LSI problem one step further, but as we can see from the experimental results, the segmentation results of SegNet and BiSeNet were over-smooth and inaccurate.

As we can see from Figure 13, the detailed performance of the MFGNet shows more consistency with the RCM. The proposed method has a more delicate edge representation and more accurate recognition accuracy than the U-shape and linear stack structures, especially on small cloud targets (Figure 14), which is an excellent proof of the advantages of the network architecture. The proposed methods can also perform well in the situation of cloud–snow coexistence. Interestingly, in some

complex cases, the segmentation results of the MFGNet were actually more accurate than that of the RCM. It may indicate that a good CNN-based method has better fault tolerance than manual labeling.

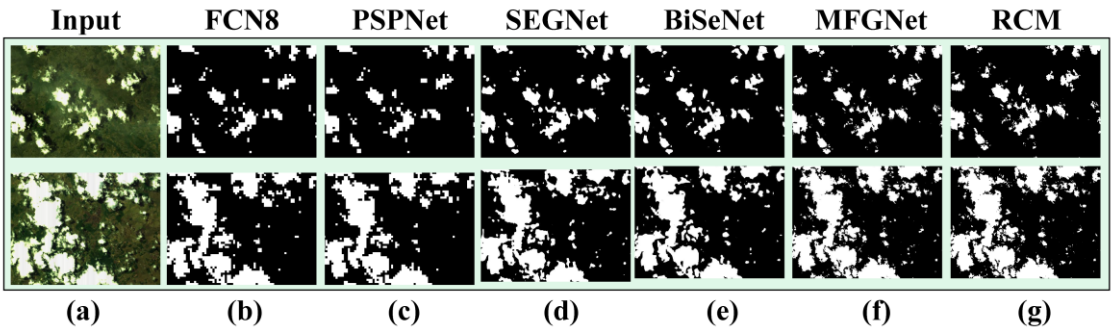

Figure 13. Detailed performance of different algorithms in cloud segmentation. (**a**) the GF-5 preview images. (**b**–**f**) are the prediction result of FCN8, PSPNet, SegNet, BiSeNet, and MFGNet. (**g**) the reference cloud mask.

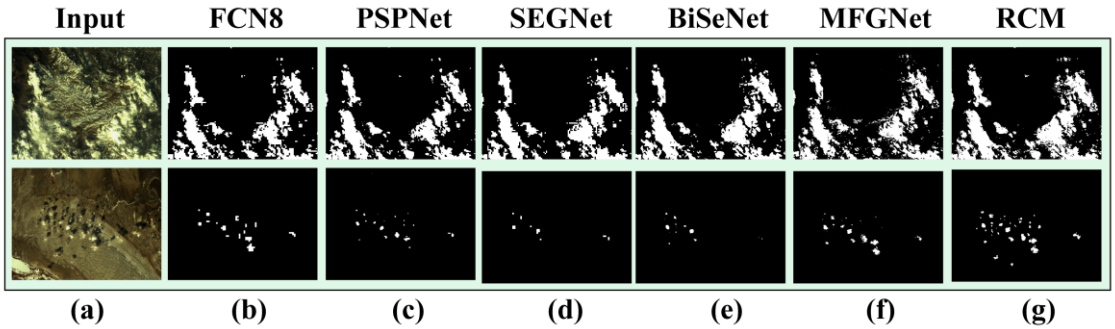

Figure 14. Detailed performance of different algorithms in tough cases. (**a**) the GF-5 preview images. (**b**–**f**) are the prediction result of FCN8, PSPNet, SegNet, BiSeNet, and MFGNet. (**g**) the reference cloud mask.

5.2. Limitation Analysis

In the process of making RCMs, we adopted a more conservative labeling strategy in order to retain as much useful data as possible for subsequent applications. Therefore, we mainly labeled thick clouds and as few as possible thin clouds. However, in actual operation, there is no reference standard, so it is challenging to label different images with a consistent judgment. It directly leads to inconsistent standards in the labeling of thin-cloud RCMs, which is also a common problem faced by deep learning applications.

Another complicated problem is that the algorithm is able to distinguish cloud from snow, but it will make mistakes in recognition of 100% cloud or snow. In an RGB image, 100% of the cloud or snow overlay on the image has often reached a saturation state in value, which means that it cannot be distinguished from features such as color, texture, and brightness. Although this situation is also challenging for artificial recognition, it can still be roughly judged by analyzing the surroundings. Under the premise of having a sufficiently large receptive field, deep learning algorithms can also realize cloud and snow recognition, but due to hardware limitations, we cannot input the entire scene of images into the network. In fact, the size of the input patches determines the upper limit of the receptive field, so it is also a limitation of the capabilities of most CNN-based algorithms.

5.3. Extended Application

The development of satellite remote sensing has increased the demand for large-scale data quality assessment. Like other satellite data, the GF-5 satellite also faces considerable challenges in data quality assessments, such as cloud detection, invalid data screening and classification, and so on (Figure 15).

The CNN-based method has natural advantages in semantic segmentation and image classification. Combined with the advantages of big data from satellite images, it can propose possible solutions for the realization of automated, full-process, high-efficiency, and high-precision satellite data quality assessment. As a lightweight network, the MFGNet can achieve high-precision cloud detection with lower computational consumption, which shows great potential for large-scale practical applications.

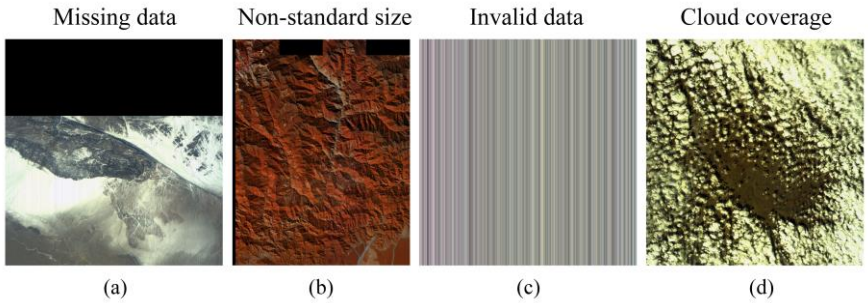

Figure 15. The main target of data quality assessment. (**a**) the GF-5 images with missing data. (**b**) the GF-5 images with non-standard image size. (**c**) the GF-5 images with invalid data. (**d**) the GF-5 images with cloud coverage.

As a data-driven method, a CNN-based model requires continuous optimization to meet the needs of global-scale applications. However, the considerable sample label workload becomes a problem. According to the previous analysis, the MFGNet has excellent performance on cloud segmentation in various environments. Therefore, the MFGNet can be used to predict the samples first, and screen out the unqualified samples with poor prediction accuracy for manual labeling. The newly generated data can be used for model training again to further improve its accuracy, thus forming a looping workflow, which can improve labeling efficiency.

6. Conclusions and Future Developments

In recent years, there has been an increasing demand for efficient cloud detection in massive satellite images, which makes for challenges to traditional cloud detection methods that mainly rely on spectral information. In this case, the combined use of RGB imagery and CNN-based methods provide a solution for efficient cloud detection in GF-5 satellite data. In this paper, we presented the MFGNet, a novel cloud segmentation model with dual-branch CNN architecture. The proposed model employs SPPA, LFSA, and GMFF modules to implement a better fusion of features from different depths and scales and strengthens the collection of useful spatial information. The MFGNet was trained on hundreds of globally distributed GF-5 satellite images in a variety of scenarios and compared with FCN8, SegNet, PSPNet, and BiSeNet. The overall accuracy, recall, F-score, precision, and IoU, were deployed to quantitatively evaluate the MFGNet and the compared methods. The experimental results show that, compared with the other models, the MFGNet can achieve promising performance for cloud recognition of GF-5 RGB imagery with an F1 score reaching 0.94 and an IoU of approximately 0.9. The efficiency test results also indicate that the proposed model has fewer parameters (#Params = 7.83×10^6) and less computational consumption (MFLOPs = 15.72). Based on these results, we believe that the use of CNN-based methods for cloud detection is a promising way forward and has practical significance for large-scale, automated, and efficient data quality assessment applications.

In our future study, we will collect as much data as possible from around the world for cloud segmentation to improve the generalizability of the algorithm. In addition, we will generalize the proposed method to other satellite data. Furthermore, to better overcome the weaknesses of the current models, we will try to use a small number of hyperspectral bands to improve the segmentation performance of targets where cloud and snow coexist.

Author Contributions: Conceptualization, J.Y., Y.L., and Y.Z.; Funding acquisition, X.Z.; Methodology, J.Y. and Y.Z.; Software, X.Z.; Writing—original draft, J.Y. and Y.L.; Writing—review and editing, X.Z., Y.Z., and P.H. All authors have read and agree to the published version of the manuscript.

Funding: This work was funded in part by the Major Projects of High-Resolution Earth Observation System (30-Y20A010-9007-17/18, 04-Y30B01-9001-18/20), and jointly by the 13th Five-Year Advance Research Project on Civil Space Technology of the National Defense Science and Technology Administration.

Acknowledgments: The authors would like to thank Wei Huang and Shiguang Wang for providing GPU resources for this work.

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

References

1. Yang, Y.; Li, H.; Du, Y.; Cao, B.; Liu, Q.; Sun, L.; Zhu, J.; Mo, F. A temperature and emissivity separation algortihm for chinese gaofen-5 satelltie data. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium(IGARSS 2018), Valencia, Spain, 22–27 July 2018; pp. 2543–2546.
2. Liu, L.; Shang, K. Mineral information extraction based on gaofen-5's thermal infrared data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-3*, 1157–1160. [CrossRef]
3. Yu, J.C.; Yan, B.K. Efficient solution of large-scale domestic hyperspectral data processing and geological application. In Proceedings of the IEEE 2017 International Workshop on Remote Sensing with Intelligent Processing, Shanghai, China, 18–21 May 2017; pp. 1–4.
4. King, M.D.; Platnick, S.; Menzel, W.P.; Ackerman, S.A.; Hubanks, P.A. Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 3826–3852. [CrossRef]
5. Irish, R.R.; Barker, J.L.; Goward, S.N.; Arvidson, T. Characterization of the landsat-7 etm+ automated cloud-cover assessment (acca) algorithm. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 1179–1188. [CrossRef]
6. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [CrossRef]
7. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images. *Remote Sens. Environ.* **2015**, *159*, 269–277. [CrossRef]
8. Rossow, W.B.; Garder, L.C. Cloud detection using satellite measurements of infrared and visible radiances for isccp. *J. Clim.* **1993**, *6*, 2341–2369. [CrossRef]
9. Gesell, G. An algorithm for snow and ice detection using avhrr data an extension to the apollo software package. *Int. J. Remote Sens.* **1989**, *10*, 897–905. [CrossRef]
10. Stowe, L.; McClain, E.; Carey, R.; Pellegrino, P.; Gutman, G.; Davis, P.; Long, C.; Hart, S. Global distribution of cloud cover derived from noaa/avhrr operational satellite data. *Adv. Space Res.* **1991**, *11*, 51–54. [CrossRef]
11. Qiu, S.; He, B.B.; Zhu, Z.; Liao, Z.M.; Quan, X.W. Improving fmask cloud and cloud shadow detection in mountainous area for landsats 4–8 images. *Remote Sens. Environ.* **2017**, *199*, 107–119. [CrossRef]
12. Qiu, S.; Zhu, Z.; He, B. Fmask 4.0: Improved cloud and cloud shadow detection in landsats 4–8 and sentinel-2 imagery. *Remote Sens. Environ.* **2019**, *231*, 1–20. [CrossRef]
13. Hagolle, O.; Huc, M.; Pascual, D.V.; Dedieu, G. A multi-temporal method for cloud detection, applied to formosat-2, venμs, landsat and sentinel-2 images. *Remote Sens. Environ.* **2010**, *114*, 1747–1755. [CrossRef]
14. Zhu, Z.; Woodcock, C.E. Automated cloud, cloud shadow, and snow detection in multitemporal landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sens. Environ.* **2014**, *152*, 217–234. [CrossRef]
15. Lin, C.H.; Lin, B.Y.; Lee, K.Y.; Chen, Y.C. Radiometric normalization and cloud detection of optical satellite images using invariant pixels. *ISPRS J. Photogramm. Remote Sens.* **2015**, *106*, 107–117. [CrossRef]
16. Di Vittorio, A.V.; Emery, W.J. An automated, dynamic threshold cloud-masking algorithm for daytime avhrr images over land. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1682–1694. [CrossRef]
17. Sun, L.; Wei, J.; Wang, J.; Mi, X.; Guo, Y.; Lv, Y.; Yang, Y.; Gan, P.; Zhou, X.; Jia, C. A universal dynamic threshold cloud detection algorithm (udtcda) supported by a prior surface reflectance database. *J. Geophys. Res. Atmos.* **2016**, *121*, 7172–7196. [CrossRef]

18. Luo, Y.; Trishchenko, A.P.; Khlopenkov, K.V. Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven modis land bands over canada and north america. *Remote Sens. Environ.* **2008**, *112*, 4167–4185. [CrossRef]
19. Frantz, D.; Haß, E.; Uhl, A.; Stoffels, J.; Hill, J. Improvement of the fmask algorithm for sentinel-2 images: Separating clouds from bright surfaces based on parallax effects. *Remote Sens. Environ.* **2018**, *215*, 471–481. [CrossRef]
20. Bian, J.; Li, A.; Liu, Q.; Huang, C. Cloud and snow discrimination for ccd images of HJ-1A/B constellation based on spectral signature and spatio-temporal context. *Remote Sens.* **2016**, *8*, 31. [CrossRef]
21. Ge, S.L.; Dong, S.Y.; Sun, G.Y.; Du, Y.M.; Lin, Y. Cloud detection algorithm for images of visual and infrared multispectral imager. *Aerosp. Shanghai* **2019**, *36*, 204–208. [CrossRef]
22. Zhan, Y.; Wang, J.; Shi, J.; Cheng, G.; Yao, L.; Sun, W. Distinguishing cloud and snow in satellite images via deep convolutional network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1785–1789. [CrossRef]
23. Wang, L.; Chen, Y.; Tang, L.; Fan, R.; Yao, Y. Object-based convolutional neural networks for cloud and snow detection in high-resolution multispectral imagers. *Water* **2018**, *10*, 1666. [CrossRef]
24. Oishi, Y.; Ishida, H.; Nakamura, R. A new landsat 8 cloud discrimination algorithm using thresholding tests. *Int. J. Remote Sens.* **2018**, *39*, 9113–9133. [CrossRef]
25. Shao, Z.; Pan, Y.; Diao, C.; Cai, J. Cloud detection in remote sensing images based on multiscale features-convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4062–4076. [CrossRef]
26. Hong, Y.; Hsu, K.L.; Sorooshian, S.; Gao, X. Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J. Appl. Meteorol.* **2004**, *43*, 1834–1853. [CrossRef]
27. Hall, D.K.; Riggs, G.A.; Salomonson, V.V. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sens. Environ.* **1995**, *54*, 127–140. [CrossRef]
28. Ghasemian, N.; Akhoondzadeh, M. Introducing two random forest based methods for cloud detection in remote sensing images. *Adv. Space Res.* **2018**, *62*, 288–303. [CrossRef]
29. Egli, S.; Thies, B.; Bendix, J. A hybrid approach for fog retrieval based on a combination of satellite and ground truth data. *Remote Sens.* **2018**, *10*, 628. [CrossRef]
30. Lee, Y.; Wahba, G.; Ackerman, S.A. Cloud classification of satellite radiance data by multicategory support vector machines. *J. Atmos. Ocean. Technol.* **2004**, *21*, 159–169. [CrossRef]
31. Ishida, H.; Oishi, Y.; Morita, K.; Moriwaki, K.; Nakajima, T.Y. Development of a support vector machine based cloud detection method for modis with the adjustability to various conditions. *Remote Sens. Environ.* **2018**, *205*, 390–407. [CrossRef]
32. 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]
33. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
34. Tsagkatakis, G.; Aidini, A.; Fotiadou, K.; Giannopoulos, M.; Pentari, A.; Tsakalides, P. Survey of deep-learning approaches for remote sensing observation enhancement. *Sensors* **2019**, *19*, 3929. [CrossRef] [PubMed]
35. Zhang, L.; Li, H.; Shen, P.Y.; Zhu, G.M.; Song, J.; Shah, S.A.A.; Bennamoun, M.; Zhang, L. Improving Semantic Image Segmentation With a Probabilistic Superpixel-Based Dense Conditional Random Field. *IEEE Access* **2018**, *6*, 15297–15310. [CrossRef]
36. Hegarat-Mascle, S.L.; Andre, C. Use of markov random fields for automatic cloud/shadow detection on high resolution optical images. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 351–366. [CrossRef]
37. 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]
38. Chai, D.; Newsam, S.; Zhang, H.K.; Qiu, Y.; Huang, J. Cloud and cloud shadow detection in landsat imagery based on deep convolutional neural networks. *Remote Sens. Environ.* **2019**, *225*, 307–316. [CrossRef]
39. Mohajerani, S.; Saeedi, P. Cloud-net+: A cloud segmentation cnn for landsat 8 remote sensing imagery optimized with filtered jaccard loss function. *arXiv* **2020**, arXiv:2001.08768.
40. Yang, J.; Guo, J.; Yue, H.; Liu, Z.; Hu, H.; Li, K. Cdnet: Cnn-based cloud detection for remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6195–6211. [CrossRef]

41. Jeppesen, J.H.; Jacobsen, R.H.; Inceoglu, F.; Toftegaard, T.S. A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sens. Environ.* **2019**, *229*, 247–259. [CrossRef]
42. Drönner, J.; Korfhage, N.; Egli, S.; Mühling, M.; Thies, B.; Bendix, J.; Freisleben, B.; Seeger, B. Fast cloud segmentation using convolutional neural networks. *Remote Sens.* **2018**, *10*, 1782. [CrossRef]
43. Liu, H.; Zeng, D.; Tian, Q. In Super-pixel cloud detection using hierarchical fusion cnn. In Proceedings of the 2018 IEEE Fourth International Conference on Multimedia Big Data, Xi'an, China, 13–16 September 2018; pp. 1–6.
44. Morales, G.; Huamán, S.G.; Telles, J. In Cloud detection in high-resolution multispectral satellite imagery using deep learning. In Proceedings of the International Conference on Artificial Neural Networks, Kuala Lumpur, Malaysia, 21–23 November 2018; pp. 280–288.
45. Guo, Z.S.; Li, C.H.; Wang, Z.M.; Kwok, E.; Wei, X. A cloud boundary detection scheme combined with aslic and cnn using zy-3, gf-1/2 satellite imagery. In Proceedings of the ISPRS Technical Commission III Midterm Symposium on "Developments, Technologies and Applications in Remote Sensing", Beijing, China, 5–7 May 2018; pp. 699–702. [CrossRef]
46. Zi, Y.; Xie, F.; Jiang, Z. A cloud detection method for landsat 8 images based on pcanet. *Remote Sens.* **2018**, *10*, 877. [CrossRef]
47. Chen, Y.; Fan, R.; Bilal, M.; Yang, X.; Wang, J.; Li, W. Multilevel cloud detection for high-resolution remote sensing imagery using multiple convolutional neural networks. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 181. [CrossRef]
48. Xie, F.; Shi, M.; Shi, Z.; Yin, J.; Zhao, D. Multilevel cloud detection in remote sensing images based on deep learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3631–3640. [CrossRef]
49. Ronneberger, O.; Fischer, P.; Brox, T. In U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 December 2015; pp. 234–241.
50. Badrinarayanan, V.; Kendall, A.; SegNet, R.C. A deep convolutional encoder-decoder architecture for image segmentation. *arXiv* **2015**, arXiv:1511.00561. [CrossRef] [PubMed]
51. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
52. Hatamizadeh, A.; Terzopoulos, D.; Myronenko, A. Edge-gated cnns for volumetric semantic segmentation of medical images. *arXiv* **2020**, arXiv:2002.04207.
53. Takikawa, T.; Acuna, D.; Jampani, V.; Fidler, S. In Gated-scnn: Gated shape cnns for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 20–26 December 2019; pp. 5229–5238.
54. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. In Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV 2018), Munich, Germany, 8–14 September 2018; pp. 325–341.
55. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. In 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.
56. Hu, J.; Shen, L.; Sun, G. In Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
57. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. In Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
58. Zhang, Z.; Zhang, X.; Peng, C.; Xue, X.; Sun, J. In Exfuse: Enhancing feature fusion for semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV 2018), Munich, Germany, 8–14 September 2018; pp. 269–284.
59. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. In Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
60. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V. In Searching for mobilenetv3. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 20–26 December 2019; pp. 1314–1324.
61. Chollet, F. In Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.

62. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. In Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV 2018), Munich, Germany, 8–14 September 2018; pp. 801–818.
63. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
64. Glorot, X.; Bengio, Y. In Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
65. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).