*remote sensing*

MDPI

*Article*

# A Cloud Detection Method Using Convolutional Neural Network Based on Gabor Transform and Attention Mechanism with Dark Channel Subnet for Remote Sensing Image

**Jing Zhang [1,*], Qin Zhou [1], Jun Wu [1], Yuchen Wang [1], Hui Wang [1], Yunsong Li [1], Yuzhou Chai [2] and Yang Liu [2]**

[1] State Key Laboratory of Integrated Service Network, Xidian University, Xi'an 710071, China; gaolate77@gmail.com (Q.Z.); wjeugene@163.com (J.W.); yc_wang@stu.xidian.edu.cn (Y.W.); sctian@stu.xidian.edu.cn (H.W.); ysli@mail.xidian.edu.cn (Y.L.)

[2] Data Transmission Institute, China Academy of Space Technology, Xi'an 710000, China; jokeyuchen@163.com (Y.C.); jzclass@163.com (Y.L.)

* Correspondence: jingzhang@xidian.edu.cn; Tel.: +86-298-820-3116

Received: 17 August 2020; Accepted: 2 October 2020; Published: 7 October 2020




**Abstract:** Cloud detection, as a crucial step, has always been a hot topic in the field of optical remote sensing image processing. In this paper, we propose a deep learning cloud detection Network that is based on the Gabor transform and Attention modules with Dark channel subnet (NGAD). This network is based on the encoder-decoder framework. The information on texture is an important feature that is often used in traditional cloud detection methods. The NGAD enhances the attention of the network towards important texture features in the remote sensing images through the proposed Gabor feature extraction module. The channel attention module that is based on the larger scale features and spatial attention module that is based on the dark channel subnet have been introduced in NGAD. The channel attention module highlights the important information in a feature map from the channel dimensions, weakens the useless information, and helps the network to filter this information. A dark channel subnet with spatial attention module has been designed in order to further reduce the influence of the redundant information in the extracted features. By introducing a "dark channel", the information in the feature map is reconstructed from the spatial dimension. The NGAD is validated while using the Gaofen-1 WFV imagery in four spectral bands. The experimental results show that the overall accuracy of NGAD reaches 97.42% and the false alarm rate reaches 2.22%. The efficiency of cloud detection using NGAD exceeds the state-of-art image segmentation network model and remote sensing image cloud detection model.

**Keywords:** cloud detection; gabor transform; attention mechanism; dark channel subnet; NGAD

## 1. Introduction

With the rapid development of remote sensing satellite technology, satellite images are increasingly being used in daily lives. An increasing amount of remote sensing data is being used in environmental protection, agricultural engineering, and others [1]. In daily life, people use remote sensing satellite maps for geological mapping, urban heat island monitoring, environmental monitoring, as well as for fire detection in forests from remote sensing images [2–5]. However, more than 66% of the world's surface is covered by clouds [6]. Therefore, cloud-covered areas will inevitably appear in remote sensing satellite images. Because of the natural environment and the angles at which the remote sensing images are taken, different types of clouds are captured in such images, including thin clouds,

thick clouds, etc. Thick clouds sometimes cover the ground completely, and this affects the subsequent ground recognition and environmental monitoring. The thin cloud is generally semi-transparent and translucent. Although the ground features are not completely blocked, the ground features and the thin cloud information are mixed together, which results in blurred or missing ground feature information. This greatly reduces the quality of the remote sensing images and affects their subsequent recognition. Therefore, cloud detection is a hot topic in the preprocessing of remote sensing images.

The traditional cloud detection methods are mainly divided into threshold-based methods and methods that are based on spatial texture characteristics. The former utilizes the difference in the brightness of the cloud and the ground objects in order to obtain the physical threshold for dividing the cloud and non-cloud areas [7,8]. Zhu et al. [9,10] proposed cloud detection methods for the Landsat imagery by setting multiple thresholds, derived from unique physical attributes, in order to extract the cloud areas. Luo et al. [11] utilized a specific scene-dependent decision matrix to identify clouds in the MODIS imagery. Jedlovec and Haines [12] synthesized several images of the same area at different times into clear sky images and classified them using the synthetic clear sky image data as the threshold. Although the threshold method is simple and effective for specific sensor data, it depends on the selection of spectral bands and physical parameters [13]. Therefore, the threshold-based methods are only applicable to specific sensors and specific scenarios and lack universal adaptability. In addition, these methods always misidentify the highly reflective non-cloud areas as cloud areas. Therefore, some cloud methods combine geometric and texture features with the spectral features in order to improve the cloud detection accuracy [14,15]. The traditional cloud detection methods used in remote sensing images, such as the threshold-based and rule-based cloud detection methods, use hand-crafted features and the special threshold to identify the cloud and non-cloud regions in the remote sensing images, which do not utilize semantic-level information[16]. Multi-temporal cloud detection methods make use of the time-series imagery in order to reduce further misidentification of the non-cloud areas [17–19]. However, these methods require multiple sets of images with and without clouds from the same location, which are difficult to collect. In recent years, a large number of remote sensing image processing algorithms that are based on deep learning have performed well in terms of object recognition [20,21] and semantic segmentation [22,23].

Cloud detection algorithms can be designed on the basis of the idea of semantic image segmentation algorithm that is based on deep learning. The deep convolutional neural network can extract various features, such as spatial features and spectral features. Long et al. [24] proposed Fully Convolutional Networks (FCN), which realized end-to-end semantic segmentation and it is the pioneering work in semantic segmentation. Ronneberger et al. [25] proposed a classic semantic segmentation algorithm, named U-Net, which is based on the encoder-decoder structure. Chen et al. [26] also designed a semantic segmentation neural network, called Deeplabv3+, based on the encoder-decoder structure, which exhibits excellent performance. Xie et al. [27] proposed a cloud detection method that combines a super-pixel algorithm and a convolutional neural network. Zi et al. [28] proposed a cloud detection method that is based on PCANet. Jacob et al. [29] proposed the Remote Sensing Network (RS-Net) with encoder-decoder convolutional neural network structure, which performs well in detecting clouds in the remote sensing images. Li et al. [30] proposed a framework that can train deep networks with only block-level binary labels. Yu et al. [31] proposed the MFGNet, which employs three different modules in order to implement a better fusion of features from different depths and scales.

Texture and color features are not only important information for the segmentation of cloud regions by the rule-based remote sensing image cloud detection algorithm, but also important information for the neural network detection algorithm to segment the cloud region [32].

A neural network model within the deep learning framework can extract a variety of image features. However, there is no distinction between the importance of the different types of features, and there will also always be some redundant information. Scientists have discovered the attention mechanism from the study of the human visual system. In cognitive science, humans selectively

focus on the part of the total available information and ignore the other visible information due to the bottleneck of information processing [33]. Such a mechanism is usually called attention mechanism and is divided into the hard attention mechanism and the soft attention mechanism. The soft attention mechanism in a convolutional neural network model is used for locating the most salient features, such that redundancy is removed for the vision tasks [34]. This mechanism is widely used in image captioning, object detection, and others [35]. Therefore, filtering the information extracted from the neural network model can help in improving the performance of the neural network model, depending on the characteristics of cloud detection.

We have proposed a network for cloud detection in remote sensing images, based on Gabor transform and spatial and channel attention mechanism, named NGAD, which is built on the encoder-decoder structure. We have designed a Gabor feature extraction module based on Gabor transform and added it into the encoder in order to enhance the encoder to pay attention to the image texture in the low-level feature map. There is some redundancy in the decoder when interpreting information. Thus, we have introduced a channel attention module to improve the abstract information at the decoder. From the perspective of color characteristics, we have created a new subnet, named Dark channel subnet, and used it as an auxiliary input of the spatial attention module to further eliminate the redundant information in the low-level feature map that assists the decoder. Cloud detection in GF-1 WFV imagery is a challenging task because of the unfixed radiometric calibration parameters and insufficient spectral information [36–38]. Therefore, we have evaluated the proposed algorithm by applying it to the GF-1 WFV data set [39].

The main innovation of this paper can be summarized, as follows. First, we designed a Gabor module that uses the texture features that were extracted by the Gabor filter to help the network perform feature learning. Second, we have introduced attention modules that provide enhanced key information to the cloud detection network based on its network structure and the characteristics of cloud detection. Third, we have proposed the Dark channel subnet in order to generate the auxiliary feature map that is required by the spatial attention module.

## 2. Methods

The NGAD is based on the encoder-decoder structure with Dark channel subnet. It is an end-to-end convolutional neural network. The width and height of the output are equal to that of the input of the encoder. The evaluation of the network model is divided into two stages, training and testing. In the training stage, the input of the encoder is a multi-spectral image, $x \in R^{w \times h \times s}$. The cloud mask corresponding to the input image is $x \in R^{w \times h \times 1}$. The input of Dark channel subnet is generated from $x \in R^{w \times h \times s}$. The output of the NGAD is a probability map $\hat{y} \in R^{w \times h \times 1}$. Next, we use a fixed threshold $t$ to perform a binary division on the output image in order to obtain the final binary segmentation map. For balancing the commission and omission errors in the cloud mask, the value of $t$ is taken to be 0.5. The loss function that is used by the network in the training stage is binary cross entropy loss function, and it is expressed as:

$$L(y, \hat{y}) = -\frac{1}{w \cdot h}(\hat{y} ln(y) + (1 - \hat{y}) ln(1 - y)) \quad (1)$$

During the training phase, the parameters in the network are continuously updated using the back-propagation algorithm [40] that is based on the loss function. We have also used the Adam optimization gradient algorithm [41] that enables convergence at high-performing local minima. The implementation of NGAD is based on Python 3.6 and employing Keras 2.2.4 and TensorFlow 1.12 deep learning framework. All of the experiments were carried out using the NVIDIA GEFORCE RTX 2080 Ti card.

In this section, we introduce our proposed network, NGAD, in detail. First, we describe the overall network architecture of NGAD. Subsequently, we analyzed the Gabor feature extraction module for

texture feature enhancement, Channel attention module based on the larger scale features, and spatial attention module based on Dark channel subnet, respectively.

### *2.1. The Framework of NGAD*

Figure 1 shows a block diagram of the proposed NGAD, which is based on the encoder–decoder structure and incorporated with Dark channel subnet. We first use the modified Unet network as our basic framework. The backbone of the network consists of a encoding path and an decoding path. The encodeing path consists the repeated application of CC block or CCD block (CC block with a dropout layer), and a $2 \times 2$ max pooling operation with stride 2 of downsampling. Similar to Unet, each downsampling step we double the number of the feature channels. Every step in the decoding path consists of an upsampling of the feature map, a concatentation with the correspondingly croped feature map from the encoding path, and CC block same with the encoding path to halves the number of feature channels. The number of convolution kernels of the CC block at the encoding end are 64, 128, 256, and 512, respectively. The decoding end is 256, 128, 64, and 32 respectively. At the final layer, a $1 \times 1$ convolution is used to map each 32 component feature vector in order to generate semantic score maps for cloud aeras and non-cloud aeras.

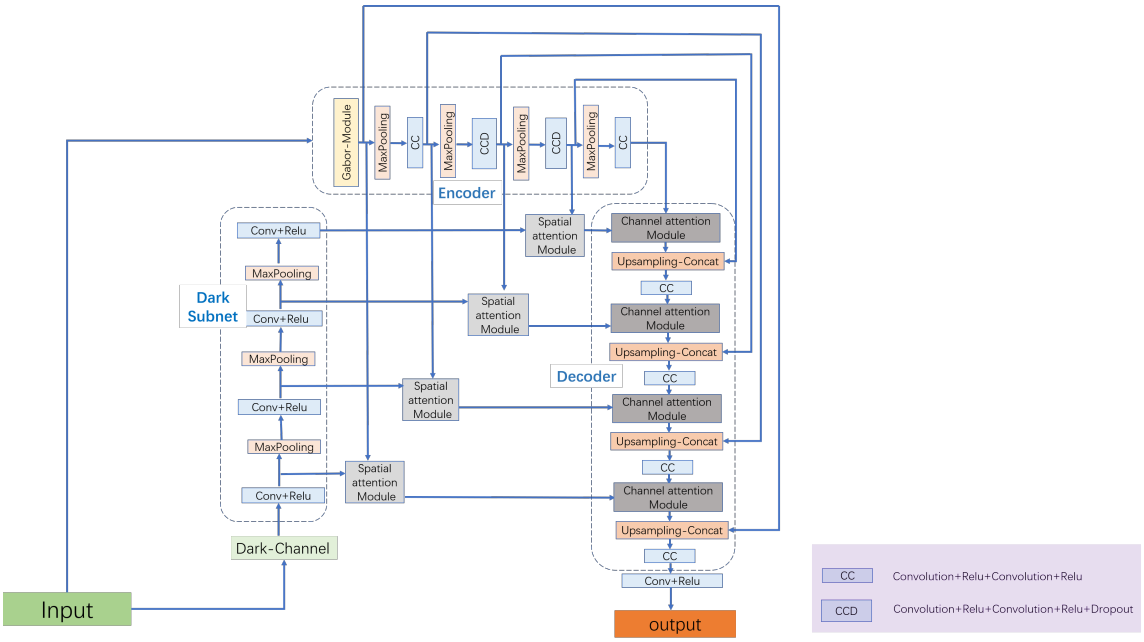



**Figure 1.** Block diagram of the Network based on the Gabor transform and Attention modules with Dark channel subnets (NGADs) architecture.

The CC block contains two layers of convolution with a rectified linear unit (ReLU) as the activation function, which is expressed as:

$$ReLu(x) = max(x, 0) \tag{2}$$

The CCD block consists of two convolutional layers with the activation function ReLU and a dropout layer, and the dropout ration was set to 0.2. The convolutional layer in the CC block and CCD block both use the filter size of $3 \times 3$. In the training stage, the dropout layer randomly produces certain neurons in the network output 0. This layer is only used in the encoder. In fact, the network structure will be slightly different in each training epoch, and the inactivation of some neurons in the encoder does not affect the feature extraction process. This can effectively prevent overfitting [42].

The traditional cloud detection methods tend to extract multiple textures and color features to enhance the performance of cloud detection. Deng et al. [32] used the Gabor transform for extracting

the texture features in cloud detection to improve the ability of the algorithm to distinguish between cloud regions and highlighted snow regions. Chethan et al. [43] also used Gabor transform to extract texture features for cloud detection primarily. Because Gabor transform can extract various texture features that are important for cloud detection, we have introduced the Gabor feature extraction module into the network in order to enhance the learning ability of the network for texture features. The network continuously down samples the feature map using the Max pooling layer to enhance the receptive field of the network, but, at the same time, the detailed information, such as the texture information, is lost, as depicted in Figure 1. Therefore, the Gabor feature extraction module is added before the first Max pooling layer in order to efficiently enhance the ability of the network to extract the texture information.

In the encoder–decoder structure, the encoder is primarily used for feature extraction and the decoder interprets the information to obtain the final result of the network recognition. The efficient interpretation capability of the decoder will greatly affect the final output of the network. Thus, inspired by the attention mechanism, we have introduced a spatial attention module that is based on Dark channel subnet and Channel attention module based on the higher-level features of the encoder to assist the decoder to enhance the key information, thereby improving the recognition ability of the decoder.

### 2.2. Gabor Feature Extraction Module

The kernel function of the Gabor wavelet is very similar to the stimulus-response of the receptive field cells in the primary visual cortex of mammals [44]. Features that were extracted using the Gabor wavelets can overcome the effects of illumination, scale, angle, etc. [45]. The Gabor transform has good local characteristics in both spatial and frequency domains [46]. A 2D-Gabor transform kernel function is expressed as:

$$\begin{gathered} g(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right) \\ \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \sin\theta & \cos\theta \\ -\cos\theta & \sin\theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \end{gathered} \tag{3}$$

where $\lambda$ represents the wavelength, $\theta$ represents the directionality of wavelet, $\psi$ is the phase offset, $\sigma$ is the standard deviation of the Gaussian envelope, and $\gamma$ is the spatial aspect ratio, and it specifies the ellipticity of the support of the Gabor function. This function consists of real and imaginary parts. The real part of this function is expressed as:

$$g_{real}(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x}{\lambda} + \psi\right) \tag{4}$$

whereas the imaginary part is expressed as:

$$g_{imag}(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \sin\left(2\pi\frac{x}{\lambda} + \psi\right) \tag{5}$$

The real part predominantly extracts the texture features in the image, while the imaginary part predominantly extracts the edge information. Because the texture information extracted from the image is the main requirement, we only use the real part of this kernel function for filtering. In the function, $\theta$ determines the direction of the filter and, thus, the feature information in different directions can be extracted, depending on the different values taken by $\theta$. The range of the $\theta$ values is $[0, \pi]$ and the interval is $\frac{\pi}{8}$ in order to obtain the texture features in multiple directions as uniformly as possible. The value of $\lambda$ is usually greater than 2. Therefore, the range of values is $[2, 5]$, and the interval is 1. The size of the Gabor filters is $3 \times 3$. Deng et al. [32] and Chethan et al. [43] both used

Gabor transform to extract texture features for cloud detection primarily, and improve the ability of the algorithm in order to distinguish between cloud regions and highlighted snow regions. They both use eight different orientations. We have also selected eight different orientations. We have selected four wavelengths in order to obtain more multi-scale information. In the parameter selection of the Gabor filter, a total of 32 features are obtained for each superpixel, using eight different orientations and four different wavelengths. The filters of the different wavelength or different directions have been shown, and the filters in each row have the same wavelength, while the filters in each column have the same direction, as shown in Figure 2. These parameters have been able to extract most of the texture features [47]. A finer range division will extract more detailed features, but this will also make the network structure more complex.

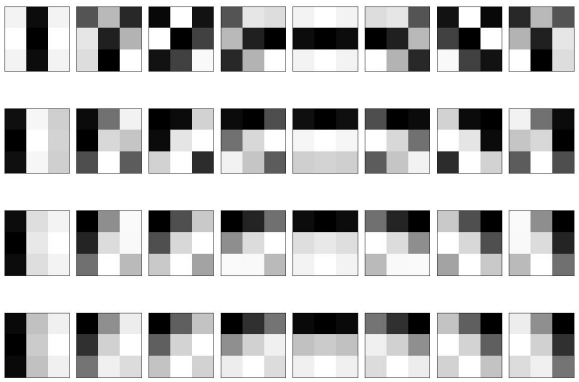


**Figure 2.** The filters of the Gabor transform.

The Gabor feature extraction module is divided into an upper and a lower branch, as shown in Figure 3. The upper branch is the texture extraction branch based on the Gabor transform, while the lower branch is a convolution information extraction branch. In general, we cannot know the type of specific features that are extracted by the convolutional layer with the activation function. However, in the upper branch, we can certainly know the specific texture features that are extracted from the feature map in multiple directions.

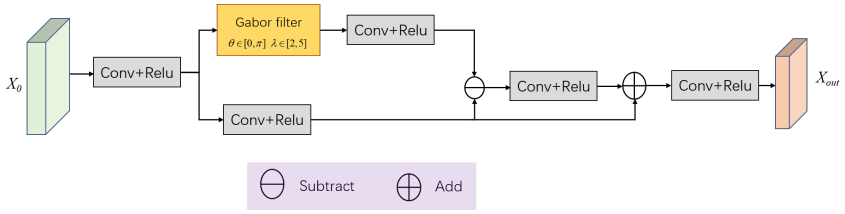



**Figure 3.** Block diagram of the Gabor feature extraction module.

The information output from the Gabor transform in the upper branch passes through a convolutional layer with the activation function once again in order to increase the variety of the texture features. The difference between the features that were obtained from the upper and the lower branch is then extracted by carrying out a subtraction between the feature maps. After passing through a convolutional layer with the activation function, the difference information is added back to the output feature map of the lower branch. In fact, the upper branch information can produce a local supervision effect on the lower branch information and guide the lower branch to learn and pay more attention to the texture information. In Gabor feature extraction module, the sizes of the convolutional filters are all $3 \times 3$. The output feature maps of the convolutional layer have the same size as the input feature maps.

### *2.3. Channel Attention Module Based on the Larger Scale Features*

Within the framework of the original encoder-decoder structure, feature maps of different scales at the encoder are introduced to the decoder in order to assist it in interpreting abstract information. However, the feature maps contain not only a large number of low-level features, but also a large amount of redundant information. Although the important low-level information can improve the performance of the decoder, the redundant information contained in it will interfere with the ability of the decoder to interpret the information. To reduce this, it is necessary for the information that is introduced into the decoder to be filtered properly. At the same time, there also is redundant information existing in the feature maps of the decoder.

In image processing that is based on deep learning, the attention mechanism is used to locate the salient features in the deep neural networks model. The attention mechanism is widely used in object detection [48] and image caption [49]. Thus, inspired by attention mechanism, we have proposed Channel attention module that is based on the larger scale features. Figure 4 shows a block diagram of the channel attention module.

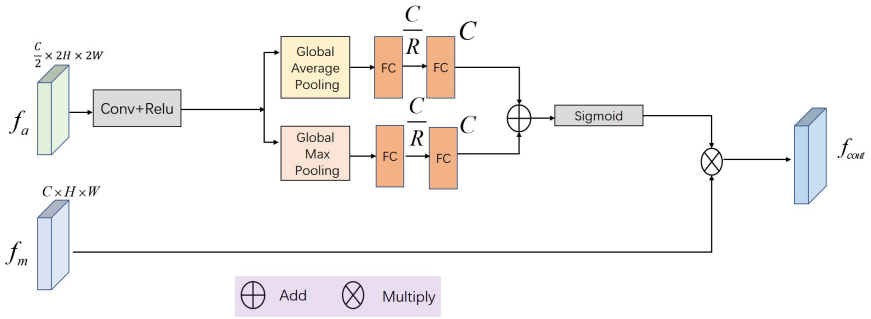



**Figure 4.** Block diagram of the Channel attention module.

There are two inputs of this module. $f_a \in \mathbb{R}^{C/2 \times 2H \times 2W}$ is the auxiliary feature map which is used for generating the weight map. $f_m \in \mathbb{R}^{C \times H \times W}$ is the feature map that needs to be reconstructed. The channel attention module is computed, as follows:

$$\begin{aligned} f_{\text{cout}} &= f_m \otimes M(f_a) \\ M(f_a) &= R\left(D\left(\text{AvgPool}\left(Conv\left(f_a\right)\right)\right) + D\left(\text{MaxPool}\left(Conv\left(f_a\right)\right)\right)\right) \\ &= R(W_1\left(W_0\left(\text{AvgPool}\left(Conv\left(f_a\right)\right)\right)\right) + W_1\left(W_0\left(\text{MaxPool}\left(Conv\left(f_a\right)\right)\right)\right)) \end{aligned} \tag{6}$$

where $R$ denotes the ReLU function, $D$ denotes the fully connected layer, $W_0 \in \mathbb{R}^{C \times C/R}$ and $W_1 \in \mathbb{R}^{C/R \times C}$, $AvgPool$ and $MaxPool$ denote the average pooling and max pooling operations, respectively. In the upper branch, the number of channels of $f_a$ is adjusted to be the same as $f_m$ using a convolution layer with an activation function. Subsequently, two sets of vectors are obtained by performing a global maximum pooling and global average pooling of the channel dimension. After that, the resulting two sets of vectors obtained via two fully connected layers are then added, and finally the weight map is obtained using the Sigmoid function as:

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{7}$$

The weight map and $f_m$ are multiplied to obtain the final reconstructed feature map. The information from the different channels of $f_m$ has different importance. In the channel attention module, each channel of $f_m$ can be associated with a weight to measure the importance of the information contained in it. The important information in $f_m$ is emphasized in $f_{cout}$ and the redundant information is weakened.

The Channel attention module is used at the decoder to reconstruct the feature maps. Thus, the auxiliary feature map of Channel attention module always comes from the information extracted

from themselves or the encoder, which contains features on the same scale. On the other hand, the larger scale features of the encoder are used as the auxiliary feature map in order to make better use of the low-level information.

### *2.4. Spatial Attention Module Based on Dark Channel Subnet*

We filter the information by adding Dark channel subnet and Spatial attention module in order to further remove the redundant information from the feature map at the encoder and highlight the important information. The Spatial attention module can perform spatial information screening of the feature map in each channel. The importance of the information that is contained in the different regions in a feature map is different [50]. A block diagram of the Spatial attention module is shown in Figure 5. In the figure, $f_a^{'}$ is the auxiliary feature map. On the one hand, it is used to generate a weight map. On the other hand, it directly supplements information in the lower branch. After a global maximum pooling and a global average pooling of the spatial dimensions, $f_a^{'}$ obtains two groups of feature maps with only a single channel. Two sets of feature maps with a single channel are obtained after performing a global maximum pooling of the spatial dimensions and a global average pooling process on $f_a^{'}$. These are concatenated in the channel dimensions and attain a weight map after a convolution layer with an activation function ReLU. The weight map and the feature maps that are generated by $f_a^{'}$ and $f_m^{'}$ are multiplied to obtain the final spatial-dimensionally reconstructed $f_{sout}$. The detail processing of Spatial attention module is expressed by Equation(8). Equation(8) is expressed as:

$$\begin{aligned} f_{sout} &= R(Conv^{3\times3}([R(Conv^{3\times3}([f_a^{'}; f_m^{'}])); f_a^{'}])) \otimes M(f_a^{'}) \\ M^{'}(f_a^{'}) &= R(Conv^{3\times3}([AvgPool^{'}(f_a^{'}); MaxPool^{'}(f_a^{'})])) \end{aligned} \tag{8}$$

where $R$ denotes the ReLU function. $Conv^{3\times3}$ denotes the convolutional layer with the filter size of $3 \times 3$. $AvgPool$ and $MaxPool$, respectively, denotes the average pooling and max pooling across the channel. The auxiliary feature map in Spatial attention module should have the ability to characterize the importance of feature maps from the encoder in the spatial dimensions. Therefore, it is necessary to introduce data from outside the encoder-decoder structure for effective information filtering.

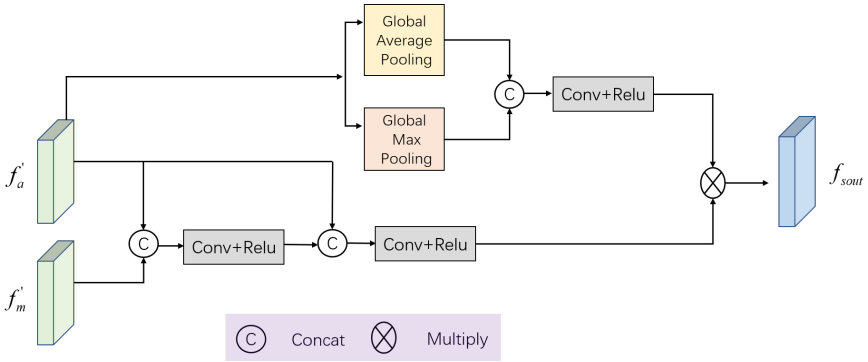



**Figure 5.** Block diagram of the Spatial attention module.

Spatial information screening is mainly related to color features. In a few rule-based cloud detection methods, a few color features are often extracted via operations between the different spectrum bands [15,51,52]. In the traditional cloud detection methods, it is generally preferred to convert the original image to a color space, in which the contrast between the cloud and non-cloud regions is more obvious. The color space transformation to the hue, saturation, and intensity (HSI) is always used in cloud detection [27,52,53]. However, we found that the Dark channel image is very similar to the cloud mask, as shown in Figure 6.

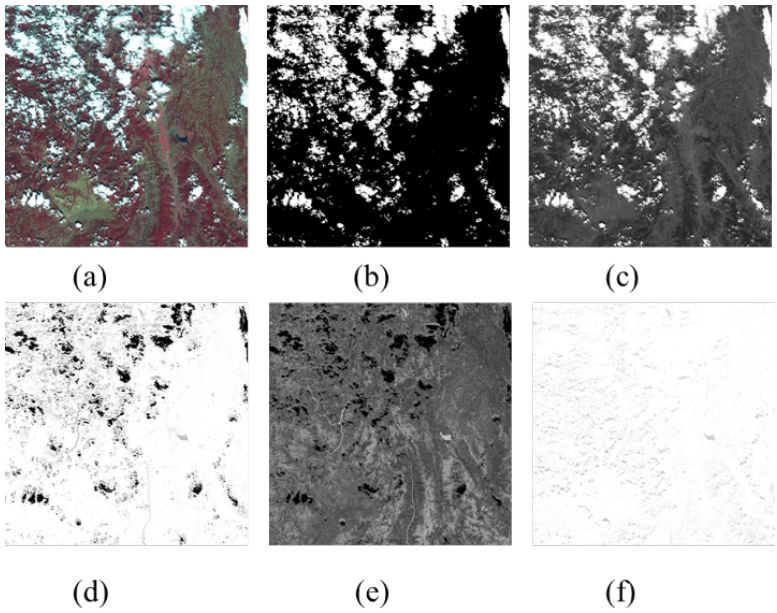



**Figure 6.** (**a**)NIR-R-G image, (**b**) ground truth, (**c**) Dark channel image, (**d**) H component of hue, saturation, and intensity (HSI), (**e**) S component of HSI, and (**f**) I component of HSI.

Thus, we have proposed the Dark channel subnet to generate the auxiliary feature map that is required by the spatial attention module. The input of Dark channel subnet is Dark channel image obtained from the three visible bands, red, blue, and green. The processing method is identified as:

$$f_{dark}(x,y) = \min_{c\in[r,g,b]} f(x,y,c) \tag{9}$$

Here, $f_{dark}(x,y)$ is the Dark channel image. $f(x,y,c)$ is the original remote sensing image that has at least three visible bands. The spatial attention module requires that the width and height of the auxiliary feature map need to be the same as the reconstructed map. In Equation (9), $f_{dark}(x,y)$ has the same width and height as the original remote sensing image. We use the same maximum pooling layer and convolutional layer with ReLU as the activation function in the Dark channel subnet in order to adjust the size of the auxiliary feature map. In the Dark channel subnet that is depicted in Figure 1, the size of the convolutional filters is $3 \times 3$ and the kernel size of Max pooling layer is $2 \times 2$. In the Dark subnet, we adopted the same strategy as the encoder, and the number of channels doubled after each downsampling. Therefore, the number of convolution kernels in the convolutional layer in the Dark subnet are two, four, eight, and 16, respectively.

## 3. Data Set and Evaluation Metrics

### *3.1. Data Processing*

The dataset that is used in the experiment is the open access GF-1 WFV imagery [37,39]. This collection of validation data includes 108 Level-2A scenes that were collected from different landscapes with varying cloud conditions, and all data have the corresponding cloud masks. This data set covers a variety of geomorphic environments, including urban areas, barren areas with little vegetation, areas covered by snow, areas covered by large amounts of vegetation, and oceans or lakes. The resolution of the images is 16 m, covering the visible and near infrared bands.

The approximate size of the images is $17000 \times 16000 \times 4$. Out of the 108 scenes, 86 were selected as the training data. The rest were selected as the test data. We evaluated the distribution of clouds in the dataset, and different kinds of cloud covers are included. Balance training and test data as much as possible. All of the images were rotated and clipped to the size of $11264 \times 11264 \times 4$ in order to

remove the surrounding black areas, as shown in Figure 7. This is because the black areas do not contain any remote sensing information, which is not helpful for feature extraction. At the same time, black areas are easy to be recognized by the network as a non-cloud area, which has a bad influence on the detection result. The training dataset was clipped to 41,624 small images by a stride of 512 × 512 in each image. The test dataset was clipped to 10,648 small images by a stride of 512 × 512 in each image. The value of the pixels in these patches was divided by 1023 in order to normalize between 0 and 1.

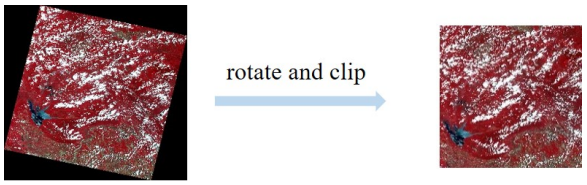



**Figure 7.** The resultant image (**b**) that was obtained by processing the original image (**a**).

### *3.2. Evaluation Metrics*

We have used the overall accuracy (OA), precision, recall, kappa, and false alarm rate (FAR) as the evaluation metrics in order to evaluate the performance of our proposed network. The OA indicates the overall accuracy of the network for cloud detection. Recall represents the ratio between the correct number of detected cloud pixels to the actual number of cloud pixels in the ground truth. FAR indicates the false detection of cloud pixels. Kappa is a robust parameter that measures the overall effectiveness of the network in cloud detection. The higher the value of Kappa, the better the cloud detection performance of the network.

The above-mentioned metrics are defined, as follows:

$$\begin{aligned} OA &= \frac{TP + TN}{TP + TN + FP + FN} \\ Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{P} \\ Kappa &= \frac{p_a - P_e}{1 - p_e} \\ p_a &= \frac{TP + TN}{TP + TN + FP + FN} \\ p_e &= \frac{P(TP + FP) + N(FN + TN)}{(P + N)^2} \\ FAR &= \frac{FP}{TN + FP} \\ mIOU &= \frac{Intersectionareasofdetectedandreferenceclouds}{Unionareasofdetectedandreferenceclouds} \end{aligned} \tag{10}$$

where $TP$ denotes the true positive outcomes, i.e., the number of cloud pixels that are correctly identified as correctly identified as cloud pixels in the generated mask, $TN$ denotes the true negative outcomes, i.e., the number of non-cloud pixels correctly identified as the non-cloud pixels in the generated mask, $FP$ denotes the false positive outcomes, i.e., the number of non-cloud pixels wrongly identified as cloud pixels in the generated mask, while $FN$ denotes the false negative outcomes, i.e., the number of non-cloud pixels identified as non-cloud pixels in the generated mask. $P$ denotes the number of cloud pixels in the ground truth and $N$ denotes the number of non-cloud pixels in the ground truth.

## 4. Results and Analysis

### *4.1. Evaluation of Gabor Feature Extraction Module*

The Gabor feature extraction module was primarily designed for guiding the encoder in order to learn the texture features. We removed the channel attention module and Dark channel subnet with Spatial attention module in NGAD and only kept Gabor feature extraction module. This network containing only the Gabor feature extraction module has been named NG. The feature maps directly obtained from the encoder are concatenated with the output of the up-sampling layer in the decoder. The structure of NG is very similar to that of U-Net [25], but there is an additional Gabor feature extraction module as compared to U-Net. Therefore, we compared the performance of NG with U-Net, which is a kind of a classic image segmentation network having an encoder-decoder structure.

The subjective results of U- Net and NG have been compared, as shown in Figure 8. The remote sensing image in the first row contains the snow-covered area, which has been marked by the red circles. Generally, a cloud-covered area has high brightness, and, in the figure, the snow-covered area also has high brightness. Thus, the two types of areas are hard to distinguish by the naked eye. However, the snow-covered area is relatively rougher than the cloud-covered area due to its rich texture feature.

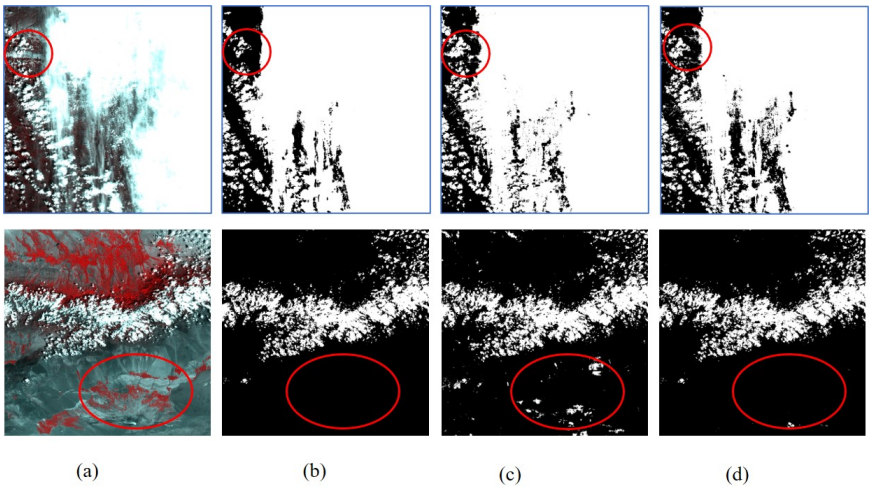



**Figure 8.** (**a**) NIR-R-G image, (**b**) ground truth, and masks generated by (**c**) U-Net and (**d**) NG.

It can be seen from the detection results presented in Figure 8 that NG has relatively few false detections because the texture information is much more efficiently utilized in NG than in U- Net. In the remote sensing image in the second row, a part of the area inside the red circle is barren. The brightness of some areas is relatively high, and this can interfere with the detection of clouds in the image. In terms of the cloud detection results, U-Net has a larger number of false detection areas when compared to NG. Therefore, it can be seen from the detection results of the two remote sensing images that Gabor feature extraction module guides the network to pay attention to the texture information, which is beneficial for the model to correctly identify the non-cloud regions that have high brightness and rich texture information.

U-Net and NG were, respectively, tested by applying them to the test set consisting of multiple scenes. Table 1 shows the average values of the different objective indicators. Under the condition that the precision of NG is higher than that of U-Net, the false detection rate of NG is lower than that of U-Net, as depicted in the table The kappa coefficient of NG is also higher than U-Net, which indicates that the cloud detection ability of NG consisting of the Gabor feature extraction module is better than that of U-Net. It can be seen from both the subjective and objective detection results that NG enhances the network's attention to texture information through the Gabor feature extraction module, which effectively improves the accuracy of the cloud detection model.

**Table 1.** Evaluation results with U-Net models and NG models on the test dataset.

| Method | OA (%) | Precision (%) | Recall (%) | FAR (%) | Kappa (%) | mIOU |
|---|---|---|---|---|---|---|
| U-Net | 96.73 | 89.89 | 89.66 | 4.46 | 82.31 | 0.83 |
| NG | 96.95 | 94.05 | 89.86 | 1.94 | 86.72 | 0.85 |

### *4.2. Evaluation of the Attention Mechanism Module and Dark Channel Subnet*

In the decoder, we proposed a Channel attention module based on the larger scale features. We have introduced the Dark channel subnet with Spatial attention module in order to further enhance the screening ability of the network and its attention to key information that are beneficial for cloud detection in space. We remove Gabor feature extraction module and Dark channel subnet with Spatial attention module within NGAD to obtain a network containing only the channel attention module (NC) in order to evaluate the effect of Channel attention module and Dark channel subnet with the Spatial attention module. Similarly, we remove Gabor feature extraction module from NGAD to obtain the network with Channel attention module and Dark channel subnet with Spatial attention module (NDSC).

For the sake of comparison, U-Net was used once again. The cloud detection results of the three models are basically close to the ground truth, as shown in Figure 9. However, the shape of the cloud region is irregular and complex, and it is difficult in many regions to judge with the naked eye whether the detection results of each model are consistent with the ground truth. A comparison of the Kappa coefficients obtained for each model indicates that both NC and NDSC perform better than U-Net, and NDSC is observed to be superior to NC. U-Net, NC, and NDSC were, respectively, tested by applying these networks to the test set that included multiple scenes.

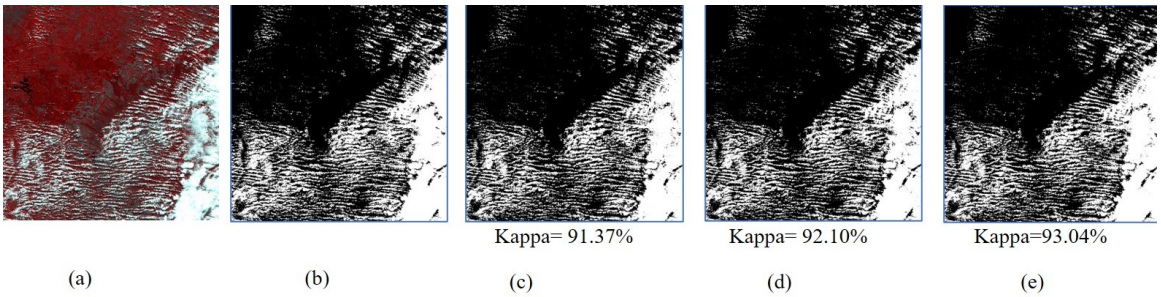



**Figure 9.** (**a**) NIR-R-G image, (**b**) ground truth, and masks generated by (**c**) U-Net, (**d**) NC, and (**e**) NDSC.

Table 2 shows the average values of the different objective indicators. From these, it can be seen that the overall accuracy of NC and NDSC is higher than that of U-Net. The recall rate of NC is slightly lower than U-Net, but its false detection rate is significantly reduced. This shows that the stability of NC is better than U-Net. Each objective indicator of NDSC has higher values than NC and U-Net, and the Kappa coefficient is higher by 1.96% when compared to that of NC. This shows that NDSC exhibits a significantly improved cloud detection ability as compared to NC. Thus, from these experimental results, it is clear that the key information in the network has been effectively enhanced by incorporating Channel attention module in it. On the basis of this module, the ability of the network to filter information is further improved after combining the Dark channel subnet with Spatial attention module.

**Table 2.** Evaluation results with U-Net, NC, and NDSC models on the test dataset.

| Method | OA (%) | Precision (%) | Recall (%) | FAR (%) | Kappa (%) | mIOU |
|---|---|---|---|---|---|---|
| U-Net | 96.73 | 89.89 | 89.66 | 4.46 | 82.31 | 0.83 |
| NC | 97.08 | 93.74 | 88.00 | 2.71 | 85.65 | 0.83 |
| NDSC | 97.39 | 94.05 | 90.28 | 2.37 | 87.61 | 0.85 |

### *4.3. Evaluation of NGAD*

Our proposed NGAD is composed of the Gabor feature extraction module, Channel attention module, and Dark channel subnet with Spatial attention module. The Gabor feature extraction module is used to guide the attention of the network to the texture information of an image. The channel attention module uses large-scale shallow information on the encoding side in order to reconstruct the information on the decoding side. The Dark channel subnet with Spatial attention module further improves the ability to filter information at the decoding end by introducing additional spatial auxiliary information.

We choose the SegNet [54], DeepLabv3+ [26], and RS-Net [29] comparison in order to evaluate the performance of the proposed cloud detection network, NGAD. Figure 10 shows the results that were obtained by applying the four models of cloud detection to remote sensing images containing five different types of land-cover.

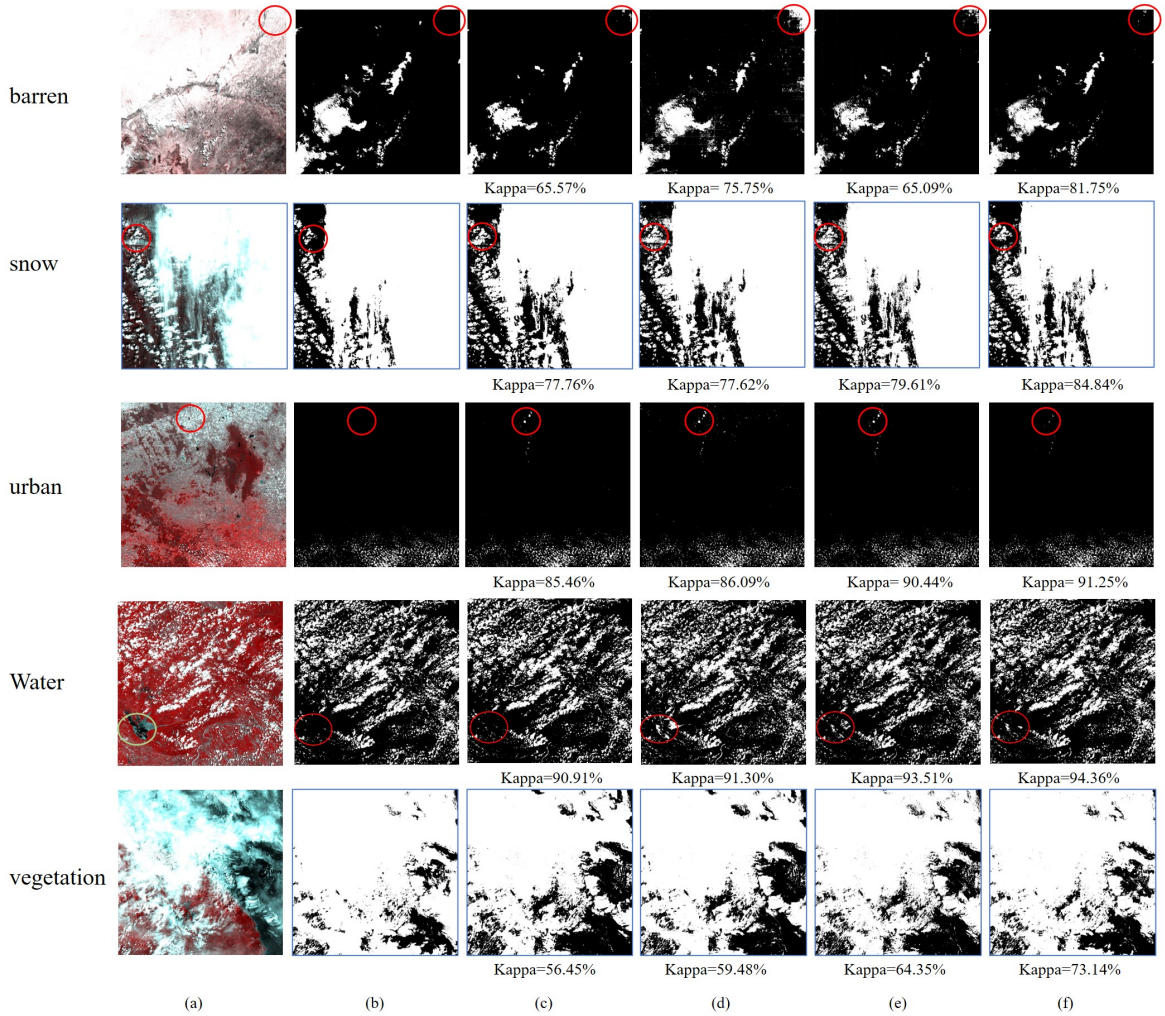



**Figure 10.** (**a**) NIR-R-G image, (**b**) the ground truth, and masks generated by the (**c**) SegNet, (**d**) DeepLabv3+, (**e**) RS-Net, and (**f**) NGAD.

In the first row, there is no object covering the ground, and some ground areas have high reflectivity similar to the cloud area. In the area that is marked by the red circle, the false alarm detections by the DeepLabv3+ and RS-Net are observed to be significantly higher than NGAD. Although these areas have high brightness, they have more texture information than the cloud areas in the figure. There are no false alarm detections in this area from the SegNet, while many cloud areas are missed from the overall view of the image. Therefore, it can be seen from the Kappa coefficient that the cloud detection ability of SegNet is poor. Similarly, results that correspond to RS-Net show obvious false alarm detections as well as obviously missed cloud detections. Thus, its overall detection ability is worse than SegNet. The results corresponding to NGAD show relatively few false alarm detections and missed detections in the cloud area and, thus, its overall detection ability is better than the other three models.

In the second row, the area inside the red circle is covered by snow, and such regions have high brightness, similar to the cloud area. However, the snow-covered ground is rougher. Because the texture information in NGAD is enhanced by Gabor feature extraction module, it helps to identify these highlighted non-cloud areas correctly. In the other areas in the image, there are thin as well as thick clouds, which makes cloud detection quite difficult. It can be seen from the Kappa coefficient that the detection ability of NGAD is better than the other three models.

In the third row, the image contains highlights of artificial buildings that can easily be erroneously recognized as clouds. There are a large number of scattered point cloud regions below the image, and all four models have been able to detect clouds in these regions. Based on the value of the Kappa, the overall detection performance of NGAD is observed to be better than the other three models.

In the fourth row, most of the vegetation area that is present in the image did not interfere with cloud detection. However, some smooth water-covered areas have been easily misidentified as clouds. In the circled area, the detection results of DeepLabv3+ and RS-Net have all false detections, whereas the results of SegNet and NGAD show fewer false detections. On the basis of the Kappa coefficient, NGAD is observed to be significantly better than SegNet in the overall cloud detection performance.

In the fifth row, a large number of the thick as well as thin clouds cover the water body, and some areas are difficult to distinguish by the naked eye. In the detection results of SegNet and DeepLabv3+, it can be seen that some cloud areas have been mistakenly identified as non-cloud areas. On the basis of the Kappa, the detection performance of NGAD is observed to be relatively stable and still better than the other three models.

The DeepLabv3+, RS-Net, and NGAD networks were tested by applying them to a test set that included multiple scenes. Table 3 sjows the average values of the objective indicators. It can be seen from the table that the false detections by the NGAD and SegNet are relatively low, which indicates that it is possible to effectively avoid interference with areas similar to clouds, such as ice and snow areas, bare ground highlight areas, etc. It can be seen from the values of OA and precision that the detection accuracy of NGAD is higher. By comparing the recall rates, it can be seen that NGAD can correctly identify a larger number of cloud areas efficiently. In terms of a lower false detection rate, the detection accuracy of NGAD is still higher, which indicates that NGAD is the most robust among the four tested models. From the comparison of the Kappa coefficient, the overall cloud detection performance of NGAD is observed to the best as compared to the other three models. Thus, this comparative study in terms of the subjective as well as objective aspects shows that NGAD exhibits excellent cloud detection performance.

**Table 3.** Evaluation results with SegNet, DeepLabv3+, RS-Net, and NGAD models on the test dataset.

| Method | OA (%) | Precision (%) | Recall (%) | FAR (%) | Kappa (%) | mIOU |
|---|---|---|---|---|---|---|
| SegNet | 96.02 | 94.25 | 84.05 | 1.38 | 82.73 | 0.80 |
| DeepLabv3+ | 96.18 | 91.31 | 85.99 | 3.25 | 82.37 | 0.79 |
| RS-Net | 96.71 | 94.34 | 87.92 | 3.97 | 84.74 | 0.83 |
| NGAD | 97.42 | 94.39 | 90.57 | 2.22 | 88.12 | 0.87 |

## 5. Discussion

### *5.1. Advantage Analysis*

The experimental results show that NGAD outperforms the reference methods of cloud segmentation. We believe that this maily depends on the CNN-based encoder-decoder structure, and the attention extraction module on this basis.

#### 5.1.1. Encoder-Decoder Structure

Long et al. [24] proposed Fully Convolutional Networks (FCN), which realized end-to-end semantic segmentation and is the pioneering work in semantic segmentation. A large number of excellent end-to-end semantic segmentation algorithms have emerged one after the other based on the FCN structure. U-net, SegNet, RS-Net, and Deeplabv3+ are based on the encoder-decoder structure, which exhibits excellent performance. The encoder-decoder structure is an efficient image segmentation network framework. The encoder in such a structure is used for extracting multiple features, while the decoder is used for interpreting the abstract information.

#### 5.1.2. Attention Module

A lot of useful information is lost during the repeated upsampling process, which leads to insufficient information interpretation. The problem of loss spatial information has led to a decrease in evaluation metrics. SegNet uses a novel upsampling strategy and added more information to the decoder; it decreases the loss of spatial information, but still leads to inaccurate bonudary definitions. U-Net concatenate the feature map of the encoding end to the decoding end to reduce the loss of spatial information. Inspired by attention mechanism, we have introduced modules that provide enhanced key information to the cloud detection network that is based on its network structure and the characteristics of cloud detection. The overall accuracy and Kappa of NC and NDSC is higher than these of U-Net, as we can see from Table 2. From these experimental results, it is clear that the key information in the network has been effectively enhanced by incorporating Channel attention module and Spatial attention module in it.

### *5.2. Limitations and Future Perspectives*

Our Gabor module and attention module can help the network perform feature extraction, but the network's features of cloud layer extraction still need to be further improved. There are still false detections on the boundary, due to the features of boundary clouds being difficult to extract.

In future work, we will continue to increase the utilization rate of the information extracted by the network. The aim will be to combine the characteristics of cloud detection and design a highly accurate information guidance module that is based on a large number of features extracted by the convolutional neural network. At the same time, we will continue to test and verify our developed network by applying it to different types of remote sensing imagery from different sensors.

## 6. Conclusions

Cloud detection methods for remote sensing images that are based deep learning use convolutional neural networks to extract a variety of image features at different scales. However, this feature information contains both information that is conducive to cloud detection as well as a large amount of redundant information. Therefore, it is necessary to further enhance the attention of the network to important information in order to reduce the impact of useless information on the cloud detection performance of the network. Traditional cloud detection methods often perform segmentation of cloud regions in remotely sensed images using the differences in the specific textures and color features between the cloud and the non-cloud regions. Therefore, the traditional feature extraction methods can be introduced into the network in order to guide it to learn important features.

At the same time, according to the characteristics of the network and the characteristics of cloud detection, an information screening module can be added to the network to effectively use the characteristic information extracted by it to further improve the network performance.

This paper proposes a cloud detection network for remote sensing images, called NGAD, which is based on Gabor transform, attention mechanism, and Dark channel subnet. The NGAD exhibits enhanced learning of the texture features, which is boosted due to the Gabor feature extraction module. The Channel attention module that is based on the larger scale features uses shallow feature maps with rich information to guide the network to interpret the abstract information, and screens the information from the channel dimension.

On the basis of the characteristics of cloud detection, the Spatial attention module based on Dark channel subnet, by establishing the Dark channel subnet and introducing the Dark channel image, is able to provide screening of feature information from the spatial dimension. The spatial attention module that is based on Dark channel subnet has been combined with the Channel attention module to filter the information in the network efficiently and comprehensively. The overall accuracy rate of NGAD is 97.42% and the false detection rate is 2.22%. The experimental results also show that NGAD improved cloud detection performance effectively.

**Author Contributions:** Conceptualization, J.Z. and Q.Z.; methodology, J.Z. and Q.Z.; software, J.Z., Q.Z., Y.W., H.W., J.W., Y.L., Y.C. and Y.L. (Yang Liu); writing–original draft preparation, Q.Z.; writing–review and editing, J.Z., Q.Z. and J.W.; supervision, Y.L. (Yunsong Li), Y.C. and Y.L. (Yang Liu). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Natural Science Foundation of China under Grant 61801359, Grant 61571345 and the Pre-Research of the "Thirteenth Five-Year-Plan" of China Grant 305020903.

**Acknowledgments:** We are very grateful for Zhiwei Li and Huanfeng Shen from the School of Resource and Environmental Sciences of Wuhan University for providing the GaoFen-1 validation dataset.

**Conflicts of Interest:** The authors declare no conflict of interst.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| MODIS | Moderate Resolution Imaging Spectrorodiometer |
| RS-Net | Remote Sensing Network |
| FCN | Fully Convolutional Networks |
| OA | Overall Accuracy |
| FAR | False Alarm Rate |

## References

1. 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]
2. Van Westen, C.J. Remote sensing and GIS for natural hazards assessment and disaster risk management. *Treatise Geomorphol.* **2013**, *3*, 259–298.
3. Adab, H.; Kanniah, K.D.; Solaimani, K. Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. *Nat. Hazards* **2013**, *65*, 1723–1743. [CrossRef]
4. Shi, T.; Xu, Q.; Zou, Z.; Shi, Z. Automatic raft labeling for remote sensing images via dual-scale homogeneous convolutional neural network. *Remote Sens.* **2018**, *10*, 1130. [CrossRef]
5. Shi, Y.; Qi, Z.; Liu, X.; Niu, N.; Zhang, H. Urban Land Use and Land Cover Classification Using Multisource Remote Sensing Images and Social Media Data. *Remote Sens.* **2019**, *11*, 2719. [CrossRef]
6. Li, Y.; Yu, R.; Xu, Y.; Zhang, X. Spatial distribution and seasonal variation of cloud over China based on ISCCP data and surface observations. *J. Meteorol. Soc. Jpn. Ser. II* **2004**, *82*, 761–773. [CrossRef]
7. Shin, D.; Pollard, J.; Muller, J.P. Cloud detection from thermal infrared images using a segmentation technique. *Int. J. Remote Sens.* **1996**, *17*, 2845–2856. [CrossRef]

8. Tapakis, R.; Charalambides, A. Equipment and methodologies for cloud detection and classification: A review. *Sol. Energy* **2013**, *95*, 392–430. [CrossRef]
9. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [CrossRef]
10. 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]
11. 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]
12. Jedlovec, G.J.; Haines, S.L.; LaFontaine, F.J. Spatial and temporal varying thresholds for cloud detection in GOES imagery. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1705–1717. [CrossRef]
13. Xiong, Q.; Wang, Y.; Liu, D.; Ye, S.; Du, Z.; Liu, W.; Huang, J.; Su, W.; Zhu, D.; Yao, X.; et al. A Cloud Detection Approach Based on Hybrid Multispectral Features with Dynamic Thresholds for GF-1 Remote Sensing Images. *Remote Sens.* **2020**, *12*, 450. [CrossRef]
14. Bai, T.; Li, D.; Sun, K.; Chen, Y.; Li, W. Cloud detection for high-resolution satellite imagery using machine learning and multi-feature fusion. *Remote Sens.* **2016**, *8*, 715. [CrossRef]
15. Zhang, J.; Zhou, Q.; Shen, X.; Li, Y. Cloud detection in high-resolution remote sensing images using multi-features of ground objects. *J. Geovis. Spat. Anal.* **2019**, *3*, 14. [CrossRef]
16. 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]
17. Tseng, D.C.; Tseng, H.T.; Chien, C.L. Automatic cloud removal from multi-temporal SPOT images. *Appl. Math. Comput.* **2008**, *205*, 584–600. [CrossRef]
18. Hagolle, O.; Huc, M.; Pascual, D.V.; Dedieu, G. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VEN$\mu$S, LANDSAT and SENTINEL-2 images. *Remote Sens. Environ.* **2010**, *114*, 1747–1755. [CrossRef]
19. Mateo-García, G.; Gómez-Chova, L.; Camps-Valls, G. Convolutional neural networks for multispectral image cloud masking. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 2255–2258.
20. Lu, X.; Zhang, Y.; Yuan, Y.; Feng, Y. Gated and Axis-Concentrated Localization Network for Remote Sensing Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 179–192. [CrossRef]
21. Li, Q.; Mou, L.; Jiang, K.; Liu, Q.; Wang, Y.; Zhu, X.X. Hierarchical region based convolution neural network for multiscale object detection in remote sensing images. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4355–4358.
22. Peng, C.; Li, Y.; Jiao, L.; Chen, Y.; Shang, R. Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2612–2626. [CrossRef]
23. Mou, L.; Zhu, X.X. RiFCN: Recurrent network in fully convolutional network for semantic segmentation of high resolution remote sensing images. *arXiv* **2018**, arXiv:1805.02091.
24. 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.
25. Ronneberger, O.; Fischer, P.; Brox, T. 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 October 2015; pp. 234–241.
26. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
27. 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]
28. Zi, Y.; Xie, F.; Jiang, Z. A cloud detection method for Landsat 8 images based on PCANet. *Remote Sens.* **2018**, *10*, 877. [CrossRef]

29. 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]
30. Li, Y.; Chen, W.; Zhang, Y.; Tao, C.; Xiao, R.; Tan, Y. Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning. *Remote Sens. Environ.* **2020**, *250*, 112045. [CrossRef]
31. Yu, J.; Li, Y.; Zheng, X.; Zhong, Y.; He, P. An Effective Cloud Detection Method for Gaofen-5 Images via Deep Learning. *Remote Sens.* **2020**, *12*, 2106. [CrossRef]
32. Deng, C.; Li, Z.; Wang, W.; Wang, S.; Tang, L.; Bovik, A.C. Cloud detection in satellite images based on natural scene statistics and Gabor features. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 608–612. [CrossRef]
33. Mnih, V.; Heess, N.; Graves, A.; kavukcuoglu, K. Recurrent models of visual attention. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2204–2212.
34. Li, L.; Xu, M.; Liu, H.; Li, Y.; Wang, X.; Jiang, L.; Wang, Z.; Fan, X.; Wang, N. A Large-Scale Database and a CNN Model for Attention-Based Glaucoma Detection. *IEEE Trans. Med. Imaging* **2019**, *39*, 413–424. [CrossRef]
35. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the International Conference on Machine Learning, Miami, FL, USA, 9–11 December 2015; pp. 2048–2057.
36. Li, J.; Feng, L.; Pang, X.; Gong, W.; Zhao, X. Radiometric cross calibration of gaofen-1 wfv cameras using landsat-8 oli images: A simple image-based method. *Remote Sens.* **2016**, *8*, 411. [CrossRef]
37. Li, Z.; Shen, H.; Li, H.; Xia, G.; Gamba, P.; Zhang, L. Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery. *Remote Sens. Environ.* **2017**, *191*, 342–358. [CrossRef]
38. Wu, X.; Shi, Z. Utilizing multilevel features for cloud detection on satellite imagery. *Remote Sens.* **2018**, *10*, 1853. [CrossRef]
39. 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]
40. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
41. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
42. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
43. Chethan, H.; Kumar, G.H.; Raghavendra, R. Texture based approach for cloud classification using SVM. In Proceedings of the 2009 International Conference on Advances in Recent Technologies in Communication and Computing, Kottayam, India, 27–28 October, 2009; pp. 688–690.
44. Bhate, D.; Chan, D.; Subbarayan, G. Non-empirical modeling of fatigue in lead-free solder joints: Fatigue failure analysis and estimation of fracture parameters. In Proceedings of the EuroSime 2006—7th International Conference on Thermal, Mechanical and Multiphysics Simulation and Experiments in Micro-Electronics and Micro-Systems, Como, Italy, 24–26 April 2006; pp. 1–7.
45. Zhang, T.; Zhang, P.; Zhong, W.; Yang, Z.; Yang, F. JL-GFDN: A Novel Gabor Filter-Based Deep Network Using Joint Spectral-Spatial Local Binary Pattern for Hyperspectral Image Classification. *Remote Sens.* **2020**, *12*, 2016. [CrossRef]
46. Xu, C.; Li, S.; Tan, T.; Quan, L. Automatic 3D face recognition from depth and intensity Gabor features. *Pattern Recognit.* **2009**, *42*, 1895–1905. [CrossRef]
47. Peng, B.; Li, W.; Xie, X.; Du, Q.; Liu, K. Weighted-fusion-based representation classifiers for hyperspectral imagery. *Remote Sens.* **2015**, *7*, 14806–14826. [CrossRef]
48. Zhu, Y.; Zhao, C.; Guo, H.; Wang, J.; Zhao, X.; Lu, H. Attention couplenet: Fully convolutional attention coupling network for object detection. *IEEE Trans. Image Process.* **2018**, *28*, 113–126. [CrossRef]
49. Yu, Y.; Choi, J.; Kim, Y.; Yoo, K.; Lee, S.H.; Kim, G. Supervising neural attention models for video captioning by human gaze data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July, 2017; pp. 490–498.
50. 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]

51. Sun, L.; Wei, J.; Wang, J.; Mi, X.; Guo, Y.; Lv, Y.; Yang, Y.; Gan, P.; Zhou, X.; Jia, C.; et al. A universal dynamic threshold cloud detection algorithm (UDTCDA) supported by a prior surface reflectance database. *J. Geophys. Res. Atmos.* **2016**, *121*, 7172–7196. [CrossRef]
52. Zhang, Q.; Xiao, C. Cloud detection of RGB color aerial photographs by progressive refinement scheme. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7264–7275. [CrossRef]
53. An, Z.; Shi, Z. Scene learning for cloud detection on remote-sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4206–4222. [CrossRef]
54. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [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/).