Cloud Detection for Satellite Imagery Using Deep Learning

Guo Yanan\textsuperscript{1,2*}, Cao Xiaoqun\textsuperscript{1,2}, Liu Bainian\textsuperscript{1,2}, Peng Kecheng\textsuperscript{1,2}

\textsuperscript{1} College of Computer, National University of Defense Technology, Changsha Hunan 410073, China
\textsuperscript{2} College of Meteorology and Oceanography, National University of Defense Technology, Changsha Hunan 410073, China
*Corresponding author’s e-mail: caoxiaoqun@nudt.edu.cn

Abstract. Cloud is the most uncertain factor in the climate system and has a huge impact on climate change. Therefore, the study of changes in cloudiness is of great importance for understanding climate and climate change. Cloud detection is also an important research area in satellite remote sensing image pre-processing. But cloud detection is a difficult task due to various noise disturbances in the remote sensing data itself, as well as factors such as ice and snow on the ground. With the rapid development of artificial intelligence technology, deep learning methods have achieved great success in methods such as image processing and classification. In this study, we use the modified U-Net architecture that introduces the attention mechanism for cloud detection. The experimental results show that the method proposed in this paper has a significant accuracy advantage over the traditional cloud detection method, especially in snowy areas and other areas covered by bright non-cloud objects. The effectiveness of this method makes it a great potential for other optical image processing as well.

1. Introduction
Cloud is the most uncertain factor in the climate system, with over 66\% of the Earth's surface regularly covered by clouds. Global cloud change is critical to climate and affects all aspects of the Earth's ecosystem [1]. Besides, clouds are an indicator of various weather phenomena. They occur during extreme weather events such as heavy rains and storms and can cause serious damage and threaten human life. The cloud also affects satellite observations of the ground, so the presence of the cloud reduces the availability of satellite data. Accurate cloud extraction from satellite imagery can help reduce the negative impact of the cloud on imagery applications [2,3]. Therefore, cloud detection is of great importance, both in geoscientific research and in applied satellite data research [4].

So far, scholars have done a lot of research on cloud detection of different types of remote sensing data and have achieved many excellent results. In recent years, deep learning techniques have achieved great success in image recognition, object detection, natural language processing, etc. The application of deep learning in the field of remote sensing is also ongoing and cloud detection methods based on deep learning are gradually emerging in recent research [5].

The study shows that deep learning has strong feature extraction and data mining capabilities, and can be used to achieve the task of cloud detection of satellite data using a limited band, and thus has a good application. To improve the accuracy of cloud detection, in this paper, we propose a deep learning-based approach for cloud detection, which introduces an attention mechanism [6-9] based on the U-Net architecture, thus paying more attention to cloud-related factors and ignoring invalid...
information during the cloud detection process. The proposed method shows a significant improvement in accuracy compared to the previous method.

2. Proposed Method

In the field of image segmentation, U-Net [10-11] is a well-known neural network architecture. Specifically, the U-Net consists of a typical down-sampled encoder and an up-sampled decoder structure and a “skip connection” between them. It combines local and global contextual information through an encoding and decoding process. Due to the excellent performance of U-Net, many remote sensing segmentation methods have been proposed in recent years based on U-Net. Despite the good performance of the U-Net architecture and its various variants, they are often complex and difficult to explain, so we introduce spatial attention in U-Net and propose a lightweight network model, which we name Spatial Attention U-Net (SAtt-UNet). Besides, batch normalization (BN) can improve the convergence speed of the network. Therefore, SAtt-UNet uses batch normalization (BN). Currently, the distinction between cloud and ice features is not evident in images with ground ice disturbance, especially between cloud and alpine ice areas. By introducing a small number of additional parameters, spatial attention can enhance the characteristics of the cloud and suppress the characteristics of non-cloud areas, thereby improving the performance of the network.

Figure 1 shows the structural diagram of our proposed SAtt-UNet, with the encoder on the left and the decoder on the right. As can be seen in Figure 1, each step of the encoder includes a structured convolution and a 2×2 max pooling operation. The convolutional layer is followed by batch normalization (BN) and rectifier linear unit (ReLU), followed by a maximum pooling operation for down-sampling with a stride size of 2. The number of feature channels is doubled after each down-sampling is completed. Each step of the decoder consists of a 2×2 transposed convolutional operation, which is used for up-sampling, halving the number of feature channels so that they can be fused to the feature map corresponding to the encoder, followed by a structured convolutional block operation. Besides, there is a spatial attention module between the encoder and the decoder. In the final layer, the final output partition is obtained using the 1×1 convolution and Sigmoid activation functions.

The spatial attention module is part of the convolutional attention module and is introduced into image classification. The spatial attention module uses spatial relationships between features to produce spatial attention maps. To compute spatial attention, we use max-pooling and average-pooling operations along the channel axis and combine them to obtain a valid feature descriptor. Figure 2 is the structural diagram of the spatial attention module.
3. Experiments and Results

3.1 Datasets and Preprocess
To test our proposed model, we use a publicly available dataset of Landsat 8 satellite images. Landsat 8 is one of the Landsat series of satellites, and compared to the previous Landsat satellites, Landsat 8 has more granular band delineation, so the data is more valuable to utilize. The data of Landsat 8 are currently being used in a wide variety of fields, so it makes practical sense to use the data of Landsat 8 to test models. This dataset is divided into a training set and a test set. Ground truths for all images in the training set were manually annotated and this dataset contains 8400 training patches and 9201 test patches.

3.2 Evaluation Metrics
We compare the segmentation results generated by the model with the corresponding ground truths. Each pixel is divided into two categories, clouded and non-cloud. The performance of our algorithms is measured quantitatively by evaluating the Overall Accuracy, Recall, Precision, Specificity, and Jaccard Index. These indicators are defined as follows:

\[
\text{Jaccard Index} = \frac{TP}{TP + FN + FP} \quad (1)
\]
\[
\text{Precision} = \frac{TP}{TP + FP} \quad (2)
\]
\[
\text{Recall} = \frac{TP}{TP + FN} \quad (3)
\]
\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (4)
\]
\[
\text{Overall Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)
\]

Here TP, TN, FP, and FN are the total number of true-positive, true-negative, false-positive, and false-negative pixels, respectively. The Jaccard index is a broadly adopted measure of the performance of many image segmentation methods.

3.3 Analysis of Experimental Results
We trained the model on the training set and then tested it on the test set. Figure 3 gives the results of several tests. The first column of Figure 3 is the RGB image, the second column is the ground truth, and the third column is the segmentation results of the method proposed in this paper. It can be seen that the segmentation results of this paper are very consistent with the ground truth, indicating that the proposed method is very effective. Table 1 gives a comparison of the test results of different methods on the test data set. As shown in Table 1, the method proposed in this paper performs better relative to other methods. Because these models are trained with the same training set. Therefore, the numerical
results of this experiment can illustrate the performance of different methods. The Fmask [12] method is a widely used cloud detection algorithm and the method proposed in this paper is superior to the Fmask method. Compared to the original U-Net architecture, the method presented in this paper also has better results, indicating the great potential of the method presented in this paper.

![Figure 3 Comparison of real images and segmentation results of the method in this paper.](image)

| Method      | Jaccard | Precision | Recall | Specificity | Overall Accuracy |
|-------------|---------|-----------|--------|-------------|------------------|
| FCN         | 71.94   | 84.16     | 81.92  | 97.04       | 94.47            |
| Fmask       | 74.26   | 80.43     | 93.31  | 94.17       | 93.56            |
| UNet        | 76.30   | 89.37     | 86.27  | 96.03       | 95.87            |
| Proposed method | 82.85   | 90.34     | 95.71  | 97.82       | 96.70            |

4. Conclusions
Cloud detection has always been an important part of satellite remote sensing data processing, but there are still many problems with cloud detection technology, such as non-cloud factor interference in complex environments. This paper presents a deep learning-based remote sensing image cloud detection algorithm that incorporates attention mechanism into the U-Net architecture. This improvement allows the model to focus more on cloudy regions in the cloud detection process and to resist interference from non-cloud factors. We compare the model with other cloud detection models and find that the method proposed in this paper has obvious advantages. We will further optimize the model in our future research and strive to apply it to the actual operational processing of satellite remote sensing data.
Acknowledgment
National key research and development program of China (Grant of No. 2018YFC1506704)

References
[1] Stephens G L. Cloud feedbacks in the climate system: A critical review[J]. Journal of climate, 2005, 18(2): 237-273.
[2] Francis A, Sidiropoulos P, Muller J P. CloudFCN: Accurate and robust cloud detection for satellite imagery with deep learning[J]. Remote Sensing, 2019, 11(19): 2312.
[3] Jeppesen J H, Jacobsen R H, Inceoglu F, et al. A cloud detection algorithm for satellite imagery based on deep learning[J]. Remote sensing of environment, 2019, 229: 247-259.
[4] Ozkan S, Efendioğlu M, Demirpolat C. Cloud detection from RGB color remote sensing images with deep pyramid networks[C]//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018: 6939-6942.
[5] Shi M, Xie F, Zi Y, et al. Cloud detection of remote sensing images by deep learning[C]//2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016: 701-704.
[6] Oktay O, Schlemper J, Folgoc L L, et al. Attention u-net: Learning where to look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.
[7] Schlemper J, Oktay O, Schaap M, et al. Attention gated networks: Learning to leverage salient regions in medical images[J]. Medical image analysis, 2019, 53: 197-207.
[8] Lian S, Luo Z, Zhong Z, et al. Attention guided U-Net for accurate iris segmentation[J]. Journal of Visual Communication and Image Representation, 2018, 56: 296-304.
[9] He N, Fang L, Plaza A. Hybrid first and second order attention Unet for building segmentation in remote sensing images[J]. Information Sciences, 2020, 63(140305): 1-140305.
[10] Li X, Chen H, Qi X, et al. H-DenseUNet: hybrid densely connected U-Net for liver and tumor segmentation from CT volumes[J]. IEEE transactions on medical imaging, 2018, 37(12): 2663-2674.
[11] Guo M, Liu H, Xu Y, et al. Building Extraction Based on U-Net with an Attention Block and Multiple Losses[J]. Remote Sensing, 2020, 12(9): 1400.
[12] Qiu S, Zhu Z, He B. Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery[J]. Remote Sensing of Environment, 2019, 231: 111205.