Applied Mathematics, Modeling and Computer Simulation
C. Chen (Ed.)

doi:10.3233/ATDE220073

Research on Cloud Detection Method of Multi-Spectral Remote Sensing Image in Ice and Snow Area

Xing WEI[a], Qianwen MU[a], Hui LI [a], Ziwen ZHANG[b,1]
[a] *Chemical Geological Survey Institute of Liaoning Province, Jinzhou, 121007, China*
[b] *Guangzhou Maritime University, Guangzhou, 510725, China*

Abstract. In order to improve the accuracy of cloud detection in multispectral remote sensing images, this paper first proposes a deep learning framework to solve the problem of cloud detection accuracy in remote sensing images. This framework benefits from the Full Convolutional Neural Network (FCN), Cloud regions in Landsat 8 images are pixel-level labeled. Secondly, in view of the difficulty in distinguishing clouds from snow and ice regions during cloud detection, a method for removing snow and ice regions based on a gradient recognition method is proposed. Experiments show that a hybrid based on the above two methods (based on gradient recognition and deep learning) can improve the performance in the cloud recognition process, and the method can be automatically generated without manual correction. The Jeckard index and recall rate increased by an average of 4.36% and 3.62%, respectively.

Keywords. Remote sensing, Deep learning, Pest Forest Inspection.

1. Introduction

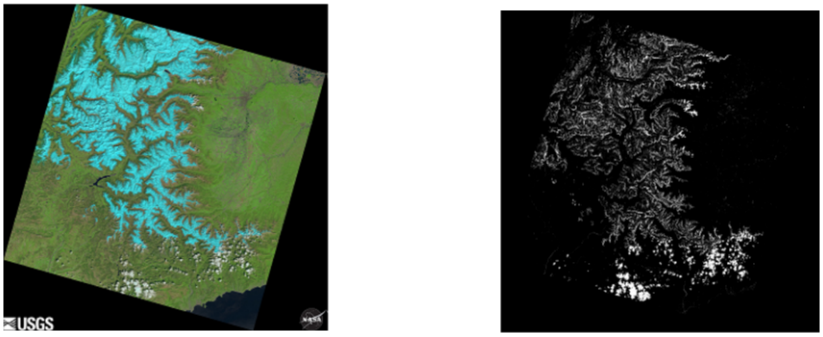
Satellite images often have cloud obscuration. The presence of clouds not only causes information loss in parts of the ground, but also causes difficulties in subsequent image production such as image registration, fusion, and correction. Cloud detection of remote sensing images has important practical production Significance[1]. Clouds have similar reflection characteristics to some other ground surfaces (such as snow, ice, and artificial white facilities), so detecting and identifying the clouds in the image and separating them from non-cloud areas is a problem in this research area . For clouds with multi-spectral bands, the cloud information can be determined and identified more accurately through other additional band reserve information (such as water, temperature, etc.). However, the data of many satellites (such as HJ-1 and GF-2) have the limitations of small spectral range and narrow range. When the channel of spectral band information is limited to red, green, blue, and near-infrared, The cloud's automated segmentation extraction will become more difficult.

[1] Corresponding Author, Ziwen ZHANG, Guangzhou Maritime University, Guangzhou, 510725, China; Email: zhzw2018@163.com.

The algorithms or methods related to remote sensing image cloud detection can be roughly divided into three categories: threshold-based segmentation methods[2-3], manual methods [4-5], and deep learning-based methods[6]. Object-oriented cloud detection algorithm (FMask) [7] and automatic cloud coverage assessment (ACCA)[8] algorithms are currently the most famous and reliable cloud segmentation algorithms based on threshold segmentation. Both methods use a decision tree to mark each pixel as cloud or non-cloud, mainly in each branch of the tree, using the results of one or more data spectral band threshold functions to make decisions. The cloud segmentation method based on threshold segmentation can not only count the cloud cover, but also has higher recognition accuracy than general methods. However, its disadvantage is that for images containing both clouds and snow, its detection accuracy is still very low. Haze optimization conversion (HOT) is one of the manual methods, which uses the relationship between the spectral response of two visible bands to separate haze and dense clouds from other pixels. Other manual methods, such as support vector machine methods, require sufficient sample training to obtain reliable cloud region features to ensure classification accuracy, and this method is easy to misjudge houses, roads, bare land, etc. as cloud regions.

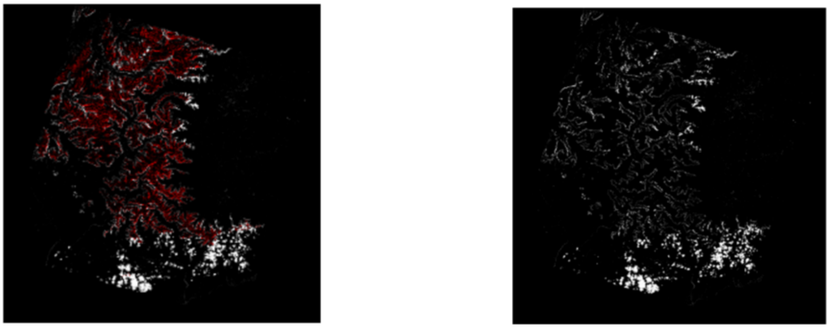
With the continuous development of deep learning algorithms for image segmentation, many methods for cloud detection using deep learning methods have been born. Xu Qiheng et al. classifierd to detect and identify cloud and non-cloud areas in high-resolution remote sensing image based on a convolutional neural networks[9]. This method first uses principal component analysis to unsupervise the pre-trained network structure to obtain the cloud characteristics of the remote sensing image to be measured, then uses the super pixel segmentation method to segment the image, and finally stitches the detection result image blocks to complete the entire image cloud detection.

The main problem of cloud detection methods based on deep learning is the lack of accurate marking of ground truth information. It turns out that most of the silent real information obtained through automatic or semi-automatic methods is not accurate enough. For example, people can easily mark areas with ice or snow as clouds. This incorrect ground information annotation has limited their use in the training set of new systems based on deep learning methods. Figure 1 shows an example of these errors in the default ground truth.



(a) Color image

(b) Cloud's default ground truth



(c) Mislabeled as cloud or snow (red) (d)Ground truth corrected using snow removal (ice) algorithm

Figure 1. Landsat 8 Remote Sensing Image Ground Annotation Example

Although the above methods show limited good results for scenes including thick clouds, they cannot provide reliable and accurate detection results in scenes where clouds and snow coexist.

In summary, this paper proposes a new method based on thresholds and deep learning to identify cloud regions and separate them from snow and ice regions in multispectral Landsat 8 images. In this paper, a threshold-based method is used to detect the snow-covered area by using Band 2 (blue light band) and image gradient in Landsat 8. This article first augments the existing Landsat 8 ground truth image by identifying icy / snow regions and then removing them from ground truth data used to train our deep learning system. The proposed deep learning system is a fully convolutional neural network (FCN), which uses cropped pictures of the training set images for training. The weights of the training network are used to detect cloud pixels in an end-to-end manner. Unlike FMask and ACCA, no blind spots will appear in this method whether it is a whole image or a partial image. In addition, because the system requires only four spectral bands for training and prediction (red, green, blue, and near-infrared), the architecture can be used intelligently for cloud detection of images acquired by all monitoring systems (satellite, airborne, etc. system). The flow chart of the research method in this paper is as follows in figure 2:

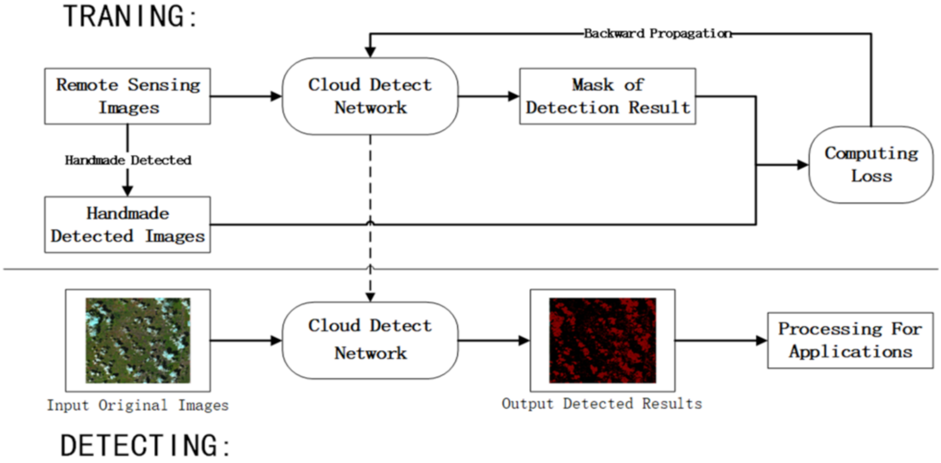


Figure 2. Processing framework

2. Theory and Method

Please follow these instructions as carefully as possible so all articles within a conference have the same style to the title page. This paragraph follows a section title so it should not be indented.

2.1. Experimental Data

Landsat 8 multispectral data consists of nine spectral bands collected from an Operational Land Imager (OLI) sensor and two tropical zones obtained from a Thermal Infrared Sensor (TIRS) sensor, each measuring a different wavelength range. Table 1 summarizes the parameter specifications for these frequency bands. In this article, we use only four spectral bands, Band 2 to Band 5. In addition, there is a quality assessment (QA) band, which was developed by the Landsat 8 cloud layer assessment (CCA) system and the FMask algorithm[10]. The default cloud / snow ground truth information of the image can be extracted from the QA band.

Table 1. Landsat8 data band parameters

Band	**Band range (μm)**	**Spatial resolution (m)**
1- Coast band	0.433–0.453	30
2- Blue band	0.450–0.515	30
3- Green band	0.525–0.600	30
4- Red band	0.630–0.680	30
5- Near infrared	0.845–0.885	30
6- Short-wave infrared 1	1.560–1.660	30
7- Short-wave infrared 2	2.100–2.300	30
8- Panchromatic band	0.500–0.680	15
9- Cirrus band	1.360–1.390	30
10- Thermal infrared 1	10.60 -11.19	100
11- Thermal infrared 2	11.50 -12.51	100

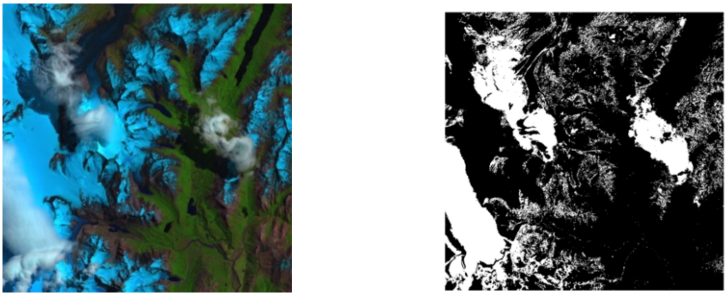
2.2. Image Snow Removal Processing Method Based on Gradient Recognition

2.2.1. Gradient Recognition Method to Remove Ice and Snow Areas.

In order to improve the accuracy of cloud region labeling for Landsat 8 training data, this paper proposes an image snow removal method based on the gradient recognition method. Figure 3 shows the original image diagram.

First, each spectral band image was divided into three different regions using Landsat 8 QA band information, namely: snow and ice regions, cloud regions, and sunny regions. Then get the gradient size of each pixel. Once calculated, the average image gradient magnitude for each of the snow, cloud, and clear areas can be determined.

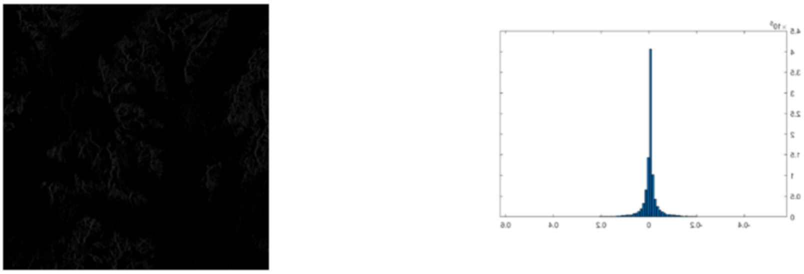
Figure 4 shows the average pixel gradient magnitudes of the snow, cloud, and clear areas in the image in the four bands of red, green, blue, and near-infrared. Comparing some averages of the four spectral bands, it can be found that there is a considerable difference between the snow-covered area and the rest of the image, and the maximum gradient difference between the average gradient of the snowfall area and the rest of the image under the blue light band shows the largest proportion difference. Its main manifestation is: in the non-ice and snow area of the image, the average gradient value of the blue light band is close to zero. In the snowfall area of the image, the average gradient of the blue light band is larger than that in other areas.



(a) Original image (b) Simple processing of cloud tags

Figure 3. Original image diagram

After determining the image gradient for Band 2, a global threshold is applied to isolate pixels with larger values and generate a binary snow mask. Figure 4 shows the schematic image of gradient interpretation, and figure 4 shows snow and ice mislabeling in this image. Figure 5 shows Snow and ice mislabeling.



(a) Gradient graph of 2 Band spectral (b) Gradient statistical histogram of 2 Band spectral gradient

Figure 4. Schematic image of gradient interpretation

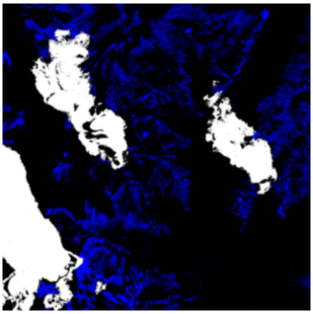


Figure 5. Snow and ice mislabeling

2.2.2. Cloud Image Preprocessing After Snow Removal.

After correcting the image ground truth scene (snow removal processing), pre-process the cloud areas in the image after removing the snow and ice interference. For an ideal single-band cloudy image, the histogram should have double peaks, as shown in figure 6 (b). A good threshold should correspond to the minimum value between the two peaks in the histogram. However, because the gray-scale histogram has large random fluctuations, the maximum value of the two peaks and the minimum value of the valley bottom between the two peaks cannot be well determined, so the histogram needs to be smoothed. Image smoothing is divided into two categories: spatial domain method and frequency domain method. In the frequency domain, various forms of low-pass filters are mainly used to eliminate noise in order to enhance certain frequency characteristics of the image, thereby changing the gray contrast between the feature object and the background [11]. Gaussian low-pass filtering can filter out gray deviation caused by isolated single-point noise, while effectively retaining image texture details [12]. Here Gaussian low-pass filtering is used for smoothing.

In image processing, Gaussian filtering is generally implemented in two ways, one is to use discretization window sliding window convolution [13], and the other is to use the Fourier transform method. In this paper, a discretized window sliding window convolution method is used to realize a moving window composed of a weight factor matrix or a coefficient matrix. These matrices are used as templates, and the size is usually an odd number of pixels. In this paper, a 5 × 5 matrix template is used.

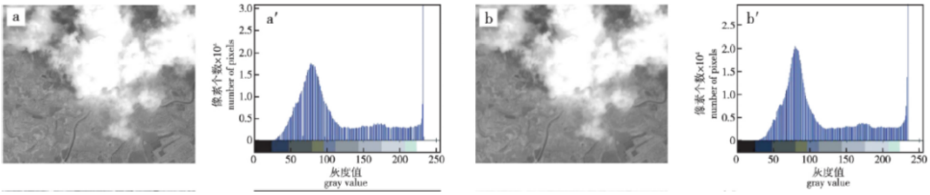


Figure 6. (a) Cloudy image; (b) Smoothed cloudy image; (a'), (b') are corresponding histograms

2.3. Cloud Detection Processing

After correcting ground truths, we will use them in a deep learning framework to identify cloud pixels in an image. In FCN, the spatial size of the output image is the same as the input image. This feature allows these types of CNNs to be used for pixel-level labeling tasks, such as image segmentation. The CNN proposed in this paper has an FCN architecture, which is inspired by U-Net [14]. In order to achieve the above purpose, the U-Net neural network-based remote sensing image cloud detection method proposed in this paper is to construct and train a UNET network that takes remote sensing images as input and outputs a mask of remote sensing image clouds. The construction of the network is to realize the conceived network structure with a deep learning platform. The training of the network includes the process of making data sets and training the parameters. At the same time, in the method of this paper, the data set does not need to be paired with pictures and labels like traditional supervised learning, but only needs to collect a large number of satellite remote sensing images and their one-to-one corresponding cloud detection mask images. The method implementation process is as follows:

1) Build a U-net network model, use the data set for training and tuning to get the final model;

2) Enter a remote sensing image of snow removal processing of any size into the U-net network model in this paper;

3) The network model performs cloud detection on the remote sensing image, and outputs the result of cloud layer detection, and the result is a picture;

4) Using the detection result as a mask, further operations can be performed on the original remote sensing image.

Specifically, the implementation of this method has the following steps:

(1) Building a UNET network model:

1) It consists of 5 downsampling layers and 5 upsampling layers (the first four downsampling layers are followed by a pooling layer and the last four upsampling layers are followed by a deconvolution layer), where the activation function is the ReLU function The UNET network has four jump connection chains, the fifth downsampling layer is connected to the first upsampling layer, and the last layer is connected to the fully connected layer output. The structure diagram is shown in figure 7.

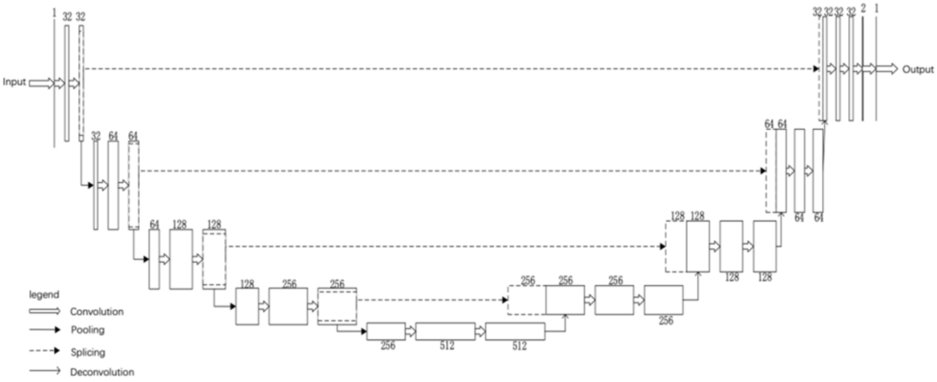


Figure 7. UNET structure framework

2) The loss function of this UNET network model is

$$E = \sum_{x\in\Omega} w(x) \log(p_{l(x)}(x)) \quad (1)$$

$$w(x) = w_c(x) + w_0 * \exp\left(-\frac{(d_1(x)+d_2(x))^2}{2\sigma^2}\right) \quad (2)$$

Among them, E is a cross-entropy calculation function, p_ (l (x)) is an approximate maximum function, w_c is a class frequency weight, and d_1 and d_2 are the closest and the penultimate distance of the pixel from the cloud region, respectively. Adjustable parameters w_0 and standard deviation σ. x is the picture input.

(2) Network training using remote sensing image datasets:

The remote sensing image data set is used to train the network using the Dropout method, and the network constructed above is repeatedly trained. After the number of training times reaches a preset threshold or the accuracy of the test reaches a target value, it indicates that the UNET network model has met the requirements of the invention..

(3) Tuning, optimizing, and retraining the model:

According to the effect of the test set in the model, adjust the parameters of the network model and retrain. Repeat step (3) until the effect reaches the desired expectation.

The image space size to be input for the training model network is set to 192x192x4 pixels. Since each spectral band size of Landsat 8 is very large (approximately 8000x8000 pixels), we must cut the image into smaller image patches. It is recommended that each spectral band image be cropped into 384x384 non-overlapping image blocks. Before training, these image blocks will be resized to 192x192 pixels. Then stack the red, green, blue, and near-infrared images to create a 4D input, and then network this input. In the last convolutional layer of the network, the output probability map is extracted by the sigmoid activation function, and then the Jeckard loss function is called and the network is optimized by reducing the Adam gradient.

$$L(h,y)=-\frac{\sum_{i=1}^{n} h_i y_i+\epsilon}{\sum_{i=1}^{n} h_i+\sum_{i=1}^{n} y_i-\sum_{i=1}^{n} h_i y_i+\epsilon}; \tag{3}$$

In the formula, h is the ground truth label, and y is the output probability map extracted from the last convolution layer through the sigmoid activation function. n is the total number of pixels of the labeled image. *yi* and hi are the *i*th pixel of y and h. Is a tiny real number to avoid division by 0.

3. Experiments and Conclusions

3.1. Data Set

This experiment uses Landsat 8 data package to make training data set. Landsat is the United States NASA's Land Satellite Program. Remote sensing images can be downloaded. This experiment uses remote sensing images from northwestern China for training, which includes the original image and cloud detection images labeled manually or by machine. Through the enhancement and expansion of the data, a certain number of picture sets are obtained, that is, a certain degree of rotation translation, elastic deformation, and gray value changes are performed to enrich the diversity of the data set and enhance the robustness of the model. The processed picture set is further divided into a training set, an evaluation set, and a test set according to 7: 2: 1.

3.2. Network Training

In this experiment, before the model training, the previously mentioned gradient recognition was used to remove the ice / snow from the image, and then two trainings were performed on the image before the ice / snow removal and the effect after the ice and snow removal. The training set obtained by the new method is compared with the manually created method training to highlight the improvements. Both the training set and test set images should be selected to cover many scene elements, such as vegetation, bare soil, buildings, urban areas, water, snow, ice, haze, different types of cloud patterns, etc., as well as in the training and test sets, The average percentage of cloud cover is about 50%.

After 600 iterations of the training process, the network converged to a local minimum. The obtained weights are used for target prediction. Before prediction, cut each spectral band in the test set into non-overlapping 384x384 pixel pictures, and then adjust these pictures to 192x192 and stack them together. After obtaining the cloud features corresponding to each picture, the output cloud probability map is adjusted to 384x384 pixels. These resized pictures are stitched together to create a cloud probability map of the entire image. Finally, through simple threshold processing, a binary cloud detection mask map of the input image is obtained. Figure 8 shows some visual examples of the shadow masks predicted in our test set sample images.

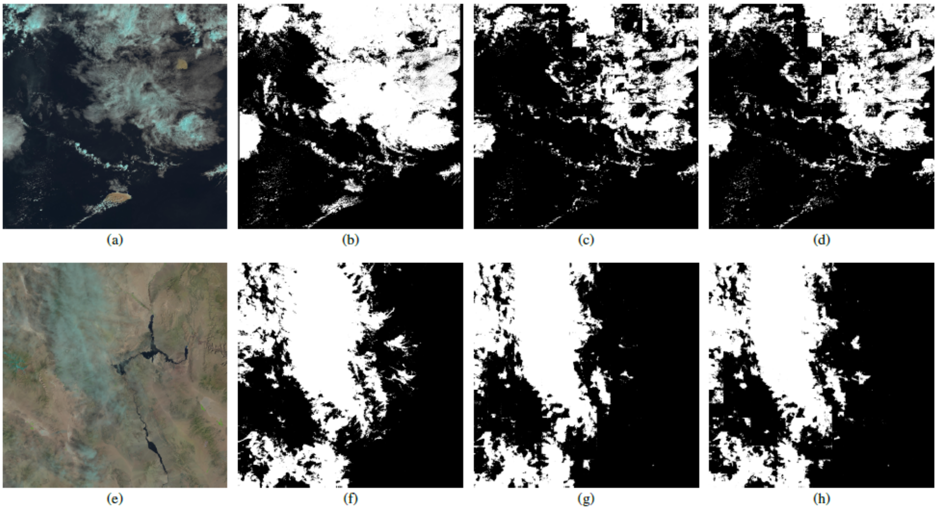


Figure 8. Example of cloud detection obtained by the proposed method.
(a), (e) true color input image, (b), (f) manual image annotation, (c), (g) are those without snow / ice correction Forecast cloud cover, (d), (h) Forecast cloud cover with snow / ice correction.

3.3. Evaluation of Detection Accuracy

This paper determines the performance of the method by evaluating the overall accuracy of the training results, recall rate, detection accuracy, and Jeckard similarity coefficient. The indicators are defined as follows:

$$\begin{aligned}
&\text{Jaccard Index} = \frac{TP}{TP+FN+FP}, \\
&\text{Precision} = \frac{TP}{TP+FP}, \\
&\text{Recall} = \frac{TP}{TP+FN}, \\
&\text{Overall Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}
\end{aligned} \tag{4}$$

Among them, TP is the number of true cloud pixels that can be accurately identified. TN is the number of non-cloud misjudged as a cloud pixel; FP is the number of pixels that are misjudged as a non-cloud by true cloud. Jeckard similarity coefficient is an index to measure the similarity of two sets.

Table 2 shows the experimental results of the proposed method. As shown in the table, the Jaccard index obtained by the method in this paper has increased by 4.36%. The recall rate has also increased by 3.62%. It shows that with the increase of clouds, the number of correctly labeled cloud pixels will increase, which proves the effectiveness of the proposed method to remove ice and snow + deep learning cloud detection framework.

Table 2. Comparison of system detection accuracy indicators (%).

How to use (Image # 1061)	Jaccard Index	Precision	Recall	Overall accuracy
FCN（without snow and ice removal）	62.63	72.59	79.39	87.81
FCN（with snow and ice removal）	65.36	73.54	82.26	88.30
Improvement rate	4.36	1.30	3.62	0.56

4. Conclusion and Outlook

This paper proposes a method based on deep learning that uses only four spectral bands of RGBNir to detect cloud pixels in Landsat 8 images. This method can accurately extract the semantic local and global features of the cloud in the image. Can be used for other segmentation tasks in satellite or airborne sensor remote sensing image applications.

After the cloud segmentation of remote sensing images, some roads and buildings still remain. Because the dimensions of buildings and roads are much smaller than the dimensions of clouds, the morphological opening operation completely deletes the object area that cannot contain structural elements, smooths the outline of the object, breaks the narrow connection, and removes the small protruding parts [15]. It is to first perform corrosion operations on the image to eliminate some isolated noise points, and then perform expansion operations to smooth the outline of the target. Therefore, you can open the cloud image to swallow smaller areas, thereby eliminating some roads and buildings, and smoothing the real clouds. The contours of the edges of the zone.

Acknowledgments

Thanks to the Landsat 8 data provided by NASA and USGS. This paper is supported by funds from the National Natural Science Foundation of China(Grant No.61974035).

Reference

[1] Zhu Z, Wang S, and Woodcock C E 2015 Improvement and expansion of the fmask algorithm: cloud, cloud shadow, and snow detection for landsats 47, 8, and sentinel 2 images *Remote Sensing of Environment* **159** pp 269 -77.

[2] Irish R R, Barker J L, Goward S, and Arvidson T 2006 Characterization of the landsat-7 etm+ automated cloud-cover assessment (acca) algorithm **72** pp 1179-88.

[3] Zhang Y, Guindon B, and Cihlar J 2002 An image transform to characterize and compensate for spatial variations in thin cloud contamination of landsat images *Remote Sensing of Environment* **82** pp 173-87.

[4] Yuan Y and Hu X 2015 Bag-of-words and object-based classification for cloud extraction from satellite imagery *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **8** pp 4197-4205.

[5] Xie F, Shi M, Shi Z, Yin J, and Zhao D 2017 Multilevel cloud detection in remote sensing images based on deep learning *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **10** pp 3631-40.

[6] D. of the Interior U.S. Geological Survey. Landsat 8 (l8) data users handbook. [Online]. Available: https://landsat.usgs.gov/documents/ Landsat8DataUsersHandbook.pdf

[7] Ronneberger O, Fischer P, and Brox T 2015 U-net: Convolutional networks for biomedical image segmentation *CoRR*, vol. abs/1505.04597.

[8] O¨ . C¸ ic¸ek, Abdulkadir A, Lienkamp S S, Brox T, and Ronneberger O 2016 3d u-net: Learning dense volumetric segmentation from sparse annotation in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds. Cham: Springer International Publishing, 2016, pp.424–432.

[9] H. Wang, X. Yang, L. Ma, and R. Liang, "Fingerprint pore extraction using u-net based fully convolutional network," in Biometric Recognition, J. Zhou, Y. Wang, Z. Sun, Y. Xu, L. Shen, J. Feng, S. Shan, Y. Qiao, Z. Guo, and S. Yu, Eds. Cham: Springer International Publishing, 2017, pp. 279–287.

[10] Nair V and Hinton G E 2010 Rectified linear units improve restricted Boltzmann machines in *Proceedings of the 27th international conference on machine learning (ICML-10)* pp 807–14.

[11] Waegeman W, Dembczy´nki K, Jachnik A, Cheng W, and H¨ullermeier E 2014 On the bayes-optimality of f-measure maximizers *J. Mach. Learn. Res.*, **15**, pp 3333–88.

[12] Yuan Y, Chao M, and Luo Y C 2017 Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance *IEEE Transactions on Medical Imaging* **36** pp 1876–86.

[13] Kingma D P and Ba J 2014 Adam: A method for stochastic optimization CoRR abs/1412.6980, 2014

[14] Chen G, E D C 2007 Support Vector Machines for Cloud Detection over Ice-snow Areas *Geo-spatial Information Science* **10** 117-20.

[15] Pan J J, Zheng X W, Sun L, Yang L N 2016 Image segmentation based on 2D OTSU and simplified swarm optimization *2016 International Conference on Machine Learning and Cybernetics (ICMLC)* **2** pp 1026-30.