A Multi-scale features-based cloud detection method for Suomi-NPP VIIRS day and night imagery

Jun Li, Chengjie Hu, Qinghong Sheng, Jiawei Xu, Chongrui Zhu, Weili Zhang

the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016

Keywords: Cloud detection, multi-scale feature extraction fusion, adaptive feature fusion, deep learning, daytime and nighttime remote sensing

Abstract

Cloud detection is a necessary step before the application of remote sensing images. However, most methods focus on cloud detection in daytime remote sensing images. The ignored nighttime remote sensing images play more and more important role in many fields such as urban monitoring, population estimation and disaster assessment. The radiation intensity similarity between artificial lights and clouds is higher in nighttime remote sensing images than in daytime remote sensing images, which makes it difficult to distinguish artificial lights from clouds. Therefore, this paper proposes a deep learning-based method (MFFCD-Net) to detect clouds for day and nighttime remote sensing images. MFFCD-Net is designed based on the encoder-decoder structure. The encoder adopts Resnet-50 as the backbone network for better feature extraction, and a dilated residual up-sampling module (DR-UP) is designed in the decoder for up-sampling feature maps while enlarging the receptive field. A multi-scale feature extraction fusion module (MFEF) is designed to enhance the ability of the MFFCD-Net to distinguish regular textures of artificial lights and random textures of clouds. An Global Feature Recovery Fusion Module (GFRF Module) is designed to select and fuse the feature in the encoding stage and the feature in the decoding stage, thus to achieve better cloud detection accuracy. This is the first time that a deep learning-based method is designed for cloud detection both in day and nighttime remote sensing images. The experimental results on Suomi-NPP VIIRS DNB images show that MFFCD-Net achieves higher accuracy than baseline methods both in day and nighttime remote sensing images. Results on daytime remote sensing images indicate that MFFCD-Net can obtain better balance on commission and omission rates than baseline methods (92.3% versus 90.5% on F1-score). Although artificial lights introduced strong interference in cloud detection in nighttime remote sensing images, the accuracy values of MFFCD-Net on OA, Precision, Recall, and F1-score are still higher than 90%. This demonstrates that MFFCD-Net can better distinguish artificial lights from clouds than baseline methods in nighttime remote sensing images. The effectiveness of MFFCD-Net proves that it is very promising for cloud detection both in day and nighttime remote sensing images.

1. Introduction

Satellite sensors are designed to collect electromagnetic wave data across various spectral bands, enabling the formation of satellite remote sensing images. These images are integral in executing a range of tasks, including feature identification, target detection, target tracking, and issuing disaster warnings. However, cloud cover significantly impacts the quality of satellite remote sensing imagery. According to a study by the International Satellite Cloud Climatology Project (ISCCP), clouds obscure about 67% of the Earth at any given time (Zhang et al., 2004). The efficacy of most remote sensing applications hinges on the precision and surface detail richness of these satellite images.

Over the past few decades, there has been considerable scholarly interest in the area of cloud detection in satellite remote-sensing images. Pioneering studies and advancements have been made (Li et al., 2017; Li et al., 2022). Presently, the techniques for cloud detection in remote-sensing images are broadly classified into four categories: physical rule-based, temporal phase-difference-based, machine learning-based, and deep learning-based methods (Li et al., 2022). These methodologies utilize a combination of spectral and spatial features, along with temporal data and integrated features, to effectively differentiate clouds from clear segments in satellite imagery.

1.1 Physical Rule-based Methods

Physical rule-based methods, initially employed for cloud detection in remote sensing imagery, capitalize on the distinct high reflectivity and low temperature characteristics of cloud layers. These methods implement specific rules tailored for different spectral channels to facilitate cloud detection. A notable example is the ISCCP cloud detection approach (Schiffer et al., 1983), which distinguishes clouds from clear skies based on the disparity in radiation between them in visible and near-infrared channels, using empirical thresholds for identification. Similarly, the Automatic Cloud-Cover Assessment (ACCA) methodology (Irish et al., 2006) leverages satellite spectral reflectance data from the second to fifth bands of the Landsat-7 ETM+ imagery. Physical rule-based methods mainly utilize cloud and surface spectral analysis to achieve cloud detection, which is simple and easy to use. However, the selection of physical rules for this type of method relies on empirical judgment and parameter sensitivity analysis. Moreover, owing to the complexity and variability of the surface environment, the diversity of cloud geometries, and the limited spectral information, it is usually difficult to determine the optimal thresholds with full consideration of the influencing factors, which leads to varying degrees of cloud-coverage estimation bias.

1.2 Temporal Phase-Difference-based Methods

In time-series imagery of the same geographical area, cloud coverage often results in abrupt shifts in reflectivity. Consequently, temporal phase-difference-based methods typically identify clouds by assessing the reflectivity contrast between cloud pixels and clear-sky pixels, or simulated clear-sky pixels. Pixels exhibiting reflectivity exceeding a predetermined reference value are classified as clouds (Zhu et al., 2012). The foundational principle of these methods is the differential analysis between pairs of images. For instance, Wang et al. (1999) initially employed a histogram-based

https://doi.org/10.5194/isprs-annals-X-1-2024-115-2024

approach to broadly distinguish between cloud-covered and clear-sky areas. Subsequently, they integrated this technique with a differential analysis of Landsat TM (Thematic Mapper) images, taken at different times, to enhance the precision of cloud detection.Another notable method, the Tmask approach (Zhu et al., 2014), exploits time-series data from Landsat to develop a model that predicts apparent reflectance. It then compares this predicted reflectance with the actual measured apparent reflectance of the pixel under examination for cloud detection. Methods based on temporal phase differences, which merge spectral and temporal data, can notably improve cloud detection accuracy. They are particularly effective in reducing the misidentification between bright surfaces and clouds. However, the efficacy of these methods is heavily contingent upon the time series data quality. Challenges arise in regions with significant surface type variations over time, leading to potential misclassifications of clouds and surface features.

1.3 Machine-learning-based Methods

Machine learning-based methods for cloud detection approach the task as a binary classification problem. These methods extract spectral and spatial features from a large dataset of training samples, iteratively refining and optimizing model parameters to develop robust classifiers for cloud identification. In traditional machine learning, the methods are broadly divided into supervised and unsupervised learning categories. Supervised learning encompasses techniques like Support Vector Machine (SVM) (Ma et al., 2017; Joshi et al., 2019), Decision Trees (Jang et al., 2012), Bayesian classifiers (Xu et al., 2017), and Random Forest (RF) (Wei et al., 2020). On the other hand, unsupervised learning includes methods such as the Principal Component Analysis (PCA) (Mackie et al.,2010), and Singular Value Decomposition (SVD) (Hurley et al., 2009), all of which have seen extensive application in cloud detection research within the realm of remote sensing. Machine learning-based methods leverage both spectral and spatial attributes of clouds and surfaces for automated cloud detection. Compared to physical rule-based and temporal phase-difference methods, machine learning approaches offer greater efficiency in implementation and automation, enhancing cloud detection accuracy to some degree. Nevertheless, in vast and highly heterogeneous regions, the generalization capabilities of traditional machine learning models can be limited. As a result, cloud identification performance may vary across different scenes.

1.4 Deep-learning-based Methods

In recent times, deep learning networks have gained significant traction in cloud detection tasks, demonstrating superior performance over traditional methods. Notable examples include the U-Net (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017), and DeepLabV3+ (Chen et al., 2017). These deep learning-based cloud detection methods not only enhance cloud recognition effectiveness but also boost the efficiency of cloud detection models and simplify the detection process (Yang et al., 2019; Li et al., 2019; Li et al., 2022).

However, purely convolutional network architectures often fall short in capturing global contextual information, a critical aspect when dealing with targets varying greatly in texture, shape, and size. To address this challenge, research efforts have focused on enlarging the receptive field to better capture contextual details. For example, Chen et al. (2017) introduced the concept of dilated convolution in DeepLab networks to expand the receptive field, coupled with a conditional random field (CRF) to enhance detail capture in the model. Similarly, Li et al. (2022) developed the Global Context Dense Block (GCDB UNet) within the UNet architecture, effectively improving the detection of thin clouds by integrating global context dense blocks. Additionally, Shi et al. (2019) presented an Enhanced Fully Convolutional Network (EFCN) based on the VGG-16 framework, substituting VGG-16’s fully connected layers with convolutional layers and employing up-sampling predictions for improved detection results.

1.5 VIIRS DNB

Launched in 2011, NPP is NASA's satellite dedicated to global environmental observation, equipped with the VIIRS (Visible Infrared Imaging Radiometer Suite) as its primary multispectral imager. VIIRS captures radiometric imagery of the atmosphere, land, and oceans across various channels, including visible and infrared bands. The sensor features 22 bands: five I-bands with a 370-meter spatial resolution, one Day/Night Band (DNB), and sixteen M-bands with a 750-meter spatial resolution. The DNB on the VIIRS sensor represents a significant advancement in remote sensing technology, capable of capturing detailed Earth's surface imagery under both day and night conditions.

The primary distinctions between day and night imaging modes lie in the lighting conditions and the adaptability of the imaging technology. During the day, the DNB utilizes sunlight as the main source of illumination to record reflected radiation from the Earth's surface and atmosphere. In contrast, the DNB's night mode showcases its exceptional low-light imaging capabilities, relying on moonlight, starlight, and terrestrial artificial light sources. Nighttime images, while dimmer and less contrasted compared to daytime, still provide valuable insights into human activity patterns and geographical distribution. In analyzing DNB images from the VIIRS sensor, manmade lights are typically characterized by bright spots with clear boundaries and regular shapes, reflecting human activities like road networks and urban centers. Conversely, clouds in remote sensing images often display irregular shapes and diffuseness, covering extensive areas with blurred edges. Spectrally, even in grayscale, artificial lights manifest as uniformly bright areas, whereas cloud-reflected light appears darker, with lower brightness and a more consistent tonal range. Leveraging these characteristics, algorithms can be tailored to effectively focus on and distinguish between these disparate elements in nighttime remote sensing images.

Although recent advancements in cloud detection using satellite remote sensing images have been considerable, there remain several areas requiring further exploration and improvement. Currently, key challenges in deep learning-based cloud detection of remote sensing images include:

1) Owing to the complexity of the cloud structure, different scales of clouds have different boundary scales, and existing deep-learning-based methods tend to discriminate clouds and backgrounds within the global scale, which leads to the model's omission and misdetection of fragmented clouds as well as clouds with smaller scales;

2) Since the VIIRS DNB image is a single-band image, the color features of clouds and lights may be similar, which increases the difficulty of differentiation in a nighttime image;

3) Fewer explorations of existing methods have aimed at the detection of night-lighting remote sensing clouds. Moreover, there is no publicly available cloud-masked dataset of night-lighting remote sensing imagery.

The information mining ability of a deep learning model is one of the factors affecting cloud detection accuracy, so mining and using more effective information are important ways to improve cloud detection accuracy. Aiming at addressing the above-mentioned problems, we propose MFFCD-Net in this paper to realize cloud detection of daytime and nighttime remote sensing images. MFFCD-Net is implemented through an encoder-

decoder architecture that acquires the multi-scale information of the image through the multi-scale feature extraction module (MFEF Module) and associates the shallow detail features with the distinguishing features of lights and clouds through the Global Feature Recovery Fusion Module (GFRF Module) to achieve a better cloud detection effect. The main contributions of this paper are as follows:

1) In this paper, the designed MFEF module is added at the lowest end of the encoder to extract image multi-scale information and select noteworthy cloud information from different scale feature maps to improve the network's ability to focus on the spatial background region that is most relevant to clouds. In addition, the designed DR-UP Block is used in the decoding stage to expand the network sensory field so that the network can capture cloud features at different scales;

2) The MFEF module can capture the large-scale features of urban lights and more dispersed lights in rural areas, filter and optimize the multi-scale feature maps, strengthen the model's ability to recognize regular textures of urban lights and random textures of clouds, and reduce the misdetection rate of lights. The edge-detail information in the encoding stage and the feature information after filtering out lights in the decoding stage are fused by the GFRF module to achieve better cloud boundary segmentation accuracy and lighting discrimination;

3) Aiming at the lack of publicly available nighttime remote sensing cloud mask datasets, we produced day-night remote-sensing cloud mask datasets by manual labeling. Remote sensing data captured by the day-night band (DNB) of the Visible Infrared Imaging Radiometer Suite (VIIRS) on board the Soumi NPP satellite were mainly used, and a total of 14,894 256 × 256 cloud mask images were obtained, including 2,708 nighttime remote sensing images.

2. Method and Experiment

2.1 Overview of the MFFCD-Net

The overall architecture of MFFCD-Net designed in this paper is shown in Fig.1. MFFCD-Net is realized based on an encoder-decoder architecture. The encoder adopts Resnet50 as the backbone network to extract the features at different layers, and the decoder recovers the image features and dimensions from the fused multi-scale feature maps by means of the dilated residual module up-sampling (DR-UP Blocks). Each DR-UP Block is composed of three convolutional layers with different dilated rates and one up-sampling layer, and the DR-UP Block expands the feature-sensing field of the network while recovering the image size. The MFEF module is added to the bottom layer of the encoder to mine the multi-scale information of the feature map and improve the network's ability to recognize the fragmented clouds as well as the clouds with different boundary scales. To better realize the combination of shallow and deep features, we replaced the traditional simple jump-joining structure with an GFRF, which fuses the edge detail information of different scales and the feature maps after the MFEF module processing to discriminate the lights, thus improving the cloud-boundary segmentation accuracy and lighting discrimination effect.

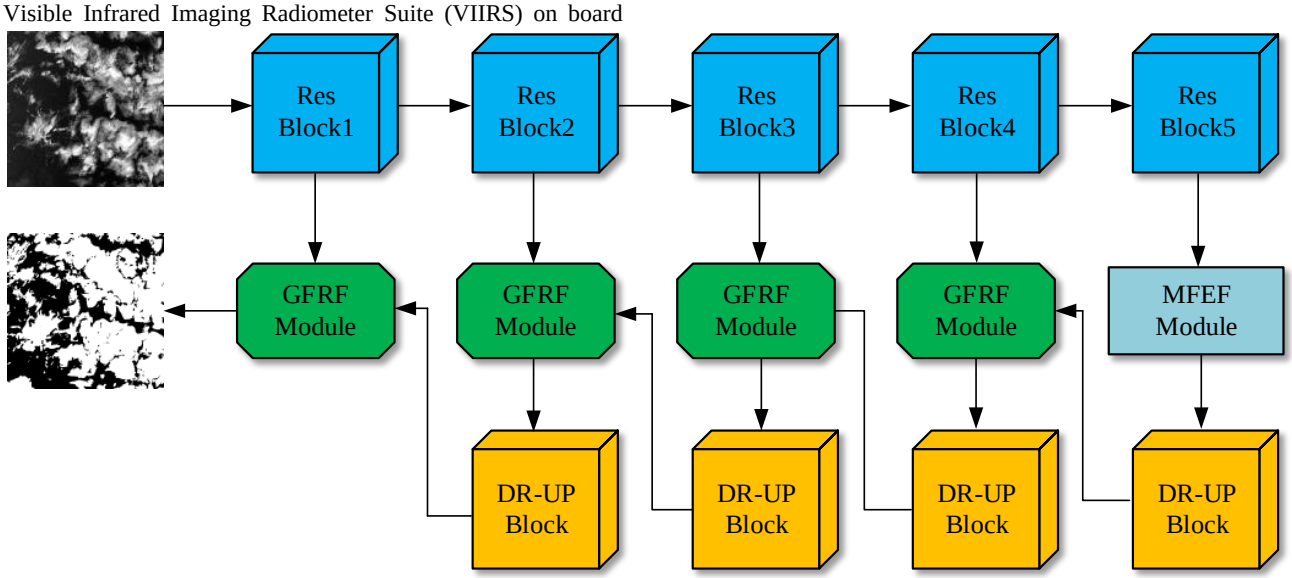

Fig.1.The overall architecture of MFFCD-Net.

2.2 DR-UP Block

Traditional convolution aims to use a convolution kernel to perform a convolution operation on the neighboring elements in the tensor, while the cavity convolution can be convolved on two non-neighboring elements, which can increase the convolution kernel's receptive field and reduce the amount of computation. Different expansion rates have different receptive fields, and multi-scale features can be extracted so that the network can capture cloud features at different scales. The DR-UP Block was designed in the decoder by combining the cavity convolution and up-sampling layers, and the DR-UP Block consists of three cavity convolution layers and one up-sampling layer. The same dilated rate of adjacent cavity convolutional layers leads to the problem of a discontinuous convolution center. To solve this problem, we designed the dilated rate as a cyclic sawtooth structure of [1, 2, 5]. Fig. 2 shows the structure of the DR-UP Block designed in this paper.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-1-2024-115-2024 | © Author(s) 2024. CC BY 4.0 License.

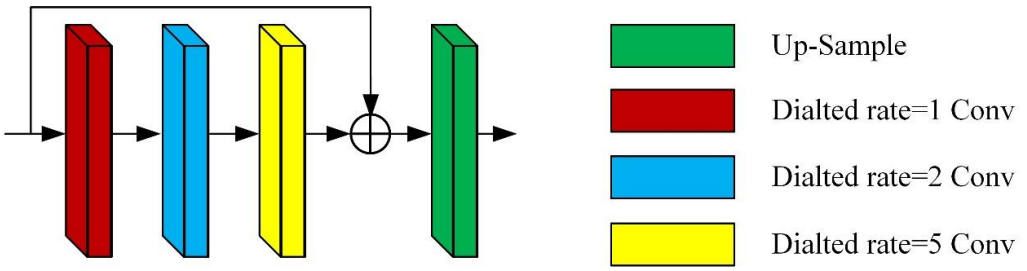

Fig. 2. Structure of DR-UP block

2.3 MFEF module

In nighttime remote-sensing-image cloud detection, the high radiation intensity of artificial lights and their similarity in color to clouds on the image often lead to their misidentification as clouds. However, the regular arrangement and consistent brightness of lights provide discriminative features between lights and clouds for nighttime cloud detection. In view of this, this study designed an MFEF module, which consists of two parts: a multi-scale spatial pyramid and a multi-scale feature selection module. In the multi-scale spatial pyramid, multilevel features are extracted through hollow convolutional layers with different expansion rates and global pooling operations, allowing the network to capture large-scale urban lights and more dispersed rural area lights. Subsequently, the multi-scale feature selection module is used to filter the multi-scale feature maps and optimize the model to recognize the regular texture of city lights and the random texture of clouds. The MFEF module realizes the distinction between lights and clouds, significantly improves the accuracy of cloud detection in nighttime remote sensing images, and provides new perspectives and methods for solving the similar remote sensing image processing problems. In addition, the MFEF module designed in this paper not only performs well in nighttime cloud detection but likewise optimizes the ability to capture clouds at different scales during daytime. This is attributed to its multi-scale spatial pyramid and feature selection mechanism, which enables the module to accurately recognize and distinguish clouds of different scales and types. This advancement significantly improves the network's cloud identification and classification performance in daytime environments, and it enhances the model's application flexibility and robustness under different environmental conditions.

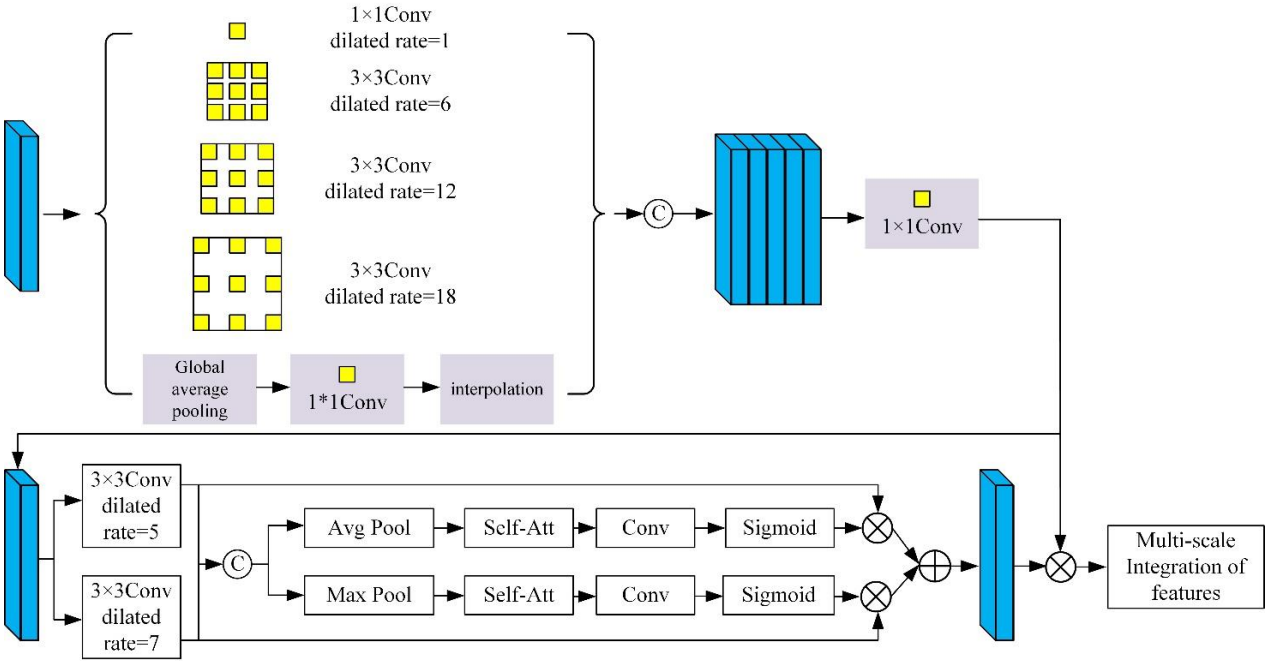

Fig. 3. Structure of MFEF module

2.4 GFRF module

Global features and inter-pixel correlation in the image play an important role in target recognition, and the lack of global information leads to the loss of effective recognition features for clouds and background, affecting cloud detection accuracy. Since the feature map processed by the encoder residual module may be from background information, and although the feature map processed by the MFEF module can achieve better separation of clouds and the background, the use of the dilated convolution in the decoder results in the loss of some of the detailed features. This lack of information leads to the model encountering difficulties in recognizing the subtle differences between the cloud and the background, which affects the segmentation of the cloud boundary effectiveness. In the decoding stage, it is necessary to effectively utilize the multilayer contextual information acquired by the network in the encoding stage as well as the semantic information extracted from the deep network. Therefore, the network adds an GFRF between the encoder and decoder to combine shallow spatial information and deep semantic information. The structure of the GFRF is shown in Fig. 4. The input image retains a lot of background information after the encoder, especially the high-bright artificial lights, which brings interference, and it loses part of the feature information after the decoder owing to the use of the dilated convolution. Through the design of GFRF to

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-1-2024-115-2024 | © Author(s) 2024. CC BY 4.0 License.

integrate the detail information in the encoding stage with the multi-scale features after separating the lights through the processing of the MFEF module, we can achieve a better cloud-boundary segmentation effect while filtering out the lights.

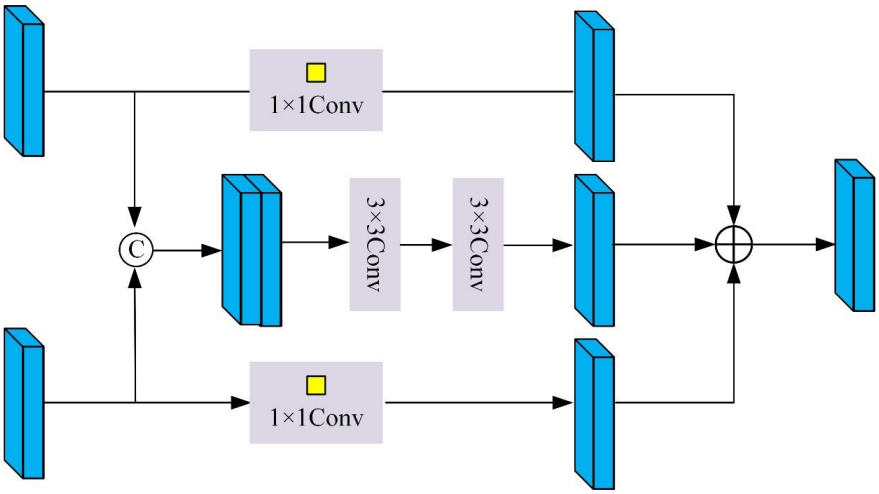

Fig. 4. Structure of AFF

2.5 Datasets

Owing to the lack of publicly available cloud mask datasets of nighttime remote sensing images, we produced a cloud mask dataset of day-night remote sensing images by manual labeling. The remote sensing data captured by the DNB of the VIIRS on board the Soumi NPP satellite were mainly used, covering bare ground and ocean, for example. The data were labeled to obtain 18 daytime raw images, four nighttime raw images, and four daytime raw images. A total of 18 views of daytime raw images and four views of nighttime raw images were labeled, of which four views of daytime images and one view of nighttime images were used to test the model accuracy, and the rest were used to train the model. The training set was chosen to segment the whole scene 4,064 pixels × 3,072 pixels night remote sensing images into 256 pixels × 256 pixels blocks, with 50% overlap in every two blocks, and the set was discarded when the whole scene image had been insufficiently segmented, and a total of 14,894 images were obtained for training. Of these, there were a total of 2,708 nighttime images. Some of the raw images with truth-labeled data are shown in Fig. 5, where the cloud mask’s black pixels indicate non-cloudy regions, and white pixels indicate cloudy regions.

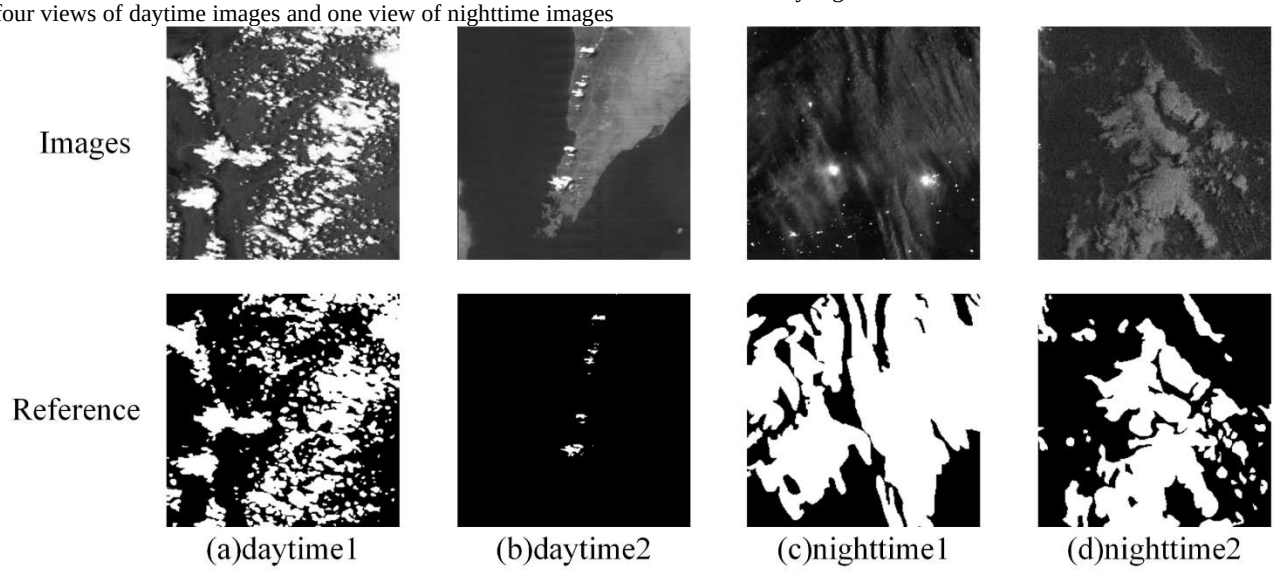

Fig. 5. Schematic of cloud-mask dataset for diurnal remote sensing imagery

2.6 Experimental setup

The experiments in this paper were conducted on a manually labeled cloud mask dataset of day and night remote sensing imagery, and the equipment used for the experiments were as follows: CPU, Intel Core i5-12400F; and GPU, NVIDIA GeForce RTX 3060Ti. The experiments and network development were conducted using Python 3.7 and Pytorch 1.7.1. In this paper, we used the Adam optimizer to train the network. The maximum learning rate was set to 0.0001, training to convergence to reduce the learning rate for training; the minimum learning rate was 0.000001, the batch_size was set to 2, and the epoch was set to 100.

2.7 Evaluation methodology

To evaluate the performance of the method proposed in this paper, we selected five models in the field of semantic segmentation for comparison with the method proposed in this paper, including U-Net, Deeplab-v3+, CDnetv2 (Guo et al., 2021), and CloudU-Net (Shi et al., 2021). CDnetV2 combines multiple attention mechanisms to adaptively fuse multi-scale features to achieve better cloud detection results; moreover, it provides cloud location information for abstract features

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-1-2024-115-2024 | © Author(s) 2024. CC BY 4.0 License.

through an advanced semantic information guided flow. Shi et al. (2021) utilized dilated convolution to improve the network-sensing field and optimized the network output by combining fully connected CRF to construct a CloudU-Net network for circadian cloud detection task. They compared it with the current state-of-the-art semantic segmentation network, and the improved model showed better cloud detection results. Experiments were conducted on a test set including 720 daytime cloud-mask data points and 180 nighttime cloud-mask data points, and the results were compared in terms of evaluation metrics and visualization comparisons.

In this paper, overall accuracy (OA), precision, recall, and F1-score were selected to evaluate the network model prediction results. They were calculated as follows:

$$\mathrm{OA} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{TN}}, \quad (1)$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \quad (2)$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \quad (3)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall}, \quad (4)$$

where TP is true positive, indicating the number of originally cloudy predictions that are also cloudy; TN is true negative, indicating the number of original background predictions that are also background; FP is false negative, indicating the number of original background predictions that are incorrectly predicted to be cloudy, and FN is false negative, indicating the number of original clouds incorrectly predicted to be background. The higher of the four evaluations metrics in the testing phase indicates higher accuracy.

2.8 Analsis of cloud detection results

The performance of different models at daytime on the evaluation metrics is shown in Table 1. As can be seen in Table 1, the Overall Accuracy(OA), precision, recall, and F1-score of MFFCD-Net in daytime image cloud detection were 92.1%, 90.8%, 93.9%, and 92.3%, respectively, which are higher than those of the comparison methods. Through quantitative analysis, it was shown that MFFCD-Net could effectively realize cloud detection in daytime remote sensing images and substantially improve the cloud detection accuracy.

Table 1 Comparison of prediction accuracy of different methods during daytime

Method	OA	Precision	Recall	F1-score
U-Net	85.8	85.3	86.2	85.7
Deeplab-v3+	87.4	85.7	93.1	89.3
CDNetv2	84.6	86.2	85.4	85.8
CloudU-Net	90.4	90.3	90.7	90.5
MFFCD-Net	**92.1**	**90.8**	**93.9**	**92.3**

Figure 6 shows the comparison results of cloud detection by MFFCD-Net and other models for daytime remote sensing images in the test set. From the first line in Fig. 6, it can be seen that there are many small broken clouds in the remote sensing image, and the gap between the clouds is small. the U-Net, Deeplab-v3+, and CDnetv2 networks can effectively detect the area where the clouds are located in the image, but the identification of the cloud boundary of the broken clouds is more ambiguous, and the MFFCD-Net model can be used to detect the clouds in the remote sensing image, and the MFFCD-Net model can be used to detect the clouds in the remote sensing image. The CloudU-Net model performs better in recognizing cloud boundaries compared to the above comparison methods, but some broken clouds are not well recognized, while MFFCD-Net can accurately recognize broken clouds in the image. In the second row of remote sensing image there are thick as well as thin clouds with large area, from the results, it can be seen that U-Net, Deeplab-v3+, CDNetv2, and CloudU-Net in the comparison experiments failed to recognize the thin cloud boundary in the image efficiently, whereas the present research method detects the cloud boundary clearly. Overall, U-Net, Deeplab-v3+, and CDnetv2 show a large number of misdetections in their detection results and are not very accurate in recognizing cloud boundaries; CloudU-Net does not show a large number of misdetections but still misses on broken clouds, and MFFCD-Net outperforms the other four models on all of the datasets.

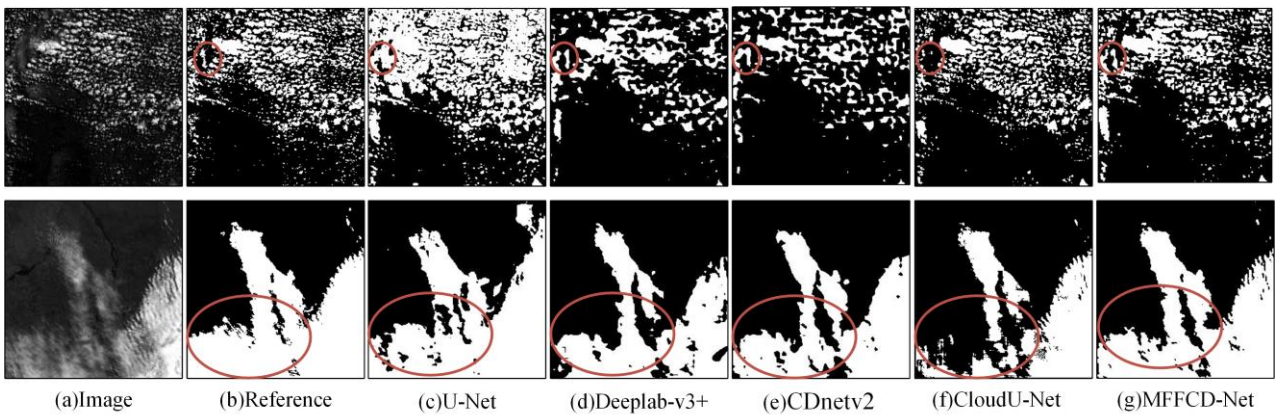

Fig.6. Comparison of daytime remote sensing cloud detection results by different methods

The performance of different models at daytime on the evaluation metrics is shown in Table 2. As can be seen in Table 2, the OA, precision, recall, and F1-score of MFFCD-Net in daytime image cloud detection were 90.4%, 90.2%, 91.5%, and 90.8%, respectively, which are higher than those of the comparison methods. Therefore, the quantitative comparison method results show that MFFCD-Net could effectively improve the cloud detection performance of the network at nighttime and accomplish the day and night cloud detection tasks at the same time.

Table 2 Comparison of prediction accuracy of different methods

Method	OA	Precision	Recall	F1-score

Method	OA	Precision	Recall	F1-score
U-Net	80.8	85.5	79.1	82.2
Deeplab-v3+	86.7	86.1	90.8	88.4
CDNetv2	81.8	85.6	81.2	83.3
CloudU-Net	89.3	89.9	85.1	87.4
MFFCD-Net	**90.4**	**90.2**	**91.5**	**90.8**

Figure 7 shows the comparison results of cloud detection between MFFCD-Net and other models for nighttime remote sensing images in the test set. From the first line of the figure, it can be seen that there are lights in the city area of the nighttime remote sensing image, and these lights are more similar to the tiny broken clouds, which are easy to be misdetected, and the lights are misdetected as clouds in the detection results of Deeplabv3+ as well as CloudU-Net in the comparison methods; at the same time, in the red ellipse area, the cloud boundary of the comparison methods is not clearly recognized, while the cloud boundary segmentation effect of MFFCD-Net is better. In the second row of the figure, there are more broken clouds and thin clouds, U-Net, Deeplab-v3+, CDnetv2 recognize the cloud boundary more ambiguously, CloudU-Net cloud boundary segmentation is clearer than the other comparative methods, but the detection of broken clouds within the red ellipse area is less effective. MFFCD-Net, on the other hand, can accurately recognize broken clouds, thin clouds, and cloud boundaries in the image. There are more broken clouds and thin clouds in the second row of the figure, U-Net, Deeplab-v3+, CDnetv2 recognize the cloud boundary more ambiguously, CloudU-Net cloud boundary segmentation is clearer compared to other comparative methods, but the detection of broken clouds within the red ellipse area is less effective. MFFCD-Net, on the other hand, can accurately recognize broken clouds, thin clouds, and cloud boundaries in the image.

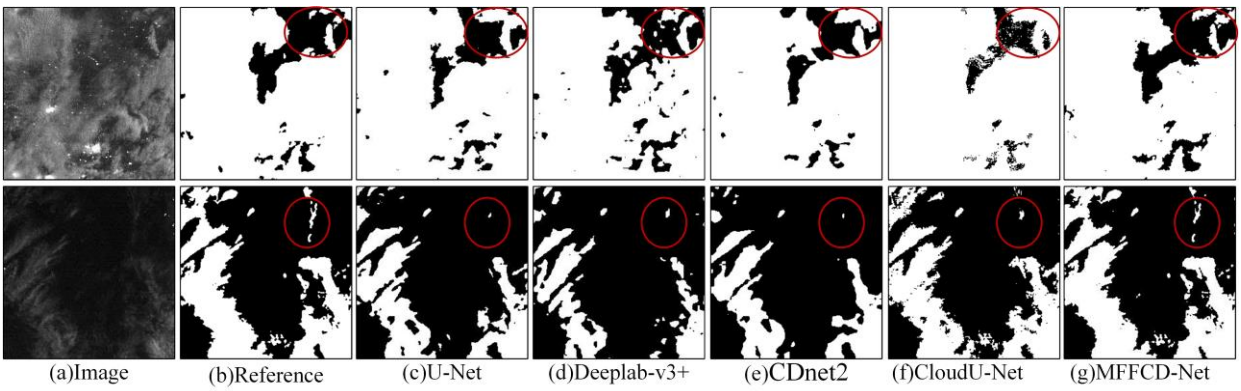

Fig.7. Comparison of nighttime remote sensing cloud detection results by different methods

3. Conclusion

In this paper, MFFCD-Net was proposed to realize cloud detection in diurnal remote sensing images. Compared with the existing methods, MFFCD-Net adds the designed MFEF module, GFRF module, and expanded residual up-sampling module, which improve the model's ability to detect clouds at different scales and achieve better cloud boundary segmentation. These enhancements also improve the ability of MFFCD-Net to distinguish between the regular texture of artificial light and the random texture of a cloud, which significantly improves the cloud detection effect at night. Moreover, owing to the lack of publicly available cloud mask datasets of night remote sensing images, we produced a cloud-mask dataset of day and night remote sensing images by manual labeling, obtained a total of 14894 256×256 cloud mask images, and conducted a precision test on the labeled dataset. Its precision was 92.1%, 90.8%, 93.9%, and 92.3%, respectively, and the precision, OA at night, recall, and F1 score were 90.4%, 90.2%, 91.5% and 90.8%, respectively. The experimental results on VIIRS DNB images indicate that MFFCD-Net can realize the cloud detection task for both daytime and nighttime remote sensing images and can effectively improve the network's ability to capture clouds at different scales, the segmentation effect of cloud boundaries, and the discrimination of light. This is the first time that a deep learning method is utilized to realize the day and night cloud detection task simultaneously. In the future, we can explore the cloud detection method that combines the reflectivity information of the ground surface with the characteristics of SAR images penetrating clouds, and we can reduce the influence of high-brightness ground surface on cloud detection by combining the reflectivity information of the ground surface.

References

Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.

C. Shi, Y. Zhou, B. Qiu, D. Guo and M. Li, "CloudU-Net: A Deep Convolutional Neural Network Architecture for Daytime and Nighttime Cloud Images’ Segmentation, " in IEEE Geoscience and Remote Sensing Letters PP.99(2020)

C. Shi.; Zhou, Y.; Qiu, B.; He, J.; Ding, M.; Wei, S. Diurnal and nocturnal cloud segmentation of all-sky imager (ASI) images using enhancement fully convolutional networks. Atmos. Meas. Tech. 2019,12, 4713–4724

Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.

Hurley J, Dudhia A, Grainger R G. Cloud detection for MIPAS using singular vector decomposition[J]. Atmospheric Measurement Techniques, 2009, 2(2): 533-547.

Irish R R, Barker J L, Goward S N, et al. Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm[J]. Photogrammetric engineering & remote sensing, 2006, 72(10): 1179-1188.

J. Guo, J. Yang, H. Yue, H. Tan, C. Hou and K. Li, "CDnetV2: CNN-Based Cloud Detection for Remote Sensing Imagery With Cloud-Snow Coexistence," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 1, pp. 700-713, Jan. 2021

Li X, Yang X, Li X, et al. GCDB-UNet: A novel robust cloud detection approach for remote sensing images[J]. Knowledge-Based Systems, 2022, 238: 107890.

Li, Z., Shen, H., Li, H., Xia, G., Gamba, P., Zhang, L., 2017. Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2017.01.026

Li, Z., Shen, H., Weng, Q., Zhang, Y., Dou, P., Zhang, L., 2022. Cloud and cloud shadow detection for optical satellite imagery: Features, algorithms, validation, and prospects. ISPRS J. Photogramm. Remote Sens. 188, 89–108. https://doi.org/10.1016/j.isprsjprs.2022.03.020

Li, J., Wu, Z., Hu, Z., Jian, C., Luo, S., Mou, L., Zhu, X.X., Molinier, M., 2022. A Lightweight Deep Learning-Based Cloud Detection Method for Sentinel-2A Imagery Fusing Multiscale Spectral and Spatial Features. IEEE Trans. Geosci. Remote Sens. 60, 1–19. https://doi.org/10.1109/TGRS.2021.3069641

Li, J., Wu, Z., Sheng, Q., Wang, B., Hu, Z., Zheng, S., Camps-Vall, G., Molinier, M., 2022. A hybrid generative adversarial network for weakly-supervised cloud detection in multispectral images. Remote Sens. Environ. 280, 113197. https://doi.org/10.1016/j.rse.2022.113197

Ma L, Li M, Ma X, et al. A review of supervised object-based land-cover image classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 130: 277- 293.

Mackie S, Embury O, Old C, et al. Generalized Bayesian cloud detection for satellite imagery. Part 1: Technique and validation for

Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.

Schiffer R A, Rossow W B. The International Satellite Cloud Climatology Project (ISCCP): The first project of the world climate research programme[J]. Bulletin of the American Meteorological Society, 1983, 64(7): 779-784.

Wei J, Huang W, Li Z, et al. Cloud detection for Landsat imagery by combining the random forest and superpixels extracted via energy-driven sampling segmentation approaches[J]. Remote Sensing of Environment, 2020, 248: 112005.

Xu L, Wong A, Clausi D A. A novel Bayesian spatial–temporal random field model applied to cloud detection from remotely sensed imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(9): 4913-4924.

Zhang Y, Rossow W B, Lacis A A, et al. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data[J]. Journal of Geophysical Research: Atmospheres, 2004, 109(D19).

Zhu Z, Woodcock C E. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change[J]. Remote Sensing of Environment, 2014, 152: 217-234.

Zhu Z, Woodcock C E. Object-based cloud and cloud shadow detection in Landsat imagery[J]. Remote sensing of environment, 2012, 118: 83-94.