A Unified Cloud Detection Method for Suomi-NPP VIIRS Day and Night PAN Imagery

Jun Li, Chengjie Hu, Qinghong Sheng, Bo Wang, Xiao Ling, Fan Gao, Yunfei Xu, Zhiwei Li, and Matthieu Molinier, *Member, IEEE*

***Abstract*— Cloud detection is a necessary step before the application of remote sensing images. However, the radiation intensity similarity between artificial lights and clouds is higher in nighttime remote sensing images than in daytime remote sensing images, making it difficult to distinguish artificial lights from clouds. This article proposes a deep learning method called multifeature fusion for cloud detection network (MFFCD-Net) to detect clouds in daytime and nighttime remote sensing images. A dilated residual upsampling module was designed for upsampling feature maps while enlarging the receptive field. A multiscale feature-extraction fusion module (MFEF) was designed to enhance the ability to distinguish regular textures of artificial lights from random textures of clouds. Moreover, an adaptive feature-fusion module (AFF) was designed to select and fuse the feature in the encoding stage and decoding stage, thus improving the cloud detection accuracy. To the best of our knowledge, this is the first time that a method is designed for cloud detection in both daytime and nighttime remote sensing images. The experimental results on Suomi-NPP Visible Infrared Imaging Radiometer Suite (VIIRS) of the panchromatic (PAN) day/night band (DNB) images show that MFFCD-Net could obtain a better balance in commission and omission rates than baseline methods (92.3% versus 90.5% on the F1-score) in daytime remote sensing images. Although artificial lights introduce strong interference in nighttime remote sensing images, MFFCD-Net can better distinguish artificial lights from clouds than baseline methods (90.8% versus 88.4% on the F1-score). The results indicate that MFFCD-Net is promising for cloud detection both in daytime and nighttime remote sensing images. The source code and dataset are available at https://github.com/Neooolee/MFFCD-Net.**

***Index Terms*— Adaptive feature fusion, cloud detection, daytime and nighttime remote sensing, deep learning, multiscale feature-extraction fusion (MFEF).**

Manuscript received 16 February 2024; revised 6 April 2024 and 22 May 2024; accepted 1 July 2024. Date of publication 11 July 2024; date of current version 24 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42301384, Grant 42271448, and Grant 42101357; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20220888 and Grant BK20231030; in part by the National Key Laboratory of Science and Technology on Space Microwave under Grant HTKJ2022KL504018; and in part by the Space Optoelectronic Measurement and Perception Laboratory of BICE under Grant LabSOMP-2021-06. The work of Matthieu Molinier was part of the Finnish Flagship Programme FCAI: Finnish Center for Artificial Intelligence by the Academy of Finland. *(Corresponding author: Qinghong Sheng.)*

Jun Li, Chengjie Hu, Qinghong Sheng, Bo Wang, and Xiao Ling are with the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: jun.li@nuaa.edu.cn; hucj_1206@nuaa.edu.cn; qhsheng@nuaa.edu.cn; wangbo_nuaa@nuaa.edu.cn; xlingsky@nuaa.edu.cn).

Fan Gao is with the Institute of Communication Engineering, Army Engineering University of PLA, Nanjing 210001, China (e-mail: fgao_lgd@163.com).

Yunfei Xu is with Beijing Institute of Control Engineering, Beijing 100190, China (e-mail: 1594603058@qq.com).

Zhiwei Li is with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong (e-mail: zhiwei.li@polyu.edu.hk).

Matthieu Molinier is with VTT Technical Research Centre of Finland Ltd., 02044 Espoo, Finland (e-mail: matthieu.molinier@vtt.fi).

Digital Object Identifier 10.1109/TGRS.2024.3426649

I. Introduction

SATELLITE sensors collect data related to electromagnetic waves in different bands to form satellite remote sensing images, and they use the collected information to accomplish tasks, such as feature identification, target detection, target tracking, and disaster warning. Satellite remote sensing images are inevitably affected by cloud cover, and a data analysis by the International Satellite Cloud Climatology Project (ISCCP) indicated that 67% of the world is covered by clouds [1]. Most remote sensing applications rely on ground surface information. However, because the image features are obscured by clouds, the radiation information reflected from the surface is not accurately received by the satellite sensors, and the spectral information of the features is altered, which affects the usable information of the optical remote sensing images and poses a great obstacle to subsequent applications [2]. In addition, remote sensing images covered by a large number of clouds could be used for the analysis of cloud statistics and properties, which are very useful for monitoring weather and climate. Therefore, cloud detection of satellite remote sensing images is an indispensable preprocessing step in remote sensing image processing and plays an important role in the subsequent application of remote sensing images.

In previous decades, scholars have carried out extensive research on daytime remote sensing image cloud detection [3], [4], [5], [6], [7]. The cloud detection methods in these works are mainly categorized into four types: physical rule-based methods, multitemporal-based methods, machine learning-based methods, and deep learning-based methods [8].

Physical rule-based cloud detection methods use the characteristics of the high reflectivity and low temperature of a cloud layer to set different rules for different spectral channels to realize cloud detection [9], [10], [11]. Fmask has been widely used for cloud detection in Landsat 7–9 [12]. It set several rules combining different bands to distinguish cloud from different backgrounds. Parmes et al. [13] developed a rule-based method for Suomi-NPP Visible Infrared Imaging Radiometer Suite (VIIRS) images based on spectral analysis of

clouds, shadows, snow, and water pixels. This method reached 94.2% correct detection rates and 11.1% false alarms for cloud detection. 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 relies on empirical judgment and parameter sensitivity analysis.

In image time series of the same region, cloud coverage can lead to sudden changes in reflectivity, so multitemporal-based methods identify clouds by comparing the reflectivity differences between cloud pixels and clear-sky pixels [14]. Hagolle et al. [15] identified clouds through the detection of the reflectivity surge in the blue band pixel by pixel, and they combined this identification with the correlation of neighboring pixels in multitemporal images. The Tmask method [16] utilizes Landsat time-series information to construct an apparent reflectance prediction model, and it utilizes the difference between the predicted and the imaged reflectance of the pixel to perform cloud detection. Multitemporal-based methods combine spectral information and temporal information, which can improve the accuracy of cloud detection. However, the performance multitemporal-based method is highly dependent on the time series, while misidentification of clouds and surfaces occurs for regions with significant changes in surface types.

Machine learning-based cloud detection methods consider cloud identification as a classification problem (often binary). Supervised learning mainly includes support vector machine (SVM) [17], decision tree [18], Bayesian [19], and random forest (RF) [20]. Unsupervised learning includes principal component analysis (PCA) [21] and singular value decomposition (SVD) [22], which have been widely used in remote sensing image cloud detection research. Machine learning methods usually need to design features manually. The model extracts limited information and cannot access advanced semantic information. The machine learning-based cloud detection method synthesizes the spectral and spatial features of a cloud and surface to achieve automatic cloud detection.

Deep learning networks have been widely used in cloud detection tasks recently. Compared to the traditional cloud detection methods, U-Net [23], SegNet [24], DeepLabV3+ [25], and other deep learning cloud detection methods have a better cloud recognition ability and, at the same time, can improve the efficiency of cloud detection [26], [27], [28], [29], [30], [31]. Li et al. [27] designed a novel multiscale feature-fusion module that fuses features of different scales into output for cloud detection in remote sensing images from different sensors. Self-attention mechanism has been proven effective in computer vision tasks and applied on cloud detection. Zhang et al. [32] proposed a cloud vision transformer (CloudViT) that utilizes dark channel priors in multispectral imagery to guide the network to learn features. Ma et al. [33] combined the strengths of both transformer and convolutional neural networks (CNNs) to extract local and global features with dual branches. One branch processes cloud images, and the other processes the differential image between cloud and corresponding cloud-free images.

In recent years, nighttime remote sensing images have been widely used and are important data for monitoring human activities, such as urbanization, disaster evaluation, and wars. Clouds and artificial lights have high brightness, which makes it difficult to distinguish them in nighttime remote sensing images. This will influence the application of nighttime remote sensing images. Thus, accurately detecting clouds in nighttime remote sensing images is very necessary and important.

Fewer works studied cloud detection in nighttime remote sensing images. Obregon et al. [34] improved the existing MODIS cloud-masking algorithm by using the brightness temperature differences between thermal and mid-infrared (IR) bands to achieve highly accurate identification of nighttime underlying clouds while reducing the false detection rate from 44.2% to 16.3%. In order to improve the cloud detection accuracy of MODIS cloud mask algorithm in polar region at nighttime, a 7.2-μm water vapor band and a 14.2-μm carbon dioxide band were involved in a few new physical rules [35]. Taking cloud product from Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) as a reference, Yang et al. [36] constructed an XGBoost machine learning model for cloud retrieval of day and night remote sensing images by using the brightness temperature of ten thermal IR channels with central wavelengths ranging from 3.9 to 13.3 μm in advanced Himawari imager (AHI) onboard the Himawari-8 satellite. Liu et al. [37] also took CALIOP cloud product as the reference of AHI images. The brightness temperatures of eight thermal bands are put into the RF algorithm for cloud detection in nighttime images. The brightness temperature differences between 3.7-, 11-, and 12-μm bands were used to obtain the cloud confidence of a pixel in Visible and Infrared Radiometer (VIRR) onboard the Chinese Feng Yun 3 (FY-3) satellite [38]. Three threshold test groups were designed and combined for cloud detection in different land cover types. Merchant et al. [39] designed a Bayesian approach using thermal IR spectral information of 3.7, 11, and 12 μm and texture features for cloud detection in nighttime image acquired by Along Track Scanning Radiometers 2 (ATSR-2). Visible and near-IR band and thermal IR band of Defense Meteorological Satellite Program (DMPS) are put into integrated active contour (IAC) model for cloud detection in nighttime images [40]. The results on two channels are not uniform, because they only used the gray-level information, ignoring the texture and shape information. Wang et al. [41] used CALIOP cloud product and three VIIRS IR bands (8.6, 11, and 12 μm) to train an all-day RF model for cloud detection in both daytime and nighttime remote sensing images. It only used the spectral information of these bands. Focusing on urban areas, Joachim and Storch [42] used the enterprise cloud mask (ECM) [43] for DNB RF cloud detection algorithm for nighttime panchromatic (PAN) day/night band (DNB) images, which achieved better detection results in urban areas.

Although the current research on cloud detection in daytime and nighttime remote sensing images has made great progress. Most methods take multispectral bands as input. Methods for nighttime cloud detection only use spectral information while ignoring the spatial and texture information. There are far few works on daytime and nighttime cloud detection for PAN

DNB images. The main difficulties for cloud detection in DNB images are given as follows.

1) Since the VIIRS DNB image is PAN image, there is little spectral feature in the DNB image. The complexity of the cloud structure and different scales of clouds have different boundary scales leading to the omission and misdetection of fragmented clouds as well as clouds with smaller scales.
2) The brightness of clouds and artificial lights is very similar in nighttime remote sensing images, which increases the difficulty of distinguishing lights from clouds in nighttime images with only PAN band. Artificial lights may be recognized as point clouds.

Aiming at addressing the abovementioned problems, we propose multifeature fusion for cloud detection network (MFFCD-Net) to realize cloud detection in daytime and nighttime DNB images in the same architecture for the first time. By analyzing the difference between daytime and nighttime images, three modules were designed in MFFCD-Net to improve the cloud detection accuracy in nighttime images. MFFCD-Net is based on an encoder–decoder architecture that acquires the multiscale information of the image through the multiscale feature-extraction module (MFEF Module) and associates the shallow detail features with the distinguishing features of lights and clouds through the adaptive feature-fusion module (AFF module) to achieve a better cloud detection effect. The main contributions of this article are given as follows.

1) An MFEF module is designed to extract multiscale information and select noteworthy cloud information from different scale feature maps, thus improving the network's ability to focus on the spatial background region that is most relevant to clouds. In addition, a dilated residual upsampling module (DR-UP) block is designed in the decoding stage to expand the network receptive field so that the network can capture cloud features at a large scale.

2) At night, the MFEF module can capture the large-scale features of urban lights and more dispersed lights in rural areas, filter and optimize the multiscale 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 AFF module to achieve better cloud-boundary segmentation accuracy and lighting discrimination.

3) Addressing the lack of manually labeled nighttime remote sensing cloud mask datasets, an NUAA-DNB-CD dataset, in which 18 daytime and four nighttime VIIRS DNB images were manually labeled, is produced. Images and corresponding cloud masks in NUAA-DNB-CD were clipped into small patches with size of 256 × 256, and a total of 14 894 256 × 256 cloud mask patches were obtained after expansion, including 2708 cloud mask patches for nighttime remote sensing images.

The rest of this article is organized as follows. Section II presents the experiment data. Section III introduces the proposed MFFCD-Net method. The experimental results and discussion are shown in Section IV, and conclusions are presented in Section V.

II. Data

A. VIIRS of the PAN DNB Imagery

The VIIRS is the key multispectral imager onboard the Suomi-NPP satellite. VIIRS collects radiometric images of the atmosphere, land, and oceans in different channels, such as visible and IR bands. The VIIRS sensor has a total of 22 bands: five I-bands at 370-m spatial resolution, one DNB, and 16 M-bands at 750-m spatial resolution. Two observations can be obtained every day globally. In addition to inheriting the band characteristics of sensors, such as MODIS and AVHRR, and improving the signal-to-noise ratio and spatial resolution of the data, VIIRS has effective control of the spatial resolution, which increases with the increase of the scanning angle. This represents the biggest improvement of VIIRS on the data quality of polar-orbiting environmental satellites.

The DNB is PAN image and a revolutionary remote sensing technology that is capable of capturing details of the Earth's surface under day and night conditions. The main differences between day and night imaging modes are the lighting conditions and the adaptability of the imaging technology. During daytime, the DNB relies on the Sun as the primary light source to capture reflected radiation from the surface and atmosphere. Daytime imaging mainly utilizes radiation in the visible-to-near-IR wavelength range in the DNB. At night, the DNB demonstrates its unique low-light imaging capability. In the night mode, it relies on moonlight, starlight, and artificial light sources on the Earth's surface for imaging. Nighttime imagery is characterized by the ability to capture surface features, such as city lights, fires, ships, and other signs of human activity, despite extremely low-light conditions. Compared with daytime, nighttime images are dimmer and have less contrast, but they still reveal the patterns and geographic distribution of human activity. The DNB has a spectral range of 500–900 nm, covering a portion of the spectrum from the visible to the near-IR. Unlike conventional multiband sensors, DNB provides highly sensitive optical imaging through a single band, both during the day and at night. This wavelength setup allows the DNB to operate continuously under varying light conditions, providing all-weather Earth observation data.

In the DNB-image analysis of the VIIRS sensor, artificial lights at night are usually characterized by bright spots with clear boundaries and regular shapes, which significantly reflect the distribution of human activities, such as the layout of road networks and urban center areas. In contrast, clouds in remote sensing images show obvious irregularities and a diffuse nature, usually covering large areas with blurred edges. In terms of spectral characteristics, even in grayscale images, artificial lights appear as relatively uniform and bright areas, while light reflected from clouds is relatively darker, with lower brightness and more uniform tonal distribution. By utilizing these features, we can design algorithms to make the network more effective in focusing on and distinguishing between these two different elements, thereby improving the interpretation accuracy of nighttime remote sensing images.

TABLE I
DETAILS OF TRAINING AND TESTING DATASETS (NUMBER OF PATCHES)

Dataset	Training	Testing	Patch pixels
Daytime	11509	720	256×256
Nighttime	2031	180	256×256

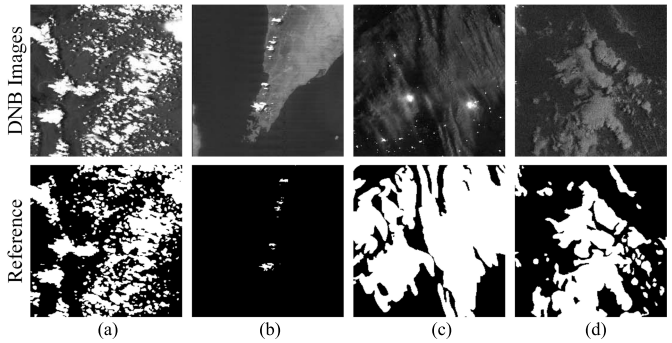

Fig. 1. Examples of manually labeled cloud masks for daytime and nighttime remote sensing images. (a) Daytime 1. (b) Daytime 2. (c) Nighttime 1. (d) Nighttime 2.

B. NUAA-DNB-CD Dataset

Owing to the lack of manually labeled 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 onboard the Suomi NPP satellite were mainly used, covering bare ground and ocean, for example. A total of 18 daytime images and four nighttime images were manually labeled for training and testing. Each DNB image was processed by three experts in this field to produce cloud mask. The DNB image was labeled by one expert, and then, the manually labeled cloud mask will be checked and corrected by the other two experts. For controversial region, if both 2 or more experts classify it as cloud, this region will be labeled as cloud, and otherwise, it will be labeled as background. The size of each image is about 4064 × 3072 pixels. Among these images, four daytime images and one nighttime image were used to test the model's accuracy, and the rest were used to train the model. The images were clipped into small patches with 256 × 256 pixels, and 50% overlap was set between every two patches. The details of training and testing patches are presented in Table I. Four manually labeled examples are shown in Fig. 1, where black pixels indicate noncloudy regions and white pixels indicate cloudy regions.

III. METHODOLOGY

A. Overview of MFFCD-Net

The overall architecture of MFFCD-Net designed in this article is shown in Fig. 2. MFFCD-Net is designed 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 multiscale feature maps by means of the dilated residual module upsampling (DR-UP). Each DR-UP Block is composed of three convolutional layers with different dilated rates and one upsampling layer, and the DR-UP Block expands the feature-sensing field of the network while recovering the image size. The multiscale feature-extraction fusion module (MFEF) module is added to the bottom layer of the encoder to mine the multiscale 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 skip-connection structure with an adaptive feature-fusion module (AFF), 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.

B. DR-UP

Traditional convolution aims to use a convolution kernel to perform a convolution operation on the neighboring elements in the tensor, while the dilated convolution can be convolved on two nonneighboring elements, which can increase the convolution kernel's receptive field and reduce the amount of computation. Different expansion rates have different receptive fields, and multiscale features can be extracted so that the network can capture cloud features at different scales. The receptive fields of different convolutional expansion rates are shown in Fig. 3. The DR-UP Block was designed in the decoder by combining the dilated convolution and upsampling layers, and the DR-UP Block consists of three dilated convolution layers and one upsampling layer. The kernel size of all layers is set to 3 × 3 in the DR-UP Block. The same dilated rate of adjacent dilated 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. 3 shows the structure of the DR-UP Block designed in this article.

C. MFEF

In nighttime remote sensing image cloud detection, the high radiation intensity of artificial lights and their similarity in color to clouds often leads 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, we designed an MFEF module, which consists of two parts: a multiscale spatial pyramid and a multiscale feature selection module. In the multiscale 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 multiscale feature selection module is used to filter the multiscale 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 similar remote sensing image processing problems. In addition, the MFEF module designed

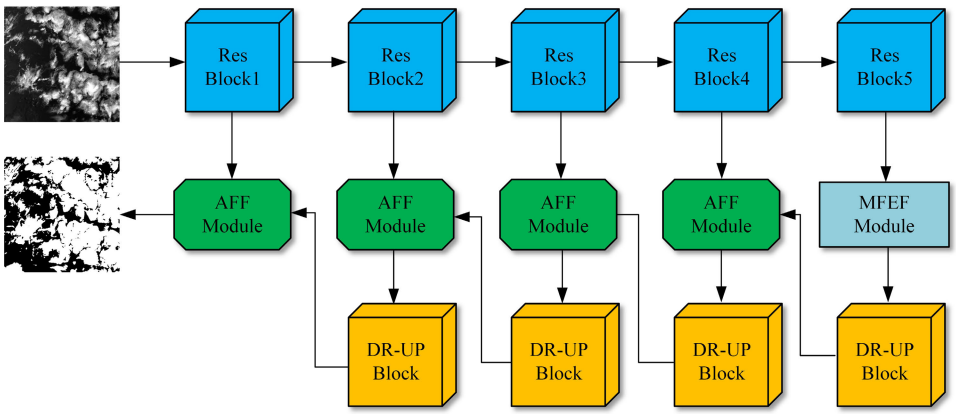

Fig. 2. Overall architecture of multiscale feature-fusion cloud detection network (MFFCD-Net).

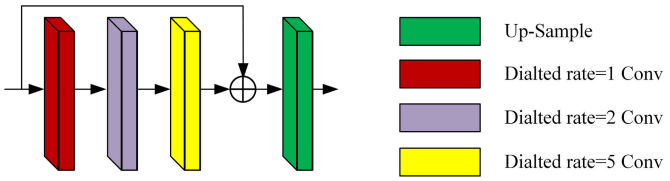

Fig. 3. Structure of DR-UP block.

in this article aims to improve performance in nighttime cloud detection and simultaneously optimize the ability to capture clouds at different scales during daytime. This is attributed to its multiscale 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.

D. AFF

Global features and interpixel 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 AFF between the encoder and the decoder to combine shallow spatial information and deep semantic information.

The structure of the AFF is shown in Fig. 5. 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. There are two input branches in the AFF module. One is for feature from encoder and the other is for feature from decoder. Features from encoder have a limited receptive field, but features from decoder have multiscale receptive field. Thus, two 1×1 Conv are used to select useful features in different branches. The 3×3 Conv is used to fuse the local and multiscale features from the encoder and decoder. The selected features and fused features are summed to produce the final result. In this way, AFF can fuse features from the encoder and decoder adaptively. Through the design of AFF to integrate the detail information in the encoding stage with the multiscale features after separating the lights through the processing of the MFEF module, we aim to achieve a better cloud-boundary segmentation effect while filtering out the lights.

The cross-entropy loss function [44] has proved very effective in image segmentation task [31], [45], [46], [47] and thus was used as the loss function of MFFCD-Net

$$L = -\frac{\mathbf{1}}{w \times h} \sum_{i=1}^{w} \sum_{j=1}^{h} \left(y_{i,j} \ln \bar{y}_{i,j} + \left(\mathbf{1} - y_{i,j}\right) \ln\left(\mathbf{1} - \bar{y}_{i,j}\right) \right) \tag{1}$$

where y is the reference cloud mask; $\bar{y}$ is the predicted mask; and w and h are the width and height of the reference mask, respectively.

E. Experimental Setting

1) Hyperparameters Setting: The experiments in this article were conducted on a manually labeled cloud mask dataset of day and night remote sensing imagery, and the equipment used for the experiments was given 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 article, 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) Accuracy Metrics: In order to evaluate the performance of cloud detection methods quantitatively, four accuracy metrics were selected: overall accuracy (OA), precision or

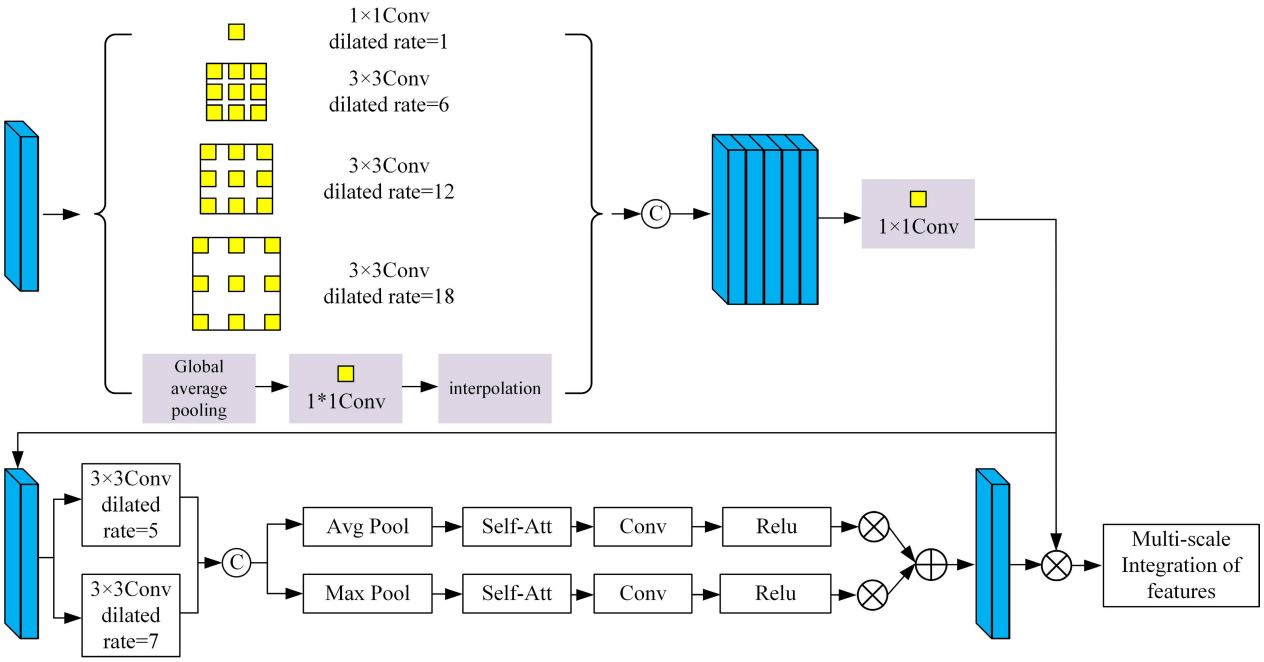

Fig. 4. Structure of MFEF module.

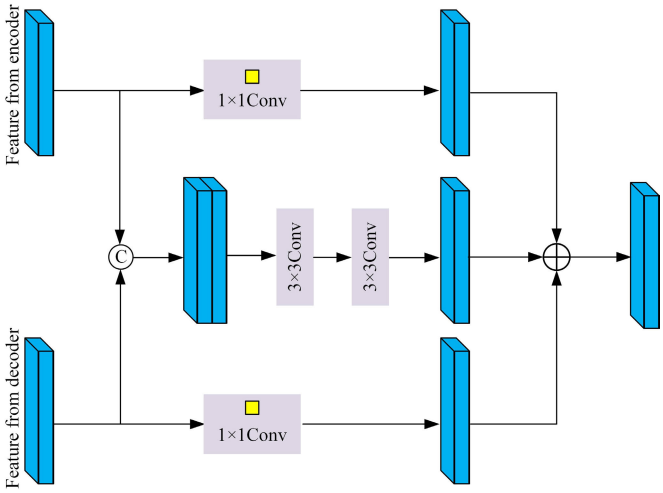

Fig. 5. Structure of AFF module.

producer's accuracy (PA), recall or user's accuracy (UA), and F1-score. The formulas of accuracy metrics are given as follows:

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

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

F. Baseline Methods

Cloud detection in VIIRS DNB images is a domain of research that has not yet been explored, which restricts the applicability of state-of-the-art cloud detection methods. For example, Zhang et al. [32] require a dark channel image and at least two bands. Ma et al. [33] require paired cloudy/cloud-free images, but there is no such dataset for DNB images yet, and they are challenging to annotate especially at night-time. Even rule-based methods are usually constructed upon thresholds using several spectral bands to discriminate clouds from different land cover classes. Thus, many state-of-the-art cloud detection methods cannot yet be applied to this research domain as such.

To evaluate the performance of the method proposed in this article, we selected seven methods for comparison with the method proposed in this article, including U-Net [23], Deeplab-v3+ [31] CDnetv2 [48], CloudU-Net [49] Improved deeplabv3+ [50], and RD-Unet [51]. CDNetV2 combines multiple attention mechanisms to adaptively fuse multiscale features to achieve better cloud detection results. Moreover, it provides cloud location information for abstract features through an advanced semantic information-guided flow. Shi et al. [49] 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 daytime and nighttime cloud detection task. However, CloudU-Net is designed to be trained on ground-based images. They compared it with the current state-of-the-art semantic segmentation network, and the improved model

TABLE II
COMPARISON OF PREDICTION ACCURACY OF DIFFERENT METHODS AT DAYTIME. THE BEST VALUES ARE MARKED IN BOLD

Methods	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
IDeeplab-v3+	87.6	83.3	**96.5**	89.4
RD-UNet	90.4	**97.2**	83.4	89.8
MFFCD-Net	**92.1**	90.8	93.9	**92.3**

showed better cloud detection results. Improved deeplabv3+ (called IDeeplabv3+) introduced an improved Xception backbone network and CBAM (convolutional block attention module) into Deeplabv3+ to enhance the model capability. RD-UNet designed a residual path to cascade two U-shaped networks for information flow from the first network to the second network. All methods were trained on daytime/nighttime images. In our experiments, we also merged daytime and nighttime images for training and then tested on daytime or nighttime images separately.

IV. EXPERIMENTAL RESULTS

A. Results at Daytime

The performance of different methods at daytime on the evaluation metrics is shown in Table II. As can be seen in Table II, 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. MFFCD-Net got higher OA and F1-score than those of the comparison methods, but ranked the second in precision and recall. Although IDeeplabv3+ obtained the highest recall, it got the lowest precision. This means that IDeeplabv3+ will classify more background as cloud. RD-UNet has the highest precision but lowest recall among all methods. This means that RD-UNet will classify more cloud as background. Through the 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.

Fig. 6 shows the comparison results of cloud detection by MFFCD-Net and other methods for daytime remote sensing images in the test set. Fig. 6(a) shows both large-scale thick clouds and small-scale broken clouds, and MFFCD-Net achieved better boundary segmentation of large-scale clouds and higher detection accuracy of small-scale broken clouds compared with the comparison methods. In Fig. 6(b), there is a large thick cloud, and at the same time, its cloud boundary is more complicated. MFFCD-Net had the best cloud-boundary segmentation delineation. In Fig. 6(c), there is a large area of tiny broken clouds, and MFFCD-Net could comprehensively capture the broken clouds in the figure with a lower false detection rate. In Fig. 6(d), there are long thick clouds, broken clouds, and thin clouds at the same time, and there is a cavity in the thin cloud underneath. MFFCD-Net had the best detection results in all three types of clouds compared with the comparison methods. Overall, U-Net, Deeplab-v3+, CDnetv2, and IDeeplabv3+ had a large number of misdetections in their detection results and were not accurate in recognizing cloud boundaries. CloudU-Net and RD-UNet did not have a large number of misdetections, but they still missed the detection of thin clouds and had poor segmentation of broken clouds. MFFCD-Net obtains the best detection results among all methods.

B. Results at Nighttime

The performance of different methods at daytime on the evaluation metrics is shown in Table III. As can be seen in Table III, the OA, precision, recall, and F1-score of MFFCD-Net in daytime image cloud detection are 90.4%, 90.2%, 91.5%, and 90.8%, respectively, which are higher than those of all baseline methods except lower than IDeeplabv3+ in recall. However, IDeeplabv3+ got the second lowest precision. RD-UNet obtained the worst performance on all accuracy metrics. Although CloudU-Net obtained close OA and precision to MFFCD-Net, F1-score and recall of it are much lower than those of MFFCD-Net. This means that MFFCD-Net achieved a better balance between commission and omission errors. Therefore, the quantitative comparison 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.

Fig. 7 shows a comparison of the results of different methods for nighttime cloud detection. It can be seen that there are lights in all four sets of test images, which affects the discrimination of nighttime cloud detection. The cloud boundary in Fig. 7(a) is more complicated. MFFCD-Net had the best delineation on the cloud boundary, as well as the least misclassification. RD-UNet missed many clouds in Fig. 7(a). There are more stray point light sources in Fig. 7(b), U-Net, Deeplab-v3, CDnetv2, and RD-UNet misdetected some of the point light sources as clouds, while CloudU-Net and IDeeplabv3+ resulted in leakage of cloud regions. There are some light sources in Fig. 7(c), and since it is a rural area, the lights are sparser, which requires the network to have a higher ability to mine the cloud information. As such, MFFCD-Net, compared with the comparative methods, could realize the detection of all the clouds in the figure with no light misdetection. However, IDeeplabv3+ and RD-UNet classified some lights into cloud. There is a large light source in Fig. 7(d); compared with baseline methods, MFFCD-Net could realize a better cloud-boundary segmentation effect, and at the same time, the false detection rate of the light was low. All baseline methods misclassified some lights into cloud. In Fig. 7, MFFCD-Net had some cloud misdetection owing to the overlap between lights and clouds, smoothing out the texture features of the clouds. Moreover, the presence of lights with cloud edges did not appear inside the clouds, and thus, it could not be detected by large-scale discrimination to realize the detection. In summary, MFFCD-Net achieved a better cloud detection effect in night remote sensing images and effectively completed the discrimination and separation of lights from clouds. Furthermore, even in the case of low image contrast, it could still effectively mine cloud information and perform cloud detection.

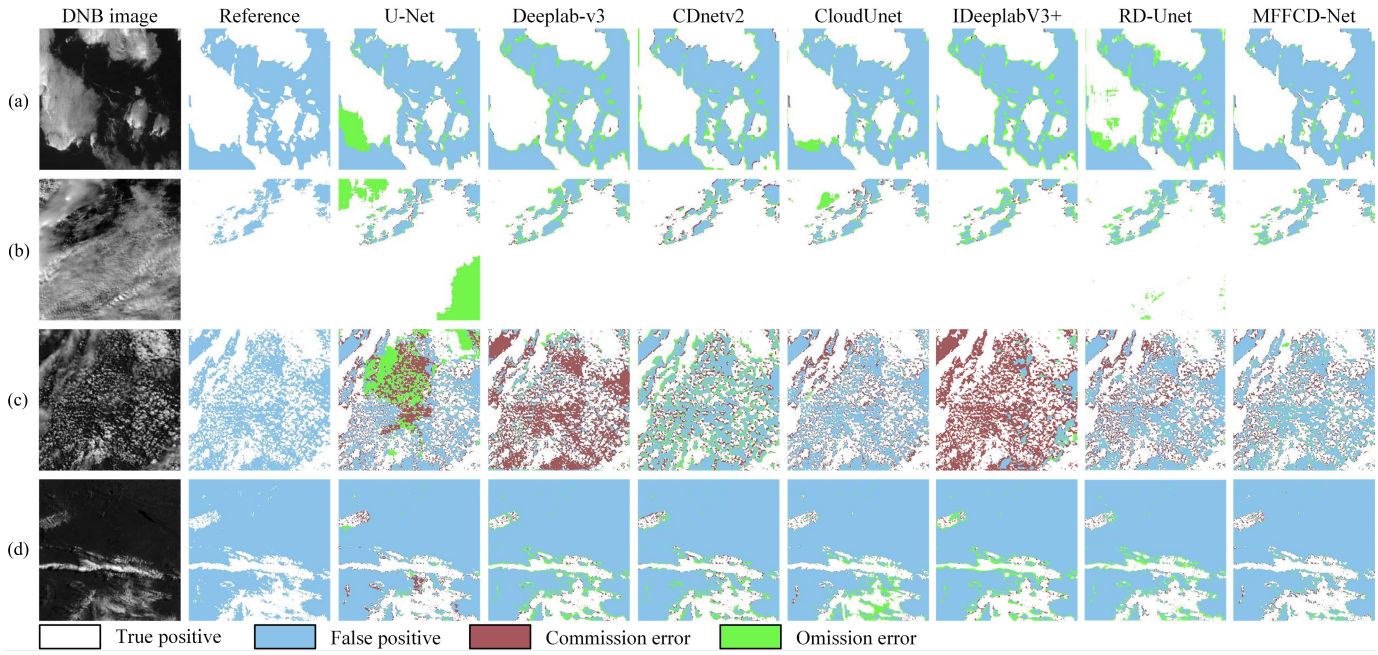

Fig. 6. Comparison of daytime remote sensing cloud detection results by different methods.

TABLE III
COMPARISON OF PREDICTION ACCURACY OF DIFFERENT METHODS AT NIGHTTIME. THE BEST VALUES ARE MARKED IN BOLD

Methods	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
IDeeplab-v3+	82.8	81.1	**97.1**	88.4
RD-UNet	76.9	76.5	62.2	68.6
MFFCD-Net	**90.4**	**90.2**	91.5	**90.8**

C. Model Efficiency Analysis

In this study, a comparison was made for the network efficiency, as shown in Table IV. The computing efficiency test method was the time required to process 900 testing patches of 256 × 256 pixels at inference. Compared with CDnetv2 and CloudU-Net, MFFCD-Net was more efficient while achieving high-precision cloud detection. RD-UNet got the same inference time as that of MFFCD-Net because MFFCD-Net is more complicated than U-Net, Deeplab-v3+, and IDeeplabv3+. MFFCD-Net is less effective than them. Although MFFCD-Net increased the complexity of the network, the detection efficiency did not lag much behind, and it had little impact on the practical cloud detection applications with almost no impact.

D. Ablation Experiment

1) Ablation Experiment for Daytime Imagery: To study the effect of the design modules of this article on MFFCD-Net, we conducted ablation experiments as well as visualization and analysis of the prediction results. Table V shows the prediction accuracies at daytime after adding different modules, from which it can be seen that the design modules in this article could effectively improve the cloud detection OA and F1-score. The addition of DR-Up Block and MFEF module significantly improved the recall, which suggests that the sensing field and multiscale information provided by these two modules can allow the network to capture more cloud information and improve the network's ability to detect clouds. After adding AFF, we found that recall was reduced, while precision was significantly improved, which indicates that AFF can well fuse local spatial details and global semantic information to achieve better cloud discrimination ability.

Fig. 8 shows the comparison of daytime detection results after adding different modules at a time. There are more complex cloud boundaries in Fig. 8(a), and there are cloud shadows due to the varying cloud heights. The false detection was improved by adding DR-Up, and the leakage detection of cloud shadows was improved by adding MFEF. In contrast, MFFCD-Net achieved the optimal detection of broken clouds as well as the segmentation accuracy of the cloud boundaries after adding AFF. The ablation experimental results prove that the method proposed in this article can significantly improve the detection effect of cloud layers of different scales and the segmentation accuracy of cloud boundaries.

2) Ablation Experiment for Nighttime Imagery: Table VI shows the prediction accuracies at night after adding different modules, from which it can be seen that the modules designed in this article could effectively improve the OA and F1-score of cloud detection. DR-Up Block can improve OA, recall, and F1-score. The MFEF module could improve the OA precision, recall, and F1-score. The addition of the AFF showed that although recall decreased, the precision improved significantly. This means that AFF can reduce the oversegmentation of Model3. Because AFF can fuse the encoder detail information and multiscale features after the MFEF module processing to filter out the lights, thus achieving better cloud detection at night.

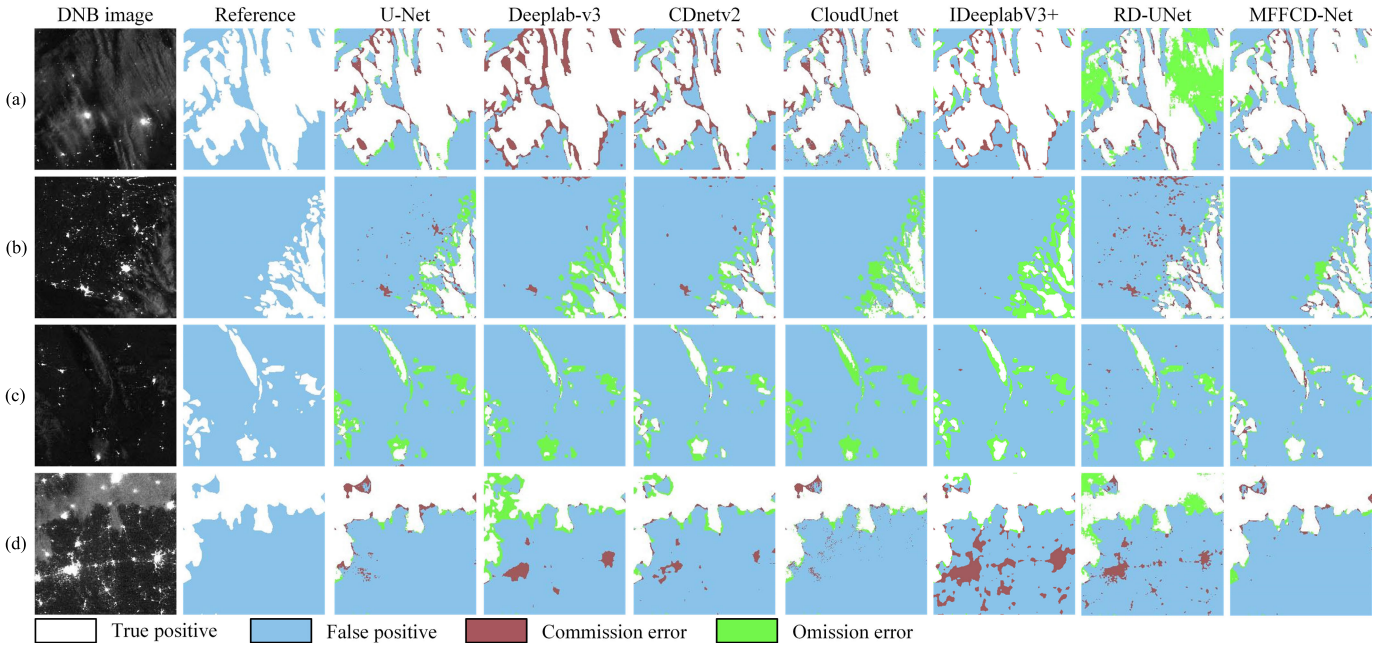

Fig. 7. Comparison of nighttime remote sensing cloud detection results by different methods.

TABLE IV
COMPARISON OF DETECTION EFFICIENCY OF DIFFERENT METHODS. THE BEST VALUES ARE MARKED IN BOLD

Method	U-Net	Deeplab-v3+	CDnetv2	CloudU-Net	IDeeplab-v3+	RD-UNet	MFFCD-Net
Efficiency (s)	**18**	18.9	32.1	138.5	18.2	26	26

TABLE V
COMPARISON OF PREDICTION ACCURACY OF DIFFERENT METHODS AT DAYTIME. THE BEST VALUES ARE MARKED IN BOLD

Method	DR-Up	MFEF	AFF	OA	Precision	Recall	F1-score
Model 1	×	×	×	85.6	85.3	86.2	85.7
Model 2	√	×	×	88.1	85.8	93.8	89.6
Model 3	√	√	×	88.7	85.2	**96.2**	90.4
MFFCD-Net	√	√	√	**92.1**	**90.8**	93.9	**92.3**

TABLE VI
COMPARISON OF PREDICTION ACCURACY OF DIFFERENT METHODS AT NIGHTTIME. THE BEST VALUES ARE MARKED IN BOLD

Method	DR-Up	MFEF	AFF	OA	Precision	Recall	F1-score
Model 1	×	×	×	80.8	85.5	79.1	82.2
Model 2	√	×	×	85.9	81.5	92.5	86.7
Model 3	√	√	×	88.5	83.5	**95.6**	89.1
MFFCD-Net	√	√	√	**90.4**	**90.2**	91.5	**90.8**

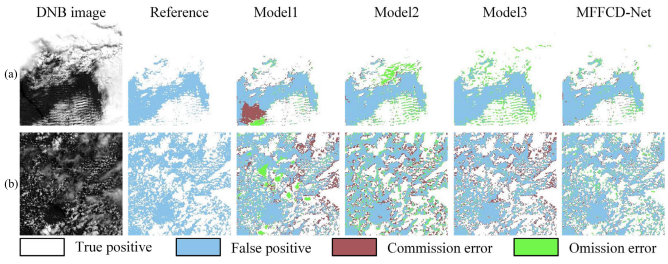

Fig. 8. Comparison of results from daytime cloud detection ablation experiments.

Fig. 9 shows the comparison of the nighttime cloud detection results after adding different modules in sequence. In Fig. 9(a), there are lights, and in the network without adding the module, it can be seen that the lights were misdetected as clouds. In the detection results obtained after the addition of the DR-UP and MFEF modules designed in this article, it can be seen that light misdetection was greatly improved. In Fig. 9(b), there are lights and clouds at the same time. The right side of the lights overlaps with the clouds, and the left side is only the lights. The detection results without adding the module were part of the misdetection and omission. After adding the MFEF module, cloud omission was greatly reduced, and at the same time, owing to the lack of detail information, the cloud-boundary segmentation effect was worse. Moreover, after adding the AFF module, we found that MFFCD-Net had the best cloud detection results. The ablation experimental results prove that the method proposed in this article can significantly improve the network's ability to discriminate lights and can also accomplish the task of luminous cloud detection.

3) Effectiveness of Feature Extraction: Fig. 10 shows the feature maps of last convolution layer of different models. The feature maps of Model1 are rough. Some thick and thin clouds got very low feature values in daytime image. It can be seen that with addition of the designed modules, the feature maps become more detailed. This means that feature-extraction ability is improved by introducing the designed modules. For

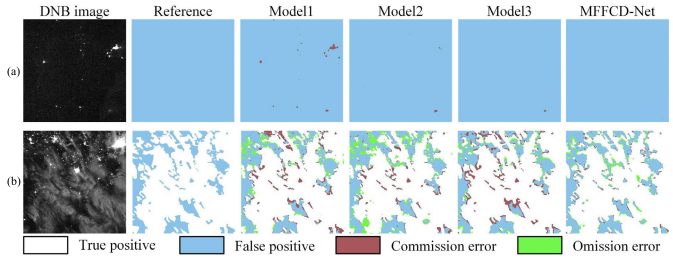

Fig. 9. Comparison of results from nighttime cloud detection ablation experiments.

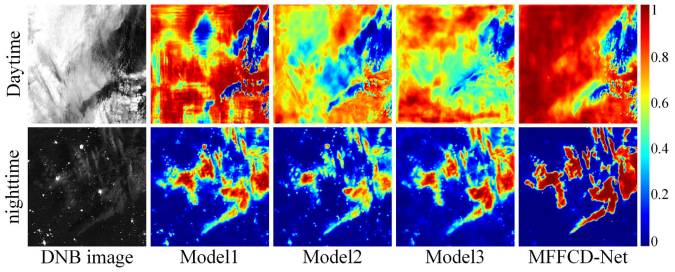

Fig. 10. Comparison of feature-extraction abilities of different models. The feature map represents the attention from clear pixels to cloudy pixels, with values between 0 and 1.

nighttime image, the ability of distinguishing light from cloud was improved by adding designed modules.

E. Comparison With ECM

The official cloud mask of VIIRS DNB images is ECM. In order to compare with ECM, two daytime and nighttime VIIRS DNB images and corresponding ECM products were downloaded. The ECM products and results of different methods are shown in Fig. 11. It can be seen that ECM always classified more background into cloud than deep learning-based methods. CloudU-Net, IDeeplabv3+, and RD-UNet misclassified some clouds into background in Fig. 11(a). U-Net, Deeplab-v3, IDeeplabv3+, and RD-UNet misclassified some lights into cloud in Fig. 11(c). U-Net, Deeplab-v3, and CDNetv2 misclassified background into cloud in Fig. 11(d), while MFFCD-Net can obtain more accurate cloud boundary than ECM and baseline methods. The results show the effectiveness of MFFCD-Net.

V. Discussion

The number of manually labeled cloud masks is very important for the performance of deep learning-based methods. Labeling cloud masks manually for remote sensing image, especially VIIRS DNB image, is very time-consuming because there is only one band in the DNB image and the feature difference between clouds and background in the DNB image is smaller than that in the multispectral image. The accuracy of deep learning-based methods also depends on the accuracy of manually labeled cloud masks. Because some backgrounds, such as barren in Fig. 10(a) and (d), have very similar brightness as clouds, it is more difficult for human to distinguish cloud and background in such scene. Also, the time for labeling such scene is much longer than other scenes such as water and vegetation. Image-level labels, such as whether

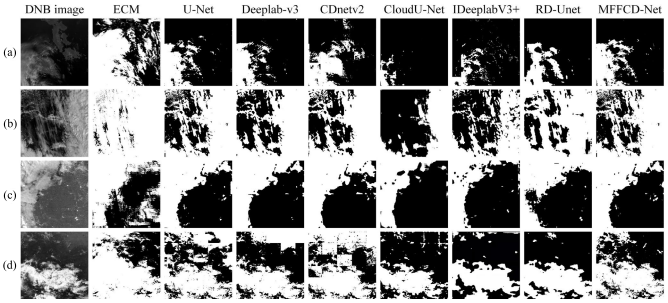

Fig. 11. Comparison of daytime and nighttime remote sensing images of different methods.

there is cloud in an image, are time-saving and easy to obtain, weakly supervised methods will be a good choice to address the above problems.

In Fig. 11, it can be seen that the cloud boundaries of ECM are rough. Many backgrounds are misclassified as clouds in ECM. This will reduce the available background data in the application of DNB image. This may be because ECM is produced by a Bayesian algorithm, which only uses the gray feature of the image. The proposed MFFCD-Net is constructed by CNN, which can extract gray, texture, and shape features simultaneously. This can improve the ability of MFFCD-Net for distinguishing cloud and background in difficult scene. The more accurate cloud masks produced by MFFCD-Net will increase the available data for the downstream applications. Although MFFCD-Net is trained on VIIRS DNB images, MFEF and AFF are designed for distinguishing cloud and artificial lights in nighttime remote sensing images. It is possible to apply MFFCD-Net on other nighttime satellites such as LuoJia1-01 [52], SDGSAT-1 [53], and Lookup-1 [54] with new training data.

Although MFFCD-Net can achieve high accuracy both on daytime and nighttime DNB images, the performance on daytime images is better than nighttime images. Some highlights signal will transfer thin cloud, which makes the feature of thin cloud unclear and is hard to detect thin cloud in this situation [red circle regions in Fig. 7(b) and (d)]. Artificial light is related to the distributions of road, building, and industry facility. In the future, we can explore improving the accuracy in nighttime images by combining the information of road, building, and industry facility, thus reducing the influence of artificial light.

VI. Conclusion

In this article, MFFCD-Net was proposed for cloud detection in daytime and nighttime remote sensing images. This is the first time that the same cloud detection model is designed to perform cloud detection both in daytime and nighttime VIIRS DNB images. MFFCD-Net injected the designed MFEF module, AFF module, and expanded residual upsampling module into U-Net architecture, 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 manually labeled cloud mask datasets of night remote sensing images, we produced a cloud detection dataset of day and night remote sensing images by manual labeling. The 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.

REFERENCES

[1] Y. Zhang, W. B. Rossow, A. A. Lacis, V. Oinas, and M. I. Mishchenko, "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. Geophys. Res., Atmos.*, vol. 109, no. D19, Oct. 2004, Art. no. D19105, doi: 10.1029/2003jd004457.

[2] A. Tayebi, S. Kasmaeeyazdi, F. Tinti, and R. Bruno, "Contributions from experimental geostatistical analyses for solving the cloud-cover problem in remote sensing data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 118, Apr. 2023, Art. no. 103236, doi: 10.1016/j.jag.2023.103236.

[3] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sens. Environ.*, vol. 118, pp. 83–94, Mar. 2012, doi: 10.1016/j.rse.2011.10.028.

[4] Z. Li, H. Shen, H. Li, G. Xia, P. Gamba, and L. Zhang, "Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery," *Remote Sens. Environ.*, vol. 191, pp. 342–358, Mar. 2017, doi: 10.1016/j.rse.2017.01.026.

[5] J. Li et al., "A hybrid generative adversarial network for weakly-supervised cloud detection in multispectral images," *Remote Sens. Environ.*, vol. 280, Oct. 2022, Art. no. 113197, doi: 10.1016/j.rse.2022.113197.

[6] Z. Wu, J. Li, Y. Wang, Z. Hu, and M. Molinier, "Self-attentive generative adversarial network for cloud detection in high resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1792–1796, Oct. 2020, doi: 10.1109/LGRS.2019.2955071.

[7] W. Li, Z. Zou, and Z. Shi, "Deep matting for cloud detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8490–8502, Dec. 2020, doi: 10.1109/TGRS.2020.2988265.

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

[9] R. A. Schiffer and W. B. Rossow, "The international satellite cloud climatology project (ISCCP): The first project of the world climate research programme," *Bull. Amer. Meteorolog. Soc.*, vol. 64, no. 7, pp. 779–784, Jul. 1983.

[10] L. L. Stowe et al., "Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data," *Adv. Space Res.*, vol. 11, no. 3, pp. 51–54, 1991, doi: 10.1016/0273-1177(91)90402-6.

[11] R. R. Irish, J. L. Barker, S. N. Goward, and T. Arvidson, "Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 10, pp. 1179–1188, Oct. 2006, doi: 10.14358/pers.72.10.1179.

[12] S. Qiu, Z. Zhu, and B. He, "Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery," *Remote Sens. Environ.*, vol. 231, Sep. 2019, Art. no. 111205, doi: 10.1016/j.rse.2019.05.024.

[13] E. Parmes, Y. Rauste, M. Molinier, K. Andersson, and L. Seitsonen, "Automatic cloud and shadow detection in optical satellite imagery without using thermal bands—Application to Suomi NPP VIIRS images over Fennoscandia," *Remote Sens.*, vol. 9, no. 8, p. 806, Aug. 2017, doi: 10.3390/rs9080806.

[14] B. Wang, A. Ono, K. Muramatsu, and N. Fujiwara, "Automated detection and removal of clouds and their shadows from Landsat TM Images," *IEICE Trans. Inf. Syst.*, vol. 82, no. 2, pp. 453–460, 1999.

[15] O. Hagolle, M. Huc, D. V. Pascual, and G. Dedieu, "A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images," *Remote Sens. Environ.*, vol. 114, no. 8, pp. 1747–1755, Aug. 2010, doi: 10.1016/j.rse.2010.03.002.

[16] Z. Zhu and C. E. Woodcock, "Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change," *Remote Sens. Environ.*, vol. 152, pp. 217–234, Jul. 2014, doi: 10.1016/j.rse.2014.06.012.

[17] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 130, pp. 277–293, Aug. 2017, doi: 10.1016/J.ISPRSJPRS.2017.06.001.

[18] J. Jang, A. A. Viau, F. Anctil, and E. Bartholomé, "Neural network application for cloud detection in SPOT VEGETATION images," *Int. J. Remote Sens.*, vol. 27, no. 4, pp. 719–736, Feb. 2006, doi: 10.1080/01431160500106892.

[19] L. Xu, A. Wong, and D. A. Clausi, "A novel Bayesian spatial–temporal random field model applied to cloud detection from remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4913–4924, Sep. 2017, doi: 10.1109/TGRS.2017.2692264.

[20] J. Wei et al., "Cloud detection for Landsat imagery by combining the random forest and superpixels extracted via energy-driven sampling segmentation approaches," *Remote Sens. Environ.*, vol. 248, Oct. 2020, Art. no. 112005, doi: 10.1016/j.rse.2020.112005.

[21] S. Mackie, O. Embury, C. Old, C. J. Merchant, and P. Francis, "Generalized Bayesian cloud detection for satellite imagery. Part 1: Technique and validation for night-time imagery over land and sea," *Int. J. Remote Sens.*, vol. 31, no. 10, pp. 2573–2594, May 2010, doi: 10.1080/01431160903051703.

[22] J. Hurley, A. Dudhia, and R. G. Grainger, "Cloud detection for MIPAS using singular vector decomposition," *Atmos. Meas. Techn.*, vol. 2, no. 2, pp. 533–547, Sep. 2009, doi: 10.5194/amt-2-533-2009.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[24] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[25] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2018, *arXiv:1706.05587*.

[26] J. Yang, J. Guo, H. Yue, Z. Liu, H. Hu, and K. Li, "CDNet: CNN-based cloud detection for remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6195–6211, Aug. 2019, doi: 10.1109/TGRS.2019.2904868.

[27] Z. Li, H. Shen, Q. Cheng, Y. Liu, S. You, and Z. He, "Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors," *ISPRS J. Photogramm. Remote Sens.*, vol. 150, pp. 197–212, Apr. 2019, doi: 10.1016/j.isprsjprs.2019.02.017.

[28] J. Li et al., "A lightweight deep learning-based cloud detection method for Sentinel-2A imagery fusing multiscale spectral and spatial features," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5401219, doi: 10.1109/TGRS.2021.3069641.

[29] X. Li, X. Yang, X. Li, S. Lu, Y. Ye, and Y. Ban, "GCDB-UNet: A novel robust cloud detection approach for remote sensing images," *Knowl.-Based Syst.*, vol. 238, Feb. 2022, Art. no. 107890, doi: 10.1016/j.knosys.2021.107890.

[30] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239, doi: 10.1109/CVPR.2017.660.

[31] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.

[32] B. Zhang, Y. Zhang, Y. Li, Y. Wan, and Y. Yao, "CloudViT: A lightweight vision transformer network for remote sensing cloud detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2022.3233122.

[33] N. Ma, L. Sun, Y. He, C. Zhou, and C. Dong, "CNN-TransNet: A hybrid CNN-transformer network with differential feature enhancement for cloud detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3288742.

[34] A. Obregon, C. Gehrig-Downie, S. R. Gradstein, and J. Bendix, "The potential distribution of tropical lowland cloud forest as revealed by a novel MODIS-based fog/low stratus night-time detection scheme," *Remote Sens. Environ.*, vol. 155, pp. 312–324, Dec. 2014, doi: 10.1016/j.rse.2014.09.005.

[35] Y. Liu, J. R. Key, R. A. Frey, S. A. Ackerman, and W. P. Menzel, "Nighttime polar cloud detection with MODIS," *Remote Sens. Environ.*, vol. 92, no. 2, pp. 181–194, Aug. 2004, doi: 10.1016/j.rse.2004.06.004.

[36] Y. Yang et al., "Machine learning-based retrieval of day and night cloud macrophysical parameters over east Asia using Himawari-8 data," *Remote Sens. Environ.*, vol. 273, May 2022, Art. no. 112971, doi: 10.1016/j.rse.2022.112971.

[37] C. Liu et al., "A machine learning-based cloud detection algorithm for the Himawari-8 spectral image," *Adv. Atmos. Sci.*, vol. 39, no. 12, pp. 1994–2007, Dec. 2022, doi: 10.1007/s00376-021-0366-x.

[38] Q. He, "Night-time cloud detection for FY-3A/VIRR using multispectral thresholds," *Int. J. Remote Sens.*, vol. 34, no. 8, pp. 2876–2887, Apr. 2013, doi: 10.1080/01431161.2012.755275.

[39] C. J. Merchant, A. R. Harris, E. Maturi, and S. Maccallum, "Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval," *Quart. J. Roy. Meteorolog. Soc.*, vol. 131, no. 611, pp. 2735–2755, Oct. 2005.

[40] K. Liu and Z. Kou, "The research on DMSP nighttime cloud image segmentation," in *Proc. 3rd Int. Conf. Comput. Res. Develop.*, Mar. 2011, vol. 3, pp. 143–145, doi: 10.1109/ICCRD.2011.5764265.

[41] C. Wang, S. Platnick, K. Meyer, Z. Zhang, and Y. Zhou, "A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations," *Atmos. Meas. Techn.*, vol. 13, no. 5, pp. 2257–2277, May 2020, doi: 10.5194/amt-13-2257-2020.

[42] L. Joachim and T. Storch, "Cloud detection for night-time panchromatic visible and near-infrared satellite imagery," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. V-2, pp. 853–860, Aug. 2020, doi: 10.5194/isprs-annals-v-2-2020-853-2020.

[43] A. Heidinger, D. Botambekov, and A. Walther, "A naive Bayesian cloud mask delivered to noaa enterprise, algorithm theoretical basis document," NOAA NESDIS Cent. Satell. Appl. Und Res., Silver Spring, MD, USA, Tech. Rep. Version 1.1., 2016. [Online]. Available: https://www.star.nesdis.noaa.gov/goesr/documents/ATBDs/Enterprise/ATBD_Enterprise_Cloud_Mask_v1.2_2020_10_01.pdf

[44] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005, doi: 10.1007/s10479-005-5724-z.

[45] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3194–3203, doi: 10.1109/CVPR.2018.00337.

[46] L. Mou and X. X. Zhu, "Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6699–6711, Nov. 2018, doi: 10.1109/TGRS.2018.2841808.

[47] F. Xie, M. Shi, Z. Shi, J. Yin, and D. Zhao, "Multilevel cloud detection in remote sensing images based on deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3631–3640, Aug. 2017, doi: 10.1109/JSTARS.2017.2686488.

[48] 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," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 700–713, Jan. 2021, doi: 10.1109/TGRS.2020.2991398.

[49] 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," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1688–1692, Oct. 2021, doi: 10.1109/LGRS.2020.3009227.

[50] J. Qian, J. Ci, H. Tan, W. Xu, Y. Jiao, and P. Chen, "Cloud detection method based on improved DeeplabV3+ remote sensing image," *IEEE Access*, vol. 12, pp. 9229–9242, 2024, doi: 10.1109/ACCESS.2024.3353205.

[51] A. Li, X. Li, and X. Ma, "Residual dual U-shape networks with improved skip connections for cloud detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024, doi: 10.1109/LGRS.2023.3337860.

[52] X. Li, X. Li, D. Li, X. He, and M. Jendryke, "A preliminary investigation of Luojia-1 night-time light imagery," *Remote Sens. Lett.*, vol. 10, no. 6, pp. 526–535, Jun. 2019, doi: 10.1080/2150704x.2019.1577573.

[53] Z. Lin et al., "Modelling the public perception of urban public space lighting based on SDGSAT-1 glimmer imagery: A case study in Beijing, China," *Sustain. Cities Soc.*, vol. 88, Jan. 2023, Art. no. 104272, doi: 10.1016/j.scs.2022.104272.

[54] X. Zhu et al., "Assessment of a new fine-resolution night-time light imagery from the Yangwang-1 ('look up 1') satellite," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3139774.

Jun Li received the B.S. degree in remote sensing science and technology, the M.S. degree in geomatics engineering, and the Ph.D. degree in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2015, 2018, and 2021, respectively.

He is currently an Associate Research Fellow with the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests include remote sensing image processing and deep learning.

Chengjie Hu received the B.S. degree in optoelectronic information science domain engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2018, where he is currently pursuing the Ph.D. degree in aeronautics and astronautics science and technology.

His research interests include remote sensing image processing and deep learning.

Qinghong Sheng received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2008.

She is currently a Professor with the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, China. Her research interests include aerospace photogrammetry and intelligent remote sensing image processing.

Bo Wang received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2015.

He is currently an Assistant Professor with the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests include spatial information extraction and remote sensing image processing.

Xiao Ling received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2017.

He is currently an Assistant Research Fellow with the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests include remote sensing image processing and 3-D reconstruction.

Fan Gao received the B.S. degree in remote sensing science and technology, the M.S. degree in photogrammetry and remote sensing, and the Ph.D. degree in cartography and geographical information engineering from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2016, 2018, and 2022, respectively.

He is currently a Lecturer with the Institute of Communication Engineering, Army Engineering University of PLA, Nanjing, China. His research interests include parallel geoprocessing, EO data infrastructure, and GIService.

Yunfei Xu, photograph and biography not available at the time of publication.

Zhiwei Li received the B.Eng. degree in geoinformation science and technology from China University of Geosciences, Wuhan, China, in 2015, and the Ph.D. degree in cartography and geographic information engineering from Wuhan University, Wuhan, in 2020.

Since July 2022, he has been a Research Assistant Professor with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong. Prior to his current position, he worked as a Post-Doctoral Fellow with Wuhan University from July 2020 to June 2022. His research primarily focuses on remote sensing of cloudy and rainy environments. The specific research interests include urban remote sensing, cloud detection and removal, multisource data fusion, land cover/use mapping, and flood monitoring. He was or is the PI of one project funded by the National Natural Science Foundation of China and two projects funded by China Postdoctoral Science Foundation.

Matthieu Molinier received the Engineering degree from the École Nationale Supérieure de Physique de Strasbourg (ENSPS), Illkirch-Graffenstaden, France, in 2004, and the M.Sc. degree in image processing from the Université Louis Pasteur (ULP), Strasbourg, France, in 2004.

Since 2004, he has worked at VTT, as a Research Scientist in Earth observation, with a main focus on machine learning and change detection for optical satellite images. He is the author of 60 articles in scientific journals and conferences. His research interests include deep learning for multispectral and hyperspectral images, unsupervised change detection, and satellite image time-series analysis, with applications to environment monitoring.