Sensors and Materials, Vol. 36, No. 10 (2024) 4473–4482
MYU Tokyo

S & M 3812

Matrix Optimization Cloud Detection Algorithm Based on Multiple Receptive Fields

Yifei Cao,[1] Yingqi Bai,[2*] Yang Lantao,[3] and Jiannan Shi[2]

[1]Beijing Institute of Surveying and Mapping,
No. 15, Yangfangdian Road, Haidian District, Beijing 100038, China
[2]SpaceWill Info. Co., Ltd.,
Building 5, Yard A2, Xisanhuan North Road, Haidian District, Beijing 100081, China
[3]Zhongke Landscape (Beijing) Technology Information Co., Ltd.,
No. 6 East Automobile Museum Road, Fengtai District, Beijing 100071, China

(Received August 9, 2024; accepted October 21, 2024)

Keywords: optical satellite data, cloud detection

In the process of remote sensing data processing, cloud data will have a considerable negative impact on data processing. Therefore, a matrix optimization cloud detection algorithm based on multiple receptive fields is proposed in this paper. First, the algorithm adopts a block reshaping model to improve the efficiency of the algorithm, while reducing the need for hardware configuration. Second, the cloud region adaptive sparse attention matrix is used to improve the characteristics of cloud regions with different concentrations. Finally, the multi-receptive field scaling module is used to improve the ability to segment cloud regions at different scales. The experimental results show that the accuracy of the proposed algorithm is 85.5%, and the recall rate is 86.4%. The proposed algorithm can basically accurately screen the location of a cloud region and provide technical support for the subsequent image application processing.

1. Introduction

Optical satellites capture Earth's surface images using visible light and near-infrared electromagnetic waves, and about 66% of the Earth's surface is covered by clouds,(1) which makes optical satellites vulnerable to cloud occlusion in the imaging process, seriously reducing the quality of the acquired optical images, and thus disturbing the subsequent classification, segmentation, three-dimensional modeling, and other processing work. Rapid and accurate identification of cloud regions in optical images not only helps extend the satellite shooting window and improve the acquisition probability of cloud-free data, but also facilitates the design of solutions for preset cloud vulnerability and reduces the difficulty of data processing.

In most present studies, panchromatic, multispectral, and hyperspectral raw data are used for cloud detection, which has the following advantages: (1) they contain much detailed information on texture and geometry and (2) they have rich spectral characteristic information. However, there are also certain challenges. For cloud region detection tasks, owing to the large amount of

*Corresponding author: e-mail: 821099463@qq.com
https://doi.org/10.18494/SAM5289

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

redundant feature information in optical images and the high complexity of information, the cloud region segmentation task not only has stringent requirements for feature extraction and feature identification capabilities, but also usually produces a considerable waste of computing power.

To reduce information complexity while retaining a large number of effective features, some scholars currently use remote sensing preview images to perform cloud segmentation tasks, which are mainly divided into the following two categories: (1) based on traditional feature correlation algorithms and (2) based on neural network correlation algorithms. For tasks based on the traditional feature correlation algorithm, the cloud region is identified by extracting basic features such as geometry, texture, and color from the image. Zhang and Xizo proposed the construction of a significance map with statistical features to extract the cloud region scope.(2) Baseski and Seneras proposed the use of color and texture features to analyze initial color and edge features for cloud region detection.(3) Chen *et al.* realized cloud region detection through the stationary wavelet transform of hue-saturation-intensity (HSI) spatial color features and spectra.(4) Xia *et al.* proposed a method that combined an extreme learning machine and a K-nearest neighbor algorithm to extract 21 features such as texture, color, and shape to identify the cloud region range.(5) For tasks based on neural network correlation algorithms, a deep neural network method is adopted to extract the feature information of cloud regions in images by extracting a convolutional network to achieve the cloud detection task. Xie *et al.* proposed to combine a branch network and multiscale information to realize the cloud detection task.(6) Yang *et al.* proposed a cloud detection neural network based on boundary-thinning blocks and obtained good detection results.(7) Guo *et al.* designed a lightweight full convolutional neural network (ClouDet) with microarchitecture extension and separable volume integration modules to achieve cloud detection tasks, which performed well in most scenarios.(8)

However, in the cloud detection task using remote sensing preview images, the correlation algorithm based on traditional features has insufficient abilities in feature extraction and feature self-discrimination. Therefore, the detection accuracy needs to be improved in an actual task. The correlation algorithm based on neural networks can obtain more high-dimensional semantic information because of its high feature extraction ability. However, the cloud in an image often has a strong randomness in concentration and scale, and the information is considerably reduced under the premise of a large field of view of remote sensing preview images, which increases the difficulty of the detection task.

In this paper, we propose a cloud detection algorithm based on cloud region adaptive multi-receptive field combination. To improve the detection accuracy of the algorithm for scenarios with different concentrations and scales in the cloud region, first, the cloud region adaptive sparse attention matrix is proposed to solve the problem of different concentrations of clouds in a shrink map. By optimizing the features of different bands in the shrink map, we can enhance the ability of the algorithm to discriminate between the cloud region and the background. Second, to enhance the precision of cloud region detection across various scales, a multi-receptive field fusion coding module is employed. This module integrates features from different receptive fields, thereby enhancing the algorithm's capacity to recognize clouds ranging from small to large regions. Consequently, this enables the detection of all clouds within the image.

2. Our Work

U-Net is often used in the field of image segmentation. The network is composed of the encoder, decoder, and long-jump connection layer, showing a U-shaped structure. The specific structure is shown in Fig 1. The characteristic of the U-Net network is that the encoder structure extracts the shallow information of an image through downsampling, while the decoder obtains the deep information of the image through upsampling and fuses the shallow and deep features of the image through the long-jump connection layer to recover the image features and realize the segmentation of the remote sensing image.

The encoder gradually reduces the resolution of the feature map by subsampling in five stages to extract the shallow features of the image. The encoder consists of multiple convolution and pooling layers, presenting a structure of decreasing image height and width and increasing number of channels to capture contextual information. At each stage, the 3*3 convolution kernel, batch normalization, and ReLU activation function were used to extract features. After feature extraction, the height and width were reduced and the number of channels was increased. After sampling under the maximum pooling layer, the feature height width is halved and the number of channels is unchanged so that the network can obtain more image feature information.

The decoder recovers the resolution of the feature map gradually by upsampling in five stages to obtain deep information from the image. The decoder also consists of a series of convolutional layers and upper sampling layers, the structure of which corresponds to the encoder part. At each upper sampling layer, the resolution of the feature map is doubled by deconvolution operation, and the feature map generated at the five stages is spliced pixel by pixel through the long-jump connection layer, which realizes the multiscale feature information fusion of remote sensing image while retaining more spatial information and improves the accuracy of remote sensing image segmentation.

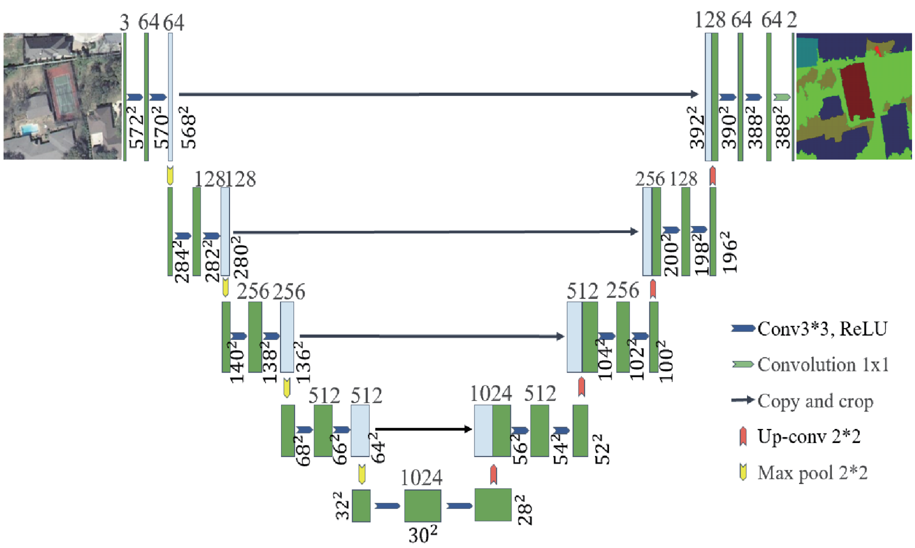

Fig. 1. (Color online) U-Net network structure.

3. Textual Algorithm

The algorithm in this paper is based on the Unet network. Aiming at the task of image segmentation of the cloud region, the block reshaping module is first adopted to preprocess large-size data. Then, the cloud region adaptive perception sparse matrix and multi-field fusion coding module are proposed to improve the sensitivity of the algorithm to the cloud region and the stability of cloud region discrimination under multiscale. The algorithm framework is shown in Fig. 2 below.

3.1 Block remodeling model based on large field of view

Remote sensing preview images, as raster representations derived from the downsampling of raw satellite sensor data, play a crucial role in the preliminary evaluation of image quality. They

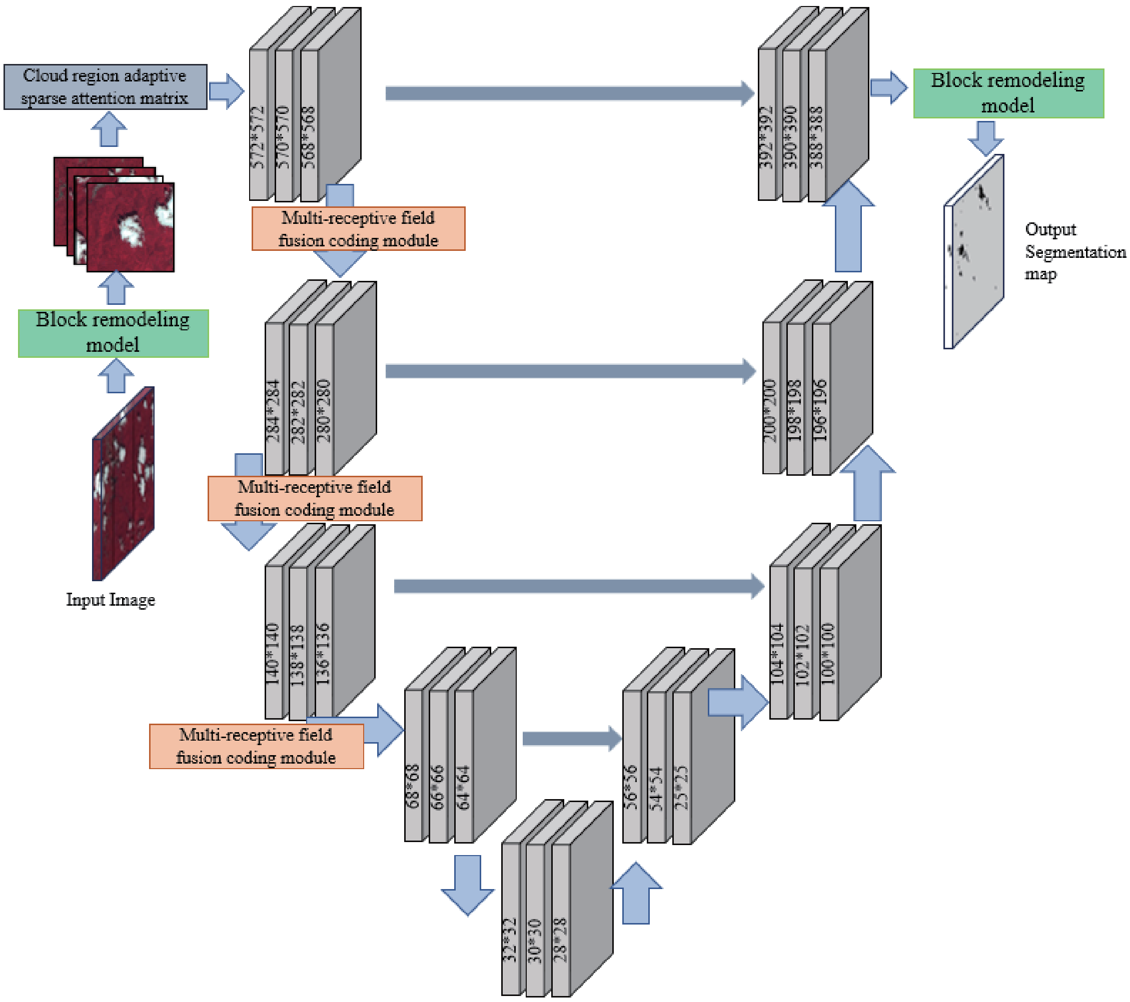

Fig. 2. (Color online) Algorithm framework.

offer a swift and comprehensive view of critical factors such as cloud coverage and surface characteristics, facilitating the determination of whether the acquisition of additional high-resolution original images is warranted. Distinct from standard remote sensing imagery, the reduced spatial resolution of preview images enables more efficient cloud segmentation processes, leading to a decrease in computational expenses. Moreover, unlike images captured by handheld devices such as mobile phones and cameras, remote sensing preview images encompass vastly larger geographical areas, thereby necessitating the implementation of image tiling methodologies to manage the substantial data inputs and mitigate the risk of memory overload during single-batch processing.

Compared with remote sensing images, the data volume of remote sensing preview images is considerably reduced, but it is usually larger than the size of conventional images. Therefore, it is necessary to adopt the method of image segmentation to alleviate the memory overload caused by the large input data of a single batch.

Compared with surface objects such as vehicles, buildings, and structures, cloud generation is clustered, and its distribution is in the form of curtains and lobes. Therefore, the size reduction will not considerably interfere with recognition. According to the characteristics of the cloud region, the following formula is used to divide the remote sensing preview image into blocks:

$$K_w^n = floor\left(K_w \,//\, K'_w\right), \tag{1}$$

$$K_w^n = floor\left(K_h \,//\, K'_h\right), \tag{2}$$

where K_w, K_h is the original size of the remote sensing preview image, K'_w, K'_h is the size of the subblock, K_w^n, K_w^n represents the final number of rows and columns, and $floor(\cdot)$ is the downward integer function.

3.2 Cloud region adaptive sparse attention matrix

The remote sensing preview image includes three bands: red, green, and blue, which have different characteristics in spectral reflection. Through dynamic perception of the cloud region, the ability of different bands to discriminate cloud regions with different concentrations is enhanced, and the weight of the cloud region is enhanced.

For each subblock(QM_{ij}), the three-band image is first transformed jointly between bands, and then the dynamic feature sparse attention matrix (DW) is obtained through adaptive pooling. The formula is as follows:

$$DW = GAP\left(SM\left(f\left(f\left(\mathrm{QM_{ij}}\right)_{3*3*3}\right)_{1*1*1}\right)\right), \tag{3}$$

where $GAP(\cdot)$ is the adaptive pooling layer, SM is the softmax activation function, and f_{n*n*c} represents c convolutional layers of $n*n$ size.

3.3 Multi-receptive field fusion coding module

The Unet network performs image segmentation through the U-shaped network and is successful in simple image segmentation. However, remote sensing images contain many details, which requires the network to process complex image scenes or images with fine-grained semantic information. At the same time, the Unet encoder is a local operation, and there are some deficiencies in the ability to control global information. In extreme cases, the cloud region will cover more than 90% of the image region. Therefore, in this paper, we propose a multi-field scaling module to improve the ability to segment the cloud region at different scales.

For each stage of the coding part, the multistage convolution kernel is first used to output the features of each layer, and the obtained features of different layers are combined by the dot multiplication method. The output is optimized by mean pooling and maximum pooling according to the depth of the layers where they are located. Finally, the results of different poolings are merged by concat. The shallow features are subjected to mean pooling to reduce the noise in the image. The deep features are maximized to improve the saliency of the features in the image. The formula is as follows:

$$FM_{sh} = Avg_{pool}\left(f\left(FM\right)_{3*3*2} \otimes f\left(FM\right)_{3*3*5}\right), \tag{4}$$

$$FM_{de} = Max_{pool}\left(f\left(FM\right)_{3*3*2} \otimes f\left(FM\right)_{3*3*8}\right), \tag{5}$$

$$Fm_{cod}^{i} = concat\left(FM_{sh}, FM_{de}\right), \tag{6}$$

where FM represents the input feature matrix, Avg_{pool} represents the mean pooling, Max_{pool} represents the maximum pooling, FM_{sh} and FM_{de} respectively represent the deep and shallow layer pooling feature maps, and $Fm^{i}_{\square}$ represents the phase i output feature map.

4. Experimental Results

4.1 Experimental data and environment

In this experiment, the self-made cloud region detection data set came from the Super View (SV) series satellite and Gao Fen (GF) series satellite, including SV-1, SVN2, GF2, and GF7 satellite preview images. Using Labelme software, we prepared visual interpretation and manual annotation. It was divided into the cloud region and cloud-free region from two aspects of cloud region scale and cloud region concentration, and the 512*512 block preview was adopted. After the block, a total of 800 images were divided. Some cloud areas were positive samples of the truth value, and the rest were negative samples of the background. Labelme was manually marked in json format, including the spatial position information of each pixel and the classification and naming of the image, and converted into the general png format of the model. The ratio of the training set to the test set was 8:2, with 640 pieces of the training set and 160 pieces of the test set.

In the experiment, we used the PC environment, the CPU was Intel Core i7-7700K, the GPU was NVidia RTX2060, the operating system was Windows 11, and the algorithm framework was Pytorch1.4. To improve the generalization performance of the model, a small batch of random gradient descent was used for offline training. The batch was 8, and a total of 50 iterations were performed. The learning rate attenuation strategy adopts a multistep attenuation strategy, from 10^{-2} to 10^{-5} gradual attenuation.

4.2 Experimental results

To verify the detection accuracy of the proposed algorithm for the cloud region, in the experiment, we randomly adopted high-view (SV) and high-score (GF) series preview images to detect the cloud region in the image, and the detection results are shown in Fig. 3.

As can be seen in Fig. 3, the proposed algorithm has a good detection capability for dense clouds and basically realizes the efficient segmentation of cloud areas, and the edge is basically consistent with reality. As evident from the first and second columns in Fig. 3, the proposed methodology demonstrates a robust capability in detecting small, punctate clouds. Furthermore, as observed in the fourth column, it effectively identifies extensive, laminar cloud formations. Consequently, the present approach exhibits commendable stability in detecting clouds of varying scales. However, there are still improvements to be made in the detection of very small fragmentary clouds and thin clouds. In the third column of Fig. 3, thin clouds visible on the ground were not successfully detected.

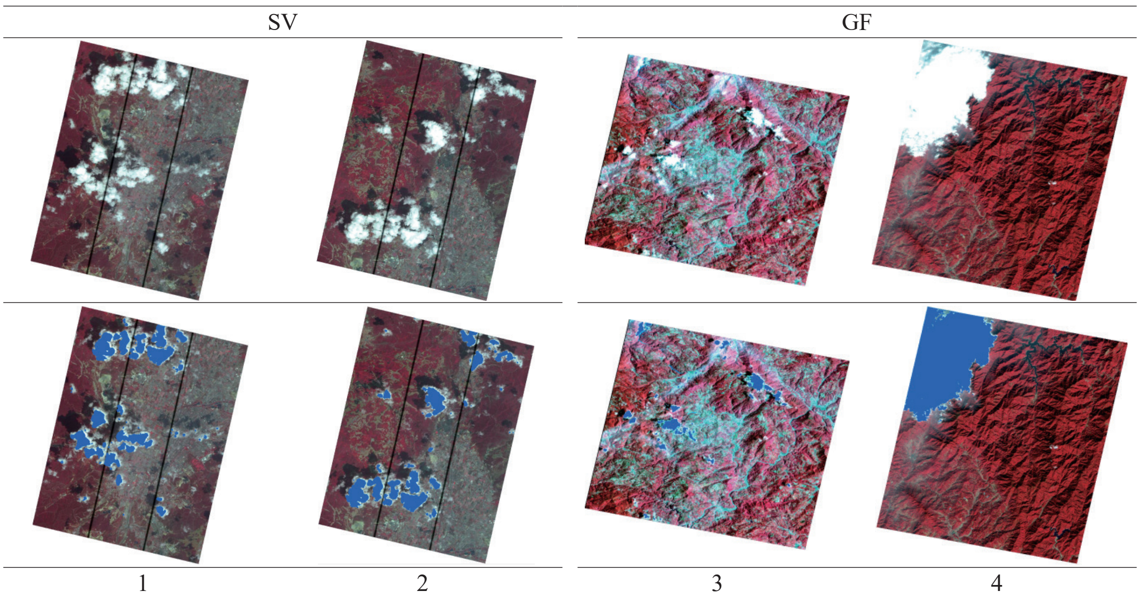

Fig. 3. (Color online) Detection result (blue is the algorithm detection cloud area).

Table 1
Experimental comparison of different algorithms

Methods	ACC (%)	Precision (%)	Recall (%)	MIOU (%)
SegNet	73.1	73.1	74.1	71.1
PSPNet	83.8	82.9	83.1	80.5
Unet	84.5	83.4	83.3	83.4
CSK	67.2	65.0	66.9	62.6
Sturck	63.2	62.7	61.3	61.6
Textual method	87.6	85.5	86.4	82.6

4.3 Experimental analysis

To verify the effectiveness of the proposed algorithm, three methods, U-net,(9) SegNet,(10) and PSPNet,(11) which are widely used in cloud detection networks at present, were selected for testing in self-made data sets, and compared with the proposed algorithm.

As can be seen in Table 1, first, the performance of Segnet is slightly lower than those of the other algorithms in terms of various indicators, and the reason is that SegNet retains insufficient feature information, resulting in rough edge segmentation, and there are certain problems of missing judgment and wrong judgment. Second, PSPNet and Unet have similar segmentation performances, and the precisions of both reach above 82%, among which the Unet network has the highest average crossover ratio among all the models. However, for cloud detection tasks, the overall accuracy of segmentation needs to be improved in complex scenarios. The proposed algorithm significantly outperforms traditional cloud detection methods across four key metrics. By enhancing the Unet network, our algorithm exhibits notable advantages over three other deep-learning-based cloud detection algorithms, particularly in terms of ACC, precision, and recall. Collectively, these diverse metrics underscore the superior performance of our algorithm in cloud detection, demonstrating its effectiveness and robustness.

5. Conclusion

To improve the segmentation accuracy of cloud detection in satellite images due to the strong randomness of cloud concentration and large-scale gap, a cloud detection algorithm based on a cloud adaptive multi-field combination is proposed in this paper. The algorithm mainly includes the block reshaping model, cloud region adaptive sparse attention matrix, and multi-receptive field fusion coding module. Among them, the block reshaping model can alleviate the memory overload caused by the large input data of a single batch and reduce the hardware requirements of the algorithm.The multi-receptive field fusion coding module achieves the segmentation of clouds of different sizes by combining receptive fields of multiple scales. The cloud region adaptive sparse attention matrix is enhanced by feature optimization, which effectively improves the ability to detect different cloud regions. The experiment verifies that the algorithm proposed in this paper has good detection performance for cloud detection tasks and can effectively extract the cloud region more accurately. However, further research is needed to compare small-scale broken clouds and thin clouds.

References

1 M. D. King, S. Plantnick, W. P. Menael, S. A. Ackerman, and P. A. Hubanks: Proc. IEEE Trans. Geosci. Remote Sens. **51** (2013) 3826.
2 Q. Zhang and C. Xiao: IEEE Trans. Geosci. Remote Sens. **52** (2014) 7264. https://doi.org/10.1109/TGRS.2014.2310240
3 E. Baseski and C. Senaras: Proc. 2015 7th Int. Conf. Recent Advances in Space Technologies (RAST) (IEEE, 2015) 311. https://doi.org/10.1109/RAST.2015.7208361
4 Z. Chen, T. Deng, H. Zhou, and S. Luo: Proc. SPIE 8919, Mippr: Pattern Recognition & Computer Vision. International Society for Optics and Photonics (2013) 891907. https://doi.org/10.1117/12.2030857
5 M. Xia, W. Lu, J. Yang, Y. Ma, W. Yao, and Z. Zheng: Neurocomputing **160** (2015) 238. https://doi.org/10.1016/j.neucom.2015.02.022
6 F. Xie, M. Shi, Z. Shi, J. Yin, and D. Zhao: IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **10** (2017) 3631. https://doi.org/10.1109/JSTARS.2017.2686488
7 J. Yang, J. Guo, H. Yue, Z. Liu, H. Hu, and K. Li: IEEE Trans. Geosci. Remote Sens. **57** (2019) 6195. https://doi.org/10.1109/TGRS.2019.2904868
8 H. Guo, H. Bai, and W. Qin: IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **14** (2021) 9743. https://doi.org/10.1109/JSTARS.2021.3114171
9 O. Ronneberger, P. Fischer, and T. Brox: Proc. Int. Conf, Medical Image Computing and Computer-Assisted Intervention (2015) 234. https://doi.org/10.1007/978-3-319-24574-4_28
10 V. Badrinarayanan, A. Kendall, and R. Cipolla: IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 2481. https://doi.org/10.1109/TPAMI.2016.2644615
11 H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2017) 6230. https://doi.org/10.1109/CVPR.2017.660

About the Authors

Yifei Cao received his B.E. degree in surveying and mapping engineering from Beijing University of Civil Engineering and Architecture, Beijing, China, in 2019 and his master's degree from Beijing University of Civil Engineering and Architecture in 2022. Since 2022, he has been an assistant engineer at Beijing Institute of Surveying and Mapping, China. His current research interests include target detection and tracking in remote sensing. (1023362448@qq.com)

Yingqi Bai received his master's degree in surveying and mapping engineering from Beijing University of Civil Engineering and Architecture, Beijing, China, in 2022. He is currently the junior project manager of SpaceWill Info Co., Ltd. His research interests include object detection tracking in remote sensing. (byq15624984098@126.com)

Lantao Yang graduated from Beijing University of Civil Engineering and Architecture in 2018 with a bachelor's degree in Geographic Information System Science. She is currently working in data support and operation and maintenance. Her current research interests include intelligent interpretation of remote sensing and GIS management. (1436720195@qq.com)

Shi Jiannan graduated from Beijing Jiaotong University in 2013, majoring in business administration. He is currently the intermediate project manager of SpaceWill Info Co., Ltd. His current research interests include intelligent interpretation of remote sensing.