An Automatic Cloud Detection Neural Network for High-Resolution Remote Sensing Imagery With Cloud–Snow Coexistence

Yang Chen, Qihao Weng, *Fellow, IEEE*, Luliang Tang, Qinhuo Liu, *Senior Member, IEEE*, and Rongshuang Fan

***Abstract*— Cloud detection is a crucial procedure in remote sensing preprocessing. However, cloud detection is challenging in cloud–snow coexisting areas because cloud and snow have a similar spectral characteristic in visible spectrum. To overcome this challenge, we presented an automatic cloud detection neural network (ACD net) integrated remote sensing imagery with geospatial data and aimed to improve the accuracy of cloud detection from high-resolution imagery under cloud–snow coexistence. The proposed ACD net consisted of two parts: 1) feature extraction networks and 2) cloud boundary refinement module. The feature extraction networks module was designed to extract the spectral–spatial and geographic semantic information of cloud from remote sensing imagery and geospatial data. The cloud boundary refinement module is used to further improve the accuracy of cloud detection. The results showed that the proposed ACD net can provide a reliably cloud detection result in cloud–snow coexistence scene. Compared with the state-of-the-art deep learning algorithms, the proposed ACD net yielded substantially higher overall accuracy of 97.92%. This letter provides a new approach to how remote sensing imagery and geospatial big data can be integrated to obtain high accuracy of cloud detection in the circumstance of cloud–snow coexistence.**

***Index Terms*— Cloud detection, deep learning, remote sensing imagery, geospatial big data.**

I. Introduction

With the development of high-resolution satellite sensors (IKONOS, QuickBird, Ziyuan-3, Gaofen-1, etc.), the optical remote sensing technology has gained a wider application in earth system science, such as global vegetation monitoring, target detection and monitoring, and global water monitoring [1].

Manuscript received April 20, 2021; revised July 5, 2021; accepted August 3, 2021. Date of publication August 20, 2021; date of current version December 29, 2021. This work was supported in part by the National Key Research and Development Plan of China under Grant 2017YFB0503604 and Grant 2016YFE0200400, and in part by the National Natural Science Foundation of China under Grant 41971405 and Grant 41671442. *(Corresponding authors: Qihao Weng; Luliang Tang.)*

Yang Chen and Luliang Tang are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: cy1017@whu.edu.cn; tll@whu.edu.cn).

Qihao Weng is with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong (e-mail: qihao.weng@polyu.edu.hk).

Qinhuo Liu is with the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China (e-mail: liuqh@radi.ac.cn).

Rongshuang Fan is with the Chinese Academy of Surveying and Mapping, Beijing 100830, China (e-mail: fanrsh@casm.ac.cn).

Digital Object Identifier 10.1109/LGRS.2021.3102970

However, optical remote sensing imagery often suffers from blurring by clouds and cloud shadows [2], restricting the application of remote sensing images. The cloud detection is thus an important task in remote sensing image processing.

The rule-based methods, such as automatic multifeature combined (MFC), and Fmask, are widely used to detect clouds [3], [4]. The MFC method is used to extract clouds and cloud shadows from GaoFen-1 multispectral imagery, and it was found that the overall accuracy (OA) of the MFC method yielded higher than 96.8% in vegetation regions [3]. Zhu *et al.* [4] developed the Fmask method for cloud, cloud shadow, and snow detection from Landsat images and found that the average OA of Fmask was as high as 89.0%. However, these rule-based methods, using low-level features, such as spectral, textural, and geometrical features, to extract cloud, may not be effective because these rule-based methods cannot extract high-level semantic features of cloud, leading to serious extraction errors.

Recently, deep learning methods, especially deep convolutional neural networks (CNNs), are gaining wide applications in cloud detection [5], water extraction [6], change detection [7], and so on. Deep CNN can extract high-level abstract features from input images using different sizes of convolution kernels, making it suitable for cloud detection from remote sensing images [8]–[10]. Wang *et al.* [11] used a novel deep CNN to distinguish cloud and snow from high-resolution multispectral imagery at the object level. It is found that the performance of the proposed method relied on segmented objects. Zhan *et al.* [12] developed a fully CNNs (FCNN) to detect cloud and snow at the pixel level and found that the OA of the FCNN method yielded higher than 91.4%. Xia *et al.* [13] developed a multidimensional deep residual network (M-ResNet) for cloud and snow detection from multispectral satellite imagery. It is found that M-ResNet can achieve an OA of 90.28% for the HJ-1A/1B satellite imagery. Yang *et al.* [14] used modified CNN with an encoder–decoder structure for cloud detection from high-resolution remote sensing imagery. They found that the modified CNN cannot cope well with the issue of cloud and snow detection.

However, the above-mentioned deep learning methods may not be effective for cloud detection when cloud and snow coexist. This is especially true when researches use only reflectance data of remote sensing images and ignore the use of prior geospatial information (altitude and location). Previous studies suggest that prior geospatial information of class events can be beneficial in satellite image classification to improve

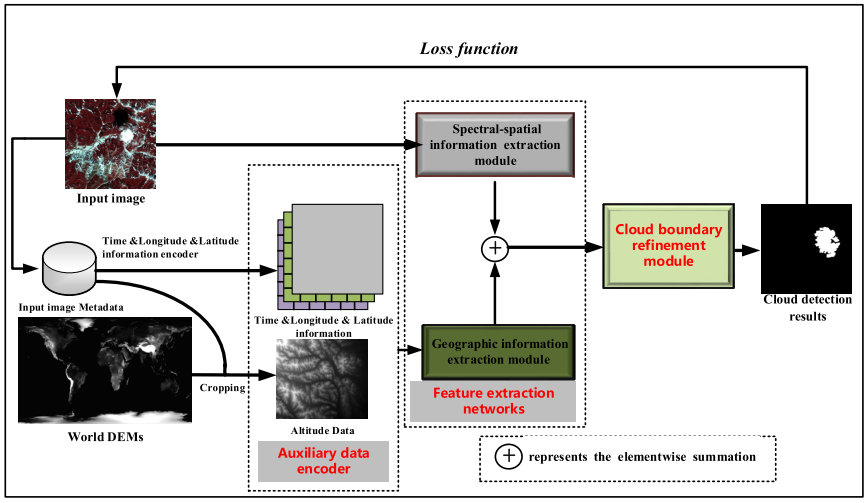

Fig. 1. Flowchart of the proposed cloud detection neural network.

the classification accuracy [15]–[17]. In addition, the existing deep learning methods often cannot cope well with the issue of cloud boundaries blurring, owning to stride convolution and pooling operations [9], [10].

In this letter, we explored how remote sensing imagery and geospatial big data can be integrated for improving the accuracy of cloud detection in an image scene of cloud and snow coexistence. Inspired by a multisource data fusion idea [18], this letter presented an automatic cloud detection neural network (ACD net) that integrated remote sensing image with geospatial data, aiming to improve the accuracy of cloud detection for high-resolution multispectral imagery in the case of cloud–snow coexistence. We proposed an ACD net with the spectral–spatial information extraction module, geographic information extraction module, and cloud boundary refinement module.

The main advantages of the proposed cloud detection neural network are included as follows.

1) An ACD net, with geospatial data and remote sensing data, is proposed for cloud detection from high-resolution satellite imagery with cloud–snow coexistence.
2) A feature extraction network, which consists of a spectral–spatial information extraction module and a geographic information extraction module, is designed to extract the spectral–spatial and geographic semantic information of cloud from remote sensing imagery and geospatial big data, aiming to distinguish cloud from snow.
3) The cloud boundary refinement module is applied to improve the accuracy of cloud detection.
4) The remainder of this letter is organized as follows: Section II introduces data preprocessing, the feature extraction networks, and cloud boundary refinement module. The experimental results are analyzed in Section III; and the conclusions are presented in Section IV.

II. Methodology

Fig. 1 describes the proposed method for cloud detection, which consists of three components: 1) auxiliary data encoder; 2) feature extraction networks; and 3) cloud boundary refinement module. The auxiliary data encoder was used to encode input image metadata (such as, longitude, latitude, time, and altitude) from four types of auxiliary maps. The feature extraction networks' model was designed to extract the spectral–spatial and geographic semantic information of cloud from remote sensing imagery and geospatial data, aiming to distinguish cloud from snow. The cloud boundary refinement module was used to improve the accuracy of cloud detection.

A. Data Preprocessing

To use geospatial data as auxiliary maps, an auxiliary data encoder was used. For the input image of size $M \times N$, we used an Affine transformation model [19] to generate the longitude map I_{Long} and the latitude map I_{Lat}. Given a certain pixel in row $Y(0 \le Y < M)$ and column $X(0 \le X < N)$, the corresponding longitude $I_{\text{Long}}(Y, X)$ and latitude $I_{\text{Lat}}(Y, X)$ can be expressed by the following equation:

$$\begin{aligned} I_{\text{Long}}(Y, X) &= I_{\text{Long}}(0, 0) + k_1 \times Y + k_2 \times X \\ I_{\text{Lat}}(Y, X) &= I_{\text{Lat}}(0, 0) + k_3 \times Y + k_4 \times X \end{aligned} \tag{1}$$

where k_1, k_2, k_3, and k_4 are the longitude and latitude pixel resolution in the input image, which can be obtained from the input image header file. $I_{\text{Long}}(0, 0)$ and $I_{\text{Lat}}(0, 0)$ are the longitude and latitude values of the top-left input image corner.

In general, snow is less likely to exist in summer (June, July, and August) for low-latitude or low-altitude areas, while clouds can exist in any season. We hypothesized the input image acquisition month t; time map I_{time} can be defined by the following equation:

$$I_{\text{time}} = \begin{bmatrix} a_t & a_t & \dots & a_t \\ a_t & a_t & \dots & a_t \\ \dots & \dots & \dots & \dots \\ a_t & a_t & \dots & a_t \end{bmatrix}_{M \times N} \tag{2}$$

$$a_t = \begin{cases} 0, & t \in [6, 8] \\ \left| \frac{t-8}{12} \right|, & t \in [1, 5] \cup [9, 12] \end{cases} \tag{3}$$

where a_t is the snow existence probability.

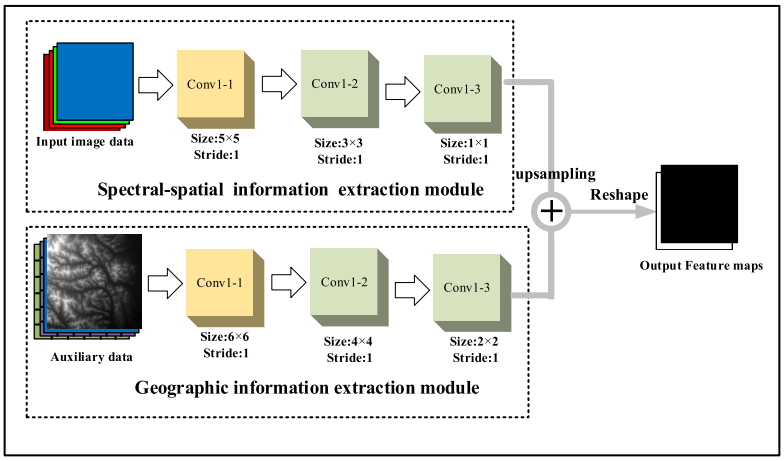

Fig. 2. Structure of the feature extraction networks.

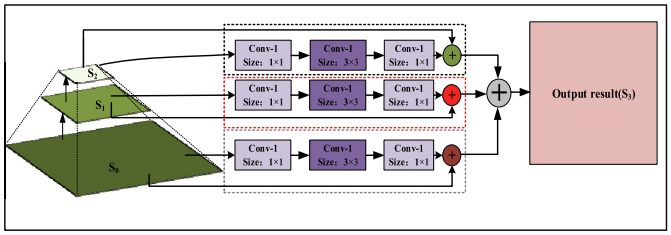

Fig. 3. Detail structure of the cloud boundary refinement module.

In addition, the strategy of pixel cropping was used to generate the altitude map I_{Alt} from the global Digital Elevation Models (http://viewfnderpanoramas.org/dem3.html). The final auxiliary map I was generated by the strategy of aggregation [22], which can be expressed by the following equation:

$$I = \mathrm{concat}(I_{\mathrm{Long}}, I_{\mathrm{Lat}}, I_{\mathrm{Alt}}, I_{\mathrm{time}}). \tag{4}$$

B. Feature Extraction Networks

Geographic and spectral–spatial semantic information on snow are generally distributed in high-latitude or high-altitude regions, which are widely used in optical satellite image processing of cloud detection [20]. This type of semantic information is key for distinguishing between cloud and snow. Following the two stream networks by Zhang and Chen [21], we built a feature extraction network to extract geographic and spectral–spatial semantic information from geospatial data and remote sensing imagery.

The feature extraction networks consisted of the spectral–spatial information extraction module and the geographic information extraction module. The former contained three convolution layers of sizes 5×5, 3×3, and 1×1, respectively, which were used to extract spectral–spatial semantic information from remote sensing images. The latter possessed three convolution layers of sizes 6×6, 4×4, and 2×2, respectively, which were used to extract semantic geographic information from geospatial data. Following the strategy of aggregated features [22], the extracted features were fused. The overall structure of the feature extraction networks is shown in Fig. 2.

C. Cloud Boundary Refinement Module

The proposed ACD net used convolution operations in feature extraction networks, potentially leading to the blurring of cloud boundaries. To solve this problem, postprocessing steps such as conditional random field (CRF) in CNN + CRF have been widely used to improve the accuracy of cloud detection. However, the CNN + CRF method cannot fully delineate global boundaries because the CRF model is suboptimal for the global schema inference process. A residual refinement module has been introduced to overcome this shortcoming of the CRF model in object boundary refinement processing [23].

However, boundary refinement models are achieved for both region and global boundaries in the score map of different spatial scales. Inspired by global-residual and local-boundary refinement networks' structure [24], we proposed a novel cloud boundary refinement module in this study to refine cloud boundaries at different scales. The overall structure of the cloud boundary refinement module is shown in Fig. 3. In mathematics, the cloud boundary refinement module can be expressed by the following equation:

$$S_3 = S_0 + S_1 + S_2 + R(S_0) + R(S_1) + R(S_2) \tag{5}$$

where S_3 is the refined score map, $R(S_0)$ is the S_0 of the residual branch, $R(S_1)$ is the S_1 of the residual branch, $R(S_2)$ is the S_2 of the residual branch, S_0 is the original coarse score map, and the coarse score map of S_1 and S_2 is half size of S_0 and one-quarter size of S_0, respectively.

The mean-squared error was used to compute the loss function, which is defined by the following equation:

$$\mathrm{Loss} = \frac{1}{2M} \sum_{i=1}^{M} \| X_1 - F_i \|_2^2 \tag{6}$$

where X_1 is the input data, F_i is the cloud label, and N is the number of cloud label.

III. Experimental Results

A. Experimental Data

The cloud dataset [3] consisted of 40 Gaofen-1 satellite scenes with four bands (Blue, Green, Red, and Near-infrared), which were mainly selected from Asia. The dataset was

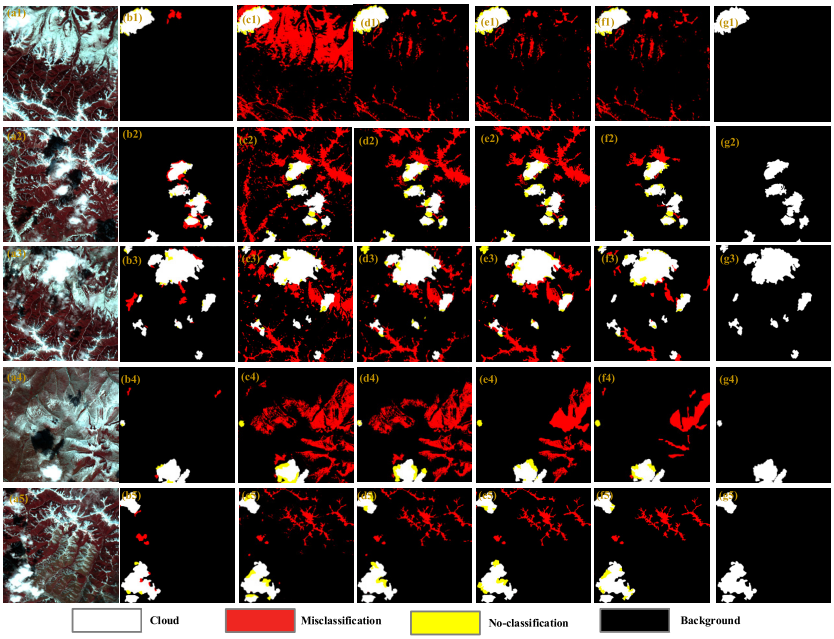

Fig. 4. Visual comparison of cloud detection results on cloud–snow coexistence scenes. (a1)–(a5) Original images. (b1)–(b5) Our method. (c1)–(c5) MFC method [3]. (d1)–(d5) CNN method [14]. (e1)–(e5) FCNN method [12]. (f1)–(f5) M-ResNet method [13].

divided into a training set of 30 scenes and a testing set of ten scenes. The ground-truth labels were obtained by manually digitizing in the ENVI platform and were composed of clouds and backgrounds. All 40 Gaofen-1 satellite scenes were cropped into image patch size of 512 × 512 pixels with a spatial resolution of 16 m. There were 10 000 image patches in the experimental dataset. The experimental dataset consisted of 7000 training, 1500 validation, and 1500 testing image patches. The image patches were selected to cover cloud–snow coexistence scenes and different types of cloud patterns. The different surface scenes were contained in the experimental images, and the testing areas covered typical cloud–snow coexistence scenes, aiming to test the performance of the proposed ACD net.

B. Network Training and Evaluation Metrics

We implemented our network on TensorFlow 2.0 (https://tensorflow.google.cn/install) and executed on a personal computer with an E3-1505M v6 at 3 GHz, 32 GB DDR4 memory, and an NVIDIA Quadro M2200. Training of the proposed method was implemented by ACD net loss function [see (5)] using the Adam optimization algorithm with a batch of 64 image patches, and the weight decay and momentum are 0.001 and 0.009, respectively. The proposed network was trained on the Gaofen-1 dataset, with an initial learning rate of 0.0005 for the first 100 epochs and 0.00025 for the next 100 epochs. In this experiment, training and testing of all comparison deep learning models were done using TensorFlow 2.0 in the same computing environments. All compared deep learning models were trained with the same training dataset.

We adopted widely used metrics to evaluate the performance of the proposed method, including OA, kappa coefficient (Kappa), producer accuracy (PA), user accuracy (UA), and intersection over union (IoU) [25]. All the evaluation metric scores were calculated based on the ground truth as the reference maps.

C. Results

To effectively validate the performance of the proposed ACD net, we compared the proposed ACD net, with the state-of-the-art algorithms, including MFC [3], CNN [14], FCNN [12], and M-ResNet [13]. The MFC method is the threshold segmentation of the cloud detection algorithm based on the multiple features and mask refinement for GF-1 wide-field-of-view (WFV) imagery. CNN, FCNN, and M-ResNet have been widely used in deep learning for cloud detection from optical satellite images.

Fig. 4 shows the results of cloud detection by the proposed ACD net along with the MFC, CNN, FCNN, and M-ResNet method under typical cloud–snow coexistence scenes. From Fig. 4(c1)–(c3), it is found that the MFC method seriously confuses snows with clouds. A large number of snows are misclassified as clouds in cloud–snow coexistence scenes by the MFC method since it cannot extract spectral–spatial semantic information. As is shown in Fig. 4(d1)–(d3), (e1)–(e3), and (f1)–(f3), the CNN, FCNN, and M-ResNet algorithms performs better when they were used to detect thick clouds and to distinguish between snow and cloud. Based on Fig. 4(b1) and (b2), the proposed ACD net can separate cloud from snow in three typical cloud–snow coexistence scenes, which is mainly attributed to integrating remote sensing imagery and geospatial data.

Table I provides quantitative evaluation results based on three types of cloud–snow coexistence scenes. Generally speaking, higher values of OA, Kappa, PA, UA, and IoU suggest good performance of a cloud detection algorithm. The proposed method yields an average OA of 97.85%, average Kappa of 91.09%, average PA of 93.21%, average UA of 90.29%, and average IoU of 91.04%, respectively.

TABLE I
QUANTITATIVE EVALUATION RESULTS OF FIVE CLOUD DETECTION METHODS

Method	OA (%)	Kappa (%)	PA (%)	UA (%)	IoU (%)
MFC	74.02	69.34	70.91	70.18	72.01
CNN	85.44	83.79	84.19	80.89	82.02
FCNN	91.87	85.96	84.81	81.16	85.21
M-ResNet	93.47	87.18	**93.95**	88.32	88.89
Our	**97.85**	**91.09**	93.21	**90.29**	**91.04**

Furthermore, the OA of the proposed ACD net is better than that of MFC, CNN, FCNN, and M-ResNet by 23.83%, 12.41%, 5.98%, and 4.38%, respectively. Compared with MFC, CNN, FCNN, and M-ResNet, the proposed ACD net achieves Kappa gain by more than 21.75%, 7.30%, 5.13%, and 3.91%, respectively. The IoU of the proposed ACD net is 19.03%, 9.02%, 5.83%, and 2.15% higher than MFC, CNN, FCNN, and M-ResNet, respectively.

Based on visual assessment and quantitative evaluation, the proposed ACD net can provide a reliable cloud detection result in cloud–snow coexistence scenes. In addition, this experiment suggests that geospatial data can be beneficial in satellite cloud detection task.

IV. CONCLUSION

In this letter, an ACD net is proposed for cloud detection from satellite imagery with cloud–snow coexistence. The proposed ACD net aims to solve the overestimation problem for cloud detection in cloud–snow coexistence scenes. Compared with previous cloud detection algorithms for optical satellite imagery, the proposed ACD net integrates both remote sensing imagery and geospatial data in cloud detection task with spectral–spatial information extraction module, geographic information extraction module, and cloud boundary refinement module. The experimental results show the proposed ACD net can provide a reliably cloud detection result, as demonstrated using the Gaofen-1 dataset.

The proposed ACD net obtained a reliable cloud detection result on Gaofen-1 images with cloud–snow coexistence. However, the proposed ACD net warrants further investigation for its applicability. First, the proposed ACD net was conducted with the Gaofen-1 image data and cannot be applied for other satellite images (such as Landsat, Sentinel-2, and QuickBird) without validation. Second, the proposed ACD net cannot cope well with the issue of thin cloud detection due to the insufficient training dataset of a variety of thin clouds.

REFERENCES

[1] J. Zhao, Y. Zhong, H. Shu, and L. Zhang, "High-resolution image classification integrating spectral-spatial-location cues by conditional random fields," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4033–4045, Sep. 2016.

[2] L. Sun *et al.*, "A cloud detection algorithm-generating method for remote sensing data at visible to short-wave infrared wavelengths," *ISPRS J. Photogram. Remote Sens.*, vol. 124, pp. 70–88, Feb. 2017.

[3] Z. Li, H. Shen, H. Li, G. Xia, 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.

[4] Z. Zhu, S. Wang, and C. E. Woodcock, "Improvement and expansion of the fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images," *Remote Sens. Environ.*, vol. 159, pp. 269–277, Mar. 2015.

[5] Y. Chen, Q. Weng, L. Tang, X. Zhang, M. Bilal, and Q. Li, "Thick clouds removing from multitemporal landsat images using spatiotemporal neural networks," *IEEE Trans. Geosci. Remote Sens.*, early access, Dec. 24, 2021, doi: 10.1109/TGRS.2020.3043980.

[6] Y. Chen, R. Fan, X. Yang, M. Wang, and A. Latif, "Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning," *Water*, vol. 10, no. 5, p. 585, 2018.

[7] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 24–41, Jun. 2016.

[8] Y. Chen, R. Fan, M. Bilal, X. Yang, J. Wang, and W. Li, "Multilevel cloud detection for high-resolution remote sensing imagery using multiple convolutional neural networks," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 5, p. 181, May 2018.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[10] 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.

[11] L. Wang, Y. Chen, L. Tang, R. Fan, and Y. Yao, "Object-based convolutional neural networks for cloud and snow detection in high-resolution multispectral imagers," *Water*, vol. 10, no. 11, p. 1666, Nov. 2018.

[12] Y. Zhan, J. Wang, J. Shi, G. Cheng, L. Yao, and W. Sun, "Distinguishing cloud and snow in satellite images via deep convolutional network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1785–1789, Oct. 2017.

[13] M. Xia, W. Liu, B. Shi, L. Weng, and J. Liu, "Cloud/snow recognition for multispectral satellite imagery based on a multidimensional deep residual network," *Int. J. Remote Sens.*, vol. 40, no. 1, pp. 156–170, Jan. 2019.

[14] 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.

[15] M. Simwanda and Y. Murayama, "Integrating geospatial techniques for urban land use classification in the developing sub-Saharan African city of Lusaka, Zambia," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 4, p. 102, Mar. 2017.

[16] M. Li, K. M. de Beurs, A. Stein, and W. Bijker, "Incorporating open source data for Bayesian classification of urban land use from VHR stereo images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 4930–4943, Nov. 2017.

[17] T. Hu, J. Yang, X. Li, and P. Gong, "Mapping urban land use by using landsat images and open social data," *Remote Sens.*, vol. 8, pp. 1–18, Feb. 2016.

[18] T. Wan, H. Lu, Q. Lu, and N. Luo, "Classification of high-resolution remote-sensing image using OpenStreetMap information," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2305–2309, Dec. 2017.

[19] S. Zhao *et al.*, "GDAL-based extend ArcGIS Engine's support for HDF file format," in *Proc. 18th Int. Conf. Geoinform.*, Beijing, China, Jun. 2010, pp. 1–3.

[20] Y. Chen *et al.*, "Cloud and cloud shadow detection based on multiscale 3D-CNN for high resolution multispectral imagery," *IEEE Access*, vol. 8, pp. 16505–16516, 2020.

[21] K. Zhang and Z. Chen, "Video saliency prediction based on spatial-temporal two-stream network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3544–3557, Dec. 2019.

[22] Z. Yu *et al.*, "Melanoma recognition in dermoscopy images via aggregated deep convolutional features," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 1006–1016, Apr. 2019.

[23] Y. Niu, G. Long, W. Liu, W. Guo, and S. He, "Boundary-aware RGBD salient object detection with cross-modal feature sampling," *IEEE Trans. Image Process.*, vol. 29, pp. 9496–9507, Oct. 2020.

[24] R. Zhang, S. Tang, M. Lin, J. Li, and S. Yan, "Global-residual and local-boundary refinement networks for rectifying scene parsing predictions," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 3427–3433.

[25] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process*, vol. 5, no. 2, pp. 1–11, 2015.