DCNet: A Deformable Convolutional Cloud Detection Network for Remote Sensing Imagery

Yang Liu, Wen Wang, *Member, IEEE*, Qingyong Li, *Member, IEEE*, Min Min, and Zhigang Yao

Abstract*— Recently, deep convolutional neural networks (CNNs) have made important progress in cloud detection with powerful representation learning capability and yield significant performance. However, most existing CNN-based cloud detection methods still face serious challenges because of the variable geometry of clouds and the complexity of underlying surfaces. It is attributed that they only use the fixed grid to extract contextual information, which lacks internal mechanisms to handle the geometric transformations of clouds. To tackle this problem, we propose a deformable convolutional cloud detection network with an encoder-decoder architecture, named DCNet, which can enhance the adaptability of a model to cloud variations. Specifically, we introduce deformable convolution blocks at the encoder to capture saliency spatial contexts adaptively based on the morphological characteristics of clouds and generate high-level semantic representations. After this, we incorporate skip-connection mechanisms into the decoder that integrate low-level spatial contexts as guidance to recover high-level semantic pixel localization and export precise cloud-detection results. Extensive experiments on the GF-1 wide field-of-view (WFV) Satellite Imagery demonstrate that DCNet outperforms several state-of-the-art methods. A public reference implementation of our proposed model in PyTorch is available at *https://github.com/NiAn-creator/deformableCloudDetection.git.

***Index Terms*— Cloud detection, deformable convolution block (DCB), encoder-decoder, semantic segmentation.**

I. Introduction

With the rapid development of remote sensing technology, high-resolution satellite imagery is available and has been widely used in environmental protection, resource exploration, military reconnaissance, and so on [1]. Because nearly 66% earth surface is covered by clouds [2], most optical imageries are inevitably contaminated and prevented from obtaining a clear view of the Earth's surface. Hence, it is essential and worthwhile to establish an efficient cloud-detection model.

Early cloud-detection algorithms mainly focus on handcrafted features, which are extracted from particular spectral bands or the front principal components of inputs. These methods detect clouds either utilizing an adaptive threshold in different spectral bands derived from the physical characteristics of clouds, such as spectrum, texture, temperature and elevation [3], [4], or utilizing manual features to train classifiers (e.g., support vector machines [5]). As the resolution of satellite remote sensing increases, images contain more complex spatial information and within-class spectral heterogeneity, making the traditional methods less feasible for practical applications.

Motivated by the great progress of deep learning in semantic segmentation tasks, convolutional neural network (CNN)-based algorithms are explored in the community of cloud detection to mine the discriminative representation of spatial details without manual prior assumptions. Dev *et al.* [6] cascaded some convolution and max-pooling layers to learn discriminative features for cloud detection. However, these consecutive convolutional strides or pooling operations reduce the spatial resolution and thus fail to capture the detailed spatial information. To alleviate this problem, Shi *et al.* [7] utilized the dilated convolution to increase the receptive field of filters and obtain more context information. Li *et al.* [8] proposed a global convolutional pooling to enhance the spatial representation ability of the feature map. Similarly, Yang *et al.* [9] and He *et al.* [10] designed a feature pyramid module to extracts multiscale contexts. Many other scholars [11]–[13] incorporated attention mechanisms into cloud-detection tasks to emphasize useful features. Through leveraging geometric and contextual information of the tagged images, these methods have achieved significant improvement on cloud-detection tasks.

In light of the specific morphological characteristics of cloud appearances varying from fiber to lump with irregular and blurry borders, these existing CNN-based methods are still insufficient to encode the complex geometry deformation of clouds as shown in Fig. 1(a) and (b). It can be attributed that they utilize fixed geometric filters and merely augment the spatial sampling locations without adequate supervision. Although for various textures, shapes, and scales of clouds, it is vital to adaptively determine the scales or receptive field sizes. Moreover, the diversity of the background also brings great difficulties for cloud detection because of the high

Manuscript received March 24, 2021; revised May 24, 2021; accepted May 26, 2021. Date of publication June 21, 2021; date of current version January 5, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62006017 and in part by the Fundamental Research Funds for the Central Universities under Grant 2020JBZD010 and Grant GFZX0404120313. *(Corresponding author: Qingyong Li.)*

Yang Liu, Wen Wang, and Qingyong Li are with the Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China (e-mail: 19112025@bjtu.edu.cn; wangwen@bjtu.edu.cn; liqy@bjtu.edu.cn).

Min Min is with the School of Atmospheric Sciences, Sun Yat-Sen University, Zhuhai 519082, China (e-mail: minm5@mail.sysu.edu.cn).

Zhigang Yao is with the Beijing Institute of Applied Meteorology, Beijing 100029, China (e-mail: yzg_biam@163.com).

Digital Object Identifier 10.1109/LGRS.2021.3086584

1558-0571 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Southeast University. Downloaded on November 28,2022 at 07:12:11 UTC from IEEE Xplore. Restrictions apply.

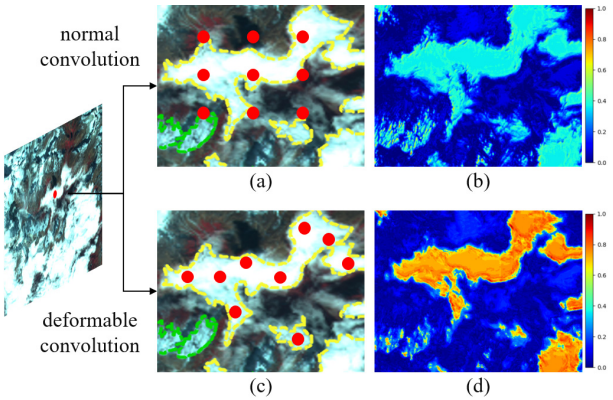

Fig. 1. Illustration of the sampling locations for (a) normal convolution and (c) deformable convolution. Note that the yellow curves of areas denote cloud and the green curves of areas denote snow. (b) and (d) Visualize the semantic activation weights after different convolutions. It is observed that deformable convolutions can adaptively extract context information of clouds and suppress the snow regions, but normal convolutions only describe the fixed receptive field.

similarity between clouds and bright regions, e.g., buildings in the middle of Fig. 4(c) and snow regions in the bottom left of Fig. 4(d). It requires that the sampling location of convolutional filters change dynamically with different image contents and are adapted to the exact cloud object rather than staying the same no matter for clouds or other distractions.

In this letter, we propose a novel deformable convolutional cloud-detection network (DCNet) to enhance the deformation modeling capability of the encoder-decoder architecture and facilitate the model to conform more closely to cloud structure. DCNet is designed to have a similar U-shaped architecture. The encoder converts input images into a dense representation, thus robust category information is more easily captured in the deeper level. The decoder is designed to perform the inverse process of encoding, where cloud details and spatial localization are gradually recovered. Especially, DCNet integrates deformable convolution blocks (DCBs) [14] into cloud-detection tasks. Different from previous CNN-based modules, DCB augments the flexibility of spatial sampling locations with learnable 2-D offsets. As illustrated in Fig. 1, DCB enables free form deformation of the sampling grid. To summarize, the main contributions of this work are listed as follows:

1) We propose a DCNet, which is based on encoder-decoder architecture, to model the morphological characteristics of clouds.
2) We explore the adaptability of deformable convolutional blocks and facilitate the DCNet to conform more closely to cloud structure.
3) Experiments on the GF-1 wide field-of-view (WFV) Satellite Imagery illustrate that DCNet achieves significant performance with state-of-the-art algorithms.

The remainder of this letter is organized as follows: Section II describes the details of the proposed DCNet framework. Experimental results are presented in Section III, followed by our conclusions in Section IV.

II. Method

DCNet is proposed to construct a self-adaptive framework for high-accuracy pixel-level cloud detection through deformable feature extraction modules. In this section, we will describe the overall framework of the proposed DCNet, as shown in Fig. 2. The key components, i.e., the DCB, loss function involved, are presented in detail.

A. Overview of DCNet

As discussed in Section I, the proposed DCNet framework incorporates the encoder-decoder architecture with the deformable convolution mechanism to tackle with the specific challenges for cloud detection in complex appearance variations and highly similar distractions. Specifically, given remote sensing imagery $\mathcal{I}$, we stack the visible and near-infrared bands as combination input to leverage various spectral and temperature information. Then, we take each pixel of combination images as the basic research unit and construct an encoder and a decoder symmetrically. As shown in Fig. 2, the encoder contains DCB, max-pooling, and down-sampling. Different from the traditional fixed convolution, we attach a DCB at the shallow layer of the encoder to catch the low-level representation implied in the geometric transformations of clouds and change the filter sampling locations with variations of the cloud adaptively. Following that, max-pooling and down-sampling are performed to extract salient features and encode global contexts. Considering the wide distribution of clouds in images, we perform four subsamplings to balance the spatial details and contextual semantic during the encoding phase. Symmetrically, the decoder includes up-sampling, skip-connection, and DCB. To maintain the local consistency of pixel localization, the dense semantic features are recovered to original resolution by bilinear up-sampling and continuous convolution. In addition, we incorporate the effective skip connection, which provides low-level spatial details as guidance of global context to weight category localization details without causing too much computation burden. Finally, we implement a DCB to refine discriminative features and generate precise cloud-segmentation results.

Although involving similar encoder-decoder architecture, our method is different from the CloudU-Net [7] by replacing the traditional fixed convolutions with DCBs. It augments the flexibility of spatial sampling locations of the geometric variation of clouds. He *et al.* proposed the DABNet [10], which applies a deformable context feature pyramid module in the highest layer of the network, whereas we devise the DCB as an underlying structure of the encoder to mine the inherent image representation. An exhaustive ablation study in Section III-B experimentally supports such analysis and further explores the mechanism of DCB in our proposed framework.

B. Deformable Convolution Block

In light of the specific morphological characteristics of clouds, we integrate deformable paradigm to automatically exploit discriminative contextual information and improve

Authorized licensed use limited to: Southeast University. Downloaded on November 28,2022 at 07:12:11 UTC from IEEE Xplore. Restrictions apply.

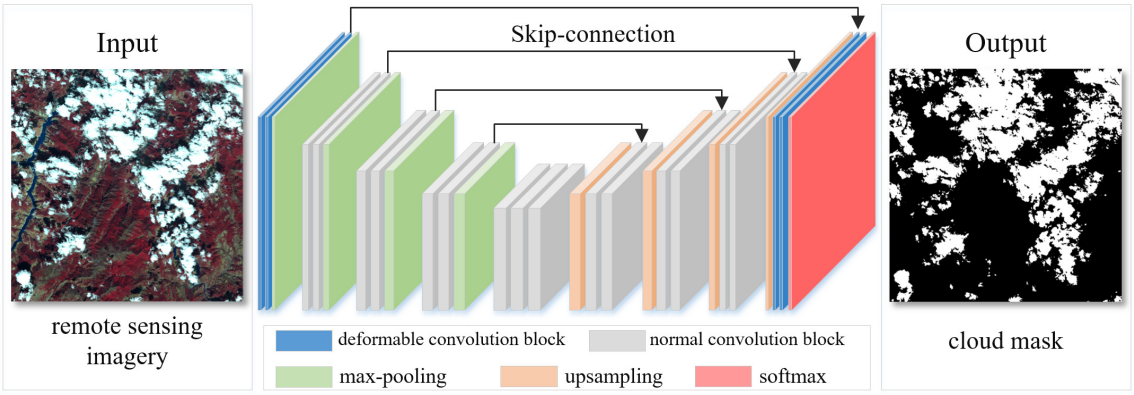

Fig. 2. Framework of the proposed DCNet. It has a U-shaped architecture with an encoder and a decoder symmetrically on two sides. DCBs are introduced at both the beginning and the end of network.

the capability of modeling geometric transformation. DCB contains a deformable convolution layer, a batch normalization layer, and a rectified linear unit (ReLU) activation layer.

In deformable convolution layers, an additional 2-D offset is added to the regular grid-sampling locations in a standard convolution. For example, given a 3×3 kernel with dilation 1, the receptive field size and dilation of a standard convolutional grid $\mathcal{R}$ can be formalized as

$$\mathcal{R} = \begin{Bmatrix} (-1,-1) & (-1,0) & (-1,1) \\ (0,-1) & (0,0) & (0,1) \\ (1,-1) & (1,0) & (1,1) \end{Bmatrix}. \quad (1)$$

Thus, for each location p_0 on output feature map $\boldsymbol{y}$, we have

$$\boldsymbol{y}(p_0) = \sum_{p_n \in \mathcal{R}} \boldsymbol{w}(p_n) \cdot \boldsymbol{x}(p_0 + p_n) \quad (2)$$

where $\boldsymbol{x}$ means the input feature map, $\boldsymbol{w}$ represents the weights of sampled value, and p_n enumerates the locations in $\mathcal{R}$. Although in deformable convolution, the regular grid $\mathcal{R}$ is augmented with offsets $\{\triangle p_n | n = 1, \ldots, N\}$, where $N = |\mathcal{R}|$. So, the modulated deformable convolution can be expressed as

$$\boldsymbol{y}(p_0) = \sum_{p_n \in \mathcal{R}} \boldsymbol{w}(p_n) \cdot \boldsymbol{x}(p_0 + p_n + \triangle p_n). \quad (3)$$

Now, the free deformation is realized by the irregular offset locations $p_n + \triangle p_n$. As illustrated in Fig. 3, the offsets are learned from the preceding feature maps, using additional convolutional layers in parallel. It has a $2N$ channel dimension that corresponds to N 2-D offsets. Because the offset $\triangle p_n$ is usually a decimal, bilinear interpolation is introduced to revise the value of the sampled points after migration.

C. Classification Layer and Loss

We formulate cloud detection as a pixel-level binary semantic segmentation task, which uses pixel-level labels to indicate whether each pixel contains a cloud or not. In the output layer, let x_i represents an unnormalized category vector of the location i, and $P(y_i = j|x_i)$ denotes the probability estimation that y_i is the ground truth of x_i. It is realized by the softmax function

$$P(y_i = j|x_i) = \frac{\exp(x_{ij})}{\sum_{c \in C} \exp(x_{ic})} \quad (4)$$

where C is the set of categories.

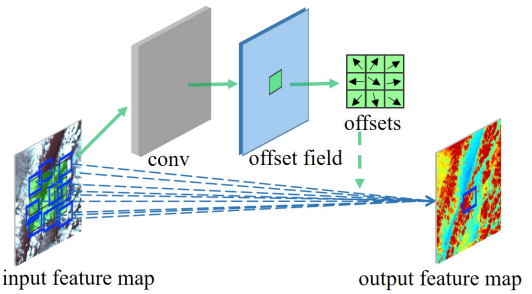

Fig. 3. Illustration of a 3×3 deformable convolution. Offset field comes from the input feature map and has the same spatial resolution as the input.

The cross-entropy loss function $J(\theta)$ is formulated as

$$J(\theta) = -\frac{1}{M} \sum_{i \in M} \sum_{j \in C} [\mathbf{1}\{y_i = j\} \log(P(y_i|x_i))] + \lambda \|\Theta\|^2 \quad (5)$$

where M is the set of pixels in input images, Θ contains parameters of the network, $\mathbf{1}\{\cdot\}$ is the indicative function. We use $L2$-norm regularization term to avoid overfitting.

III. Experiments

In this section, we evaluate the proposed DCNet on the GF-1 WFV satellite imagery. Specifically, we first present data preparation and experimental settings. Then, we discuss the performance and variants of the proposed module. Additionally, we compare DCNet with some other recently published approaches.

A. Dataset and Experimental Settings

1) GF-1 WFV Satellite Imagery: GF-1 WFV Cloud and Cloud Shadow Cover Validation Data [15] include 108 WFV level-2A scenes and their reference masks. The approximate size of the images is $15\,700 \times 16\,200 \times 4$. In this experiment, we use 40 (train), 40(val), and 28 (test) scenes for training, validation, and testing, respectively. We evaluate the distribution of clouds in the dataset, and different kinds of cloud covers are included. Then, all original images are divided into subimages of size $320 \times 320 \times 4$ at the step of 60. Finally, we randomly select 50 000 subimages as the training data.

Authorized licensed use limited to: Southeast University. Downloaded on November 28,2022 at 07:12:11 UTC from IEEE Xplore. Restrictions apply.

TABLE I
ABLATION STUDIES ON DCBS

w/o DCB	Depths			Accuracy	F1-score	Params
	L1	L2	L3			
no				94.03	82.19	6.74M
+L1	√			**96.94**	**90.12**	6.79M
+L2		√		96.77	89.35	6.92M
+L3			√	96.41	88.51	7.36M
+L12	√	√		96.21	86.91	6.97M
+L123	√	√	√	95.98	85.80	7.59M

2) Evaluation Metrics: We use the accuracy, $F1$-score, and mean intersection-over-union (mIoU) as the quantitative metrics to evaluate the performance of DCNet. Additionally, we adopt the kappa coefficient (Kappa) as a robust parameter that measures the interrater agreement for qualitative (categorical). The value of Kappa is usually less than or equal to 1, and the larger means the better.

These quantitative metrics are defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{6}$$

$$\text{mIoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \tag{7}$$

$$F1\text{-score} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FN} + \text{FP}} \tag{8}$$

$$\text{Kappa} = \frac{p_0 - p_e}{1 - p_e} \tag{9}$$

where TP represents the number of the true-positive pixels; TN represents for the number of the true-negative pixels; FP represents the number of the false-positive pixels; and FN represents the number of the false-negative pixels. Besides, $p_0 = (\text{TP} + \text{FP})/(P + N)$ and $p_e = (P(\text{TP} + \text{FP}) + N(\text{TN} + \text{FN}))/(P + N)^2$, where P and N denote the number of cloud and noncloud pixels in the ground truth, respectively.

3) Implementation Details: All models are trained with PyTorch framework and optimized by the stochastic gradient descent (SGD) algorithm. The operating system is Ubuntu 16.04 equipped with NVIDIA Tesla P100 16G GPUs. We train on two GPUs in parallel for accelerating the optimization process, with a learning rate of 1×10^{-4}, which is update by the "poly" strategy. The minibatch sizes, momentum, and total iteration are 4, 0.9, and 1×10^6, respectively. Comparison methods are trained with the same experimental settings as DCNet. Only one GPU is used in the test phase.

B. Ablation Studies on DCBs

How to bring out the superior competence of deformable convolutions? We conduct a detailed investigation with different combinations of DCBs to analyze the performance of DCNet. First, we verify the contribution of DCB at different depths to our cloud-detection task. We stand on standard convolution without DCB as the self-contrast. On this basis, DCB is incorporated into the different depths of the network incrementally. As summarized in Table I, depths (L1–L3) indicate whether the DCB is adopted symmetrically at the first, second, and third levels in the DCNet. It is obvious that when DCB is introduced at the shallowest level symmetrically (DCB+L1), DCNet achieves the best accuracy, which outperforms the self-contrast (no DCB) and DCB at the deepest level (DCB+L3) by 7.93% and 1.61% in terms of $F1$-score, respectively. Furthermore, we found that the performance of DCNet is inversely proportional to the depth of DCB. When DCB are incorporated into the depths of L1, L2, and L3 simultaneously (DCB+L123), its detection result is much worse than DCB+L1 and consumes more computing resources. It is because the DCB prediction is jointly affected by learned offsets and network weights during training. But not all pixels within the receptive field have an equal contribution to the network response. In high-level feature maps, the equal size kernel corresponds to a larger receptive field, so the sampling locations may be interfered by redundant features, including background areas irrelevant for detection. What is more, we evaluate computational complexity of DCNet with the number of trainable parameters (Params). It can be seen that the configuration of DCB+L1 only adds bits of parameters, whereas the accuracy and $F1$ score have been greatly improved. This indicates that the significant performance improvement is from the capability of modeling geometric transformations, other than increasing model parameters.

From these studies, we make the observation that using deformable convolution at the shallow level can enhance the deformation modeling capability of DCNet and facilitate the model to conform more closely to cloud structure.

TABLE II
COMPARISON OF DCNET WITH OTHER ARCHITECTURES

Method	Accuracy	F1-score	mIoU	Kappa
FCN	96.12	88.08	87.08	85.76
U-Net	93.93	82.22	81.38	78.60
SegNet	96.06	88.07	87.04	85.72
DeepLabv3+	94.18	79.85	79.94	76.47
CloudSegNet	95.19	85.29	84.38	82.42
CDNet	96.44	89.01	88.02	86.89
CloudU-Net	96.72	90.06	89.09	88.17
DABNet	96.07	87.75	86.53	85.59
DCNet	**96.94**	**90.12**	**89.19**	**88.26**

C. Cloud-Detection Results on GF-1 WFV Satellite Imagery

We extend the evaluation of the proposed model with several state-of-the-art approaches on the GF-1 Satellite imageries. Table II and Fig. 4 demonstrate quantitative and qualitative results, respectively. It can be seen that our proposed DCNet outperforms the state-of-the-art methods, yielding 96.94% accuracy and 90.12% $F1$-score. As depicted in Fig. 4, these general semantic segmentation models [16]–[19] only focus on regional accuracy but pay less attention to the boundary quality, leading to the blurred and smooth boundary in the detection results. As for recently published cloud-detection models [6], [7], [9], [10], they pay more attention to obtain more abundant features by augmenting the receptive field and combining the multiscale contextual information. Even their results are much better than general methods, but it is still tough to distinguish cloud-like targets resulting in a higher false alarm rate. This can be seen from Fig. 4(c) and (d).

In summary, the offset learned in the deformable convolutional layer is very suitable to capture the specific morphological characteristics of clouds. Benefiting from this, DCNet

Authorized licensed use limited to: Southeast University. Downloaded on November 28,2022 at 07:12:11 UTC from IEEE Xplore. Restrictions apply.

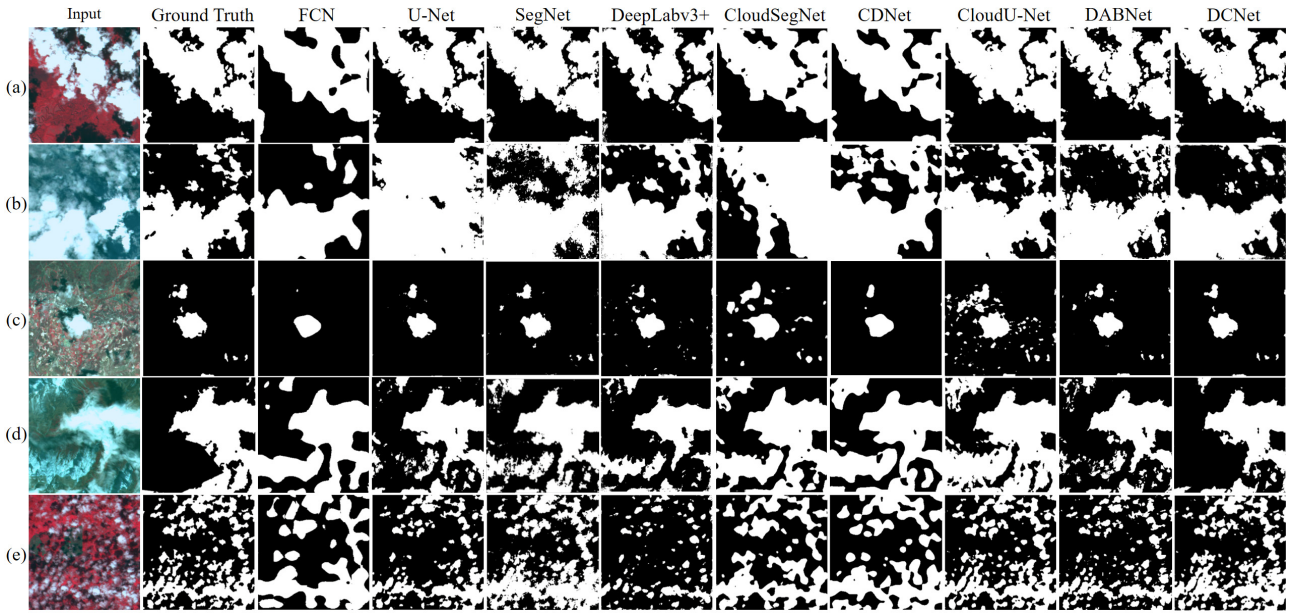

Fig. 4. Qualitative comparisons of cloud-detection results in the partial scene of GF-1 WFV imageries. (a)–(e) are cropped subimages. Among of them, (a), (b), and (e) are cloud-only cases, (c) is a cloud image with bright buildings, and (d) is a cloud image with snow. Black represents the pixels classified as clear and white represents the pixels classified as cloud.

is enforced to adaptively generate clearer and more explicit cloud-detection results.

IV. Conclusion

In this letter, we propose a DCNet for cloud mask segmentation. By introducing DCBs, the capability of a network to describe geometric transformation is considerably enhanced. Additionally, we implement skip connections between the encoder and decoder. It exploits low-level spatial details as guidance of global context to weight category-localization details and recover high-level semantic pixel localization. Experimental results show that DCNet achieves precise cloud-detection results and outperforms state-of-the-art methods.

Despite the satisfactory results, DCNet needs massive pixel-level annotation labels, which require a great deal of manual annotation labor. In future work, we will conduct an indepth study on this issue.

References

[1] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, Jun. 2019.

[2] Y. Zhang, "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.*, vol. 109, no. D19, pp. 1–27, 2004.

[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.

[4] C. Cucu-Dumitrescu, "A simple method of determining cloud-masks and cloud-shadow-masks from satellite imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 10–13, Jan. 2014.

[5] H. Ishida, Y. Oishi, K. Morita, K. Moriwaki, and T. Y. Nakajima, "Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions," *Remote Sens. Environ.*, vol. 205, pp. 390–407, Feb. 2018.

[6] S. Dev, A. Nautiyal, Y. H. Lee, and S. Winkler, "CloudSegNet: A deep network for Nychthemeron cloud image segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1814–1818, Dec. 2019.

[7] 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.*, early access, Jul. 23, 2020, doi: 10.1109/LGRS.2020.3009227.

[8] Y. Li, W. Chen, Y. Zhang, C. Tao, R. Xiao, and Y. Tan, "Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning," *Remote Sens. Environ.*, vol. 250, Dec. 2020, Art. no. 112045.

[9] 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.

[10] Q. He, X. Sun, Z. Yan, and K. Fu, "DABNet: Deformable contextual and boundary-weighted network for cloud detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, early access, Jan. 5, 2021, doi: 10.1109/TGRS.2020.3045474.

[11] 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.

[12] 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.

[13] Y. Guo, X. Cao, B. Liu, and M. Gao, "Cloud detection for satellite imagery using attention-based U-Net convolutional neural network," *Symmetry*, vol. 12, no. 6, p. 1056, Jun. 2020.

[14] J. Dai *et al.*, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 764–773.

[15] 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.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assisted Intervent.*, 2015, pp. 234–241.

[18] 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.

[19] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 801–818.

Authorized licensed use limited to: Southeast University. Downloaded on November 28,2022 at 07:12:11 UTC from IEEE Xplore. Restrictions apply.