Automatic Cloud Segmentation Based on Fused Fully Convolutional Networks

Jie An, Jingfeng Wu, and Jinwen Ma(✉)

Department of Information Science,
School of Mathematical Sciences and LMAM, Peking University,
Beijing 100871, China
{jie.an,pkuwjf}@pku.edu.cn, jwma@math.pku.edu.cn

Abstract. Cloud detection and segmentation of remote sensing images is a pivotal task in the area of weather forecast. Many meteorologic applications such as precipitation forecast, extreme weather forecast, etc., depend on the results of the cloud detection. In this paper, based on the satellite remote sensing image dataset, we propose an image segmentation model to address the cloud detection problem. Our model is derived from the fully convolutional neural network, which achieves pixel-level cloud segmentation results on high resolution, large scale, multi-channel satellite images. We introduce Deep Feature Aggregation and Model Fuse strategies to improve the cloud segmentation results. Compared with the traditional methods, our proposed algorithm has the advantages that is independent of the expert knowledge, totally data motivated, and more robust in hard cases. The testing results show that the proposed model can satisfy the requirements of the weather forecast, thus has a strong potential to be put into business usage.

Keywords: Remote sensing images · Semantic segmentation · FCN · Deep learning · Cloud detection

1 Introduction

There is no need to say more about the crucial role that the remote sensing data has played in people's daily life of contemporary society. As the development of the high-precision sensors, a growing amount of high-resolution remote sensing data is collected every second, raising an increasing demanding on new effective techniques to handle such vast amounts of data of extremely high dimensionality. Among which, meteorological satellite remote sensing data is the satellite images collected by meteorological satellites for weather forecasting missions. Every day, tons of this kind of satellite images are generated by the meteorological satellites in orbit. Since the detection of meteorological satellite remote sensing image is the preliminary of meteorological forecasting, it becomes an important and attractive topic to researchers (Fig. 1).

D.-S. Huang et al. (Eds.): ICIC 2019, LNCS 11643, pp. 520–528, 2019.
https://doi.org/10.1007/978-3-030-26763-6_50

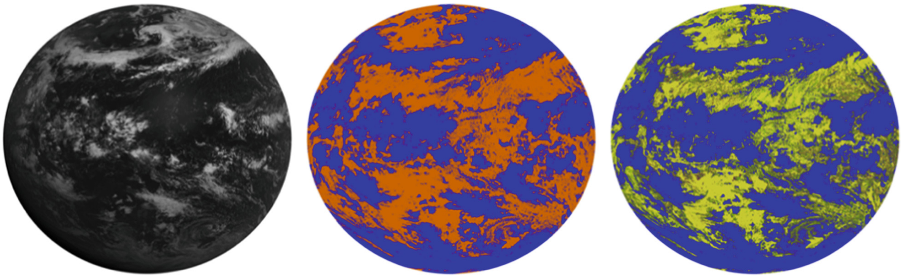

Fig. 1. Satellite images and two types of ground truth masks. Left: the image of visible channel. Middle: the two classes (cloudy/non-cloudy) ground truth. Right: the four classes (cloudy/probably cloudy/clear/probably clear) ground truth.

Concisely, the cloud detection task for meteorological satellite images is to identify the type of each pixel of the image, i.e., cloudy, probably cloudy, clear and probably clear. Or on the other hand, one can treat it as a four-classes semantic segmentation task based on remote sensing data.

Traditionally, cloud detection is treated with the thresholds-based models [1, 16]. The quality of the acquired thresholds depends on the sensor accuracy and the experts' understanding of the physical meaning of the collected signals in addition with the careful statistical analysis. Being simple and reliable, the physical models for cloud segmentation have already been employed in commercial products, e.g., the MODIS cloud product [16]. Such advantages, however, come with price. The reliance on the experts' accumulated experience causes the model too costy to obtain and almost impossible to transfer to other similar datasets. More importantly, threshold methods are unusual sensitive to the data noise. Thus, researchers are looking for effective machine learning models learned purely from the collected data, which are known for robustness.

Classical machine learning models have been well studied in the area of remote sensing data analysis, including cloud segmentation. Viewing segmentation as pixel level classification, various classification machine learning methods can be employed in segmenting remote sensing data directly, e.g., histogram thresholds [4], support vector machine [3], sparse representation classifier [13], extreme learning machine [18], autoencoder [5] and etc. Since these models classify each pixel separately, they use to suffer from high computational burden when it comes to high-resolution remote sensing images, which has large amounts of pixels. Thus, they can hardly satisfy the requirement of real-time weather forecasting.

Recently, there are also works trying to apply deep learning methods into segmenting the remote sensing images, motivated by the success of Fully Convolutional Networks (FCN) [14] and its variants [2, 10, 11, 17, 23], which achieve remarkable performance on the normal RGB images. And it has also been demonstrated that deep learning approaches could indeed outperform classical machine learning models in the segmentation on various types of remote sensing images [7, 9, 20, 22]. Another line is based on super resolution [6, 8, 12, 19, 21, 24]. They proposed to address cloud detection by using CNN to classify the super pixels. Though the super resolution approach helps, it still cannot achieve good pixel-level cloud detection performance [15]. suggests to use FCN [14] for cloud detection. However, they have not carefully designed the network structure for cloud detection task. Also, they model is limited into small input data and it cannot solve the challenge of the exceeding of GPU memory when taking the extremely high-resolution remote sensing data as input.

As far as we know, there is no literature studying the cloud detection on extremely high resolution (5500×5500) and multiple channels (16 channels) meteorological satellite remote sensing images.

To bridge the gap, we propose a novel FCN-based model for the cloud detection task on such extremely high resolution and multiple channels images. We use Deep Feature Aggregation (FA) to improve cloud segmentation results by adopting multi-scale features. Moreover, we fused the segmentation results by two structurally-distinct models based on FCN and hereby achieve state-of-the-art performance. Our proposed model is able to obtain pixel-level cloud detection results, i.e., to assign each pixel belonging the image a label of cloud, probably cloud, clear or probably clear. Experiments on practical datasets show that our method outperforms state-of-the-art FCN networks in terms of accuracy, precision, recall, F1 and IoU, while it is also efficient enough to support real-time weather forecasting.

2 Methodology

2.1 Datasets and Cloud Detection

The concerned cloud remote sensing images are collected by Himawari-8 satellite[1], which collects data for the Earth every ten minutes. The original data has 16 spectrum channels (6 visible channels and 10 infrared light channels), and 5500×5500 pixels per image. The data is open for downloading. We choose one-month data as the training set, and four-day data from another month as the test set.

To obtain the annotation of the images, we use the expert-corrected label from commercial MODIS [16] cloud product. In our annotation, every pixel of data is labeled as either *cloudy, probably cloudy, clear or probably clear*. Thus, the issue of cloud detection could be viewed as a four categories semantic segmentation problem. And our model is expected to output a label for each pixel. The major challenge of cloud detection is, due to the extremely high resolution and multiple channels of the input data and limited by the hardware resources, one cannot directly apply semantic segmentation strategies to address the problem.

2.2 The Segmentation Model for Remote Sensing Images

Overview of Our Framework. We design a novel FCN based framework to handle the segmentation tasks on extremely high resolution and multiple channels remote sensing images. As shown in Fig. 2, our framework consists of three parts:

1 http://www.data.jma.go.jp/mscweb/en/himawari89/cloud_service/cloud_service.html.

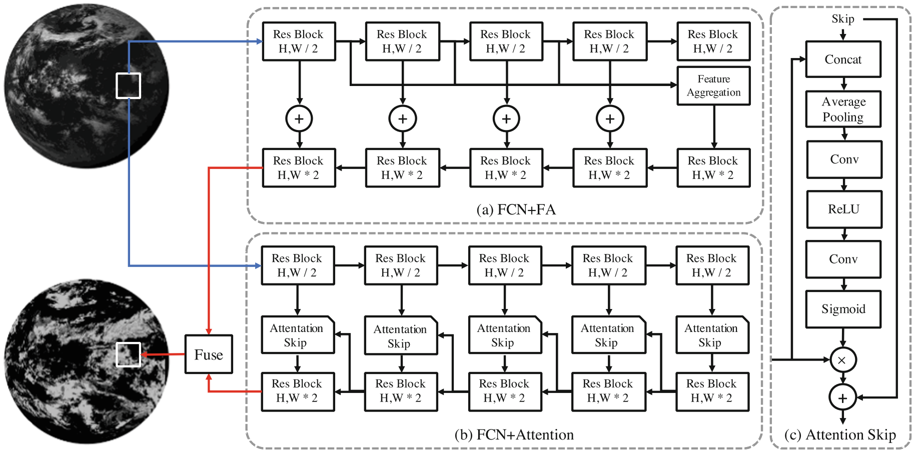

Fig. 2. Our proposed framework and the network structure.

1. Splitting the input images. we cut the original image (5500 × 5500 × 16) into small patches (256 × 256 × 16);
2. Segmentation of patches. We use an FCN-like network to derive the segmentation results of small patches;
3. Fusion of patch predictions. The final output of our model is the fusion of the predictions of the small patches of one original input.

We now turn to elaborate each part.

Cropping the Input Images. Generally, the meteorological satellite remote sensing images is of huge size, e.g., 5500 × 5500 × 16. Thus, due to the limit of GPU memory, usually less than 128 Gb, it is not possible to directly train or evaluate semantic segmentation networks on the whole original images. Therefore, it is necessary to split the original image into small patches and use the small patches as the input of the model.

In practice, to augment the training set and better training the model, we randomly sample 256 × 256 × 16 patches from the original image as our patched training set. There are two remarks. First, in order to eliminate the noise of the patched training set, we delete patches which do not intersect with the Earth. Second, for helping the fusion process, we keep an overlap of boundary pixels for two contiguous patches, as illustrated in Fig. 2.

Segmentation of Patches. The structure of our segmentation networks is shown in Fig. 2. The network is based on FCN [14] and U-Net [9]. We use two networks with Feature Aggregation and Attention Skip strategies separately and fuse the segmentation results of them. The encoder part of each network uses residue network to extract image features, while the decoder part uses symmetrical trans-convolution layers to enlarge features. We add shortcut connection between the encoder and the decoder to make the model as a “U” shape, which is known to be beneficial to semantic segmentation.

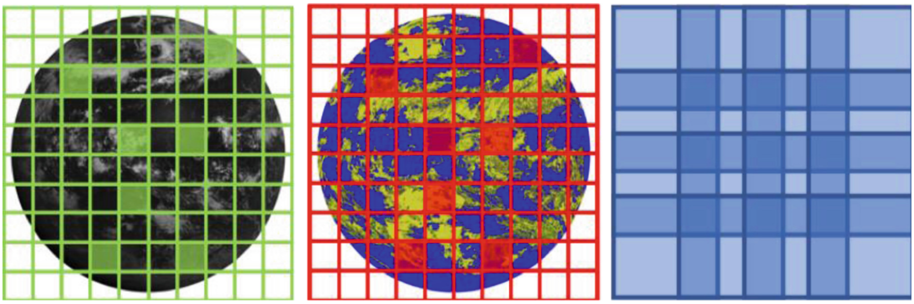

Fig. 3. Image crop and result mosaic to get the whole prediction. We crop patch image as shown in the left image, and mosaic the prediction patch results as at corresponding locations as shown in the middle figure. We mosaic the final patches with overlap to avoid local mismatch.

Stitch of Patch Predictions. After we have the prediction of each patches, we need to fuse them into the original size to make the model valid for evaluation and usage. The fusion process is by combining the patched prediction into whole, and vote the results of the shared boundary pixels of the patches. Figure 3 illustrates this process.

3 Experimental Results

3.1 Settings

In this section we describe the detailed model training and evaluation procedures. As stated before, during training, we use randomly sampled $256 \times 256 \times 16$ patches as the input of our model. We only use the patches intersected with the Earth. During evaluation, we evenly split one $5500 \times 5500 \times 16$ test image into $256 \times 256 \times 16$ small patches with shared boundary pixels. The total number of full training images is $100,000$, and the total number of full test images is 277.

We choose mean square loss as our loss function

$$L := \| \mathit{Prediction} - \mathit{Groundtruth} \|_2^2 .$$

We use Adam to optimize the loss function, with the learning rate adjust strategy as

$$\begin{aligned} &LR_0 = 10^{-4}, \\ &LR_{i+1} \leftarrow LR_i \times 0.95, \quad i \text{ indicates epoch status.} \end{aligned}$$

For all experiments, we save model parameters every epoch, and report the best of them in test set as the final results.

3.2 Results Comparison

We measure the performance of our results in multiple indicators, including IoU, accuracy, precision, recall and F1. The IoU of the class i is computed as

$$IoU_i := \frac{Prediction_i \cap Groundtruth_i}{Prediction_i \cup Groundtruth_i}.$$

And mean IoU is the mean of IoU of each class.

To fully compare our model with others, we report the above indicators in two cases: the cloudy/non-cloudy two classes segmentation and the cloudy/probably cloudy/clear/probably clear four classes segmentation.

As illustrated in Table 1, our proposed model outperforms the result by the Vanilla FCN network structure in literature and FCN [14] coupled with feature aggregation operator only. Meanwhile, our algorithm could simulate the segmentation results of commercial software like MODIS [16] cloud product. What is more, driven by data, our approach is more robust in extreme cases, compared with MODIS [16] cloud product. Last but not least, since our method does not rely on expert knowledge, it could be easily transferred into other similar detection tasks on remote sensing data.

Table 1. Testing results of the proposed algorithm. **FA** denotes Feature Aggregation, **Att** represents Attention Skip.

Metrics	IoU	Accuracy	Precision	Recall	F1 value
FCN Class#2	0.8826	0.9376	0.9382	0.9358	0.9368
FCN+FA Class#2	0.8857	0.9394	0.9389	0.9386	0.9387
FCN+FA+Att Class#2	**0.8879**	**0.9406**	**0.9407**	**0.9393**	**0.9400**
FCN Class#4	0.7862	0.9402	0.7964	0.7665	0.7787
FCN+FA Class#4	0.7873	0.9405	0.8002	0.7633	0.7787
FCN+FA+Att Class#4	**0.7938**	**0.9425**	**0.8053**	**0.7741**	**0.7871**

3.3 Correcting the Wrong Annotation

Surprisingly, we find that our algorithm could even correct the wrong annotation of MODIS [16] cloud product in some extreme cases. Figure 4 shows some scenarios where our predictions are better similar to expert's prediction, than the results of MODIS [16] cloud product. This indicates the commercial potential of our model (Fig. 5).

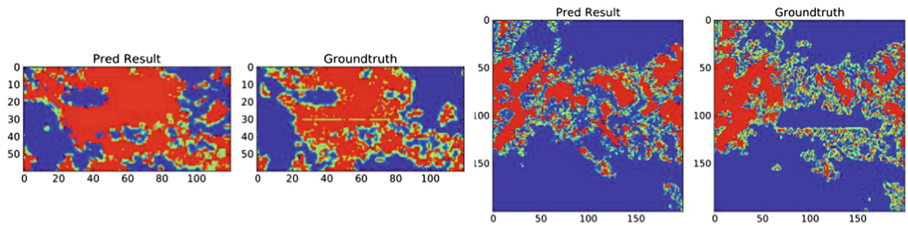

Fig. 4. Our model can improve the results of the MODIS cloud production. As left two image shows, the ground truth image has an artifact that a straight line in the middle of the image. Such an issue is also shown in the right figure.

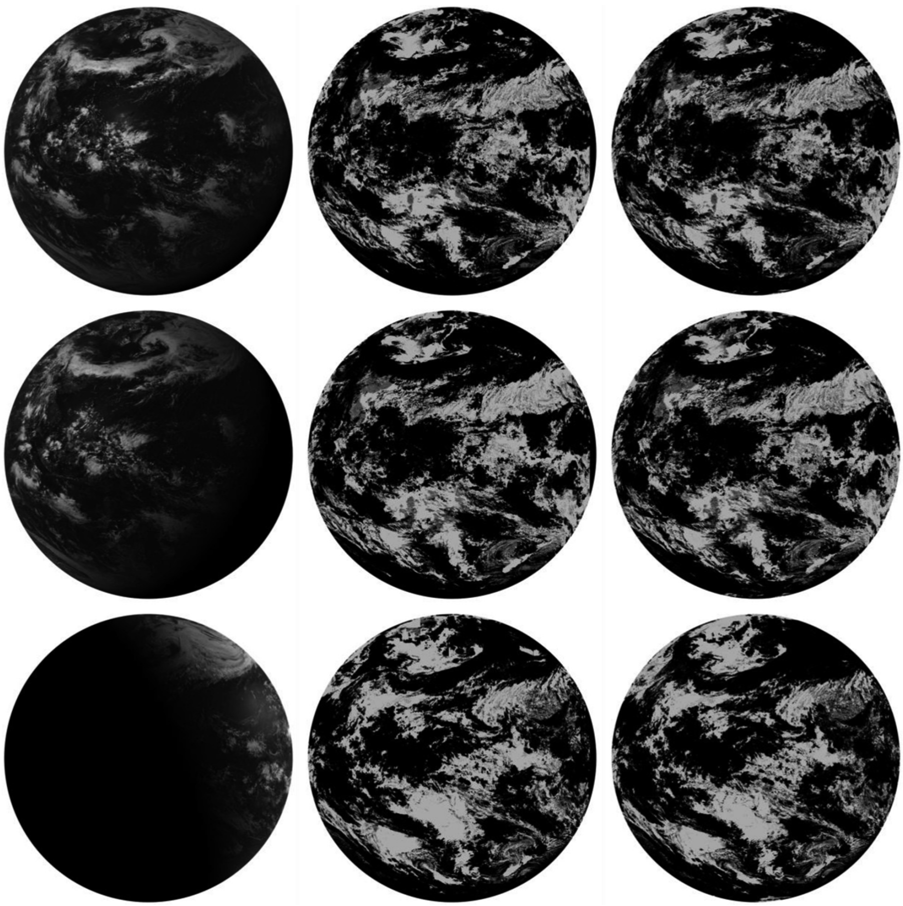

Fig. 5. Four class prediction results. The left images are the input image. The middle images are four class prediction results. The right images are four class ground truth.

4 Conclusion

The cloud detection task based on meteorological satellite remote sensing data is the key for many follow-up applications. In this work, based on FCN [14], we come up with a novel framework for the cloud detection of extremely large remote sensing images. Specifically, the original image is first split into small patches. Then, the segmentation network outputs the prediction of the small patches. Last, we mosaic the prediction of the small patches along their shared boundaries, hence obtaining the segmentation of the extremely large remote sensing images. As a result, our approach can achieve the state-of-the-art performance in cloudy/non-cloudy two classes segmentation task and cloudy/probably cloudy/clear/probably clear four classes segmentation task. Thanks to the data driven properties, our method does not rely on expert knowledge, and could be easily to extend into other similar problems. Experiments indicates our model has strong commercial potential in business usage.

Acknowledgement. This work is supported by the *Natural Science Foundation of China for Grant U1604153* and *High-performance Computing Platform of Peking University*.

References

1. Chepfer, H., et al.: The GCM-oriented calipso cloud product (CALIPSO-GOCCP). J. Geophy. Res.: Atmos. **115**(D4) (2010)
2. Dai, J., He, K., Sun, J.: BoxSup: exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1635–1643 (2015)
3. Dumitru, C.O., Datcu, M.: Information content of very high resolution sar images: study of feature extraction and imaging parameters. IEEE Trans. Geosci. Remote Sens. **51**(8), 4591–4610 (2013)
4. El Zaart, A., Ziou, D., Wang, S., Jiang, Q.: Segmentation of sar images. Pattern Recogn. **35**(3), 713–724 (2002)
5. Geng, J., Fan, J., Wang, H., Ma, X., Li, B., Chen, F.: High-resolution SAR image classification via deep convolutional autoencoders. IEEE Geosci. Remote Sens. Lett. **12**(11), 2351–2355 (2015)
6. Henry, C., Azimi, S.M., Merkle, N.: Road segmentation in SAR satellite images with deep fully-convolutional neural networks. arXiv preprint arXiv:1802.01445 (2018)
7. Jiao, L., Liang, M., Chen, H., Yang, S., Liu, H., Cao, X.: Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **55**(10), 5585–5599 (2017)
8. Le Goff, M., Tourneret, J.Y., Wendt, H., Ortner, M., Spigai, M.: Deep learning for cloud detection (2017)
9. Li, R., et al.: DeepUNet: a deep fully convolutional network for pixel-level sea-land segmentation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **11**(11), 3954–3962 (2018). Fused-FCN for Automatic Cloud Segmentation 9
10. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.: Semantic image segmentation with deep convolutional nets and fully connected CRFs. In: International Conference on Learning Representations (2015)

11. Lin, G., Milan, A., Shen, C., Reid, I.: RefineNet: multi-path refinement networks with identity mappings for high-resolution semantic segmentation. arXiv preprint arXiv:1611.06612 (2016)
12. Liu, H., Zeng, D., Tian, Q.: Super-pixel cloud detection using hierarchical fusion CNN. In: 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), September 2018. https://doi.org/10.1109/bigmm.2018.8499091
13. Liu, M., Wu, Y., Zhao, W., Zhang, Q., Li, M., Liao, G.: Dempster–Shafer fusion of multiple sparse representation and statistical property for SAR target configuration recognition. IEEE Geosci. Remote Sens. Lett. **11**(6), 1106–1110 (2014)
14. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
15. Mohajerani, S., Krammer, T.A., Saeedi, P.: Cloud detection algorithm for remote sensing images using fully convolutional neural networks. arXiv preprint arXiv:1810.05782 (2018)
16. Platnick, S., et al.: The modis cloud products: algorithms and examples from terra. IEEE Trans. Geosci. Remote Sens. **41**(2), 459–473 (2003)
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Samat, A., Gamba, P., Du, P., Luo, J.: Active extreme learning machines for quad-polarimetric SAR imagery classification. Int. J. Appl. Earth Obs. Geoinf. **35**, 305–319 (2015)
19. Shi, M., Xie, F., Zi, Y., Yin, J.: Cloud detection of remote sensing images by deep learning. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 701–704. IEEE (2016)
20. Wang, Y., He, C., Liu, X., Liao, M.: A hierarchical fully convolutional network integrated with sparse and low-rank subspace representations for PolSAR imagery classification. Remote Sens. **10**(2), 342 (2018)
21. Xie, F., Shi, M., Shi, Z., Yin, J., Zhao, D.: Multilevel cloud detection in remote sensing images based on deep learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **10**(8), 3631–3640 (2017)
22. Yao, W., Marmanis, D., Datcu, M.: Semantic segmentation using deep neural networks for SAR and optical image pairs. In: Proceedings of the Big Data from Space, pp. 1–4 (2017)
23. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
24. Zi, Y., Xie, F., Jiang, Z.: A cloud detection method for landsat 8 images based on PCANet. Remote Sens. **10**(6), 877 (2018)