

Earth and Space Science



RESEARCH ARTICLE

10.1029/2019EA001040

An Efficient Solution for Semantic Segmentation of Three Ground-based Cloud Datasets

Qianqian Song¹ , Zhihui Cui², and Pu Liu¹ 

¹Key Laboratory of Meteorological Disaster, Ministry of Education (KLME)/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disaster (CIC-FEMD), Nanjing University of Information Science and Technology, Nanjing, China, ²Zhengzhou Meteorological Bureau, Zhengzhou, China

Key Points:

- Twelve kinds of semantic segmentation techniques are implemented for meteorological image segmentation and performance evaluation
- A novel framework is used to generate cloud pixel-level class annotations in an end-to-end way
- Comprehensive performance evaluation on the proposed method with three public datasets exhibits robustness and generalization results

Correspondence to:

P. Liu,
pu_liu@nuist.edu.cn

Citation:

Song, Q., Cui, Z., & Liu, P. (2020). An efficient solution for semantic segmentation of three ground-based cloud datasets. *Earth and Space Science*, 7, e2019EA001040. <https://doi.org/10.1029/2019EA001040>

Received 10 DEC 2019

Accepted 20 MAR 2020

Accepted article online 4 APR 2020

Abstract The machine learning approach has shown its state-of-the-art ability to handle segmentation and detection tasks. It is increasingly employed to extract patterns and spatiotemporal features from the ever-increasing stream of Earth system data. However, there is still a significant challenge, which is the generalization capability of the model on cloud images in different types and weather conditions. After studying several popular methods, we propose a semantic segmentation neural network for cloud segmentation. It extracts features learned by source and target domains in an end-to-end behavior, which can address the problem of significant lack of labels in the observed cloud image data. It is further evaluated on the Singapore Whole Sky Image Segmentation (SWIMSEG) dataset by using Mean Intersection-over-Union, recall, F-score, and accuracy matrices. The scores of these matrices are 86%, 97%, 92%, and 96%, which prove that it has excellent efficiency and robustness. Most importantly, a new benchmark based on the SWIMSEG dataset for the task of cloud segmentation is introduced. The others, BENCHMARK, Cirrus Cumulus Stratus Nimbus are evaluated through the model trained from the SWIMSEG dataset by way of visualization.

Plain Language Summary The machine learning approaches offer a new view about how to effectively and comprehensively understand ground-based cloud datasets. The essential advantage of deep learning methods is that it can extract more critical cloud features automatically than traditional algorithms, such as spatiotemporal features. Therefore, it is worth exploring the possibility of cloud segmentation with the help of deep learning techniques. We first introduce a semantic segmentation neural network for cloud segmentation problems after measuring a few classic neural networks. The results exceed the traditional methods by a large margin with standard evaluation matrices, such as Mean Intersection-over-Union, recall, F-score, and accuracy. The scores achieved here may accomplish as a baseline for competitive development. Second, the trained model is used to produce a few cloud masks in two public datasets: BENCHMARK, Singapore Whole Sky Image Segmentation, and their respective performance is further evaluated. Finally, the segmentation results show the excellent performance and generalization in another untrained dataset, Cirrus Cumulus Stratus Nimbus.

1. Introduction

The Earth's average temperature is mainly regulated by clouds, which absorb and reflect shortwave radiation as well as absorb and emit longwave radiation (Huang & Su, 2008; Jiang et al., 2018; Su et al., 2017). The heat generated by clouds influences the atmospheric circulation and water content, which in turn affect cloud shapes. Clouds show in a variety of forms and always stay in a continuous process of evolution, and there are several principals of discriminations in which clouds are determined by their shapes or interior distinctive features. Clouds are thought as a kind of natural texture due to their texture similarity, brightness similarity, and contour continuity in an image. Therefore, it is possible to identify 10 main cloud groups known as cloud genera by means of cloud segmentation based on recognition of certain features. The traditional cloud segmentation methods are mainly based on the threshold segmentation technology, which combines the fixed/dynamic/mixed threshold to distinguish different categories (Heinle et al., 2010; Li et al., 2011; Long et al., 2006; Souza-Echer et al., 2006). However, these methods rarely use the spatial information and rely heavily on the weather condition. Therefore, considerable mis-segmentation between the source domain and the target domains still can be found. Besides, employing the color and texture features can also acquire

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

some promising detection results (Mantelli Neto et al., 2010; Richards & Sullivan, 1992), but there are issues that need to be resolved such as smooth regions of cloud and cloud at borders with sky.

Recent developments in deep convolutional neural networks make relentless progress on semantic segmentation (Badrinarayanan et al., 2017; Chen et al., 2014; Long et al., 2015; Noh et al., 2015; Wu et al., 2019; Reichstein et al., 2019; Zhao et al., 2017). Semantic segmentation is regarded as a fundamental task in computer vision, which can help scene understanding completely. It can predict dense labels of all pixels in the image to generate the corresponding mask. Cloud segmentation can be treated as an application of image segmentation, and therefore, applying semantic segmentation techniques for cloud detection is a reasonable consideration. Moreover, existing approaches based on deep learning for cloud segmentation have largely concentrated on satellite data (Drönner et al., 2018; Lu et al., 2019). Hence, it deserves to exploring the cloud segmentation performance by means of deep learning on the ground-based cloud dataset. Recently, a few methods based on deep learning have achieved better results than traditional algorithms on the ground-based cloud dataset. Dev et al. (2019) propose a model, which can effectively solve the problem of thin clouds with wrong labels. A SegCloud model trained on the 400 whole sky images with manual-marked labels is proposed for the cloud image segmentation (Xie et al., 2019). Neither comprehensively compares the effects of other models on their datasets, leaving the generalization of the models on other datasets unclear. Thus, it is a pressing need for a method based on deep learning in terms of effective, robustness, and speedup runtime. Furthermore, a labeling algorithm for reducing errors induced by human subjective judgment is also indispensable. Comparing the performances of various previous benchmark models will contribute to selecting neural network frameworks for other tasks.

In the following sections, we first describe three publicly available datasets. Second, the experimental designs and parameter settings are further presented. Third, the proposed method is thoroughly compared with the traditional methods. Finally, we summarize the paper and draw conclusions about this work.

2. Three Ground-Based Cloud Datasets

Typically, the annotated dataset is one of the essential prerequisites for a rigorous evaluation of any traditional or any modern segmentation algorithm. It is well known that a well-organized dataset that includes corresponding ground truth contributes to improving accuracy and avoiding overfitting concerning machine learning techniques. The existing three public cloud datasets are used for cloud segmentation.

2.1. The BENCHMARK Dataset

The BENCHMARK dataset is a subset of UTILITY dataset collected by Li et al. (2011) from three sources: whole sky camera, photographers, and Internet. The UTILITY dataset includes 1,000 cloud images, which are cropped of various dimensions to avoid areas near the sun and the horizon. After the visual screening, 32 images are selected from UTILITY including different cloud forms, such as the cumuliform, cirriform, and stratiform. And then, the Voronoi polygonal region generator is further used to generate the binary mask image as the ground truth with average pixels of 682×512 . A few sample images in BENCHMARK are displayed in Figure 1a. Given the limited samples of this dataset, using deep learning techniques may lead to overfitting.

2.2. The Singapore Whole Sky Image Segmentation Dataset

Another publicly available cloud dataset annotated with masks is the Singapore Whole Sky Image Segmentation (SWIMSEG) dataset created by Dev et al. (2017). Figure 1b illustrates a few representative cloud samples. It consists of 1,013 images with the dimension of 600×600 , which is obtained by the sky imaging system that is developed and deployed at Nanyang Technological University in Singapore. Each cloud patch is manually chosen from the 2-year observation images series. Cloud masks for the dataset are generated in discussion with cloud experts from the Singapore Meteorological Services. It has more samples than BENCHMARK dataset and is more suitable for machine learning training.

2.3. The Cirrus Cumulus Stratus Nimbus Dataset

Zhang et al. (2018) introduces Cirrus Cumulus Stratus Nimbus (CCSN) dataset, which is grouped into 11 categories under the meteorological criteria. It includes 2,543 cloud images with 256×256 pixels in the JPEG format and contains the 10 most concerned cloud genera in the cloud observation and contrails. The



Figure 1. Some samples from two datasets: (a) the BENCHMARK dataset, (b) the SWIMSEG dataset.

contrails may cause the warming effect and therefore arouse the concern of many scientists. However, compared to the two datasets mentioned before, this dataset currently does not have the cloud masks. Therefore, we try to use the model trained from the SWIMSEG dataset to generate the corresponding cloud masks.

3. Experimental Setup and Implementation Details

The following 12 kinds of start-of-the-art and effective neural networks have been introduced for semantic segmentation: SegNet (Badrinarayanan et al., 2017), MobileNets (Howard et al., 2017), PSPNet (Zhao et al., 2017), FCDensnet (Jégou et al., 2017), DeepLabV3 (Chen et al., 2017), RefineNet (Lin et al., 2017), FRRN (Pohlen et al., 2017), AdapNet (Valada et al., 2017), DeepLabV3+ (Chen et al., 2018), DenseApp (Yang et al., 2018), Global Convolutional Network (GCN) (Peng et al., 2017), and BiSeNet (Yu et al., 2018).

These neural network frameworks are first deployed into cloud segmentation, and their performances are evaluated on the SWIMSEG dataset. In addition, there are many evaluation metrics that have been introduced and frequently adopted to assess the effectiveness of semantic segmentation techniques. In the performance evaluation, we not only adopt the precision, recall, F-score, and accuracy metric but also calculate the Mean Intersection-over-Union (MIoU) (Garcia-Garcia et al., 2017), which is a widely used metric for the task of semantic segmentation and quantifies a ratio of overlap between the intersection and the union of two sets. The percentage can also be interpreted as the number of true positives over the sum of true positives, false negatives, and false positives.

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}, \quad (1)$$

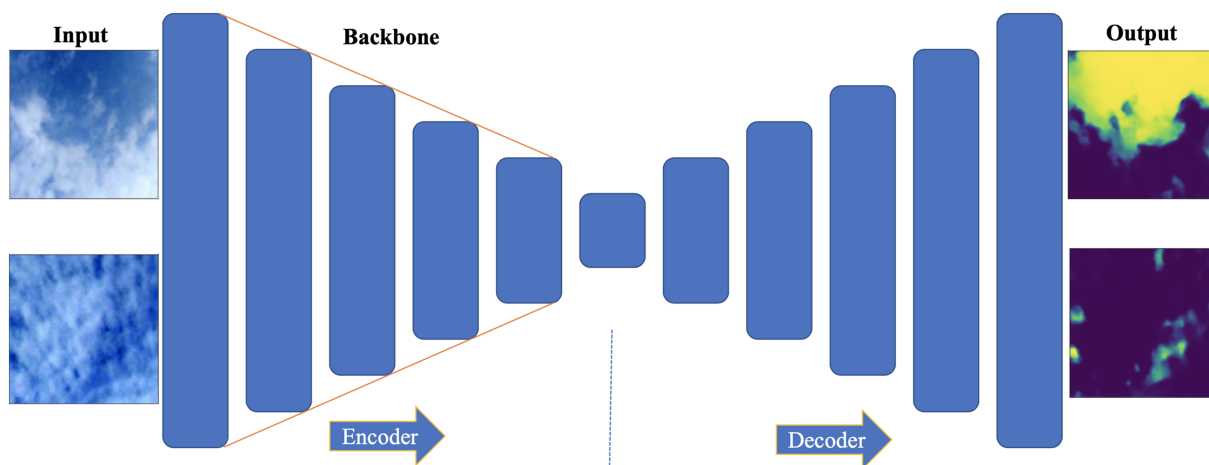


Figure 2. The overall illustration of model design is composed by encoder and decoder. The input images are observation cloud images, and the output images are corresponding cloud masks using yellow and dark purple to represent sky and cloud, respectively.

where $k+1$ denotes the number of categories in SWIMSEG dataset, p_{ij} is the number of false positives, which mean the number of pixels of class i inferred to belong to class j . Similarly, p_{ii} represents the number of true positives, p_{ji} is the number of false negatives.

The 12 neural networks are implemented by training on the SWIMSEG dataset with Tensorflow (Abadi, 2015) (version 1.12) running on a single NVIDIA GeForce GTX1080Ti. The SWIMSEG dataset, which includes 1,013 cloud images, is divided into three parts according to the common practice. Consequently, 911 cloud images are used as the training set. The remaining cloud samples are used for the testing and validation set, with 51 and 51 cloud images, respectively. In the training process, the batch size is 1, and the image is resized into 512×512 pixels, according to the number of input layer nodes setting. Figure 2 illustrates the overall structure, which consists of encoder and decoder. The former is similar to the convolutional neural network without the final fully connected layer. Here, we apply the ResNet101 (He et al., 2015), a state-of-the-art neural network, pretrained on the ImageNet dataset (Deng et al., 2009) as backbone. The latter uses several special networks to generate a complete cloud images. The hyperparameters of the network in each layer keep its original setting. The entire network parameters are fine-tuned on the SWIMSEG dataset by RMSProp optimizer with the learning rate of 0.0001 and decay of 0.995. The same settings are used to train 300 epochs in the 12 kinds of encoder-decoder architectures. To check the effect of data augmentation, we do not employ the data augmentation techniques in the previous experiments.

A classic Unet (Ronneberger et al., 2015) is introduced with SE-ResNet-152 (Hu et al., 2017) as backbone pretrained on the ImageNet to train the SWIMSEG dataset. Here the batch size is 4 and the learning rate is 0.0001 with Adam (Kingma & Ba, 2014) optimizer. The other settings are the same as the previous experiments, such as partitioning of the SWIMSEG dataset. The SWIMSEG dataset is trained 50 epochs to get the final model. Since the dataset is very small, a large number of augmentation methods are adopted to increase the quantity of cloud image, such as horizontal flip, affine transforms, perspective transforms, brightness/contrast/colors manipulations, image blurring and sharpening, gaussian noise, and random crops. These transformations can be easily implemented using the albumentation package (Buslaev et al., 2018). Some neural networks with a slight modification are used to compare their final results. For example, the MobileNets-skip indicates that it uses the skip connection in the network layer connection. The only difference between FC-Densnet-56 and FC-Densnet-67 is neural network depth, with 56 and 67 layers, respectively.

4. Experimental Result and Discussion

Compared with the results of existing methods, this section proves the effectiveness of our proposed method.

4.1. Evaluation of Several Semantic Segmentation Methods

The results of different methods are summarized in Table 1. The IoU is 86% in our proposed method, which acquires the best result of all methods. Besides the IoU criterion, the results of recall, F-score, and accuracy

Table 1
Evaluation Matrix Results of Cloud Segmentation Based on SWIMSEG Dataset

Methods	Precision	Recall	F-score	Accuracy	IoU
SegNet	0.91	0.90	0.89	0.90	0.79
MobileNets	0.90	0.90	0.89	0.90	0.79
MobileNets-skip	0.85	0.85	0.85	0.83	0.71
PSPNet	0.85	0.83	0.82	0.83	0.71
FC-Densnet-56	0.87	0.86	0.85	0.86	0.74
FC-Densnet-67	0.84	0.82	0.80	0.82	0.68
DeepLabV3	0.85	0.86	0.85	0.86	0.73
DeepLabV3+	0.90	0.90	0.89	0.90	0.77
RefineNet	0.89	0.89	0.88	0.89	0.76
FRRN-A	0.92	0.89	0.89	0.89	0.78
FRRN-B	0.87	0.87	0.86	0.87	0.75
AdapNet	0.85	0.84	0.82	0.84	0.71
DenseApp	0.87	0.87	0.85	0.87	0.74
GCN	0.91	0.91	0.91	0.91	0.82
BiSeNet	0.86	0.88	0.87	0.88	0.76
Proposed method	0.95	0.97	0.92	0.96	0.86

are also higher than that of other methods. Moreover, MobileNets-skip has achieved lower results comparing with MobileNets. The performance of the DeepLabV3+ is better than its original network design. Two versions of the full-resolution residual network (FRRN) are implemented. We discover that more complicated networks cannot always yield a more satisfying result. For example, the FRRN-B network structure is more in-depth than FRRN-A, but the results of the FRRN-A outperform FRRN-B. The performance of FC-Densnet-67 is also worse than the shallow FC-Densnet-56 network structure. The reason may be overfitting caused by deep networks and the lack of data augmentation technology deployed on this small dataset.

4.2. Comparison to Traditional Works

The performances of our proposed method have been extensively analyzed and compared with traditional works. Detailed results as shown in Table 2, the results outperform the current best results proposed by Dev et al. (2017), which is utterly dependent on the color characteristics. It acquires high scores to the precision, recall, and F-score, with 95%, 97%, and 92%, respectively. The segmentation speed of producing a cloud mask is also faster than traditional methods, only 0.90s. Meanwhile, the model trained on the SWIMSEG dataset is used to produce a few cloud masks in the BENCHMARK and SWIMSEG datasets. The predicted masks are

Table 2
Cloud Segmentation Performance Comparison Based on SWIMSEG Dataset

Methods	Precision	Recall	F-score	Misclassification rate	Time (s)
Li et al.	0.90	0.86	0.88	0.11	2.06
Souza et al.	0.95	0.76	0.81	0.14	2.04
Long et al.	0.71	0.98	0.80	0.23	1.83
Mantelli-Neto et al.	0.70	0.97	0.79	0.24	2.16
SLIC + DBSCAN	0.72	0.79	0.65	0.33	5.39
GRAY + SVM	0.87	0.56	0.64	0.33	2.61
LBP + SVM	0.62	0.65	0.63	0.36	4.73
ColorHIST + SVM	0.81	0.64	0.66	0.31	2.63
dSIFT + BOW + SVM	0.65	0.88	0.72	0.28	5.04
Texture + BOW + SVM	0.82	0.71	0.70	0.31	2.74
Dev et al.	0.92	0.90	0.90	0.09	1.89
Proposed method	0.95	0.97	0.92	0.04	0.90

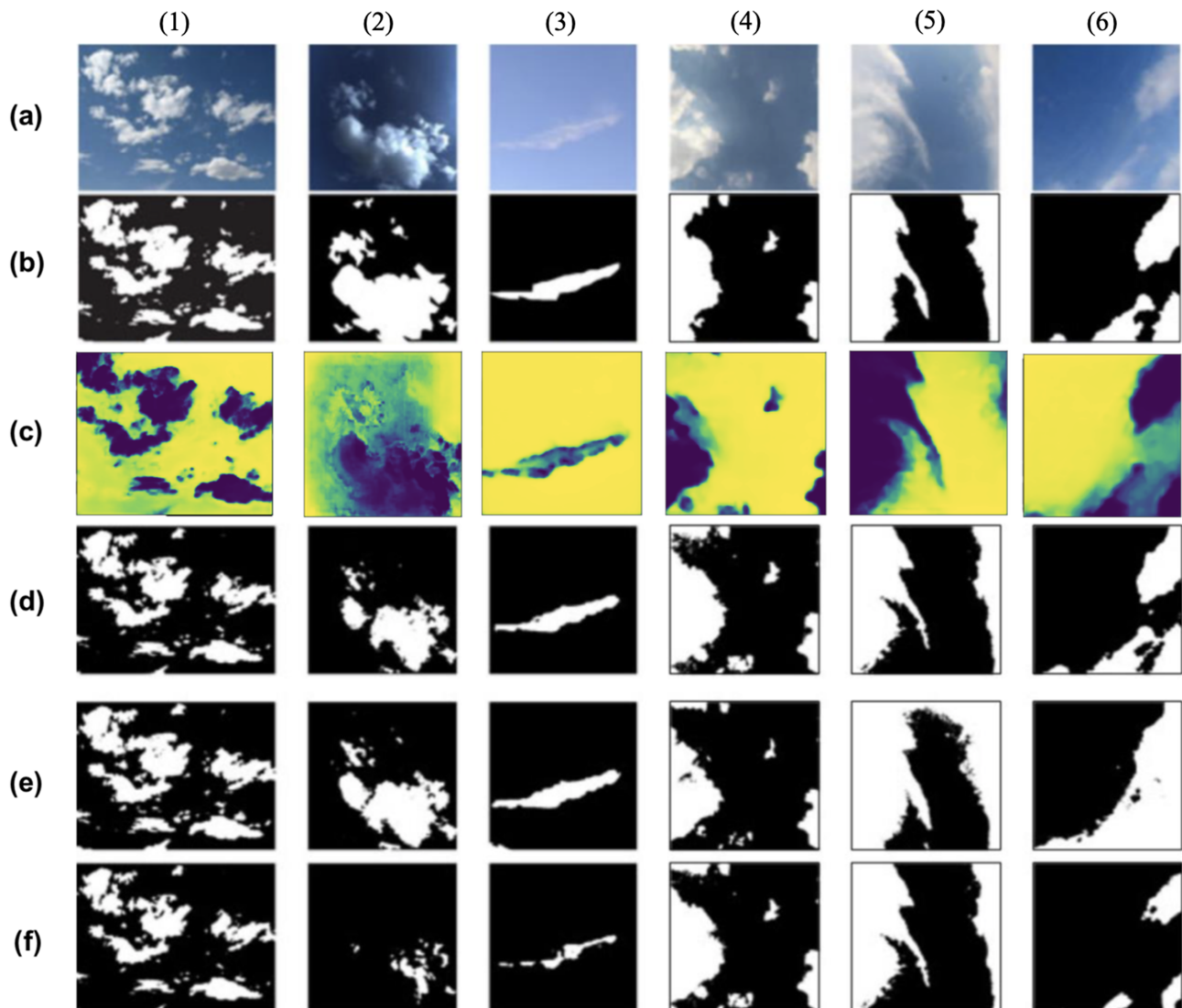


Figure 3. Cloud segmentation visualization results on the BENCHMARK (left three columns) and SWIMSEG (right three columns) datasets. (a) Input images. (b) Ground truth. (c) Our segmentation annotation results use yellow and dark purple to represent sky and cloud, respectively. (d) Dev et al. (2017). (e) Li et al. (2011). (f) Souza-Echer et al. (2006). Note that (d), (e), and (f) are directly quoted from Dev et al. (2017) in Figure 12 for comparison.

visualized in Figure 3. Overall, this model can grasp the core parts and profiles of the cloud images. That is to say, our model has learned cloud features from the SWIMSEG dataset during the training process. Although the results of the BENCHMARK dataset (Figure 3 left three columns) fail at some points compared with the ground truth such as in the (c)1 top middle, the performance of the cloud masks is more reliable by visual comparison with traditional results and ground truth. The reason why it fails may be that the quality of the generated cloud labels largely depends on the specific knowledge of experts. Those experts can not present the cloud masks of different color depths, such as the dark cloud. Unsurprisingly, the results on the SWIMSEG dataset (Figure 3 right three columns) are quite well and even more real than the ground truth at some points, such as (a)4 lower middle where the corresponding ground truth [(b)4] does not include the thin cloud. However, our generated cloud mask contains the thin cloud and can truly reflect the depth of different parts of the cloud, which can help us to distinguish different weather phenomena.

4.3. The Performance of Generating CCSN Dataset Labels

The labeling process is a quite cumbersome and time-consuming task for domain experts and it is not feasible to create all labels manually for datasets beyond dozens of petabytes. Therefore, the model trained on the SWIMSEG dataset is utilized to segment a few samples in the CCSN dataset. The segmentation performance

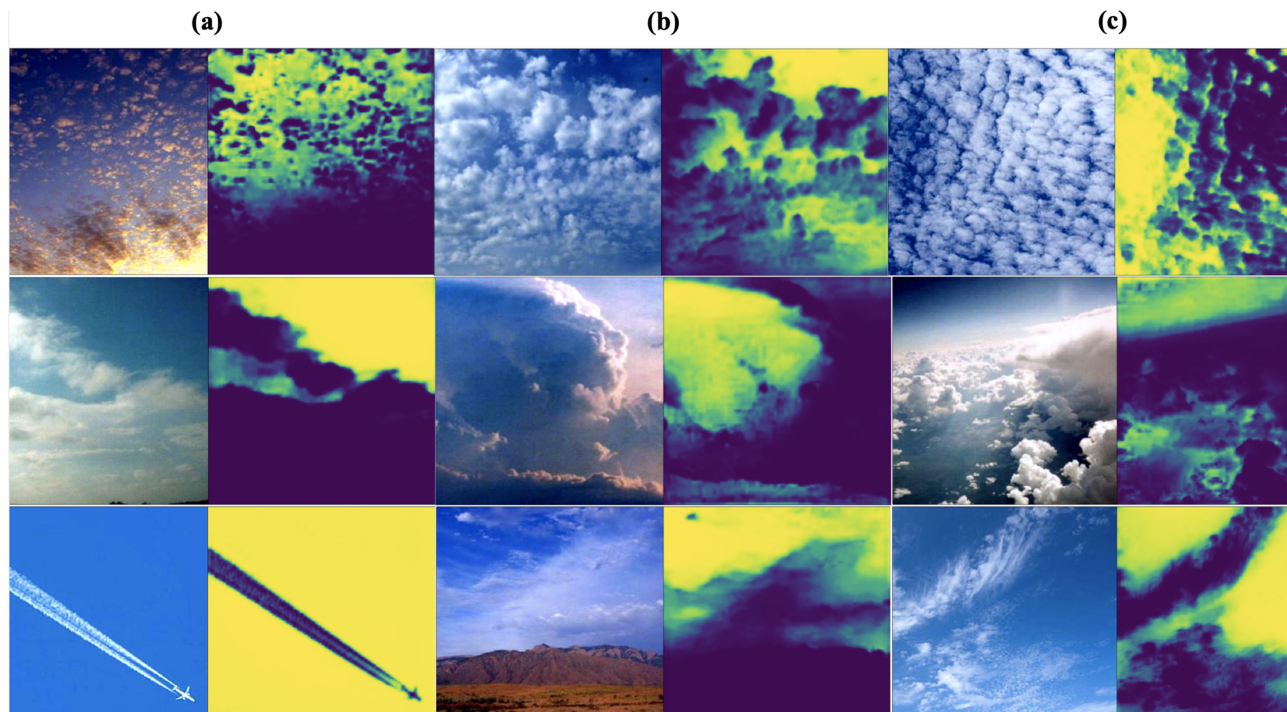


Figure 4. Segmentation results of our model on the CCSN dataset several samples. The Columns (a), (b), and (c) are the input images followed by corresponding its output segmentation labels. The dark purple represents the cloud, while the yellow represents the no cloud items.

is visualized in Figure 4. Observing the generated results, most of them can effectively produce reliable cloud masks. For example, the contrail [the bottom of Column (a) in Figure 4] presents a nearly perfect mask as well as clear boundaries. The deficiencies or mis-predictions only appear on the relatively unclear and dark boundaries among adjacent structures. For example, the SWIMSEG dataset does not include the object like the mountain [in the bottom of Column (b) of Figure 4], which will cause mis-segmentation. It also can be found that the color information in cloud images will greatly affect the final segmentation results. In the middle of the Column (b) of Figure 4, the deep color part on the top of the cumulonimbus cloud cannot be identified as a cloud.

5. Conclusion

In this paper, we study the challenging yet essential task of ground-based cloud segmentation in the pixel level based on deep learning. The 12 kinds of semantic segmentation networks and some variants performances are thoroughly evaluated on the SWIMSEG dataset. This is the first time that multiple deep learning methods have been used to comprehensively evaluate and compare the segmentation performance of ground-based cloud datasets. A novel neural network combination is introduced, which surpasses traditional methods by a large margin. And then we use trained model to the CCSN dataset for producing cloud masks. The experimental results have validated the effectiveness of our approach and shown the generalization capability and robustness of the model. Although the results are promising, there is still room for improvement. For example, the current ground truth only includes cloud or no cloud, but the actual observation may contain other objects, such as buildings and mountains. Moreover, we contribute a benchmark on the specific application of cloud segmentation to facilitate future research on the cloud segmentation problem. We believe that the cloud segmentation task will witness rapid developments with the help of a well-labeled dataset.

References

- Abadi, M. (2015). Tensorflow: Large-scale machine learning on heterogeneous systems. Retrieved from <https://www.tensorflow.org/> (Software available from tensorflow.org).
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495.
- Buslaev, A., Khvedchenya, E., Iglovikov, V. I., Parinov, A., & Kalinin, A. A. (2018). Albumentations: Fast and flexible image augmentations. ArXiv e-prints.

Acknowledgments

The authors would like to thank the editors and anonymous reviewers for their insightful comments and suggestions, which have significantly improved this paper. This work was carried out with Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJKY19_0960). The BENCHMARK and SWIMSEG datasets can be downloaded from <https://doi.org/10.7910/DVN/HEJTK1>. The CCSN dataset can be downloaded from <https://doi.org/10.7910/DVN/CADDPD>. The code in the paper is based in part on the source code of the GitHub page (<https://doi.org/10.5281/zenodo.3715378>).

- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected CRFS. arXiv preprint arXiv:1412.7062.
- Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, pp. 801–818.
- Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Li, F.-F. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 248–255.
- Dev, S., Lee, Y. H., & Winkler, S. (2017). Color-based segmentation of sky/cloud images from ground-based cameras. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(1), 231–242.
- Dev, S., Manandhar, S., Lee, Y. H., & Winkler, S. (2019). Multi-label cloud segmentation using a deep network. CoRR abs/1903.06562. Retrieved from <http://arxiv.org/abs/1903.06562>
- Drönnner, J., Korfhage, N., Egli, S., Mühling, M., Thies, B., Bendix, J., et al. (2018). Fast cloud segmentation using convolutional neural networks. *Remote Sensing*, 10(11), 1782.
- Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Rodríguez, J. G. (2017). A review on deep learning techniques applied to semantic segmentation. CoRR, abs/1704.06857. Retrieved from <http://arxiv.org/abs/1704.06857>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. CoRR, abs/1512.03385. Retrieved from <http://arxiv.org/abs/1512.03385>
- Heinle, A., Macke, A., & Srivastav, A. (2010). Automatic cloud classification of whole sky images. *Atmospheric Measurement Techniques*, 3(3), 557–567.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. CoRR, abs/1704.04861. Retrieved from <http://arxiv.org/abs/1704.04861>
- Hu, J., Shen, L., & Sun, G. (2017). Squeeze-and-excitation networks. CoRR abs/1709.01507. Retrieved from <http://arxiv.org/abs/1709.01507>
- Huang, X., & Su, H. (2008). Cloud radiative effect on tropical troposphere to stratosphere transport represented in a large-scale model. *Geophysical Research Letters*, 35, L21806. <https://doi.org/10.1029/2008GL035673>
- Jégou, S., Drozdal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Hawaii, pp. 11–19.
- Jiang, J. H., Su, H., Huang, L., Wang, Y., Massie, S., Zhao, B., et al. (2018). Contrasting effects on deep convective clouds by different types of aerosols. *Nature Communications*, 9(1), 3874.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Li, Q., Lu, W., & Yang, J. (2011). A hybrid thresholding algorithm for cloud detection on ground-based color images. *Journal of Atmospheric and Oceanic Technology*, 28(10), 1286–1296. <https://doi.org/10.1175/JTECH-D-11-00009.1>
- Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, pp. 1925–1934.
- Long, C. N., Sabburg, J. M., Calbó, J., & Pages, D. (2006). Retrieving cloud characteristics from ground-based daytime color all-sky images. *Journal of Atmospheric and Oceanic Technology*, 23(5), 633–652.
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, pp. 3431–3440.
- Lu, J., Wang, Y., Zhu, Y., Ji, X., Xing, T., Li, W., & Zomaya, A. Y. (2019). P_segnet and np_segnet: New neural network architectures for cloud recognition of remote sensing images. *IEEE Access*, 7, 87,323–87,333.
- Mantelli Neto, S. L., von Wangenheim, A., Pereira, E. B., & Comunello, E. (2010). The use of euclidean geometric distance on RGB color space for the classification of sky and cloud patterns. *Journal of Atmospheric and Oceanic Technology*, 27(9), 1504–1517.
- Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520–1528.
- Peng, C., Zhang, X., Yu, G., Luo, G., & Sun, J. (2017). Large kernel matters—Improve semantic segmentation by global convolutional network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4353–4361.
- Pohlen, T., Hermans, A., Mathias, M., & Leibe, B. (2017). Full-resolution residual networks for semantic segmentation in street scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, pp. 4151–4160.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743), 195.
- Richards, K., & Sullivan, G. D. (1992). Estimation of cloud cover using colour and texture, *BMVC92* (pp. 436–442). Springer, London: Springer.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich, Germany, pp. 234–241.
- Souza-Echer, M. P., Pereira, E. B., Bins, L., & Andrade, M. (2006). A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera. *Journal of Atmospheric and Oceanic Technology*, 23(3), 437–447.
- Su, H., Jiang, J. H., Neelin, J. D., Shen, T. J., Zhai, C., Yue, Q., et al. (2017). Tightening of tropical ascent and high clouds key to precipitation change in a warmer climate. *Nature Communications*, 8, 15,771.
- Valada, A., Vertens, J., Dhall, A., & Burgard, W. (2017). Adapnet: Adaptive semantic segmentation in adverse environmental conditions. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, pp. 4644–4651.
- Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognition*, 90, 119–133.
- Xie, W., Liu, D., Yang, M., Chen, S., Wang, B., Wang, Z., et al. (2019). Segcloud: A novel cloud image segmentation model using deep convolutional neural network for ground-based all-sky-view camera observation. *Atmospheric Measurement Techniques Discussions*, 2019, 1–16. <https://doi.org/10.5194/amt-2019-356>
- Yang, M., Yu, K., Zhang, C., Li, Z., & Yang, K. (2018). Denseaspp for semantic segmentation in street scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, pp. 3684–3692.

- Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., & Sang, N. (2018). Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, pp. 325–341.
- Zhang, J., Liu, P., Zhang, F., & Song, Q. (2018). Cloudnet: Ground-based cloud classification with deep convolutional neural network. *Geophysical Research Letters*, 45, 8665–8672. <https://doi.org/10.1029/2018GL077787>
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, pp. 2881–2890.