

A Novel Ground-Based Cloud Image Segmentation Method by Using Deep Transfer Learning

Zecheng Zhou^{ID}, Feng Zhang^{ID}, Haixia Xiao^{ID}, Fuchang Wang, Xin Hong, Kun Wu, and Jinglin Zhang^{ID}

Abstract—Cloud segmentation is fundamental in obtaining many parameters of clouds. However, traditional cloud segmentation performs far from satisfactory, due to the fuzzy boundaries and complex textures of clouds. Although deep learning methods have shown superior performance in cloud segmentation, they are constrained by limited labels in ground-based cloud image data sets. This letter established a new Ground-Based Cloud Segmentation (GBCS) data set with 1742 accurately labeled images. Then to evaluate how well deep learning models perform in cloud segmentation, 12 state-of-the-art semantic segmentation networks are selected, among which DeepLabV3+ outperformed all others. Since 1742 images are not enormous, a novel Transfer learning (TL)-DeepLabV3+ model was developed by TL: DeepLabV3+ network was trained with the PASCAL VOC 2012 data set, then retrained in GBCS. TL-DeepLabV3+ showed a high ability of cloud segmentation, scoring the Mean Intersection-over-Union (MIoU) of 91.05% in GBCS and further verified in the UTILITY data set and the Cirrus Cumulus Stratus Nimbus (CCSN) data set.

Index Terms—Cloud segmentation, ground-based cloud data set, transfer learning (TL).

I. INTRODUCTION

CLOUDS, which cover approximately 67% of the global surface [1], can help understand and identify weather phenomena and forecast meteorological changes. Due to the

Manuscript received December 18, 2020; revised March 10, 2021; accepted April 8, 2021. Date of publication April 28, 2021; date of current version January 5, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC1507002 and in part by the National Natural Science Foundation of China under Grant 41675003. (*Corresponding author: Feng Zhang*.)

Zecheng Zhou and Kun Wu are with the Key Laboratory of Meteorological Disaster, Ministry of Education (KLME)/Collaborative Innovation Center on Forecast & Evaluation of Meteorological Disaster (CIC-FEMD), Nanjing University of Information Science and Technology, Nanjing 210044, China, and also with the Shanghai Qi Zhi Institute, Shanghai 200232, China.

Feng Zhang is with the Department of Atmospheric & Oceanic Sciences and the Institute of Atmospheric Sciences, Fudan University, Shanghai 200438, China, and also with the Shanghai Qi Zhi Institute, Shanghai 200232, China (e-mail: fengzhang@fudan.edu.cn).

Haixia Xiao is with the Key Laboratory of Transportation Meteorology, China Meteorological Administration (CMA), Nanjing 210008, China, and also with the Nanjing Joint Institute for Atmospheric Sciences, Nanjing 210041, China.

Fuchang Wang is with the Department of Atmospheric & Oceanic Sciences and the Institute of Atmospheric Sciences, Fudan University, Shanghai 200438, China, and also with the Shanghai Typhoon Institute (STI), China Meteorological Administration (CMA), Shanghai 200030, China.

Xin Hong is with the Key Laboratory of Big Data Science on Knowledge Graph, Yunnan University of Software Institute, Kunming 650091, China.

Jinglin Zhang is with the College of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China.

This article has supplementary material provided by the authors and color versions of one or more figures available at <https://doi.org/10.1109/LGRS.2021.3072618>.

Digital Object Identifier 10.1109/LGRS.2021.3072618

interaction between solar and terrestrial radiation, clouds play a fundamental role in the hydrological cycle and energy balance of the atmosphere-Earth surface system [2]. Thus, cloud conditions have become vital meteorological factors in various weather phenomena, such as rainstorms and hurricanes. Most cloud-related studies are based on cloud observations by human [3]–[5]. Inevitably, such cloud observations are embedded with subjectivity and inconsistency, which may lead to different results for the same cloud conditions [6].

In recent years, cloud segmentation, which classifies each pixel of a cloud image into “sky” or “cloud,” becomes a fundamental task for further application of cloud images [5]. However, cloud segmentation is a challenging task because of the fuzzy boundaries and complex textures of clouds. Traditional cloud segmentation works generally derive from threshold methods. However, threshold-based methods heavily depend on the camera’s specification, weather conditions, and light conditions. The adaptive and automatic graph-cut method [7], [8] and superpixel segmentation algorithm [9] are proposed for cloud image segmentation to overcome the defects. Although specific progress can be achieved, there are still unsatisfactory practical measurement applications. For instance, these threshold-based methods are challenging to distinguish thin cloud, fragmentary cloud, and cloud edge. Moreover, it is difficult to differentiate between cloud and high-brightness objects due to the dependence of these methods on spectral features. Therefore, it is necessary to develop a stable and accurate cloud segmentation method.

Convolutional neural network (CNN), which is a representative algorithm in deep neural networks, has been applied in cloud analysis and achieved extraordinary progress. For cloud segmentation, due to the lack of corresponding annotations of clouds, most deep learning methods are based on satellite data [1], [10], [11], [12], and only a few methods are based on ground-based cloud data set. Dev *et al.* [13] designed a lightweight deep learning model named CloudSegNet and achieved cloud image segmentation for both daytime and nighttime. Shen *et al.* [14] proposed a symmetrical densely connected CNN model-based cloud image segmentation method to achieve further accurate segmentation of the ground-based cloud images. Moreover, SegCloud, a deep learning model trained on 400 whole sky images with annotations, shows effective and accurate results for ground-based cloud segmentation (GBCS) [15]. Additionally, an end-to-end cloud segmentation neural model was proposed by Song *et al.* [16] to extract the features learned from the source and target domain.

Although extensive studies have shown that deep learning can achieve good cloud segmentation results, there are still not enough labeled cloud image data sets in practical application. These models usually show the disadvantage of insufficient generalization ability. Transfer learning (TL) aims to improve the performance of target networks on target domains by transferring the knowledge contained in other source domains [17]. In this way, the dependence on extensive target-domain data can be reduced when constructing the target network. A commonly adopted TL approach is to utilize the weights pretrained on a large data set instead of random initialization. In the field of cloud-related research, TL has already achieved outstanding performance on satellite data [12], [18]. However, to the best of our knowledge, few studies on ground-based cloud image segmentation utilize TL.

In this letter, we first built up the GBCS data set consisting of 1742 images, each of which was accurately labeled by manual operation and threshold method. Next, we trained 12 classical semantic segmentation models from scratch and then compare their ability in extracting cloud features. The model with the best segmentation performance was selected for TL.

The letter is organized as follows. The details about the GBCS data set are shown in Section II. Section III introduces the methods, the experimental details, and parameter settings. The experimental results and analysis are presented in Section IV. Finally, the conclusions and future works are summarized in Section V.

II. DATA SET

The quality of data set will directly affect the performance of cloud segmentation. Therefore, it is necessary to use a well-organized data set with as accurate labels as possible for deep learning. However, it is very time-consuming to obtain the corresponding labels of cloud segmentation. At present, there are some problems in the current data sets, e.g., a small amount of data and insufficient annotation. For instance, the UTILITY data set, collected by Li *et al.* [4], includes 1000 cloud images but only 32 images have labels; the Whole-Sky Image Segmentation (WSISEG) [19], which consists of 400 whole-sky annotated images. The two data sets are insufficient for model training in deep learning. Moreover, the Singapore Whole Sky Image Segmentation (SWIMSEG) built by Dev *et al.* [20] contains 1013 images, and the data set is annotated with binary masks. Although SWIMSEG is much larger than the former two and more suitable for deep learning tasks, its quantity of samples is small.

To improve these existing problems, we built a labeled GBCS dataset¹ that was collected from two sources: 1) the machine learning competition of data fountain and 2) internet. The data set contains 1742 ground-based cloud images, and each image has its corresponding label. These cloud images were cropped to ensure that only clear sky and cloud elements are included. Additionally, all the images were resized to 512 × 512 pixels. The labels were processed as binary images

¹Some examples are provided in the supplementary material.

in the same size, in which the cloud pixels were marked as 1, and the sky pixels were marked as 0. Two methods were adopted to generate these labeled images: A) the cloud regions were manually annotated utilizing Labelme software. However, for the clouds with complex boundaries such as fragmentary clouds and flocculent clouds, it was difficult to operate manually and likely to result in errors. Hence, for the left cloud images B) was applied: we converted a number of images into grayscale images with a single channel. We then generated binary images via the threshold method. Note that, method B can only be used for clouds with clear boundaries, consistent brightness, and no cloud shadows due to the shortness of the threshold-based method mentioned above. Consequently, only 75 images were selected for method B by visual interpretation to guarantee the precision of labels. Based on the two methods, the labels of the GBCS data set can reach high accuracy.

III. METHODS AND IMPLEMENTATION DETAILS

We selected 12 different state-of-the-art semantic segmentation models for cloud segmentation. These models are FCN-8s, FCN-16s, FCN-32s [21], UNet [22], SegNet [23], PSPNet [24], RefineNet [25], PAN [26], DeepLabV3 [27], DeepLabV3+ [28], DenseASPP [29], and BiSeNet [30]. The cloud segmentation models were developed by training these models with GBCS data set.

To quantitatively estimate the capability and effectiveness of different models in extracting and decoding cloud features, four evaluation metrics called Pixel accuracy (PA), Mean PA (MPA), F1-score, and Mean Intersection-over-Union (MIoU) were calculated.²

The data set was randomly divided into training set, validation set, and testing set with a ratio of 8:1:1. To reduce the occupation of video memory, we set a random crop for each image with the size of 256 × 256 and used the batch size of 4. Subsequently, the parameters of our experiments were updated by the Adam optimizer [31] with an initial learning rate of 0.001. The 12 different models were trained from scratch with Tensorflow framework [32] on an NVIDIA GeForce GTX1080Ti graphics card. We selected the binary cross-entropy (BCE) loss as the loss function in the training stage. The BCE represents the distance between the true distribution and the predicted distribution. It is calculated as follows:

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log p(y_i) + (1 - y_i) \cdot \log (1 - p(y_i))] \quad (1)$$

where y is the true label (1 for cloud elements and 0 for sky elements), and $p(y)$ is the predicted probability of the pixel being positive for all N pixels. Note that, we did not apply any data augmentation techniques.

Next, by quantitative performance analysis of the training, validation, and testing set, the most state-of-the-art model called DeepLabV3+ with the best segmentation performance

²The details of these metrics are introduced in the supplementary material.

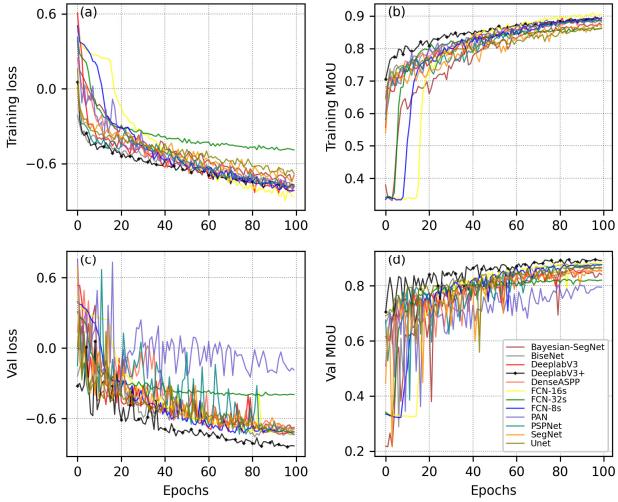


Fig. 1. BCE loss values (in log scale) at (a) training stage and (c) validation stage. MIoU at (b) training and (d) validation stage.

was selected for deep TL. We initialized the backbone of DeepLabV3+ with the weights pretrained on the PASCAL VOC 2012 [33] instead of random initialization. Finally, we reported the final results of DeepLabV3+ for TL (hereinafter called TL-DeepLabV3+).

IV. EXPERIMENT RESULTS

A. Evaluate the Performance of Different Models

To compare the performance of the 12 models, the BCE loss and MIoU values were calculated during the training and validation stages. The trend of BCE loss and MIoU are shown in Fig. 1. According to the performance of validation set, we finally set the epochs as 100 to ensure that the validation loss and MIoU become stable. In the training stage, most models exhibit similar and stable performance of the cross-entropy loss and MIoU after a few epochs, and MIoU of the 12 models are convergent at approximately 84%–88% after dozens of epochs. However, for the validation set, these models show apparent differences. As the epoch increases, the cross-entropy loss of several models becomes very unstable, such as PAN and PSPNet. Meanwhile, PAN performs the worst with the highest cross-entropy loss values and lowest MIoU, while DeepLabV3+ performs the best due to its lowest loss values and highest MIoU.

Semantic segmentation task is a pixel-level classification task. Thus, to further compare the performance of these classifiers, the receiver operating characteristic (ROC) curve [34] was employed as well. First, the softmax layer was added at the last layer of each model to acquire the category probability of each pixel. Then we gradually altered the threshold from 0 to 1 in steps of 0.01, then calculated the false positive rate (FPR) and true positive rate (TPR) of each threshold, respectively. The FPR and TPR can be calculated as follows:

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (2)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (3)$$

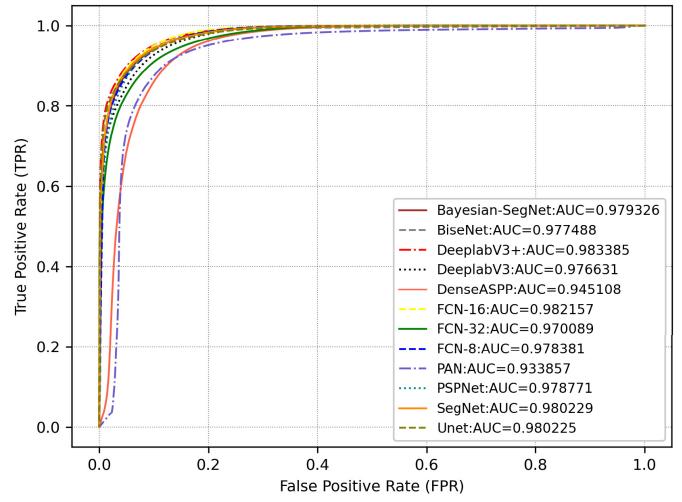


Fig. 2. ROC curve of the 12 models. The AUC represents the area under the curve.

TABLE I
EVALUATION METRICS OF THE 12 MODELS FOR THE TESTING SET. NOTE THAT TL-DEEPLABV3+ IS THE RESULT OF SECTION IV-B

Models	PA(%)	MPA(%)	F1-Score (%)	MIoU(%)
SegNet	94.39	93.96	94.60	86.88
Bayesian-SegNet	94.00	94.17	94.16	86.01
DeepLabV3	94.02	94.21	94.26	85.98
DeepLabV3+	95.43	94.78	95.30	88.85
BiSeNet	95.27	95.11	95.42	88.32
DenseASPP	91.08	92.19	91.86	81.46
FCN-8s	95.03	94.65	95.11	87.79
FCN-16s	95.47	94.77	95.51	88.70
FCN-32s	92.48	89.92	92.65	82.56
PAN	90.50	92.63	90.86	81.58
UNet	94.08	93.90	94.18	86.54
PSPNet	93.87	92.74	89.95	86.14
TL-DeepLabV3+	96.28	95.44	96.26	91.05

For an ideal classifier, the FPR value should be 0 and the TPR value should be 1, which means the larger the area under the curve (AUC) is, the better the model performs. The ROC results of the 12 models are shown in Fig. 2. The AUC values of all these models are above 0.9, which means these models can all achieve outstanding segmentation performance. Especially, DeepLabV3+ has the largest AUC value of 0.983385 and PAN has the smallest AUC value of 0.933857, which indicates that the DeepLabV3+ has the best segmentation performance for the training and validation stages.

For the testing set, we calculated the evaluation metrics mentioned in Section III, respectively. The results of different models are summarized in Table I. In general, the evaluation metrics of DeepLabV3+, BiSeNet, and FCN-16s are significantly higher than other models, in which PA and F1-Score reach more than 95%, and only the MPA value of BiSeNet reaches 95%. Besides, the performance of DeepLabV3+, BiSeNet, and FCN-16s is immensely close, and the MIoU of DeepLabV3+ is 88.85%, which is the best result in all the models. Overall, in terms of comprehensive evaluation of the model performance on training set, validation set, and

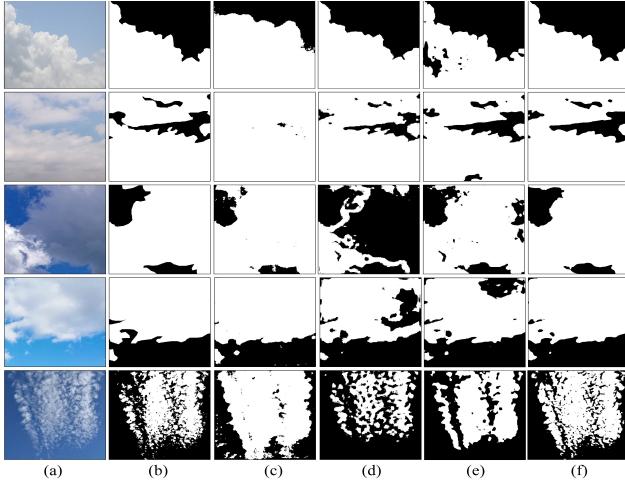


Fig. 3. Visualization results on GBCS data set. (a) Input images. (b) Ground truth. (c) SegCloud. (d) CloudSegNet. (e) DeepLabV3+. (f) TL-DeepLabV3+.

testing set, DeepLabV3+ appears to be the reasonable choice for cloud segmentation.

B. Developed the Cloud Segmentation Model by TL

According to the evaluation above, DeepLabV3+ was selected for cloud segmentation. We transferred the weights, which were pretrained on the PASCAL VOC 2012 data set and had achieved outstanding performance of 87% MIoU, to initialize the backbone of DeepLabV3+ to finetune our model, TL-DeepLabV3+. We reset the ratio of training set and testing set to 4:1, without a validation set. Moreover, the momentum optimizer and poly learning rate schedule with the momentum of 0.9 and initial learning rate of 0.0001 were used. The other experimental settings are the same as Section III.

Similarly, after training 100 epochs, we also calculated the PA, MPA, F1-Score, and MIoU, respectively. When the knowledge learned in the source domain has a negative effect on the target domain, it is called negative transfer [35]. Compared with DeepLabV3+ with random initialization in Table I, the values of all these evaluation metrics have significantly increased. The values of PA, MPA, and F1-Score increased by about 1%, and MIoU even increased by 3%. The results demonstrate that TL can effectively improve the performance of DeepLabV3+ in cloud segmentation task, which achieve positive transfer.

We visualized the output of cloud masks from three deep learning methods [SegCloud [15], CloudSegNet [13], DeepLabV3+, column (c)–(e)], and TL-DeepLabV3+ (f) to compare their performance intuitively, and displayed some samples in Fig. 3. Overall, these four deep learning methods can achieve precise segmentation for the testing set. Besides, the proposed TL-DeepLabV3+ performs better than the other methods, especially for the cloud shadows and fragmentary clouds. For example, CloudSegNet and DeepLabV3+ have certain misjudgment for cloud shadows in the dark areas in the third and fourth row [column (d) and (c)]. Additionally, SegNet has a higher missing rate for the cloud [first and second rows in Fig. 3(c)]. The fifth row illustrates a sample labeled by method B. Adopting method B can improve the ability of

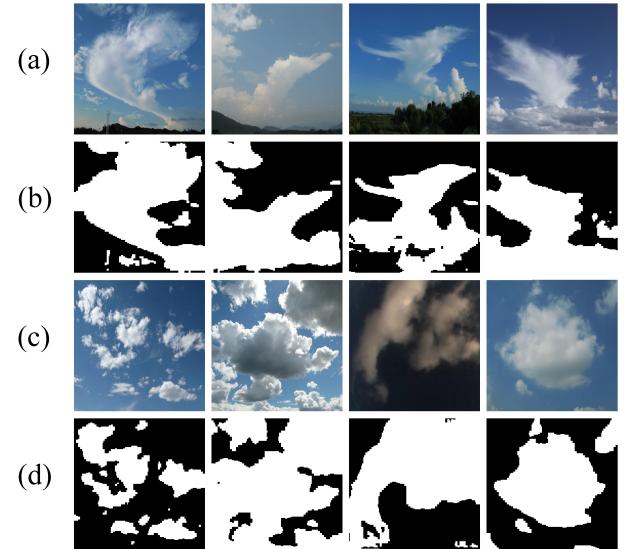


Fig. 4. Output images of TL-DeepLabV3+ with the other data sets. (a) Samples of CCSN data set. (b) Corresponding prediction results of CCSN data set. (c) Samples of UTILITY data set. (d) Corresponding prediction results of UTILITY data set.

these models to extract the features of fragmentary clouds. The prediction of TL-DeepLabV3+ expresses more details and is most consistent with the ground truth image. Furthermore, we can focus on Fig. 3(e) and (f) to assess the positive efficiency of TL. DeepLabV3+ are easier to misjudge cloud shadows, e.g., first to fourth rows in Fig. 3(e). After utilizing TL, TL-DeepLabV3+ has fewer errors and omissions and can extract more detailed cloud information [column (f)].

C. Segmentation Performance on Other Data Sets

Strong generalization ability and robustness are also quite indispensable attributes of a state-of-the-art model. Therefore, The Cirrus Cumulus Stratus Nimbus (CCSN) data set and the UTILITY data set were put into the TL-DeepLabV3+. A few samples of segmentation results were visualized in Fig. 4. The results show that most of the clouds in the images can be accurately recognized by the model, such as the first column and second column in Fig. 4(a) and (b). Especially the picture in the fourth column of Fig. 4 obtains nearly perfect performance. The high-accuracy prediction results can be processed to the cloud masks of these nonannotated data sets, increasing the amount of labeled cloud segmentation data sets. For those images that contain elements other than clouds or sky [left three columns in Fig. 4(a)], the model keeps a relatively accurate recognition rate for the parts of cloud and sky. However, the other elements such as mountains and vegetation are obviously misjudged by the model because the GBCS data set only includes sky and clouds, and the model does not learn the features of other elements. The image in the third column of Fig. 4(c) has the worst performance in these samples. It may be because this image is a nighttime cloud image with bad illumination, which is not contained in the GBCS data set. Thus, it is indispensable to add cloud images under diverse light conditions in future data sets. Overall, the performance on other data sets proves the excellent generalization ability and robustness of TL-DeepLabV3+.

V. CONCLUSION

In this letter, a well-annotated GBCS data set containing 1742 images has been released, and we adopt two methods to generate annotations for these images: one is the manual operation with visual interpretation by Labelme software; the other is to use threshold method with manual screening to generate the binary masks of thin or fragmentary clouds efficiently. Twelve outstanding models were employed to train from scratch with the GBCS data set for segmenting clouds, then the performance of these models was compared quantitatively. Finally, the DeepLabV3+ model that performed the best was selected for transferring knowledge of parameters. The TL-DeepLabV3+ was further trained and evaluated, and the results prove that deep TL has a positive effect and efficiently improves the accuracy of cloud segmentation tasks. The other two existing data sets are also input into TL-DeepLabV3+, and the prediction results show that the partial output can be used as the labels of clouds after visual comparison. This can significantly reduce the time consumption and huge workload in the annotation process, and further increase the data set volume of cloud segmentation. The cloud images must include other elements like cloud shadows, water bodies, vegetation, mountains, and buildings in practical application. Therefore, in future work, we will face certain challenges to add various categories of labels instead of only two classes (cloud and sky).

REFERENCES

- [1] A. Garcia-Garcia, S. Orts-Escalano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," Apr. 2017. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2017arXiv170406857G>
- [2] G. L. Stephens, "Cloud feedbacks in the climate system: A critical review," *J. Climate*, vol. 18, no. 2, pp. 237–273, Jan. 2005.
- [3] J. Calbó and J. Sabburg, "Feature extraction from whole-sky ground-based images for cloud-type recognition," *J. Atmos. Ocean. Technol.*, vol. 25, no. 1, pp. 3–14, Jan. 2008.
- [4] Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *J. Atmos. Ocean. Technol.*, vol. 28, no. 10, pp. 1286–1296, Oct. 2011.
- [5] C. N. Long, J. M. Sabburg, J. Calbó, and D. Pagès, "Retrieving cloud characteristics from ground-based daytime color all-sky images," *J. Atmos. Ocean. Technol.*, vol. 23, no. 5, pp. 633–652, May 2006.
- [6] D. V. Hoyt, "Interannual cloud-cover variations in the contiguous united states," *J. Appl. Meteorol.*, vol. 17, no. 3, pp. 354–357, Mar. 1978.
- [7] S. Liu and Z. Zhang, "Adaptive graph cut based cloud detection in wireless sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 2015, no. 6, pp. 1–7, 2015.
- [8] S. Liu, Z. Zhang, B. Xiao, and X. Cao, "Ground-based cloud detection using automatic graph cut," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1342–1346, Jun. 2015.
- [9] S. Liu, L. Zhang, Z. Zhang, C. Wang, and B. Xiao, "Automatic cloud detection for all-sky images using superpixel segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 354–358, Feb. 2015.
- [10] J. Lu *et al.*, "P_SegNet and NP_SegNet: New neural network architectures for cloud recognition of remote sensing images," *IEEE Access*, vol. 7, pp. 87323–87333, 2019.
- [11] J. Drönner *et al.*, "Fast cloud segmentation using convolutional neural networks," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1782.
- [12] G. Mateo-García, V. Laparra, D. López-Puigdolers, and L. Gómez-Chova, "Transferring deep learning models for cloud detection between Landsat-8 and Proba-V," *ISPRS J. Photogramm. Remote Sens.*, vol. 160, pp. 1–17, Feb. 2020.
- [13] 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.
- [14] H. Shen, X. Min, S. Bicheng, and L. Jia, "Ground-based cloud image segmentation method based on symmetric convolutional neural network with dense connection," *Comput. Eng. Appl.*, vol. 55, no. 17, pp. 207–213, 2019.
- [15] W. Xie *et al.*, "SegCloud: A novel cloud image segmentation model using a deep convolutional neural network for ground-based all-sky-view camera observation," *Atmos. Meas. Techn.*, vol. 13, no. 4, pp. 1953–1961, Apr. 2020.
- [16] Q. Song, Z. Cui, and P. Liu, "An efficient solution for semantic segmentation of three ground-based cloud datasets," *Earth Space Sci.*, vol. 7, no. 4, Apr. 2020, Art. no. e2019EA001040.
- [17] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.
- [18] C. Gonzales and W. Sakla, "Semantic segmentation of clouds in satellite imagery using deep pre-trained U-Nets," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2019, pp. 1–7.
- [19] T. Fa, W. Xie, Y. Wang, and Y. Xia, "Development of an all-sky imaging system for cloud cover assessment," *Appl. Opt.*, vol. 58, no. 20, pp. 5516–5524, Jul. 2019.
- [20] S. Dev, Y. H. Lee, and S. Winkler, "Color-based segmentation of sky/cloud images from ground-based cameras," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 231–242, Jan. 2017.
- [21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
- [22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
- [23] 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.
- [24] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.
- [25] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5168–5177.
- [26] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," May 2018, *arXiv:1805.10180*. [Online]. Available: <http://arxiv.org/abs/1805.10180>
- [27] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," Jun. 2017, *arXiv:1706.05587*. [Online]. Available: <http://arxiv.org/abs/1706.05587>
- [28] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," Feb. 2018, *arXiv:1802.02611*. [Online]. Available: <http://arxiv.org/abs/1802.02611>
- [29] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "DenseASPP for semantic segmentation in street scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3684–3692.
- [30] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiSeNet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 334–349.
- [31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [32] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," Mar. 2016, *arXiv:1603.04467*. [Online]. Available: <https://arxiv.org/abs/1603.04467>
- [33] M. Everingham, S. Eslami, L. Van Gool, C. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, pp. 98–136, Jun. 2014.
- [34] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [35] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.