Response Letter

Dear editor and reviewers,

Sincerest thanks for your response and reviewers' comments on our manuscript entitled "SegCloud: A novel cloud image segmentation model using deep convolutional neural network for ground-based all-sky-view camera observation" (amt-2019-356). These comments are very valuable and helpful for revising and improving our paper, and they also have important guiding significance to our researches. We have studied comments carefully and made revisions which we hope meet with approval. This manuscript also has been modified by a professor whose native language is English to improve the English writing. Revised portion are marked in red in the resubmitted manuscript.

The main revisions in the paper and the responses to the reviewers' comments are as following:

Reviewer #1:

1. As viewing the sample images, there are some doubtful regions found around the sun in the original sample image for "Clear sky" shown in Fig.4a (First row). While these look like small clouds or reflected images of dome, the system judges these are no cloud.

Reply: Thanks for your concern. The sample image shown in Fig.4a is indeed the image of clear sky. The whole sky images used in experiment are captured by our all-sky imaging instrument as introduced in the Section 2. A plastic protective dome is added to the front of the imaging system to protect the device from dust and rain. However, after long-term field observation work, dust may fall on the protective dome and the protective dome has been some aged, which affects the imaging process and results in the existence of gray pixels that similar to cloud points.

2. An original image to be analyzed is fused with ten different exposure-photos (P3, L28-30). By using such a technique, it has a possibility to include cloud information around the sun. However, the paper shows no description on this effect when analyzing it. If the accuracy of cloud segmentation is improved by using the technique, please discuss it more, especially for aureole areas.

Reply: Thanks for your comment. The high dynamic range technique do guarantee better imaging and high-quality whole sky images, but it belongs to the part of the imaging system of all-sky imaging instrument. This manuscript presented in this paper mainly focuses on the cloud image segmentation algorithm, so the high dynamic range technique is not described in detail.

3. In the Tao et al.'s paper they have discussed the analysis system of cloud

segmentation called the optimized U-Net, which looks like similar to the present SegCloud. Because several authors are overlapped with those of that paper and one is the leading author of the present paper, they should discuss the difference of both systems and analyzed results in this paper clearly.

Reply: Thanks for your comment. The manuscript presented in this paper is different from the Tao et al.'s paper. Tao et al.'s paper focuses on the all-sky imaging instrument (ASC), and the ASC hardware system is mainly detailed, including the imaging system, data analysis module, et al. On the basis of these hardware system, the U-Net model is introduced into the cloud cover analysis. However, this paper focuses on the theoretical analysis of whole sky image segmentation algorithm. The new convolutional neural network model named SegCloud is mainly proposed for accurate whole sky image segmentation. The proposed SegCloud has been compared to other algorithms qualitatively and quantitatively according to ground truth images, which demonstrates its accuracy and superiority. Thus, these two papers are totally different.

4. Tao et al. (2019) have described the new database created for the system. This database must be the same as in this paper. Therefore, in this paper they should write "This database has been used in the present analysis, created in Tao et al. (2019)". The referee thinks that it is not suitable to include the contribution of the database production for training the SegCloud (P3, L15-16).

Reply: Thanks for your reminder. The data are actually mentioned and briefly described in Tao et al.'s paper, but no further details are presented. We realize it's not appropriate to emphasize the contribution in this paper. We have removed the relative content about the contribution of the database production in this paper, and have corrected the text to "In this paper, the database used in Tao et al. (2019) is applied to train and test the proposed SegCloud model". But in order to keep the integrity of the proposed algorithm, this manuscript still presents the database in detail, including its advantage and features.

5. Tao et al. (2019) have discussed the relation between the U-Net results and the human observers' ones in detail. The similar discussion is found in the section 4.3 basically by using the similar data frame (This paper used the data only of July, 2018, and Tao et al. used the data of August to November, 2018. The data site is the same.) If both analysis systems are different, it is useful and effective to discuss the difference between both results, but if not, this section may not be required in the paper because the detailed discussion has been already performed by using much more data.

Reply: Thanks for your comment. In Tao et al.'s paper, the ASC instrument has been running stably at the airport (data site) for cloud cover observation in real time. The accuracy of real-time cloud cover observation is affected by both the U-Net algorithm and the observation system. So, in Tao et al.'s paper, the purpose of the cloud cover comparison between the ASC instrument and human observer is to verify the stability and reliability of the ASC instrument. As to the manuscript presented in this paper, the

database used to train and test the proposed SegCloud algorithm is made by ourselves. To further demonstrate the persuasiveness and advantages of the proposed whole sky image segmentation algorithm, the human observer's cloud cover data is treated as ground-truth. At the same time, the whole sky images from ASC were downloaded and segmented, and then the cloud cover were computed and compared to with the data of human observation. As mentioned above, the purpose of the comparison with human observer's data in those two papers is different.

6. While the accuracy computed with respect to the ground truth is important and correctly reported, I would like to see more work regarding the comparison with other algorithms. The fixed threshold algorithm cannot be directly compared with the CNN output simply because it has been built to be used with a camera with a shadow band. Then, the actual fixed threshold algorithm (Long et al., 2006) is much more complicated than a simple R/B threshold shown in Fig. 4, and accounts (at least partially) for most of the issues mentioned in the current manuscript (e.g., solar obstruction).

Reply: Thanks for your comment. During the experiments, in addition to the R/B threshold algorithm and the Otsu algorithm, some other clustering algorithms were tested for whole sky image segmentation, such as k-means algorithm and mean-shift algorithm. None of those algorithms has satisfactory performance. The reasons for these results are that they require pixels of the same class to have similar gray value but clouds appear to be opposite (the same reason with Otsu algorithm, as introduced in Line 25, Page 7). We also test the other threshold segmenting algorithm using Red and Blue channel values, such as R-B, but their performance is similar to R/B threshold algorithm. So in this paper, we just pick two typical algorithms and compare our algorithm with them. For most cameras, in order to protect the CCD from the direct sunlight and avoid the large sun circle, shadow band will be added during the imaging process. Different from other cameras, our instrument uses the light-cutting module and high dynamic range technique, sun circle is small. So we think it's appropriate to use the R/B threshold algorithm as the comparison algorithm and it will not result in much bias. In Long et al.'s paper, the R/B threshold value is set as 0.6 through several tests performed on training images. But in our test experiment, we found the best threshold value is 0.77. Therefore, we choose 0.77 as the final threshold.

7. Since Tao et al. (2019) already showed validations results with respect to visual observations for August-November 2018, please remove this part from the manuscript. I suggest including some applications exploiting the R/B method and its potentialities. For example, comparison with cloud fraction (or cloud mask) estimated by satellites (e.g., Himawari-8, MODIS), analysis of trend/changes at different locations, adapting the same algorithm to images recorded by other cameras (since the method is based on CNN, it should be no so difficult).

Reply: Thanks for your suggestion. As we mentioned in comment #5, the purpose of comparison is different. Your suggestion all are great and they will be considered in our

further study. Thank you again.

8. The reference of the second segmentation method is Otsu (2007). However, I think that the authors actually refer to Otsu (1979).

Reply: Thanks for your reminder. We have corrected the reference information.

9. The name of the two previous datasets showed in Fig. 2 are not mentioned in the text of the manuscript. In section 2, please include their names when their respective references are mentioned

Reply: Thanks for your reminder. The names of the two databases have been added to the location where the respective references are mentioned.

Reviewer #3:

Page 1:
1. Line 29-31, one more reference could be added, Yang et al. (2017, doi: 10.1002/2016JD025954), which shows the application of ground-based cloud observation for evaluating satellite-based observations.

Reply: Thanks for your comment. The related reference has been added in Line 28.

2. Line 32-34, "are used to detect clouds". Also, one more references could be added, Yang et al. (2018, doi:10.1016/j.atmosres.2017.11.021).

Reply: Thanks for your comment. The sentence has been corrected to "are used to detect clouds", and the reference also has been added.

Page 2
3. Line 3-5, "… for recent years".

Reply: Thank you for pointing it out. The word "in recent years" has been corrected to "for recent years".

4. Line 7-9, I am a little confused with this sentence. I understand that the aerosol particles along with some small cloud droplets follows Mie scattering. However, for air molecules, they generally follow Rayleigh scattering.

Reply: Thank you for pointing it out. Now, the sentence has been corrected as "Traditional segmentation methods generally use "color" as a distinguishing factor between clouds and clear sky because cloud particles have similar scattering intensity in blue and red bands due to the Mie scattering effort. By contrast, air molecules have

more scattering intensity in the blue band than in the red band due to the Rayleigh scattering theory".

5. Line 13, "treated" Line 14-15, this seems not a complete sentence.

Reply: Thank you for your comment. The related content all have been changed to the past tense.

6. Line 21-23, "technology" -> "technologies". Also, this sentence seems with grammar error.

Reply: Thank you for pointing it out. The word "technology" has been corrected into "technologies". The relative sentence has been corrected as "Convolutional neural networks (CNNs) are outstanding and powerful object recognition technologies, which have been widely applied in many fields, such as computer vision and pattern recognition"

7. Line 26-27, "is" -> "was"

Reply: Thanks for your careful check. The word "is" has been corrected into "was".

Page 3
8. Line 31, where is description for Fig. 1(a)?

Reply: Thanks for your reminder. In Line 26, "The appearance and functional specifications of ASC are shown in Fig. 1", Fig. 1 actually means Figs. 1(a) and 1(c). Now, we has corrected the expression.

Page 4:
9. Line 4-5, where is description for Fig. 2 (a) and (b)?

Reply: Thanks for your reminder. We has added the description about Figs. 2(a) and 2(b) in Line 26-27.

Page5
10. Line 7, "accepts"

Reply: Thank you for pointing it out. The word "accept" has been corrected into "accepts".

11. Line 9, "are" ->"is"

Reply: Thank you for pointing it out. The word "are" has been corrected into "is".

12. Line 27-29, "are the locations",

Reply: Thank you for pointing it out. The word "is the location" has been corrected into "are the location".

13. Line 30, "ensure effectively to restore …"

Reply: Thank you for pointing it out. The sentence has been corrected.

14. Line 32, "achieve … cost …" ?

Reply: Thanks for your comment. The sentence has been changed as "Although pooling indices have advantage in computational time, they may lead to a slight loss of cloud boundary details".

Page 6
15. Line 13, "is"->"are"

Reply: Thank you for pointing it out. The word "is" has been corrected into "are".

16. Line 14, the last sentence seems not a complete sentence.

Reply: Thank you for pointing it out. The sentence has been corrected as "Thus, the final segment results are outputted".

17. Line 17-20, please modify the description to make them more concise.

Reply: Thanks for your suggestion. The related description has been modified as "SegCloud is then trained on NVIDIA GeForce GTX1080 hardware and machine learning software package named TensorFlow. Mini-batch gradient descent is used as an optimization algorithm to find the appropriate model weights. During training, the number of whole-sky images fed to the SegCloud model per batch is 10, and momentum parameter with a decay of 0.9 is used (Sutskever et al., 2013)". We hope these expression would be more concise.

Page 7
18. Lines 23-27, the performances are great. However, if you could provide some explanations or discussions regarding those that are not accurately classified, it would be more useful.

Reply: Thanks for your suggestion. Although SegCloud shows advantages in whole-sky image segmentation, some misidentification remains due to decreased recognition for extremely thin clouds, which should be investigated in the future. The related content has been added in Line 8-9, Page 8.

Page 8

19. Line 7, why do you only provide "some representative segmentation results"? How do you make the choice of "some", subjectively or objectively?

Reply: Thanks for your comment. In this work, we test 60 images, we cannot show all segmented results in the manuscript. So we randomly choose four whole sky images and their segmentation results under the precondition of including different cloud cover. Therefore, these four images separately show clear sky, partial cloudy sky and overcast sky.

20. Line 7, "poorly" -> "poor"

Reply: Thank you for pointing it out. The word "poorly" has been corrected into "poor".

21. Line 13-14, do you mean "more excellent"/"more accurate"?

Reply: Yes, the sentence has been corrected as "The R/B threshold method has more accurate segmenting results compared with Otsu algorithm".

SegCloud: A novel cloud image segmentation model using deep convolutional neural network for ground-based all-sky-view camera observation

Wanyi Xie[1, 2], Dong Liu[1, 2], Ming Yang[4], Shaoqing Chen[4], Benge Wang[4], Zhenzhu Wang[1, 2], Yingwei Xia[3], Yong Liu[2, 3], Yiren Wang[3], and Chaofang Zhang[3]

[1]Key Laboratory of Atmospheric Optics, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230088, China
[2]University of Science and Technology of China, Hefei 230026, China
[3]Opto-Electronics Applied Technology Research Center, Institute of Applied Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, 230031, China
[4]Anhui Air Traffic Management Bureau, Civil Aviation Administration of China, Hefei, 230094, China

Correspondence: Yiren Wang (wyiren90@mail.ustc.edu.cn) and Chaofang Zhang (zcf0413@mail.ustc.edu.cn)

Abstract. Cloud detection and cloud properties have substantial applications in weather forecast, signal attenuation analysis, and other cloud-related fields. Cloud image segmentation is the fundamental and important step in deriving cloud cover. However, traditional segmentation methods rely on low-level visual features of clouds and often fail to achieve satisfactory performance. Deep convolutional neural networks (CNNs) can extract high-level feature information of objects and have achieved remarkable success in many image segmentation fields. On this basis, a novel deep CNN model named SegCloud is proposed and applied for accurate cloud segmentation based on ground-based observation. Architecturally, SegCloud possesses a symmetric encoder–decoder structure. The encoder network combines low-level cloud features to form high-level, low-resolution cloud feature maps, whereas the decoder network restores the obtained high-level cloud feature maps to the same resolution of input images. The Softmax classifier finally achieves pixel-wise classification and outputs segmentation results. SegCloud has powerful cloud discrimination capability and can automatically segment whole-sky images obtained by a ground-based all-sky-view camera. The performance of SegCloud is validated by extensive experiments, which show that SegCloud is effective and accurate for ground-based cloud segmentation and achieves better results than traditional methods do. The accuracy and practicability of SegCloud is further proven by applying it to cloud cover estimation.

1 Introduction

Clouds are among the most common and important meteorological phenomena, covering over 66% of the global surface (Rossow and Schiffer, 1991; Carslaw and Ken, 2009; Stephens and Graeme, 2005; Zhao et al., 2019; Wang et al., 2017). The analysis of cloud condition and cloud cover plays a key role in various applications (Papin et al., 2002; Yang et al., 2014; Yuan et al., 2015; Li et al., 2018; Ma et al., 2018; Bao et al., 2019). Localized and simultaneous cloud condition can be accurately acquired with high temporal and spatial resolution of ground-based observed clouds. Many ground-based cloud measurement devices, such as radar and lidar, are used to detect clouds (Zhao et al., 2014; Garrett and Zhao, 2013; Huang et al., 2012; Wang et al., 2017). Especially, ground-based all-sky-view imaging devices have been increasingly developed in the recent decades (Long et al., 2001; Genkova et al., 2004; Feister and Shields, 2005; Tapakis and Charalambides, 2013) because of their large field of view and low cost. Accurate cloud segmentation is a primary precondition for the cloud analysis of ground-based all-

sky-view imaging equipment, which can improve the precision of derived cloud cover information and help meteorologists further understand climatic conditions. Therefore, accurate cloud segmentation has become a topic of interest, and many algorithms have been recently proposed for the cloud analysis of ground-based all-sky-view imaging instrument (Long et al., 2006; Kreuter et al., 2009; Heinle et al., 2010; Liu et al., 2014; Liu et al., 2015).

5

Traditional segmentation methods generally use "color" as a distinguishing factor between clouds and clear sky because cloud particles have similar scattering intensity in blue and red bands due to the Mie scattering effort. By contrast, air molecules have more scattering intensity in the blue band than in the red band due to the Rayleigh scattering theory. Thus, the blue and red channel values of a cloud image are available for identifying features as cloud segmentation. Long et al. (2006) and Kreuter

10 et al. (2009) proposed a fixed-threshold algorithm using the ratio of red and blue channel values to identify clouds from whole-sky images. Particularly, pixels whose ratio of red and blue channel values are greater than the defined fixed threshold are identified as cloud, and as clear sky otherwise. Similarly, Heinle et al. (2010) treated the difference of red and blue channel values as a judgment condition to detect clouds. On the basis of red and blue channel values, Souzaecher et al. (2004) selected saturation as a complementary characteristic for cloud identification. These fixed-threshold algorithms, which strongly depend

15 on cameras' specifications and atmospheric conditions, are not adaptable for varied sky conditions (Long and Charles, 2010). Graph-cut method (Liu et al., 2015) and superpixel segmentation algorithm (Liu et al., 2014) are also applied for cloud segmentation to overcome the drawback of the aforementioned fixed-threshold algorithms. Although certain improvement can be achieved, the performance of such algorithms remains unsatisfactory in real measurement applications. Therefore, accurate and robust cloud segmentation algorithms must be well developed.

20

Convolutional neural networks (CNNs) are outstanding and powerful object recognition technologies, which have been widely applied in many fields, such as computer vision and pattern recognition (LeCun and Bengio, 1998; Taigman et al., 2014). CNNs have also achieved breakthrough progress in cloud analysis (Xiao et al., 2019; Shi et al., 2016; Shi et al., 2017; Liang et al., 2017) due to their strong capability in cloud feature representation and advanced cloud feature extraction for accurate

25 cloud identification (Lecun et al., 2015). Yuan et al. (2018) proposed an edge-aware CNN for satellite remote-sensing cloud image segmentation, which was proven to have superior detection results near cloud boundaries. Xiao et al. (2019) proposed an automatic classification model, namely, TL-ResNet152, to achieve accurate recognition of ice crystal in clouds. Zhang et al. (2018) proposed a CloudNet model for ground-based observed cloud categorization; the model could surpass the progress of other traditional approaches. However, few studies have evaluated the accuracy of CNNs in segmenting cloud images from

30 ground-based all-sky-view imaging instruments.

In this study, we present a CNN model named SegCloud for the accurate segmentation of cloud images based on ground-based observation. The architecture of the proposed SegCloud is straightforward and clear; it comprises an encoder network, a corresponding decoder network, and a final Softmax classifier. SegCloud is characterized by powerful cloud discrimination

and can automatically segment the obtained whole-sky images. It improves the accuracy of cloud segmentation and avoids misrecognition caused by traditionally color-based threshold methods. The SegCloud model is trained and tested by a database that consists of 400 whole-sky images and corresponding annotated labels. Extensive experimental results show that the proposed SegCloud model has effective and superior performance for cloud segmentation and has the advantage of recognizing

5   the area near the sun. Moreover, the local cloud cover calculated by SegCloud model has high correlation with human observation, which not only further proves the accuracy of SegCloud but also provides a practical reference for future automatic cloud cover observation.

The remainder of this paper is organized as follows. Section 2 describes the cloud segmentation database used in the experiment.

10   Section 3 introduces the SegCloud architecture. Section 4 presents the experimental details and results. Finally, conclusion and future work are depicted in Section 5.

2 Data description

Sufficient whole-sky images are necessary to evaluate cloud segment algorithms comprehensively. In this study, the database used by Tao et al. (2019) is applied to train and test the proposed SegCloud model. The database called Whole-Sky Image

15   SEGmentation (WSISEG) consists of 400 whole-sky images. These whole-sky images are captured by a ground-based all-sky-view camera (ASC). The appearance and functional specifications of ASC are shown in Figs. 1(a) and 1(c), respectively. The basic imaging component of ASC is a digital camera equipped with a fish-eye lens with 180° field view angle. The digital camera faces the sky directly to capture complete hemispheric sky images, and no traditional solar occulting devices, such as sun tracker or black shading strip, are required. High dynamic range technique is also applied to obtain clear whole-sky images

20   by fusing 10 photos with different (from low to high) exposure times to one image. Fig. 1(b) presents one whole-sky image sample, which has RGB color with a resolution of 2000×1944 pixels. WSISEG database collects whole-sky images that cover various cloud covers, times of day, and azimuth and elevation angles of the sun. Details about the database are shown in Fig. 2(c). The images are resized to 480×450 resolution to accommodate the input size of the CNN model, and their labels are manually created using photograph editing software. Different from the existing public database whose images are manually

25   cropped from whole-sky images and areas near or including sun and horizon are avoided, such as the SWIMSEG (Dev et al., 2017) and HYTA (Li et al., 2011) databases shown in Figs. 2(a) and 2(b), respectively, WSISEG database reflects a complete whole-sky condition. Thus, the whole-sky condition can be analyzed, and the performance of the segmentation algorithm in areas near the sun and horizon can be verified on this database. We divide the annotated images of WSISEG database into a training set (340 annotated images) and a test set (60 annotated images) and ensure that no image overlaps between the two.

30   Complex air condition images, such as cloud images under heavy haze and fog, are not included in the WSISEG database.

3 Cloud image segmentation approach

Detecting clouds from whole-sky images remains challenging for traditional cloud segmentation methods based on low-level visual information. CNNs can mine high-level cloud feature representation and are naturally considered as a novel choice for solving cloud segmentation problems. Thus, the SegCloud model is proposed and applied for whole-sky image segmentation. In this section, we initially describe an overall layout of the SegCloud model (Section 3.1) and then elaborate its training process (Section 3.2).

3.1 SegCloud architecture

SegCloud is an optimized CNN model that focuses on end-to-end cloud segmentation task. Fig. 3 shows that SegCloud consists of an encoder network, a corresponding decoder network, and a final Softmax classifier. SegCloud evolves from an improvement of VGG-16 network (Simonyan and Zisserman, 2014). VGG-16 network is one of the best CNN architectures for image classification, which achieved huge success on ImageNet Challenge in 2014 and has been applied in many other fields (Wang et al., 2016). In this study, we improve the VGG-16 network and propose our SegCloud model by replacing the fully connected layers of the original VGG-16 network with the decoder network to achieve end-to-end cloud image segmentation. SegCloud can take a batch of fixed-size whole-sky images as inputs. The encoder network transforms the input images to high-level cloud feature representation, whereas the decoder network enlarges the cloud feature maps extracted from the encoder network to the same resolution of input images. Finally, the outputs of the decoder network are fed to a Softmax classifier to classify each pixel and produce segmentation prediction.

3.1.1 Encoder network

The encoder network of SegCloud consists of 10 convolutional layers and 5 max-pooling layers. Each convolutional layer contains three operations, namely, convolution, batch normalization (Ioffe and Szegedy, 2015), and rectified linear units (ReLU) activation (Hinton, 2010). First, the convolution operation accepts the input feature maps and produces a set of output feature maps using a trainable filter bank with a 3×3 window size and a stride of 1. Batch normalization is then used to normalize these obtained feature maps to accelerate the convergence of the SegCloud model and alleviate the vanishing gradient problem during later training. The following ReLU activation is applied to achieve nonlinear transformation and add the expression capability of the SegCloud model. Generally, the shallow convolutional layers tend to capture the fine texture, such as shape and edge, whereas the deeper convolutional layers compute more high-level and complex semantic features using these obtained shallow layer features (Liang et al., 2017). The max-pooling layers are also an essential content of the encoder network. They are located separately after the convolutional layers and achieve increased translation invariance for robust cloud image segmentation. Each max-pooling layer subsamples the input feature maps with 2×2 window size and a stride of 2. Thus, the size of output feature maps is reduced by half, and many predominant features are extracted. High-level

and small-size cloud feature maps are finally formed for further pixel-wise cloud segmentation through 10 convolutional layers and 5 max-pooling layers.

3.1.2 Decoder network

5 The goal of the decoder network is to restore the obtained high-level cloud feature maps to the same resolution of input images and achieve end-to-end cloud image segmentation. The decoder network consists of 5 upsampling layers and 10 convolutional layers. Each upsampling layer upsamples input feature maps and produces sparse double-size feature maps, whereas each convolutional layer tends to densify these obtained sparse feature maps to produce many dense segmentation results. Upsampling layer uses feature information from the encoder network to perform upsampling and ensure segmentation accuracy

10 of whole-sky images. As shown by the black arrows in Fig. 3, the first four upsampling layers use pooling indices from the corresponding max-pooling layers to perform upsampling. The pooling indices are the locations of the maximum value in each max-pooling window of feature maps, which was proposed by Badrinarayanan et al. (2017) to ensure accurate feature restoration with less computational cost during inference. Thus, we introduce pooling indices into our first four upsampling layers and ensure effective restoration of high-level cloud feature maps.

15

Although pooling indices have advantage in computational time, they may lead to a slight loss of cloud boundary details. High-level semantic information and edge property play important roles in segmenting whole-sky images. Thus, the last upsampling layer directly use whole feature maps duplicated by the first max-pooling layers to perform upsampling and improve cloud boundary recognition. The specific operation is divided into three steps. (1) The upsampling layer initially uses bilinear

20 interpolation method to upsample the feature maps of the nearest convolutional layers and obtains output feature maps that are double in size. (2) The feature maps from the first max-pooling layer of the encoder network are duplicated. (3) The feature maps obtained from Steps (1) and (2) are concatenated to produce the final upsampling feature maps. In summary, high-level cloud feature maps are restored to the same resolution of input images step by step through the five upsampling layers.

3.1.3 Softmax classifier

25 The Softmax classifier is located after the decoder network to achieve final pixel-wise cloud classification (i.e., cloud image segmentation). Softmax classifier calculates the class probability of every pixel from feature maps through the following classification formula:

$$\text{softmax}(x_i) = \frac{exp(x_i)}{\sum_j exp(x_j)}, \tag{1}$$

30 where x_i and x_j are the feature values of classes i and j, respectively. The output of the Softmax classifier is a three-channel probability image, where 3 is the number of categories (cloud, sky, and background areas, such as sun and surrounding

building). The predicted results of each pixel are classes with maximum probability. Thus, the final segment results are outputted.

3.2 Training details of SegCloud model

5 SegCloud must be initially trained using the training set of the WSISEG database, such that it can be applied to cloud image segmentation. Before training the SegCloud, local contrast normalization is performed on the training set to further accelerate the training convergence of SegCloud. SegCloud is then trained on NVIDIA GeForce GTX1080 hardware and machine learning software package named TensorFlow. Mini-batch gradient descent is used as an optimization algorithm to find the appropriate model weights. During training, the number of whole-sky images fed to the SegCloud model per batch is 10, and 10 momentum parameter with a decay of 0.9 is used (Sutskever et al., 2013). SegCloud is trained in 26,000 epochs, with a learning rate of 0.006. It uses cross-entropy loss function defined in the following equation as the objective optimization function:

$$\text{loss} = -\frac{1}{N}\sum_{j=1}^{N} y_j^t \log y_j, \tag{2}$$

where N is the batch size, y_j^t is the ground truth images (i.e., the corresponding labels from the training set), and y_j is the predicated segmentation of the inputted whole-sky images. The calculated cross-entropy loss values are continuously 15 optimized using a backpropagation algorithm (Lecun et al., 2014) until the loss values converge. The final trained model and the best model parameters are saved for actual whole-sky image segmentation.

4 Experiment

Various experiments are designed and conducted to validate the performance of the proposed SegCloud model. First, vast segmentation experiments are performed to evaluate the effectiveness of the SegCloud. The superiority of the SegCloud is 20 demonstrated by comparing it with other traditional cloud segmentation methods. Finally, the accuracy and practicability of the SegCloud are proven by applying it to cloud cover estimation.

4.1 Effectiveness of SegCloud model

The well-trained SegCloud model is evaluated using the test set of the WSISEG database to verify its effectiveness in segmenting whole-sky images. A series of whole-sky images is fed into the SegCloud, and segmented images are outputted. 25 Some representative segmentation results are illustrated in the second row of Fig. 4. The SegCloud successfully segments the whole-sky images, as evident by the clouds, sky, and the sun in Fig. 4, which are colored white, blue, and black, respectively.

Furthermore, the effectiveness of the proposed SegCloud model is objectively quantified by calculating segmentation accuracy. The labels of the test set are set to the ground truths, and segmentation accuracy of the SegCloud is calculated by comparing

6

the segmentation result with the ground truth, as defined in Eq. (3). Average segmentation accuracy is also determined using Eq. (4).

$$Accuracy = \frac{T}{M} = \frac{T_{tcloud} + T_{tsky}}{M},$$ (3)

$$Average\_accuracy = \frac{1}{n} \sum_{i=1}^{n} Accuracy_i,$$ (4)

5   where T denotes the number of correctly classified pixels, which is the sum of true cloud pixels T_{tcloud} and true clear sky pixels T_{tsky}; M denotes the total number of pixels (excluding background regions) in the corresponding cloud image; and n is the number of test images. In this experiment, we have 60 test images: 10 for clear sky images, 10 for overcast sky images, and 40 for partial cloud images.

10   Table 1 reports that SegCloud achieves a high average accuracy of 96.24%, which further objectively proves its effectiveness. Moreover, SegCloud performs well on whole-sky images with different cloud cover conditions and achieves 96.98% accuracy on clear sky images, 95.26% accuracy on partial cloud images, and 99.44% near-perfect accuracy on overcast sky images. These experimental results show that the SegCloud is effective and accurate for cloud segmentation and can provide a reference for future cloud segmentation research.

15   **4.2 Comparison with other methods**

To verify its superiority, SegCloud is compared with two other conventional methods using the test set of the WSISEG database. (1) R/B threshold method: Given the camera parameters and atmospheric environment, the fixed ratio of red and blue channel values is set to 0.77 (Kreuter et al., 2009). (2) Otsu algorithm (the adaptive threshold method): The threshold is automatically calculated according to the whole-sky

20   images to be segmented (Otsu, 1979).

We use the proposed SegCloud model, R/B threshold method, and Otsu algorithm to segment the whole-sky images of the test dataset. Fig. 4 shows some representative cloud segmentation results. The Otsu algorithm has poor performance in segmenting whole-sky images because it requires pixels of the same class to have similar gray value, but clouds exhibit the opposite

25   behavior. The R/B threshold method has more accurate segmenting results compared with the Otsu algorithm. However, segmenting accuracy is poor especially for circumsolar areas (Figs. 4(a), 4(b) and 4(c)) because circumsolar areas often have texture and intensity similar to the clouds due to the forward scattering of visible light by aerosols/haze and dynamic range limitation of the detectors of the sky imager (Long and Charles, 2010). Thus, traditional methods merely utilize low-level vision cues to segment images and tend to result in misclassification once the boundary between clouds and sky is unclear.

30   However, SegCloud performs excellent segmentation compared with the two other conventional methods and exhibits segmentation advantages in the area near the sun. SegCloud learns from the given calibration database and constantly mines

deep cloud features. Thus, it has improved capability to identify circumsolar areas, although these pixels have textures and intensities that are similar to those of clouds.

The average segmentation accuracy of the three algorithms is also calculated on the test dataset to objectively verify the performance of the SegCloud. Table 1 shows that SegCloud obtains a higher average accuracy than the two other methods do and achieves better segmentation performance for clear sky, partial cloud, or overcast sky images. Although SegCloud shows advantages in whole-sky image segmentation, some misidentification remains due to decreased recognition for extremely thin clouds, which should be investigated in the future.

4.3 Application of SegCloud in cloud cover estimation

We apply the SegCloud model in cloud cover estimation and compare its derived cloud cover with human observation from 08:00 to 16:00 LT on July 1–31, 2018 to verify its accuracy and practicability. First, SegCloud derives cloud cover information by segmenting corresponding whole-sky images acquired by ASC100. For comparison, cloud covers of simultaneous human observation are provided by Anhui Air Traffic Management Bureau, Civil Aviation Administration of China, where the ASC is located. Well-trained human observers record cloud cover every hour during day and night by dividing the sky in oktas. Thus, cloud cover estimated by SegCloud is multiplied by 8 to be consistent with human observation.

A total of 279 pairs of cloud cover data are available, and the comparison results are shown in Table 2. The correlation coefficient between SegCloud and human observation is high (0.84). Statistically, the error (i.e., cloud cover difference between SegCloud and human observation) within ±1/8 oktas is 75.3%, and the error within ±2/8 oktas is 90.9% for all cases. These results clearly show that the proposed SegCloud model provides accurate cloud image segmentation and reliable cloud cover estimation. Moreover, the cloud cover derived by SegCloud and human observation demonstrate consistency in cases of clear sky (0/8 oktas), indicating that the SegCloud has better cloud segmentation performance for circumsolar areas compared with the other methods. These experimental results prove the accuracy and practicability of the SegCloud and show its practical significance for future automatic cloud cover observations.

5 Conclusion

In this paper, a deep CNN model named SegCloud is proposed and applied for accurate cloud image segmentation. Extensive segmentation results demonstrate that the SegCloud model is effective and accurate. Given its strong capability in cloud feature representation and advanced cloud feature extraction, the SegCloud outperforms the traditional methods significantly. In comparison with human observation, the SegCloud provides reliable cloud cover information by calculating the percentage of cloud pixel to all pixels, which further proves its accuracy and practicability. In our next work, certain challenges must be investigated. SegCloud must be further optimized to improve recognition accuracy in extremely thin clouds. On the basis of

cloud segmentation, other important cloud parameters, such as cloud base altitude estimation, cloud phase identification, and cloud movement, are also crucial for meteorological applications and must be elaborated in the future.

10   **References**

Badrinarayanan, V., Kendall, A., and Cipolla, R.: Segnet: a deep convolutional encoder-decoder architecture for scene segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1-1, 2017.

Bao, S., Letu, H., Zhao, C., Tana, G., Shang, H., Wang, T., Lige, B., Bao, Y., Purevjav, G., He, J., Zhao, J.: Spatiotemporal Distributions of Cloud Parameters and the Temperature Response Over the Mongolian Plateau During 2006–2015 Based

15   on MODIS Data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 2, pp. 549-558, https://doi.org/10.1109/JSTARS.2018.2857827, 2019.

Carslaw, and Ken: Atmospheric physics: cosmic rays, clouds and climate, Nature, 460(7253), 332-333, 2009.

Dev, S., Lee, Y. H., and Winkler, S.: 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, 2017.

20   Feister, U., and Shields, J.: Cloud and radiance measurements with the vis/nir daylight whole sky imager at lindenberg (germany), Meteorologische Zeitschrift, 14(5), 627-639, 2005.

Garrett, T. J. and Zhao, C.: Ground-based remote sensing of thin clouds in the Arctic, Atmos. Meas. Tech., 6, 1227–1243, https://doi.org/10.5194/amt-6-1227-2013, 2013.

Genkova, I., Long, C., Besnard, T., and Gillotay, D.: Assessing cloud spatial and vertical distribution with cloud infrared

25   radiometer cir-7, Proceedings of SPIE - The International Society for Optical Engineering, 5571, 1-10, 2004.

Heinle, A., Macke, A., and Srivastav, A.: Automatic cloud classification of whole sky images, Atmos. Meas. Tech., 3, 557-567, https://doi.org/10.5194/amt-3-557-2010, 2010.

Hinton, G. E.: Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair, International Conference on International Conference on Machine Learning. Omnipress, 2010.

30   Huang, D., Zhao, C., Dunn, M., Dong, X., Mace, G. G., Jensen, M. P., Xie, S., and Liu, Y.: An intercomparison of radar-based liquid cloud microphysics retrievals and implications for model evaluation studies, Atmos. Meas. Tech., 5, 1409–1424, https://doi.org/10.5194/amt-5-1409-2012, 2012.

Ioffe, S., and Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. International Conference on International Conference on Machine Learning JMLR.org, 2015.

Kreuter, A., Zangerl, M., Schwarzmann, M., and Blumthaler, M.: All-sky imaging: a simple, versatile system for atmospheric research, Appl Opt, 48(6), 1091-1097, 2009.

5   Lecun, Y., and Bengio, Y.: Convolutional networks for images, speech, and time series, the handbook of brain theory and neural networks, MIT Press, 1998.

Lecun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., and Hubbard, W., et al.: Backpropagation applied to handwritten zip code recognition, Neural Computation, 1(4), 541-551, 2014.

Lecun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521(7553), 436, 2015.

10   Li, J., Lv, Q., Jian, B., Zhang, M., Zhao, C., Fu, Q., Kawamoto, K., and Zhang, H.: The impact of atmospheric stability and wind shear on vertical cloud overlap over the Tibetan Plateau, Atmospheric Chemistry Physics, 18, 7329–7343, https://doi.org/10.5194/acp-18-7329-2018, 2018.

Li, Q., Lu, W., and Yang, J.: A hybrid thresholding algorithm for cloud detection on ground-based color images, Journal of Atmospheric and Oceanic Technology, 28(10), 1286-1296, 2011.

15   Liang, Y., Cao, Z., and Yang, X.: Deepcloud: ground-based cloud image categorization using deep convolutional features, IEEE Transactions on Geoscience and Remote Sensing, PP(99), 1-12, 2017.

Liu, S., Zhang, L., Zhang, Z., Wang, C., and Xiao, B.: Automatic cloud detection for all-sky images using superpixel segmentation, IEEE Geoscience and Remote Sensing Letters, 12(2), 354-358, 2014.

Liu, S., Zhang, Z., Xiao, B., and Cao, X.: Ground-based cloud detection using automatic graph cut, IEEE Geoscience and
20   Remote Sensing Letters, 12(6), 1342-1346, 2015.

Long, C., Slater, D., and Tooman, T.: Total sky imager model 880 status and testing results, Office of Scientific and Technical Information Technical Reports, 2001.

Long, C. N., Sabburg, J. M., Calbó, J., and Pagès, D.: Retrieving cloud characteristics from ground-based daytime color all-sky images, Journal of Atmospheric and Oceanic Technology, 23(5), 633-652, 2006.

25   Long, C. N., and Charles, N.: Correcting for circumsolar and near-horizon errors in sky cover retrievals from sky images, Open Atmospheric Science Journal, 4(1), 45-52, 2010.

Ma, Z., Liu, Q., Zhao, C., Shen, X., Wang, Y., Jiang, J. H., Li, Z., Yung, Y.: Application and evaluation of an explicit prognostic cloud-cover scheme in GRAPES global forecast system, Journal of Advances in Modeling Earth Systems, 10, 652– 667, https://doi.org/10.1002/2017MS001234, 2018.

30   Otsu, N.: A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62-66, 1979

Papin, C., Bouthemy, P., and Rochard, G.: Unsupervised segmentation of low clouds from infrared meteosat images based on a contextual spatio-temporal labeling approach, IEEE Transactions on Geoscience and Remote Sensing, 40(1), 104-114, 2002.

Rossow, W.B. and R.A. Schiffer.: ISCCP Cloud Data Products, Bulletin of the American Meteorological Society, 72, 2–20, 1991.

Shi, C., Wang, C., Yu, W., and Xiao, B.: Deep convolutional activations-based features for ground-based cloud classification, IEEE Geoscience and Remote Sensing Letters, PP(99), 1-5, 2017.

5 Shi, M., Xie, F., Zi, Y., and Yin, J.: Cloud detection of remote sensing images by deep learning, IEEE International Geoscience and Remote Sensing Symposium, 2016.

Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition. Computer Science, 2014.

Souzaecher, M. P., Pereira, E. B., Bins, L. S., and Andrade, M. A. R.: A simple method for the assessment of the cloud cover

10 state in high-latitude regions by a ground-based digital camera, Journal of Atmospheric and Oceanic Technology, 23(23), 437, 2004.

Stephens, and Graeme, L.: Cloud feedbacks in the climate system: a critical review, Journal of Climate, 18(2), 237-273, 2005.

Sutskever, I., Martens, J., Dahl, G., and Hinton, G.: On the importance of initialization and momentum in deep learning, International Conference on International Conference on Machine Learning. JMLR.org, 2013.

15 Taigman, Y., Ming, Y., Ranzato, M., and Wolf, L.: DeepFace: Closing the Gap to Human-Level Performance in Face Verification, IEEE Conference on Computer Vision and Pattern Recognition, 2014.

Tapakis, R., and Charalambides, A. G.: Equipment and methodologies for cloud detection and classification: a review, Solar Energy, 95(5), 392-430, 2013.

Tao Fa, Wanyi Xie, Yiren Wang, and Yingwei Xia.: Development of an all-sky imaging system for cloud cover assessment,

20 Applied Optics, 58, 5516-5524, 2019.

Wang D, Khosla A, Gargeya R, et al.: Deep Learning for Identifying Metastatic Breast Cancer [J]. 2016.

Wang, Y., and Zhao, C.: Can MODIS cloud fraction fully represent the diurnal and seasonal variations at DOE ARM SGP and Manus sites, Journal of Geophysical Research: Atmospheres, vol. 122, no. 1, pp. 329–343, Jan. 2017.

Wang, Y. , Zhao, C. , Dong, Z. , Li, Z. , Hu, S. , and Chen, T. , et al.: Improved retrieval of cloud base heights from ceilometer

25 using a non-standard instrument method, Atmospheric Research, 202, 148-155, 2018.

Xiao, H., Zhang, F., He, Q., Liu, P., Yan, F., Miao, L., and Yang, Z.: Classification of ice crystal habits observed from airborne Cloud Particle Imager by deep transfer learning, Earth and Space Science, 6, https://doi.org/10.1029/2019EA000636, 2019.

Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., Chi, W. C., and Ghonima, M., et al.: Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego, Solar Energy, 103(6), 502-524, 2014.

30 Yuan, F., Lee, Y. H., and Meng, Y. S.: Comparison of cloud models for propagation studies in Ka-band satellite applications, International Symposium on Antennas and Propagation, 2015.

Yuan, K., Yuan, K., Meng, G., Cheng, D., Bai, J., and Pan, X. C.: Efficient cloud detection in remote sensing images using edge-aware segmentation network and easy-to-hard training strategy, IEEE International Conference on Image Processing, 2018.

Zhang, J. L., Liu, P., Zhang, F., and Song, Q. Q.: CloudNet: Ground-based cloud classification with deep convolutional neural network, Geophysical Research Letters, 45, 8665–8672, https://doi.org/10.1029/2018GL077787, 2018.

Zhao, C., Chen, Y., Li, J., Letu H., Su, Y., Chen, T., Wu, X.: Fifteen-year statistical analysis of cloud characteristics over China using Terra and Aqua Moderate Resolution Imaging Spectroradiometer observations, International Journal of Climatology, 39: 2612– 2629, https://doi.org/10.1002/joc.5975, 2019.

Zhao, C., Wang, Y., Wang, Q., Li, Z., Wang, Z., and Liu, D.: A new cloud and aerosol layer detection method based on micropulse lidar measurements, Journal of Geophysical Research Atmospheres, 119, 6788– 6802, doi:10.1002/2014JD021760, 2014.

5
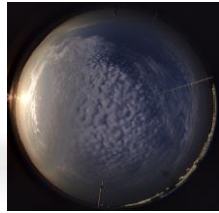
Functional specifications

Field of view	180°
Imaging period	10 min
Image resolution	2000×1944
Operating temperature	-45° – 55°
Dimension and weight	30×30×40 cm and 30kg
Power supply	AC220V/50Hz

(a) (b) (c)

Figure 1: (a) Apperance of ASC100. (b) Whole-sky images captured by ASC100. (c) Functional specifications of ASC100.
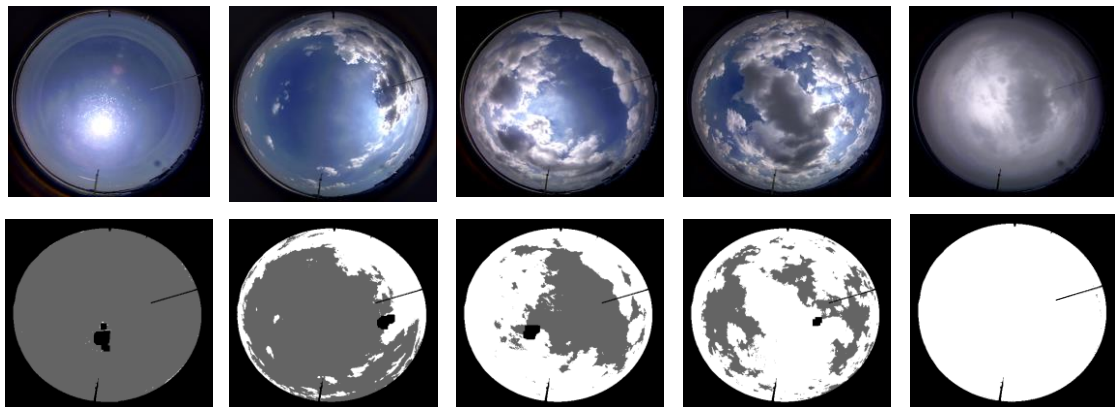
5


(a) SWIMSEG Database (b) HYTA Database

(c) WSISEG Database

13

Figure 2: Representative sky images and their corresponding labels from (a) SWIMSEG database, (b) HYTA database, and (c) WSISEG database. The labels of SWIMSEG and HYTA databases are binary images, where zero represents clear sky, and one represents cloud. The labels of the proposed WSISEG database, namely, clouds, sky, and undefined areas (including sun and backgrounds) are marked with gray values 255, 100, and 0, respectively. In comparison with SWIMSEG and HYTA databases, the proposed WSISEG database has more advantages to reflect the whole-sky condition.
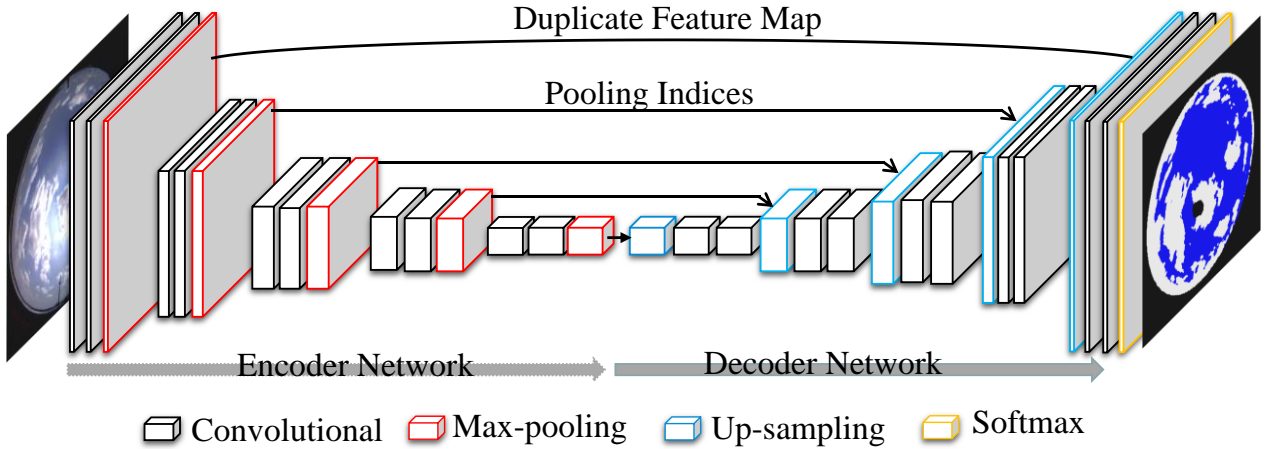
Figure 3: Illustration of the proposed SegCloud architecture. Overall, the networks contain an encoder network, a corresponding decoder network, and a final Softmax classifier.
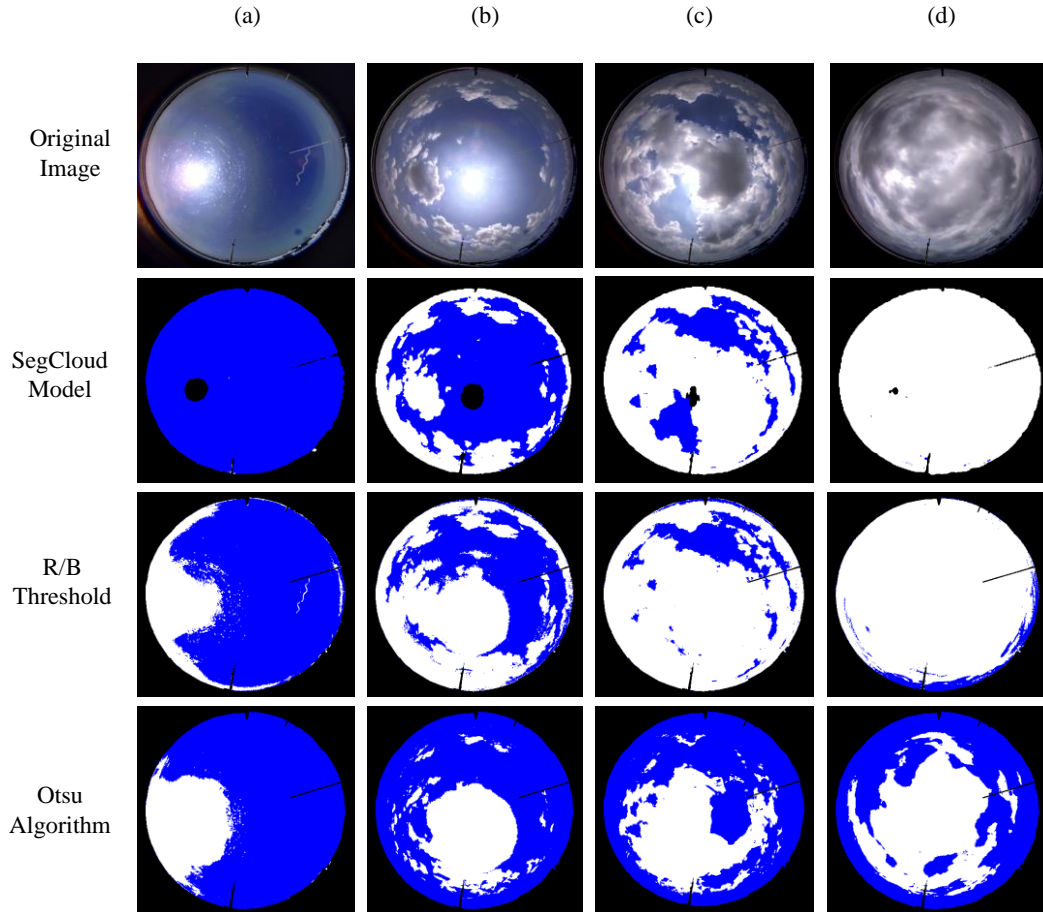
Figure 4: Examples of segmentation results with three algorithms. The original images are presented in the top line; and the segmentation results of SegCloud model, R/B threshold approach, and Otsu algorithm are presented in the second, third, and last lines, respectively. Clouds, sky, and sun are colored white, blue, and black, respectively. Masks are finally used in all result images to remove buildings around the circle for improved comparison of the sky and clouds.

Table 1: Segmentation accuracy (%) of three methods for clear sky, partial cloud, and overcast sky images.

Method	Clear sky	Partial cloud	Overcast sky	Average
Otsu	72.54	56.74	41.64	56.86
R/B threshold	70.63	81.36	90.95	81.17
SegCloud	**96.98**	**95.26**	**99.44**	**96.24**

15

Table 2: Comparison of derived cloud cover between SegCloud and human observation. The cloud cover provided by human observation is set to ground truth, and the error is defined as the cloud cover estimated by SegCloud minus the ground truth. The table presents the percentages of the error within ±1/8 oktas and the error within ±2/8 oktas. The correlation coefficient between SegCloud and human observation is also calculated.

Human observation (eighths)	Error within ±1/8 oktas (%)	Error within ±2/8 oktas (%)	Correlation
0	97.0	100	
1	91.6	100	
2	64.7	100	
3	52.6	94.7	
4	50.0	75.0	
5	68.3	85.4	
6	83.3	91.7	
7	90.9	95.5	
8	98.1	100	
All	**75.3**	**90.9**	**0.84**

5