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

Wanyi Xie1,2, Dong Liu1,2, Ming Yang1, Shaoqing Chen4, Benge Wang4, Zhenzhu Wang1,2, Yingwei Xia3, Yong Liu2,3, Yiren Wang1,2, and Chaofang Zhang3

1Key Laboratory of Atmospheric Optics, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230088, China
2Science Island Branch of Graduate School, University of Science and Technology of China, Hefei 230026, China
3Opto-Electronics Applied Technology Research Center, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China
4Anhui 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)

Received: 25 September 2019 – Discussion started: 27 November 2019
Revised: 12 March 2020 – Accepted: 16 March 2020 – Published: 17 April 2020

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 are 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, 2009; Stephens, 2005; Zhao et al., 2019; Wang and Zhao, 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 conditions can be accurately acquired with a 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 and Zhao, 2017). Especially, ground-based all-sky-view imaging devices have been increasingly developed in recent decades (Long et al., 2001; Genkova et al., 2004; Feister and Shields, 2005; Tapakis and Charalam-bides, 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 imag-
traditional approaches. However, few studies have evaluated the
accuracy of CNNs in segmenting cloud images from 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 ob-
tained whole-sky images. It improves the accuracy of cloud
segmentation and avoids misrecognition caused by tradition-
ally 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 segmenta-
tion and has the advantage of recognizing the area near
the sun. Moreover, the local cloud cover calculated by Seg-
Cloud model has a 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. Sec-
tion 2 describes the cloud segmentation database used in the
experiment. Section 3 introduces the SegCloud architecture.
Section 4 presents the experimental details and results. Fi-
ally, the conclusion and future work are depicted in Sect. 5.

2 Data description

Sufficient whole-sky images are necessary to evaluate cloud
segmentation 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 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 Fig. 1a
and 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 a sun tracker or black shading
strip, are required. The high-dynamic-range technique is also
applied to obtain clear whole-sky images by fusing 10 photos
with different (from low to high) exposure times to one im-
age. Figure 1b presents one whole-sky image sample, which
has RGB color with a resolution of 2000 pixels \( \times \) 1944 pix-
els. The 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. 2c. The images are resized to 480 \( \times \) 450 res-
olution to accommodate the input size of the CNN model,
and their labels are manually created using photograph edit-
W. Xie et al.: A novel cloud image segmentation model

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

Table 1. Functional specifications

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field of view</td>
<td>180°</td>
</tr>
<tr>
<td>Imaging period</td>
<td>10 min</td>
</tr>
<tr>
<td>Image resolution</td>
<td>2000×1944</td>
</tr>
<tr>
<td>Operating temperature</td>
<td>-45° – 55°</td>
</tr>
<tr>
<td>Dimension and weight</td>
<td>30×30×40 cm and 30kg</td>
</tr>
<tr>
<td>Power supply</td>
<td>AC220V/50Hz</td>
</tr>
</tbody>
</table>

Figure 2. Representative sky images and their corresponding labels from the (a) SWIMSEG database, (b) HYTA (hybrid thresholding algorithm) database, and (c) WSISEG database. The labels of the 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.

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 (Sect. 3.1) and then elaborate its training processes (Sect. 3.2).
3.1 SegCloud architecture

SegCloud is an optimized CNN model that focuses on end-to-end cloud segmentation task. Figure 3 shows that SegCloud consists of an encoder network, a corresponding decoder network, and a final Softmax classifier. SegCloud evolves from an improvement of the VGG-16 network (Simonyan and Zisserman, 2015). The VGG-16 network is one of the best CNN architectures for image classification, which achieved huge success on the 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 unit (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

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. The upsampling layer uses feature information from the encoder network to perform upsampling and ensure the segmentation accuracy 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 the effective restoration of high-level cloud feature maps.

Although pooling indices have an advantage in computational time, they may lead to a slight loss of cloud boundary details. High-level semantic information and edge prop-
softmax plays important roles in segmenting whole-sky images. Thus, the last upsampling layer directly uses 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 the bilinear interpolation method to upsample the feature maps of the nearest convolutional layers and obtain 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

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 function:

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_{j} \exp(x_j)}, \quad (1)$$

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 three 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

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 the 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 the 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 the cross-entropy loss function defined in the following equation as the objective optimization function:

$$\text{loss} = -\frac{1}{N} \sum_{j=1}^{N} y_j' \log y_j, \quad (2)$$

where $N$ is the batch size, $y_j'$ is the ground truth images (i.e., the corresponding labels from the training set), and $y_j$ is the predicted segmentation of the inputted whole-sky images. The calculated cross-entropy loss values are continuously 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 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 are fed into the SegCloud, and segmented images are outputted. 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 the segmentation result with the ground truth, as defined in Eq. (3). Average segmentation accuracy is also determined using Eq. (4).

$$\text{Accuracy} = \frac{T}{M} = \frac{T_{\text{cloud}} + T_{\text{sky}}}{M}, \quad (3)$$

$$\text{Average\_accuracy} = \frac{1}{n} \sum_{i=1}^{n} \text{Accuracy}_i, \quad (4)$$

where $T$ denotes the number of correctly classified pixels, which is the sum of true cloud pixels $T_{\text{cloud}}$ and true clear-sky pixels $T_{\text{sky}}$; $M$ 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.

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.

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**

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Clear sky</th>
<th>Partial cloud</th>
<th>Overcast sky</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otsu</td>
<td>72.54</td>
<td>56.74</td>
<td>41.64</td>
<td>56.86</td>
</tr>
<tr>
<td>R/B threshold</td>
<td>70.63</td>
<td>81.36</td>
<td>90.95</td>
<td>81.17</td>
</tr>
<tr>
<td>SegCloud</td>
<td>96.98</td>
<td>95.26</td>
<td>99.44</td>
<td>96.24</td>
</tr>
</tbody>
</table>

3. **R/B threshold approach**

4. **Otsu algorithm**

5. **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.

<table>
<thead>
<tr>
<th>Human observation (eighths)</th>
<th>Error within ±1/8 oktas (%)</th>
<th>Error within ±2/8 oktas (%)</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>97.0</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>91.6</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>64.7</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>52.6</td>
<td>94.7</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>50.0</td>
<td>75.0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>68.3</td>
<td>85.4</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>83.3</td>
<td>91.7</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>90.9</td>
<td>95.5</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>98.1</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>75.3</td>
<td>90.9</td>
<td>0.84</td>
</tr>
</tbody>
</table>
2. Otsu algorithm (the adaptive threshold method). The threshold is automatically calculated according to the whole-sky 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. Figure 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 a similar gray value, but clouds exhibit the opposite 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 (Fig. 4a, b and 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. 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 verify the performance of the SegCloud objectively. Table 1 shows that SegCloud obtains 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 a human observation from 08:00 to 16:00 LT on 1–31 July 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 the 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, the 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 \(\pm 1/8\) oktas is 75.3\%, and the error within \(\pm 2/8\) 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 demonstrates 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 Conclusions

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, SegCloud provides reliable cloud cover information by calculating the percentage of cloud pixels 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.

Data availability. Our database examples are available at https://github.com/CV-Application/WSISEG-Database (last access: 24 January 2019).

Author contributions. WX and YW designed the cloud image segmentation method. MY, SC, and BW prepared and provided the cloud cover data of human observation. DL, YL, YX, ZW, and CZ provided useful suggestions for the experiment and support for the writing. WX prepared the article with contributions from all the coauthors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We would like to thank Jun Wang and three anonymous reviewers for their helpful comments.
Financial support. This research has been supported by the Science and Technology Service Network Initiative of the Chinese Academy of Sciences (grant no. KFJ-STS-QYZD-022); the Youth Innovation Promotion Association, CAS (grant no. 2017482); and the Research on Key Technology of Short-Term Forecasting of Photovoltaic Power Generation Based on All-sky Cloud Parameters (grant no. 201904b11020031).

Review statement. This paper was edited by Jun Wang and reviewed by three anonymous referees.

References


W. Xie et al.: A novel cloud image segmentation model


