Postprint: All-Sky Ground-Based Cloud Image Classification Using Convolutional Neural Networks

Authors: Cui Shun, Xu Yunfei, Su Liying, Cui Chenzhou, Fan Dongwei, Han Jun, Wang Chuanzhong, Zhang Lei

Date: 2018-11-01T00:00:00+00:00

Abstract

All-sky ground-based cloud images captured by all-sky cameras can reflect real-time local cloud cover information, which constitutes one of the primary factors to be considered in astronomical site selection. Prior to cloud cover detection, automating the classification of all-sky ground-based cloud images based on factors such as image quality and application background, and implementing a robust and highly adaptable automated classification algorithm, will reduce manual costs in the image classification process and provide significant assistance for astronomical site selection. This paper implements an automated classification method for all-sky ground-based cloud images based on a convolutional neural network classification model. The classification model was trained using data from all-sky cameras onboard the Xue Long icebreaker and validated using data from the all-sky camera at the Lijiang Observatory of Yunnan Astronomical Observatory, Chinese Academy of Sciences, achieving favorable classification results in both cases and demonstrating that the method possesses good transferability.

Full Text

Preamble

Classification of All-Sky Ground-Based Cloud Images Using Convolutional Neural Networks

Cui Shun[12], Xu Yunfei[23], Su Liying[1], Cui Chenzhou[2], Fan Dongwei[2], Han Jun[2], Wang Chuanzhong[12], Zhang Lei[23]

[1] College of Mechanical Engineering and Applied Electronic Technology, Beijing University of Technology, Beijing 100124, China

[2] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China
[3] University of Chinese Academy of Sciences, Beijing 100049, China

Abstract

All-sky ground-based cloud images captured by all-sky cameras can reflect local cloud cover information in real time, making them a primary data source for automatic cloud amount detection—a critical factor in astronomical site selection. Automating the classification of these images based on image quality and application context can significantly reduce manual labor costs and improve the robustness and adaptability of cloud detection systems. This paper presents an automated classification method for all-sky ground-based cloud images based on convolutional neural network (CNN) models. The classification model was trained using data from an all-sky camera deployed on the XueLong polar research vessel and validated using data from the Lijiang Observatory of the Chinese Academy of Sciences. The method achieved favorable classification results and demonstrated strong transferability across different datasets.

Keywords: astronomical site selection; cloud image classification; pattern recognition; convolutional neural network

1. Research Background and Significance

Cloud amount is one of the foremost considerations in astronomical site selection. Light from astronomical targets is scattered and absorbed by atmospheric clouds before reaching ground-based telescopes, and the degree of cloud cover directly determines observation data quality, usable observation time, and visible sky coverage. However, current cloud observations remain predominantly manual, with results heavily influenced by subjective human factors. Consequently, implementing automatic cloud detection is of paramount importance.

All-sky cameras serve as the primary instruments for automatic cloud detection. Comprising a charge-coupled device (CCD) and fisheye lens, these cameras capture all-sky ground-based cloud images with high spatial and temporal resolution. Their analysis results accurately reflect local cloud cover characteristics and variations, making them well-suited for all-sky cloud monitoring.

During image acquisition, all-sky cameras are inevitably affected by temporal, meteorological, and illumination factors, leading to significant variations in image quality. Moreover, cloud detection methods differ for daytime versus nighttime images. Therefore, prior to cloud amount detection, it is essential to remove unreliable images (those affected by rain, dew, or snow, as well as overexposed images) and distinguish image scenes (day or night) to enable automated statistics of clear days and clear nights. This preprocessing forms the foundation of automated cloud analysis.

Addressing these requirements, this paper establishes a detailed classification scheme for all-sky ground-based cloud images based on image quality and application scenarios. Using a dataset from an all-sky camera deployed on the XueLong vessel, we trained a classification model based on convolutional neural networks (CNNs) and propose an automated processing method for all-sky ground-based cloud images. Validation using data from the Lijiang Astronomical Observatory of the Chinese Academy of Sciences yielded excellent classification results, demonstrating strong generalization performance and applicability for astronomical site selection.

This paper first elaborates on the background and significance of all-sky cloud image classification. Section 2 introduces the dataset and classification criteria used for model training. Section 3 provides a detailed description of existing CNN models and the classification model employed in this study. Section 4 describes the model training process. Section 5 presents validation results using Lijiang data with detailed analysis. Section 6 concludes with a summary and future outlook, analyzing the model’ s strengths and weaknesses and proposing directions for improvement.

2. Dataset and Classification Criteria

All-sky ground-based cloud images exhibit complex diversity and varying quality. Utilizing large, representative training datasets enhances model generalization capability, enabling robust automated classification across different regions and supporting the implementation of automated cloud detection systems for astronomical site selection.

2.1 XueLong Polar Research Vessel All-Sky Camera Data

The XueLong is China’ s specialized polar comprehensive research vessel. On July 20, 2017, XueLong departed from Shanghai, entered the Arctic Circle on July 31, and returned to the Yangtze River estuary on October 9, completing its 8th Arctic scientific expedition after traveling 20,590 nautical miles. The all-sky camera onboard collected data every five minutes, acquiring 89.8 GB of data comprising 13,484 images with a resolution of 1000$×$625 pixels.

The voyage covered extensive regions with complex environmental factors, yielding rich and comprehensive all-sky image data suitable for model training.

2.2 Classification Criteria

Based on application requirements, cloud images are classified into six categories: **dark** (nighttime), **rain_{snow}** (affected by precipitation), **overlight** (overexposed), **bluesky** (clear sky), **cloudimage** (partly cloudy), and **overcast** (heavily clouded).

Day and night conditions impart different characteristics to images, necessitating distinct research approaches. For daytime all-sky images, cloud points are typically identified using visible light properties, while nighttime images require comparison between observed star fields and calculated star maps to identify cloud-obscured bright stars. Therefore, separating daytime and nighttime images facilitates subsequent scientific analysis, with all nighttime images classified as a single category, as illustrated in [Figure 1: see original paper].

Daytime images are further subdivided. During acquisition, lenses are often affected by adverse weather conditions such as rain, dew, or snow coverage, which severely degrades image quality and renders these images unsuitable for cloud detection. Consequently, images affected by precipitation are classified as a separate category, as shown in [Figure 2: see original paper].

At noon, abundant sunlight often causes overexposure. Classifying these images separately helps identify more valuable data for research. Images are categorized as overlight when the overexposed area exceeds one-third of the total image area, with typical examples shown in [Figure 3: see original paper].

After removing lower-quality images, the remaining images are classified into three categories based on cloud amount: **bluesky** (cloud-free), **overcast** (sky nearly completely covered by clouds), and **cloudimage** (all other cases). Some images exhibit overexposure near the sun position, making it difficult to distinguish between solar glare and clouds. In such cases, the region is identified based on capture time and region shape.

Representative images for each category are shown in [Figure 4: see original paper] (bluesky), [Figure 5: see original paper] (overcast), and [Figure 6: see original paper] (cloudimage). A total of 6,160 images were manually classified and annotated with the following distribution: overlight (1,124), overcast (1,128), rain_{snow} (1,117), cloudimage (1,115), dark (1,119), and bluesky (557).

2.3 Lijiang Astronomical Observatory All-Sky Camera Data

The Lijiang Observatory of the Yunnan Astronomical Observatories, Chinese Academy of Sciences, is an important astronomical observation base in southern China. The observatory houses multiple telescopes (2.4m, 1.8m, BOOTES-4, and TAT), with the all-sky camera providing real-time cloud information for nighttime observations and playing an essential auxiliary role. This study uses Lijiang all-sky camera data to validate the classification model.

The dataset comprises 5,208 all-sky cloud images captured between November 26 and December 26, 2016, with a resolution of 720$×$480 pixels. Following the classification criteria described in Section 2.2, the Lijiang data were manually classified to evaluate model generalization capability on different all-sky ground-based cloud image datasets.

3. Classification Model Introduction

Convolutional neural networks offer significant advantages in image recognition, using pixel values directly as input to extract the most effective features through training. They exhibit invariance to scaling, translation, and rotation distortions while achieving excellent generalization. The weight-sharing structure of convolutions substantially reduces network parameters, preventing overfitting and reducing model complexity. This study adapts and improves the classic LeNet5 and AlexNet models for all-sky camera data classification.

3.1 LeNet5 and AlexNet

LeNet5, designed by Yann LeCun in 1998 for handwritten digit recognition, is one of the earliest CNN models. Its features include: each convolutional layer comprising convolution, pooling, and nonlinear activation; spatial feature extraction via convolution; dimensionality reduction through average pooling; use of hyperbolic tangent or sigmoid activation functions; an MLP classifier; and sparse connections between layers to reduce computational complexity. Many of these characteristics remain fundamental in modern CNNs.

AlexNet, developed by Hinton and Alex Krizhevsky, extended LeNet's principles to a deeper and wider network, winning the 2012 ImageNet competition by a significant margin. AlexNet successfully employed ReLU activation, which outperformed sigmoid functions in deeper networks. Dropout mitigated overfitting, while overlapping max pooling avoided the blurring effects of average pooling. The model used pooling kernels smaller than the stride to enhance feature richness and performed random cropping from 256$×256*imagesto*224×$224 regions, maximizing data volume and improving generalization.

3.2 Model Architecture and Training

This study combines and optimizes the LeNet5 and AlexNet architectures with modified parameters, as illustrated in [Figure 7: see original paper]. The network comprises two convolutional layers with 5$×5*kernels*, *eachfollowedbya*3×$3 max pooling layer with stride 2, and two fully connected layers. A Softmax layer determines the final classification. Dropout is applied in the fully connected layers to reduce overfitting, and ReLU activation functions follow the convolutional layers.

4. Classification Model Training

Model training aims to obtain optimal network weights by fitting the training dataset to enable accurate classification of future data. The process 主要包括 training data preprocessing, fitting computation, and model performance evaluation.

4.1 Data Preprocessing

Before training, data must be preprocessed to meet model input requirements while preserving maximum effective information. Preprocessing includes four components: data splitting, region of interest (ROI) extraction, data augmentation, and standardization.

4.1.1 Training and Test Set Division Among the 6,160 manually annotated images, the bluesky class contains only 557 images, while other classes exceed 1,000 images each. Since large training sets are crucial for preventing overfitting, we selected 500 bluesky images and 1,000 images from each remaining class for training, with the remainder serving as the test set. Training and test sets contain no overlapping images; test data are used exclusively for evaluating classification accuracy on unseen images.

4.1.2 Region of Interest (ROI) Extraction All original images at 1000$×625resolution([Figure8 : seeoriginalpaper]left)werecenter − croppedtoremoveperipheralirrelevantinformation, resultingin800×$625 images ([Figure 8: see original paper] right).

4.1.3 Data Augmentation To prevent overfitting, sufficient data volume is required. When data are limited, augmentation techniques can geometrically transform images while preserving essential features. Common operations include flipping, scaling, translation, rotation, contrast adjustment, noise perturbation, and color transformation.

Based on data characteristics, we applied the following augmentation: training images were cropped at five positions (lower-right, upper-right, center, lower-left, upper-left) to extract 700$×$550 regions, then horizontally flipped, expanding the dataset tenfold ([Figure 9: see original paper]). For the smaller bluesky class, vertical flipping was additionally applied. This yielded 10,000 images per class, balancing sample distribution.

Images were then resized to 128$×128toreducecomputationalburden.Randomcroppingto100×100furtherexpan
100)^{2}$ = 784 times. Finally, random vertical flipping and adjustments to contrast, brightness, saturation, and hue were applied. Experiments demonstrate that these augmentation techniques significantly mitigate overfitting and improve generalization. Notably, fixed objects from the XueLong vessel that appear in images become noise after flipping and translation, further enhancing model robustness.

Test set preprocessing involved center-cropping 700$×550regions, resizingto128×128, andcenter−croppingto100×$100, ensuring test images retain sufficient information while matching training input dimensions.

4.1.4 Standardization Standardizing images before network input reduces redundancy and decreases sensitivity to dynamic range variations, as shown in formula (1):

$$img_{out} = \frac{img_{in} - mean}{adjusted\_stddev}$$

where img_{in} represents RGB pixel values, $mean$ denotes per-channel means, and $adjusted\_stddev$ is defined in formula (2):

$$adjusted\_stddev = \max\left(stddev, \frac{1}{\sqrt{NumElements}}\right)$$

where $stddev$ is the per-channel standard deviation and $NumElements$ is the number of pixels per channel. Both training and test sets undergo this standardization before model input.

4.2 Experimental Results and Analysis

During training, loss values decreased while accuracy increased with successive epochs ([Figure 10: see original paper]). With appropriate batch size and learning rate, the model exhibited stable loss reduction, converging noticeably after 50 epochs. By epoch 100, loss approached 0.15 with minimal further decrease. Training and validation accuracy trends were similar, with validation accuracy stabilizing after 50 epochs. Training accuracy fluctuated between 95% and 100% due to random contrast and brightness transformations, indicating incomplete fitting to all possible image variations at each iteration. After 220 epochs, validation accuracy stabilized at 95.5%, confirming model convergence.

Given multiple classes with imbalanced samples, we evaluated performance using Precision, Recall, and F1-Score. Precision measures the proportion of true positives among positive predictions (TP/(TP+FP)). Recall measures the proportion of actual positives correctly identified (TP/(TP+FN)). F1-Score provides a harmonic mean of precision and recall (2TP/(2TP+FP+FN)). Per-class metrics are shown in [Figure 11: see original paper], demonstrating excellent performance with most categories exceeding 95% precision, recall, and F1-score.

5.1 Validation Process

To evaluate model generalization, we validated classification performance using the Lijiang Observatory dataset. Preprocessing ensured validation data dimensions matched training data to maintain classification accuracy. The 5,208 Lijiang images were manually annotated as: dark (4,257), bluesky (549), overlight (519), cloudimage (321), overcast (115), and rain_{snow} (46).

Original 720$×480*images*([*Figure*12 : *seeoriginalpaper*]*left*)*werecenter* − *croppedto*400×320*toobtaineffectiveinformationwithaspectratiosimilartotrainingdata.Afterresizingto*128×
*croppingto*100×$100, images were standardized as described in Section 4.1.4 before model input.

5.2 Prediction Results and Analysis

Precision, recall, and F1-Score for each validation class are shown in [Figure 13: see original paper]. The model performed well on the new dataset, accurately classifying the predominant dark class and effectively distinguishing among cloudimage, bluesky, overlight, and overcast categories. However, rain_{snow} classification performed poorly, with relatively low recall for bluesky and overcast, and lower precision for cloudimage.

The confusion matrix ([Figure 14: see original paper]) reveals that most rain_{snow} images were misclassified as overlight. Misclassified examples ([Figure 15: see original paper]) show sparse raindrops with severe overexposure, suggesting the model prioritized overexposure features. Additionally, the validation set contained fewer raindrops than the training set, contributing to lower rain_{snow} accuracy.

Bluesky recall and cloudimage precision suffered primarily because some bluesky images were misclassified as cloudimage. Representative misclassifications ([Figure 16: see original paper]) show images darker than training bluesky samples with noticeable artifacts, likely causing the errors.

Overcast recall was also low, with most misclassifications as overlight. Misclassified examples ([Figure 17: see original paper]) exhibit high brightness similar to training overlight images, with differences attributable to varying camera parameters.

6. Summary and Outlook

This paper proposes a CNN-based automated classification method for all-sky ground-based cloud images applicable to astronomical site selection. Training on XueLong data and validation on Lijiang data demonstrated strong generalization and transferability across different acquisition equipment, locations, and image scales. This method enables automated classification of ground-based cloud images from any region, substantially reducing manual preprocessing effort and labor costs in astronomical site selection.

However, hardware variations, camera parameters, meteorological conditions, and environmental factors create significant image quality differences. Although XueLong data are diverse and comprehensive, validation results indicate the training set does not encompass all possible data types, leading to misclassifications such as rain_{snow} as overlight and bluesky as cloudimage.

Future work should employ transfer learning with more diverse data to improve classification accuracy and enhance performance across various all-sky ground-based cloud image datasets.

Acknowledgments

This work received data resources and technical support from the Chinese Virtual Observatory and the Chinese Astronomical Data Center, funded by the National Science and Technology Infrastructure Platform projects “Earth System Science Data Sharing Service Platform” and “National Basic Science Data

Sharing Service Platform" (DKA2017-12-02-07) from the Ministry of Science and Technology. We thank the National Astronomical Observatories-Alibaba Cloud Astronomy Big Data Joint Research Center for their support, and the Lijiang Observatory of the Yunnan Astronomical Observatories for providing data resources.

References

[1] Shi Y, Yao Y Q, Liu L Y. A Method for Observing and Counting Day-time Cloud Amounts in an Astronomical Site Survey[J]. Astronomical Research & Technology, 2008, 5(4):415-419.

[2] Peng H, Xin Y, He S, et al. An Introduction to the All-Sky Camera at the YNAO Lijiang Astronomical Station[J]. Astronomical Research & Technology, 2015, 12(1):000089-95.

[3] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436.

[4] YANG Jun, LU Weitao, MA Ying, et al. An automatic ground-based cloud detection method based on the local threshold interpolation[J]. Acta Meteorologica Sinica, 2010, 68(6):1007-1017.

[5] Sui-Ping Q I, Liu T, Tong H U, et al. Ground cloud observation technology and its instrument research advances[J]. Shandong Science, 2014, 27(6):1-9.

[6] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[7] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.

[8] Bishop C. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer[M]// Stat Sci. 2006:140-155.

Note: Figure translations are in progress. See original paper for figures.