

RESEARCH

Open Access



Deep multimodal fusion for ground-based cloud classification in weather station networks

Shuang Liu^{1,2*} and Mei Li^{1,2}

Abstract

Most existing methods only utilize the visual sensors for ground-based cloud classification, which neglects other important characteristics of cloud. In this paper, we utilize the multimodal information collected from weather station networks for ground-based cloud classification and propose a novel method named deep multimodal fusion (DMF). In order to learn the visual features, we train a convolutional neural network (CNN) model to obtain the sum convolutional map (SCM) by using a pooling operation across all the feature maps in deep layers. Afterwards, we employ a weighted strategy to integrate the visual features with multimodal features. We validate the effectiveness of the proposed DMF on the multimodal ground-based cloud (MGC) dataset, and the experimental results demonstrate the proposed DMF achieves better results than the state-of-the-art methods.

Keywords: Ground-based cloud, Multimodal information, Convolutional neural networks, Weather station networks

1 Introduction

Clouds, as one of the major meteorological phenomena, play a profound role in climate predictions and services [1, 2]. Cloud classification is a crucial task for cloud observation, and it is currently undertaken by the professional observers [3]. However, manual observation is time-consuming and labor-intensive. Furthermore, the observation results are unreliable due to large dependency on subjective judgements. Hence, there is a high demand for automatic classification of ground-based cloud.

In recent years, many attempts have been made to classify the ground-based cloud. One trend is to develop the ground-based sky imagers such as whole-sky imager (WSI) [4], total-sky imager (TSI) [5], infrared cloud imager (ICI) [6], all-sky imager (ASI) [7, 8], whole-sky infrared cloud measuring system (WSIRCMS) [9], and day/night whole sky imagers (D/N WSIs) [10]. Benefiting from these devices, a number of ground-based cloud images are available for developing automatic classification algorithms. Calbo et al. [1] extracted statistical

texture features based on Fourier transform to classify the cloud images into eight categories. Heinle et al. [11] distinguished seven sky conditions based on twelve kinds of features and k -nearest neighbor classifier. Ghonima et al. [12] treated the pixel red-blue ratio (RBR) between the test image and clear sky image as the feature. Zhuo et al. [13] combined texture and structure features for cloud representation and obtained a high classification accuracy. Kazantzidis et al. [14] took into account the static color, the solar zenith angle, and the existence of raindrops in sky images. Cheng et al. [15] divided cloud images into several blocks and conducted the classification task on blocks. Xiao et al. [16] fused the texture, structure, and color features as the multi-view cloud visual features.

It is observed that appearance of clouds can be treated as a kind of natural texture. Therefore, it is reasonable to describe cloud appearances using texture and image descriptors. Sun et al. [17] utilized the local binary pattern (LBP) to classify the cloud images into five predefined types. Liu et al. [18–20] proposed several algorithms for extracting texture and image descriptors, such as the multiple random projections, the salient local binary pattern, and the group pattern learning. Recently, convolutional neural networks (CNNs) have shown remarkable

*Correspondence: shuangliu.tjnu@gmail.com

¹Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin, China

²College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin, China

performances in several fields, such as visual classification [21], object detection [22], and speech recognition [23]. The success of CNNs is attributed to their ability to learn rich representations. CNNs could achieve some degree of shift and deformation invariance by using local receptive fields, shared weights, and spatial sub-sampling. In particular, the shared weights are supportive for improving the generalization of CNNs, because they reduce the number of parameters of CNNs. Several researchers have resorted to training CNNs using cloud images for ground-based cloud classification. For example, Ye et al. [24] utilized cloud visual features from the convolutional layers of the network. Afterward, they employed Fisher vector encoding to further improve the cloud classification results. Shi et al. [25] extracted visual features from both the shallow and deep convolutional layers of the network. They also evaluated the performance of fully connected (FC) layer for cloud classification.

However, it is difficult to solve the problems of ground-based cloud classification by using one kind of sensor, i.e., image sensors. This is because the cloud type is determined by many factors, such as temperature, humidity, pressure, and wind speed. We are inspired by the recent development in weather station networks which is a kind of wireless sensor network (WSN) [26–28], and we consider to classify the ground-based cloud by using weather station networks. The weather station networks consist of many kinds of sensors [29], for example, image sensors, thermal sensors, moisture sensors, and wind speed sensors. These sensors have the abilities to obtain multimodal information of clouds. The visual, thermal, moisture, and wind speed information could collect more complete information for ground-based cloud information with the

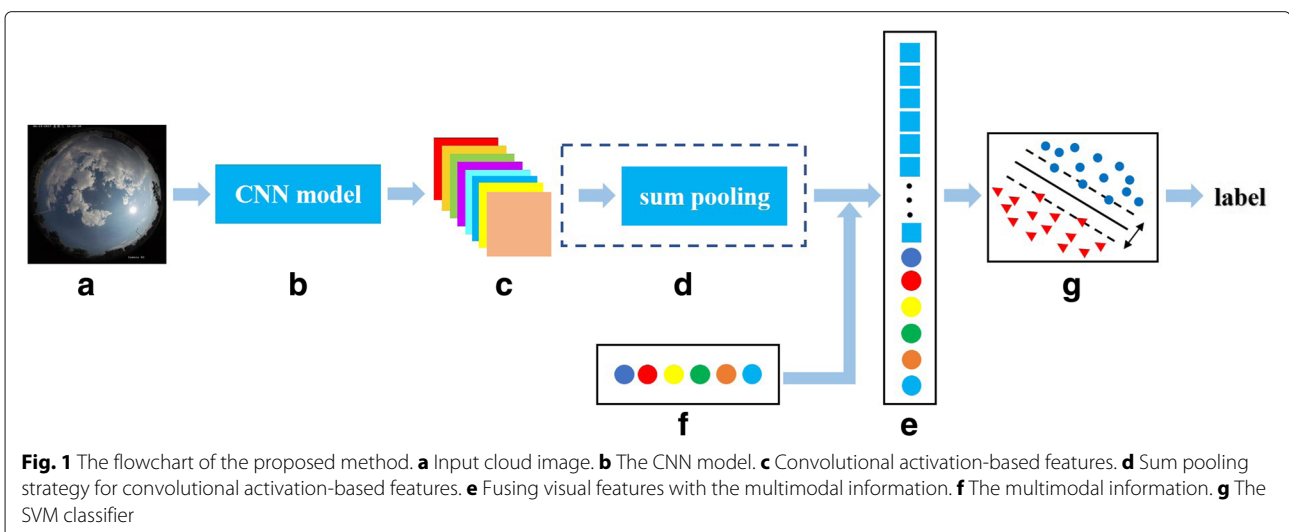
help of weather station networks, so the limitation of each kind of information could be compensated.

In this paper, we propose a novel method named deep multimodal fusion (DMF) for ground-based cloud classification in weather station networks. Concretely, we first fine-tune the pre-trained CNN model to adjust the parameters for visual cloud information. Then, the visual features are extracted from convolutional layers. Different from the other activation-based features, we apply a pooling strategy across all the feature maps to reserve the spatial information of cloud images. Furthermore, we also evaluate the performance of FC features. After obtaining the visual information of clouds, we fuse the multimodal information collected from weather station networks, e.g., temperature, humidity, pressure, and wind speed, into the final representations. This fusion strategy learns the complementary information between visual and multimodal features, which could further improve the performance. Finally, support vector machine (SVM) [30] is selected as the classifier.

The rest of this paper is organized as follows. Section 2 introduces the proposed approach in details, Section 3 illustrates the experimental results, and Section 4 draws the conclusion for this paper.

2 Method

In this section, we present the proposed method in detail, and the flowchart is illustrated in Fig. 1. The cloud images are first utilized to train a CNN model, and then the sum pooling is applied to aggregate the feature maps of one convolutional layer so as to obtain the visual features. Afterwards, the visual features and multimodal cloud features are integrated. Finally, we utilize the SVM to train the classification model.



2.1 Deep convolutional neural networks

In recent years, CNNs have achieved great success in classification task [21] due to large-scale databases and efficient computational abilities. Simonyan and Zisserman [31] proposed very deep neural networks which consist of more convolutional and pooling layers. The deep CNNs not only obtain outstanding performance on imagenet large-scale visual recognition challenge (ILSVRC), but also show promising performance on other classification tasks. Hence, we employ a deep CNN to extract features for cloud images.

Training a CNN model needs a very large number of annotated images to learn millions of parameters [32, 33]. Thus, it is running the risk of overfitting to train a CNN model from scratch by only utilizing a few thousand cloud images. To overcome this drawback, we first fine-tune a deep CNN model named imagenet-vgg-f [34] to transfer the cloud information to the deep model. The imagenet-vgg-f comprises five convolutional layers and three fully connected layers, and the detail configuration is shown in Table 1. The last FC layer has 1000 dimensionality, and we replace it with a new one with N dimensionality to start the fine-tuning procedure, where N is the number of cloud classes. Meanwhile, for the bias, all the parameters in the new FC layer are initialized to zero. For the weight, the parameters obey a Gaussian distribution. In addition, three max pooling layers follow the first, second, and fifth convolutional layer, respectively, with the size of 3×3 in conjunction with a downsampling factor 2. Moreover, two local response normalization layers are after the first two convolutional layers, respectively.

Table 1 The configuration of imagenet-vgg-f. conv i denotes the i -th convolutional layer

Config.	Receptive fields	Stride	Padding	Filter banks
conv1	11×11	4	0	64
Max pooling				
conv2	5×5	1	2	256
Max pooling				
conv3	3×3	1	1	256
conv4	3×3	1	1	256
conv5	3×3	1	1	256
Max pooling				
Neurons				
fc6	4096			
fc7	4096			
fc8	1000			

2.2 Deep features for cloud images

The appearance of clouds can be treated as a kind of natural texture, and therefore, it is rational to describe cloud appearance using texture descriptors. The CNN models have been applied to capture texture information and have achieved promising results [35, 36]. The features extracted from deeper layers possess several desirable properties such as invariance and discrimination. On the contrary, the shallower layers tend to be more sensitive to small transformations, which is challenging for unpredictable and changeful cloud.

Based on the analysis mentioned above, we adopt deeper layers to extract the cloud features. For the convolutional layer, we aggregate the raw activations by sum pooling and then obtain the sum convolutional map (SCM). The activation value y_{ij} in SCM at position (i, j) is defined as

$$y_{ij} = \sum_{k=1}^C x_{ij}^k, \quad (1)$$

where x_{ij}^k is the activation at position (i, j) in the k -th feature map and C is the number of feature maps in the convolutional layer. Suppose the size of each feature map is $H \times W$ and the SCM is also with the size of $H \times W$. The SCM preserves the spatial information because the pooling operation is conducted across all the feature maps, while the other traditional pooling operations aggregate one feature map into a feature. As a result, the convolutional activation-based features V_{conv} for each image is acquired by transforming the SCM into a vector

$$V_{conv} = [y_{11}, y_{21}, \dots, y_{H1}, y_{12}, y_{22}, \dots, y_{H2}, \dots, \dots, y_{1W}, y_{2W}, \dots, y_{HW}]^T. \quad (2)$$

The dimensionality of the vector is $H \times W$. The above procedure is summarized in Fig. 2. On the other hand, we do not use any pooling strategies in FC layers. The FC layer-based features could be considered as a special case of convolutional layers. It utilizes rather smaller filter banks with the size of 1×1 . The feature vector V_{fc} for a FC layer is indicated as

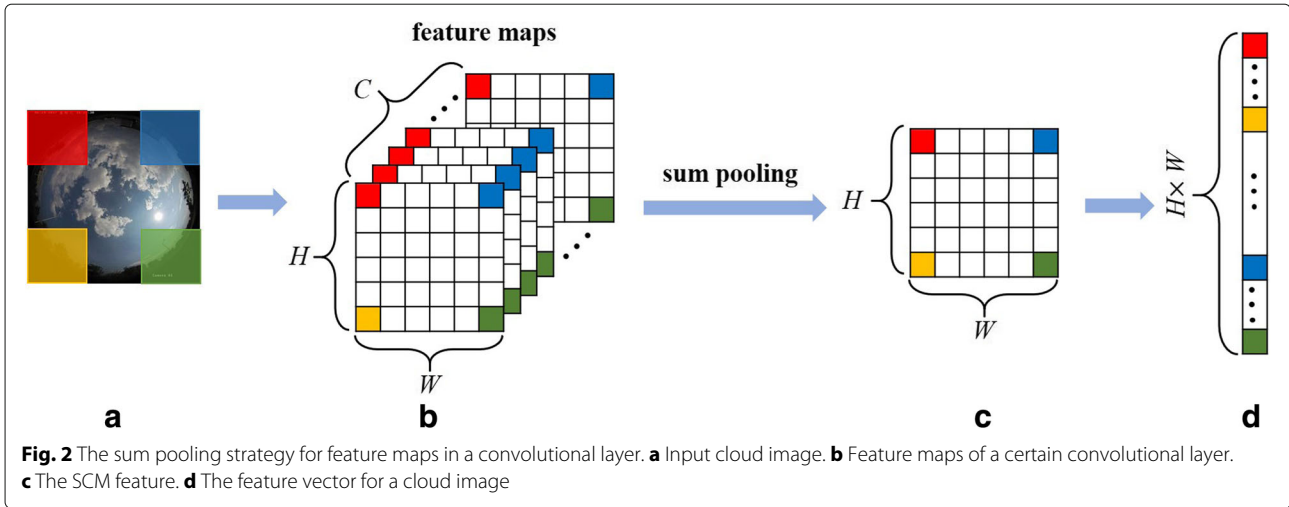
$$V_{fc} = [v_1, v_2, \dots, v_k, \dots, v_K]^T, \quad (3)$$

where v_k is the output of the k -th neuron and K is the number of neurons in the FC layer. Finally, V_{conv} and V_{fc} are normalized by L_2 -norm.

2.3 Multimodal fusion

To capture the complete cloud information, we integrate multimodal cloud information collected from weather station networks. The integration features can be formulated as

$$Q = f(V, M), \quad (4)$$



where V is the visual feature vector, i.e., activations-based or FC-based feature vector, and $M = [m_1, m_2, \dots, m_p]^T$ denotes the multimodal feature vector. For simplicity and efficiency, we directly concatenate the visual feature vector with multimodal feature vector

$$f(V, M) = [\alpha V^T, \beta M^T] \quad (5)$$

where $[\cdot, \cdot]$ indicates to concatenate two vectors, and α and β are the parameters to balance the importance between visual features and multimodal features. Note that the multimodal information M should be normalized by L_2 -norm before fusion.

3 Experimental results

In this section, we conduct a series of experiments on the multimodal ground-based cloud (MGC) dataset to evaluate the effectiveness of the proposed DMF. We first introduce the MGC dataset and the implementation details of experiments. Then, we compare the proposed DMF with the other methods. Finally, we evaluate the influence of visual features extracted from different layers.

3.1 Dataset and experimental setup

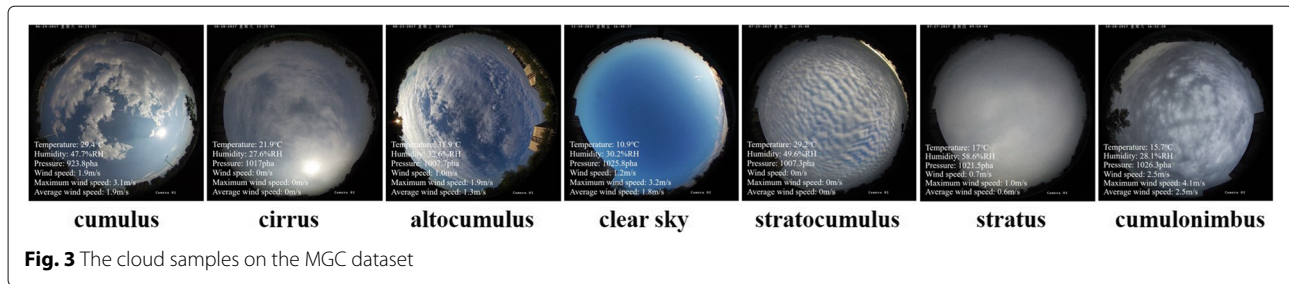
The MGC dataset collected in China consists of cloud images and multimodal cloud information. The cloud images are captured by a sky camera with a fisheye lens under a variety of conditions. The fisheye lens could scan the sky with a wide angle. In the interim, we utilize a weather station to capture the multimodal information of clouds, that is to say temperature, humidity, pressure, and wind speed. We should note that the cloud images and multimodal information are collected at the same time. Therefore, each cloud image corresponds to a set of multimodal data. The MGC dataset is a challenging dataset, because it covers a wide range of sky conditions and possesses large intra-class variations. The MGC dataset comprises a total number of 1720 cloud data. According to

the International cloud classification system criteria published in the World Meteorological Organization (WMO), considering the visual similarity in practice, the sky conditions are divided into seven classes, i.e., cumulus, cirrus, altocumulus, clear sky, stratus, stratocumulus, and cumulonimbus. Note that the clear sky is the condition that the cloud accounts for no more than 10% of the total sky. The number of cloud samples of each class varies from 140 to 350, and the detailed numbers are listed in Table 2. Herein, cloud classes are labeled using Arabic numerals from 1 to 7. Figure 3 shows some cloud samples from each class where each cloud image is with the size of 1056×1056 .

The MGC dataset is randomly partitioned into 120 training samples for each class and the remaining ones as the test set. The partition process is implemented 10 times independently, and the final classification accuracy is reported as the average accuracy over these 10 random splits. For fair comparison, the same experimental setup is used for all the experiments. In the training stage, we first resize the original cloud images into 256×256 pixels with preserved aspect ratio by bilinear interpolation. Then, in order to learn more cloud information, we centrally crop the training images into 224×224 pixels. In addition, each

Table 2 The sample number of each cloud class on the MGC dataset

Label	Cloud type	Number of samples
1	Cumulus	160
2	Cirrus	300
3	Altocumulus	340
4	Clear sky	350
5	Stratocumulus	250
6	Stratus	140
7	Cumulonimbus	180
Total number		1720



training image subtracts the mean RGB values computed on training set from each pixel.

We shuffle the training set images to fine-tune the pre-trained imagenet-vgg-f model. To learn the parameters (weights and bias), we train the network using the backpropagation gradient-descent procedure [21] with a mini-batch size of 48. The fine-tuning procedure is terminated at 20 epochs. For the first 10 epochs, the learning rate is set to 0.001. While for the remaining 10 epochs, it is reduced by a factor of 10 to 0.0001. The weight decay is with the value of 0.0005. In the test stage, the images have the same pre-processing with those in the training stage. For the multimodal information fusion, we empirically set α and β in Eq. (5) to 1 and 0.8, respectively. The multimodal information vector M is $[m_1, m_2, \dots, m_6]^T$, where m_1, m_2, \dots, m_6 denotes the temperature, humidity, pressure, wind speed, average wind speed, and maximum wind speed in one minute, respectively. Finally, we treat SVM with radial basis function (RBF) kernel as the classifier.

3.2 Baselines

We compare the proposed method with the state-of-the-art methods which are listed as follows.

- (1) BoW [37] model: The bag-of-words (BoW) model represents cloud images as histograms over a discrete codebook of local features. We choose SIFT [38] descriptors as the local features. The codebook size for each cloud class is set to 200, which results in 1400 dimensionality histogram for each cloud image.
- (2) PBoW [39] model: The pyramid BoW (PBoW) model is that the BoW model incorporates with the spatial pyramid which could learn the spatial information of cloud images. We divide each cloud image into three levels, i.e., 1, 2, and 4, which results in 1, 4, and 16 cells, respectively. Thus, for each cloud image, it contains a total of 21 cells. The PBoW model also represents cloud images as histograms based on each cell. Herein, the codebook is obtained in the same way as BoW. Hence, the histogram for each cloud image is 29400 dimensionality.
- (3) LBP [40]: The local binary pattern (LBP) labels each pixel by computing the sign of the difference between the intensities of that pixel and its

neighboring pixels. In our experiments, we utilize the uniform invariant LBP and set the parameter (P, R) to (8, 1), (16, 2), and (24, 3), respectively. Here, P is the total number of involved neighbors in a circle and R is the radius of the circle. Then we combine the representations from these three different conditions. Hence, the dimensionality of the representation for each cloud image is 54.

- (4) CLBP [41]: The completed LBP (CLBP) is an extension of LBP and has shown to perform well in image analysis and texture classification. In CLBP, a local region is represented by its center pixel, and the signs and magnitudes of the local differences. We combine these three components into joint distribution s to obtain completed cloud representation. The parameter (P, R) is also set to (8, 1), (16, 2), and (24, 3), respectively. We concatenate the three scales into one feature vector resulting in a 2200 dimensional vector.

For all of the above feature extraction techniques, we use the same training set and test set as the proposed method. The only difference is that each cloud image is converted to gray scale with the size of 300×300 .

3.3 Comparison with other methods

We first compare the deep visual features (DVF) learned from CNN with the other state-of-the-art methods, and the results are shown in Table 3. We extract the DVF from conv5, and therefore, the dimensionality of DVF is 169. From Table 3, we can see that the DVF obtains the best result. Especially, the classification accuracy of DVF is over 15% better than that of PBoW which achieves the

Table 3 Classification accuracy (%) using visual features

Method	Accuracy (%)
BoW	63.14
PBoW	69.77
LBP	56.77
CLBP	67.47
DVF	82.52

Table 4 Classification accuracy (%) with multimodal information

Method	Accuracy (%)
BoW+M	65.47
PBoW+M	70.65
LBP+M	58.77
CLBP+M	69.47
DMF	86.30

second best result among all the compared methods. Furthermore, the dimensionality of DVF is much smaller than PBoW. The improvement of the proposed DVF is because the CNN model could learn more discriminative features than other learning-based methods (BoW and PBoW) and hand-crafted features (LBP and CLBP). Moreover, the sum pooling strategy ensures that the DVF possesses more cloud spatial information.

Then, we compare the proposed DMF with the other state-of-the-art methods for the multimodal information fusion, and the results are listed in Table 4. Note that “+M” indicates concatenating the visual features with multimodal information. From Table 4, we can see that the proposed DMF outperforms the other methods and the classification accuracy achieves over 86%. A comparison between Tables 3 and 4 shows that the classification accuracies in the latter are all better than those in the former. This demonstrates that the multimodal cloud information provides support for the ground-based cloud classification. This demonstrates that the multimodal cloud information is helpful for the ground-based cloud classification. The visual features and the multimodal cloud information are complementary and therefore fusing them could obtain the completed information of ground-based cloud. The improvement of DMF exceeds other methods, which verifies the effectiveness of the fusion algorithm.

In this paper, we focus on the feature representation of cloud, and any classifiers could be chosen, such as the 1-nearest neighbor (1NN) classifier. In Table 5, we summarize the classification accuracy with 1NN classifier. From the table, we can observe that the proposed method obtains the best results when utilizing 1NN classifier.

Table 5 Classification accuracy (%) with 1NN classifier for different methods

Method	Accuracy (%)	Method	Accuracy (%)
BoW	61.79	BoW+M	63.77
PBoW	68.25	PBoW+M	67.54
LBP	56.29	LBP+M	58.09
CLBP	64.85	CLBP+M	66.65
DVF	79.33	DMF	81.29

Table 6 The feature dimensionality of different layers

Layers	SCM dimensionality	Features dimensionality
conv3	13×13	169
conv4	13×13	169
conv5	13×13	169
Layers	Neurons number	Features dimensionality
fc6	4096	4096
fc7	4096	4096

3.4 Influence of different parameters

In this subsection, we evaluate the performance of different layers in CNN for ground-based cloud classification. For the convolutional layers, the feature dimensionality is equal to that of SCM. For example, the size of SCM from conv5 is 13×13 , and therefore, the dimensionality of V_{cov5} is 169. For the FC layers, the feature dimensionality is equal to the number of neurons. For example, fc6 has 4096 neurons, and therefore, the dimensionality of V_{fc6} is 4096. The dimensionality of visual features extracted from the trained CNN is concluded in Table 6. Then, we directly concatenate visual features with multimodal features to obtain the final features.

The classification performance of DVF and DMF are summarized in Table 7. Several conclusions can be drawn from the results presented in Table 7. First, the accuracy of conv5 achieves the best in both DVF and DMF. Second, comparing conv5 with other shallower convolutional layers, we can see that deeper convolutional layers could learn more semantic information. Third, the accuracy of conv5 is higher than that of fc6 and fc7, while the feature dimensionality of conv5 is much less than that of fc6 and fc7. It is because the sum pooling strategy across over all feature maps could keep more spatial information. Forth, the classification accuracies of DMF are all higher than that of DVF, which validates the effectiveness of the fusion between the visual features and the multimodal information.

Additionally, we compare the classification results of V_{cov5} for different α and β settings. Since the ratio of α to β is important to fusion performance, we fix α and change β . The comparison results are listed in Table 8.

Table 7 The classification accuracy (%) of DVF and DMF in different layers

Layers	DVF	DMF
conv3	69.71	70.10
conv4	80.72	80.77
conv5	82.52	86.30
fc6	76.91	77.91
fc7	74.57	75.68

Table 8 Classification accuracy (%) of *conv5* for different α and β settings

(α, β)	(1, 2)	(1, 1.5)	(1, 1)	(1, 0.9)	(1, 0.8)	(1, 0.7)	(1, 0.6)
	59.29	67.02	82.84	83.63	86.30	82.95	62.70

From the table, we can see that when α and β are set to 1 and 0.8, respectively, the best classification accuracy is obtained. α is larger than β in the optimal situation which indicates that the visual features are more important than the multimodal features.

4 Conclusions

In this paper, the integration of the deep visual features and the multimodal information has been proposed for ground-based cloud classification in weather station networks. We first fine-tune the pre-trained deep CNN model using the cloud images, followed by extraction of deep visual features and then fused with the multimodal information. A series of comparative experiments have been conducted to test the effectiveness of the proposed DMF, and the results show that the accuracy of the proposed DMF is higher than the state-of-the-art methods.

Abbreviations

BoW: Bag of words; CLBP: Completed LBP; CNN: Convolutional neural network; DMF: Deep multimodal fusion; DVF: Deep visual features; FC: Fully connected; LBP: Local binary pattern; MGC: Multimodal ground-based cloud; PBoW: Pyramid BoW; SCM: Sum convolutional map

Acknowledgements

The authors would like to thank the editor and the anonymous reviewers for their helpful comments and suggestions in improving the quality of this paper. This work was supported by the National Natural Science Foundation of China under Grant No. 61501327, the Natural Science Foundation of Tianjin under Grant No. 17JCZDJC30600 and No. 15JCQNJC01700, the Open Projects Program of National Laboratory of Pattern Recognition under Grant No. 201800002, and the China Scholarship Council No. 201708120039.

Funding

This work was supported by the National Natural Science Foundation of China under Grant No. 61501327, the Natural Science Foundation of Tianjin under Grant No. 17JCZDJC30600 and No. 15JCQNJC01700, the Open Projects Program of National Laboratory of Pattern Recognition under Grant No. 201800002, and the China Scholarship Council No. 201708120039.

Availability of data and materials

The basic codes are available via email to the corresponding author.

Authors' contributions

SL contributed to the conception and algorithm design of the study. ML took charge of the implementation of the experiment, and SL verified all the results. ML and SL drafted and revised the manuscript. Both authors read and approved the final manuscript.

Authors' information

Shuang Liu received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences. She is currently an Associate Professor at Tianjin Normal University. Mei Li is currently pursuing the M.S. degree at Tianjin Normal University.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 January 2018 Accepted: 20 February 2018

Published online: 27 February 2018

References

1. J Calbo, J Sabburg, Featu3re extraction from whole-sky ground-based images for cloud-type recognition. *J. Atmos. Ocean. Technol.* **25**(1), 3–14 (2008)
2. AJ Illingworth, RJ Hogan, EJ O'connor, D Bouniol, J Delanoë, J Pelon, A Protat, ME Brooks, N Gaussiat, DR Wilson, et al., Cloudnet: continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bull. Am. Meteorol. Soc.* **88**(6), 883–898 (2007)
3. L Liu, X Sun, F Chen, S Zhao, T Gao, Cloud classification based on structure features of infrared images. *J. Atmos. Ocean. Technol.* **28**(3), 410–417 (2011)
4. CN Long, DW Slater, T Tooman, Total sky imager model 880 status and testing results. Technical report, DOE/SC-ARM/TR-006 (2001)
5. CN Long, JM Sabburg, J Calbo, D Pagès, Retrieving cloud characteristics from ground-based daytime color all-sky images. *J. Atmos. Ocean. Technol.* **23**(5), 633–652 (2006)
6. JA Shaw, B Thurairajah, in *ARM Science Team Meeting*. Short-term arctic cloud statistics at NSA from the infrared cloud imager (ARM, Broomfield, 2013), pp. 1–7
7. J Huo, D Lu, Comparison of cloud cover from all-sky imager and meteorological observer. *J. Atmos. Ocean. Technol.* **29**(8), 1093–1101 (2012)
8. F Zhao, B Li, H Chen, X Lv, Joint beamforming and power allocation for cognitive MIMO systems under imperfect CSI based on game theory. *Wirel. Pers. Commun.* **73**(3), 679–694 (2013)
9. X Sun, T Gao, D Zhai, S Zhao, J Lian, Whole sky infrared cloud measuring system based on the uncooled infrared focal plane array. *Infrared Laser Eng.* **37**(5), 761–764 (2008)
10. JE Shields, ME Karr, RW Johnson, AR Burden, Day/night whole sky imagers for 24-h cloud and sky assessment: history and overview. *Appl. Opt.* **52**(8), 1605–1616 (2013)
11. A Heinle, A Macke, A Srivastav, Automatic cloud classification of whole sky images. *Atmos. Meas. Tech.* **3**(3), 557–567 (2010)
12. MS Ghonima, B Urquhart, CW Chow, JE Shields, A Cazorla, J Kleissl, A method for cloud detection and opacity classification based on ground based sky imagery. *Atmos. Meas. Tech.* **5**(11), 2881–2892 (2012)
13. W Zhuo, Z Cao, Y Xiao, Cloud classification of ground-based images using texture–structure features. *J. Atmos. Ocean. Technol.* **31**(1), 79–92 (2014)
14. A Kazantzidis, P Tzoumanikas, AF Bais, S Fotopoulos, G Economou, Cloud detection and classification with the use of whole-sky ground-based images. *Atmos. Res.* **113**, 80–88 (2014)
15. HY Cheng, CC Yu, Block-based cloud classification with statistical features and distribution of local texture features. *Atmos. Meas. Tech.* **8**(3), 1173–1182 (2015)
16. Y Xiao, Z Cao, W Zhuo, L Ye, L Zhu, mcloud: A multiview visual feature extraction mechanism for ground-based cloud image categorization. *J. Atmos. Ocean. Technol.* **33**(4), 789–801 (2016)
17. X Sun, L Liu, S Zhao, Whole sky infrared remote sensing of cloud. *Procedia Earth Planet. Sci.* **2**(Supplement C), 278–283 (2011)
18. S Liu, C Wang, B Xiao, Z Zhang, Y Shao, in *International Conference on Computer Vision in Remote Sensing*. Ground-based cloud classification using multiple random projections (IEEE, Xiamen, 2012), pp. 7–12
19. S Liu, C Wang, B Xiao, Z Zhang, Y Shao, Salient local binary pattern for ground-based cloud classification. *Acta Meteorol. Sin.* **27**(2), 211–220 (2013)
20. S Liu, Z Zhang, Learning group patterns for ground-based cloud classification in wireless sensor networks. *EURASIP J. Wirel. Commun. Netw.* **2016**(1), 69 (2016)
21. A Krizhevsky, I Sutskever, GE Hinton, in *Advances in Neural Information Processing Systems*. Imagenet classification with deep convolutional neural networks (NIPS Foundation, Lake Tahoe, 2012), pp. 1097–1105
22. P Sermanet, D Eigen, X Zhang, M Mathieu, R Fergus, Y LeCun, Overfeat: Integrated recognition, localization and detection using convolutional networks. *CoRR*. [abs/1312.6229](https://arxiv.org/abs/1312.6229) (2013). <https://arxiv.org/abs/1312.6229>

23. O Abdel-Hamid, L Deng, D Yu, in *Interspeech*. Exploring convolutional neural network structures and optimization techniques for speech recognition (ISCA Archive, Lyon, 2013), pp. 3366–3370
24. L Ye, Z Cao, Y Xiao, Deepcloud: ground-based cloud image categorization using deep convolutional features. *IEEE Trans. Geosci. Remote Sens.* **55**(10), 5729–5740 (2017)
25. C Shi, C Wang, Y Wang, B Xiao, Deep convolutional activations-based features for ground-based cloud classification. *IEEE Geosci. Remote Sens. Lett.* **14**(6), 816–820 (2017)
26. F Zhao, L Wei, H Chen, Optimal time allocation for wireless information and power transfer in wireless powered communication systems. *IEEE Trans. Veh. Technol.* **65**(3), 1830–1835 (2016)
27. F Zhao, X Sun, H Chen, R Bie, Outage performance of relay-assisted primary and secondary transmissions in cognitive relay networks. *EURASIP J. Wirel. Commun. Netw.* **2014**(1), 60 (2014)
28. F Zhao, H Nie, H Chen, Group buying spectrum auction algorithm for fractional frequency reuse cognitive cellular systems. *Ad Hoc Netw.* **58**, 239–246 (2017)
29. F Zhao, W Wang, H Chen, Q Zhang, Interference alignment and game-theoretic power allocation in mimo heterogeneous sensor networks communications. *Sig. Process.* **126**, 173–179 (2016)
30. CC Chang, CJ Lin, Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**(3), 27 (2011)
31. K Simonyan, A Zisserman, Very deep convolutional networks for large-scale image recognition. *CoRR*. **abs/1409.1556** (2014). <https://arxiv.org/abs/1409.1556>
32. M Oquab, L Bottou, I Laptev, J Sivic, in *IEEE Conference on Computer Vision and Pattern Recognition*. Learning and transferring mid-level image representations using convolutional neural networks (IEEE, Columbus, 2014), pp. 1717–1724
33. S Lawrence, CL Giles, AC Tsoi, AD Back, Face recognition: a convolutional neural-network approach. *IEEE Trans. Neural Netw.* **8**(1), 98–113 (1997)
34. K Chatfield, K Simonyan, A Vedaldi, A Zisserman, Return of the devil in the details: delving deep into convolutional nets. *CoRR*. **abs/1405.3531** (2014). <https://arxiv.org/abs/1405.3531>
35. Y Song, Q Li, D Feng, J Zou, W Cai, Texture image classification with discriminative neural networks. *Comput. Vis. Media.* **2**(2), 367–377 (2016)
36. Z Lu, J Yang, Q Liu, Face image retrieval based on shape and texture feature fusion. *Comput. Vis. Media.* **3**(4), 359–368 (2017)
37. FF Li, P Perona, in *IEEE Conference on Computer Vision and Pattern Recognition*. A bayesian hierarchical model for learning natural scene categories (IEEE, San Diego, 2005), pp. 524–531
38. DG Lowe, Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
39. S Lazebnik, C Schmid, J Ponce, in *IEEE Conference on Computer Vision and Pattern Recognition*. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories (IEEE, New York, 2006), pp. 2169–2178
40. T Ojala, M Pietikäinen, D Harwood, A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **29**(1), 51–59 (1996)
41. Z Guo, L Zhang, D Zhang, A completed modeling of local binary pattern operator for texture classification. *IEEE Trans. Image Process.* **19**(6), 1657–1663 (2010)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
