

Received April 25, 2019, accepted May 8, 2019, date of publication May 15, 2019, date of current version May 28, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2916905

Dual Guided Loss for Ground-Based Cloud Classification in Weather Station Networks

MEI LI, SHUANG LIU^{ID}, (Member, IEEE), AND ZHONG ZHANG^{ID}, (Member, IEEE)

Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China

Corresponding author: Shuang Liu (shuangliu.tjnu@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61501327 and Grant 61711530240, in part by the Natural Science Foundation of Tianjin under Grant 17JCZDJC30600, in part by the Fund of Tianjin Normal University under Grant 135202RC1703, in part by the Open Projects Program of National Laboratory of Pattern Recognition under Grant 201800002, and in part by the Tianjin Higher Education Creative Team Funds Program.

ABSTRACT In this paper, we propose a novel loss function named dual-guided loss (DGL) for ground-based cloud classification in weather station networks. The proposed DGL could integrate the knowledge of different convolutional neural networks (CNNs) in the process of optimization, which improves the discriminative ability of ground-based cloud feature representations. To this end, we add a modulation term into the DGL, which assigns large weights to the hard-classified ground-based cloud samples. As a result, the deep network is forced to pay more attention to these hard-classified samples, and therefore, the performance of the deep network gets improved. We demonstrate the effectiveness of the proposed DGL with the extensive experiments on two ground-based cloud datasets, and the experimental results of the DGL outperform the state-of-the-art methods.

INDEX TERMS Dual guided loss, ground-based cloud classification, weather station networks.

I. INTRODUCTION

Accurate ground-based classification is vital for many real-world applications, such as climate monitoring, weather forecast, agriculture and aviation industry. The use of weather station networks equipped with image sensors is one of the most effective ways to collect substantial ground-based cloud images. Correspondingly, various approaches have been developed for automatic ground-based cloud classification with these collected samples.

Some approaches [1]–[6] usually employ hand-crafted features, such as texture, structure, color and so on, to represent clouds. However, hand-crafted features are designed based on the expert knowledge, and the discriminative ability is limited. It is because hand-crafted features are difficult to adapt to great variations of cloud appearances caused by different sensor locations in weather station networks.

Over the past few years, deep learning has significantly boosted the state-of-the-art performance in a variety of visual classification tasks, for example, object recognition [7]–[9] and face recognition [10], [11], and has pointed out a new way for ground-based cloud classification [12]–[18]. The convolutional neural network (CNN) is a representative of

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang.

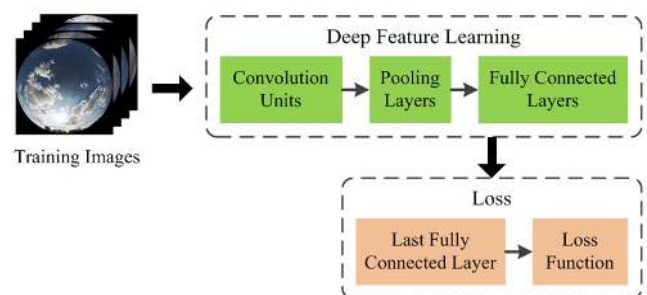


FIGURE 1. The general pipeline of CNN model.

deep learning algorithms and has been used to learn discriminative features. The advantage of CNN lies in the ability to learn highly complex non-linear features owing to its multiple convolutional layers and pooling layers, and therefore CNN is able to adapt to different data distributions.

The CNN models can be viewed as the combination of deep feature learning component and loss component as shown in Fig. 1. It should be noticed that since the outputs of the last fully connected layer are usually employed to compute the CNN loss, we define the loss function together with the last fully connected layer as the loss component

of CNN. The parameters of CNN is optimized by minimizing the loss using the back-propagated algorithm, and accordingly the representation ability of each layer is strengthened. Currently, the common-used loss functions are contrastive loss [19], triplet loss [20] and cross-entropy loss [21]. These loss functions aim to strengthen intra-class compactness and inter-class separability. For ground-based cloud classification task, the cross-entropy loss is the most commonly used benefiting from its simplicity and excellent performance. However, the cross-entropy loss treats each class equally and does not take into consideration the knowledge of other CNN models, which results in the knowledge poverty for the hard-classified cloud samples.

In this paper, we propose a novel loss named dual guided loss (DGL) for ground-based cloud classification in weather station networks, which could embed the knowledge of different CNN models into the process of CNN optimization. To this end, the DGL is obtained by imposing a modulation term on the cross-entropy loss. The modulation term reflects the distribution of each cloud category learned from another CNN model, and therefore the DGL is restricted by both the prior knowledge and current prediction probabilities. The modulation term assigns a large weight to the cloud image with low classification accuracy. This in turn increases the importance of specific cloud classes, and reaches a balanced learning ability of the CNN model between the easy-classified cloud images and the hard-classified cloud images. In the experiments, we implement the proposed algorithm on the ResNet-50 [22] and verify the performance on two ground-based cloud datasets. The experimental results verify that the proposed DGL has an advantage over the cross-entropy loss as well as the hand-crafted features.

The remainder of this paper is constructed as follows. Section II briefly reviews the related work. Section III presents a description of the proposed DGL in detail. Section IV exhibits comprehensive experimental results on two ground-based cloud datasets and Section V finally terminates this paper.

II. RELATED WORK

This work is closely connected to the deep learning-based ground-based cloud classification and the loss of CNNs. We briefly introduce the two kinds of works in this section.

A. DEEP LEARNING-BASED CLOUD CLASSIFICATION

Since clouds have a great texture characteristic, many researchers employ texture classification techniques for ground-based cloud classification. Zhuo *et al.* [23] combined the texture, local structure and global rough structure information to classify ground-based cloud images, where the former two are captured by the color census transform (CCT) and the later one is captured by the block assignment method. Liu *et al.* [24] presented the weighted local binary

patterns (WLBP) which assigns each pixel an adaptive weight in the process of accumulating histogram to distinguish different ground-based cloud categories. In stable LBPs [25], the averaged ranks of the occurrence frequencies of all rotation invariant patterns in the LBPs are learned in order to obtain stable representations for classifying ground-based cloud images.

Recently, deep learning-based features have been demonstrated the effectiveness for ground-based cloud classification. For instance, Taravat *et al.* [26] verified that multilayer perceptron (MLP) achieves better performance than support vector machine (SVM) for ground-based cloud classification in most situations. Xia *et al.* [12] applied a hybrid scheme on the basis of neural network and K -nearest neighbor to distinguish various cloud types. Shi *et al.* [13] conducted the max-pooling or sum-pooling strategy on convolutional feature maps of CNN to produce the deep convolutional activations-based features to represent each ground-based cloud image. Ye *et al.* [14] exerted the cloud pattern mining and selection method to discover meaningful local descriptors from convolutional activation maps, and then encoded them using the Fisher vector. Zhang *et al.* [15] proposed the CloudNet to learn texture, structure and shape features from CNN for cloud classification.

Apart from visual features extracted from cloud images, some approaches investigate to fuse the cloud multi-modal information under the framework of CNN. For example, Liu and Li [16] exerted the sum-pooling operation across convolutional activation maps and flattened them as cloud visual features. Then the visual features are directly concatenated with the multimodal information to reform the ground-based cloud representation. Afterward, the joint fusion convolutional neural network (JFCNN) [17] was proposed to integrate visual features and multi-modal information using the joint fusion layer in a unified framework. Furthermore, Liu *et al.* [18] presented the multimodal generative adversarial network (Multimodal GAN) to generate the cloud images and the corresponding multimodal information simultaneously, which could improve the generation ability of the classification network.

B. LOSS FUNCTION

Loss functions aim to estimate how much the predicted value deviates from the corresponding true value. The parameters of deep neural networks are optimized by minimizing the loss and therefore the representative features can be learned. The loss functions utilized for classification tasks are the contrastive loss, triplet loss, cross-entropy loss as well as a series of their variants [27]–[29]. At present, the CNN models of ground-based cloud classification are mainly supervised by the multinomial logistic regression [13], [15] and the cross-entropy loss [16], [17]. Normally, the outputs of the last fully connected layer are first passed through the softmax function and then computed by the corresponding loss function. To improve the generalization ability of CNN, Liu *et al.* [18] first generated the fake cloud samples, and

then trained them using the label smoothing regularization for outliers (LSRO) together with the cross-entropy loss.

III. APPROACH

Deep learning methods especially CNN models have achieved huge success in terms of ground-based cloud classification. The majority of them focus on how to design the CNN structure, and few involves the loss function of CNN. This paper proposes the DGL to enhance the discriminative ability of features by explicitly integrating the knowledge of different CNN models in the loss function. Since the DGL is derived from the cross-entropy loss, we introduce it starting from the cross-entropy loss.

A. CROSS-ENTROPY LOSS

For ground-based classification, the cross-entropy loss is one of the most commonly employed losses in CNN and formulated as

$$L = - \sum_{i=1}^K q_i \log p_i, \quad (1)$$

where K is the ground-based cloud category number, $q_i = 1$ if i is the ground-truth label and 0 otherwise, and $p_i \in [0, 1]$ indicates the predicted probability of the i -th cloud category. p_i is calculated by the softmax function

$$p_i = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}}, \quad (2)$$

where x_i is the output result of the i -th neuron of the last fully connected layer. In the training procedure, the loss calculated by Equation (1) is back-propagated through the whole network. As a result, the network assigns a high value to x_i when i is the ground-truth label, otherwise x_i is assigned to a low value. Hence, the predicted probability of the ground-truth label in Equation (2) is maximized.

From Equation (1), it is observed that the cross-entropy loss treats each category sample equally. However, there exists imbalance in terms of classification difficulty among different categories. That is, the number of hard-classified samples in different cloud categories is significantly different as shown in Fig. 2(a). Hence, the loss should consider the cloud images from different categories in an unequal way in the training process.

B. DUAL GUIDED LOSS

To overcome the weakness of cross-entropy loss, we propose the DGL which utilizes the modulation term to adjust the learning weights for hard-classified samples. The modulation term of the i -th class is defined as

$$r_i = \alpha_i^\gamma, \quad (3)$$

where α_i is an indicator of classification difficulty, and γ is a positive tunable parameter. In this paper, α_i is computed by

$$\alpha_i = \frac{1}{c_i}, \quad (4)$$

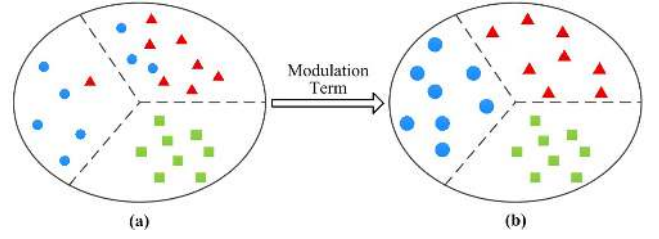


FIGURE 2. (a) Indicates the classification result of the cross-entropy loss, and (b) indicates the classification result of the proposed DGL which is imposed by the modulation term. Herein, the circle, triangle and square represent three different categories of cloud samples. In (a), the cloud categories represented by circle and triangle include many hard-classified samples, and the category represented by square is well classified. In (b), after imposing the modulation term, the three categories are all well classified by learning in an unequal way.

where c_i is the classification accuracy of the i -th class which is learned by the CNN model with the traditional cross-entropy loss. Hence, a high weight will be assigned to the sample with the low classification accuracy.

Combining Equation (1) and Equation (3), the DGL is formulated as

$$L_{DGL} = - \sum_{i=1}^K r_i q_i \log p_i. \quad (5)$$

There are two properties of the DGL. Firstly, the DGL assigns different weights according to the degree of classification difficulty, which could force the CNN model to tilt the hard-classified cloud samples. Specifically, when a cloud image is from the easy-classified class, the modulation term is near 1 and the loss is almost unchanged. On the contrary, when a cloud image is from the hard-classified class, the modulation term is larger than 1 and the loss value rises due to the increased weight of the hard-classified cloud sample. Secondly, the DGL could smoothly adjust the influence of the hard-classified cloud samples using the parameter γ . When $\gamma = 0$, the DGL degenerates into the traditional cross-entropy loss. As γ becomes larger, the effect of the modulation term increases. As a result, the sample is penalized heavily when it is misclassified. This in turn forces the network to concentrate more on the hard-classified samples and therefore we learn an optimal CNN model to classify different types of ground-based cloud images.

IV. EXPERIMENTS

In this section, we start with presenting the description of two ground-based cloud datasets, and then give an introduction of the experimental setup. Afterwards, we show the classification results of two cloud datasets. Eventually, the important parameters are analyzed.

A. GROUND-BASED CLOUD DATASETS

GCD-A. The ground-based cloud dataset A (GCD-A) contains 8000 ground-based cloud images captured by the sky camera with fisheye lens. Since the collection of the dataset undergoes a long period of time, the cloud images have large

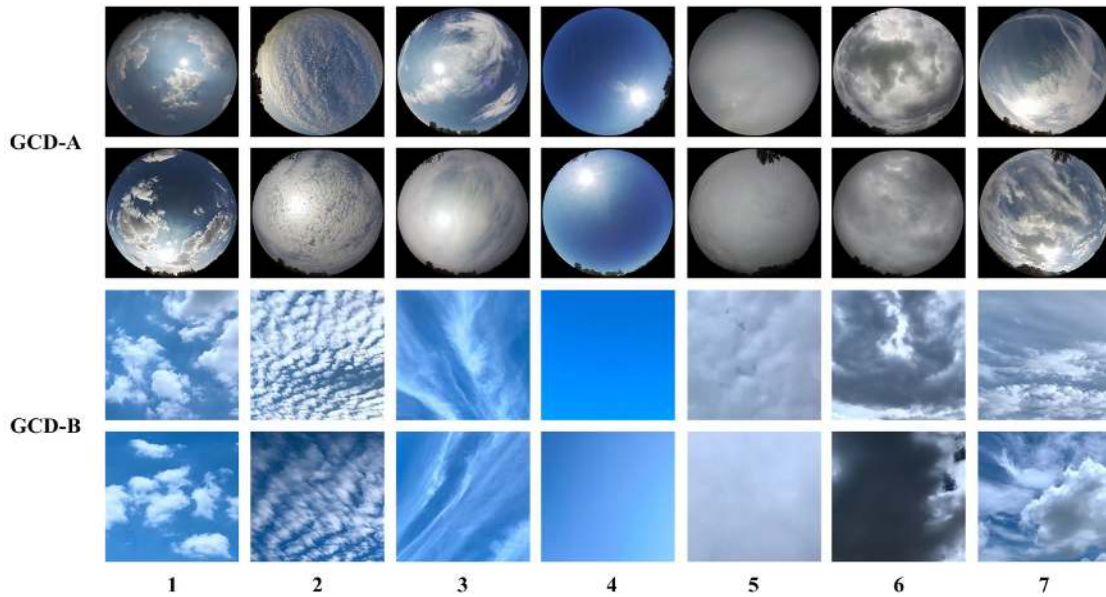


FIGURE 3. Some representative cloud samples from each category in GCD-A and GCD-B. The Arabic numerals from 1 to 7 indicate 1) cumulus, 2) altocumulus and cirrocumulus, 3) cirrus and cirrostratus, 4) clear sky, 5) stratocumulus, stratus and altostratus, 6) cumulonimbus and nimbostratus, and 7) mixed cloud, respectively.

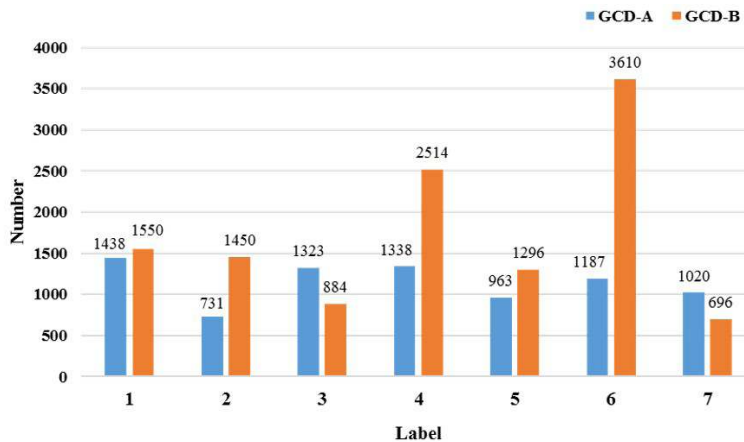


FIGURE 4. The detailed sample numbers of each category on GCD-A and GCD-B.

variations in illumination and appearance. Each cloud image is with the resolution of 1024×1024 pixels and preserved in the JPEG format. The dataset is partitioned into 7 categories, including 1) cumulus, 2) altocumulus and cirrocumulus, 3) cirrus and cirrostratus, 4) clear sky, 5) stratocumulus, stratus and altostratus, 6) cumulonimbus and nimbostratus, and 7) mixed cloud. Herein, the categories are classified in terms of the genera-based classification recommendation of the World Meteorological Organization (WMO) and the visual similarities of cloud appearance in practice. Additionally, cloud images with cloudiness less than 10% are treated as the clear sky. The GCD-A is divided into the training set and the test set, both of which are with 4000 samples.

GCD-B. The ground-based cloud dataset B (GCD-B) is collected in Zhangjiakou, Hebei, China from 2017 to 2018.

It contains 12000 ground-based cloud images and is classified into seven sky types according to the same criteria as GCD-A. The GCD-B is separated into the training set and the test set in which both of them contain 6000 cloud images with seven sky types. Furthermore, each cloud image is captured by a camera sensor and preserved in the JPEG format with the pixel resolution of 512×512 . Some representative cloud samples from GCD-A and GCD-B are shown in Fig. 3, and the detailed sample numbers of each category on the two datasets are illustrated in Fig. 4.

B. EXPERIMENTAL SETUP

As for GCD-A, we initially resize the ground-based cloud images into 252×252 , and then subtract the mean RGB values calculated from the training set for each cloud image.

To augment the number of training examples, we randomly crop the cloud images into 224×224 , and randomly flip the cloud images horizontally. The ResNet-50 is utilized as the training model, and the last fully connected layer is substituted with a new one with 7 neurons which are the number of cloud categories.

In the training process, the parameters of the ResNet-50 except the last fully connected layer are initialized by the pre-trained model on ImageNet. As for the last fully connected layer, the biases are initialized to 0 and the weights are initialized to a normal distribution with the standard deviation of 0.01. The parameter α_i in Equation (3) is assigned to the reciprocal of the i -th category classification accuracy learned from the ResNet-50 model with the traditional cross-entropy loss. We set γ in Equation (3) to 1 and 4 for GCD-A and GCD-B, respectively. The ResNet-50 is optimized by the stochastic gradient descent (SGD) with the batch size of 32. The momentum and the weight decay are set to 0.9 and 2×10^{-4} , respectively. In addition, the learning rate starts at 0.0001 and is reduced by a factor of 10 at the 30-th epoch of 50 epochs.

The experimental setup of GCD-B is the same as GCD-A with the exception of the momentum and the weight decay being set to 0.95 and 5×10^{-4} , respectively.

C. RESULT ANALYSIS

We first compare the proposed DGL with other methods, i.e., bag-of-visual-words (BoVW) [30], local binary patterns (LBP) [31], the completed LBP (CLBP) [32] and the ResNet-50 optimized by the cross-entropy loss. The BoVW densely extracts the SIFT descriptors [33] from each cloud image and then generates a codebook with 300 clusters using k -means algorithm. In addition, we employ the spatial pyramid matching scheme [34] where each cloud image is partitioned into three levels with 1, 4, and 16 sub-regions, respectively. Therefore, the final representation of BoVW for each cloud image is a 6300-dimensional histogram. The LBP is represented by a series of binary codes which are the difference signs between the center pixel and its neighbors. We utilize the uniform invariant LBP descriptor $LBP_{P,R}^{riu2}$ to estimate the performance of the ground-based cloud classification, where P and R represent the neighbor number of the center pixel and the circle radius. Herein, (P, R) is set to (8, 1), (16, 2) and (24, 3), respectively. Accordingly, the dimensions of the feature vectors are 10, 18 and 26 under the three (P, R) conditions. The CLBP is evolved from the LBP, and it additionally considers the center pixel and the magnitude of local differences. The final cloud image is represented by combining these three descriptors jointly. The parameter (P, R) is set to (8, 1), (16, 2) and (24, 3) respectively, and therefore each cloud image is represented as the feature vectors with the dimensions of 200, 648 and 1352, respectively. To demonstrate the effectiveness of the proposed DGL, we also train the ResNet-50 using the traditional cross-entropy in Equation (1) which has the same initialization and learning strategy as the proposed DGL.

TABLE 1. The classification accuracies (%) of different methods on GCD-A and GCD-B.

Methods	GCD-A	GCD-B
BoW	66.13	79.18
$LBP_{8,1}^{riu2}$	45.38	61.67
$LBP_{16,2}^{riu2}$	49.00	64.08
$LBP_{24,3}^{riu2}$	50.20	64.38
$CLBP_{8,1}^{riu2}$	65.10	70.42
$CLBP_{16,2}^{riu2}$	68.20	74.20
$CLBP_{24,3}^{riu2}$	69.18	75.58
ResNet-50 + CEL	83.15	91.65
ResNet-50 + DGL	85.28	92.80

The comparison results of different methods on GCD-A and GCD-B are shown in Table 1. Herein, ResNet-50 + CEL indicates the ResNet-50 is trained by the traditional cross-entropy loss, and ResNet-50 + DGL represents the ResNet-50 is trained by the proposed DGL. From the table we can observe that the CNN-based methods (ResNet-50 + CEL, ResNet-50 + DGL) outperform the hand-crafted methods (LBP and CLBP) and the learning-based methods (BoVW) on both GCD-A and GCD-B. It is because CNN is composed of multiple convolutional and pooling layers which could learn complex nonlinear representations according to different distributions of ground-based datasets. Such an advantage endows the deep features with more discriminative ability. Furthermore, the accuracies of ResNet-50 + DGL are superior to those of ResNet-50 + CEL on the two datasets, which indicates the effectiveness of assigning higher weights to the hard-classified samples.

Then, we analyze the classification performance of the cross-entropy loss and DGL for each cloud category, and the confusion matrices of them on GCD-A and GCD-B are presented in Fig. 5 and Fig. 6, respectively. From the left parts of the two figures, we can see that the classification results of the cross-entropy loss exist imbalance among these cloud categories on both GCD-A and GCD-B. Specifically, as for GCD-A, class 3 and class 7 only achieve the classification accuracies of 67.90% and 63.73%, while class 4 achieves the accuracy of 100%. Hence, class 3 and class 7 are the hard-classified cloud categories. Similarly, as for GCD-B, class 5 and class 7 are hard-classified cloud categories due to their low classification accuracies. These hard-classified cloud categories are with larger intra-class variations and smaller inter-class variations, which deteriorates the learning ability of the network, and therefore it is unreasonable for cross-entropy loss to treat each cloud samples equally. On the other side, the classification results of the DGL on GCD-A and GCD-B are shown in the right parts of Fig. 5 and Fig. 6. According to the figures, the classification accuracies of class 3 and class 7 on GCD-A achieve 74.15% and 68.63% which lead to the improvements of 6.25% and 4.90% compared with ResNet-50 + CEL. Similarly, the improvements achieve 6.69% and 5.74% for class 5 and class 7 on GCD-B, respectively. It is because the samples from the hard-classified categories are penalized by DGL heavily when they are

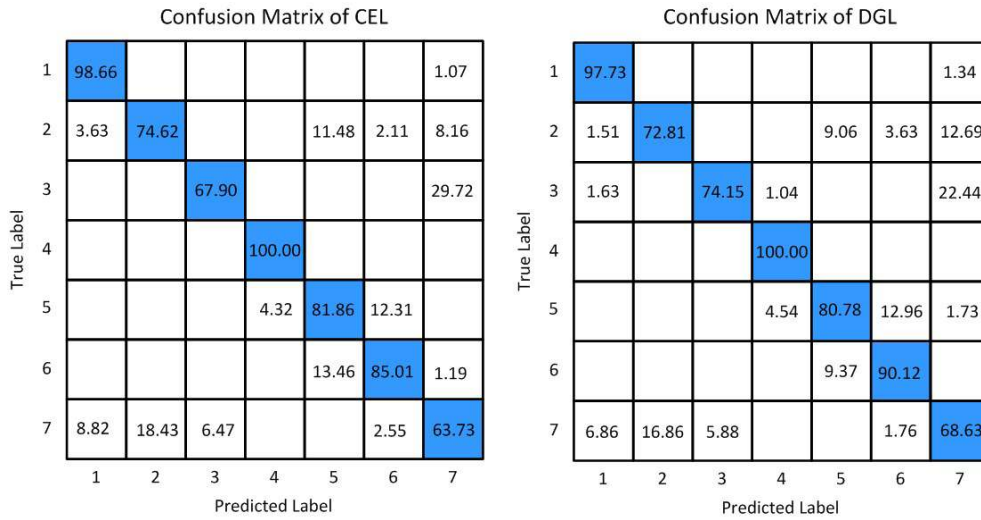


FIGURE 5. The confusion matrices of CEL and DGL on GCD-A.

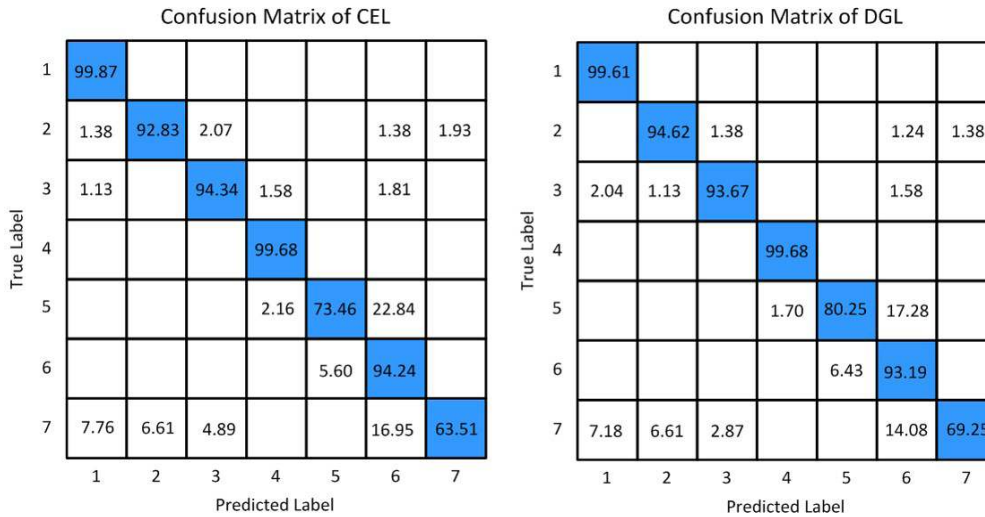


FIGURE 6. The confusion matrices of CEL and DGL on GCD-B.

TABLE 2. The classification accuracies (%) of the proposed DGL with different γ on GCD-A and GCD-B.

γ	0	0.5	1	2	3	4	5
GCD-A	83.15	84.85	85.28	84.68	84.73	85.00	84.55
GCD-B	91.65	92.32	92.48	92.37	92.77	92.80	92.02

misclassified, and therefore the CNN model pays more attention on these samples. Meanwhile, the classification accuracies of the easy-classified categories, such as classes 1, 2 and 4 in GCD-A, and classes 1, 3 and 5 in GCD-B, are almost unchanged. That is to say, the improvement of the overall classification accuracy derives from the improvement of hard-classified categories.

D. PARAMETER ANALYSIS

The parameter γ in Equation (3) is introduced to regulate the strength of the modulation term. We analyze the classification

performance of the proposed DGL with different γ and the results are illustrated in Table 2. From the table, when $\gamma = 0$, it is equivalent to the cross-entropy loss and obtains 83.15% classification accuracy. When γ increases, the hard-classified cloud samples are assigned to larger weights, and accordingly the DGL enforces the network to pay more attention to the hard-classified samples. Hence, the DGL shows obvious gains over the cross-entropy loss as γ increases, and it achieves the peak classification results of 85.28% and 92.80% when γ is equal to 1 and 4 for GCD-A and GCD-B. Moreover, when γ is greater than 1 and 4 on GCD-A and

GCD-B respectively, the classification performance of the DGL declines, but it is still superior to the cross-entropy loss. In a word, the optimal γ is set to 1 and 4 for GCD-A and GCD-B, respectively.

V. CONCLUSION

In this paper, we have presented the DGL to supervise the training process of CNN for ground-based cloud classification in weather station networks. The proposed DGL integrates the prior knowledge of each cloud category using the modulation term. By increasing the weights of the hard-classified cloud samples, the proposed DGL forces the network to pay more attention to the hard-classified cloud samples so as to learn discriminative and robust features for ground-based cloud representations. To validate the effectiveness of the proposed DGL, we have conducted a series of comparative experiments on two ground-based datasets and the results show that the proposed DGL achieves better classification performance than the other methods.

REFERENCES

- [1] A. Kazantzidis, P. Tzoumanikas, A. F. Bais, S. Fotopoulos, and G. Economou, "Cloud detection and classification with the use of whole-sky ground-based images," *Atmos. Res.*, vol. 113, pp. 80–88, Sep. 2012.
- [2] H.-Y. Cheng and C.-C. Yu, "Block-based cloud classification with statistical features and distribution of local texture features," *Atmos. Meas. Techn.*, vol. 8, no. 3, pp. 1173–1182, 2015.
- [3] S. Dev, Y. H. Lee, and S. Winkler, "Categorization of cloud image patches using an improved texon-based approach," in *Proc. IEEE Int. Conf. Image Process.*, Quebec City, QC, Canada, Sep. 2015, pp. 422–426.
- [4] Y. Xiao, Z. Cao, W. Zhuo, L. Ye, and L. Zhu, "mCLOUD: A multiview visual feature extraction mechanism for ground-based cloud image categorization," *J. Atmos. Ocean. Technol.*, vol. 33, no. 4, pp. 789–801, 2015.
- [5] S. Liu and Z. Zhang, "Learning group patterns for ground-based cloud classification in wireless sensor networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 1, p. 69, 2016.
- [6] Q. Li, Z. Zhang, W. Lu, J. Yang, Y. Ma, and W. Yao, "From pixels to patches: A cloud classification method based on a bag of micro-structures," *Atmos. Meas. Techn.*, vol. 9, no. 2, pp. 753–764, 2016.
- [7] A. Li, Z. Wu, H. Lu, D. Chen, and G. Sun, "Collaborative self-regression method with nonlinear feature based on multi-task learning for image classification," *IEEE Access*, vol. 6, pp. 43513–43525, 2018.
- [8] X. Yang, W. Wu, K. Liu, P. W. Kim, A. K. Sangaiah, and G. Jeon, "Long-distance object recognition with image super resolution: A comparative study," *IEEE Access*, vol. 6, pp. 13429–13438, 2018.
- [9] A. Caglayan and A. B. Can, "Volumetric object recognition using 3-D CNNs on depth data," *IEEE Access*, vol. 6, pp. 20058–20066, 2018.
- [10] J. Kong, M. Chen, M. Jiang, J. Sun, and J. Hou, "Face recognition based on CSGF(2D)²PCANet," *IEEE Access*, vol. 6, pp. 45153–45165, 2018.
- [11] L. Yang, J. Ma, J. Lian, Y. Zhang, and H. Liu, "Deep representation for partially occluded face verification," *EURASIP J. Image Video Process.*, vol. 2018, no. 1, p. 143, 2018.
- [12] M. Xia, W. Lu, J. Yang, Y. Ma, W. Yao, and Z. Zheng, "A hybrid method based on extreme learning machine and k-nearest neighbor for cloud classification of ground-based visible cloud image," *Neurocomputing*, vol. 160, pp. 238–249, Jul. 2015.
- [13] C. Shi, C. Wang, Y. Wang, and B. Xiao, "Deep convolutional activations-based features for ground-based cloud classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 6, pp. 816–820, Jun. 2017.
- [14] L. Ye, Z. Cao, and Y. Xiao, "DeepCloud: Ground-based cloud image categorization using deep convolutional features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5729–5740, Oct. 2017.
- [15] J. Zhang, P. Liu, F. Zhang, and Q. Song, "CloudNet: Ground-based cloud classification with deep convolutional neural network," *Geophys. Res. Lett.*, vol. 45, no. 16, pp. 8665–8672, Aug. 2018.
- [16] S. Liu and M. Li, "Deep multimodal fusion for ground-based cloud classification in weather station networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 48, 2018.
- [17] S. Liu, M. Li, Z. Zhang, B. Xiao, and X. Cao, "Multimodal ground-based cloud classification using joint fusion convolutional neural network," *Remote Sens.*, vol. 10, no. 6, p. 822, 2018.
- [18] S. Liu and M. Li, "Multimodal GAN for energy efficiency and cloud classification in Internet of Things," *IEEE Internet Things J.*, to be published. doi: 10.1109/JIOT.2018.2866328.
- [19] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Jun. 2006, pp. 1735–1742.
- [20] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 815–823.
- [21] R. Zhang, P. Isola, and A. A. Efros, "Split-brain autoencoders: Unsupervised learning by cross-channel prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 645–654.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
- [23] W. Zhuo, Z. Cao, and Y. Xiao, "Cloud classification of ground-based images using texture–structure features," *J. Atmos. Ocean. Technol.*, vol. 31, no. 1, pp. 79–92, 2014.
- [24] S. Liu, Z. Zhang, and X. Mei, "Ground-based cloud classification using weighted local binary patterns," *J. Appl. Remote Sens.*, vol. 9, no. 1, 2015, Art. no. 095062.
- [25] Y. Wang, C. Shi, C. Wang, and B. Xiao, "Ground-based cloud classification by learning stable local binary patterns," *Atmos. Res.*, vol. 207, pp. 74–89, Jul. 2018.
- [26] A. Taravat, F. D. Frate, C. Cornaro, and S. Vergari, "Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 666–670, Mar. 2015.
- [27] S. Huang, Y. Jiang, Y. Gao, Z. Feng, and P. Zhang, "Automatic modulation classification using contrastive fully convolutional network," *IEEE Wireless Commun. Lett.*, to be published. doi: 10.1109/LWC.2019.2904956.
- [28] X. Liang, X. Wang, Z. Lei, S. Liao, and S. Z. Li, "Soft-margin softmax for deep classification," in *Proc. Int. Conf. Neural Inf. Process.*, Guangzhou, China, 2017, pp. 413–421.
- [29] J. Huang, Y. Li, J. Tao, and Z. Lian, "Speech emotion recognition from variable-length inputs with triplet loss function," in *Proc. INTERSPEECH*, Hyderabad, India, 2018, pp. 3673–3677.
- [30] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. Eur. Conf. Comput. Vis. Workshop Stat. Learn. Comput. Vis.*, Prague, Czech Republic, 2004, pp. 1–22.
- [31] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [32] Z. Guo and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jan. 2010.
- [33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [34] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Jun. 2006, pp. 2169–2178.



MEI LI is currently pursuing the master's degree with Tianjin Normal University, Tianjin, China. Her research interests include computer vision, pattern recognition, and deep learning.



SHUANG LIU (M'18) received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China. She is currently an Associate Professor with Tianjin Normal University, Tianjin, China. Her research interests include computer vision and deep learning.



ZHONG ZHANG (M'14) received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is currently an Associate Professor with Tianjin Normal University, Tianjin, China. He has published about 90 papers in international journals and conferences, such as the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Pattern Recognition*, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, *Signal Processing* (Elsevier), CVPR, ICPR, and ICIP. His research interests include computer vision, pattern recognition, and deep learning.

...