Cloud Type Classification Using Multi-modal Information Based on Multi-task Learning

Yaxiu Zhang[1,2], Jiazu Xie[1,2(✉)], Di He[1,2], Qing Dong[1,2], Jiafeng Zhang[1,2], Zhong Zhang[1,2], and Shuang Liu[1,2(✉)]

[1] Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin, China
veryaz@tjnu.edu.cn, shuangliu.tjnu@gmail.com
[2] College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin, China

Abstract. Cloud classification is an important and challenging task in cloud observation technology. For better classification, we present a method based on multi-task learning using multi-modal information. We utilize different loss functions to conduct multi-task learning. We implement a series of experiments on multi-modal ground-based cloud datasets for different tasks. Experimental results show that multi-task learning is effective for cloud image classification using multi-modal information, and it can improve the results of cloud image classification.

Keywords: Cloud type classification · Multi-modal information · Multi-task learning

1 Introduction

Cloud is an important meteorological phenomenon, and it affects the energy balance locally and globally by the interaction of solar and terrestrial radiation. There are some characteristics in the process of cloud formation and evolution, such as cloud water vapor, degree of stability, cloud height, cloud thickness and so on. These characteristics are all key features for predicting weather. Therefore, accurate cloud observation technology is of great significance in estimating precipitation, forecasting weather conditions, air traffic control etc. Cloud classification using cloud visual information (cloud images) is an important and challenging task in cloud observation technology. At present, the research on cloud image classification is mainly carried out on two kinds of data samples: satellite cloud images and ground-based cloud images. The latter can not only describe low cloud and regional cloud, but also reflect the texture information of cloud. Hence, cloud images classification method based on ground-based cloud has a strong application demand.

Q. Liang et al. (Eds.): CSPS 2021, LNEE 879, pp. 39–45, 2022.
https://doi.org/10.1007/978-981-19-0386-1_6

Many works have been done in this field. Liu et al. [1] extracted structural features such as cloud blocks, cloud mean and edge sharpness from edge images and segmentation images to realize the classification of ground-based clouds. Xiao et al. [2] take the fused features as the visual features of the clouds. Fusion features include: texture, structure, and color. Afterwards, because local binary patterns (LBP) [4] has the advantages of gray level invariance and rotational invariance, LBP is widely used in ground-based cloud classification. Oikonomou et al. [3] used improvement of LBP to extract local and global features of ground-based cloud images.

All the above methods are based on the manual features, and they do not work well for different distributed databases. At the present stage, researchers employ convolutional neural network to automatically classify cloud images. Automatic acquisition of high-level features from raw data is the most prominent advantage of convolutional neural network [5–7]. Therefore, valid information can be captured to a large extent. For example, Zhang et al. [8] proposed a significant dual activation clustering algorithm. The algorithm extracts the salient vectors from the shallow convolutional layer and the corresponding weights from the high convolutional layer. However, most existing methods ignore the fact that cloud formation and evolution are also related to multi-modal information. The mullti-modal information include temperature, humidity, air pressure, instantaneous wind speed, maximum wind speed, and average wind speed. The accuracy of ground cloud classification can be greatly improved by making full use of multi-modal information.

In this paper, we use multi-modal information to study the impact of different tasks on cloud classification:

- Verification task using triplet loss function.
- Classification task using cross-entropy loss function.
- Multi-task learning is realized by combining cross-entropy loss function and triplet loss function.

The paper is organized as follows: In Sect. 2, we describe the principles of the approach. The implementation details are provided in Sect. 3. We conclude in Sect. 4.

2 Approach

In this section, we first introduce the network we used. Secondly, we introduce the implementation of multi-task learning and the loss functions used in this paper.

2.1 Framework

The multi-modal fusion network combines visual sub-network and multi-modal sub-network. The framework is shown in Fig. 1. We treat resnet-50 [9] as the main part of the visual sub-network. The visual sub-network is used to extract

the deep visual features from cloud images. Because of the deep composition of resnet-50, the deep visual features contain complex nonlinear visual information. We treat the output of the averaging pool as a deep visual feature. The deep visual feature is a vector with a dimension of 2048. The multi-modal sub-network is constructed with multi-layer perceptron (MLP). The number of neurons in six FC layers of MLP is 64, 128, 256, 512, 1024 and 2048 respectively. The multi-modal feature dimension of the multi-modal sub-network corresponds to the feature dimension of deep vision. Finally, the deep multi-modal feature and deep vision feature are fused in series. The fused features are fed into the loss function through the fully connected layer.

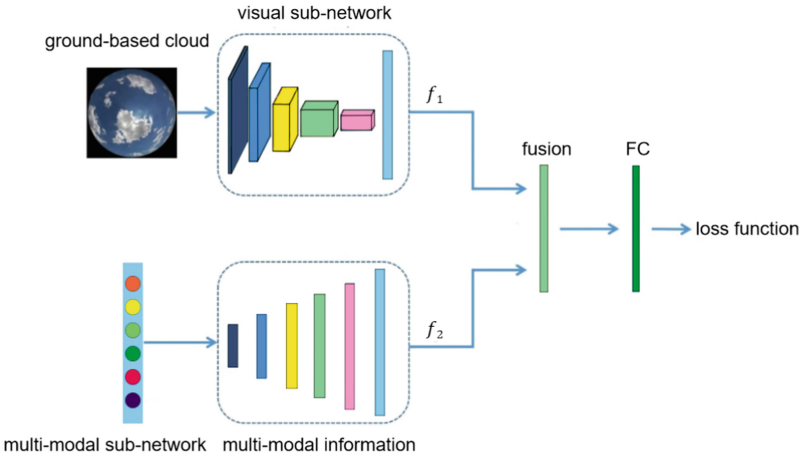

Fig. 1. The framework of the multi-modal fusion network.

2.2 Loss Function

The loss function can reflect the difference between the predicted results and the actual data. So it measures the quality of the model's predictions. The appropriate loss function is helpful to the subsequent model optimization. Here is the loss function used in this article:

Cross-Entropy Loss. The cross-entropy loss function is a loss function often used in classification problems. It evaluates the difference between the probability distribution given by current training and the true distribution. The ultimate goal of optimization is to make the trained probability distribution as close as possible to the real probability distribution. The formula is as follows:

$$L_{cross} = -\sum_{i=1}^{I} q_i \log y_i \tag{1}$$

where q_j is the probability of the actual probability distribution. When i is the true label, q_i is equal to 1, otherwise to zero.

Triplet Loss. The triplet loss is first proposed in the literature [10]. It could learn better embedding. Similar images are similar in embedded space, and different images are far apart in the embedding space. The input of triplet loss is $< a, p, n >$, where a is the chosen anchor, p is a sample of the same class as a and n is a different kind of sample from a. The final optimization goal is: close the distance between a and p and pull the distance between a and n. The metric function for measuring the distance of the embedding space is $D(x, y) : R^D \times R^D \rightarrow R$. The formula is as follows:

$$L_{tri} = \sum_{a,p,n,y_a=y_p\neq y_n} [m + D_{a,p} - D_{a,n}]_+. \tag{2}$$

This formula ensures that the distance from a sample of the same class to the anchor is at least m closer than that from a sample of a different class to the anchor in the mapping space. Such an approach allows all mapping points of the same class to eventually form a cluster without having to fold to a point. They just need to get closer to each other.

2.3 Multi-task Learning

We use the multi-loss function to conduct multi-task learning. The multi-loss function is combined cross-entropy loss function with triplet loss function. The formula is as follows:

$$L = L_{cross} + \alpha L_{tri} \tag{3}$$

where α is the weight parameter.

3 Experiments

In this section, we evaluate the effects of three loss functions on classification of multi-modal ground cloud images. Firstly, we introduce the dataset used in the experiment. Next, we describe detailedly the parameters of the experiment. The results and analysis of our experiment are presented in the end.

3.1 Dataset

The dataset we used is the multi-modal ground-based cloud images dataset (MGCD). The size of each image is 1536×1536. A cloud image and its corresponding multi-modal information constitute a multi-modal ground-based cloud sample. The samples are shown in Fig. 2. The MGCD includes seven categories: cumulus, altocumulus, cirrus, clearsky, stratocumulus, cumulonimbus and miexed.

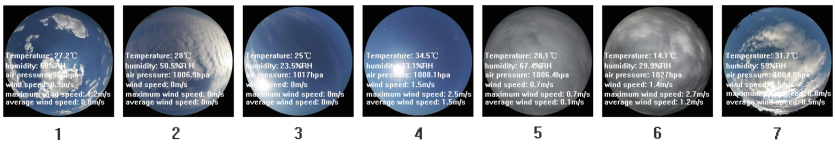

Fig. 2. Ground-based cloud samples.

3.2 Experiment Settings

The ground-based cloud images are normalized to 256×256. In addition, since our goal is to test the effect of different learning tasks on classification rather than to achieve optimal performance, we do not extend the data in our experiment. To improve the compatibility, the values of these multi-modal information are all mapped to the range of $[0-255]$. Then, the mapped value is subtracted from the corresponding mean value.

The initialization parameters for multi-modal fusion network include weight and bias. Weights are initialized to normal distribution and biases are initialized to zero. We employ Adam [12] to optimize the multi-modal fusion network. We set Epoch to 50. The size of each batch is 32. Each batch contains four categories of ground cloud images and each categorie contains eight ground cloud images. Set the learning rate to 3×10^{-4}. And according to the literature [11], we set m to 0.3.

3.3 Analysis

Results. In order to compare the ground-based cloud classification method used multi-task learning with other methods, we used three loss function separately. The network model, optimization method and parameters of the three experiments are consistent. The classification accuracy is shown in Table 1.

We apply the accuracy of the test results to evaluate the classification effect of different task learning. From Table 1, we can draw the following conclusion. Firstly, compared with other task learning mentioned above, the best classification accuracy is obtained by multi-task learning. Therefore, the validity of this method is verified. Secondly, we found that the cross-entropy loss function achieve final accuracy in the first few iterations. In subsequent iterations, they

Table 1. The accuracy of different learning task classification methods.

Method	Accuracy(%)
L_cross (Epoch=50)	83.12
L_tri (Epoch=50)	32.56
L (Epoch=50)	86.60
L (Epoch=200)	**89.43**

keep this accuracy up and down in the experiment. But it is worth noting that as the number of iterations increased to 200, the classification accuracy increased to 89.43% in the experiment of multi-tasking learning.

Parameter. The multi-task learning method presented in this paper involves an important parameter that appears in Eq. 3. This parameter is used to adjust the ratio between the cross-entropy loss function and triple loss function. The accuracy of cloud classification for multi-task learning can be optimized with appropriate settings. Figure 3 shows the classification accuracy when the α is set to different values. And the classification results are best when α is set to 0.05.

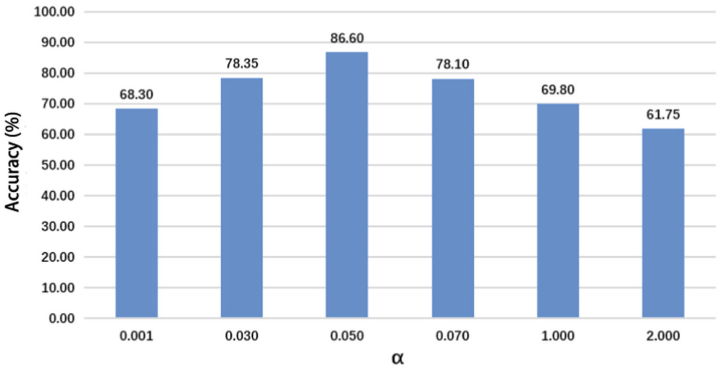

Fig. 3. Accuracy of different values of α

4 Conclusion

We evaluate the cloud type classification using multi-modal information based on multi-task learning in this paper. We compare the efficiency of multi-task learning with other task learning in the classification of multi-modal information ground-based cloud and prove that multi-task learning is effective for cloud type classification which use multi-modal information. In addition, we have done a lot of experiments based on multi-task learning.

Acknowledgement. This work was supported by College Student Research and Career-creation Program of Tianjin City for Undergraduates under Grant No. 202010065088, Natural Science Foundation of Tianjin under Grant No. 20JCZDJC00180 and No. 19JCZDJC31500, the Open Projects Program of National Laboratory of Pattern Recognition under Grant No. 202000002, and the Tianjin Higher Education Creative Team Funds Program.

References

1. Singh, M., Glennen, M.: Automated ground-based cloud recognition. Pattern Anal. Appl. **8**(3), 258–271 (2005)
2. Xiao, Y., Cao, Z., Zhuo, W., Ye, L., Zhu, L.: A multiview visual feature extraction mechanism for ground-based cloud image categorization. J. Atmos. Oceanic Tech. **33**(4), 789–801 (2016)
3. Oikonomou, S., Kazantzidis, A., Economou, G., Fotopoulos, S.: A local binary pattern classification approach for cloud types derived from all-sky imagers. Int. J. Remote Sens. **49**(7), 2667–2682 (2019)
4. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
5. Zhong, Z., Chunheng, W., Baihua, X., Wen, Z., Shuang, L., Cunzhao, S.: Cross-view action recognition via a continuous virtual path. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2690–2697 (2013)
6. Shuang, L., Mei, L., Zhong, Z., Baihua, X., Xiaozhong, C.: Multimodal ground-based cloud classification using joint fusion convolutional neural network. Remote Sens. **10**(6), 822 (2018)
7. Shuang, L., Linbo, Z., Zhong, Z., Chunheng, W., Baihua, X.: Automatic cloud detection for all-sky images using superpixel segmentation. IEEE Geosci. Remote Sens. Lett. **12**(2), 354–358 (2014)
8. Zhang, Z., Li, D., Liu, S.: Salient dual activations aggregation for ground-based cloud classification in weather station networks. IEEE Access **6**, 59173–59181 (2018)
9. Theckedath, D., Sedamkar, R.R.: Detecting affect states using VGG16, ResNet50 and SE-ResNet50. SN Comput. Sci. **1**(2), 1–7 (2020)
10. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)
11. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet Loss for person re-identification (2017). arXiv preprint, arXiv:1703.07737
12. Da, K.: A Method for Stochastic Optimization (2015). arXiv preprint, arXiv:1412.6980