Weather Recognition based on Attention Image Search Method

Yue Zhang[1,2], ChengZhang[1,2], ShichaoKan[1,2], YigangCen[1,2], LinnaZhang[3]
1 Institute of Information Science, Beijing Jiaotong University, Beijing, China
2 Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing, China
3 School of Mechanical Engineering, Guizhou University, Guiyang, China
{17112065,19125149,16112062,ygcen}@bjtu.edu.cn, zln770808@163.com

***Abstract.* Weather monitoring plays a vital role in intelligent traffic transportation, and the improvement of weather recognition accuracy can effectively improve driving safety. At present, classification-based and segmentation-based algorithms for weather recognition have achieved good performance, but it is still full of challenges in real applications. On the one hand, the number of classes in public data sets is insufficient, which cannot identify the conditions such as stagnant water and debris flow. On the other hand, the current weather recognition methods have poor generalization ability, the model needs to be retrained when classes are changed. In this paper, we first propose a new multi-traffic weather (MTW) data set for weather recognition, it contains much richer classes. Then, a new weather recognition method based on attention image retrieval (AIR) is proposed to improve the performance of recognition. Compared with the previous methods, our method can obtain better generalization performance.**

***Keywords:* weather recognition, image retrieval, attention**

I. Introduction

Weather recognition in natural scenes plays an important role in the field of intelligent traffic. Especially bad weather will not only greatly weaken the efficiency of transportation, but also directly cause traffic accidents, endangering people's lives and property. Different from general object recognition [1], [2], weather recognition is to recognize the entire image and requires an understanding of complex phenomena, such as light and reflection on the surface of the object. The current system relies on either a series of expensive sensors or on manual assistance to identify weather conditions. Since the weather condition varies locally from place to place in the same region, expensive sensors and a lack of human assistance limit the availability of local weather measurements.

Recently, weather recognition method based on computer vision has become more and more popular. The traditional weather image recognition method uses histogram features for representation [3], [4] and applies support vector machine (SVM) method for classification [5], [6]. With the development of deep learning, the accuracy of weather recognition based on deep learning has also been greatly improved, Cewu *et al.* [6] combined the data-driven CNN feature and well-chosen weather independent features to train a latent SVM classifier. This method is insensitive to global intensity transfer. Elhoseiny [7] focused on studying the feature spaces in the weather classification task. Lin *et al.* [8] proposed a region selection and concurrency model (RSCM) to help discover regional properties and concurrency for multi-class weather recognition. However, these methods need to retrain the model when adding or deleting some classes for different regions. Besides, the recognition accuracy for specific scenarios needs to be further improved to meet the requirements of actual traffic road weather recognition.

Existing datasets (e.g. TWI [6] and MWD [8]) include different scenes and can not identify some conditions of traffic jams caused by bad weather (i.e. stagnant water and debris flow). Thus, we propose a finer traffic scene weather data set called the multiple traffic weather (MTW) data set. Compared with the previous data sets, our data set contains more classes (11 classes) with 2,444 images, including sunny (norm), rainy (heavy rain, light rain), snowy (heavy snow, light snow), and haze (mist and dense fog), others(flood, debris flow, stgnant water, and landslide).

In order to meet the requirements of real scenarios, in addition to proposing more comprehensive data sets, it is necessary to build a more accurate and flexible recognition algorithm. The image recognition based on image retrieval method has been proved to be very effective in face recognition [2] and person re-identification [10]. Image retrieval method can effectively calculate the similarity between query image features and image features of the prior images, and further obtain output results according to the label of the nearest prior image. Concretely, we propose a weather recognition method based on attention image retrieval (AIR), which combines the ResNet network structure and the attention mechanism as the encoder to learn the features of each class. Then the classification loss and metric loss are employed to update the weight of the encoder. In the testing phase, we adopt the encoder to extract features and compare them with the features of the prior images to determine the weather class of the input image. Compared with the previous methods based on classification or segmentation, our method not only has a better recognition rate, but also has a stronger generalization ability such that new categories can be added without retraining the

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 2021QY002, in part by the National Natural Science Foundation of China under Grant 61872034, 62062021 and 62011530042, in part by the Beijing Municipal Natural Science Foundation under Grant 4202055, in part by the Natural Science Foundation of Guizhou Province under Grant [2019]1064. (Corresponding author: Yigang Cen.)

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

model.

The main contributions of this paper are as follows. We propose a new open-source traffic road weather data set called the multi-traffic weather (MTW), which contains richer and finer classes (11 classes).We propose a new weather recognition method based on attention image retrieval (AIR), and demonstrate its superiority on the public weather data set and our MTW data set.

II. Weather Recognition Data Set

In this section, we first revisit some existed weather recognition data sets, Martin *et al.* [3] developed a database with three classes, i.e. clear, light rain, and heavy rain. Then Cewu *et al.* [6] proposed two-classes weather images (TWI) data set, which includes 5,000 sunny and 5,000 cloudy images. Lin *et al.* [8] constructed a multiclass weather data set (MWD) with 65,000 images from 6 common weather categories, i.e., sunny, cloudy, rainy, snowy, haze, and thunder. In addition, Villarreal *et al.* [9] proposed a rain, fog, snow (RFS) data set collected online that depict scenes of three weather conditions, and each condition contains 1100 images. A comparison of weather images from different data sets shows that current data sets are limited in two ways: (1) The number of weather class is not enough. Moreover, the classification is not detailed enough. For example, dense fog and mist have different effects on traffic, which should be identified and given different warnings. (2) Only some common weather categories are included, which ignores road safety problems (i.e. stagnant water and debris flow) caused by bad weather.

In order to reduce the impact of bad weather on traffic, we construct a new open-source data set called the multi-traffic weather (MTW) data set. In our daily life, common weather conditions include sunny (norm), rainy (heavy rain, light rain), snowy (heavy snow, light snow), and haze (mist, dense fog). Common natural disasters that affect traffic include flood, stagnant water, debris flow, and landslide. By using the web crawler, we collected a total of 11 categories and 500 images for each class. After our screening, there were finally $3,254$ images left, including $2,444$ images for the training set, and 810 images for the test set. As shown in Table. I, we split the data set into the training set and testing set for each class. The ratio of the training set and the testing set is about $3:1$.

III. Methodology

A. Overall Network Structure

We propose a novel weather recognition framework based on the attention image retrieval method (AIR). In the training phase, as shown in Fig. 1, it includes an encoder and two losses. The images are input into the backbone network to extract the features. Due to the large difference in data scenes collected, the network needs to extract some features required to exist in a specific class. Therefore, before the global average pooling (GAP) of the benchmark network, we proposed a weather attention (WA) module to make the network pay more attention to the features of the weather. Then the obtained features are divided into two branches, one branch takes the classification loss, namely the cross-entropy loss. Another branch takes the metric loss. Here, we introduce the multi-similarity loss, which can effectively learn the relationship between positive and negative examples such that the network can learn robust features. In the test phase, the features obtained by the benchmark network are used to calculate the similarity with the prior images, and the class of the image is determined according to the class of its nearest neighbor in prior images.

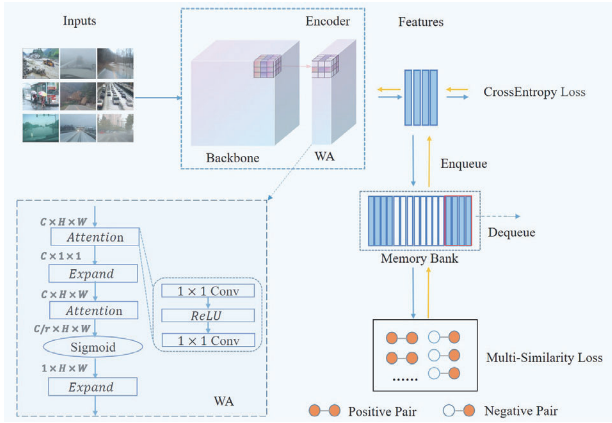

Fig. 1. The overall architecture of the training network, it includes an encoder and two loss functions

B. Weather Attention Module

In this work, in order to make the network focused on weather features, we propose a weather attention (WA) module, which considers channel attention (global feature) and spatial attention (local features) information. Specifically, our channel attention adopt an 1×1Conv-ReLU-1×1 Conv operation as the SENet [11]. The channel attention maps are dotted with the feature maps, their results are input into 1×1Conv-ReLU-1×1 Conv to reduce the channel dimensions. In Fig. 1, according to the enhancement of channel attention and spatial attention, the network can filter out some weather independent features that can improve the accuracy of weather recognition.

C. Loss Functions

As shown in Fig. 1, we extract feature maps from the backbone and WA module, then a softmax function is used to predict the classes of the input image features. The cross-entropy loss is employed to compute loss between predict label $\hat{y}$ and real label y, it is computed as:

$$L_{cls} = -\left[y log \hat{y} + (1-y)\, log\,(1-\hat{y}) \right]. \tag{1}$$

In addition, we introduce a pair-based metric loss function, the multi-similarity (MS) loss [12], to realize more efficient sample training through two iterations of sampling and weighting. For each sample pair, we need to consider not only the self-similarity of the sample pair itself, but also its relative similarity with other sample pairs. Therefore,

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

TABLE I
PROPOSED MULTI-TRAFFIC WEATHER (MTW) DATA SET, IT MAINLY INCLUDES FIVE CATEGORIES, INCLUDING RAIN, SNOW, HAZE, NORM, AND OTHERS. OTHERS REFER TO CATEGORIES THAT ARE NOT COMMONLY USED. IN ADDITION TO THE NORM CLASS, EACH CLASS CAN CONTINUE TO BE FINELY DIVIDED INTO DIFFERENT CLASSES.

class	snow		haze		rain		others				norm
	light snow	heavy snow	mist	dense fog	light rain	heavy rain	stgnant water	debris flow	flood	landslide	norm
train	146	333	136	34	269	134	436	154	355	149	398
test	48	111	44	11	89	44	144	51	118	50	100

TABLE II
WEATHER IDENTIFICATION RESULTS BASED ON DIFFERENT BENCHMARK NETWORK STRUCTURES ON THE PROPOSED MTW DATA SETS.

Methods		Accuracy(%)
Backbone	ResNet18	75.2
	ResNet34	75.6
	ResNet50	**76.5**
	ResNet101	76.3
Component analysis	baseline	76.5
	baseline+MS	77.0
	baseline+MS+Attention	**77.5**
Our AIR method	classification	77.5
	image search	**79.9**

important sample pairs can be adopted and weighted more efficiently and accurately. The MS loss is formulated as:

$$L_{MS} = \frac{1}{m}\sum_{i=1}^{m}\left\{\frac{1}{\alpha}log\left[1+\sum_{k\in P_i}e^{-\alpha(S_{ik}-\lambda)}\right] + \frac{1}{\beta}log\left[1+\sum_{k\in N_i}e^{\beta(S_{ik}-\lambda)}\right]\right\} \tag{2}$$

P_i reprsents a positive sample, and N_i represents a negative sample. m represents the sample number. λ, α, β are fixed hyper-parameters. S_{ik} represents the similarity of two features. Specifically, $S_{ik} :=< f(x_i;\theta), f(x_k;\theta) >$, where $< \cdot >$ denotes dot product, which results in an $m \times m$ similarity matrix S with the element at (i,k) is S_{ik}.

IV. EXPERIMENTAL RESULTS

A. Implementation Details and Evaluation Protocol

We perform our experiment on one 1080Ti GPU and PyTorch 1.1.0 platform. We adopt an ImageNet-pre-trained ResNet-X as the backbone, Adam is used with a learning rate 0.000003, and weight decays $5e^{-4}$. Each training mini-batch contains 10 images of 5 instances. For the k nearest neighbor (KNN), we choose k=1, which means the label of the nearest neighbor is the class of the tested image.

B. Comparison With the State-of-the-Art Methods

In this paper, for our proposed MTW data set, we first conduct some experiments on different backbone networks. The results are shown in Table. II. The recognition accuracy of the mentioned data set increases as the network deepens. But on the ResNet50 network framework, the result achieves the highest value of 76.5%. So for subsequent improvements, we will adopt ResNet50 as our baseline.

TABLE III
PERFORMANCE COMPARISON WITH THE STATE-OF-THE-ART METHODS ON TWI DATA SET.

Method		Accuracy (%)
Yan *et al.* [13]	ISNN(2009)	27.2
Roser *et al.* [14]	IV (2008)	28.5
Lalonde *et al.* [15]	IJCV (2012)	41.8
Lu *et al.* [6]	TPAMI (2017)	55.3
Elhoseiny *et al.* [7]	ICIP (2015)	92.9
SE-ResNet101 [11]	CVPR (2018)	94.5
CBAM-ResNet101 [16]	ECCV(2018)	94.6
BnInception	-	89.8
ResNet18	-	93.8
ResNet34	-	94.6
ResNet50	-	**94.7**
Our AIR method	-	**95.6**

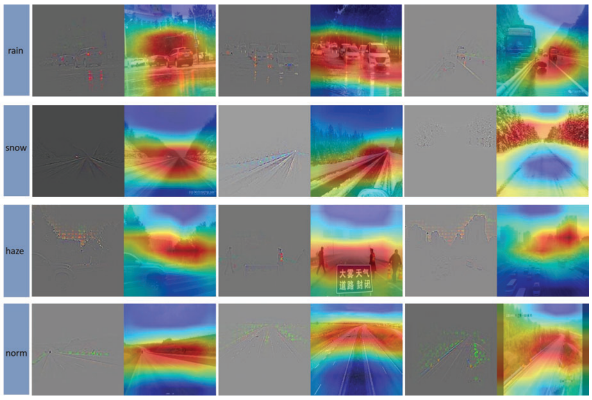

Fig. 2. Heat maps of attention of the proposed models in the proposed common weather categories. Each row represents a class, for each image, we show the GAP map (left) and Cam ++ heat map (right), respectively

Then we conduct a series of experiments and analyze the effectiveness of our proposed method in Table. II. Here, the backbone network is ResNet50, and the cross-entropy loss function is used to calculate the classification loss. The recognition accuracy rate is 76.5%. When different modules are added, the accuracy will be improved correspondingly. From Table. II, we can see that search recognition accuracy reaches 79.9%, which is improved 2.4% compared with the classification.

In order to verify the effectiveness of our proposed method, we test it on the TWI data set and compare it with some state-of-the-art recognition methods, the results are shown in Table. III. Here, we choose the highest results of the previous methods to compare with our result. We also conducted

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

TABLE IV
THE PROPOSED MTW DATA SET IS DIVIDED INTO DIFFERENT SUBSETS, AND THE SUBSETS ARE TRAINED TO TEST THE ACCURACY, RESPECTIVELY.

dataset	the number of the class	classes	Recall@1(%)	Accuracy(%)
CC	4	rain snow haze norm	93.5	93.1
CF	7	heavyrain lightrain heavysnow lightsnow densefog mist norm	85.9	87.0
UF	5	debris-flow flood landslide stgnant-water norm	86.6	87.8

some experiments on different ResNet networks, including two networks with attention mechanisms (i.e. SE and CBAM). From Table. III, we can see that the ResNet101 network structure combines with different attention mechanisms have achieved good performance. And ResNet50 achieved a better classification result than other ResNet networks. Our AIR method achieves the highest accuracy, which is superior to the current best results of about 0.9%. It demonstrates the effectiveness of the proposed AIR method.

C. Discussion

In addition, the proposed MTW data set is divided into different subsets: common coarse classification (CC), common fine classification (CF), and uncommon fine classification (UF), and the subsets are trained to test the accuracy, respectively. As shown in the Table.IV, in the CC subset, there exists four classes: norm, rain, snow, and haze. Experimental results show that the model can still get considerable performance for classes that have never been trained. In addition, we draw the GAP (global average pooling) map and Grad-CAM++ map of common weather classes images, respectively. As shown in Fig. 2, the first row shows images in the rain class, and the model focuses on the foreground and where the road reflects light. The second row shows the images in the snow class, for which the model focuses on the road-side position. In the third row and the fourth row, the model focuses on the upper part of the road, the attention region includes the sky. This demonstrates that our model will pay attention to different information, so as to distinguish different categories.

V. CONCLUSIONS

In this paper, we first proposed a multi-traffic weather data set, which mainly concentrates on traffic road scenarios and contains richer and finer classes. It enables new experiments both for training better models and as a new benchmark. Then, we proposed a new weather recognition method based on the attention image retrieval (AIR) method. It effectively improves the accuracy of weather recognition, and can be widely used in different weather recognition scenarios. In other words, it is well extensible and does not require retraining the model when adding or subtracting some classes. Our future work will continuously improving and supplementing our data set. In addition, we need to further explore the solution of some images containing multiple classes.

REFERENCES

[1] Kan, S., Cen, L., Zheng, X., Cen, Y., Zhu, Z., Wang, H. : A supervised learning to index model for approximate nearest neighbor image retrieval. Signal Processing: Image Communication, 78, 494–502 (2019).

[2] Li, S., Liu, B., Chen, D., Chu, Q., Yuan, L., Yu, N. : Density-Aware Graph for Deep Semi-Supervised Visual Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13400-13409 (2020).

[3] Roser, M., Moosmann, F. : Classification of weather situations on single color images. In 2008 IEEE Intelligent Vehicles Symposium, pp. 798-803 (2008, June). IEEE.

[4] Yan, X., Luo, Y., Zheng, X. : Weather recognition based on images captured by vision system in vehicle. In International Symposium on Neural Networks,pp. 390-398 (2009, May). Springer, Berlin, Heidelberg.

[5] Chen, Z., Yang, F., Lindner, A., Barrenetxea, G., Vetterli, M. : How is the weather: Automatic inference from images. In 2012 19th IEEE International Conference on Image Processing pp. 1853-1856 (2012, September). IEEE.

[6] Lu, C., Lin, D., Jia, J., Tang, C. K. : Two-class weather classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3718-3725 (2014).

[7] Elhoseiny, M., Huang, S., Elgammal, A. : Weather classification with deep convolutional neural networks. In 2015 IEEE International Conference on Image Processing (ICIP), pp. 3349-3353 (2015, September). IEEE.

[8] Lin, D., Lu, C., Huang, H., Jia, J. : RSCM: Region selection and concurrency model for multi-class weather recognition. IEEE Transactions on Image Processing, **26**(9), 4154-4167 (2017).

[9] Guerra, J. C. V., Khanam, Z., Ehsan, S., Stolkin, R., McDonald-Maier, K. : Weather Classification: A new multi-class data set, data augmentation approach and comprehensive evaluations of Convolutional Neural Networks. In 2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS), pp. 305-310 (2018, August). IEEE.

[10] Zhang, Y., Jin, Y., Chen, J., Kan, S., Cen, Y., Cao, Q. : PGAN: Part-based nondirect coupling embedded GAN for person reidentification. IEEE MultiMedia, **27**(3), 23-33 (2020).

[11] Hu, J., Shen, L., Sun, G. : Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132-7141 (2018).

[12] Wang, X., Han, X., Huang, W., Dong, D., Scott, M. R. : Multi-similarity loss with general pair weighting for deep metric learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5022-5030 (2019).

[13] Yan, X., Luo, Y., Zheng, X. : Weather recognition based on images captured by vision system in vehicle. In International Symposium on Neural Networks, pp. 390-398 (2009, May). Springer, Berlin, Heidelberg.

[14] Roser, M., Moosmann, F. : Classification of weather situations on single color images. In 2008 IEEE Intelligent Vehicles Symposium, pp. 798-803 (2008, June). IEEE.

[15] Lalonde, J. F., Efros, A. A., Narasimhan, S. G. : Estimating the natural illumination conditions from a single outdoor image. International Journal of Computer Vision, 98(2), 123-145 (2012).

[16] Woo, S., Park, J., Lee, J. Y., Kweon, I. S. : Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), pp. 3-19 (2018).

[17] Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V. N.: Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) pp. 839-847 (2018, March). IEEE.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.