Neural Computing and Applications (2021) 33:11107–11123
https://doi.org/10.1007/s00521-020-05650-8

ORIGINAL ARTICLE

Graph convolutional networks with attention for multi-label weather recognition

Kezhen Xie[1] · Zhiqiang Wei[1,2] · Lei Huang[1,2] · Qibing Qin[1] · Wenfeng Zhang[1]

Received: 7 July 2020 / Accepted: 16 December 2020 / Published online: 5 January 2021

Abstract

Weather recognition is a significant technique for many potential computer vision applications in our daily lives. Generally, most existing works treat weather recognition as a single-label classification task, which cannot describe the weather conditions comprehensively due to the complex co-occurrence dependencies between different weather conditions. In this paper, we propose a novel Graph Convolution Networks with Attention (GCN-A) model for multi-label weather recognition. To our best knowledge, this is the first attempt to introduce GCN into weather recognition. Specifically, we employ GCN to capture weather co-occurrence dependencies via a directed graph. The graph is built over weather labels, where each node (weather label) is represented by word embeddings of a weather label. Furthermore, we design a re-weighted mechanism to build weather correlation matrix for information propagation among different nodes in GCN. In addition, we develop a channel-wise attention module to extract informative semantic features of weather for effective model training. Compared with the state-of-the-art methods, experiment results on two widely used benchmark datasets demonstrate that our proposed GCN-A model achieves promising performance.

Keywords Multi-label weather recognition · Weather co-occurrence dependencies · Directed graph · Attention

1 Introduction

Weather has a great significance to human activities and affects our daily lives in many ways, ranging from solar technology to outdoor sports, dressing, traveling, etc [1, 2]. Current weather detection is largely dependent on sensors, which are normally expensive and require professional maintenance [3, 4]. As surveillance cameras become almost ubiquitous [5–7], weather detection may turn to be a powerful and cost-effective computer vision application [1]. Additionally, some robotic vision applications have been benefited with better weather understanding [8, 9]. For example, assistant driving system relies on weather understanding to make more reasonable decisions [10, 11].

Despite its remarkable value, weather recognition from a single outdoor image has not been thoroughly studied. In the previous literatures, most of weather recognition researches could be divided into two categories according to scene: scene-specific weather recognition methods [11–13] and scene-free weather recognition methods [1, 3, 14, 15]. Both of them mainly focus on two-class weather classification and multi-class weather classification.

Although the above works have achieved good performances, they treat weather recognition as a simple single-label classification task, which limits the practical application. A single label cannot describe the weather comprehensively because some weather conditions occur simultaneously due to the complexity of weather itself [16]. For example, in Fig. 2, the changes from Fig. 2a to Fig. 2f display a series of states between an obvious sunny weather (Fig. 2a) and a pure cloudy weather (Fig. 2f). It's difficult to recognize whether it's sunny or cloudy referring to intermediate states (i.e., Fig. 2c, d and e) since they show both sunny and cloudy.

✉ Lei Huang
huangl@ouc.edu.cn

[1] College of Information Science and Engineering, Ocean University of China, Qingdao 266000, China

[2] Pilot National Laboratory for Marine Science and Technology, Qingdao, China

Thus, considering the co-occurrence of different weather conditions, it is of great necessity to study multi-label weather recognition for a comprehensive description of the weather in an image. Comparing to single-label weather classification, multi-label weather recognition is more challenging due to the combinatorial nature of the output space. Since weather conditions normally co-occur in the physical world, the key for multi-label weather recognition is to model the weather co-occurrence dependencies, as shown in Fig. 1. This stimulates researches to capture and explore weather dependency relationships, but there are still few approaches so far. Recently, some researchers proposed Recurrent Neural Networks (RNNs)-based approaches [4] to explicitly model weather dependencies for multi-label weather recognition. However, they predict the weather labels step by step in a sequential fashion and require the pre-defined fixed orders, which cannot exploit the weather dependencies thoroughly.

Therefore, how to explore the inherent co-occurrence dependency patterns of different weather conditions and perform accurate weather recognition is still a very challenging issue.

In this paper, inspired by the success of GCN in modeling dependencies [17–19], we propose a novel end-to-end trainable deep learning model: Graph Convolution Networks with Attention (GCN-A) to capture and explore more important weather dependencies for multi-label weather recognition, which is flexible and scalable. To the best of our knowledge, this is the first attempt to employ GCN to model weather dependencies. **Firstly**, a directed graph is constructed over weather labels, which utilizes the word embeddings of weather labels as nodes. The edge between two nodes is built by weather co-occurrence information. We propose to map this graph into a set of inter-dependent weather classifiers via GCN learning, where each node corresponds to a weather classifier. This implicitly models the weather dependencies through weather labels. The generated inter-dependent weather classifiers are then applied to weather features extracted from images to enable end-to-end training. **Secondly**, most previous works built the weather correlation matrix with binarization to alleviate over-fitting problem. However, the binarized correlation matrix may lead to over-smoothing since the updated node in GCN is the weighted sum of the representations of the node itself and its neighbor nodes. To alleviate both over-fitting problem and over-smoothing problem, we design a re-weighted mechanism to build an effective weather correlation matrix. The constructed matrix guides information propagation among nodes in GCN, which explicitly models weather dependencies for GCN training. **Thirdly**, most researches directly utilize convolutional neural network to extract weather representations while ignoring that different image regions are of different importance for weather recognition. To address this problem, we develop a channel-wise attention module for the model to adaptively recalibrate feature responses to exploit more discriminative semantic features of weather for effective learning.

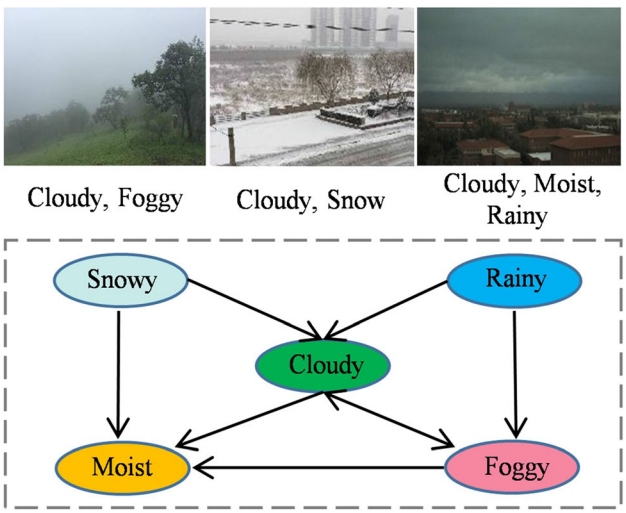


Fig. 1 We employ GCN to model weather co-occurrence dependencies via a directed graph over the weather labels for multi-label weather recognition. In this figure, "$A \rightarrow B$" stands for that when weather A occurs, weather B is likely to occur, but the reverse may not be true

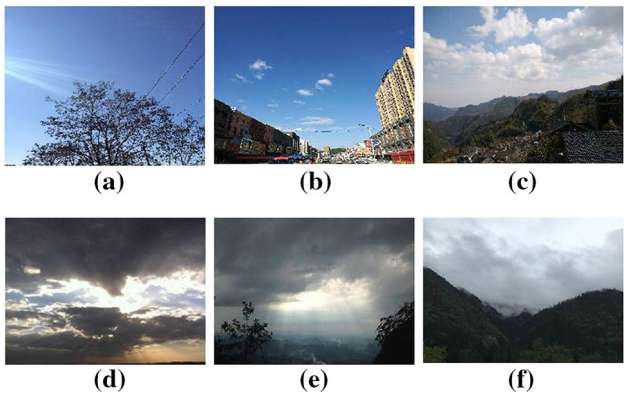


Fig. 2 Several images from [4]. **a** Pure sunny weather without clouds. **b** Sunny weather with a few clouds. **c** Sunny weather with more clouds, compared with (**b**). **d** and **e** display both sunny and cloudy weather. **f** Obvious cloudy weather

Our primary contributions in this paper are summarized as follows:

(1) We propose a novel end-to-end Graph Convolution Networks with Attention (GCN-A) model for multi-label weather recognition, which takes the first attempt to employ GCN to explore and capture important weather co-occurrence dependencies.

(2) We design a re-weighted mechanism to build effective weather correlation matrix to guide information propagation of nodes in GCN, which can alleviate over-fitting and over-smoothing problems.
(3) We develop a channel-wise attention module for GCN-A model to extract informative semantic representations of weather for effective model training.

The rest of this paper is organized as follows: In Sect. 2, the related works about weather recognition and multi-label image recognition are reviewed. In Sect. 3, the proposed method is introduced in details. In Sect. 4, implementation details and experimental results are presented. In Sect. 5, the conclusion of this work is provided.

2 Related work

In this section, we present some classic and state-of-the-art works on weather recognition. Several of the most popular related researches on multi-label image recognition are also reviewed.

2.1 Weather recognition from an image

In this paper, we roughly classify weather recognition researches into two categories: recognition with designed hand-crafted weather features and recognition with CNN extracted weather features.

2.1.1 Weather Recognition with Hand-crafted Features

Weather recognition plays a vital role in assistant driving systems. Recognizing weather conditions in real-time automatically can enhance driving safety [12, 20–22] by controlling driving speed in time, opening the car's windshield wipers, etc. Therefore, some researches focus on the weather recognition for images captured by in-vehicle cameras. Since raindrop is a critical clue for rainy weather, various approaches are proposed in [10, 12, 20, 23] to recognize rainy weather from in-vehicle images. Specifically, Roser and Moosmann [12] divide the image into thirteen regions equally and extract local descriptors in each region via feature histograms to detect raindrops. Kurihata [10, 23] and Yan [20] proposes methods to detect raindrops by template matching for rainy weather judgment. In fact, it not only recognizes rainy weather in [20], but also distinguish cloudy and sunny weathers through global features consists of HSV color histogram, gradient amplitude histogram and road information. Additionally, fog and haze detection are studied from in-vehicle images since they also affect driving safe directly. In [11, 22], Gabor filters are utilized on the power spectrum of the image for fog detection. While Bronte et al. [24] decides whether it is foggy based on image edges detection via Sobel filter.

Apart from the works on weather recognition from in-vehicle images, there are also a lot of weather recognition researches concentrate on generic outdoor images. Lu et al. [1] designs multiple local features, such as sky, reflection and shadow to achieve two-class weather recognition, while Li et al. [25] estimates 4 weather conditions by extracting a set of global features, combining SVM and decision tree as weather classifier. Song et al. [26] also exploits a series of global features, i.e., inflection point information, image noisy, edge gradient energy, power spectral slope, contrast, saturation, to recognize 4 weather conditions. Unlike previous work, in [14, 15], both local and global features are explored to tackle the problem of multi-class weather classification. In addition, a multiple kernel learning method is developed in [15] to fuse these features.

Though the researches have designed various hand-crafted features for weather recognition and they have shown good performance in their specific applications, the features are usually constrained to specific weather conditions or perspectives, which cannot be applied to general applications.

2.1.2 Weather recognition with CNNs

More recently, deep CNNs have made a series of breakthroughs in computer vision tasks, e.g., image classification [27], semantic segmentation [28] and object detection [29]. Excellent CNN architectures, such as AlexNet [27], VGG [30] and ResNet [31] have been designed and outperform traditional methods. Inspired by the overwhelming performance of deep CNNs, several researchers introduced CNNs to weather recognition tasks. Specifically, Elhoseiny et al. [32] have applied AlexNet [27] to the task of two-weather classification and obtained superior performance compared with [1]. Shi et al. [33] utilize VGG [30] to extract features of the original image and the foreground of the image for weather recognition, which effectively improves the four-weather classification accuracy. Both VGG [30] and ResNet [31] are evaluated for weather recognition on the proposed dataset in [34]. The VGG-based method achieves the best performance for sunny weather, while the ResNet based method works best for the other weather conditions. Besides, Lin et al [16] proposed a deep learning framework based on CNN for weather recognition, and achieved a better result on multi-class weather dataset. Furthermore, in [3], CNN extracted features and hand-crafted features are integrated for two-

weather classification, and has achieved the state-of-the-art.

As discussed in [3, 4], different kinds of weather conditions may co-occur due to the complexity of the weather phenomenon itself. Thus, there exists information loss when describe weather conditions by the aforementioned approaches, viewing weather recognition as simple single-label classification problem. Li et al. [2, 35] proposed multi-task frameworks to provide auxiliary semantic segmentation of weather cues for comprehensive weather description. The frameworks alleviate the problem of information loss, while the auxiliary information is not intuitive for humans. Zhao et al. [4] made an attempt to treat weather recognition as a multi-label classification task, and proposed a CNN-RNN architecture for it. CNN is utilized for weather feature extraction, while RNN mining the weather dependency relationships among weather categories. However, RNN predicts weather labels in a sequential manner and relies on the pre-defined order of weather classes, which is nonflexible.

2.2 Multi-label image recognition

Recognizing multiple labels of images is a practical and challenging task in computer vision and multimedia fields. Plenty of researches on multi-label recognition have been carried out and remarkable progress has been achieved in the past few years. We briefly review some classical works on the multi-label image recognition methods.

Recently, some researchers have studied multi-label learning based on theoretical technology. Specifically, to exploit the relationship between features and labels, Ma et al. [36] propose a three label-wise relationship (feature-label, instance-label and label–label correlations) learning approach with considering the semantic gap. By constructing a low-dimensional embedding based on the original feature space to capture label correlations, Zhang et al. [37] develop an embedded multi-label feature selectin algorithm with manifold regularization. By extracting label-specific features in local and global levels and training label-specific classifiers in individual and joint levels, Ma et al. [38] present a unified framework combining two-level label-specific features and two-level specific classifiers for multi-label classification.

With the fast development of deep convolutional networks [31, 39–42] in these years, lots of efforts have been made to extending deep convolutional networks for multi-label image recognition. Most of such approaches train independent binary classifiers for each class, which does not take the relationships between classes into consideration. However, these approaches are limited because of ignoring the topology structure among different classes [19]. The topology structure can reflect the object co-occurrence patterns of different classes. Thus, many researches concentrate on exploring the object co-occurrence dependencies (i.e., class dependencies). Specifically, in [43], multi-label image recognition is considered as a sequential prediction problem, and the label dependencies are learned utilizing RNN via transforming labels into embedded label vectors. Additionally, attention mechanisms are also deployed to capture label correlation for multi-label image recognition. In [44], a spatial regularization network is proposed to exploit both semantic and spatial relations between labels based on weighted attention maps. Besides, Wang et al. [45] developed a recurrent memorized-attention module, utilizing a spatial transformer layer and a Long Short-Term Memory (LSTM) subnetwork to capture label dependencies.

Compared with the structure learning methods mentioned above, modeling label correlations via graph has been proven to be a more effective approach [19]. Li et al. [46] proposed Conditional Graphical Lasso (CGL) to produce image-dependent conditional label structures for Multi-label Image Classification. Lee et al. [47] developed a model incorporates structured knowledge graphs to describe the relations between multiple labels. In this paper, we propose to build a directed graph structure to explore weather co-occurrence dependencies. Specifically, based on the graph, GCN is employed to propagate information among weather labels and consequently learn a set of inter-dependent classifiers for each of weather categories. The classifiers absorb the dependency information from the label graph. Then, they are further applied to weather representations extracted from images for multi-label weather recognition.

3 Method

In this section, we firstly present the overall framework of the proposed model for multi-label weather recognition. Then, we illustrate the GCN model in details for exploring weather dependencies, which is followed by the description of weather correlation matrix construction. Finally, we depict the channel-wise attention in the model to extract informative weather features for effective training.

3.1 The overall framework

In order to describe weather conditions comprehensively, we propose a Graph Convolutional Networks with Attention (GCN-A) model for multi-label weather recognition. The overall framework is shown in Fig. 3, consisting of two major parts, i.e., graph convolutional network and the semantic representation learning network.

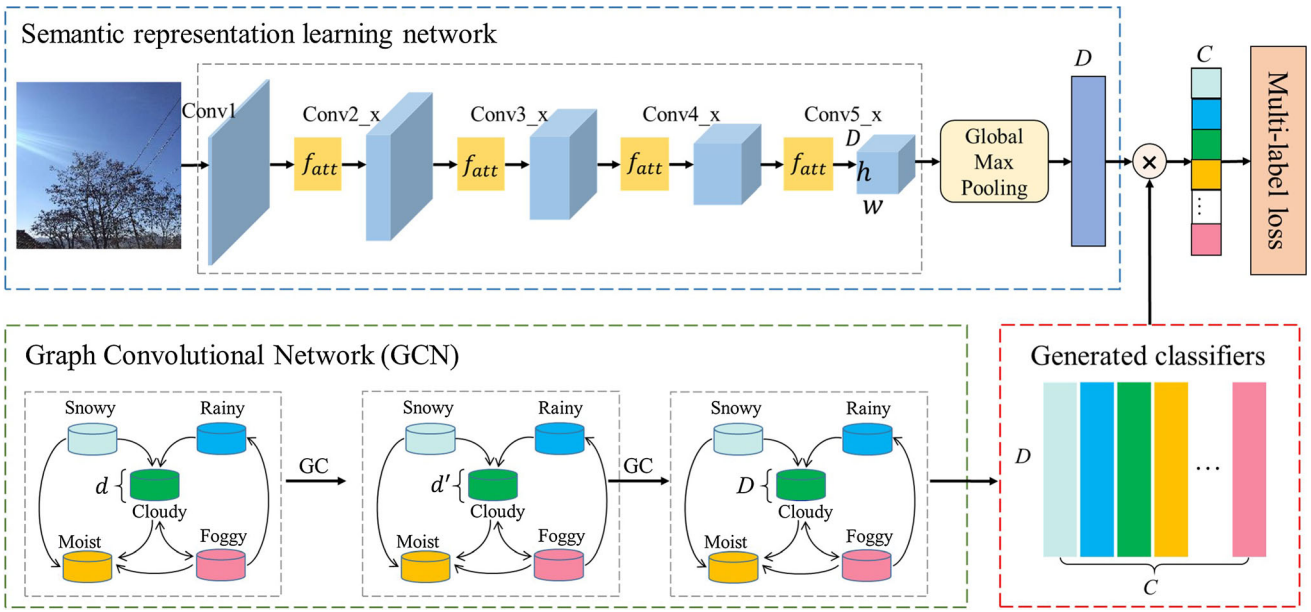


Fig. 3 Overall framework of our GCN-A model for multi-label weather recognition. The weather labels are represented by word embeddings ($\boldsymbol{V} \in \mathbb{R}^{C\times d}$, where C denotes the number of weather categories and d means the dimensionality of word-embedding vector). It constructs a directed graph over the weather label representations in GCN model and each node of the graph represents a weather label. GCNs are learned over the graph to map the weather labels representations into a set of inter-dependent weather classifiers, i.e., $\boldsymbol{W} \in \mathbb{R}^{C\times D}$ (where D represents the dimensionality of the image representation). Then, the classifiers are applied to informative weather features extracted from the input image through an attention-based convolutional network for multi-label weather recognition

In this paper, GCN is employed to model weather co-occurrence dependencies for multi-label weather recognition via a directed graph. In order to better define the graph structure, we propose to construct an effective weather correlation matrix with a reweighted mechanism. GCNs are learned to model dependencies between weather conditions in the graph and map the graph into a set of inter-dependent classifiers. The classifiers are then applied to weather representations extracted from the input images. Channel-wise attention is proposed to adaptively weight different channels and recalibrate the feature responses to extract more informative weather representations for effective GCN training.

3.2 GCN for multi-label weather recognition

How to explore and capture the inter-dependencies between different weather conditions is significant for multi-label weather recognition. Fortunately, in recent years, Graph Convolutional Network (GCN) has been proved effective at tasks with rich relational structure. It can preserve global structure information of a graph in graph embeddings [17]. In this paper, we employ GCN to model the correlations between different weather labels via a graph to capture the topological structure in the label space flexibly. Specifically, the nodes of the graph are represented by weather label word embeddings, and GCN is trained to map the embeddings into a set of inter-dependent weather classifiers for multi-label weather recognition. The motivations to propose GCN to model weather dependencies for multi-label weather recognition are the following factors. On one hand, the GCN learned inter-dependent weather classifiers can keep weak semantic structure in word embedding space, in which semantic-related concepts are close to each other. Therefore, the classifiers generation function can implicitly model the weather label dependencies. On the other hand, GCN can explicitly model the weather dependencies with weather correlation matrix via updating nodes representations with the absorbed information from correlated nodes.

Graph Convolutional Network recap. The essence of GCN is to update the node representations by propagating information between nodes [19]. GCN aims to learn a function $f(\cdot,\cdot)$ on a graph. The function takes nodes representations $\boldsymbol{X}^l \in \mathbb{R}^{n\times d}$ and corresponding correlation matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ as inputs, where n is the number of nodes and d denotes the dimensionality of node representations, and then updates the nodes representations to $\boldsymbol{X}^{l+1} \in \mathbb{R}^{n\times d'}$. Each GCN layer can be represented as a nonlinear function as follows:

$$\boldsymbol{X}^{l+1} = f(\boldsymbol{X}^{l}, \boldsymbol{M}) \tag{1}$$

With the convolution operation of [48], the function can be rewritten as:

$$\boldsymbol{X}^{l+1} = h(\hat{\boldsymbol{M}}\boldsymbol{X}^{l}\boldsymbol{T}^{l}) \tag{2}$$

where $h(\cdot)$ represents a nonlinear operation, $\boldsymbol{T}^{l} \in \mathbb{R}^{d \times d'}$ denotes the transformation matrix which will be learned during training stage, and $\hat{\boldsymbol{M}} \in \mathbb{R}^{n \times n}$ is the normalization of matrix $\boldsymbol{M}$. Therefore, we can stack multiple GCN layers to learn and model inter-relationships between the weather nodes. For more details, we refer interested readers to [48].

GCN-based weather classifier learning. The inter-dependent weather classifiers, i.e., $\boldsymbol{W} = \{w_i\}_{i=1}^{C}$ (where C represents the number of weather classes) are learned from label representations through the mapping function based on GCN. Stacked GCN layers are employed and each layer l takes the node representations $\boldsymbol{X}^{l}$ of previous layer as input and generates updated node representations $\boldsymbol{X}^{l+1}$ as output. The first layer takes the weather label embeddings as input and the last layer outputs $\boldsymbol{W} \in \mathbb{R}^{C \times D}$, where D represents the dimensionality of weather feature representations extracted from images.

The predicted category score can be obtained via applying the learned classifiers to image representations as follows:

$$\hat{y} = \boldsymbol{W}x \tag{3}$$

where $\boldsymbol{W}$ is the classifiers and x denotes the image representations.

Assume that the images' ground-truth label is $y \in \mathbb{R}^{C}$, where $y^{i} = \{0, 1\}$. 1 indicates the corresponding weather condition appears in the image and 0 indicates not. The traditional multi-label classification loss is used for the multi-label weather recognition network as follows:

$$L = \sum_{c=1}^{C} y^{c} log(\sigma(\hat{y}^{c})) + (1 - y^{c}) log(1 - \sigma(\hat{y}^{c})) \tag{4}$$

where $\sigma(\cdot)$ represents the *sigmoid* function.

In fact, GCN learning also promotes feature extraction. Because the parameters of feature extraction network (i.e., the semantic representation learning network) are optimized by the loss which is calculated by dot production between the classifiers learned by GCN and the weather features. In the backpropagation of training stage, not only the parameters in the GCN network will be updated, but also the parameters in the feature extraction network will be updated simultaneously. Therefore, GCN also contributes to feature extraction during training phase.

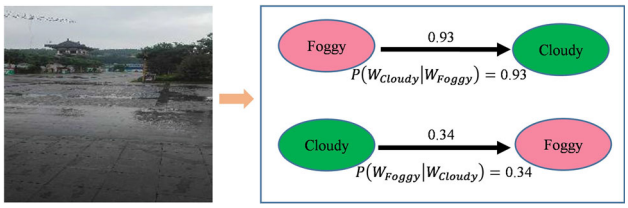


Fig. 4 Illustration of the conditional probability between two weather conditions. As normal, "Cloudy" will occur with a high probability in the condition of "Foggy" appearing in the image while "Foggy" will not necessarily occur when "Cloudy" occurs in the image

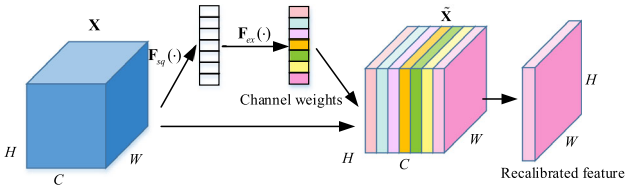


Fig. 5 The illustration of weather feature recalibration by channel-wise attention in our model. C denotes the channels of the features, W and H are width and height of the feature maps, respectively. It takes the visual feature as input and calculate channel-wise attention weight for each channel via squeeze and excitation. The feature maps are then scaled with the weights and reweighted to the recalibrated feature, which is the input of next convolution layer

In testing phase, semantic features of weather are extracted by the semantic representation learning network which is optimized during training stage.

3.3 Weather correlation Matrix of GCN-A

How to construct weather correlation matrix is crucial for weather dependencies modeling since GCN works via information propagation among nodes based on the matrix. In this paper, we construct the weather correlation matrix through mining the weather co-occurrence patterns in the dataset.

We use conditional probability $P(W_j|W_i)$ to simulate weather dependencies, which represents the probability that weather W_j will occur when the weather W_i occurs. As shown in Fig. 4, $P(W_j|W_i)$ does not equal to $P(W_i|W_j)$, which means the weather co-occurrence relationship matrix is asymmetric. To build the correlation matrix of training dataset, we firstly count the occurrences of each weather pair and build the concurrence matrix as follows:

$$cooc(W_i, W_j) = \begin{cases} 1, & Label(W_i) = 1 \wedge Label(W_j) = 1 \\ 0, & others \end{cases} \tag{5}$$

$$\boldsymbol{N}_{ij} = \sum_{\Omega} cooc(W_i, W_j) \tag{6}$$

where $Label(W_i) = 1$ denotes that weather W_i (i.e., label i) appears in an image, and $cooc(W_i, W_j)$ indicates whether weather W_i and weather W_j co-occur or not, 1 means two weather conditions appear simultaneously and 0 means not. Ω denotes all the samples in the dataset. $\boldsymbol{N}_{ij}$ represents the number of times that weather W_i and weather W_j co-occur in image. $\boldsymbol{N} \in \mathbb{R}^{C \times C}$, where C is the number of weather conditions (i.e., weather labels) in the dataset. Therefore, we calculate the conditional probability matrix using the weather co-occurrence matrix as follows:

$$\boldsymbol{P} = \boldsymbol{N}_i / T_i \tag{7}$$

where T_i means the total number of times that weather W_i appears in the training dataset, and $\boldsymbol{P}_{ij} = P(W_j|W_i)$ indicates the occurrence probability of weather W_j when weather W_i occurs.

The weather occurrence distribution in training dataset and testing dataset may be inconsistent, which may result in the correlation matrix over-fitting to the training dataset and influence the generalization capability. Additionally, some weather pair occurrences may show a long tail distribution, in which the few co-occurrences may be noise. Thus, we binarize the correlation matrix $\boldsymbol{P}$ with a threshold t to filter the noise as follows:

$$\boldsymbol{B}_{ij} = \begin{cases} 0, & \boldsymbol{P}_{ij} < t \\ 1, & \boldsymbol{P}_{ij} \geq t \end{cases} \tag{8}$$

where $\boldsymbol{B}$ is the binary matrix of weather correlations.

It can be summarized from Eq. (2) that the new representation of a node after a GCN layer is the weighted sum of its own representation and the representations of its related neighbor nodes, which may cause over-smoothing and lead to the difficulty to distinguish different nodes using the binary correlation matrix.

Thus, we propose to re-weight the binary weather correlation matrix as follows:

$$\boldsymbol{B'}_{ij} = \begin{cases} w / \sum_{j=1, i \neq j}^{C} \boldsymbol{B}_{ij}, & i \neq j \\ 1 - w, & i = j \end{cases} \tag{9}$$

where $\boldsymbol{B}'$ denotes the re-weighted matrix, and w determines the weights assigned to the node itself and its adjacent nodes.

With the re-weighted weather correlation matrix, when updating a node's representation, a fixed weight will be assigned to the node itself, and its adjacent nodes' weights will be calculated by the neighborhood distribution. We can conclude from Eq. (9) that the weight of the node itself will be smaller as w approaches 1. However, the weights of it's neighbor nodes will be smaller as w approaches 0 when updating the node representation.

3.4 Channel-wise attention for GCN-A

Following [19, 47, 49], we utilize ResNet-101 model [31] as the backbone to learn weather representations, which is pre-trained on ImageNet [50]. Thus, $2048 \times 7 \times 7$ feature maps can be obtained from "*conv5_x layer*" if the resolution of an input image is 224×224. Then, global max-pooling is employed to get the image-level weather feature x:

$$\boldsymbol{x} = f_{pooling}(f_{cnn}(\boldsymbol{I}, p_{cnn})) \in \mathbb{R}^D \tag{10}$$

where $f_{pooling}$ is the function of global max-pooling and f_{cnn} indicates the ResNet-101 model. p_{cnn} denotes ResNet-101 model parameters and D is 2048.

However, different regions of an image will be activated in different channels of the feature maps, and different regions have different importance for weather estimation. When extracting weather visual features from images, different kinds of convolution filters are adopted and the corresponding feature maps are generated. Each channel map of high-level semantic features can be regarded as a semantic-specific response. By weighting different channels, we could emphasize the feature maps with more weather-related features and improve the feature representation capability of the network. This is of great importance for GCN training since GCN learns a set of inter-dependent weather classifiers using weather features. Thus, we propose to introduce channel-wise attention module into the base model to adaptively calculate the attention weight of each channel feature response when extracting weather features in different layers.

Inspired by [51], we employ Squeeze and Excitation module as the channel-wise attention module for more informative weather features extraction. The attention module adaptively recalibrates channel-wise feature responses by explicitly modeling inter-dependencies between channels [51]. Through the dependencies, the channel-wise attention module learns to utilize global information to emphasize more informative features and weaken less useful features for weather features extraction.

Specifically, the channel information is firstly squeezed into a channel descriptor via the channel-wise statistics generated by global average pooling. Given the convolutional feature maps $\boldsymbol{X}$, the statistic $\boldsymbol{z} \in \mathbb{R}^C$ (where C denotes the number of channels) is generated by compressing $\boldsymbol{X}$ via spatial dimensions $H \times W$, with the c-th element calculated by:

$$z_c = \boldsymbol{F}_{sq}(\boldsymbol{X}_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j) \tag{11}$$

where $\boldsymbol{F}_{sq}(\cdot)$ denotes the squeeze operation, $\boldsymbol{x}_c$ is the c-th channel map of $\boldsymbol{X}$, and $x_c(i,j)$ means the element in row i and column j of the c-th channel map.

Then, to capture channel-wise dependencies, the excitation operation $\boldsymbol{F}_{ex}(\cdot)$ is performed via a gating mechanism using the information aggregated in the squeeze operation. The gating mechanism is realized with a *sigmoid* activation:

$$\boldsymbol{s} = \boldsymbol{F}_{ex}(\boldsymbol{z}, \boldsymbol{W}) = \sigma(g(\boldsymbol{z}, \boldsymbol{W})) = \sigma(\boldsymbol{W}_2 \delta(\boldsymbol{W}_1 \boldsymbol{z})) \tag{12}$$

To limit model complexity and improve generalization, a dimensionality-reduction layer, a ReLU and a dimensionality-increasing layer are employed in sequential to parameterize the mechanism. Both layers are fully connected (FC) layers. In Eq. (12), $\boldsymbol{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ is the parameters of the dimensionality-reduction FC layer with reduction ratio r, $\boldsymbol{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ is the parameters of the dimensionality-increasing FC layer, δ denotes the ReLU [52] function and $\boldsymbol{s}$ is the scalar vector (i.e., channel weights). The final output of the attention model is obtained by rescaling $\boldsymbol{X}$:

$$\hat{\boldsymbol{x}}_c = \boldsymbol{F}_{scale}(\boldsymbol{x}_c, \boldsymbol{s}_c) = \boldsymbol{s}_c \cdot \boldsymbol{x}_c \tag{13}$$

where $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \cdots, \hat{\boldsymbol{x}}_C]$ and $\boldsymbol{F}_{scale}(\boldsymbol{x}_c, \boldsymbol{s}_c)$ means channel-wise multiplication between the feature map $\boldsymbol{x}_c \in \mathbb{R}^{H \times W}$ and the corresponding channel weight $\boldsymbol{s}_c$.

4 Experiments

4.1 Datasets description

4.1.1 The transient attribute dataset

This dataset is transformed from Transient Attributes Database [53], which is originally constructed for outdoor scene understanding. Transient Attributes Database contains 8571 images, and exhibits different scales and views to guarantee the diversity across scenes. Furthermore, the database contains a variety of scenes, including urban scenes, town scenes, etc. The images in the database are annotated with 40 attribute labels including weather attribute labels. Moreover, the labels of each image are not mutually exclusive. Therefore, it is appealing to transform the dataset to perform multi-label weather recognition research. In this paper, six attribute labels are selected from

Fig. 6 Illustration of some examples in the transient attribute dataset. The right of the image depicts the studied weather labels in this paper, and only labels with scores greater than 0.5 are shown. The bold labels are the ones with the highest scores among these weather labels

the dataset for multi-label weather recognition, which are sunny, cloudy, foggy, snowy, moist and rainy. Some examples of the dataset are displayed in Fig. 6.

We re-annotate each image in the dataset for each of the six labels selected in this paper according to the score, which is annotated by the annotation works for each attribute label to evaluate the probability that the attribute appears in an image. If the score is less than 0.5, the corresponding label of the image is considered not positive enough and we re-annotate it to 0. Otherwise, re-annotate it to 1. For some images in the dataset, scores of the six labels are all less than 0.5 and not positive enough, which means the six labels are all re-annotated to 0 by us. For this situation, we define a new label for the dataset, which is "other". If the specified six labels of an image are all 0, then the label "other" is set to 1 for the image. Otherwise, set to 0. Therefore, the final transient attribute dataset used in this paper contains 8,571 images and 7 weather labels.

4.1.2 The multi-label weather classification dataset

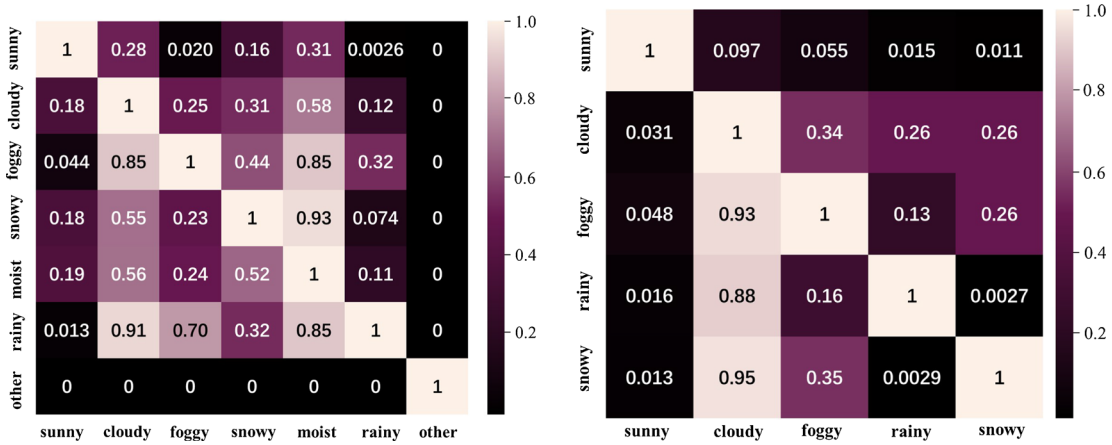
The multi-label weather classification dataset [4] contains 10,000 images, including 5 most common weather conditions in daily life, i.e., sunny, cloudy, foggy, snowy, rainy and rainy. The dataset is constructed from different kinds of scenes, including cities, suburbs, villages, etc., and the images are of different scales and views [4]. What's more, the labels of the dataset are nonexclusive and can provide a variety of weather information for each image, which is suitable for the research in this paper.

Each image in the dataset is labeled by not less than 5 annotation workers. The workers annotate the strength for each weather label from 0 to 1, in which the weather can be judged not in the image with the strength lower than 0.5, and assign the average strength as the final strength of the label [4]. To generate the final labels for images, the labels with strengths less than 0.5 are set to 0 and the labels with strengths equal or more than 0.5 are set to 1. Considering the image selection and labeling process of the dataset, we think it is reliable and suitable for the research in this paper.

4.1.3 The weather correlation matrix of both datasets

The asymmetric weather correlation matrix of the transient attribute dataset and the multi-label weather classification dataset are depicted in Fig. 7a and b, respectively. The matrix represents the probability that one weather condition will occur when another weather condition occurs for the weather conditions in the dataset.

As described in Sect. 3.3, the weather correlation matrix is constructed via the statistics of occurrences of each weather pair and the total number of occurrences that each weather appears in the dataset. From Fig. 7, we can simply draw the following conclusions. As with human intuition, there are strong co-occurrence relationships between different weather conditions, such as rainy and cloudy, foggy, cloudy, etc. From Fig. 7a and b, we can see that the probability of cloudy achieves 0.91 and 0.88, respectively, in the transient attribute dataset and the multi-label weather classification dataset when the weather condition of rainy appears. The probability of cloudy achieves 0.85 and 0.93, respectively, in the transient attribute dataset and the multi-label weather classification dataset when the weather condition of foggy appears. By analyzing the matrixes, it demonstrates there are indeed weather dependencies in weather recognition task. It is necessary to consider such



(a) Weather correlation matrix of the transient attribute dataset

(b) Weather correlation matrix of the multi-label weather classification dataset

Fig. 7 The illustration of weather co-occurrence relationships between different weather conditions **a** for the transient attribute dataset and **b** for the multi-label weather classification dataset

weather dependencies when predicting multiple weather labels.

4.2 Evaluation metrics

Following the previous work settings [4], we firstly employ per-class precision and recall as evaluation metrics, which means calculating precision and recall for each given weather label. Then, we report the average class precision (CP) and average class recall (CR), which are the average of per-class precision and recall respectively. In addition, overall precision (OP) and overall recall (OR) are computed, which measure the true predictions for all images over all weather classes. Finally, class F1 (CF1) and overall F1 (OF1) are adopted, which are the harmonic average of precision and recall. Generally, the CF1 and OF1 are of more importance for performance evaluation.

4.3 Implementation details

In the proposed model, our GCN consists of 2 layers and the output dimensionality are 1024 and 2048, respectively. We adopt 300-dim GloVe [54] to generate word embedding as weather label representations. Additionally, we employ ResNet-101 [31] as backbone extended with channel-wise attention model for image representation learning to extract weather visual features. The backbone is pre-trained on ImageNet dataset [50]. Before training, the input images are randomly cropped and resized to 224×224, and also randomly flipped for data augmentation. To optimize the network, SGD is exploited as network optimizer. The momentum is set to 0.9 and weight decay is set to 10^{-4}. The learning rate is initialized as 0.01.

4.4 Experimental results

To demonstrate the promising performance of our method, we conduct sufficient experiments on two public available datasets: the transient attribute dataset [53] and the multi-label weather classification dataset [4].

4.4.1 Comparisons with state-of-the-arts

Results on transient attribute dataset. For transient attribute dataset, 1000 images are randomly selected as testing set and the remains are as training set. Abundant experimental results are presented in Table 1. We compare the proposed approach with the state-of-the-art multi-label weather recognition methods, including AlexNet [27], VGGNet [30], Resnet101 [32], ML-KNN [55], ML-ARAM [56] and CNN-RNN method [4]. The results of AlexNet, VGGNet, ML-KNN, ML-ARAM and CNN-RNN are from [4]. Comparing with the state-of-the-arts, our method achieves promising performance. The main reason is that GCN better models the complex weather dependencies for inter-dependent weather classifiers learning and attention-aware weather representations provide more informative weather features for effective GCN training and weather estimation.

As shown in Table 1, our approach achieves the best results of CP, CR, CF1 and OR, which shows the promising performance of the proposed method. Specifically, our approach achieves CP =0.8445, CR = 0.7751, CF1 = 0.8084 and OR = 0.8342 on the transient attribute dataset. Comparing with the state-of-the-art multi-label weather recognition method CNN-RNN [4], the proposed approach improves the performance by +3.54% in CP, +3.23% in CR, +3.24% in CF1 and +1.11% in OR,

Table 1 Experimental results and comparisons with the state-of-the-arts on the transient attribute dataset (Per-class result: precision/recall)

Method	Sunny	Cloudy	Foggy	Snowy	Moist	Rainy	Other	CP	CR	CF1	OP	OR	OF1
AlexNet [27]	0.756/ 0.892	0.802/ 0.868	0.688/ 0.688	0.948/ 0.803	0.840/ 0.903	0.625/ 0.392	0.789/ 0.224	0.7783	0.6815	0.7267	0.8967	0.0800	0.8455
VGGNet [30]	0.777/ 0.836	0.847/ 0.803	0.767/ 0.717	0.848/ 0.920	0.873/ 0.899	0.880/ 0.431	0.622/ 0.552	0.8022	0.7369	0.7682	0.9043	0.8155	0.8576
Resnet101 [31]	0.805/ 0.832	0.864/ 0.834	0.756/ 0.727	0.919/ 0.943	0.936/ 0.897	0.607/ 0.638	0.675/ 0.593	0.7945	0.7808	0.7876	0.8519	0.8341	0.8429
ML-KNN [55]	0.720/ 0.892	0.898/ 0.742	0.866/ 0.609	0.887/ 0.944	0.818/ 0.933	0.700/ 0.412	0.663/ 0.425	0.7927	0.7080	0.7479	0.9001	0.8037	0.8492
ML-ARAM [56]	0.790/ 0.826	0.784/ 0.853	0.903/ 0.609	0.968/ 0.855	0.938/ 0.842	0.889/ 0.314	0.550/ 0.739	0.8318	0.7197	0.7717	0.9044	0.8047	0.8517
CNN-RNN [4]	0.857/ 0.785	0.851/ 0.852	0.837/ 0.682	0.952/ 0.896	0.913/ 0.911	0.656/ 0.454	0.585/ 0.628	0.8091	0.7428	0.7760	**0.9167**	0.8231	**0.8678**
GCN-A(ours)	0.853/ 0.816	0.859/ 0.858	0.825/ 0.735	0.94/ 0.908	0.911/ 0.893	0.761/ 0.564	0.763/ 0.651	**0.8445**	**0.7751**	**0.8084**	0.8730	**0.8342**	0.8532

Table 2 Experimental results and comparisons with the state-of-the-arts on multi-label weather classification dataset (Per-class result: precision/recall)

Method	Sunny	Cloudy	Foggy	Snowy	Rainy	CP	CR	CF1	OP	OR	OF1
AlexNet [27]	0.840/0.740	0.896/0.942	0.735/0.890	0.784/0.685	0.876/0.905	0.8263	0.8325	0.8294	0.9007	0.8668	0.8834
VGGNet [30]	0.772/0.851	0.927/0.915	0.867/0.728	0.814/0.701	0.887/0.931	0.8533	0.8252	0.839	0.9087	0.8494	0.8780
Resnet101 [31]	0.903/0.719	0.922/0.936	0.841/0.855	0.776/0.882	0.947/0.938	0.8780	0.8661	0.8720	0.8876	0.8861	0.8868
ML-KNN [55]	0.837/0.724	0.912/0.940	0.819/0.834	0.794/0.736	0.918/0.934	0.8562	0.8336	0.8447	0.9138	0.8766	0.8948
ML-ARAM [56]	0.772/0.810	0.853/0.936	0.952/0.783	0.641/0.762	0.979/0.840	0.8397	0.8262	0.833	0.8988	0.865	0.8816
SRN [44]	–	–	–	–	–	0.8700	0.8670	0.8680	0.8820	0.8910	0.8870
CNN-RNN [4]	0.838/0.843	0.917/0.953	0.856/0.861	0.856/0.758	0.894/0.938	0.8721	0.8702	0.8705	**0.9263**	0.8946	**0.9135**
GCN-A(ours)	0.881/0.851	0.933/0.948	0.902/0.840	0.925/0.796	0.948/0.955	**0.9178**	**0.8779**	**0.8974**	0.9222	**0.8976**	0.9097

respectively. Furthermore, it achieves 0.8532 of another important metric OF1, which is also comparable with the state-of-the-arts.

Results on multi-label weather classification dataset. For multi-label weather classification dataset, 2000 images are randomly selected as testing set to evaluate the performance of the methods and the remains as training set to train our model. Experimental results are reported in Table 2. It is observed from Table 2 that our proposed approach achieves the best results on the metrics of CP, CR, CF1 and OR compared with the state-of-the-arts, which shows the promising performance of the proposed model. Specifically, it improves the performance of CP, CR, CF1 and OR by +4.57%, +0.77%, +2.69% and +0.3%, respectively, compared with the CNN-RNN [4]. Moreover, our approach also gets good results on OP and OF1, which are comparable with the state-of-the-arts.

For per-class results, all the methods (which provide per-class results) achieve better performance on 'cloudy' and 'snowy' classes. That's because cloudy usually co-occurs with other weather conditions, and samples of 'cloudy' has a large number in multi-label weather classification dataset [4]. However, most methods perform a little worse on 'rainy' class. The reason may be that the dataset contains images of different views and it's difficult to recognize 'rainy' weather from distant view. Moreover, 'rainy' is the class with relatively less number of samples in the dataset. As shown in the table, our approach achieves higher precision and recall in general for each class, which also implies the superiority of our approach.

Overall, the proposed approach outperforms the multi-label version of AlexNet [27], VGGNet [30], ResNet [31],

Table 3 Ablation studies on the transient attribute dataset

Method	Sunny	Cloudy	Foggy	Snowy	Moist	Rainy	Other	CP	CR	CF1	OP	OR	OF1
A-CNN	0.874/0.765	0.884/0.840	0.759/0.765	0.959/0.875	0.973/0.805	0.573/0.741	0.682/0.643	0.8148	0.7765	0.7952	**0.8749**	0.8010	0.8363
CNN-GCN	0.814/0.869	0.893/0.853	0.797/0.712	0.925/0.928	0.923/0.891	0.607/0.638	0.702/0.607	0.8088	**0.7855**	0.7970	0.8633	**0.8411**	0.8520
GCN-A (not re-weighted)	0.849/0.795	0.867/0.836	0.820/0.690	0.935/0.928	0.919/0.908	0.674/0.534	0.632/0.65	0.8137	0.7631	0.7876	0.8614	0.8277	0.8442
GCN-A	0.853/0.816	0.859/0.858	0.825/0.735	0.94/0.908	0.911/0.893	0.761/0.564	0.763/0.651	**0.8445**	0.7751	**0.8084**	0.8730	0.8342	**0.8532**

We adopt ResNet101 as the backbone to extract weather features of an image, therefore, the CNN in the table denotes ResNet101. A-CNN: ResNet101with channel-wise attention. GCN-CNN: GCN with ResNet101. GCN-A: GCN with channel-wise attention-based ResNet101

Table 4 Ablation studies on the multi-label weather classification dataset

Method	Sunny	Cloudy	Foggy	Snowy	Rainy	CP	CR	CF1	OP	OR	OF1
A-CNN	0.869/ 0.859	0.955/ 0.893	0.922/0.8	0.889/ 0.836	0.881/ 0.981	0.9030	0.8738	0.8882	0.9182	0.8766	0.8969
CNN-GCN	0.862/ 0.842	0.903/ 0.969	0.874/ 0.826	0.873/ 0.809	0.944/ 0.929	0.8914	0.8745	0.8828	0.8946	0.9015	0.8980
GCN-A (Not re-weighted)	0.833/ 0.874	0.914/ 0.968	0.937/ 0.739	0.841/ 0.867	0.929/ 0.967	0.8909	**0.8830**	0.8870	0.8972	**0.9044**	0.9008
GCN-A	0.881/ 0.851	0.933/ 0.948	0.902/ 0.840	0.925/ 0.796	0.948/ 0.955	**0.9178**	0.8779	**0.8974**	**0.9222**	0.8976	**0.9097**

We adopt ResNet101 as the backbone to extract weather features of an image, therefore, the CNN in the table denotes ResNet101. A-CNN: ResNet101 with channel-wise attention. GCN-CNN: GCN with ResNet101. GCN-A: GCN with channel-wise attention-based ResNet101

the general multi-label methods ML-KNN [55], ML-ARAM [56], SRN [44], CNN-RNN [4] and achieves promising results.

4.4.2 Ablation study

In this section, we conduct ablation studies from three different aspects: effectiveness of Graph Convolutional Network, effectiveness of re-weighted mechanism and effectiveness of channel-wise attention in the proposed approach.

Effectiveness of Graph Convolutional Network (GCN) To evaluate the superiority of GCN in proposed GCN-A model, we conduct contrast experiments on the transient attribute dataset and the multi-label weather classification dataset. As shown in Table 3, our proposed GCN-A model achieves 0.8084 and 0.8532 of the most important metrics CF1 and OF1 on the transient attribute dataset, which outperforms the A-CNN model by +1.32% and +1.69%, respectively. As depicted in Table 4, the GCN-A model achieves 0.8974 and 0.9097 of CF1 and OF1 on the multi-label weather classification dataset, which improves the performance by +0.92% and +1.28%, respectively, compared with the A-CNN model. What's more, to further verify the effectiveness of GCN in weather recognition, we also perform the comparison experiments for CNN-GCN and ResNet101. Specifically, CNN-GCN achieves higher results of CF1 and OF1 than ResNet101 by +0.94% and +0.91% on the transient attribute dataset, and by +1.08% and +1.12% on the multi-label weather dataset. Overall, it achieves better results whether GCN is introduced into ResNet101 or A-CNN, which demonstrates the effectiveness of GCN in multi-label weather recognition. The main reason is that GCN can model the complex multi-label weather dependencies, and based on that, a set of inter-dependent weather classifiers are learned.

Effectiveness of re-weighted mechanism. To evaluate the influence of re-weighted mechanism in weather correlation matrix construction, we conduct the experiment with GCN-A (not re-weighted) on both datasets, in which the weather correlation matrix is not re-weighted. From Table 3, it can be observed that the proposed GCN-A model outperforms the GCN-A (not re-weighted) model on CP, CR, CF1, OP, OR and OF1 on the transient attribute dataset. Specifically, the proposed GCN-A model achieves higher results of the most important metrics CF1 and OF1 than the GCN-A (not re-weighted) model by +2.08% and +0.9%, respectively. From Table 4, we can see that the proposed GCN-A model achieves better performance on CP, CF1, OP and OF1 on the multi-label weather classification dataset. Specifically, the GCN-A model improves the performances of CF1 and OF1 by +1.04% and +0.89%, respectively, compared with the GCN-A (not re-weighted) model. In short, the experimental results verified the effectiveness of the re-weighted

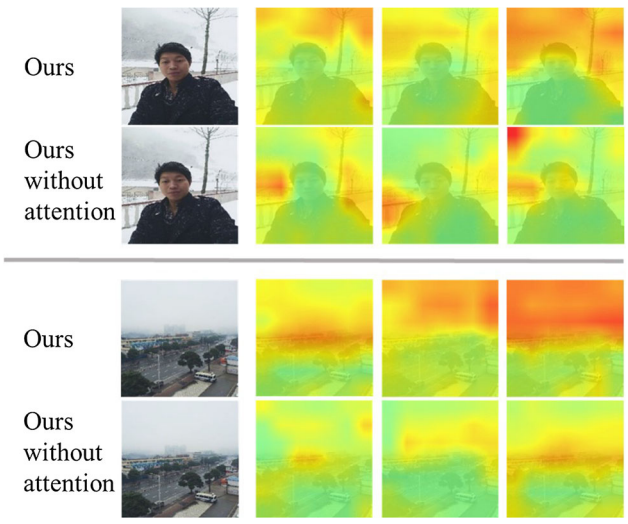


Fig. 8 Visualization of activation maps learned by our model with and without attention module. Each row contains one original image and three activation maps from the layer Conv5

mechanism for weather correlation matrix construction. That's because the developed re-weighted mechanism can select useful edges appropriately, and weigh the information of the node itself and its correlated nodes comprehensively in nodes information propagation of GCN. Thus, it can alleviate over-fitting and over-smoothing problems.

From the comparison in Tables 3 and 4, we can find that GCN-A (in which the correlation matrix is re-weighted) outperforms GCN-A (not re-weighted) on CP, CR, CF1, OP, OR and OF1, which indicates the effectiveness of the re-weighted mechanism for weather correlation matrix construction.

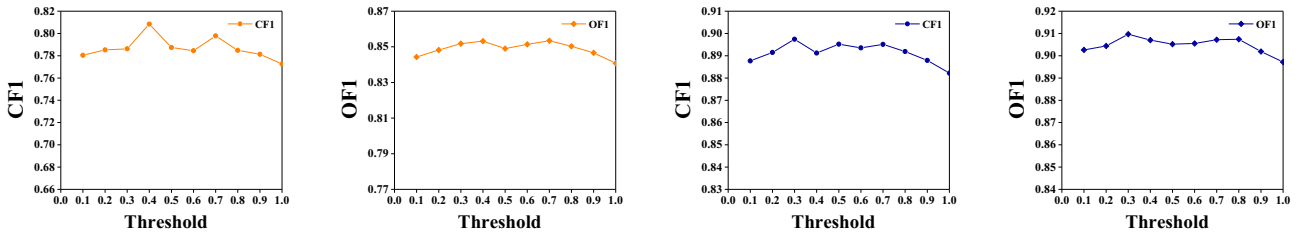
Effectiveness of channel-wise attention. To evaluate the effectiveness of channel-wise attention in our approach, we compare our approach with CNN-GCN (which only removes the channel-wise attention from our approach) on both datasets. As depicted in Table 3, our proposed GCN-A model achieves 0.8084 and 0.8532 of the most important metrics CF1 and OF1 on the transient attribute dataset, which improves the performance by +1.14% and +0.12%, respectively, compared with the CNN-GCN model. As shown in Table 4, the GCN-A model achieves 0.8974 and 0.9097 of CF1 and OF1 on the multi-label weather dataset, which outperforms the CNN-GCN model by +1.46% and +1.17%, respectively. In brief, it achieves better results with the channel-wise attention in the proposed GCN-A model, which demonstrate the effectiveness of the channel-wise attention. In addition to the above quantitative results, we present the visualization of some attention maps in Fig. 8 obtained from layer Conv5 of our model. As in the first image, obviously, the sky and snow are strong cues for the weather conditions. The sky and snow regions are more activated in our method than the method without attention. Moreover, the activated weather regions are more accurate in our approach. As in the second image, it can be observed that the sky region and foggy regions connected with the sky are more activated and accurate in our method than those in the method without attention. Overall, the better activation results demonstrate the effectiveness of the attention module and better performance of our approach. This is due to that the channel-wise attention can focus on the crucial weather-related regions to extract more discriminative weather features for weather estimation.

4.4.3 Important parameters

In this section, we evaluate three important parameters in our proposed model: parameter t in weather correlation matrix binarization, parameter w in weather correlation matrix reweighting and the depth of GCN layers. When evaluating one parameter, the others are fixed.

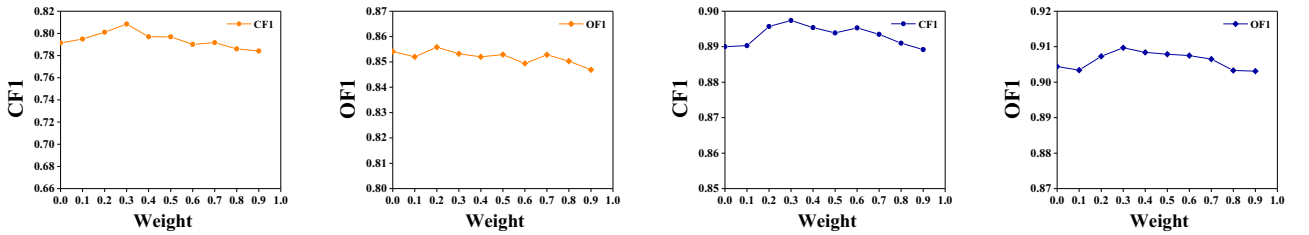
Effects of different t for weather correlation matrix binarization. We explore the effects of different threshold t in Eq. 8 for weather correlation matrix binarization by changing the values of t in a set of $\{0.0, 0.1, \ldots, 0.9, 1.0\}$. The results are shown in Fig. 9. The model will not converge if no edges are filtered. That's the reason why there is no results for $t = 0.0$. From the figure, we can see that generally the CF1 and OF1 on both datasets boost when small probabilities (i.e., noisy edges) are filtered. However, CF1 and OF1 on both datasets drop as more edges are filtered. This is due to that some correlated neighbors (weather nodes) may be ignored in node update in GCN when too many edges are filtered (i.e., when t approaches 1.0). Specifically, for transient attribute dataset, it performs best when $t = 0.4$, with CF1 of 0.8084 and OF1 of 0.8532. For multi-label weather classification dataset, CF1 and OF1 peak at 0.8974 and 0.9097, respectively, when $t = 0.3$.

Effects of different w for weather correlation matrix re-weighting. To explore the effect of w in Eq. 9 on multi-label weather recognition, we set the w from 0.0 to 1.0 and observe the changes of CF1 and OF1 on two weather datasets, as shown in Fig. 10. It presents the effect of balancing a node and its neighbors in node update in GCN. Generally, when w is too small or approaches to 1.0, the results of CF1 and OF1 are lower than middle ones. That's because when w is too small, the node cannot absorb sufficient information from its correlated neighbors in node update; while when w approaches 1.0, the node will ignore the information of itself which will result in over-smoothing. Specifically, for transient attribute dataset, it



(a) CF1 on transient attribute dataset **(b)** OF1 on transient attribute dataset **(c)** CF1 on multi-label weather classification dataset **(d)** OF1 on multi-label weather classification dataset

Fig. 9 Comparisons of CF1 and OF1 with different values of t. Note that the model does not converge when $t = 0$



(a) CF1 on transient attribute dataset **(b)** OF1 on transient attribute dataset **(c)** CF1 on multi-label weather classification dataset **(d)** OF1 on multi-label \weather classification dataset

Fig. 10 Comparisons of CF1 and OF1 with different values of w. Note that the model does not converge when $w = 1.0$

Table 5 Comparisons with different depth of GCN in our model

# Layer	Transient attribute dataset		Multi-label weather classification dataset	
	CF1	OF1	CF1	OF1
2-layer	**0.8084**	**0.8532**	**0.8974**	**0.9097**
3-layer	0.7902	0.853	0.8829	0.8975
4-layer	0.7644	0.8371	0.878	0.8849

achieves the best CF1 and a second-best OF1 when $w = 0.2$ and gets the best OF1 with a second-best CF1 when $w = 0.3$. In short, it performs better on transient attribute dataset when $w = 0.2$ or $w = 0.3$. For multi-label weather classification dataset, both CF1 and OF1 get the best results when $w = 0.3$. From the figure, we can see that we get CF1 of 0.7914 on transient attribute dataset and 0.89 on multi-label weather classification dataset, respectively, when w is set to 0, which still outperforms existing approaches. In fact, when set w to 0, the neighbor nodes are not taken into account in node update. The improvement may attribute to the inter-dependent weather classifiers, which are mapped from the nodes representations via a GCN-based mapping function. The inter-dependent weather classifiers implicitly model the weather dependencies through weather labels, as discussed in Section III-B.

Depth of GCN layers. We present the results of different GCN layers in our model in Table 5. For the two-layer model, the output dimensionalities are 1024 and 2048 for the sequential layers. For the three-layer model, the output dimensionalities are 1024, 1024 and 2048, respectively. For four-layer model, the dimensionalities are 1024, 1024, 1024 and 2048. As shown, it achieves better CF1 and OF1 on both datasets with two GCN layers. Both CF1 and OF1 drop as the number of GCN layers increases, which means that the performance of multi-label weather recognition reduces when using more GCN layers. The probable cause of performance reduction may be that the propagation between different nodes will be accumulated when employing more GCN layers, which may lead to over-smoothing.

4.4.4 Discussion

From Tables 1 and 2, it can be observed that our method achieves the best results for most metrics on both datasets compared with the state-of-the-arts, which shows the promising performance of our method. The main reason is that GCN better models the complex weather dependencies based on the re-weighted weather correlation matrix for classifiers learning. Moreover, attention-aware weather representations provide more informative weather features for training. We conduct ablation studies in Tables 3 and 4 to verify the effectiveness of Graph Convolutional Network, re-weighted mechanism, and channel-wise attention in the proposed approach, respectively. The main reason of the effectiveness of GCN is that GCN can model the complex multi-label weather dependencies, and based on that, a set of inter-dependent weather classifiers are learned. Re-weighted mechanism is effective in the proposed GAN-A model mainly because it can select useful edges appropriately and weigh the information of the node itself and its correlated nodes in nodes information propagation of GCN. The effectiveness of the channel-wise attention is due to that it can focus on the important weather related regions of an image to extract more discriminative weather features. To provide a better understanding of the results, some visualized activation maps are displayed in Fig. 8, which demonstrates the superiority of the attention-based model.

In addition, each of the important parameters in our proposed model is discussed, respectively. As shown in Fig. 9, for parameter t in weather correlation matrix binarization, noisy edges cannot be filtered when it is too small, and some useful edges may be filtered when it is too large,

which will result in performance degradation. For parameter w in weather correlation matrix reweighting shown in Fig. 10, some correlated nodes information may be ignored for node update in GCN when it is too small and the node itself may be ignored when it is too large. For another important parameter the depth of GCN layers shown in Table 5, too many GCN layers may affect the results negatively because nodes propagation in the layers will be accumulated, which can lead to over-smoothing.

Overall, GCN can be employed to model complex multi-label weather dependencies and learn the inter-dependent weather classifiers, and it may inspire the researchers to further explore multi-label weather recognition work in the future.

5 Conclusions

In this paper, to overcome the limitations of single-label weather recognition task, we propose a novel end-to-end Graph Convolutional Networks with Attention (GCN-A) model for multi-label weather recognition. GCN-A consists of GCN and semantic representation learning network. Specifically, we employ GCN to model weather co-occurrence dependencies, and design a re-weighted mechanism to construct weather correlation matrix to guide information propagation of nodes in GCN. For semantic representation learning network, we employ a channel-wise attention module based on the backbone to extract informative weather features from images. Practically, extensive experiments are conducted on different datasets. Experimental results on both the transient attribute dataset and the multi-label weather classification dataset show that the proposed GCN-A model outperforms the-state-of-arts in most metrics. The ablative experimental results are also analyzed to verify the effectiveness of the GCN, re-weighted mechanism and channel-wise attention, respectively. From the ablation experiments, we found that the proposed GCN-A model significantly improves the multi-label weather recognition task since GCN can explore important weather dependencies. Moreover, the attention mechanism performs the ability of promoting the GCN-A model to focus on discriminative regions. In future work, we will take other modality information, such as temperature and humidity, into consideration to obtain more abundant weather information.

Acknowledgements This work is supported by the National Key R&D Program of China (2019YFC1408405), National Natural Science Foundation of China (No.61672475, 61872326).

Compliance with ethical standards

Conflict of Interest The authors declared that they have no conflict of interest.

References

1. Lu C, Lin D, Jia J, Tang C (2014) Two-class weather classification, In: Proceedings of the Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 3718–3725
2. Zhao B, Hua L, Li X, Lu X, Wang Z (2019) Weather recognition via classification labels and weather-cue maps. Pattern Recognit 95:272–284
3. Lu C, Lin D, Jia J, Tang C (2017) Two-class weather classification. IEEE Trans Pattern Anal Mach Intell 39(12):2510–2524
4. Zhao B, Li X, Lu X, Wang Z (2018) A CNN-RNN architecture for multi-label weather recognition. Neurocomputing 322:47–57
5. Sun Q, Liu H, Harada T (2017) Online growing neural gas for anomaly detection in changing surveillance scenes. Pattern Recognit 64:187–201
6. Li X, Ye M, Liu Y, Zhang F, Liu D, Tang S (2017) Accurate object detection using memory-based models in surveillance scenes. Pattern Recognit 67:73–84
7. De-la-Torre M, Granger E, Sabourin R, Gorodnichy DO (2015) Adaptive skew-sensitive ensembles for face recognition in video surveillance. Pattern Recognit 48(11):3385–3406
8. Katsura H, Miura J, Hild M, Shirai Y (2005) A view-based outdoor navigation using object recognition robust to changes of weather and seasons. J Robot Soc Japan 23(1):75–83
9. Loncomilla P, Ruiz-del-Solar J, Martínez LM (2016) Object recognition using local invariant features for robotic applications: a survey. Pattern Recognit 60:499–514
10. Kurihata H, Takahashi T, Ide I, Mekada Y, Murase H, Tamatsu Y, Miyahara T (2005) Rainy weather recognition from in-vehicle camera images for driver assistance. In: Proceedings of the Intelligent Vehicles Symposium. IEEE, pp. 205–210
11. Pavlic M, Rigoll G, Ilic S (2013) Classification of images in fog and fog-free scenes for use in vehicles, In: Proceedings of the Intelligent Vehicles Symposium. IEEE, pp. 481–486
12. Roser M, Moosmann F (2008) Classification of weather situations on single color images, In: Proceedings of the Intelligent Vehicles Symposium. IEEE, pp. 798–803
13. Hautiére N, Tarel J-P, Lavenant J, Aubert D (2006) Automatic fog detection and estimation of visibility distance through use of an onboard camera. Mach Vision Appl 17(1):8–20
14. Zhang Z, Ma H, Fu H, Zhang C (2016) Scene-free multi-class weather classification on single images. Neurocomputing 207:365–373
15. Zhang Z, Ma H (2015) Multi-class weather classification on single images, In: Proceedings of the International Conference on Image Processing. IEEE, pp. 4396–4400
16. Lin D, Lu C, Huang H, Jia J (2017) Rscm: region selection and concurrency model for multi-class weather recognition. IEEE Trans Image Process 26(9):4154–4167
17. L. Yao, C. Mao, and Y. Luo, Graph convolutional networks for text classification, In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 7370–7377
18. S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, Attention based spatial-temporal graph convolutional networks for traffic flow forecasting, In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 922–929
19. Chen Z, Wei X, Wang P, Guo Y (2019) Multi-label image recognition with graph convolutional networks, In: Proceedings of the Conference on Computer Vision and Pattern Recognition. Computer Vision Foundation / IEEE, pp. 5177–5186
20. Yan X, Luo Y, Zheng X (2009) Weather recognition based on images captured by vision system in vehicle, In: Proceedings of the International Symposium on Neural Networks. Springer, pp. 390–398

21. Fathy M, Siyal MY (1995) An image detection technique based on morphological edge detection and background differencing for real-time traffic analysis. Pattern Recognit Lett 16(12):1321–1330
22. Pavlic M, Belzner H, Rigoll G, Ilic S (2012) Image based fog detection in vehicles, In: Proceedings of the Intelligent Vehicles Symposium. IEEE, pp. 1132–1137
23. Kurihata H, Takahashi T, Mekada Y, Ide I, Murase H, Tamatsu Y, Miyahara T (2006) Raindrop detection from in-vehicle video camera images for rainfall judgment, In: Proceedings of the First International Conference on Innovative Computing, Information and Control-Volume I (ICICIC'06), vol. 2. IEEE, pp. 544–547
24. Bronte S, Bergasa LM, Alcantarilla PF (2009) Fog detection system based on computer vision techniques, In: Proceedings of the International IEEE Conference on Intelligent Transportation Systems. IEEE, pp. 1–6
25. Li Q, Kong Y, Xia S-m (2014) A method of weather recognition based on outdoor images, In: Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP), vol. 2. IEEE, pp. 510–516
26. Song H, Chen Y, Gao Y (2014) Weather condition recognition based on feature extraction and k-nn, In: Proceedings of the Foundations and Practical Applications of Cognitive Systems and Information Processing. Springer, pp. 199–210
27. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks, In: Proceedings of the Advances in neural information processing systems, pp. 1097–1105
28. Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X, Cottrell GW (2018) Understanding convolution for semantic segmentation, In: Proceedings of the Winter Conference on Applications of Computer Vision. IEEE Computer Society, pp. 1451–1460
29. Lin T, Dollár P, Girshick RB, He K, Hariharan B, Belongie SJ (2017) Feature pyramid networks for object detection, In: Proceedings of the Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 936–944
30. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition, In: Proceedings of the International Conference on Learning Representations, Y. Bengio and Y. LeCun, Eds.,
31. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition, In: Proceedings of the Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 770–778
32. Elhoseiny M, Huang S, Elgammal AM (2015) Weather classification with deep convolutional neural networks, In: Proceedings of the International Conference on Image Processing. IEEE, pp. 3349–3353
33. Shi Y, Li Y, Liu J, Liu X, Murphey YL (2018) Weather recognition based on edge deterioration and convolutional neural networks, In: Proceedings of the 24th International Conference on Pattern Recognition (ICPR). IEEE, pp. 2438–2443
34. Guerra JCV, Khanam Z, Ehsan S, Stolkin R, McDonald-Maier K (2018) Weather classification: A new multi-class dataset, data augmentation approach and comprehensive evaluations of convolutional neural networks, In: Proceedings of the NASA/ESA Conference on Adaptive Hardware and Systems (AHS). IEEE, pp. 305–310
35. Li X, Wang Z, Lu X (2017) A multi-task framework for weather recognition, In: Proceedings of the International Conference on Multimedia. ACM, pp. 1318–1326
36. Ma J, Chow TW, Zhang H (2020) Semantic-gap-oriented feature selection and classifier construction in multilabel learning, In: IEEE Transactions on Cybernetics
37. Zhang J, Luo Z, Li C, Zhou C, Li S (2019) Manifold regularized discriminative feature selection for multi-label learning. Pattern Recognit 95:136–150
38. Ma J, Zhang H, Chow TW (2019) Multilabel classification with label-specific features and classifiers: A coarse-and fine-tuned framework, In: IEEE Transactions on Cybernetics
39. Zhang X, Zhou X, Lin M, Sun J (2018) Shufflenet: An extremely efficient convolutional neural network for mobile devices, In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6848–6856
40. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions, In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251–1258
41. Liu W, Liu X, Ma H, Cheng P (2017) Beyond human-level license plate super-resolution with progressive vehicle search and domain priori gan, In: Proceedings of the 25th ACM international conference on Multimedia, pp. 1618–1626
42. He L, Wang Y, Liu W, Zhao H, Sun Z, Feng J (2019) Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification, In: Proceedings of the IEEE International Conference on Computer Vision, pp. 8450–8459
43. Wang J, Yang Y, Mao J, Huang Z, Huang C, Xu W (2016) Cnn-rnn: A unified framework for multi-label image classification, In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2285–2294
44. Zhu F, Li H, Ouyang W, Yu N, Wang X (2017) Learning spatial regularization with image-level supervisions for multi-label image classification, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5513–5522
45. Wang Z, Chen T, Li G, Xu R, Lin L (2017) Multi-label image recognition by recurrently discovering attentional regions, In: Proceedings of the IEEE international conference on computer vision, pp. 464–472
46. Li Q, Qiao M, Bian W, Tao D (2016) Conditional graphical lasso for multi-label image classification, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2977–2986
47. Lee C-W, Fang W, Yeh C-K, Frank Wang Y-C (2018) Multi-label zero-shot learning with structured knowledge graphs, In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1576–1585
48. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:1609.02907
49. Ge W, Yang S, Yu Y (2018) Multi-evidence filtering and fusion for multi-label classification, object detection and semantic segmentation based on weakly supervised learning, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1277–1286
50. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database, In: Proceedings of the IEEE conference on computer vision and pattern recognition. Ieee, pp. 248–255
51. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks, In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141
52. Nair V, Hinton GE (2010) Rectified linear units improve restricted boltzmann machines, In: Proceedings of the ICML,
53. Laffont P-Y, Ren Z, Tao X, Qian C, Hays J (2014) Transient attributes for high-level understanding and editing of outdoor scenes. ACM Trans Graph (TOG) 33(4):1–11
54. Pennington J, Socher R, Manning CD (2014) Glove: Global vectors for word representation, In: Proceedings of the conference on empirical methods in natural language processing (EMNLP), pp. 1532–1543
55. Zhang M-L, Zhou Z-H (2007) Ml-knn: a lazy learning approach to multi-label learning. Pattern Recognit 40(7):2038–2048

56. Benites F, Sapozhnikova E (2015) Haram: a hierarchical aram neural network for large-scale text classification, In: Proceedings of the IEEE international conference on data mining workshop (ICDMW). IEEE, pp. 847–854