Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# Human activity recognition by manifold regularization based dynamic graph convolutional networks

Weifeng Liu [a,*], Sichao Fu [a], Yicong Zhou [b], Zheng-Jun Zha [c], Liqiang Nie [d]

[a] College of Control Science and Engineering, China University of Petroleum (East China), Qingdao 266580, China
[b] Faculty of Science and Technology, University of Macau, Macau 999078, China
[c] School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China
[d] School of Computer Science and Technology, Shandong University, Qingdao 266237, China

ARTICLE INFO

*Article history:*
Received 12 June 2019
Revised 29 September 2019
Accepted 13 December 2019
Available online xxxx

*Keywords:*
Graph convolutional networks
Semi-supervised learning
Human activity recognition

ABSTRACT

Deep learning has shown superiority to extract more representative features from multimedia data in recent years. Recently, the most typical graph convolutional networks (GCN) has achieved excellent performance in the semi-supervised framework-based data representation learning tasks. GCN successfully generalizes traditional convolutional neural networks to encode arbitrary graphs by exploiting the graph Laplacian-based sample structure information. However, GCN only fuses the static structure information. It is difficult to guarantee that its structure information is optimal during the training process and applicable for all practical applications. To tackle the above problem, in this paper, we propose a manifold regularized dynamic graph convolutional network (MRDGCN). The proposed MRDGCN automatically updates the structure information by manifold regularization until model fitting. In particular, we build an optimization convolution layer formulation to acquire the optimal structure information. Thus, MRDGCN can automatically learn high-level sample features to improve the performance of data representation learning. To demonstrate the effectiveness of our proposed model, we apply MRDGCN on the semi-supervised classification tasks. The extensive experiment results on human activity datasets and citation network datasets validate the performance of MRDGCN compared with GCN and other semi-supervised learning methods.



## 1. Introduction

With the advancement of science and technology, and the popularity of smart terminals such as smartphones and notebook computers, large scale multimedia data (e.g. document, picture, audio and video) are generated and uploaded to the Internet every day. Image is one of the largest, fastest growth speed and most informative multimedia data carriers in the current society. Therefore, images classification and recognition, such as human activity recognition (HAR) [1–3], face recognition [4–6], pedestrian detection [7–9] and object detection [10–12], have become an important part of computer vision, pattern recognition and machine learning in recent years, which can effectively analyze the content of digital images and give correct judgments. With the development of virtual reality and augmented reality, HAR has attracted much attention in many areas including video surveillance and accident warning. However, traditional conventional shallow learning algorithms including support vector machine [13], kernel least squares [14] and logistic regression [15] cannot extract more representative sample features and meet development needs, which directly affect the results of the image classification tasks.

To acquire high-level sample features from massive images, deep learning (DL) was introduced and has been demonstrated to be an effective method. In practical life, a small amount of labeled samples are readily available, whereas massive labeled samples cannot be directly obtained because it require a lot of manpower, material resources and financial resources. The most successful method is semi-supervised learning with manifold regularization (MRSSL), which uses the manifold structure information of unlabeled and labeled samples distribution by regarding it as a regularization term of the objective function. That is to say, any two samples with the closer space distance generally belong to the same category. Liu et al. [16] presented a kernel logistic regression with Laplacian regularization for web image annotation by employing the graph Laplacian to preserve the local geometry of the potential manifold. Tao et al. [17] proposed a Hessian regularized support vector machines model to improve the performance

* Corresponding author.
*E-mail address:* liuwf@upc.edu.cn (W. Liu).

https://doi.org/10.1016/j.neucom.2019.12.150
0925-2312/

of image annotation due to the richer null space of Hessian. Liu et al. [18] combined p-Laplacian with support vector machines and kernel least squares by utilizing $p$-Laplacian to express high-order manifold distribution. Ma et al. [19] utilized hypergraph $p$-Laplacian to capture the complex relationships among the different samples.

MRSSL methods are only effective for regular Euclidean data. There exists vast amounts of non-Euclidean data or graph data of arbitrary structures. In recent years, spectral convolution methods have received an increasing attention and achieved better performance including text classification [20–22] and image recognition [23–25]. Each node of graphs gathers its neighbors information by the convolution operation of the Fourier domain, in other words, this methods do not directly make the convolution on the graphs. Kipf and Welling [26] presented a graph convolutional networks to learn the sample features by fusing the direct neighbors relationships of each node. Fu et al. [27] considered the direct and indirect neighbors relationships to learn richer sample features, which improved the performance of semi-supervised classification. Yadati et al. [28] proposed a hypergraph-based GCN for document classification by using hypergraph to describe the multicultural relationships among samples.

However, the above methods depend on the static samples distribution, which limit the range of its application. Confronting this challenging problem, it is important to design a dynamic graph structure learning model, which aims to automatically optimize the local geometry of samples. In this paper, we propose a dynamic graph convolutional network based on manifold regularization (MRDGCN) for semi-supervised classification. We introduce a manifold regularization term to the objective function, which can drive the objective function to change over the potential sample distribution manifold. When the objective function value cannot meet a specific threshold, MRGCN separately updates or optimizes its manifold structure information (except the first convolution layer) and network weight matrix until model fitting. After many times training iteration, our proposed MRGCN can acquire optimal structure information. In addition, we make an optimization and derivation for the convolution layer formulation of GCN, and then propose a general graph structure learning framework. Finally, MRDGCN can extract more high-level sample features by fusing its dynamic structure information to improve the performance of the semi-supervised classification. To verify the performance of our proposed MRDGCN model, we have tested on CAS-YNU-MHAD, unstructured social activity attribute (USAA), toy human activity dataset (2moons dataset) for human activity recognition and Citeseer, Cora datasets for citation networks classification. The experiments results prove that the proposed MRDGCN algorithm shows better classification performance compared to GCN and other semi-supervised learning models.

The remainder of this paper is arranged as follows: Section 2 briefly summarizes the related works of spectral approaches on graph domain. Section 3 details describes our proposed MRDGCN framework. Large-scale experimental results and some discussions are presented in Section 4 and Section 5. At the end of this paper, we give the conclusion.

## 2. Spectral approaches on graph domain

The initial spectral graph convolution [23] is expressed as the multiplication of signal $X$ and filter $g$. It can be denoted as the following form:

$$g_\theta(L)\star X = U\left(\left(U^T g_\theta(L)\right)\odot\left(U^T X\right)\right) = U g_\theta(\Lambda)U^T X \tag{1}$$

Here, $U$ and $\Lambda$ denotes the eigenvector matrix and the diagonal matrix with non-negative eigenvalues in the normalized graph Laplacian $L$, respectively. $U^T X$ represent the frequency domain signals in the Fourier domain. That is to say, spectral graph convolution converts the convolution of the time domain to the point multiplication of the frequency domain.

However, this method is not suitable for large graphs and has a very high computation cost. To over this problem, Defferrard et al. [20] utilized the Chebyshev polynomials about the normalized graph Laplacian to approximate the filter $g$, and then proposed a spectral convolution with K-order polynomials on graphs, i.e.

$$g_\theta(L)\star X = \sum_{k=0}^{k}\theta_k T_k\left(\widetilde{L}\right)X \tag{2}$$

In this method, $\widetilde{L}$ is rescaled according to $\frac{2}{\lambda_{max}}L - I_N$. $\lambda_{max}$ represents the maximum eigenvalue of the normalized graph Laplacian $L, L$ is equal to $I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}, D_{ii} = \sum_j A_{ij}$. $A$ denotes the similarity matrix among different samples. The Chebyshev polynomials is recursively expressed according to $T_0(X) = 1,\ T_1(X) = X$ and $T_k(X) = 2XT_{k-1}(X) - T_{k-2}(X)$.

To further build a linear and deep model, Kipf and Welling limited the order of the Chebyshev polynomials ($K = 1$), in other words, it only considered the direct relationships between any two samples. Finally, it acquired a linear convolution layer formulation, i.e. $H^{(L+1)} = \sigma\left(\widetilde{B}H^{(L)}W^{(L)}\right)$. $\widetilde{B}$ is equal to $\widetilde{D}^{-\frac{1}{2}}(A + I_N)\widetilde{D}^{-\frac{1}{2}}$. $W^{(L)}$ is weight parameter matrix to be learned in the training iteration process. The detailed evolution process can be found in [26]. During the training process of GCN, it increased the number of training iteration according to value of cross entropy loss objective function and will stop until model fitting.

## 3. Manifold regularized dynamic graph convolutional networks

MRDGCN can learn more effective sample features by continuously updating manifold distribution information of samples apart from the first convolution layer during the convolution process, which yields better classification performance than GCN model. We first introduce the traditional manifold regularized framework to original objective function of GCN, and then propose a dynamic graph structure learning (DGSL) method. Following, we give the optimization scheme of DGSL on each training iteration. To solve the application limitation of GCN on different datasets, in the last, we present the derivation and optimization scheme of each convolution layer formulation for DGSL. And then we analyze the implement process of a two-layer MRDGCN. Fig. 1 shows the general framework of our proposed MRDGCN.

### 3.1. Dynamic graph structure learning

GCN relied on a static sample manifold distribution information $\widetilde{B}$ that is computed according to the input data. That is to say, this method only optimized weight matrix $W^{(L)}$ of each convolution layer. In real life, the local geometry distribution between data always in changing, thus GCN cannot get the most effective sample features. To tackle this issue, we apply traditional manifold regularization framework to the cross entropy loss function of GCN, and then propose a DGSL model to encode the complex data structure. In other words, we update the structure relationships (apart from the first convolution layer) and weight parameters simultaneously by adding an additional manifold regularization term to the cross entropy loss function. Therefore, our proposed DGSL can be denoted as the following optimization issue, i.e.

$$C(W,A) = C_1(W,A) + \beta C_2(W,A) = -y\log Y + \beta tr\left(Y^T M Y\right) \tag{3}$$

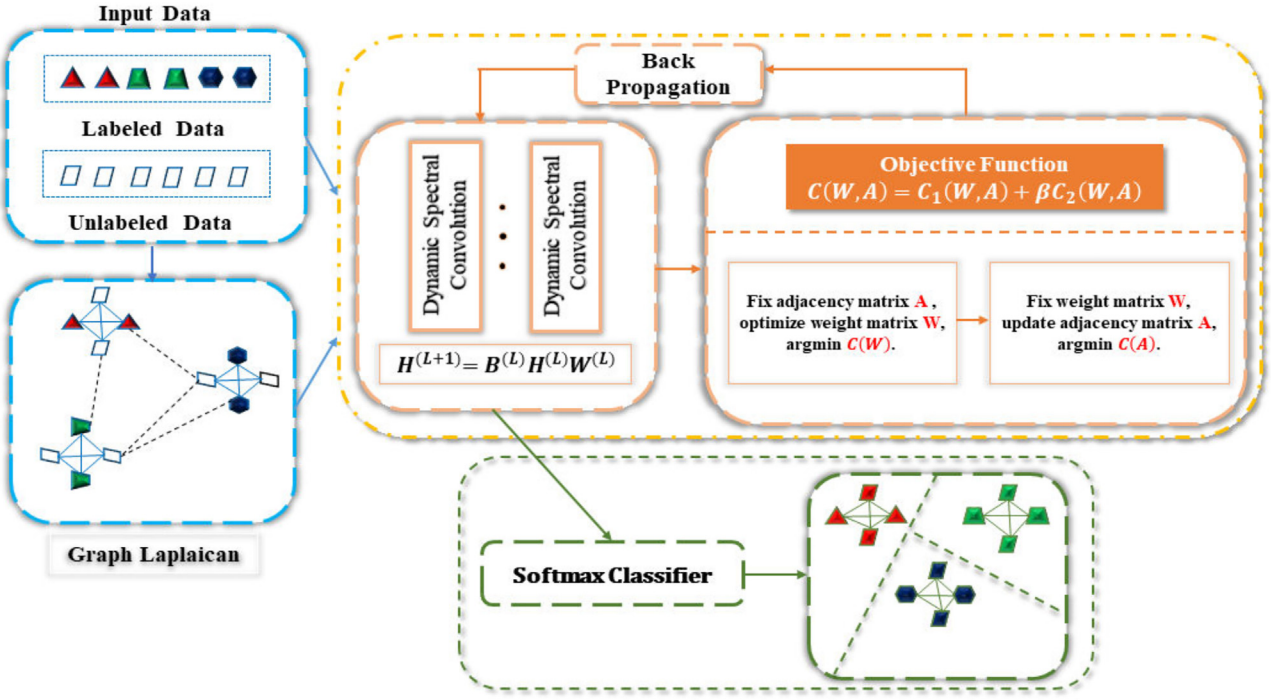



**Fig. 1.** The framework of the manifold regularized dynamic graph convolutional networks.

In DGSL, $C_1(W,A)$ denotes the cross entropy loss. $tr\left(Y^TMY\right)$ represents the manifold regularization term. $M$ denotes that how to use more effective method including graph Laplacian, Hessian and $p$-Laplacian, to better preserve the local geometric distribution of samples. In this paper, we use universally used graph Laplacian matrix. $y$ is the label information matrix of the input sample. $Y$ denotes the probability distribution matrix of samples, i.e. we can get Y by taking the extracted sample features $H^{(L+1)}$ of the last convolution layer as the input of the Softmax function. Thus, the above objective function can be written as:

$$C(W,A) = -y\log Y + \beta tr\left(Y^TLY\right) \tag{4}$$

We can regard the objective function of DGSL as a dual optimization issue, i.e. we need to update or optimize two variable $W$ and $A$. $\beta$ is a balance parameter of the objective function.

### 3.2. Optimization scheme

In this paper, we use the alternative optimization strategy to update adjacency matrix and weight matrix, i.e. when we update a specific variable, we fix others variable until the objective function of DGSL meet a specific threshold or remain unchanged for many times. We first fix the adjacency matrix $A$ and optimize weight matrix $W$, thus the above objective function can be simplified as the following form:

$$C(W) = -y\log Y + \beta tr\left(Y^TLY\right) \tag{5}$$

where the initial $A$ of each convolution layer is calculated according the input samples by k-NearestNeighbor with the Euclidean distance. In this paper, this problem can be solved by the gradient descent method [29].

Following, we fix weight matrix $W$, the above objective function on updating adjacency matrix $A$ (except for the first convolution layer) can be denoted as

$$C(A) = -y\log Y + \beta tr\left(Y^TLY\right) \tag{6}$$

The $A$ can be updated according to output sample features of final convolution layer of the last training iteration. With the update of adjacency matrix $A$, we also further make an optimization for the weight matrix $W$. Our proposed DGSL will repeat this process until the objective function value do not come down for many epochs.

### 3.3. A two-layer MRDGCN

To increase the scale adaptability of model, GCN let $\lambda_{max} = 2$. With the diversification of adjacency matrix $A$ computing methods and the differences between different datasets, $\lambda_{max} = 2$ has the limitation. To solve this problem, we make an optimization for the one-order spectral graph convolution again, and then the spectral convolution with one-order polynomial can be simplified as the following form:

$$g_\theta(L)\star X = \sum_{k=0}^{k=1}\theta_k T_k\left(\tilde{L}\right)X = \theta_0 X + \theta_1\left(\frac{2}{\lambda_{max}}L - I_N\right)X \tag{7}$$

where $\theta_0$ and $\theta_1$ denote the filter parameters, i.e. weight matrix $W^{(L)}$ of each convolution layer. To further avoid the overfitting issue of the above formula, we can get the following expression by using a single parameter $\theta$, i.e.

$$\begin{aligned} g_\theta(L)\star X &= \left(\theta_0 - \theta_1 + \frac{2}{\lambda_{max}}\theta_1\left(I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)\right)X \\ &= \left(\theta_2 + \frac{2}{\lambda_{max}}\theta_1\left(I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)\right)X \\ &= \left(\frac{2}{\lambda_{max}}\left(\frac{\lambda_{max}}{2}\theta_2 + \theta_1\right) - \frac{2}{\lambda_{max}}\theta_1 D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)X \\ &= \frac{2}{\lambda_{max}}\left(\theta_3 - \theta_1 D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)X = \theta\frac{2}{\lambda_{max}}\left(I_N + D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)X \end{aligned} \tag{8}$$

where, $\theta_0 - \theta_1 = \theta_2, \frac{\lambda_{max}}{2}\theta_2 + \theta_1 = \theta_3$ and $\theta_3 = -\theta_1 = \theta$. To further avoid vanishing gradients and instabilities by stacking the above form to build a deep network model, we also introduce the renormalization trick that is suggested in [26]. And then we get an optimization convolution layer formulation for a signal $X \in R^{n\times m}$ with $n$

samples and $m$ dimensional sample features, which is named the manifold regularized dynamic graph convolutional networks (MRDGCN), i.e.

$$g_\theta(L)\star X = \frac{2}{\lambda_{max}}\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}X\theta = \frac{2}{\lambda_{max}}B^{(L)}H^{(L)}W^{(L)} \tag{9}$$

where $\theta \in R^{m\times z}$ or $W^{(L)}$ is filter parameter of each network layer. $\widetilde{A} = A + I_N$ and $B^{(L)} = \widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}$.

The dynamic spectral convolution can be interpreted as the following process: (1) During the first iteration training process, when the loss values of the objective function Eq. (4) meet a specific threshold, MRDGCN will stop training. Otherwise, MRDGCN will constantly optimize the adjacency matrix and weight matrix of each layer by the alternative optimization strategy, and then will implement the second iteration training. (2) MRDGCN will repeat the above process until the objective function Eq. (4) meets a specific threshold. (3) Finally, MRDGCN can acquire the optimal adjacency matrix and weight matrix. After multi-layer spectral convolutions, it can learn richer data information to improve the classification performance.

**Algorithm 1**: A two-layer manifold regularized dynamic graph convolutional networks

**Input**: Datasets X
**Parameter**: $\beta$, training iteration epoch k, learning rate etc.
**Output**: Recognition accuracy
1: $H^{(1)} = RELU\left(\frac{2}{\lambda_{max}}B^{(0)}XW^{(0)}\right)$.
2: $H^{(2)} = \frac{2}{\lambda_{max}}B^{(1)}H^{(1)}W^{(1)}$.
3: **for** $j = 0 \rightarrow k-1$
4: Fix $A$, optimize $W$ by Eq. (5).
5: Fix $W$, update $A$ by Eq. (6)
6: **end for**
7: Get the optimal $A, W$ and $H^{(2)}$.
8: Give the recognition accuracy.

In this paper, we use a two-layer MRDGCN for semi-supervised classification to demonstrate the effectiveness of our proposed method. And then, we can get the following formulation by stacking two-layer Eq. (9), i.e.

$$H^{(2)} = \frac{2}{\lambda_{max}}B^{(1)}RELU\left(\frac{2}{\lambda_{max}}B^{(0)}XW^{(0)}\right)W^{(1)} \tag{10}$$

where $X$ denotes the initial sample features. After two-layer convolution operation, MRDGCN can efficiently learn the sample features with $n$ samples and $Q$ dimensional (The number of $Q$ is equal to classes). Following, we can get the probability distribution matrix $Y$, and then can get the optimal structure information of the second layer and weight matrix by our proposed optimization scheme after many training iteration. Finally, MRDGCN can extract more representative sample features by fusing the optimal structure information and give the recognition accuracy by Sofmax classifier. Algorithm 1 summarize a two-layer manifold regularized dynamic graph convolutional networks.

## 4. Experiments

To demonstrate the effectiveness of our proposed algorithm, in this section, we conduct substantial experiments to test the proposed MRDGCN, MRDGCN-1 and other semi-supervised learning methods, such as the representative GCN [26], graph attention networks (GAT) [30], Hypergraph p-Laplacian graph convolutional networks (HpLapGCN) [31], Chebyshev ($K = 2$) [20], Chebyshev ($K = 3$) [20], HyperGCN [28], manifold regularization (ManiReg) [32], semi-supervised embedding (SemiEmb) [33], multi-layer perceptron (MLP) [26], skip-gram based graph embeddings (DeepWalk) [34], on several human activity datasets and citation network datasets. The baseline MRDGCN-1 denotes the combination of optimized GCN variation Eq. (9) and cross entropy loss function [26], which still use a static sample structure information. MRDGCN is constructed by the optimized convolution layer formulation Eq. (9) and our proposed objective function Eq. (4). Following, we detailed present the used datasets, experiment parameters setting and comparison results, respectively.

### 4.1. Experiment datasets

CAS-YNU-MHAD dataset [2], a multimodal human activity database, was created by the Chinese Academy of Sciences and Yunnan University. It was collected by utilizing a Microsoft Kinect camera and 17 inertial sensors, which were placed in spine, left shoulder, right shoulder, left arm, right arm, left forearm, right forearm, left hand, right hand, spine, hips, right up leg, left up leg, left leg, right leg, left foot and right foot. It is composed of 1086 human actions with 10 different classes, such as typing, lying down, walking S, walking quickly, walking, sitting up, sitting down, running, jumping up and jumping forward. Each category was collected from 10 people between 20 and 30 years old. The original depth human action image is resized to $512 \times 424$ pixels with 16-bit. We also use the same feature extraction method that is suggested in [2] to acquire the STCP features [35] of each depth human action image, i.e. we can get a feature matrix with 1086 samples and 81648 dimensional features.

Unstructured social activity attribute (USAA) dataset [1,36] is composed of totally 1600 videos collected from the Columbia Consumer Video (CCV) [37] database. USAA consists of eight semantic category videos collected from birthday party, music performance, non-music performance, parade, wedding dance, wedding ceremony, graduation party and wedding reception. USAA provides the static SIFT features and spatial–temporal interest points (STIP) features. In our experiments, we use the video features that concatenate SIFT and STIP to learn more effective sample features via MRDGCN, i.e. each video has 10000 dimensional features.

The 2moons database is a virtual human action database constructed by humans, which contains 200 images. All images are selected into 2 categories, which is distinguished by label $+1$ (moon images) or labeled $-1$ (non-moon images).

Citeseer dataset [38] consists of 3327 machine learning publications with 3703 dimensional features, which is collected from six classes including Agents, AI, IR, HCI, DB and ML.

Cora dataset [39] is composed of 2708 publications with 1433 dimensional features, which are totally seven categories, such as reinforcement-learning, case-based, probabilistic-methods, neural-networks, theory, genetic-algorithms and rule learning.

### 4.2. Experiment parameters

In our semi-supervised classification experiments, for CAS-YNU-MHAD dataset, 200 samples are randomly selected as test set, 200 samples are randomly choose as validation set, and the remaining part for training set. For USAA database, we use 300 samples for validation set, 300 samples for testing and the rest videos for training. To demonstrate the classification performance of our proposed method under few labeled training data, in its training set (CAS-YNU-MHAD and USAA), we randomly chose 10%, 20%, 30%, 40% and 50% samples as labeled data and the rest samples for unlabeled data. Because the 2moons dataset is quite small, we randomly select 50 samples as validation samples, 50 samples for test samples and the rest as training samples. In addi-

tion, we randomly assign a specific label rate including 5%, 10%, 15%, 20% and 25% as labeled samples and the rest part as unlabeled samples on the 2moons database. For Citeseer and Cora dataset, we choose 1000 samples as test samples, 500 samples as validation samples and the rest as training samples.

To minimize the loss value of our proposed objective function, the Adam optimizer [40] with an initialized learning rate of 0.01 and 0.001 (USAA) is used to train a two-layer MRDGCN. The maximum training iteration up to 200 epochs and MRDGCN will early stopping training if the loss value of objective function do not descend for 10 times continuously on its validation set. We use the weights initialization method that is suggested in [41] during the training process. The L2 regularization with a regularization parameter of $5 \times 10^{-4}, 5 \times 10^{-5}$ (USAA) and the dropout [42] with a dropout rate of 0.5 is used to solve the overfitting problem of MRDGCN. The balance parameter $\beta$ of objective function is 0.001 and 1 (USAA). In addition, on the CAS-YNU-MHAD, USAA, 2moons datasets and citation network datasets, the output feature dimension of the first convolution layer is $128, 32, 16$ and 16, respectively.

### 4.3. Citation networks classification – toy experiment

In the GCN and most of existing GCN variant models, the commonly-used datasets are citation networks datasets. Thus, we also make a comparison on the Citeseer and Cora datasets for our proposed MRDGCN. We use 120 and 140 labeled samples on the Citeseer and Cora training samples respectively and give the mean recognition rate with 100 random runs. As shown in Table 1, we can find that our proposed MRDGCN model performs better than the state-of-the-art methods, which also indicates that MRDGCN can better preserve and utilize the structure relationships of the samples via manifold regularization method. It also reveals that MRDGCN can learn more representative sample features by dynamically updating the manifold distribution of the samples during the training process.

**Table 1**
Comparison of the different algorithms on citation network datasets.

| Method | Citeseer (120) | Cora (140) |
|---|---|---|
| ManiReg | 60.1 | 59.5 |
| SemiEmb | 59.6 | 59 |
| DeepWalk | 43.2 | 67.2 |
| Chebyshev ($K = 2$) | 53.6 | 49.8 |
| Chebyshev ($K = 3$) | 53.7 | 50.5 |
| GCN | 52.8 | 57.2 |
| MLP | 46.5 | 55.1 |
| MRDGCN | 66.6 | 64.8 |

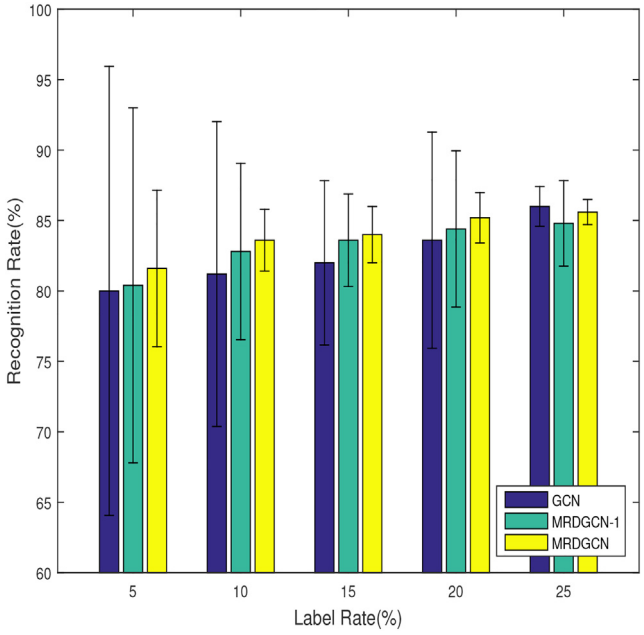



**Fig. 2.** Recognition rates of all categories on the 2moons database.

### 4.4. Human activity recognition

Figs. 2–4 show the histogram of the recognition rates for all classes on the 2moons, CAS-YNU-MHAD and USAA datasets respectively. Reported numbers denote the mean recognition accuracy with five run experiments in percent. The y-axis denotes the mean recognition accuracy over all categories for GCN, MRDGCN-1 and MRDGCN model. The x-axis denotes the number of labeled

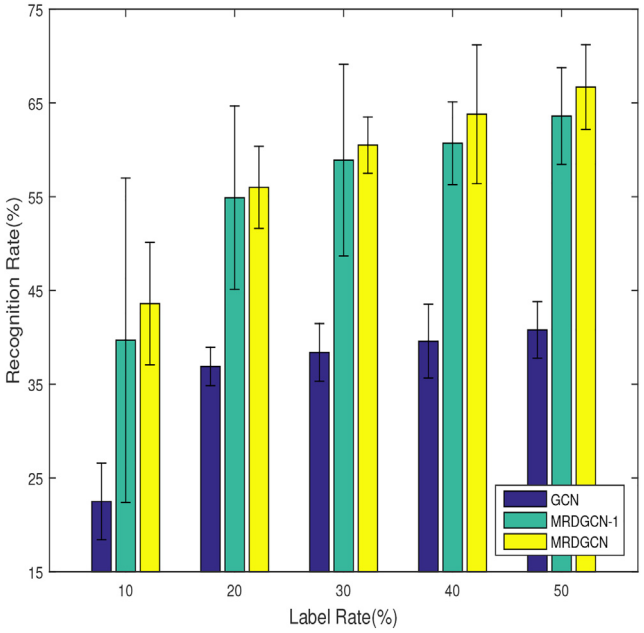



**Fig. 3.** Recognition rates of all categories on the CAS-YNU-MHAD database.

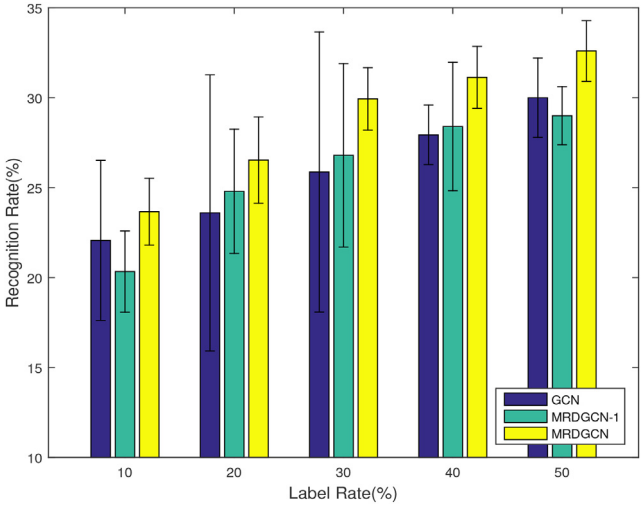



**Fig. 4.** Recognition rates of all categories on the USAA database.

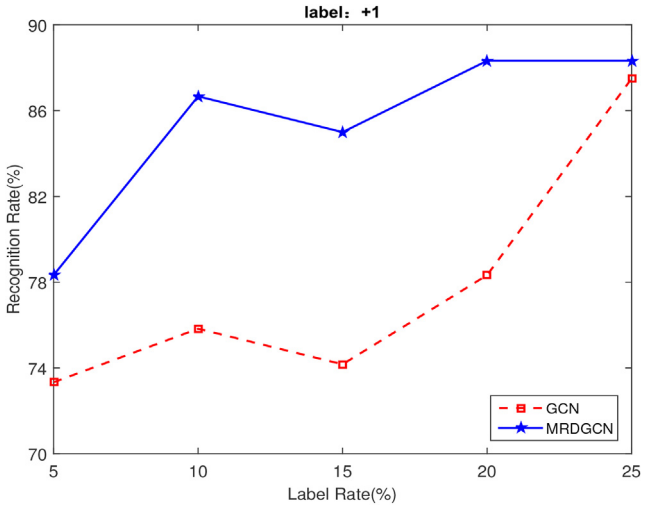



**Fig. 5.** Recognition rate of label +1 on the 2moons database.

samples randomly chose from training samples. From Figs. 2–4, we can find that the proposed MRDGCN achieves a higher recognition rate than other models, especially only a few labeled samples. In addition, it also reveals the effectiveness of optimization method for each convolution layer formulation.

Fig. 5 reveals the mean recognition rate of the 2moons database about the label +1 (moon images). Fig. 6 shows the mean recognition rate of the CAS-YNU-MHAD database over some human actions, such as jumping forward, jumping up, walking quickly and lying down. Fig. 7 shows the average classification accuracy of the USAA database over some human actions, such as graduation party, non-music performance, parade and wedding dance. The y-axis is the mean recognition rate of each class on different datasets. Each subfigure corresponds one human action. From the results of the human activity datasets, we can see that MRDGCN also gets the best classification performance.

To further demonstrate the effectiveness of dynamic graph structure learning method and the optimization of convolution layer formulation, we compare many state-of-the-art semi-supervised learning algorithms. In addition, we report the mean classification performance under the CAS-YNU-MHAD and 2moons datasets with the 30% and 5% labeled samples respectively. From Figs. 2–7 and Table 2, we can find that our proposed MRDGCN shows the best classification performance. Due to the generalization of the proposed convolution rule Eq. (9) on different datasets, MRDGCN-1 can acquire the optimal static data structure information and effective sample features compared with GCN. Compared with MRDGCN-1 and other semi-supervised learning methods, MRDGCN can get the optimal dynamic space structure information of each layer by the effective combination of our proposed convolution rule Eq. (9) and objective function with manifold regularization term Eq. (4). Finally, MRDGCN can extract more reasonable and effective data features via the convolution fusion of the acquired data structure relationships and input data features. These data indicates that MRDGCN can better preserve and utilize the local manifold structure of data during the training process of model.

### 4.5. Parameters sensitivity analysis

In our proposed MRDGCN, there exists one parameter $\beta$ to balance the cross entropy loss value and manifold regularization term

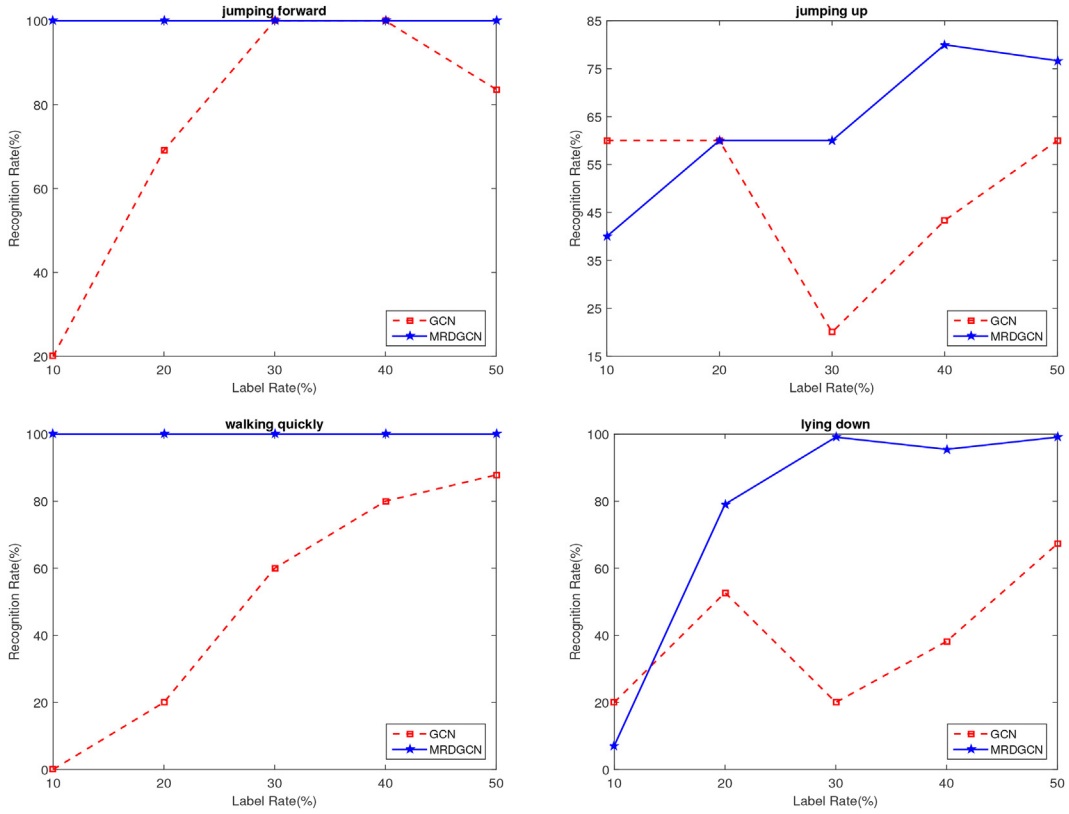



**Fig. 6.** Recognition rate of each category on the CAS-YNU-MHAD database, including jumping forward, jumping up, walking quickly and lying down. Each subfigure corresponds on single class.

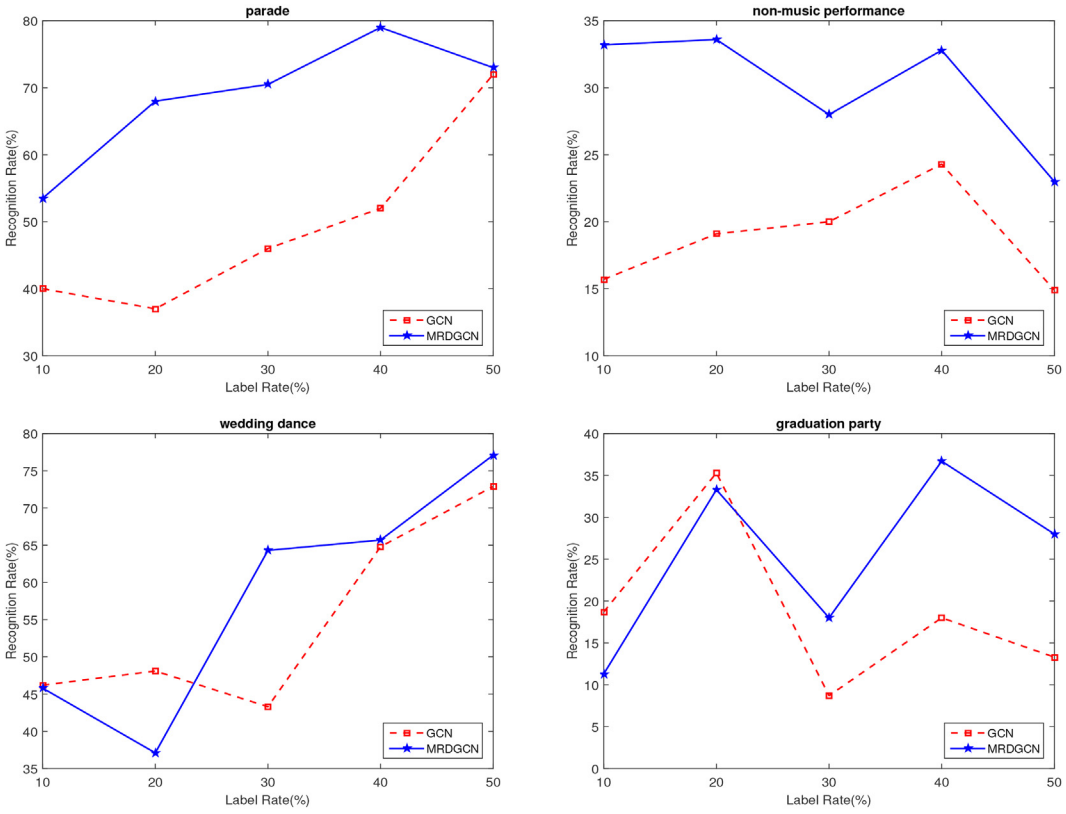



**Fig. 7.** Recognition rate of each category on the USAA database, including graduation party, non-music performance, parade and wedding dance. Each subfigure corresponds on single class.

**Table 2**
Comparison of the different algorithms on human activity datasets.

| Method | 2moons (5%) | CAS-YNU-MHAD (30%) |
|---|---|---|
| Chebyshev ($K = 2$) | 78.8 | 47.9 |
| Chebyshev ($K = 3$) | 79.2 | 55 |
| GCN | 79.6 | 38.5 |
| GAT | 80 | 48.8 |
| HyperGCN | 80.4 | 47 |
| HpLapGCN | 80.8 | 49 |
| MRDGCN | 81.6 | 60.5 |

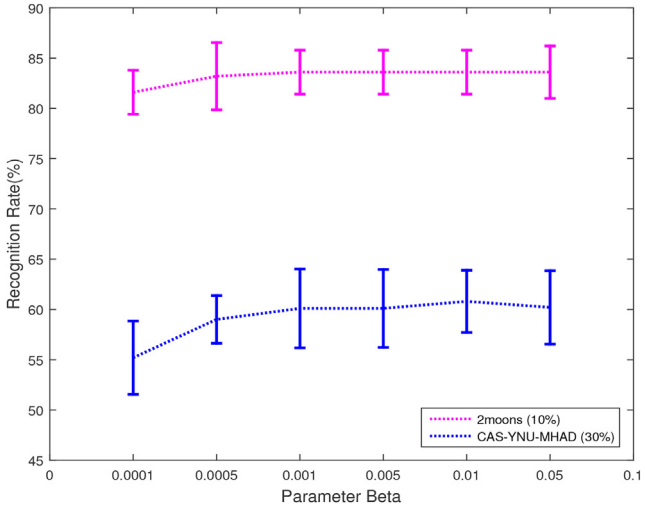



**Fig. 8.** Parameter sensitivity analysis on the CAS-YNU-MHAD and 2moons.

in the objective function. Thus, we conduct extensive experiments to analyze the sensitivity of the parameter $\beta$. We run MRDGCN with different $\beta$ values (from 0.0001 to 0.05) and give its mean average recognition accuracy under 30% and 10% labeled samples of the CAS-YNU-MHAD and 2moons. As shown in Fig. 8, we can see that when $\beta$ varies from 0.001 to 0.05, the proposed method has a stable classification performance. It indicates that MRDGCN is less sensitive to the selection of parameter $\beta$.

## 5. Discussions

In the future works, we will explore the graph manifold regularization method to maintain the local invariance constraint between data, and then propose the manifold regularization based graph convolutional networks (MRGCN) model. By the combination of our proposed objective function variant Eq. (11) and the effective convolution layer rule of GCN [26], MRDGCN can better utilize the adjacency relationships between data compared with GCN [26]. Thus, under utilizing the static structure relationships between samples, how to choose appropriate objective function to acquire the effective parameters is vital. In other words, MRGCN still cannot optimize the structure relationships between samples.

$$C(W) = C_1(W) + \beta C_2(W) = -y \log Y + \beta tr\left(Y^T M Y\right) \tag{11}$$

## 6. Conclusion

In recent years, deep learning-based graph structure learning models, especially graph convolutional networks (GCN), have captured widespread attention of researchers in data representation learning area with complex structure. Nevertheless, existing GCN

and its variants only utilized a static manifold distribution information of samples by spectral convolution operation and give little attention to its optimization problem. Thus, how to preserve and optimize the manifold structure of data still is a challenging problem. In this paper, we present a dynamic GCN model (MRDGCN) for human activity recognition and citation networks classification by combining the traditional manifold regularization framework. During the training process, we employ the alternative solution method to learn the optimal filter parameter and structure information, leading to a dynamic graph structure learning model. In addition, we further make a generalization for the layer-wise propagation rule of GCN, and then MRDGCN can extract richer sample features. Substantial experiment results on human activity datasets including CAS-YNU-MHAD, 2moons, USAA and citation networks datasets demonstrate that MRDGCN model outperforms the typical GCN.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

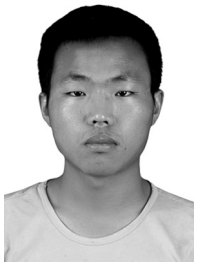
This work was supported in part by the National Natural Science Foundation of China under Grant 61671480, in part by the Fundamental Research Funds for the Central Universities, China University of Petroleum (East China) under Grant 18CX07011A and YCX2019080, in part by the the Open Project Program of the National Laboratory of Pattern Recognition (NLPR) (Grant No.202000009), in part by the Major Scientific and Technological Projects of CNPC under Grant ZD2019-183-008, in part by the Science and Technology Development Fund, Macau SAR (File No. 189/2017/A3), and by the Research Committee at University of Macau under Grants MYRG2016-00123-FST and MYRG2018-00136-FST.

## References

[1] W. Liu, Z.-J. Zha, Y. Wang, K. Lu, D. Tao, p-laplacian regularized sparse coding for human activity recognition, IEEE Transactions on Industrial Electronics 63 (2016) 5120–5129.
[2] Y. Guo, D. Tao, W. Liu, J. Cheng, Multiview cauchy estimator feature embedding for depth and inertial sensor-based human action recognition, IEEE Transactions on Systems, Man, and Cybernetics: Systems 47 (2016) 617–627.
[3] F. Baradel, C. Wolf, J. Mille, G.W. Taylor, Glimpse clouds: Human activity recognition from unstructured feature points, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 469–478.
[4] J. Yu, D. Tao, M. Wang, Adaptive hypergraph learning and its application in image classification, IEEE Transactions on Image Processing 21 (2012) 3262–3272.
[5] M. Alkanhal, G. Muhammad, Polynomial correlation filters for human face recognition, in: 2012 11th International Conference on Machine Learning and Applications, vol. 1, IEEE, 2012, pp. 646–650.
[6] Y. Duan, J. Lu, J. Feng, J. Zhou, Context-aware local binary feature learning for face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (2018) 1139–1153.
[7] I. Sutskever, G.E. Hinton, A. Krizhevsky, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems (2012) 1097–1105.
[8] H. Naceur, A. Delameziere, J. Batoz, Y. Guo, C. Knopf-Lenoir, Some improvements on the optimum process design in deep drawing using the inverse approach, Journal of Materials Processing Technology 146 (2004) 250–262.
[9] Z. Cai, M.J. Saberian, N. Vasconcelos, Learning complexity-aware cascades for pedestrian detection, IEEE Transactions on Pattern Analysis and Machine Intelligence (2019).
[10] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (2006) 504–507.
[11] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks, Advances in Neural Information Processing Systems (2007) 153–160.
[12] C.R. Qi, W. Liu, C. Wu, H. Su, L.J. Guibas, Frustum pointnets for 3d object detection from rgb-d data, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 918–927.
[13] O.L. Mangasarian, E.W. Wild, Multisurface proximal support vector machine classification via generalized eigenvalues, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (2005) 69–74.
[14] J. Huang, X. Yan, Quality-driven principal component analysis combined with kernel least squares for multivariate statistical process monitoring, IEEE Transactions on Control Systems Technology (2018) 1–8.
[15] E. Ing, W. Su, M. Schonlau, N. Torun, Support vector machines and logistic regression to predict temporal artery biopsy outcomes, Canadian Journal of Ophthalmology 54 (2019) 116–118.
[16] W. Liu, H. Liu, D. Tao, Y. Wang, K. Lu, Manifold regularized kernel logistic regression for web image annotation, Neurocomputing 172 (2016) 3–8.
[17] D. Tao, L. Jin, W. Liu, X. Li, Hessian regularized support vector machines for mobile image annotation on the cloud, IEEE Transactions on Multimedia 15 (2013) 833–844.
[18] W. Liu, X. Ma, Y. Zhou, D. Tao, J. Cheng, p-laplacian regularization for scene recognition, IEEE Transactions on Cybernetics (2018) 1–14.
[19] X. Ma, W. Liu, S. Li, D. Tao, Y. Zhou, Hypergraph p-laplacian regularization for remotely sensed image recognition, IEEE Transactions on Geoscience and Remote Sensing 57 (2019) 1585–1595.
[20] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, Advances in Neural Information Processing Systems (2016) 3844–3852.
[21] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, M.M. Bronstein, Geometric deep learning on graphs and manifolds using mixture model cnns, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5115–5124.
[22] H. Peng, J. Li, Y. He, Y. Liu, M. Bao, L. Wang, Y. Song, Q. Yang, Large-scale hierarchical text classification with recursively regularized deep graph-cnn, in: Proceedings of the 2018 World Wide Web Conference on World Wide Web, 2018, pp. 1063–1072.
[23] J. Bruna, W. Zaremba, A. Szlam, Y. LeCun, Spectral networks and locally connected networks on graphs, in: International Conference of Learning Representation, 2014..
[24] X. Liang, L. Lin, X. Shen, J. Feng, S. Yan, E.P. Xing, Interpretable structure-evolving lstm, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1010–1019.
[25] X. Wang, Y. Ye, A. Gupta, Zero-shot recognition via semantic embeddings and knowledge graphs, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6857–6866.
[26] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: International Conference of Learning Representation, 2017..
[27] S. Fu, W. Liu, S. Li, Y. Zhou, Two-order graph convolutional networks for semi-supervised classification, IET Image Processing (2019).
[28] N. Yadati, M. Nimishakavi, P. Yadav, A. Louis, P. Talukdar, Hypergcn: Hypergraph convolutional networks for semi-supervised classification, International Conference on Multimedia and Expo (2019).
[29] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, G.N. Hullender, Learning to rank using gradient descent, in: Proceedings of the 22nd International Conference on Machine Learning, 2005, pp. 89–96.
[30] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, in: International Conference of Learning Representation, 2017..
[31] S. Fu, W. Liu, Y. Zhou, L. Nie, Hplapgcn: Hypergraph p-laplacian graph convolutional networks, Neurocomputing 362 (2019) 166–174.
[32] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: A geometric framework for learning from labeled and unlabeled examples, Journal of Machine Learning Research 7 (2006) 2399–2434.
[33] J. Weston, F. Ratle, H. Mobahi, R. Collobert, Deep learning via semi-supervised embedding, in: Neural Networks: Tricks of the Trade, Springer, 2012, pp. 639–655.
[34] B. Perozzi, R. Al-Rfou, S. Skiena, Deepwalk: Online learning of social representations, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2014, pp. 701–710.
[35] X. Ji, J. Cheng, W. Feng, Spatio-temporal cuboid pyramid for action recognition using depth motion sequences, in: 2016 Eighth International Conference on Advanced Computational Intelligence, IEEE, 2016, pp. 208–213.
[36] Y. Fu, T.M. Hospedales, T. Xiang, S. Gong, Learning multimodal latent attributes, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (2013) 303–316.
[37] Y.-G. Jiang, G. Ye, S.-F. Chang, D. Ellis, A.C. Loui, Consumer video understanding: A benchmark database and an evaluation of human and machine performance, in: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, ACM, 2011, p. 29.
[38] C. Caragea, J. Wu, A. Ciobanu, K. Williams, J. Fernández-Ramírez, H.-H. Chen, Z. Wu, L. Giles, Citeseer x: A scholarly big dataset, European Conference on Information Retrieval, Springer (2014) 311–322.
[39] C. Cabanes, A. Grouazel, K.v. Schuckmann, M. Hamon, V. Turpin, C. Coatanoan, F. Paris, S. Guinehut, C. Boone, N. Ferry, et al., The cora dataset: validation and diagnostics of in-situ ocean temperature and salinity measurements, Ocean Science 9 (2013) 1–18..
[40] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: International Conference of Learning Representation, 2014..

[41] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, 2010, pp. 249–256.
[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The Journal of Machine Learning Research 15 (2014) 1929–1958.

**Weifeng Liu (M12CSM17)** received the double B.S. degrees in automation and business administration and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2002 and 2007, respectively. He was a Visiting Scholar with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, Australia, from 2011 to 2012. He is currently a Full Professor with the College of Information and Control Engineering, China University of Petroleum, Qingdao, China. He has authored or coauthored a dozen papers in top journals and prestigious conferences, including four Essential Science Indicators (ESI) highly cited papers and two ESI hot papers. His research interests include computer vision, pattern recognition, and machine learning. Prof. Liu serves as an Associate Editor for the Neural Processing Letters, the Co-Chair for the IEEE SMC Technical Committee on Cognitive Computing, and a Guest Editor for special issue of the Signal Processing, the IET Computer Vision, the Neurocomputing, and the Remote Sensing. He also serves over 20 journals and over 40 conferences.



**Sichao Fu** is currently pursuing the master's degree with the College of Control Science and Engineering, China University of Petroleum, Qingdao, China. His research interests include pattern recognition and computer vision.

**Yicong Zhou (M'07-SM'14)** received the B.S. degree in electrical engineering from Hunan University, Changsha, China, and the M.S. and Ph.D. degrees in electrical engineering from Tufts University, Medford, MA, USA. He is currently an Associate Professor and the Director with the Vision and Image Processing Laboratory, Department of Computer and Information Science, University of Macau, Macau, China. His research interests include chaotic systems, multimedia security, image processing and understanding, and machine learning. Dr. Zhou was a recipient of the Third Price of Macau Natural Science Award in 2014. He served as an Associate Editor for the Neurocomputing, the Journal of Visual Communication and Image Representation, and the Signal Processing: Image Communication. He is a Co-Chair of Technical Committee on Cognitive Computing in the IEEE Systems, Man, and Cybernetics Society.

**Zheng-Jun Zha** received the B.E. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2004 and 2009, respectively. He is currently a Full Professor with the School of Information Science and Technology, University of Science and Technology of China, the Vice Director of National Engineering Laboratory for Brain-Inspired Intelligence Technology and Application. He was a Researcher with the Hefei Institutes of Physical Science, Chinese Academy of Sciences, from 2013 to 2015, a Senior Research Fellow with the School of Computing, National University of Singapore (NUS), from 2011 to 2013, and a Research Fellow there from 2009 to 2010. He has authored or coauthired more than 100 papers in these areas with a series of publications on top journals and conferences. His research interests include multimedia analysis, retrieval and applications, as well as computer vision etc. Prof. Zha was the recipient of multiple paper awards from prestigious multimedia conferences, including the Best Paper Award and Best Student Paper Award in ACM Multimedia, etc.

**Liqiang Nie** received the B.E. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, and National University of Singapore (NUS), Singapore, in 2009 and 2013, respectively. He is currently a Professor with the School of Computer Science and Technology, Shandong University, Jinan, China. He was with the NUS as a Research Fellow for more than three years. His research interests include information retrieval, multimedia computing, and their applications in the field of healthcare.