

Group-Wise Learning for Aurora Image Classification With Multiple Representations

Jun Zhang¹, Mingxia Liu², Ke Lu, and Yue Gao, *Senior Member, IEEE*

Abstract—In conventional aurora image classification methods, it is general to employ only one single feature representation to capture the morphological characteristics of aurora images, which is difficult to describe the complicated morphologies of different aurora categories. Although several studies have proposed to use multiple feature representations, the inherent correlation among these representations are usually neglected. To address this problem, we propose a group-wise learning (GWL) method for the automatic aurora image classification using multiple representations. Specifically, we first extract the multiple feature representations for aurora images, and then construct a graph in each of multiple feature spaces. To model the correlation among different representations, we partition multiple graphs into several groups via a clustering algorithm. We further propose a GWL model to automatically estimate class labels for aurora images and optimal weights for the multiple representations in a data-driven manner. Finally, we develop a label fusion approach to make a final classification decision for new testing samples. The proposed GWL method focuses on the diverse properties of multiple feature representations, by clustering the correlated representations into the same group. We evaluate our method on an aurora image data set that contains 12 682 aurora images from 19 days. The experimental results demonstrate that the proposed GWL method achieves approximately 6% improvement in terms of classification accuracy, compared to the methods using a single feature representation.

Index Terms—Aurora image classification, group-wise learning (GWL), multiple representations, statistical image features.

I. INTRODUCTION

AS A PHYSICAL process happening in the near-Earth environment, aurora is a natural phenomenon that can

Manuscript received May 29, 2018; revised December 10, 2018 and February 24, 2019; accepted February 25, 2019. Date of publication March 28, 2019; date of current version August 4, 2021. This work was supported in part by the National Natural Science Foundations of China under Grant 61703301, Grant U1701262, Grant 61671267, and Grant 61571353, in part by the National Key Research and Development Program of China under Grant 2017YFC0113000, in part by the Beijing Natural Science Foundation under Grant 4182022, in part by the Taishan Scholar Program of Shandong Province in China, and in part by the Scientific Research Foundation of Taishan University under Grant Y-01-2018019. This paper was recommended by Associate Editor X. Li. (Corresponding authors: Mingxia Liu; Yue Gao.)

J. Zhang is with the AI Platform Department, Tencent, Shenzhen 518057, China (e-mail: xdzhangjun@gmail.com).

M. Liu is with the Department of Information Science and Technology, Taishan University, Tai'an 271021, China (e-mail: mxliu1226@gmail.com).

K. Lu is with the School of Computer and Communication Engineering, University of Chinese Academy of Science, Beijing 100049, China (e-mail: luke0313@gmail.com).

Y. Gao is with the School of Software, Tsinghua University, Beijing 100084, China (e-mail: gaoyue@tsinghua.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCYB.2019.2903591>.

Digital Object Identifier 10.1109/TCYB.2019.2903591

be directly viewed on the ground [1]. Besides, the visual feast produced by the colorful fluctuating, aurora also provides an effective way to investigate the activities of the Sun, because it is mostly and ultimately driven by the solar wind [2]–[4]. Particularly, analyzing morphologies of aurora images (e.g., aurora classification) is significantly meaningful, because of the close correlation between the aurora morphology and the inherent physical mechanism.

However, manually classifying aurora into different categories is time-consuming and tedious. Luckily, there exist many state-of-the-art feature extraction and classification methods that can be potentially used for automatic aurora image classification. An optimistic outlook on the aurora classification task is that it provides a large image dataset, creating an interesting synergy between computer vision and space science [5]. Specifically, the classification methods in the domains of machine learning and computer vision provide solutions for automatic aurora data analysis. Meanwhile, the challenging dataset provided by aurora analysis also motivates researchers to develop advanced learning models in the domain of computer vision. Existing aurora image classification methods usually extract different types of features to represent those aurora images. Specifically, single representation-based methods only extract one single feature representation [5]–[8], while multiple representations-based approaches adopt multiple feature representations for aurora images [9]. In general, the major challenges in existing aurora image classification can be summarized as follows.

- 1) Single representation-based methods cannot effectively capture the characteristics of eventful aurora morphology.
- 2) Multiple representation-based methods usually neglect the inherent correlation among multiple types of feature representations.

To address those problems, we propose a group-wise learning (GWL) method for automatic aurora image classification. Specifically, we first adopt several state-of-the-art feature extraction methods to extract multiple representations for aurora images, including local binary pattern (LBP) [10], local energy pattern (LEP) [11], VZ-maximum response 8 (MR8) [12], VZ-Joint [12], principal curvatures (PCs) [13], and basic image features (BIFs) [14]. In this way, more discriminative structural information can be possibly extracted by different feature representations. We then construct multiple graphs based on multiple representations, with each graph corresponding to a specific representation. Then, we develop a feature representation clustering method to partition multiple

representations into different groups via a spectral clustering algorithm [15], through which the underlying correlation among representations can be modeled explicitly. We further propose a multigraph-based classification model, where multiple graphs in the same group are integrated into a unified objective function. Moreover, the optimal weight for each representation can be automatically learned from our method. Finally, we develop a label fusion method to assemble the classification results obtained from different representations.

The major contribution of this paper is twofold. First, we investigate the efficacy of using different state-of-the-art feature representations and classifiers for aurora image classification. It demonstrates that multiple feature representations provide complementary information that can improve the classification performance. Second, to fuse multiple representations efficiently, we propose a GWL model to automatically learn optimal weights for multiple feature representations within each group, as well as estimated class labels for aurora images, in a data-driven manner. The merit of the proposed GWL is that we focus on making use of diverse properties of multiple representations, by first clustering correlated representations into the same group and then treating each group equally for making final classification decisions. In particular, the optimal weights for those correlated representations in the same group can be learned from data automatically. The proposed GWL framework can also be potentially used for other learning problems with multiple feature representations.

The remainder of this paper is organized as follows. In Section II, we briefly review the related work. Section III presents our proposed GWL framework for aurora image classification in detail. In Section IV, we first introduce the aurora image database used in this paper, and then report the experimental results achieved by our proposed method and several state-of-the-art methods. Section V discusses several important issues and possible future research directions. Finally, the conclusion is given in Section VI.

II. RELATED WORK

A. Aurora Image Classification

In the literature, existing studies for aurora image classification can be roughly divided into two categories: 1) single representation-based methods and 2) multiple representations-based approaches. In the first category, there are several important works. For example, Syrjäsuo *et al.* [16] adopted shape skeletons of the aurora forms and hierarchical attribute trees to classify aurora images. Syrjäsuo *et al.* [5]–[7] proposed a serious work that focuses on using Fourier descriptors to represent the aurora images. Wang *et al.* [8] extracted LBP features with a block partition scheme, and adopted the nearest neighbor classifier for classification. Wang *et al.* [17] proposed a feature extraction method named X-gray level aura matrices, and the support vector machine (SVM) classifier was employed to perform classification. However, these methods only adopt single feature representation for aurora images, where characteristics of eventful aurora morphology cannot be effectively captured by only one single feature representation. As a typical exemplar of the second category,

Yang *et al.* [9] proposed a multiple representation based on polar scale-invariant feature transform and polar deep LBP descriptor with an application for aurora image retrieval. In their work, features from two representations were simply concatenated, where the correlation among multiple feature representations are not incorporated into the learning model. Later, Yang *et al.* [18], [19] further employed the convolutional neural network (CNN) features for aurora image indexing. Generally, features were seldom combined according to their associations. Intuitively, modeling the inherent correlation among different representation could further boost the learning performance for aurora image classification.

B. Image Feature Representation

In recent years, many feature extraction approaches have been proposed for image representation in computer vision domain. For example, LBP [10] and its variants [13], [20], [21]; texon dictionary-based representations (or named codebook-based representations) [12], [13], [22], [23]; and wavelet-based representations [24]–[28]. Among them, several feature representations have been used in aurora image classification. In the following, we briefly introduce several widely used feature representation methods that have been or can be potentially applied to aurora images.

1) *Local Binary Pattern*: Image representations based on LBP have been broadly used for texture classification [10], face recognition [29], and object detection [20], due to its simplicity and efficiency. The basic idea of LBP can be concluded into two major steps: 1) calculating local derivatives and 2) binary coding. Specifically, assume that the intensity value of centroid pixel is g_c , the intensity values of its neighboring pixel is g_p ($p = 1, 2, \dots, P$), where P is the number of neighboring pixels. The joint distribution is

$$T = t(s(g_0 - g_c), s(g_1 - g_c), \dots, s(g_{P-1} - g_c)) \quad (1)$$

where

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0. \end{cases} \quad (2)$$

Using the binary coding, the distribution can be transformed into a decimal number

$$\text{LBP}_P = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p. \quad (3)$$

Generally, the statistical histogram of the numbers is regarded as a feature vector. By merging some histogram bins with the uniform and rotation invariant constraints, the representation can be further rectified to be more compact and has the property of rotation invariance for specific applications.

2) *Local Energy Pattern*: LEP adopts the oriented energies to generate local descriptors and uses an n -ray coding strategy for vector quantization (VQ). Specifically, the oriented energies are generated by the responses of second-order Gaussian-like derivative filters. Due to the steerable property of Gaussian-like filters, the filtering response in any arbitrary orientations (i.e., θ) can be obtained by the linear combination of the responses of basis filters, that is, G_{xx} , G_{xy} , and G_{yy} .

Therefore, the filtering response $G_{2\theta}$ is defined as

$$G_{2\theta} = \cos^2(\theta)G_{xx} - 2\cos(\theta)\sin(\theta)G_{xy} + \sin^2(\theta)G_{yy} \quad (4)$$

where θ can be sampled at proper intervals with the bound of $0 \leq \theta \leq \pi$.

Then, the filtering responses are rectified into energies by a serious operation of nonlinearity, smoothing, and normalization. In this way, the local descriptor has the ability of brightness and contrast invariance. In order to calculate the appearing frequency of the local descriptor, the local descriptor is transformed into a coding number by n -ray coding, similar to the operation in LBP. Finally, the statistical histogram of the coding numbers is also regarded as the presentation feature vector.

3) *VZ-MR8, VZ-Joint, and Principle Curvatures*: Varma and Zisserman [12], [22] proposed a texton dictionary-based image representation method using the responses of MR8 filter set (it includes 38 Gaussian and Laplacian of Gaussian filters, but only 8 filter responses) and image patches (the raw pixel intensities of a square neighborhood around the selected point are taken, being flattened to form a vector), respectively. For convenience, the two types of features are named VZ-MR8 and VZ-Joint, respectively. PCs are generated by calculating local PCs with multiple scales [13]. These three approaches adopted the similar texton-based framework for image representation. Specifically, local structures of an image are directly described by the responses MR8 filter set, image patches, or PCs, where each local descriptor (v) is normalized by Weber's law

$$v \leftarrow v[\log(1 + \|v\|_2/0.03)]/\|v\|_2. \quad (5)$$

Massive local patches from each category are aggregated together and clustered into different groups. The clustering centers are used to construct a texton dictionary. The textons thereby directly represent the local structures. For each image, each local structure can be labeled by its most similar texton, and the appearing frequency of these textons is regarded as the representation.

4) *Basic Image Features*: BIFs are constructed with six basis filtering responses from one zero-order (i.e., c_{00}), two first-order (i.e., c_{01} and c_{10}), and three second-order (i.e., c_{20} , c_{11} , and c_{02}) Gaussian derivative filters. Then, the local descriptor is calculated as

$$\left\{ \varepsilon s_{00}, 2\sqrt{s_{10}^2 + s_{01}^2}, \pm\lambda, 2^{-\frac{1}{2}}(\gamma \pm \lambda), \gamma \right\} \quad (6)$$

where $s_{ij} = \frac{\sigma^{i+j}c_{ij}}{\sqrt{(s_{20} - s_{02})^2 + 4s_{11}^2}}$, $\lambda = s_{20}^2 + s_{02}^2$, and $\gamma = \sqrt{(s_{20} - s_{02})^2 + 4s_{11}^2}$. In general, each element corresponds to a specific local structure. In a real-world application, the authors suggested to set $\varepsilon = 0$ and adopt a multiscale representation. For each scale, the maximum value is selected from six elements. If there are four scales, it will result in a visual dictionary $6^4 = 1296$ BIF-columns.

C. Graph Learning

In the past decades, graph learning [30]–[38] have achieved promising performance in many applications. In a graph, a

sample is represented by a vertex in a graph, and one edge connects a pair of vertices based on some distance measure metric (e.g., Euclidean distance), through which the pairwise relationships among samples can be captured. Such a pairwise relationship can partially represent the data structure *locally*. In graph learning methods, a smoothing constraint is usually implemented *globally* with the spectral transformation of the graph Laplacian [39]–[47].

In the literature, there are many representative graph learning methods. In an unsupervised manner, normalized cuts [48] considers both dissimilarity between different groups and similarity within the same group. Different from normalized cuts, graph cuts [49] uses a max-flow/min-cut optimization instead of traditional graph partitioning algorithms to minimize the energy function. Both of them are very basic tools for solving the perceptual grouping problem in the domain of computer vision. In a semisupervised manner, graph mincuts [50] minimizes the number of similar pairs of vertices that are given different labels. Harmonic energy minimization [51] formulates a Gaussian random field on the graph, where the mean of the field is characterized in terms of harmonic functions. Spectral graphical transducer [52] provides a transductive version of the K -nearest-neighbor (KNN) classifier using graph learning. Deng *et al.* [53] proposed a weakly supervised multigraph learning for image ranking by considering both intragraph and intergraph constraints.

III. PROPOSED METHOD

Fig. 1 illustrates our proposed GWL method for aurora image classification using multiple feature representations. As shown in Fig. 1, we first extract M feature representations for each aurora image. Then, multiple graphs can be built based on these features, with each graph corresponding to a specific feature representation. A clustering process is performed to partition M representations into G groups. We then perform multigraph-based classification in each group, followed by a label fusion process to make a final classification decision for a new testing sample.

A. Feature Extraction

We extract six types of feature representations for each aurora image, including LBP, LEP, VZ-MR8, VZ-Joint, PCs, and BIFs features. Note that these six feature representations are histogram-based features that calculate statistical occurrence frequencies of local structures. There are three main steps in our feature extraction approach (see Fig. 2), which are listed as follows.

- 1) *Local Structure Description*: We describe local structures of aurora images using different local descriptors, that is, local derivatives for LBP, oriented energies for LEP, MR8 filtering responses for VZ-MR8, image patches for VZ-Joint, principal curvatures for PCs, and basic image features for BIFs.
- 2) *Statistical Frequencies Computation*: In order to count the statistical frequencies of these local descriptors in an image, several VQ strategies are applied to transform the local descriptor vector into a coding number. For example, binary coding for LBP, n -ary coding for LEP, texton

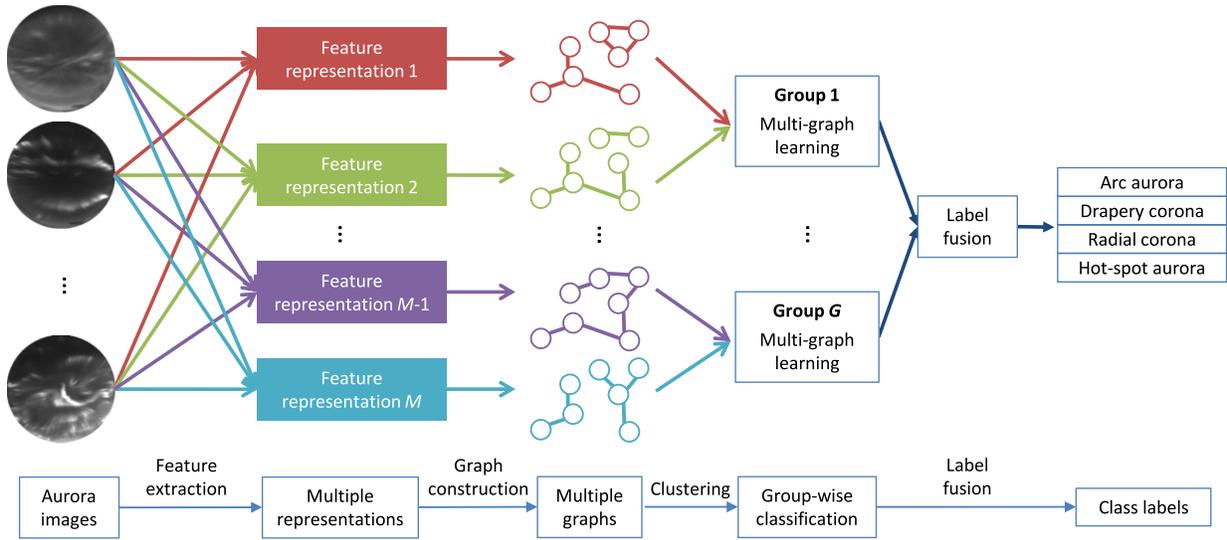


Fig. 1. Framework of the proposed GWL method for aurora image classification.

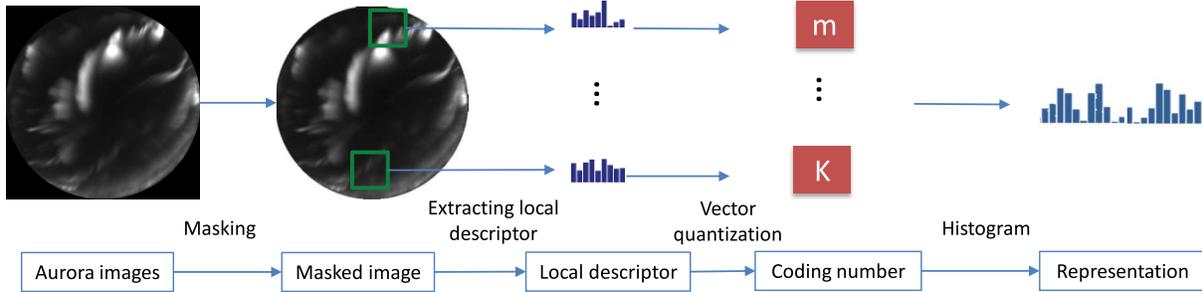


Fig. 2. Pipeline of feature extraction for multiple representations of aurora images.

dictionary coding for VZ-MR8, VZ-Joint, and PCs, and maximum response selection for BIFs. Particularly, in this paper, we only count the frequency within a circle mask where the aurora appears in the image (see Fig. 2).

3) *Image Representation*: The histogram of the coding numbers in the image is regarded as the image representation. Given six types of image representation, we can finally obtain six feature vectors as the representation for each aurora image.

B. Multigraph Construction

To incorporate the structural information of data into subsequent learning model, we propose to use the graph structure to explicitly model the pairwise relationship among individual samples. Specifically, we first construct a graph using one specific feature representation of aurora images, and then compute the graph Laplacian matrix. We denote boldface uppercase letters, boldface lowercase letters, and normal italic letters as matrices, vectors, and scalars, respectively.

Assume the studied aurora images with the m th ($m = 1, \dots, M$) representation are represented by $\mathbf{X}^m = \{\mathbf{x}_1^m, \mathbf{x}_2^m, \dots, \mathbf{x}_N^m\} \in \mathbb{R}^{N \times d^m}$, where M is the feature representation number, N denotes the total number of aurora images, and d^m is the feature dimension. Here, $\mathbf{x}_n^m \in \mathbb{R}^{d^m}$ is the m th

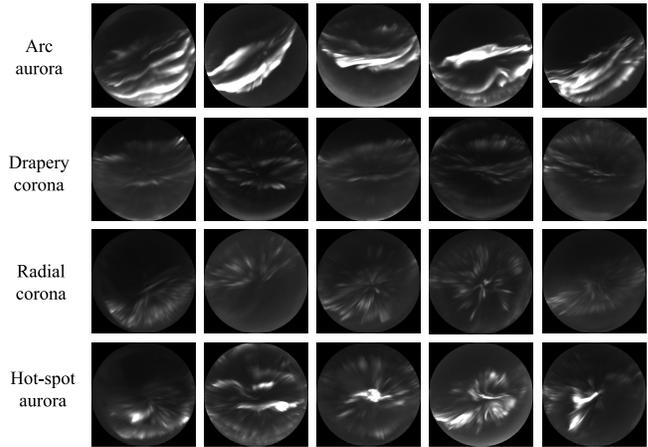


Fig. 3. Samples of four categories from the aurora database.

feature representation for the n th aurora image. In graph learning, each aurora image is regarded as a vertex, where the data structure is modeled by edges that connect samples according to some similarity measurement.

Given a vertex set \mathcal{V} where each vertex represents a sample, an edge set \mathcal{E} with each one connecting two vertices, and a weight vector $\mathbf{w} \in \mathcal{R}^{N_e}$ for N_e edges, an undirected weighted graph is represented by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ without

loops (i.e., \mathcal{G} is a simple graph). It is worth noting that an edge in a graph connects two vertices, through which pairwise relationship among samples can be modeled explicitly [30], [31], [46], [54]. There are non-negative weights $w_{i,j} \geq 0$ for the edge, satisfying $w_{i,j} = 0$ if \mathbf{x}_i and \mathbf{x}_j are not connected. In general, a larger value of the edge weight $w_{i,j}$ suggests that the i th and the j th samples are more similar to each other, and vice versa. Hence, the edge weight is usually assigned by the similarity between a pair of samples [46]. The matrix $\mathbf{W} = (w_{i,j})$ is called the *adjacency matrix*, and the Laplacian matrix \mathbf{L} is defined as [30], [54]

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \quad \mathbf{D} = \text{diag} \left(\sum_{j \neq 1} w_{1,j}, \dots, \sum_{j \neq N} w_{N,j} \right) \quad (7)$$

where $\mathbf{L} = (l_{i,j})$ is called the *graph Laplacian*. In an undirected weighted graph, the matrices \mathbf{W} and \mathbf{L} are symmetric.

A key problem in graph learning is to construct edges based on a specific similarity/dissimilarity matrix, to ensure that two similar samples can be connected by an edge. Since we adopt histogram-based representation for aurora images in this paper, conventional distance measurements (e.g., Euclidean distance) are not suitable for our case. In the literature, there are several ways to measure the distance between a pair of samples represented by histogram-based features, for example, Chi-square distance [55], histogram intersection [56], and log-likelihood statistic. In this paper, we calculate the dissimilarity between representations for two aurora images (i.e., \mathbf{u} and \mathbf{v}) with Chi-square distance, which is defined as follows:

$$\chi^2(\mathbf{u}, \mathbf{v}) = \sum_k \frac{(u^k - v^k)^2}{u^k + v^k} \quad (8)$$

where k is the index of the feature vector for each image.

Based on the Chi-square distance defined in (8), we adopt the widely used heat-kernel for adjacency matrix construction. Specifically, the element $w_{i,j}^m$ of the adjacency matrix \mathbf{W}^m is defined as

$$w_{i,j}^m = e^{-\frac{\chi^2(\mathbf{x}_i^m, \mathbf{x}_j^m)}{\sigma^2}}. \quad (9)$$

Given M types of feature representations, we can compute M adjacency matrices $\{\mathbf{W}^m\}_{m=1}^M$, where each one is constructed based on a specific feature representation.

C. Graph Grouping via Clustering

In this paper, we extract M types of feature representations to represent each aurora image. While each representation may own its unique discriminative information in representing morphologies of aurora images, some of them could be correlated to some extent. To explicitly model such correlation information could benefit the subsequent classification model, since more prior information can be incorporated into the learning process. Here, we propose to adopt a clustering algorithm to partition M representations into G groups. Specifically, based on the adjacency matrices $\{\mathbf{W}^m\}_{m=1}^M$ constructed in M representations, we first compute the correlation coefficients of those M adjacency matrices. In this way, we

can obtain a correlation matrix $\Omega \in \mathcal{R}^{M \times M}$, with each element denoting the correlation coefficient of a pair of adjacency matrices. Based on the correlation matrix Ω , we adopt the spectral clustering algorithm [15] to partition M representations into G groups. Then, in each group, we perform group-wise classification using the following proposed model.

D. Group-Wise Classification

Since it is difficult to determine which representation can better represent the morphology of aurora images, we now propose a group-wise classification model, by integrating multiple graphs (constructed in multiple feature representations) into a unified objective function in each of G groups.

1) *Graph Laplacian Regularization*: Assume we have a label matrix $\mathbf{Y} \in \mathbb{R}^{N \times C}$, where each element $y_{i,j} = 1$ if the sample \mathbf{x}_i is labeled as the j th category and $y_{i,j} = 0$ otherwise, and C is the total number of categories. Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be a matrix with element $a_{i,j}$. We denote the Frobenius norm of \mathbf{A} as $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} |a_{i,j}|^2}$, and the trace of \mathbf{A} as $\text{tr}(\mathbf{A}) = \sum_{i=1}^N a_{i,i}$. Let \mathbf{b} represent a vector, and its l_2 -norm is defined as $\|\mathbf{b}\|_2 = \sqrt{\sum_{i=1}^N |b_i|^2}$.

Suppose $\mathbf{F}^m \in \mathbb{R}^{N \times C}$ is the estimated label matrix in the m th representation, where the element in the c th column \mathbf{f}_c^m of \mathbf{F}^m is the estimated probability score belonging to the c th class for N samples. Intuitively, if two samples \mathbf{x}_i^m and \mathbf{x}_j^m are similar in the m th feature space, their estimated class labels should also be similar. Based on this assumption, we propose the following *graph Laplacian regularizer*:

$$\sum_{c=1}^C \sum_{i,j=1}^N w_{i,j}^m (\mathbf{f}_{c,i}^m - \mathbf{f}_{c,j}^m)^2 = \sum_{c=1}^C (\mathbf{f}_c^m)^\top \mathbf{L}^m \mathbf{f}_c^m \quad (10)$$

where the weight $w_{i,j}^m$ is the (i, j) th element of the adjacency matrix \mathbf{W}^m . Note that, a large value of $w_{i,j}^m$ suggests that the i th and the j th samples are similar in the m th feature space. In this way, the graph Laplacian regularizer defined in (10) is used to encourage that, if the i th and the j th samples are similar, their estimated probability scores should also be close to each other in the label space.

Although many of existing studies have shown that multiple feature representations can provide complementary information [11], it is still hard to determine the optimal weights for different feature representations (with respect to different graphs). To address that problem, we propose to learn the weights for different representations from data automatically, with details given in the following section.

2) *Classification Model*: Given the true class label matrix \mathbf{Y} and the estimated class label matrix $\{\mathbf{F}_g^m\}_{m=1}^{M_g}$ in the g th group, we define the following cost function:

$$\sum_{m=1}^{M_g} \sum_{c=1}^C \left\| \mathbf{f}_{g,c}^m - \mathbf{y}_c \right\|_2^2 = \sum_{m=1}^{M_g} \left\| \mathbf{F}_g^m - \mathbf{Y} \right\|_F^2 \quad (11)$$

where $\mathbf{f}_{g,c}^m$ and \mathbf{y}_c are the c th column of \mathbf{F}_g^m and the c th column of \mathbf{Y} , respectively.

Denote $\boldsymbol{\alpha}_g \in \mathbb{R}^{M_g}$ as a weight vector in the g th group, with its element α_g^m representing the weight for the m th

graph constructed in the m th feature space. By considering the structure information conveyed by graph Laplacian regularizer defined in (10), we propose the following multigraph-based classification model:

$$\begin{aligned} \min_{\{\mathbf{F}_g^m, \alpha_g^m\}_{m=1}^{M_g}} & \mu \sum_{m=1}^{M_g} \left\| \mathbf{F}_g^m - \mathbf{Y} \right\|_F^2 \\ & + \sum_{m=1}^{M_g} (\alpha_g^m)^2 \sum_{c=1}^C (\mathbf{f}_{g,c}^m)^\top \mathbf{L}_g^m \mathbf{f}_{g,c}^m \\ \text{s.t.} & \sum_{m=1}^{M_g} \alpha_g^m = 1, \alpha_g^m \geq 0 \end{aligned} \quad (12)$$

where the first term is the empirical loss in the g th group with M_g feature representation, and the second one is the graph Laplacian regularizer [34]. The regularization parameter $(\alpha_g^m)^2$ is used to prevent the degenerate solution of $\alpha_g = \{\alpha_g^1, \dots, \alpha_g^{M_g}\}$, and μ is employed to tradeoff the contribution of two terms in (12). The constraints in (12) are used to penalize the complexity of the weights (i.e., α_g) for different representations. With (12), one can jointly learn the class probability scores $\{\mathbf{F}_g^m\}_{m=1}^{M_g}$, as well as the optimal weights α_g for multiple feature representations.

As mentioned in Section III-C, we cluster M representations as G groups, where the inherent correlation among different representations can be explicitly modeled to some extent. Then, we propose to perform classification via the model defined in (12) in each of G groups separately, and finally assemble the classification results using a label fusion method (see Section III-D4).

3) *Alternating Optimization Algorithm*: Since the problem defined in (12) is not jointly convex with respect to $\{\mathbf{F}_g^m\}_{m=1}^{M_g}$ and α_g , in this paper, we propose an alternating optimization approach to solve the proposed objective function. To be specific, we first optimize \mathbf{F}_g^m , given a fixed α_g . Then, we optimize α_g with a fixed \mathbf{F}_g^m .

In the first step, given α_g , the objective function in (12) can be rewritten as

$$\min_{\mathbf{F}_g^m} \mu \sum_{m=1}^{M_g} \left\| \mathbf{F}_g^m - \mathbf{Y} \right\|_F^2 + \sum_{m=1}^{M_g} (\alpha_g^m)^2 \text{tr} \left((\mathbf{F}_g^m)^\top \mathbf{L}_g^m \mathbf{F}_g^m \right). \quad (13)$$

The partial derivative of the objective function in (13) with respect to \mathbf{F}_g^m is as follows:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{F}_g^m} & \left\{ \mu \sum_{m=1}^{M_g} \left\| \mathbf{F}_g^m - \mathbf{Y} \right\|_F^2 + \sum_{m=1}^{M_g} (\alpha_g^m)^2 \text{tr} \left((\mathbf{F}_g^m)^\top \mathbf{L}_g^m \mathbf{F}_g^m \right) \right\} \\ & = 2\mu \mathbf{F}_g^m - 2\mu \mathbf{Y} + 2(\alpha_g^m)^2 \mathbf{L}_g^m \mathbf{F}_g^m = 0 \end{aligned} \quad (14)$$

$$\Rightarrow \mathbf{F}_g^m = \left(\mathbf{I} + \frac{(\alpha_g^m)^2}{\mu} \mathbf{L}_g^m \right)^{-1} \mathbf{Y} \quad (15)$$

where $\mathbf{I} \in \mathbb{R}^{N \times N}$ is an identity matrix.

Given a fixed \mathbf{F}_g^m , we then optimize α_g in the second step. In such a case, the problem in (12) can be rewritten as

$$\begin{aligned} \min_{\{\alpha_g^m\}_{m=1}^{M_g}} & \sum_{m=1}^{M_g} (\alpha_g^m)^2 \text{tr} \left((\mathbf{F}_g^m)^\top \mathbf{L}_g^m \mathbf{F}_g^m \right) \\ \text{s.t.} & \sum_{m=1}^{M_g} \alpha_g^m = 1, \forall \alpha_g^m \geq 0. \end{aligned} \quad (16)$$

The partial derivative of (16) with respect to α_g^m is as follows:

$$\begin{aligned} \frac{\partial}{\partial \alpha_g^m} & \left\{ \sum_{m=1}^{M_g} (\alpha_g^m)^2 \text{tr} \left((\mathbf{F}_g^m)^\top \mathbf{L}_g^m \mathbf{F}_g^m \right) + \tau \left(\sum_{m=1}^{M_g} \alpha_g^m - 1 \right) \right\} \\ & = 0 \end{aligned} \quad (17)$$

$$\begin{aligned} \Rightarrow \alpha_g^m & = - \frac{\tau}{2 \text{tr} \left((\mathbf{F}_g^m)^\top \mathbf{L}_g^m \mathbf{F}_g^m \right)} \\ \tau & = - \frac{2 \prod_{m=1}^{M_g} \text{tr} \left((\mathbf{F}_g^m)^\top \mathbf{L}_g^m \mathbf{F}_g^m \right)}{\sum_{m=1}^{M_g} \prod_{p=1, p \neq m}^{M_g} \text{tr} \left((\mathbf{F}_g^p)^\top \mathbf{L}_g^p \mathbf{F}_g^p \right)}. \end{aligned} \quad (18)$$

4) *Label Fusion*: Given M_g feature representations, we can obtain M_g estimated probability scores for new testing samples in the g th group via (12). With the optimal weights for different representations learned in (12), we develop a label fusion approach to make a final classification result for testing samples. Specifically, given a new testing sample \mathbf{z} , its class label can be estimated by the weighted mean of its estimated class probability scores via

$$l(\mathbf{z}) = \underset{c=1, \dots, C}{\text{argmax}} \sum_{g=1}^G \sum_{m=1}^{M_g} \alpha_g^m f_{g,c}^m(\mathbf{z}) \quad (19)$$

where α_g^m is the optimal weight of the m th feature representation learned from (12) in the g th group, and $f_{g,c}^m(\mathbf{z})$ is the estimated probability score in the g th group belonging to the c th category achieved in the m th feature representation space for the sample \mathbf{z} . Note that M_g is the number of representations in the g th group, and $\sum_{g=1}^G M_g = M$.

5) *Computational Complexity Analysis*: Now, we analyze the computational complexity of the proposed method that contains four main steps as shown in Algorithm 1. In the first step, we need to construct M graphs based on M types of feature representations and compute the correlation matrix, requiring $O(MN^2)$ operations. In the second step, we perform representation clustering using the spectral clustering algorithm, which needs requires $O(N^2)$ operations. In the third step, we compute \mathbf{F}_g^m via (15) that requires $O(M_g N^2)$ operations. Also, the last step requires $O(CM)$ operations, respectively. Since both the number of representations (i.e., $M = 6$) and the number of categories (i.e., $C = 4$) are much smaller than the number of sample size (i.e., N), the overall computational complexity of Algorithm 1 is $O(N^2)$.

Algorithm 1: GWL for Aurora Image Classification

Input: Labeled aurora images with M ($M = 6$ in this study) feature representations, including LBP, LEP, VZ-MR8, VZ-Joint, PCs, and BIFs features; class label matrix $\mathbf{Y} \in \mathbb{R}^{N \times C}$; parameters μ and σ .

- 1 Step 1: Initialization
 - 2 1.1: Compute the adjacency matrices $\{\mathbf{W}^m\}_{m=1}^M$ based on M types of feature representations via Eq. (9), and the Laplacian matrices via Eq. (7);
 - 3 1.2: Construct the correlation matrix $\Omega \in \mathcal{R}^{M \times M}$ between M adjacency matrices $\{\mathbf{W}^m\}_{m=1}^M$.
 - 4 1.3: Set $\{\alpha_g\}_{g=1}^2$ with initial values;
- 5 Step 2: Clustering of multiple representations. We perform spectral clustering based on the correlation matrix Ω to partition M representations into G groups.
- 6 Step 3: Group-wise classification. In each of G groups, we perform multi-graph based classification via Eq. (12), with details listed below.
- 7 **repeat**
 - 8 3.1: Label update. Compute \mathbf{F}_g^m using Eq. (15);
 - 9 3.2: Representation weight update. Compute the weight for different feature representations (i.e., $\alpha_g = \{\alpha_g^1, \dots, \alpha_g^{M_g}\}$ via Eq. (18);
- 10 **until convergence**;
- 11 Step 4: Label fusion. For a given testing sample \mathbf{z} , we can obtain its class label via Eq. (19).

Output: $\{\mathbf{F}_g^m\}_{m=1}^{M_g}$, α_g , and class label $l(\mathbf{z})$ for a given testing sample \mathbf{z} .

IV. EXPERIMENTS

A. Database

The aurora image data used in this paper were collected from the all-sky imagers at Chinese arctic station, Yellow River Station (YRS), in Ny-Ålesund, Svalbard. The all-sky images were acquired using three-wavelength (i.e., 427.8, 557.7, and 630.0 nm) all-sky cameras. According to [57], the dayside oval was divided into four auroral active regions, including the prenoon green warm spot (i.e., region W), the midday green gap (i.e., region M), the post-noon hot spot for auroral emission at three wavelengths (i.e., region H), and the dusk green aurora sector (i.e., region D). Generally, each region is associated with the typical structures of aurora. For example, the quickly poleward moving or east-west brightening discrete rayed arcs usually appear in region W , drapery-like corona or red radial corona appear in region M , quasi-periodic brightening rayed bands with poleward moving, isolated brightening rayed bundles or brightening arcs appear in region H , and multiple, parallel east-west extended arcs appear in region D [8]. These images were acquired from December 2003 to February 2009. The images in 557.7 nm were selected as the dataset.

Following [8], we preprocessed all aurora images by the following three steps.

- 1) *Subtracting Dark Current*: Dark current is deemed as system noise caused by equipment. The operation was a linear stretching that each image was stretched with a cutoff value of 4000 and rescaled to the range of [0, 255].

- 2) *Image Cropping*: A circle mask with a radius of 220 pixels was used to mask out the noisy regions where wide-angle distortion happens and may contain YRS lights. Then, the original image size of 512×512 pixels was further cropped to 440×440 pixels.
- 3) *Rotation*: We rotated the images counter-clockwise by 61.11 with north being upward since east-west structure is dominant.

We labeled 12682 images that happened in 19 days into four categories, namely, *arc aurora*, *drapery corona aurora*, *radial corona aurora*, and *hot-spot aurora*. Specifically, there are 6142 images from arc aurora, 2722 images from drapery corona aurora, 2554 images from radial corona aurora, and 1264 images from hot-spot aurora. The typical images of four types of aurora are shown in Fig. 3.

B. Experimental Settings

Since aurora is a gradually changing process, the aurora images that have short time interval may have strong appearance relationship. If we randomly select images as training and testing, there is a high possibility that there exist similar images in training dataset that are related to the testing images. Therefore, we conduct experiments using a leave-one-day-out strategy. Specifically, given images from 19 days, we employ images from 18 days for training and the rest for testing. The process is repeated 19 times until the images from all 19 days have been used for testing. In doing so, all images can be used for training and testing.

In the feature extraction stage, we set parameters for each type of feature as follows.

- 1) *LBP Representation*: We adopt the uniform LBP, and select three scales to calculate the local differences of 8, 16, and 32 neighboring pixels. The total dimension of the LBP feature is 849.
- 2) *LEP Representation*: We adopt four filtering orientations and select $N = 4$ for n -ary coding. For multiscale representation, we adopt four scales with the standard deviation of [1, 2, 4, 8]. The final feature dimension is 1024.
- 3) *Texton-Based Representations*: That is, VZ-MR8, VZ-Joint, and PCs. The size of texton dictionary is 400. Therefore, the feature dimension is 400 for each type of representation. Specifically, the MR8 filter set used in our experiment is exactly the same with [12], and we use the image patch with a size of 7×7 and four scales of [1, 2, 4, 8] for PCs.
- 4) *BIFs Representation*: We use four scales of [1, 2, 4, 8] for the filters, and the basic local descriptor is 6 rectified filtering response. Therefore, the final feature dimension is $6^4 = 1296$.

In this paper, we employ the affinity propagation clustering algorithm [58] to partition those M feature representations into G groups. Specifically, $M = 6$ representations are automatically clustered into ($G = 3$) groups, where LBP and LEP representations are clustered into the first group, the second group contains VZ-MR8 and VZ-Joint representations, while PCs and BIFs are in the third group. In graph construction

Predicted Class \ Ground Truth Class	Ground Truth Class				Accuracy (%)
	Arc	Drapery corona	Radial corona	Hot-spot	
Arc	5568 43.9%	377 3.0%	186 1.5%	349 2.8%	85.9% 14.1%
Drapery corona	263 2.1%	1795 14.2%	681 5.4%	7 0.1%	65.4% 34.6%
Radial corona	119 0.9%	539 4.3%	1482 11.7%	143 1.1%	64.9% 35.1%
Hot-spot	192 1.5%	11 0.1%	205 1.6%	765 6.0%	65.2% 34.8%
Overall	90.7% 9.3%	65.9% 34.1%	58.0% 42.0%	60.5% 39.5%	75.8% 24.2%

Fig. 4. Confusion matrix achieved by the proposed GWL method for aurora image classification.

and multigraph classification steps, there are two parameters σ in (9) and μ in (12) to be tuned, respectively. We select those two parameters using a grid search strategy via twofold cross validation in the training dataset.

In the experiments, we first compare the proposed GWL method to the baseline method, i.e., graph-based classification model [50] (the graph is built based on a specific representation). It is worth noting that M graph-based classifiers corresponding to M representations are learned separately, while those in GWL are trained jointly by considering their underlying relationship. We also compare GWL to several state-of-the-art methods, including KNNs, SVM [59], and random forests (RFs) [60]. Specifically, we first feed each representation to each classifier for obtaining an individual classification result for basic comparison. Then, we also concatenate all features as a joint representation, and feed such joint representation to different classifiers. For a fair comparison, we further adopt a classifier ensemble method using the majority voting strategy to assemble basic classification results from different feature representations. In the experiments, 19 nearest neighbors are experimentally used for the KNN classifier with Chi-square distance, and 100 trees are used for the RF classifier. For SVM classifier, we employ the Chi-square kernel and select the parameter (i.e., C value) using a search strategy via twofold cross validation in the training dataset.

C. Classification Results

Fig. 4 shows the confusion matrix for each aurora category achieved by the proposed GWL method. From Fig. 4, we can figure out the class-wise classification performance among these four types auroral categories, including fourfold. First, GWL achieves very high classification accuracy (i.e., 90.7%) for arc aurora images. Such phenomenon is reasonable, since the arc aurora has definite arc features and can be easily distinguished by a classifier. Second, it is more challenging to classify the drapery corona aurora, radial corona

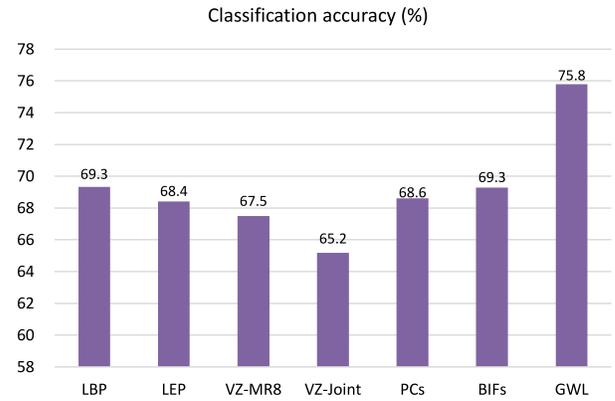


Fig. 5. Classification accuracies achieved by graph-based classification methods using different feature representations.

aurora, and hot-spot aurora images, compared to arc aurora. Third, many hot-spot aurora images have been wrongly classified as arc aurora. The reason should be that only arc aurora and hot-spot aurora have very strong intensity and which is a key feature to distinguish these two classes from the others. Fourth, it is easy to wrongly classify the drapery corona aurora to radial corona aurora, and vice versa. As shown in Fig. 3, these two classes share very similar local structures, the interclass difference between these two classes is much smaller than the difference between them and other classes.

In Fig. 5, we report the classification results achieved by the conventional graph-learning method and the proposed GWL method, where the first six bars denote results obtained by graph-based classifier using each of multiple feature representations. As can be seen from Fig. 5, our GWL method improves the classification accuracy by 6%, compared to the conventional graph-based classifier. It implies that GWL takes advantage of multiple features to improve the performance of aurora image classification. Also, it can be observed that graph-based methods using LBP, LEP, VZ-MR8, PCs, and BIFs features achieve much better classification results than that using VZ-Joint features. It is worth that aurora images often have different brightness and contrast due to various weather condition. By directly using image patches as a local descriptor, VZ-Joint features are not robust to the invariance of different imaging conditions. In contrast, the other five features are very robust to local brightness and contrast variations, and thus methods using such features achieve relatively better classification performance.

D. Comparison With State-of-the-Arts

We further conduct a group of experiments to compare the proposed GWL method with several state-of-the-art methods, with results shown in Fig. 6. From Fig. 6, we have the following observations.

- 1) The proposed GWL method achieves the best classification result. It suggests the proposed GWL method is more effective than methods using the concatenation of multiple features and methods using majority voting of multiple classification results.

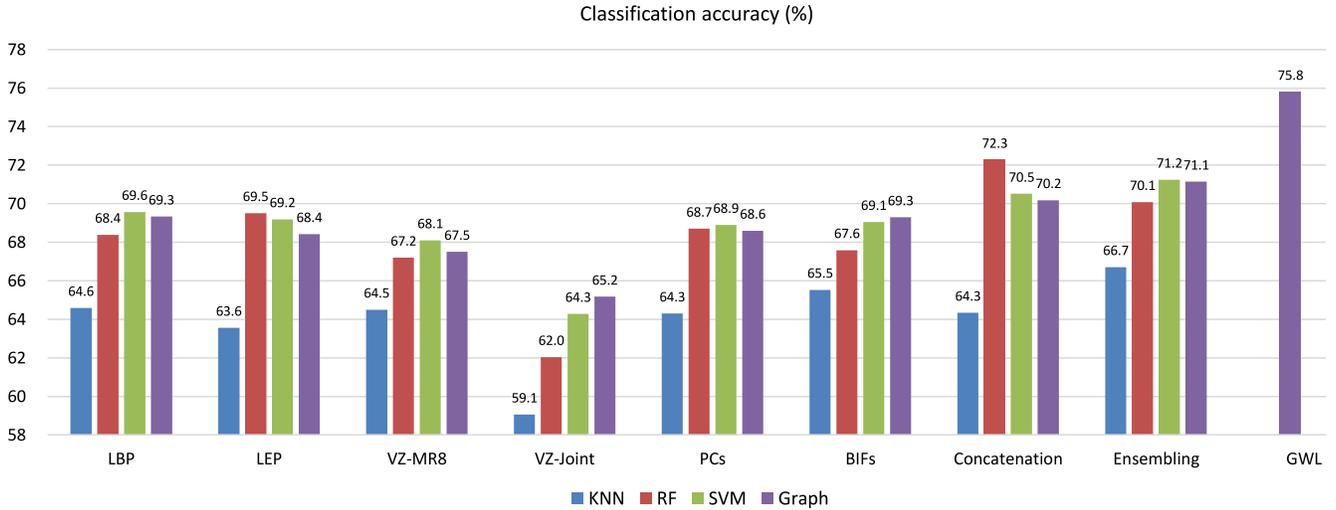


Fig. 6. Comparison with the state-of-the-art classifiers using multiple feature representations.

2) Generally, classifier ensemble-based methods (with majority voting strategy) achieve better classification results than that of using feature concatenation strategy. Because different features usually lie on different manifolds, it is not a good way to simply concatenate them into a long vector. Also, RF using feature concatenation strategy outperforms RF using ensemble strategy. The possible reason could be that RF treats each feature separately. That is, in each splitting node of a decision tree, one feature or a small subset of features in RF is used to make a decision. Therefore, feature concatenation provides a good solution for RF to make use of multiple representations.

E. Parameter Analysis

In this part, we analyze several important parameters in the proposed GWL method, including μ in (12) and σ in (9). We select the two parameters by leave-one-day-out cross validation within the training data. Fig. 7 shows the results of using aurora images from the first 18 days as training. From this figure, we could observe that the optimal parameter values lie in a relatively small range, that is, [10, 1000] for μ and [0.1, 10] for σ . Note that, the optimal parameters selected via cross validation could be slightly different for each fold, even though they all lie in the above ranges.

F. Diversity Analysis

For those ensemble-based methods and our proposed GWL method, the final classification results are made based on the outputs of multiple base classifiers. We further investigate the diversity of those base classifiers in an ensemble, using a kappa measure [61]. In this paper, each ensemble contains six individual classifiers concerning six representations, and hence there are 15 pairwise classifier combinations (corresponding to 15 kappa values). In Fig. 8, we report the diversity-error diagram achieved by different classifier ensemble-based methods, where the x -axis denotes the kappa value and the y -axis is

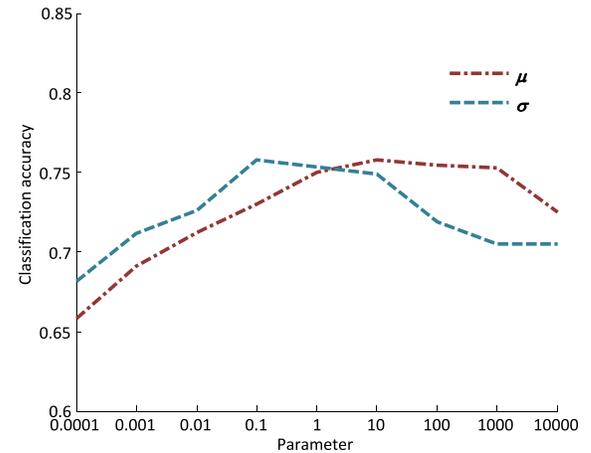


Fig. 7. Classification accuracies achieved by the proposed GWL method using different parameter values of μ and σ .

the averaged classification error of a pair of classifiers, respectively. A small kappa value indicates better diversity between a pair of classifiers, and a small averaged error indicates a better classification accuracy. Hence, the most desirable pairs of classifiers will be close to Fig. 8(bottom-left corner). To visually evaluate the relative positions of different clouds, we plot the centroids of different clouds achieved by five methods in Fig. 8, by computing the average position of all points in each cloud.

From Fig. 8, we could observe that the proposed GWL method outperforms KNN in terms of classification error, and achieves better diversity in terms of kappa measure than all the competing methods. That is, GWL builds a classifier ensemble based on reasonable accurate individual components but markedly good diversity. On the other hand, we can see that the mean classification errors achieved by those competing methods are usually small, compared to the proposed GWL method. However, their diversities are generally worse than that of GWL. The underlying reason could be that, in four

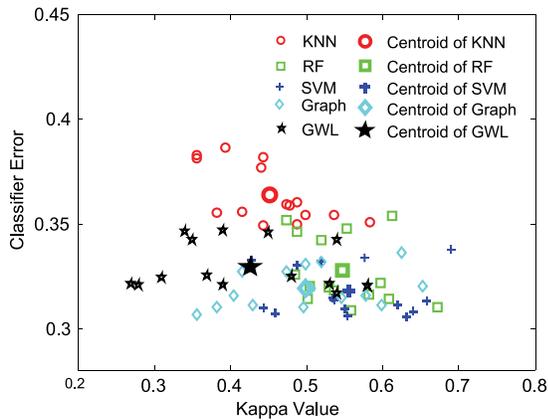


Fig. 8. Kappa-error diagram achieved by five classifier ensemble-based methods for aurora image classification.

competing methods (i.e., KNN, RF, SVM, and graph), each classifier is built in a specific feature representation space separately, where the complementary information conveyed by different representations is neglected. In GWL, the complementary information is expected to be reserved via a joint learning model, although the classification accuracy of the classifier in each representation may drop a little.

V. DISCUSSION

A. Feature Representation

We extract six types of feature representations to represent aurora images. The experimental results in Fig. 6 indicate that these features are effective for aurora representation, even with traditional classifiers (i.e., KNN, RF, and SVM). Moreover, as shown in Fig. 6, using multiple feature representations for aurora images do promote the classification performance. The underlying reason is that multiple types of representations may provide complementary information to distinguish different aurora categories. Therefore, it is important to develop efficient multiple representation fusion methods for aurora image classification.

B. Group-Wise Learning Strategy

For problems with multiple feature representations, there are generally two straightforward ways to combine multiple representations, that is, concatenation of multiple representations and classifier ensemble (e.g., via majority voting strategy), where each classifier is built on a specific feature representation. In general, feature concatenation methods neglect the complementary information conveyed in different representations, while classifier ensemble methods usually ignore the underlying correlation among different representations. In order to reasonably assemble all classification results obtained from multiple representations, we propose a GWL method for the fusion of multiple feature representations. Using the proposed method, we can automatically learn the weights for multiple representations from data, thus leading to a stable classification performance.

TABLE I
COMPARISON WITH THREE VARIANTS OF OUR METHOD

Method	GWL-individual	GWL-combine	GWL-random	GWL
ACC (%)	72.7	70.5	72.1	75.8

TABLE II
CLASSIFICATION PERFORMANCE FOR AD CLASSIFICATION

Method	AUC (%)	ACC (%)	SEN (%)	SPE (%)
KNN	78.3	75.0	71.8	77.7
RF	82.5	79.2	76.4	81.7
SVM	86.2	82.7	75.4	89.1
GWL (Ours)	92.9	89.8	88.9	90.4

In order to further explore the group-wise characteristics of our method, we conduct additional experiments using three variants of our method. In the first variant (denoted as GWL-individual), we treat six types of feature representations as six individual groups. In the second variant (denoted as GWL-combine), we combine all six feature representations to one single group. In the third variant (denoted as GWL-random), we randomly group six types of feature representations into three groups. For the fair comparison, our GWL method and its three variants share the same model defined in (12), as well as the same experimental settings. The experimental results on the classification of aurora images are shown in Table I.

As shown in Table I, our group-wise-based method (i.e., GWL) achieves superior performance compared to the other variants. The underlying reason could be organized below. First, without group partition, GWL-individual treat each feature representation individually which does not combine correlated representations. Each feature representation is not perfect and may have noise. If we treat each feature representation individually, the potential noise cannot be decreased. On the contrary, reasonable combination which combines correlated feature representations and will help decrease the effect of noise. Second, GWL-combine simply combines all feature representations together without considering the inherent correlations. Simple combination will lose the diversity of each representation. Therefore, GWL-combine cannot achieve satisfactory classification performance. Third, GWL-random makes the combination randomly which cannot decrease the feature noise effectively but lose the feature diversity.

C. Potential Applications

Besides using the aurora image dataset, we also evaluate our method on a public dataset, named Alzheimer's disease neuroimaging initiative (ADNI) [62], for Alzheimer's disease (AD) diagnosis. The public dataset has been broadly used for AD patient classification [63]–[68] and clinical score regression [69], [70]. In our experiment, we employ all baseline T1-weighted structural magnetic resonance (MR) images from ADNI-1, and these brain MR images are acquired from 199 AD subjects and 229 age-matched normal controls (NCs).

TABLE III
PERFORMANCE COMPARISON BETWEEN OUR METHOD AND SEVERAL
STATE-OF-THE-ART METHODS FOR AD DIAGNOSIS BASED ON
STRUCTURAL MRI DATA IN THE ADNI DATABASE

Method	Subject	AUC (%)	ACC (%)	SEN (%)	SPE (%)
Korolev <i>et al.</i> [63]	62 NC + 50 AD	87.0	80.0	-	-
Khvostikov <i>et al.</i> [64]	58 NC + 48 AD	-	85.4	88.3	90.0
Hinrichs <i>et al.</i> [65]	183 (NC + AD)	87.9	82.0	85.0	80.0
Koikkalainen <i>et al.</i> [66]	115 NC + 88 AD	-	86.0	81.0	91.0
Zhang <i>et al.</i> [67]	229 NC + 199 AD	88.2	83.7	80.9	86.7
Zhang <i>et al.</i> [68]	207 NC + 154 AD	94.0	88.3	79.6	94.7
GWL (Ours)	229 NC + 199 AD	92.9	89.8	88.9	90.4

Generally, the gray matter (GM) features based on regions-of-interest (ROIs) are usually used for brain image representation. In our experiment, we extracted 90 individual feature representations according to predefined 90 ROIs. To partition each brain sMR image to 90 ROIs, we use the automated anatomical labeling (AAL) map. The AAL map is originally defined on the Montreal Neurological Institute single subject brain MR image [71]. Specifically, we first linearly align all images to the same template. Then, we perform the nonlinear alignment and map the AAL GM tissue and ROIs to the images. Finally, for each subject, we extract the normalized volumes of GM tissue inside the 90 ROIs as feature representations.

In our experiment, we perform a tenfold cross validation to evaluate the classification performance. For comparison, we concatenate 90 ROI representations to train the KNN, RF, and SVM classifiers. Specifically, we employ $K = 5$ for KNN, 100 trees for RF, and $C = 1$ for linear SVM. We evaluate the binary classification performance using area under ROC (AUC), accuracy (ACC), sensitivity (SEN), and specificity (SPE), respectively. The experimental results achieved by different methods in AD versus NC classification can be found in Table II. As shown in Table II, our method achieves a very competitive classification accuracy of 89.8%, which is much higher than the other methods using KNN (75.0%), RF (79.2%), and SVM (82.7%). It suggests that our method can be potentially used for some other applications for merging multiple representations.

We further compare our GWL method with six state-of-the-art methods on ADNI, including two deep-learning-based methods [63], [64], a voxel-level-based method [65], a region-level-based method [66], and two anatomical-landmark-based methods [67], [68]. The experimental results are reported in Table III. Note that the results in Table III are not fully comparable, because different numbers of studies are used in different studies. By roughly comparing our method with these methods, we can see that GWL yields competitive results, further validating the effectiveness of our method in multirepresentation-based classification.

D. Limitations and Future Work

Although our method achieves superior classification results compared to the competing methods, there are still some limitations in this paper.

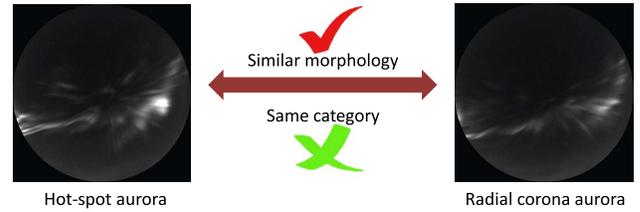


Fig. 9. Illustration of confounded aurora.

- 1) The classification performance for four aurora categories needs to be further improved. Currently, the automatic aurora classification still remains an open problem because it is a very challenging task. One type of challenge is that there exist various transition changes from a kind of aurora to another. In such a case, it is difficult to label such aurora into a specific category. For example, as shown in Fig. 9, the image has both features of hot-spot and radial corona aurora. A possible solution is to assign multiple labels for such aurora and generalize the aurora classification to a multilabel classification problem.
- 2) We simply capture the pairwise relationships among samples via conventional graph structure, while the true relationships could be more complex than pairwise. It is very interesting to incorporate such complex data structure (e.g., via hypergraph) into the learning model, in order to further promote the classification performance of aurora images.
- 3) Since labeling the aurora images is a time-consuming process, we only have 12 682 labeled images in this paper. As a future work, we will annotate more images (e.g., even millions of images). In this way, the learning-based feature extraction method, for example, deep CNNs [72], can also be used for aurora image classification.

VI. CONCLUSION

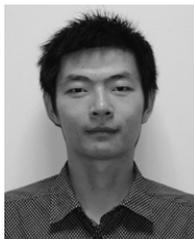
In this paper, we propose a GWL method for aurora image classification. To be specific, we first extract multiple feature representations for each aurora image, and then adopt a graph structure to model the pairwise relationship among samples. To capture the correlation among different representations, we cluster multiple representations into several groups. Then, we propose a GWL model to automatically estimate the class labels for aurora images and optimal weights for different representations in a data-driven manner. Finally, we develop a label fusion approach to make a final classification decision for a new testing sample. The experimental results suggest that our method achieves better performance than several state-of-the-art methods in aurora image classification.

REFERENCES

- [1] A. S. Jursa *et al.*, *Handbook of Geophysics and the Space Environment*. Cambridge, MA, USA: NTIS, 1985.
- [2] S.-I. Akasofu, "Energy coupling between the solar wind and the magnetosphere," *Space Sci. Rev.*, vol. 28, no. 2, pp. 121–190, 1981.

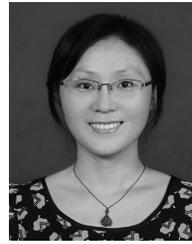
- [3] X. Yang, X. Gao, D. Tao, and X. Li, "Improving level set method for fast auroral oval segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2854–2865, Jul. 2014.
- [4] X. Yang *et al.*, "Shape-constrained sparse and low-rank decomposition for auroral substorm detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 32–46, Jan. 2016.
- [5] M. Syrjäsuo, E. Donovan, L. Cogger, and A. Calgary, "Content-based retrieval of auroral images-thousands of irregular shapes," in *Proc. IASTED Int. Conf. Visual. Imag. Image Process. (VIIP)*, 2004, pp. 224–228.
- [6] M. Syrjäsuo, E. Donovan, X. Qin, and Y. Yang, "Automatic classification of auroral images in substorm studies," in *Proc. 8th Int. Conf. Substorm Calgary Alberta Canada Univ. Calgary*, 2007, pp. 309–313.
- [7] M. Syrjäsuo and E. Donovan, "Using relevance feedback in retrieving auroral images," in *Proc. 4th IASTED Int. Conf. Comput. Intell.*, 2005, pp. 420–425.
- [8] Q. Wang *et al.*, "Spatial texture based automatic classification of dayside aurora in all-sky images," *J. Atmos. Solar Terrestrial Phys.*, vol. 72, nos. 5–6, pp. 498–508, 2010.
- [9] X. Yang, X. Gao, and Q. Tian, "Polar embedding for aurora image retrieval," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3332–3344, Nov. 2015.
- [10] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [11] J. Zhang, J. Liang, and H. Zhao, "Local energy pattern for texture classification using self-adaptive quantization thresholds," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 31–42, Jan. 2013.
- [12] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *Int. J. Comput. Vis.*, vol. 62, nos. 1–2, pp. 61–81, 2005.
- [13] J. Zhang, H. Zhao, and J. Liang, "Continuous rotation invariant local descriptors for texton dictionary-based texture classification," *Comput. Vis. Image Understand.*, vol. 117, no. 1, pp. 56–75, 2013.
- [14] M. Crosier and L. D. Griffin, "Using basic image features for texture classification," *Int. J. Comput. Vis.*, vol. 88, no. 3, pp. 447–460, 2010.
- [15] A. Y. Ng, M. I. Jordan, Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 2, 2002, pp. 849–856.
- [16] M. Syrjäsuo, E. F. Donovan, and M. Peura, "Using attribute trees to analyse auroral appearance over Canada," in *Proc. 6th IEEE Workshop Appl. Comput. Vis.*, Orlando, FL, USA, 2002, pp. 289–295.
- [17] Y. Wang, X. Gao, R. Fu, and Y. Jian, "Dayside corona aurora classification based on X-gray level aura matrices," in *Proc. ACM Int. Conf. Image Video Retrieval*, 2010, pp. 282–287.
- [18] X. Yang, X. Gao, B. Song, N. Wang, and D. Yang, "ASI aurora search: An attempt of intelligent image processing for circular fisheye lens," *Opt. Exp.*, vol. 26, no. 7, pp. 7985–8000, 2018.
- [19] X. Yang, X. Gao, B. Song, and D. Yang, "Aurora image search with contextual CNN feature," *Neurocomputing*, vol. 281, pp. 67–77, Mar. 2018.
- [20] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, "Discriminative local binary patterns for human detection in personal album," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
- [21] T. Wang *et al.*, "Jumping and refined local pattern for texture classification," *IEEE Access*, vol. 6, pp. 64416–64426, 2018.
- [22] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2032–2047, Nov. 2009.
- [23] Y. Dong, H. Wu, X. Li, C. Zhou, and Q. Wu, "Multiscale symmetric dense micro-block difference for texture classification," *IEEE Trans. Circuits Syst. Video Technol.*, to be published. doi: [10.1109/TCSVT.2018.2883825](https://doi.org/10.1109/TCSVT.2018.2883825).
- [24] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Process.*, vol. 4, no. 11, pp. 1549–1560, Nov. 1995.
- [25] Y. Dong, D. Tao, X. Li, J. Ma, and J. Pu, "Texture classification and retrieval using shearlets and linear regression," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 358–369, Mar. 2015.
- [26] Y. Dong, D. Tao, and X. Li, "Nonnegative multiresolution representation-based texture image classification," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 1, p. 4, 2015.
- [27] A. Depersinge, Z. Püspöki, J. P. Ward, and M. Unser, "Steerable wavelet machines (SWM): Learning moving frames for texture classification," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1626–1636, Apr. 2017.
- [28] Y. Dong, J. Feng, L. Liang, L. Zheng, and Q. Wu, "Multiscale sampling based texture image classification," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 614–618, May 2017.
- [29] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
- [30] W. N. Anderson, Jr., and T. D. Morley, "Eigenvalues of the Laplacian of a graph?" *Linear Multilinear Algebra*, vol. 18, no. 2, pp. 141–145, 1985.
- [31] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, no. 16, 2004, pp. 321–328.
- [32] B. Du, X. Tang, Z. Wang, L. Zhang, and D. Tao, "Robust graph-based semisupervised learning for noisy labeled data via maximum correntropy criterion," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1440–1453, Apr. 2019.
- [33] F. Wang and C. Zhang, "Label propagation through linear neighborhoods," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 55–67, Jan. 2008.
- [34] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1601–1608.
- [35] Y. Gao, M. Wang, D. Tao, R. Ji, and Q. Dai, "3-D object retrieval and recognition with hypergraph analysis," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4290–4303, Sep. 2012.
- [36] M. Liu and D. Zhang, "Sparsity score: A novel graph-preserving feature selection method," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 28, no. 4, 2014, Art. no. 1450009.
- [37] M. Liu and D. Zhang, "Feature selection with effective distance," *Neurocomputing*, vol. 215, pp. 100–109, Nov. 2016.
- [38] J. Wu, S. Pan, X. Zhu, C. Zhang, and X. Wu, "Positive and unlabeled multi-graph learning," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 818–829, Apr. 2017.
- [39] O. Chapelle, J. Weston, and B. Schölkopf, "Cluster kernels for semi-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 585–592.
- [40] X. Zhu, X. Li, and S. Zhang, "Block-row sparse multiview multilabel learning for image classification," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 450–461, Feb. 2016.
- [41] M. Liu, L. Miao, and D. Zhang, "Two-stage cost-sensitive learning for software defect prediction," *IEEE Trans. Rel.*, vol. 63, no. 2, pp. 676–686, Jun. 2014.
- [42] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Learning Theory and Kernel Machines*. Heidelberg, Germany: Springer, 2003, pp. 144–158.
- [43] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, to be published. doi: [10.1109/TCYB.2018.2810806](https://doi.org/10.1109/TCYB.2018.2810806).
- [44] M. Liu, D. Zhang, and S. Chen, "Attribute relation learning for zero-shot classification," *Neurocomputing*, vol. 139, pp. 34–46, Sep. 2014.
- [45] X. Zhang and W. S. Lee, "Hyperparameter learning for graph based semi-supervised learning algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1585–1592.
- [46] M. Liu and D. Zhang, "Pairwise constraint-guided sparse learning for feature selection," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 298–310, Jan. 2016.
- [47] X. Li, G. Cui, and Y. Dong, "Graph regularized non-negative low-rank matrix factorization for image clustering," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3840–3853, Nov. 2017.
- [48] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [49] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
- [50] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *Proc. Int. Conf. Mach. Learn.*, 2001, pp. 19–26.
- [51] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. Int. Conf. Mach. Learn.*, vol. 3, 2003, pp. 912–919.
- [52] T. Joachims *et al.*, "Transductive learning via spectral graph partitioning," in *Proc. Int. Conf. Mach. Learn.*, vol. 3, 2003, pp. 290–297.
- [53] C. Deng, R. Ji, D. Tao, X. Gao, and X. Li, "Weakly supervised multi-graph learning for robust image reranking," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 785–795, Apr. 2014.

- [54] C. Berge and E. Miniéka, *Graphs and Hypergraphs*, vol. 7. Amsterdam, The Netherlands: North-Holland, 1973.
- [55] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *J. Amer. Stat. Assoc.*, vol. 62, no. 318, pp. 399–402, 1967.
- [56] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vis.*, vol. 7, no. 1, pp. 11–32, 1991.
- [57] Z.-J. Hu *et al.*, "Synoptic distribution of dayside aurora: Multiple-wavelength all-sky observation at Yellow River Station in Ny-Ålesund, Svalbard," *J. Atmospheric Solar Terrestrial Phys.*, vol. 71, nos. 8–9, pp. 794–804, 2009.
- [58] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.
- [59] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [60] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [61] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.
- [62] C. R. Jack, Jr., *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Resonance Imag.*, vol. 27, no. 4, pp. 685–691, 2008.
- [63] S. Korolev, A. Safullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Melbourne, VIC, Australia, 2017, pp. 835–838.
- [64] A. Khvostikov, K. Aderghal, J. Benois-Pineau, A. Krylov, and G. Catheline, "3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies," *arXiv preprint arXiv:1801.05968*, 2018.
- [65] C. Hinrichs *et al.*, "Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset," *NeuroImage*, vol. 48, no. 1, pp. 138–149, 2009.
- [66] J. Koikkalainen *et al.*, "Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease," *NeuroImage*, vol. 56, no. 3, pp. 1134–1144, 2011.
- [67] J. Zhang, Y. Gao, Y. Gao, B. C. Munsell, and D. Shen, "Detecting anatomical landmarks for fast Alzheimer's disease diagnosis," *IEEE Trans. Med. Imag.*, vol. 35, no. 12, pp. 2524–2533, Dec. 2016.
- [68] J. Zhang, M. Liu, L. An, Y. Gao, and D. Shen, "Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1607–1616, Nov. 2017.
- [69] B. Jie, M. Liu, J. Liu, D. Zhang, and D. Shen, "Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 238–249, Jan. 2017.
- [70] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Deep multi-task multi-channel learning for joint classification and regression of brain status," in *Proc. Int. Conf. Med. Image Comput. Assisted Intervention*, 2017, pp. 3–11.
- [71] N. Tzourio-Mazoyer *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.
- [72] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.



Jun Zhang was born in Shaanxi Province, China. He received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2009 and 2014, respectively.

He is a Senior Research Engineer at AI Platform Department, Tencent, China. His current research interests include image processing, machine learning, pattern recognition, and medical image analysis.



Mingxia Liu received the B.S. and M.S. degrees from Shandong Normal University, Shandong, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2015.

Her current research interests include neuroimaging analysis, machine learning, pattern recognition, and data mining.



Ke Lu was born in Ningxia, China, in 1971. He received the master's and Ph.D. degrees from the Department of Mathematics and the Department of Computer Science, Northwest University, Xi'an, China, in 1998 and 2003, respectively.

He was a Post-Doctoral Fellow with the Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2003 to 2005. He is currently a Professor with the University of Chinese Academy of Sciences, Beijing. His current research interests include computer vision, 3-D image reconstruction, and computer graphics.



Yue Gao (SM'14) received the B.E. degree from the Department of Electronic Information Engineering, Harbin Institute of Technology, Harbin, China, the master's degree from the School of Software, Tsinghua University, Beijing, China, and the Ph.D. degree from the Department of Automation, Tsinghua University.

He is currently an Associate Professor with the School of Software, Tsinghua University. He has been with the School of Computing, National University of Singapore, Singapore, and the School

of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, from 2012 to 2016. He has published over 90 papers in premier journals and conferences, such as the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MEDICAL IMAGING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, ACM Transactions on Multimedia Computing, Communications, and Applications, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, TIE, Human Brain Mapping, Medical Image Computing and Computer Assisted Interventions, IEEE Conference on Computer Vision and Pattern Recognition, International Joint Conferences on Artificial Intelligence, Association for the Advancement of Artificial Intelligence, European Conference on Computer Vision, and ACM Multimedia. His current research interests include computer vision, medical image analysis, machine learning and its applications, 3-D vision, visual search, computer-aided diagnosis, and network scheduling.

Dr. Gao was a recipient of the 1000 Youth Talent Plan Grant of China. He also serves as an Associate Editor for the *Journal of Visual Communication and Image Representation* and *Neurocomputing*, the Program/Special Session/Session Chair for International Conference on Internet Multimedia Computing and Service 2016, ACM International Conference on Multimedia Retrieval 2014, Pacific-Rim Conference on Multimedia 2015, and the IEEE Visual Communications and Image Processing Conference 2014, and a TPC Member for AAAI 2017–2018, IJCAI 2016–2017, Association for Computational Linguistics 2015, and MICCAI 2015–2016. He is a member of ACM.