IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 15, 2019, accepted August 2, 2019, date of publication August 14, 2019, date of current version September 10, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2935420

Scene Recognition via Semi-Supervised Multi-Feature Regression

CAIXIA ZHENG[1,2], JIANYU CHEN[1], JUN KONG[1,3], YUGEN YI[4], YINGHUA LU[3], JIANZHONG WANG[1], AND CHONG LIU[4]

[1]College of Information Science and Technology, Northeast Normal University, Changchun 130024, China
[2]Key Laboratory of Applied Statistics of MOE, Northeast Normal University, Changchun 130024, China
[3]College of Humanities and Sciences, Institute for Intelligent Elderly Care, Northeast Normal University, Changchun 130024, China
[4]School of Software, Jiangxi Normal University, Nanchang 330022, China

Corresponding authors: Yinghua Lu (luyh@nenu.edu.cn) and Jianzhong Wang (wangjz019@nenu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702092, Grant 61672150, Grant 61602221, and Grant 61806126, in part by the Fund of the Jilin Provincial Science and Technology Department under Grant 20180520215JH, Grant 20190201305JC, Grant 20170204018GX, Grant 20180201089GX, and Grant 20180520027JH, in part by the Fund of Education Department of Jilin Province under Grant JJKH20190294KJ and Grant JJKH20190291K, in part by the Natural Science Foundation of Jiangxi Province under Grant 20171BAB212009, in part by the Science and Technology Research Project of Jiangxi Provincial Department of Education under Grant GJJ160333, and in part by the Fundamental Research Funds for the Central Universities under Grant 2412019FZ049.

ABSTRACT With the development of visual sensor equipment (e.g., personal smart phones, vehicle cameras, surveillance videos and camcorders), scene recognition technology has attracted much attention due to its latent applications in visual surveillance, intelligent traffic and aerial remote sensing. Although some progress has been made in the field of scene recognition in recent years, the complexity of scene images and the inadequate numbers of labeled data pose challenges in this area. Hence, to effectively fuse the multiple features of each image and employ the information of both labeled and unlabeled images for scene recognition, we proposed a semi-supervised multi-feature regression (SSMFR) model in this paper. The SSMFR model possesses three advantages. First, the model propagates the labels of labeled data to unlabeled data by utilizing graph-based semi-supervised learning techniques so that both the information regarding unlabeled data and labeled data can be exploited to gain better performance. Second, SSMFR employs multiple graphs to characterize the structures of multiple feature spaces and adaptively assigns the weight to different graphs. Therefore, SSMFR can efficiently preserve the manifold structure of samples in each feature space and adequately exploit the complementary information of multiple features. Moreover, SSMFR adopts a $l_{2,1}$-norm constraint to learn a sparse and robust classifier for scene recognition. To solve the SSMFR model, we proposed a simple and efficient iterative update optimization scheme. Finally, we also proved the convergence of SSMFR by theoretical analysis and experiments. Experiments were conducted on several benchmark scene datasets, and the experimental results demonstrated that the proposed SSMFR model can obtain better performance for scene recognition than some other state-of-the-art algorithms.

INDEX TERMS Scene recognition, semi-supervised recognition model, multi-features regression.

I. INTRODUCTION

In the past decades, the rapidly developing multimedia brought the explosion of image data on the Internet, making it difficult for people to determine what they need or are interested in. Hence, methods for using computers to automatically manage images, especially to classify and query images in the same way as humans understanding high-level image semantics, has drawn increasing attention. As one of the classical issues in the field of image understanding, scene recognition is the process of categorizing images into semantic types of scenes, which can be widely used in various applications, e.g., action recognition, context-aware object detection, intelligent vehicle/robot navigation and aerial remote sensing applications [1]–[5]. Although the accuracy and robustness of scene recognition models have been greatly improved in the

The associate editor coordinating the review of this article and approving it for publication was Peng Liu.

past several years, scene recognition is still challenging not only because the ambiguity, variability and a wide range of illumination in scene images, but also because the number of labeled scene images is much smaller than the unlabeled scene images in the real world [6]–[8]. There is a fascinating property of the human visual system: we recognize images by using very few labeled samples, and we accomplish the image recognition task by integrating multiple features, such as color, shape and the objects that appear on images. Is it possible that computers acquire such a capability through machine learning techniques? This is the main motive of our study in this paper. To reduce the human effort in labeling data and construct an effective classifier that can utilize multiple features of data, we developed a novel semi-supervised multi-feature regression (SSMFR) model for scene recognition.

The visual content of a given scene image can be described by different features, such as color, texture and shape. Intuitively, it is beneficial for scene recognition tasks to utilize different features simultaneously [9]. Therefore, how to effectively use multiple pieces of evidence from heterogeneous or independent features becomes an interesting issue. Previous studies have indicated that comparing with utilizing only one kind of feature or directly integrating multiple kinds of features into one feature vector, better performance could be obtained for image recognition if the information of different features is fused properly [10]. Recently, there have been many efforts to develop efficient scene recognition methods by fusing multiple features of data. Sun *et al.* [11] proposed a two-stage probabilistic classification framework to utilize multiple features for scene classification. Song *et al.* [12] proposed a joint multi-feature spatial context (JMSC) model for recognizing scene images in the semantic manifold. In JMSC, two types of contextual relations, including multi-feature relations obtained from different low-level visual features and local spatial relations between neighboring patches, are exploited to enhance consistent scene category co-occurrence patterns and eliminate noise patterns. With the rise of large data sets and the convolutional neural network (CNN), Song *et al.* [13] developed a new multi-scale multi-feature context model (MMC) by extending JMSC. MMC builds the semantic manifold on top of multiscale CNNs and combines spatial relations, various scales and features to construct a rich context model for scene recognition. Although these approaches achieved good performances in scene recognition by utilizing multiple visual features, they all need a large number of labeled training samples. However, collecting a mass of high-quality labeled training images is very difficult and time-consuming in practical applications [14], [15]; for instance, 111 researchers from different institutes spent more than 220 hours labeling only 63 hours of TRECVID 2003 development corpus [16], [17]. Conversely, unlabeled images are often easy to obtain [18]. Therefore, integrating the labeled and unlabeled images together is crucial for improving the accuracy of a scene classifier. This motivates us to develop a SSMFR method. Compared with other algorithms, the contributions of our SSMRF are threefold. First, by utilizing graph-based semi-supervised learning techniques, SSMFR can combine multiple features from both labeled and unlabeled images to infer a more accurate classifier. Second, by employing multiple graphs to characterize the structures of multiple feature spaces and adaptively assigning the weight to different graphs, SSMFR not only preserves the manifold of each feature but also exploits the complementary information contained in different features. Moreover, SSMFR adopts a $l_{2,1}$-norm constraint to learn a sparse and robust scene classifier.

The remainder of this paper is organized as follows. Section II briefly reviews related work. In Section III, we introduce the details of our proposed SSMFR and its iterative solution algorithm. The convergence analysis of SSMFR is also provided in this section. Section IV gives the experimental results and analysis. Finally, a conclusion is drawn in Section V.

II. RELATED WORK

Our work is devoted to achieving more satisfactory scene recognition performance. Since multi-feature learning and semi-supervised learning are both effective ways to obtain good performance of image or video understanding tasks, we combine them in a unified framework for scene recognition. In this section, we briefly review three related research issues including scene recognition, multi-feature learning, and semi-supervised learning.

A. SCENE RECOGNITION

Scene recognition is an important research topic in the area of computer vision and pattern recognition [19], which has been studied from various viewpoints by researchers, and numerous methods have been proposed in the recent years for acquiring high scene recognition accuracy. Bosch *et al.* [20] roughly divided these approaches into two main categories: low-level and semantic modeling-based methods. Low-level-based methods focus on developing the discriminative visual features (e.g., SIFT [21], GIST [22] and CENTRIST [23]) to represent the color, texture or shape information of scene images, while semantic modeling-based approaches involve learning semantic intermediate representations (e.g., local semantic concepts [24], [25] and semantic objects [26]–[28]) to model the content of scene images. Recently, with the proliferation of deep learning, there has been a surge of research interests in exploring deep architectures to solve the scene recognition problem [29], [30]. Khan *et al.* [7] presented a deep learning-based spectral feature for image recognition. This spectral descriptor is a spectral domain representation of convolution features from deep network architectures, which can enhance the discriminative ability of deep networks and achieve satisfactory performance of scene recognition. Herranz *et al.* [31] utilized scale-specific convolutional neural networks (CNNs) to construct a multiscale architecture, which is an effective way to combine ImageNet-CNNs and Places-CNNs in different scale ranges and address dataset bias to improve scene recognition performance.

Guo *et al.* [32] proposed a locally supervised deep hybrid model (LS-DHM) to encode and enhance the convolutional features. LS-DHM collaboratively explores the Fisher convolutional vector and the fully connected layer of the convolutional neural network (CNN) to gain significant scene recognition improvements. Zhu *et al.* [33] proposed a discriminative multimodal fusion framework for RGBD indoor scene recognition; this framework considers the intra- and inter-modality correlation of all data and regularizes the learned features to be simultaneously compact and discriminative. Although the abovementioned methods achieve good scene recognition performance, they require a tremendous amount of labeled data to learn the representation of images or train the classification model. However, collecting massive data, especially labeled image or video data, is a very time consuming process [17]. Moreover, these methods focus on designing more accurate features for scene image representation but they do not pay attention to how to fuse independent or heterogeneous features effectively and explicitly, and they cannot use the information of the unlabeled data [10].

To use the unlabeled data to improve the recognition performance, some semi-supervised learning methods are proposed for scene recognition tasks [34]–[36]. Liu *et al.* [34] proposed p-Laplacian regularization to preserve the local geometry of data for identifying the scene images. Han *et al.* [35] proposed a semi-supervised generative framework (SSGF) to achieve the task of scene recognition. These methods can gain good performance by using the information of both labeled and unlabeled data, but they just use one feature or directly connecting several features into one feature vector, which limit the performance of algorithms to some extent.

B. MULTI-FEATURE LEARNING

Effectively fusing multiple features for image understanding tasks has attracted increasing attention from the researchers in the machine learning community. Concatenating all types of features into one vector is a common method of manipulating the multiple features. However, feature concatenation leads to the dimensionality of the feature vector to be very high and is less effective in the task of image understanding [37]–[39]. To address this issue, the concept of multi-feature leaning has been developed by researchers, and many multi-feature learning algorithms have been proposed to exploit the structural information of each feature. Canonical correlation analysis (CCA) [40] and the two-view-based support vector machine (SVM-2k) [41] are two well-known multi-feature learning algorithms. CCA is a statistical algorithm that maximizes the relations between two modalities of data, while SVM-2k learns one SVM classifier based on two views of data. The various improved versions of CCA and SVM-2k have been commonly applied in image recognition tasks [10]. Recently, Yang *et al.* [42] proposed an efficient multi-task feature selection model (FSSI) for multimedia content analysis. FSSI considers the feature correlation by evaluating the importance of different features jointly. Xia *et al.* [43] developed a multiview spectral embedding (MSE) approach, which learns a physically meaningful and low-dimensional embedding of all views simultaneously for encoding different features. Ren *et al.* [44] proposed a maximum margin multimodal deep neural network (3mDNN) to discriminatively fuse multiple features. Wang *et al.* [45] developed an adaptive multiview feature selection (AMFS) method, which automatically learns Laplacian graphs for multiview features and adaptively assigns a weight to each feature by adopting a local linear regression model. These approaches are superior to feature concatenation. However, they require vast amounts of labeled data for training, which is usually time-consuming and seldom available in practical applications.

C. SEMI-SUPERVISED LEARNING

Previous studies have shown that it is beneficial for image semantics understanding when labeled and unlabeled data are simultaneously employed [10]. Therefore, many semi-supervised learning techniques have been proposed to utilize the information of both unlabeled and labeled data in the training phase to learn better classifiers.

According to [46], the existing semi-supervised learning methods can be grouped into five categories, including EM with generative mixture models, self-training, cotraining, transductive support vector machines, and graph-based approaches. Among these approaches, graph-based semi-supervised learning have been a topic of recent interest [47], [48]. Graph-based semi-supervised approaches construct a weighted graph in which the nodes represent the samples in the dataset, the edges connect the similar samples, and the weights of edges reflect the similarity between samples [48]. Different strategies have been developed for graph-based semi-supervised learning [47], [49]–[54]. For example, by assuming that the neighboring samples should have similar class labels, Zhou *et al.* [50] proposed a local and global consistency (LGC) algorithm, which estimates the labels of unlabeled samples by iteratively spreading the label information of every samples of graph to its neighbors. Cai *et al.* [55] developed a semi-supervised discriminant analysis (SDA) algorithm, which finds an optimal low-dimensional subspace by constructing a graph to preserve the geometric structure of both labeled and unlabeled samples. Though the abovementioned approaches have achieved good performances for image recognition (i.e., face recognition) tasks, they all adopt one feature or simply concatenate multiple features as one feature vector. Therefore, as we have discussed earlier, when applying these algorithms in the field of scene recognition, their performances may be limited because using one feature cannot represent the content of scene images well and concatenating multiple feature vectors into one feature neglects the potential complementary information among multiple features.

To overcome the aforementioned limitations, we presented a novel learning model named the semi-supervised multi-feature regression (SSMFR), which combines the advantages

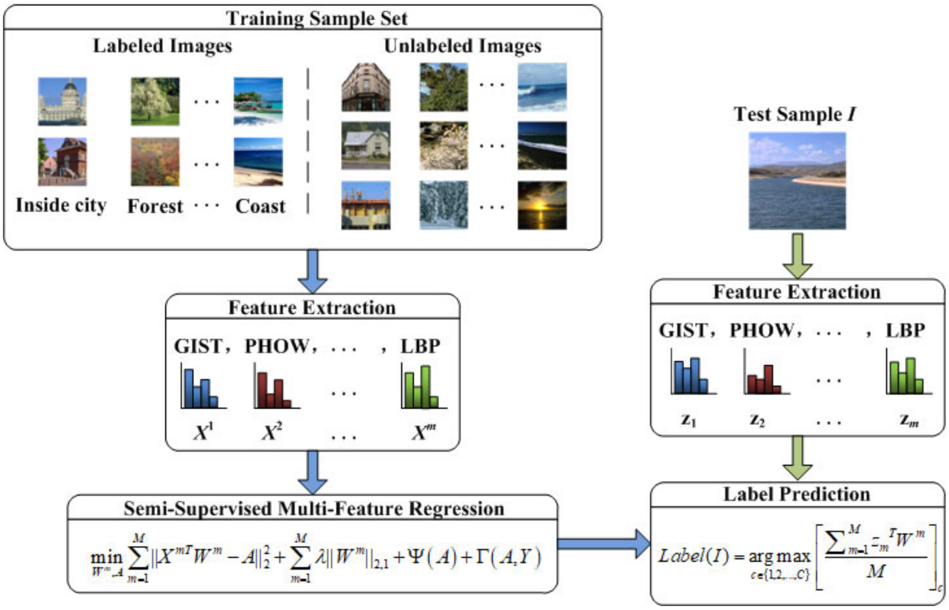


FIGURE 1. The overall procedure of the proposed SSMFR.

of multi-feature learning and semi-supervised learning for scene recognition.

III. THE PROPOSED SSMFR ALGORITHM

A. FORMULATION OF SSMFR

In SSMFR, the potentially related and complementary structure information among multiple features are exploited by jointly learning multiple subclassifiers (also called mapping matrices) for all features, and the global consistency of the prediction labels in the process of label propagation is guaranteed through the weighted graph regularization operator. The overall SSMFR procedure is shown in Fig. 1.

Let us denote $X = \{x_1, x_2, \ldots, x_l, x_{l+1}, \ldots, x_N\} \in \Re^{D\times N}$ as the training set of image samples in D dimensional space, where $x_i \in \Re^{D\times 1}(1 \leq x_i \leq N)$ is the feature vector of i-th image and N is the total number of training samples. Without loss of generality, we suppose that the first l samples, i.e., $X_l = \{x_1, x_2, \ldots, x_l\} \in \Re^{D\times l}$ in X are labeled samples, and rest of the samples, i.e., $X_u = \{x_{l+1}, x_{l+2}, \ldots, x_N\} \in \Re^{D\times(N-l)}$ in X are unlabeled samples. Let $Y = [y_1, y_2, \ldots, y_N] \in \{0, 1\}^{N\times C}$ denote the labeled matrix of X, where C is the total number of classes. In labeled matrix Y, if x_i is labeled as a sample in the j-th category, then we set the element $y_{ij} = 1$ in $y_i \in \Re^C$, and set other elements as $y_{ik} = 0(k \neq j)$. If x_i is an unlabeled sample, we set all elements in $y_i \in \Re^C$ as 0. Suppose each image sample is described by M different features, $X^m = \left[x_1^m, x_2^m, \ldots, x_N^m\right] \in \Re^{d_m\times N}$ represents the set of the m-th features of all samples, where x_i^m is the m-th feature of the sample x_i and d_m is the dimension of this feature.

In the field of semi-supervised learning, there is an important manifold assumption which demonstrates that data in a small local region should have similar properties, and the labels of them should also be similar [56]. To better preserve the label consistency of data in the manifold structure and efficiently integrate multi-feature learning and semi-supervised learning into a unified framework, we propose the SSMFR model as

$$\min_{W^m,A} \sum_{m=1}^{M} ||X^{mT}W^m - A||_2^2 + \sum_{m=1}^{M} \lambda ||W^m||_{2,1} + \Psi(A) + \Gamma(A, Y), \quad (1)$$

where $A = [a_1, a_2, \ldots, a_N] \in \Re_+^{N\times C}$ is the prediction label matrix of the training dataset, the element $a_i \in \Re^C$ of A represents the prediction label vector of sample x_i; $W^m \in \Re^{d_m\times C}$ denotes the mapping between the m-th feature set X^m and the prediction label matrix A; $\sum_{m=1}^{M} ||X^{mT}W^m - A||_2^2$ is the global error of label predicting; $||W^m||_{2,1}$ is the constraint term to enforce the row sparsity on the learned W^m and make W^m more robust; $\Psi(\cdot)$ is the graph regularization term that exploits the potential complementary information from various features of the image and ensures that the labels predicted by the different features are consistent; $\Gamma(\cdot)$ is the penalty term to ensure that the estimated labels of the labeled samples are same with their real labels; and λ is a trade-off parameter. In the following, we will introduce the specific form and the derivation process of each term in Eq. (1).

To improve the performance of the learned classification model by making use of both labeled and unlabeled data in the training dataset, we first utilize the label propagation technology based on adaptive weighted multiple graphs to propagate label information from labeled data to unlabeled data. Specifically, for each feature set X^m, we construct a weighted undirected graph G^m in which the weight S_{ij}^m between the

i-th node (sample x_i^m) and the j-th node (sample x_j^m) is defined as

$$S_{ij}^m = \begin{cases} \exp\left(-\|x_i^m - x_j^m\|_2^2/2\sigma^2\right), \\ \qquad \text{if } x_j^m \in \Delta_k\left(x_i^m\right) \text{ or } x_i^m \in \Delta_k\left(x_j^m\right) \\ 0, \qquad \text{otherwise,} \end{cases} \tag{2}$$

where $\Delta_k\left(x_i^m\right) = \left[x_{i,1}^m, x_{i,2}^m, \ldots, x_{i,k}^m\right]$ denotes the set of k-nearest neighbors of the i-th sample x_i^m in the m-th feature set X^m; and σ is a parameter.

In the process of label propagation, the manifold structure of each feature set should be maintained. That is, the adjacent samples should be assigned similar labels [48] in the prediction label matrix A. This goal can be achieved by minimizing the following objective function:

$$\begin{aligned} \sum_{i,j=1}^{N} \left\| \frac{a_i}{d_{ii}^m} - \frac{a_j}{d_{jj}^m} \right\|_2^2 S_{ij}^m &= tr\left(A^T\left(I - D^{m-\frac{1}{2}} S^m D^{m-\frac{1}{2}}\right)A\right) \\ &= tr\left(A^T L^m A\right), \\ s.t.\ A \geq 0 \end{aligned} \tag{3}$$

where D^m is the diagonal matrix whose diagonal elements are defined as $d_{ii}^m = \sum_{j=1}^{N} S_{ij}^m$; The matrix $L^m = I - D^{m-\frac{1}{2}} S^m D^{m-\frac{1}{2}}$ is the normalized Laplace matrix of the undirected weighted graph G^m constructed by the m-th feature set.

In addition to preserving the manifold structure of each feature set, we also exploit the underlying complementary information among different features to take full advantage of various features. Therefore, all normalized Laplace matrixes of M feature sets are combined with adaptive weights, and this combination can be achieved by defining the following graph regularization term:

$$\begin{aligned} \Psi(A) = \min_{A,\omega} \sum_{m=1}^{M} \omega_m\, tr\left(A^T L^m A\right) + \beta\, \|\omega\|_2^2, \\ s.t.\ A \geq 0, \quad \sum_{m=1}^{M} \omega_m = 1,\ \omega \geq 0 \end{aligned} \tag{4}$$

where ω_m is the weight of the m-th undirected graph G^m, and the weights of all undirected graphs are combined as a weight vector ω=[$\omega_1, \ \omega_2, \ldots, \ \omega_M$]. The regularization term $\|\omega\|_2^2$ is to avoid ω overfitting any normalized Laplace matrixes [55]; $\beta \geq 0$ is a trade-off parameter.

When the label information of the labeled samples is propagated over the graph, we need to ensure that the prediction labels of the labeled training samples are not change too much from their initial assigned labels. Hence, a penny term is defined as

$$\Gamma\left(A, Y\right) = \min_{A} \sum_{i=1}^{N} \|a_i - y_i\|_2^2 q_{ii}, \tag{5}$$

where q_{ii} is calculated in the following manner: if the i-th sample is a labeled sample, q_{ii} is set as a very large value, otherwise q_{ii} is set as 0.

After label propagation, we aim to learn the mapping matrix between the samples and their labels based on the multiple features of both the labeled and unlabeled samples. Because the regularization term based on the $l_{2,1}$-norm constraint can guarantee row sparsity on the mapping matrix and ensure that the mapping matrix is more robust [57], we adopt the $l_{2,1}$-norm as the regularization term in the proposed SSMFR model. By combining the information of multiple features and minimizing the global label prediction error, SSMFR learns the mapping matrix $W = [W^1, W^2, \ldots, W^M]$ between the samples and the labels by the following formula:

$$\min_{W^m} \sum_{m=1}^{M} \|X^{m^T} W^m - A\|_2^2 + \lambda\, \|W^m\|_{2,1}, \tag{6}$$

where X^m represents the m-th feature set of the samples; $W^m \in \Re^{d_m \times C}$ denotes the mapping matrix between the sample set X^m and its corresponding label set (W^m can also be denominated as the subclassifier corresponding to the m-th feature set). To achieve the minimum of global prediction error of all subclassifiers learned from multiple features, the label prediction error $\|X^{m^T} W^m - A\|_2^2$ of all subclassifiers are summed up and minimized by the term $\min\limits_{W^m} \sum\limits_{m=1}^{M} \|X^{m^T} W^m - A\|_2^2$; $\|W^m\|_{2,1} = \sum_{i=1}^{N} \|W_i^m\|_2$ is used to enforce that W is a row sparse matrix and more robust; $\lambda \geq 0$ is a parameter used to avoid overfitting.

By combining Eqs. (4), (5) and (6), the objective function of SSMFR in Eq. (1) is reformulated as

$$\begin{aligned} &\min_{W^m,A,\omega} \varepsilon\left(W^m, A, \omega\right) \\ &= \sum_{m=1}^{M} \|X^{m^T} W^m - A\|_2^2 + \lambda \|W^m\|_{2,1} \\ &\quad + \alpha\left(\sum_{i=1}^{N} \|a_i - y_i\|_2^2 q_{ii} + \sum_{m=1}^{M} \omega_m\, tr\left(A^T L^m A\right)\right) \\ &\quad + \beta\, \|\omega\|_2^2. \\ &s.t. A \geq 0, \quad \sum_{m=1}^{M} \omega_m = 1,\ \omega \geq 0 \end{aligned} \tag{7}$$

Because $\sum\limits_{m=1}^{M} \omega_m\, tr(A^T L^m A) = tr(A^T(\sum\limits_{m=1}^{M} \omega_m L^m)A)$ and $\sum\limits_{i=1}^{N} \|a_i - y_i\|_2^2 q_{ii} = tr\left(A^T Q A - 2A^T Q Y + Y^T Q Y\right)$ where Q is a diagonal matrix and its diagonal element is q_{ii}, Eq. (7) can be further rewritten as

$$\begin{aligned} &\min_{W^m,A,\omega} \varepsilon\left(W^m, A, \omega\right) \\ &= \sum_{m=1}^{M} \|X^{m^T} W^m - A\|_2^2 + \lambda\, \|W^m\|_{2,1} \end{aligned}$$

$$+\alpha\, tr\left(A^T\left(\sum_{m=1}^{M}\omega_m L^m\right)A+A^TQA-2A^TQY+Y^TQY\right)$$
$$+\beta\|\omega\|_2^2,$$
$$s.t.\ A\geq 0,\ \sum_{m=1}^{M}\omega_m=1,\ \omega\geq 0 \tag{8}$$

where α is an introduced trade-off parameter.

Equation (8) is the ultimate objective function of the proposed SSMFR. Through it, SSMFR spreads the label of the labeled data to the unlabeled data. Consequently, SSMFR can utilize both the information of labeled and unlabeled data to improve the accuracy of the learned classifier. Moreover, SSMFR utilizes adaptive non-negative weighted multigraph label propagation to exploit the latent complementary information contained in different features. The SSMFR model also provides an explicit mapping matrix between each feature set and its prediction label matrix so that the ''out-of-sample'' problem can be effectively avoided [56], [58].

B. OPTIMIZATION

From Eq. (8), it can be seen that there are three variables (i.e., A, ω and W^m) need to be optimized and the objective function of SSMFR is convex to each variable but not convex to them jointly. Therefore, it is unrealistic to find the global optimal solution of Eq. (8). To address this issue, we propose a simple and effective optimization algorithm based on an alternative updating scheme to solve the objective function of the proposed SSMFR. In our scheme, we update each variable by fixing others, and this process is iteratively executed until the objective function achieves a stable value.

1) FIXING ω AND W^m TO UPDATE A

When fixing the weight vector ω and the mapping matrix W^m $(m=1,2,\ldots,M)$, the optimization problem with respect to the prediction label matrix A becomes

$$\min_A \varepsilon(A)=\min_A\sum_{m=1}^{M}||X^{m^T}W^m-A||_2^2$$
$$+\alpha\, tr\left(A^TL\,A+A^TQA-2A^TQY+Y^TQY\right)$$
$$s.t.\ \ A\geq 0, \tag{9}$$

where $L=\sum_{m=1}^{M}\omega_m L^m$.

By introducing the Lagrangian multiplier matrix ξ and removing the unrelated terms with respect to A, we obtain the Lagrange equation of Eq. (9) as

$$\phi(A,\xi)=\sum_{m=1}^{M}tr\left(A^TA-2A^TX^{m^T}W^m\right)$$
$$+\alpha\, tr\left(A^TL\,A+A^TQA-2A^TQY\right)+tr(\xi A). \tag{10}$$

Setting the derivative of Eq. (10), with respect to A, to 0, we obtain

$$\frac{\partial\phi(A,\xi)}{\partial A}=2MA-2\sum_{m=1}^{M}X^{m^T}W^m+2\alpha\,L\,A$$
$$+2\alpha\,QA-2\alpha\,QY+\xi=0. \tag{11}$$

According to the Karush-Kuhn-Tucker condition $\xi_{ij}A_{ij}=0$ [30], we get

$$[MA-F+\alpha\,L\,A+\alpha\,QA-\alpha\,QY]_{ij}A_{ij}=0 \tag{12}$$

where $F=\sum_{m=1}^{M}X^{m^T}W^m$.

We define the matrixes $L=L^+-L^-$ and $F=F^+-\mathrm{F}^-$ according to [16], and then the predicted label matrix A can be solved as

$$A_{ij}\leftarrow A_{ij}\frac{\left[F^++\alpha L^-A+\alpha QY\right]_{ij}}{\left[F^-+\alpha L^+A+\alpha QA+MA\right]_{ij}}. \tag{13}$$

2) FIXING W^m AND A TO UPDATE ω

To solve the weight vector ω, we fix the mapping matrix W^m $(m=1,2,\ldots,M)$ and the prediction label matrix A, then the optimization problem of ω becomes

$$\min_\omega \varepsilon(\omega)=\min_\omega\ \ \rho(\omega)\ =q^T\omega+\beta\,||\omega||_2^2,$$
$$s.t.\sum_{m=1}^{M}\omega_m=1,\quad \omega\geq 0 \tag{14}$$

where $q = (q_1,q_2,\ldots,q_M)^T$, and the element $q_m=tr(A^TL^mA)$.

It can be easily seen that Eq. (14) is a convex quadratic programming problem. By referring to the solution method in [55], we adopt a fast coordinate gradient descent algorithm to solve Eq. (14) quickly and effectively. Based on the constraint conditions $\sum_{m=1}^{M}\omega_m=1$ and $\omega_m\geq 0$, if we only update any pairs of elements ω_i and $\omega_j (i\neq j)$ in ω and fix other elements $\omega_m(m\neq i,j)$ in each iteration solution, we will obtain

$$\omega_j=1-\sum_{\substack{m=1\\m\neq i,j}}^{M}\omega_m-\omega_i. \tag{15}$$

Let $\rho(\omega_i)$ represent the objective function, it can be expressed by

$$\rho(\omega_i)=\sum_{\substack{m=1\\m\neq i,j}}^{M}\omega_m q_m+\beta\sum_{\substack{m=1\\m\neq i,j}}^{M}\omega_m^2+\omega_i q_i+\omega_j q_i+\beta\left(\omega_i^2+\omega_j^2\right)$$
$$=\sum_{\substack{m=1\\m\neq i,j}}^{M}\omega_m q_m+\beta\sum_{\substack{m=1\\m\neq i,j}}^{M}\omega_m^2+\omega_i q_i+\left(1-\sum_{\substack{m=1\\m\neq i,j}}^{M}\omega_m-\omega_i\right)q_i$$
$$+\beta\left[\omega_i^2+\left(1-\sum_{\substack{m=1\\m\neq i,j}}^{M}\omega_m-\omega_i\right)^2\right]. \tag{16}$$

Differentiating Eq. (16) with respect to ω, we get

$$\frac{\partial \rho(\omega_i)}{\partial \omega_i} = q_i - q_j + 2\beta\left(\omega_i - \omega_j\right) = 0. \tag{17}$$

According to Eq. (17), we can obtain

$$\omega_i^* - \omega_j^* = \frac{1}{2\beta}\left(q_j - q_i\right). \tag{18}$$

where ω_i^* and ω_j^* are the results of updating ω_i and ω_j, respectively. Since $\omega_i^* + \omega_j^* = \omega_i + \omega_j$, the updated ω_i^* can be expressed as

$$\omega_i^* = \frac{1}{4\beta}\left(q_j - q_i\right) + \frac{\omega_i + \omega_j}{2}. \tag{19}$$

To ensure that ω_i^* is non-negative, the problem of solving Eq. (19) can be further decomposed into the following form:

1) if $\frac{q_j - q_i}{4\beta} + \frac{\omega_i + \omega_j}{2} \leq 0$, then

$$\begin{cases} \omega_i^* = 0 \\ \omega_j^* = \omega_i + \omega_j. \end{cases} \tag{20}$$

2) According to the symmetry of the variables i and j, we can achieve that if $\frac{q_i - q_j}{4\beta} + \frac{\omega_i + \omega_j}{2} \leq 0$, then

$$\begin{cases} \omega_i^* = \omega_i + \omega_j \\ \omega_j^* = 0, \end{cases} \tag{21}$$

3) otherwise

$$\begin{cases} \omega_i^* = \frac{1}{4\beta}\left(q_j - q_i\right) + \frac{\omega_i + \omega_j}{2} \\ \omega_j^* = \omega_i + \omega_j - \omega_i^*. \end{cases} \tag{22}$$

By using Eqs. (20) to (22), the variables in the weight vector ω are iteratively optimized in pairs until the value of the objective function in Eq. (14) is convergent.

3) FIXING ω AND A TO UPDATE W^m

By fixing the weight vector ω and the prediction label matrix A, the optimization problem of the mapping matrix W^m $(m = 1, 2, \ldots, M)$ can be reduced to

$$\min_{W^m} \varepsilon\left(W^m\right) = \min_{W^m} \sum_{m=1}^{M} ||X^{m^T} W^m - A||_2^2 + \lambda\, ||W^m||_{2,1}. \tag{23}$$

According to the properties of the matrix, we know that the equation $||A||_{2,1} = tr(A^T G A)$ is satisfied for any matrix $A \in \Re^{n \times m}$, where G is a diagonal matrix, and its i-th diagonal element is $g_{ii} = 1/(2||a^i||_2)$. Therefore, Eq. (23) is equivalent to

$$\begin{aligned} &\min_{W^m} \varepsilon\left(W^m\right) \\ &= \min_{W^m} \sum_{m=1}^{M} tr\left(\left(X^{m^T} W^m - A\right)^T \left(X^{m^T} W^m - A\right)\right) \\ &\quad + \lambda tr\left(W^{m^T} G^m W^m\right) \\ &= \min_{W^m} \sum_{m=1}^{M} tr\left(W^{m^T} X^m X^{m^T} W^m - 2W^{m^T} X^m A + A^T A\right) \\ &\quad + \lambda tr\left(W^{m^T} G^m W^m\right), \end{aligned} \tag{24}$$

where G^m is a diagonal matrix, and its i-th diagonal element is $g_{ii} = 1/(2||w_i^m||_2)$; w_i^m is the i-th row of W^m. By setting the derivative of the above objective function with respect to W^m to 0, we have

$$X^m X^{m^T} W^m + \lambda G^m W^m = X^m A. \tag{25}$$

Let Eq. (25) left multiply $(X^m X^{m^T} + \lambda G^m)^{-1}$, we have:

$$W^m = \left(X^m X^{m^T} + \lambda G^m\right)^{-1} X^m A. \tag{26}$$

Since $||w_i^m||_2$ may be zero in practice, it needs to redefine g_{ii}^m as

$$g_{ii}^m = \frac{1}{2||w_i^m||_2 + \upsilon}, \tag{27}$$

where υ is an extremely small constant.

Finally, the W^m in Eq. (23) can be solved by an alternative updating scheme. In the t-th iteration, we first fix G_{t-1}^m to update W_t^m, then we fix W_t^m to update G_t^m. This iteration procedure is repeated until the algorithm converges. The procedure of optimizing W^m is shown in Algorithm 1.

Algorithm 1 Optimizing W^m

1: For $m = 1$: M
2: $t = 1$;
3: **Repeat**
4: Calculate W^m by Eq. (26);
5: Calculate G^m by Eq. (27);
6: Update $t = t + 1$;
7: **Until** convergence;
8: End
9: **Output**: W^m $(m = 1, 2, \ldots, M)$.

In conclusion, the optimization process of our proposed SSMFR algorithm is shown in Algorithm 2. As seen from Algorithm 2, A, ω and W^m are updated alternately and iteratively, which demonstrates that the proposed SSMFR executes label propagation and classification model learning jointly. Therefore, the proposed SSMFR can achieve more accurate classifiers and predicted labels because the predicted labels and the mapping matrix affect each other through each iteration.

C. RECOGNITION CRITERION

The classifiers W^m $(m = 1, 2, \ldots, M)$ can be obtained by using Algorithm 2 with the training dataset. In the recognition phase, the label of a given testing image I represented by M features $z_m \in \Re^{d^m \times 1}(m = 1, 2, \ldots, M)$ can be estimated by

$$Label(I) = \underset{c \in \{1,2,\ldots,C\}}{\arg\max} \left[\frac{\sum_{m=1}^{M} z_m^T W^m}{M}\right]_c, \tag{28}$$

Algorithm 2 SSMFR Algorithm

1: **Input**: The data and label matrices $X^m\,(m = 1, 2, \ldots, M)$ and Y, and the parameters λ, α and β;
2: **Initialize**: $A = rand(N, C)$, $\omega = 1/M$ and $T = 1$;
3: Calculating the weight matrix $S^m\,(m = 1, 2, \ldots, M)$ by using Eq. (2);
4: **Repeat**
5: Update the prediction label matrix A by using Eq. (13);
6: Calculating the weight vector ω through the coordinate gradient descent method;
7: Update the mapping matrix W^m according to Algorithm 1;
8: Update $T = T + 1$;
9: **Until** the value of the objective function in Eq. (8) does not change;
10: **Output:** The mapping matrix $W^m\,(m = 1, 2, \ldots, M)$, the prediction label matrix A and the weight vector ω.

where $\frac{\sum_{m=1}^{M}(z_m)^T W^m}{M}$ is a label prediction vector with C dimensions. Specifically, the recognition process of our proposed SSMFR is shown in Algorithm 3.

Algorithm 3 The Recognition Process of SSMFR

1: **Input**: The training dataset $X^m\,(m = 1, 2, \ldots, M)$ and its label set Y, M features $z^m \in \Re^{d^m \times 1}(m = 1, 2, \ldots, M)$ of a given testing image I;
2: Calculate the mapping matrix $W^m\,(m = 1, 2, \ldots, M)$ according to Algorithm 2;
3: Compute the label of the testing sample I by using Eq. (28);
4: **Output**: The prediction.

D. CONVERGENCE ANALYSIS

To prove the convergence of SSMFR, we need to prove that the subproblems as shown in Eqs. (9), (14) and (23) will decrease the objective function value of our SSMFR. Since our proposed SSMFR essentially follows the similar fashion of many existing algorithms, e.g., the algorithms proposed in Refs. [16] and [59], we exclude the proof that Eq. (9) will not be increased when the variable A is updated by Eq. (13), please refer the literature [16] and [59] for details. Furthermore, Eq. (14) has been proven to be strictly convex in the literature [17], [55], which ensures that the value of the objective function of SSMFR is decreasing by utilizing the coordinate gradient descent to solve it at each iteration. Hence, to demonstrate that SSMFR algorithm is convergent, it is just need to prove that Eq. (23) will not be increased when variable W^m is updated by Algorithm 1.

According to Algorithm 1, if we fix G^m to G_t^m and update W_{t+1}^m in the t-th iteration, we have

$$\varphi(W_{t+1}^m, G_t^m) \leq \varphi(W_t^m, G_t^m). \tag{29}$$

Equation (29) can be rewritten as

$$\begin{aligned} &tr\left(\left(X^{m^T} W_{t+1}^m - A\right)^T \left(X^{m^T} W_{t+1}^m - A\right)\right) \\ &+ \lambda tr\left(W_{t+1}^{m^T} G_t^m W_{t+1}^m\right) \leq tr\left(\left(X^{m^T} W_t^m - A\right)^T \left(X^{m^T} W_t^m - A\right)\right) \\ &+ tr\left(W_t^{m^T} G_t^m W_t^m\right). \end{aligned} \tag{30}$$

Since $||W^m||_{2,1} = \sum_{i=1}^{m} ||w_i^m||_2$, we have

$$\begin{aligned} &tr\left(\left(X^{mT} W_{t+1}^m - A\right)^T \left(X^{mT} W_{t+1}^m - A\right)\right) \\ &+ \lambda \sum_i \frac{||(w_i^m)_{t+1}||_2^2}{2||(w_i^m)_t||_2} \leq tr\left(\left(X^{mT} W_{t+1}^m - A\right)^T \left(X^{mT} W_{t+1}^m - A\right)\right) \\ &+ \lambda \sum_i \frac{||(w_i^m)_t||_2^2}{2||(w_i^m)_t||_2}, \end{aligned} \tag{31}$$

and

$$\begin{aligned} &tr\left(\left(X^{mT} W_{t+1}^m - A\right)^T \left(X^{mT} W_{t+1}^m - A\right)\right) \\ &\quad + \lambda \sum_i ||(w_i^m)_{t+1}||_2 - \lambda\Big(\sum_i ||(w_i^m)_{t+1}||_2 - \sum_i \frac{||(w_i^m)_{t+1}||_2^2}{2||(w_i^m)_t||_2}\Big) \\ &\leq tr\left(\left(X^{mT} W_{t+1}^m - A\right)^T \left(X^{mT} W_{t+1}^m - A\right)\right) + \lambda \sum_i ||(w_i^m)_t||_2 \\ &\quad - \lambda\Big(\sum_i ||(w_i^m)_t||_2 - \sum_i \frac{||(w_i^m)_t||_2^2}{2||(w_i^m)_t||_2}\Big). \end{aligned} \tag{32}$$

Given any i, we can clearly find that

$$||(w_i^m)_{t+1}||_2 - \frac{||(w_i^m)_{t+1}||_2^2}{2||(w_i^m)_t||_2} \leq ||(w_i^m)_t||_2 - \frac{||(w_i^m)_t||_2^2}{2||(w_i^m)_{t+1}||_2}. \tag{33}$$

Thus, the following inequation is satisfied.

$$\begin{aligned} &\sum_i ||(w_i^m)_{t+1}||_2 - \sum_i \frac{||(w_i^m)_{t+1}||_2^2}{2||(w_i^m)_t||_2} \\ &\qquad \leq \sum_i ||(w_i^m)_t||_2 - \sum_i \frac{||(w_i^m)_t||_2^2}{2||(w_i^m)_t||_2}. \end{aligned} \tag{34}$$

By combining Eqs. (31) and (34), we get

$$\begin{aligned} &||X^{mT} W_{t+1}^m - A||_2^2 + \lambda ||W_{t+1}^m||_{2,1} \\ &\qquad \leq ||X^{mT} W_t^m - A||_2^2 + \lambda ||W_t^m||_{2,1}. \end{aligned} \tag{35}$$

Therefore, the following inequation holds.

$$\begin{aligned} &\sum_{m=1}^{M} ||X^{mT} W_{t+1}^m - A||_2^2 + \lambda ||W_{t+1}^m||_{2,1} \\ &\qquad \leq \sum_{m=1}^{M} ||X^{mT} W_t^m - A||_2^2 + \lambda ||W_t^m||_{2,1}. \end{aligned} \tag{36}$$

Eq. (36) demonstrates that the objective function in Eq. (23) is not increased in each iteration.

We now discuss the convergence of our proposed SSMFR algorithm shown in Algorithm 2. Let $\eta(A^t, \omega^t, W^t)$ represent the objective function value of our SSMFR in the

t-th iteration. Based on the above convergence analysis, it can be easily seen that by using Eq. (13), the coordinate gradient descent method and Algorithm 1 given above, we can obtain the optimal solution of each variable in the $(t+1)$-th iteration by fixing two others. Meanwhile, we have $\eta(A^{t+1}, \omega^t, W^t) \leq \eta(A^{t+1}, \omega^t, W^t)$, $\eta(A^t, \omega^{t+1}, W^t) \leq \eta(A^t, \omega^t, W^t)$ and $\eta(A^t, \omega^t, W^{t+1}) \leq \eta(A^t, \omega^t, W^{t+1})$, and by combining them together, we get $\eta(A^{t+1}, \omega^{t+1}, W^{t+1}) \leq \eta(A^t, \omega^t, W^t)$. Therefore, Algorithm 2 can ensure that the value of $\eta(A, \omega, W)$ is nonincreasing. Moreover, because all terms in Eq. (8) is not smaller than 0, the function $\eta(A^t, \omega^t, W^t)$ has the lower bound. According to $\eta(A^{t+1}, \omega^{t+1}, W^{t+1}) \leq \eta(A^t, \omega^t, W^t)$ and Cauchy's Convergence Rule [60], the proposed optimization algorithm for SSMFR is convergent.

IV. EXPERIMENTS

In this section, the effectiveness of SSMFR is evaluated by extensive experiments.

A. DATASET

To test the performance of our proposed SSMFR, we evaluated it on five popular scene datasets, including three small datasets (Scene8 [22], UIUC Sports [61] and Scene15 [24]) and two large dataset (MIT Indoor [62] and SUN397 [2]).

The Scene8 dataset [22] consists of 2688 images across 8 different scene categories, and the size of each image is 256 $\times$ 256. The UIUC Sports dataset [61] contains 1585 images of 8 sport scene classes, and the minimum resolution of the images is about 800 $\times$ 600. The Scene8 dataset [24] consists of 15 scene categories with a total of 4485 images, which are approximately 300$\times$250 in average resolution. The MIT Indoor dataset [62] contains 15620 images labeled into 67 indoor scene categories, and all images have a minimum resolution of 200 pixels on the smallest axis. It should be noted that we just use a subset containing 6700 images of MIT Indoor dataset in the experiments. The SUN397 dataset [2] contains 397 scene categories and each category includes at least 100 images. The total number of images is 108754 which make it extremely challenging for scene recognition task [32]. Similar to the literatures [63]–[65], we also use a subset of SUN397 to evaluate our proposed method. The subset is created by randomly selecting 100 scene categories and each category has 100 randomly selected images. Fig. 2 shows some images from the different datasets.

The abovementioned datasets are fairly challenging for scene recognition because 1) the same or similar objects usually appear in the different scene categories; 2) the appearances, sizes and numbers of objects are very diverse within the same scene class; and 3) the backgrounds of the images belonging to the same scene class are very different [17].

FIGURE 2. Example images from different datasets.

B. EXPERIMENTAL SETUP

In the experiments, we extracted several features to describe the color, texture and shape information of the scene images. For the datasets Scene8, UIUC Sports, Scene15 and MIT Indoor, we computed several different features for each dataset. Specifically, when extracting GIST [22] and the

TABLE 1. Details of the experimental datasets.

Dataset	Number of samples	Number of classes	Features
Scene8	2688	8	GIST, PHOW, LBP, Gradient, HSV
UIUC Sports	1579	8	GIST, PHOW, LBP, Gradient, HSV
Scene15	4485	15	GIST, PHOW, LBP, Gradient
MIT Indoor	6700	67	GIST, PHOW, LBP
SUN397	10000	100	GIST, Geo-Texton, LBP

Pyramid Histogram of Visual Words (PHOW) [66] feature descriptors, the parameters were set referring to [22] and [66], respectively. Local Binary Pattern (LBP) [67], gradient histogram and HSV color histogram were extracted from nonoverlapping 32×32 subregions of each image, and the bag-of-words model [68] was used to quantify them as feature vectors. For SUN397 dataset, we used the features (GIST [22], Geo-Texton [2] and LBP [69])[1] precomputed by the Xiao *et al.* [2]. Table 1 gives the details of each dataset and its feature descriptors. Note that the HSV color histogram was not extracted for the Scene15 dataset since it includes gray images, and HSV color and gradient histograms are not extracted for the MIT Indoor dataset since it is very time-consuming to extract these two features for this large-scale dataset. In each dataset, 40%, 30% and 30% of samples are randomly selected as the labeled training samples, unlabeled training samples and testing samples, respectively. The process of random sample selection was repeated 10 times, and the average recognition rate and standard deviation are provided in the following subsections.

C. EXPERIMENTAL RESULTS AND ANALYSIS

We executed two experiments to illustrate the validity of our proposed SSMFR.

First, to prove that our proposed SSMFR model can effectively fuse multiple features to improve the recognition performance, the SSMFR model with the single feature was compared with the SSMFR model with multiple features on five datasets. The comparison results are shown in Table 2. As shown in Table 2, when only adopting a single feature, the best recognition performances of Scene8, Scene15 and MIT Indoor datasets were obtained by adopting GIST features, the best recognition performance of UIUC Sports was obtained by PHOW features, and the best recognition performances of SUN397 was gained by LBP feature. The highest recognition accuracies (%) and standard deviations (%) of the five datasets were 78.08±0.88, 64.34±1.48, 62.84±1.03, 16.32±0.62 and 19.76±0.48 respectively. While employing the fusion of multiple features, SSMFR can improve the recognition rates

[1]https://vision.princeton.edu/projects/2010/SUN/

TABLE 2. Average recognition rates (%) and standard deviations (%) of SSMFR by using a single feature and multiple features.

Feature \ Dataset	Scene8	UIUC Sports	Scene15	MIT Indoor	SUN397
GIST	78.08±0.88	63.73±1.71	62.84±1.03	16.32±0.62	17.84±0.45
PHOW	73.42±1.28	64.34±1.48	54.53±1.17	13.87±1.07	—
LBP	56.54±1.57	34.15±2.01	34.49±0.88	4.47±0.34	19.76±0.48
Gradient	50.39±1.18	30.38±1.81	28.43±0.89	—	—
HSV	42.96±1.91	34.53±2.79	—	—	—
Geo-Texton	—	—	—	—	18.44±0.89
All Features	84.76±1.15	79.62±1.19	76.80±0.89	22.78±1.05	29.92±0.72

NOTE: "—" means that the feature (row) is not adopted to describe the content of images in that dataset (column).

on the five datasets to 84.76±1.15, 79.62±1.19, 76.80±0.89, 22.78±1.05 and 29.92±0.72 respectively. The average recognition precision was significantly increased by approximately 10%. These experimental results demonstrate that the accuracy of scene recognition will be limited when only a single feature is used and the proposed SSMFR algorithm can greatly improve the recognition performance by effectively fusing the multiple features.

TABLE 3. Recognition rates (%) and standard deviations (%) of the different methods on five datasets.

Method \ Dataset	Scene8	UIUC Sports	Scene15	MIT Indoor	SUN397
FSSI	79.45±1.77	71.69±1.41	69.85±1.08	21.56±0.91	20.75±0.74
LRRADP	78.51±1.33	70.30±2.14	73.18±0.95	19.35±1.22	10.84±0.72
MMSSL	81.38±1.18	71.78±1.76	70.17±1.14	16.34±0.34	18.60±0.84
SFSS	82.03±1.08	74.96±1.92	74.22±0.97	20.11±1.20	26.96±0.88
MLAN	82.23±1.18	75.23±1.77	74.58±1.02	18.67±0.92	23.38±0.48
SSMFR	**84.76±1.15**	**79.62±1.19**	**76.80±0.89**	**22.78±1.05**	**29.92±0.72**

Next, we compared the proposed SSMFR algorithm with several state-of-the-art algorithms, including Feature Selection with Shared Information among Multiple Tasks (FSSI) [42], Low Rank Representation with Adaptive Distance Penalty (LRRADP) [70], Multi-Modal Semi-Supervised Learning Model (MMSSL) [6], Structural Feature Selection with Sparsity (SFSS) [71], Multi-view Learning with Adaptive Neighbours (MLAN) [72]. FSSI is a supervised multi-feature learning method, LRRADP and SFSS are semi-supervised learning method without multi-feature learning, and the other two algorithms are semi-supervised multi-feature learning methods. The results of all methods are described in Table 3.

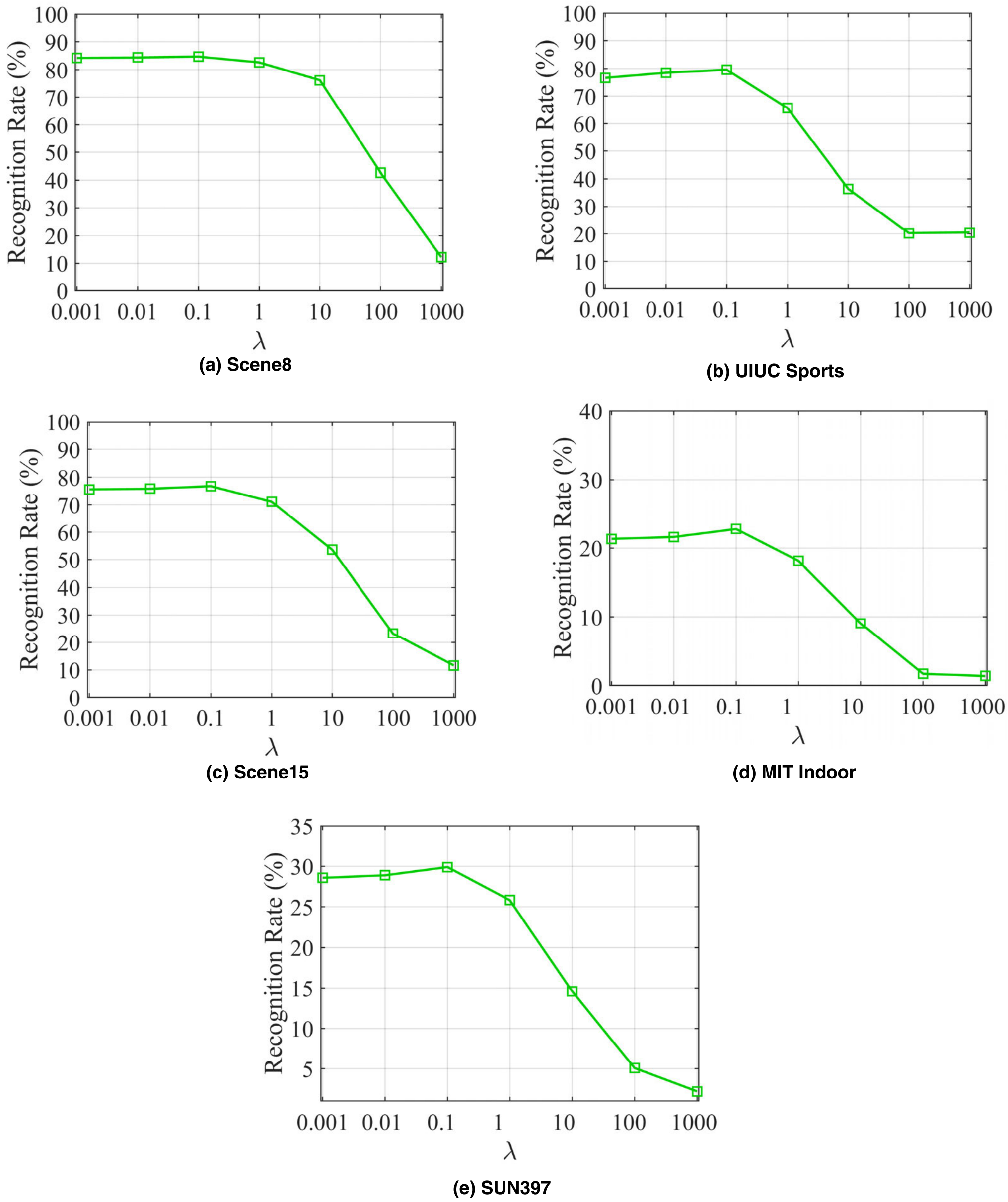
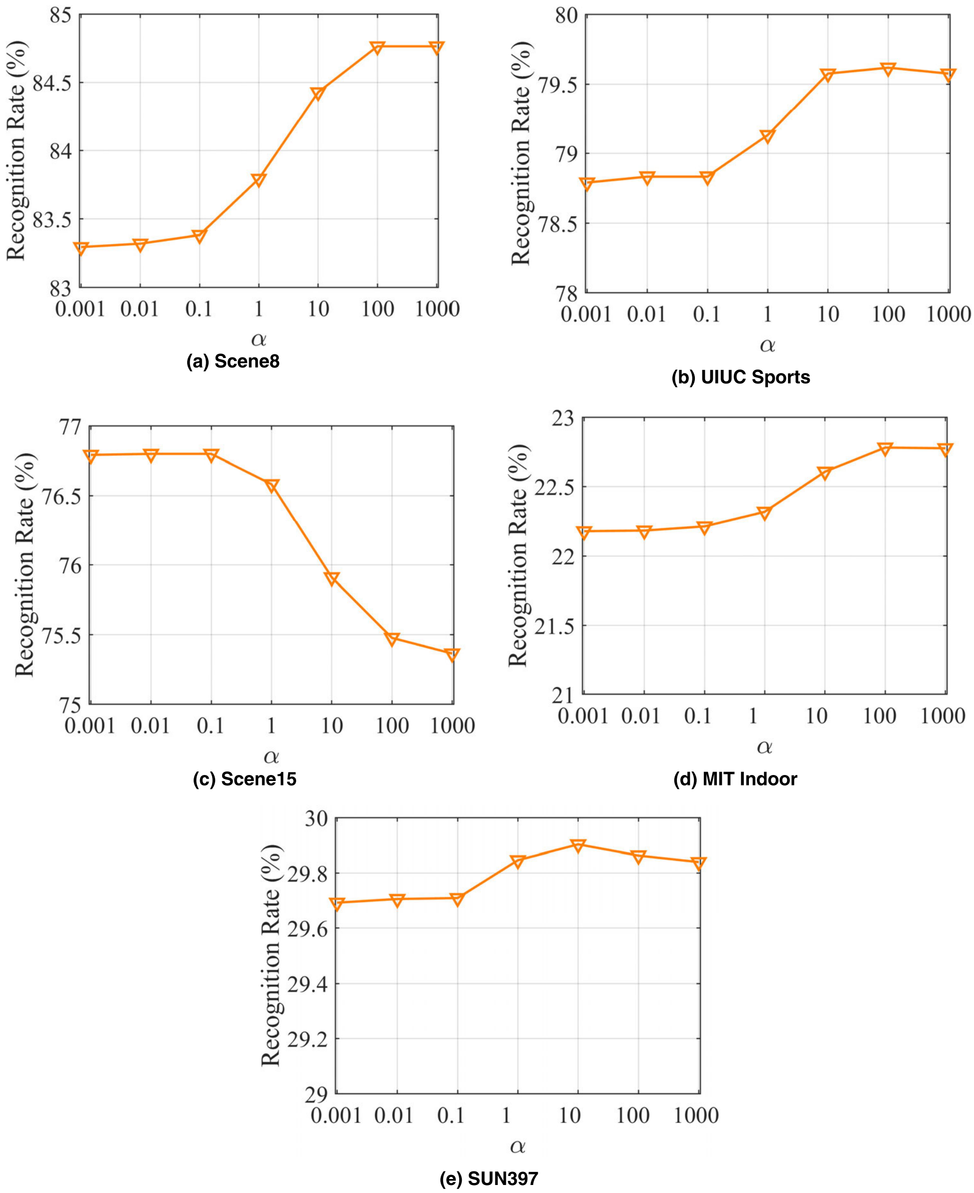


FIGURE 3. **Example images from different datasets. Recognition rate of the proposed SSMFR with different λ values in the five datasets.**

By comparing the results in Table 3, we have the following observations. First, the overall performance of the supervised algorithm FSSI was generally worse than the other five semi-supervised algorithms (LRRADP, MMSSL, SFSS, MLAN and SSMFR) because FSSI cannot utilize the useful information of unlabeled samples, which demonstrates that it is beneficial to use unlabeled data for scene recognition tasks, especially when the amount of labeled data is not large. Second, the proposed SSMFR always gained better performance than the other four semi-supervised algorithms. This maybe because that SFSS and LRRADP directly connecting multiple features into one feature vector, which cannot efficiently fuse multiple features to recognize scene images. For MLAN, since there is no label propagation mechanism in it, the label information of labeled samples cannot be fully used. Although MMSSL employs the label propagation technique to take full advantage of the label information, it does not adopt $l_{2,1}$-norm as the regularization. Thus, the classifier obtained by this approach may lack of robustness to the potential noise or outliers in the data.

Finally, to further verify the performance of our proposed SSMFR is superior to other methods, we used the one-tailed Wilcoxon rank sum test to validate whether SSMFR performs significantly better than other compared methods. In this one-tailed Wilcoxon rank sum test, the null hypothesis is that the performance of SSMFR is the same as that of the



(a) Scene8 (b) UIUC Sports (c) Scene15 (d) MIT Indoor (e) SUN397

FIGURE 4. Recognition rate of the proposed SSMFR with different α values in the five datasets.

compared methods, and the alternative hypothesis is that the performance of SSMFR is better than that of the compared methods. For instance, if we want to compare the recognition rate of SSMFR with that of FSSI (SSMFR vs. FSSI), the null hypotheses can be defined as H_0: $M_{\text{SSMFR}} = M_{\text{FSSI}}$, and the alternative hypothesis can be defined as H_1: $M_{\text{SSMFR}} > M_{\text{FSSI}}$, where M_{SSMFR} and M_{FSSI} represent the medians of the recognition rates of SSMFR and FSSI, respectively, on all datasets. The significance level of this test is set as 0.05 and the test results are shown in Table 4. From this Table, it can be found that the p-values obtained by all pair-wise one-tailed Wilcoxon rank sum tests are less than 0.05, which means the alternative hypotheses are accepted in all tests and our proposed SSMFR significantly outperforms other compared methods in this paper.

D. PARAMETER SENSITIVITY ANALYSIS

The parameter sensitivity analysis of three parameters λ, α and β in the proposed SSMFR model are discussed in this section. λ is used to avoid the over-fitting of the proposed model, α is used to control the importance of the adaptive multiple graphs-based label propagation term in Eq. (8), and β is used to control the sparse of the non-negative weight vector ω. In experiments, λ, α and β are tuned from the set $\{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$. Specifically, we have tested the influence of the combination of three parameters (λ, α and β) together, but for the ease of understanding and concision, when we plot the Figs. 3-5, we fix two of the three parameters as their optimal values and report the average recognition accuracy and standard deviation while the other parameter is changing. The recognition performance

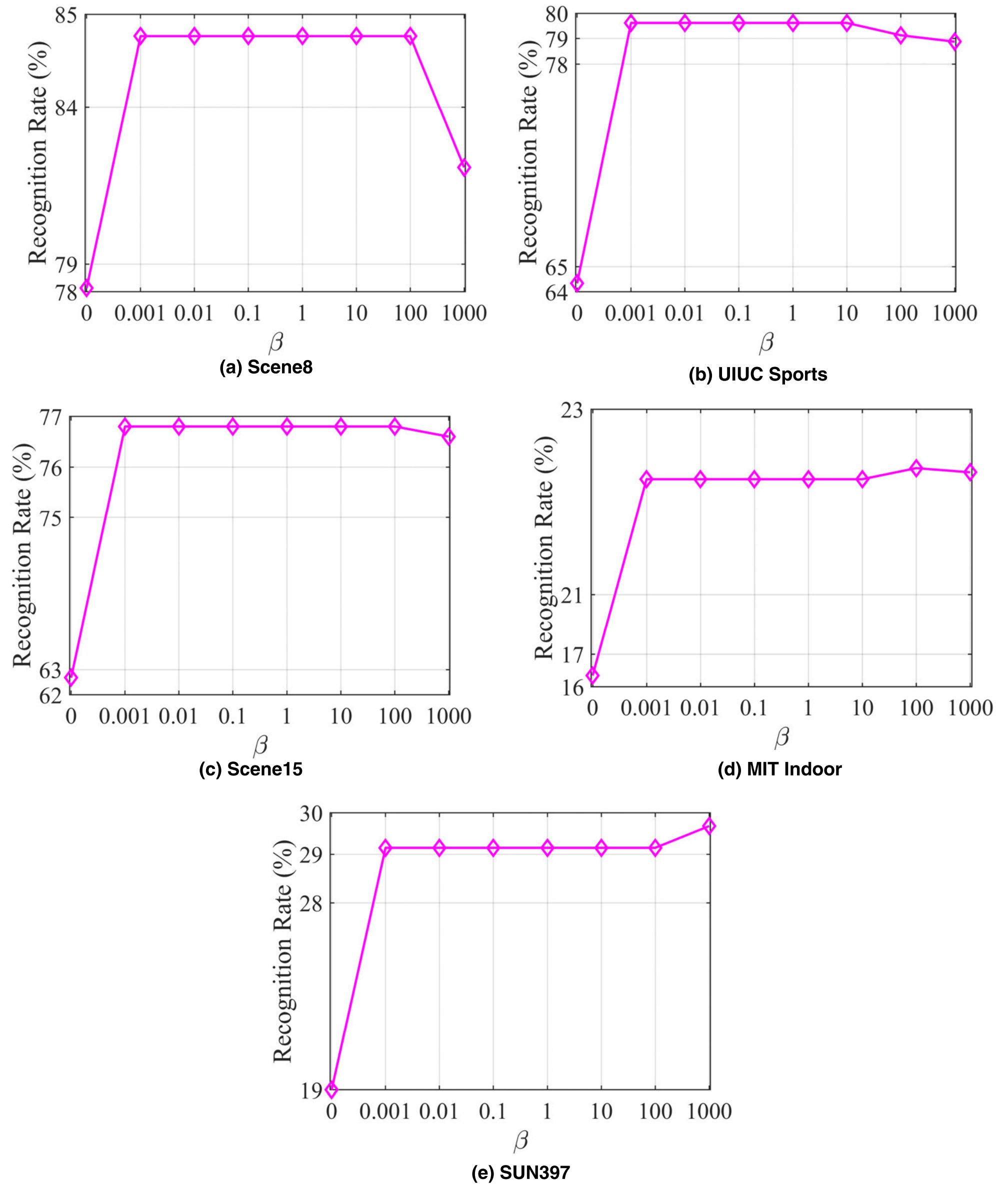


FIGURE 5. Recognition rate of the proposed SSMFR with different β values in the five datasets.

TABLE 4. The *P*-values of the pair-wise one-tailed wilcoxon rank sum tests on the five datasets.

FSSI vs. SSMFR	3.84×10^{-4}
LRRADP vs. SSMFR	1.11×10^{-4}
MMSSL vs. SSMFR	6.50×10^{-4}
SFSS vs. SSMFR	0.0120
MLAN vs. SSMFR	0.0072

of SSMFR with different values of λ, α and β for different datasets are described in Figs. 3-5, from which we can conclude the following points: 1) A small value of parameter λ contributes to a much better performance (see Fig. 3). This is because if λ is too large, only the regularization term $||W^m||_{2,1}$ in the SSMFR model takes a dominant role, which leads the mapping relationship between the feature set and the label matrix to not be accurately learned and thus decreases the recognition accuracy. 2) With the increase of parameter α values, the recognition rate shows a trend of increasing first and then descending (see Fig. 4). This is because, as the value of α increases, the information of unlabeled data can be more fully used by SSMFR, which is beneficial for improving the recognition performance. When the optimum recognition rate is reached, if the value of α continues to increase, the role of the other terms in Eq. (8) will be ignored, resulting in a decline in the performance of the SSMFR. 3) From Fig. 5, it can be generally observed that the proposed SSMFR achieves its best performances under moderate values of parameter β which controls the sparse of

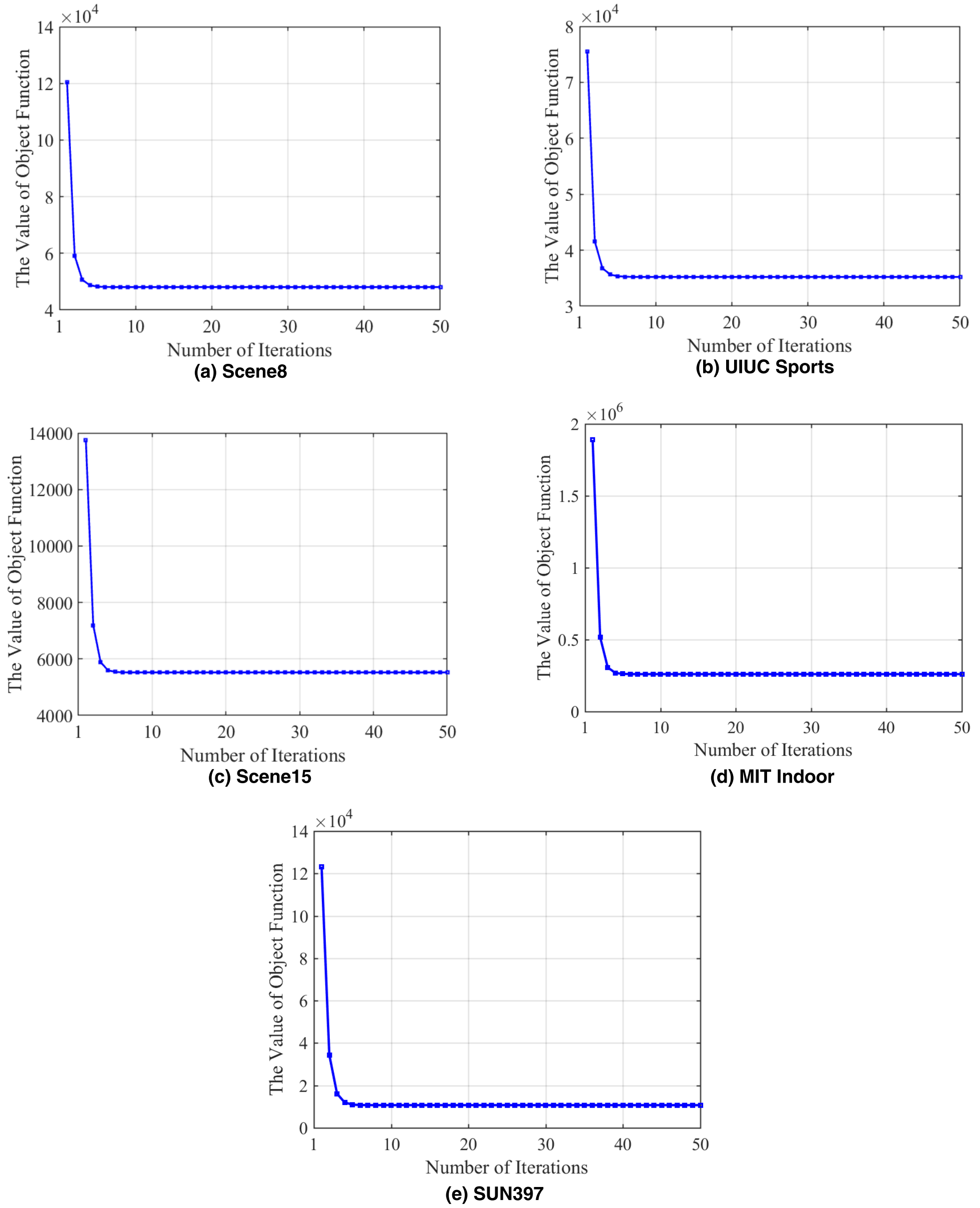


FIGURE 6. Convergence curves of the proposed SSMFR in the five datasets.

non-negative weight vector. This is because $\beta = 0$ will lead a trivial solution of Eq. (14) as

$$\omega_i = \begin{cases} 1, & \text{if } q_i = \min_{m=1,\ldots,M} q_m \\ 0, & \text{otherwise.} \end{cases} \tag{37}$$

This extremely sparse solution is undesirable since SSMFR only adopts one feature and ignores the useful complementary information among different features. Conversely, a uniform weight vector (i.e.,$\omega_i = 1/M(i = 1, 2, \ldots, M)$) will be obtained when β is set as a very large value, and the different contribution of multiple features is neglected. Moreover, we can also find that the performance of our algorithm is insensitive to the β values when it is set neither too small nor too large.

E. CONVERGENCE EVALUATION

The convergence proofs of the proposed SSMFR are presented in Section III.D. Here, the convergence curves of our approach on the five datasets are shown in Fig. 6, in which the x-axis represents the iteration numbers and y-axis represents the objective function value in Eq. (8). It can be easily seen from Fig. 6 that the value of the objective function of our proposed algorithm gradually decreases with the increase of iteration times, and the curve of the objective function value eventually reaches flat. Therefore, our proposed algorithm is convergent on the five datasets.

V. CONCLUSION

Scene recognition is helpful in narrowing the gap between computers and human beings when exploring an understanding of a scene. However, due to the complexity of scene images and the limited amount of labeled data, scene recognition is challenging work. Specifically, in the scene recognition problem, different information of scene images (color, texture, shape, etc.) can be described by different features, while a large number of scene images are lack of manual category labels. Hence, integrating multi-feature learning and semi-supervised learning into a unified model cannot only fuse different features of images, but also use the information of unlabeled images to improve the efficiency of scene recognition.

In this paper, we proposed a new semi-supervised scene recognition method called SSMFR. SSMFR can fully exploit the complementary information contained in various features and learn a unified global label matrix and the subclassifiers corresponding to each feature jointly. By introducing the adaptive weighted multigraph label propagation, the information of unlabeled samples can be adopted to improve the performance of the classifier. By using $l_{2,1}$-norm to constrain the learning process of the classifier, SSMFR can learn a more robust classifier for scene recognition. In addition, we present an effective iterative algorithm to solve the SSMFR model. A large number of experiments were executed on five classical databases, and the experimental results verify that our proposed SSMFR is valid.

At last, it should be pointed out that our SSMFR is a linear approach. Thus, it may not be able to process highly nonlinear distributed data. In our future work, we will try to combine our proposed algorithm with nonlinear kernel function or deep learning framework for solving this problem.

REFERENCES

[1] S. H. Khan, M. Hayat, M. Bennamoun, R. Togneri, and F. A. Sohel, "A discriminative representation of convolutional features for indoor scene recognition," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3372–3383, Jul. 2016.

[2] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "SUN database: Large-scale scene recognition from abbey to zoo," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3485–3492.

[3] I. Ulrich and I. Nourbakhsh, "Appearance-based place recognition for topological localization," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2000, pp. 1023–1029.

[4] H. Li, Y. Shi, B. Zhang, and Y. Wang, "Superpixel-based feature for aerial image scene recognition," *Sensors*, vol. 18, no. 1, p. 156, Jan. 2018.

[5] M. Liu, R. Chen, D. Li, Y. Chen, G. Guo, and Z. Cao, "Scene recognition for indoor localization using a multi-sensor fusion approach," *Sensors*, vol. 17, no. 12, p. 2847, 2017.

[6] X. Cai, F. Nie, W. Cai, and H. Huang, "Heterogeneous image features integration via multi-modal semi-supervised learning model," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1737–1744.

[7] S. H. Khan, M. Hayat, and F. Porikli, "Scene categorization with spectral features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2017, pp. 5638–5648.

[8] C. Zheng, Y. Yi, M. Qi, F. Liu, C. Bi, J. Wang, and J. Kong, "Multicriteria-based active discriminative dictionary learning for scene recognition," *IEEE Access*, vol. 6, pp. 4416–4426, 2017.

[9] Z.-Z. Lan, L. Bao, S.-I. Yu, W. Liu, and A. G. Hauptmann, "Double fusion for multimedia event detection," in *Proc. Int. Conf. MultiMedia Modeling*, Jan. 2012, pp. 173–185.

[10] Y. Yang, J. Song, Z. Huang, Z. Ma, N. Sebe, and A. G. Hauptmann, "Multi-feature fusion via hierarchical regression for multimedia analysis," *IEEE Trans. Multimedia*, vol. 15, no. 3, pp. 572–581, Apr. 2013.

[11] Z.-L. Sun, D. Rajan, and L.-T. Chia, "Scene classification using multiple features in a two-stage probabilistic classification framework," *Neurocomputing*, vol. 73, nos. 16–18, pp. 2971–2979, Oct. 2010.

[12] X. Song, S. Jiang, and L. Herranz, "Joint multi-feature spatial context for scene recognition in the semantic manifold," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1312–1320.

[13] X. Song, S. Jiang, and L. Herranz, "Multi-scale multi-feature context modeling for scene recognition in the semantic manifold," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2721–2735, Jun. 2017.

[14] Y. Yi, W. Zhou, Q. Liu, G. Luo, J. Wang, Y. Fang, and C. Zheng, "Ordinal preserving matrix factorization for unsupervised feature selection," *Signal Process., Image Commun.*, vol. 67, pp. 118–131, Sep. 2018.

[15] Y. Yi, W. Zhou, C. Bi, G. Luo, Y. Cao, and Y. Shi, "Inner product regularized nonnegative self representation for image classification and clustering," *IEEE Access*, vol. 5, pp. 14165–14176, 2017.

[16] C. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, Jan. 2010.

[17] R. F. Brem, J. A. Rapelyea, G. Zisman, J. W. Hoffmeister, and M. P. DeSimio, "Evaluation of breast cancer with a computer-aided detection system by mammographic appearance and histopathology," *Cancer*, vol. 104, no. 5, pp. 931–935, Sep. 2005.

[18] Y. Han, Y. Yang, Y. Yan, Z. Ma, N. Sebe, and X. Zhou, "Semisupervised feature selection via spline regression for video semantic recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 2, pp. 252–264, Feb. 2015.

[19] Y. Li, M. Dixit, and N. Vasconcelos, "Deep scene image classification with the MFAFVNet," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2017, pp. 5757–5765.

[20] A. Bosch, X. Muñoz, and R. Martí, "Which is the best way to organize/classify images by content?" *Image Vis. Comput.*, vol. 25, no. 6, pp. 778–791, Jun. 2007.

[21] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[22] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[23] J. Wu and J. M. Rehg, "Centrist: A visual descriptor for scene categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1489–1501, Aug. 2011.

[24] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 2169–2178.

[25] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 524–531.

[26] S.-W. Choi, C. H. Lee, and I. K. Park, "Scene classification via hypergraph-based semantic attributes subnetworks identification," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 361–376.

[27] L.-J. Li, H. Su, Y. Lim, and L. Fei-Fei, "Objects as attributes for scene classification," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2010, pp. 57–69.

[28] L.-J. Li, H. Su, L. Fei-Fei, and E. P. Xing, "Object bank: A high-level image representation for scene classification & semantic feature sparsification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1378–1386.

[29] Y. Yuan, L. Mou, and X. Lu, "Scene recognition by manifold regularized deep learning architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2222–2233, Oct. 2015.

[30] W. Guo, R. Wu, Y. Chen, and X. Zhu, "Deep learning scene recognition method based on localization enhancement," *Sensors*, vol. 18, no. 10, p. 3376, Oct. 2018.

[31] L. Herranz, S. Jiang, and X. Li, "Scene recognition with CNNs: Objects, scales and dataset bias," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 571–579.

[32] S. Guo, W. Huang, L. Wang, and Y. Qiao, "Locally supervised deep hybrid model for scene recognition," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 808–820, Feb. 2017.

[33] H. Zhu, J.-B. Weibel, and S. Lu, "Discriminative multi-modal feature fusion for RGBD indoor scene recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2969–2976.

[34] W. Liu, X. Ma, Y. Zhou, D. Tao, and J. Cheng, "*p*-Laplacian regularization for scene recognition," *IEEE Trans. Cybern.*, vol. 49, no. 8, pp. 2927–2940, Aug. 2019.

[35] W. Han, R. Feng, L. Wang, and Y. Cheng, "A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 23–43, Nov. 2018.

[36] X. Dai, X. Wu, B. Wang, and L. Zhang, "Semi-supervised scene classification for remote sensing images based on CNN and ensemble learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 4732–4735.

[37] L. Zhang and D. Zhang, "Visual understanding via multi-feature shared learning with global consistency," *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 247–259, Feb. 2016.

[38] H. Ma, J. Zhu, M. R.-T. Lyu, and I. King, "Bridging the semantic gap between image contents and tags," *IEEE Trans. Multimedia*, vol. 12, no. 5, pp. 462–473, Aug. 2010.

[39] J. Song, Y. Yang, Z. Huang, H. T. Shen, and R. Hong, "Multiple feature hashing for real-time large scale near-duplicate video retrieval," in *Proc. 19th ACM Int. Conf. Multimedia*, Nov./Dec. 2011, pp. 423–432.

[40] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, nos. 3–4, pp. 321–377, 1936.

[41] J. Farquhar, D. Hardoon, H. Meng, J. S. Shawe-Taylor, and S. Szedmak, "Two view learning: SVM-2K, theory and practice," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 355–362.

[42] Y. Yang, Z. Ma, A. G. Hauptmann, and N. Sebe, "Feature selection for multimedia analysis by sharing information among multiple tasks," *IEEE Trans. Multimedia*, vol. 15, no. 3, pp. 661–669, Apr. 2013.

[43] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 6, pp. 1438–1446, Dec. 2010.

[44] Z. Ren, Y. Deng, and Q. Dai, "Local visual feature fusion via maximum margin multimodal deep neural network," *Neurocomputing*, vol. 175, pp. 427–432, Jan. 2016.

[45] Z. Wang, Y. Feng, T. Qi, X. Yang, and J. J. Zhang, "Adaptive multi-view feature selection for human motion retrieval," *Signal Process.*, vol. 120, pp. 691–701, Mar. 2016.

[46] X. Zhu, "Semi-supervised learning literature survey," Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, Wisconsin, Tech. Rep. TR1530, 2005.

[47] M. Zhao, Z. Zhang, T. W. S. Chow, and B. Li, "Soft label based Linear Discriminant Analysis for image recognition and retrieval," *Comput. Vis. Image Understand.*, vol. 121, pp. 86–99, Apr. 2014.

[48] Y. Yi, C. Bi, X. Li, J. Wang, and J. Kong, "Semi-supervised local ridge regression for local matching based face recognition," *Neurocomputing*, vol. 167, pp. 132–146, Nov. 2015.

[49] O. Chapelle, J. Weston, and B. Schölkopf, "Cluster kernels for semi-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2003, pp. 601–608.

[50] 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.*, 2004, pp. 321–328.

[51] Y. Yi, Y. Shi, H. Zhang, J. Wang, and J. Kong, "Label propagation based semi-supervised non-negative matrix factorization for feature extraction," *Neurocomputing*, vol. 149, pp. 1021–1037, 2015.

[52] Z. Wang, Q. Song, Y. C. Soh, and K. Sim, "An adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation," *Comput. Vis. Image Understand.*, vol. 117, no. 10, pp. 1412–1420, Oct. 2013.

[53] Y. Yi, S. Qiao, W. Zhou, C. Zheng, Q. Liu, and J. Wang, "Adaptive multiple graph regularized semi-supervised extreme learning machine," *Soft Comput.*, vol. 22, no. 11, pp. 3545–3562, Jun. 2018.

[54] J. Wang, R. Zhao, Y. Wang, C. Zheng, J. Kong, and Y. Yi, "Locality constrained graph optimization for dimensionality reduction," *Neurocomputing*, vol. 245, pp. 55–67, Jul. 2017.

[55] P. Li, J. Bu, C. Chen, Z. He, and D. Cai, "Relational multimanifold coclustering," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1871–1881, Dec. 2013.

[56] M. Zhao, Z. Zhang, T. W. S. Chow, and B. Li, "A general soft label based linear discriminant analysis for semi-supervised dimensionality reduction," *Neural Netw.*, vol. 55, pp. 83–97, Jul. 2014.

[57] F. Nie, H. Huang, X. Cai, and C. H. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1813–1821.

[58] M. Zhao, B. Li, Z. Wu, and C. Zhan, "Image classification via least square semi-supervised discriminant analysis with flexible kernel regression for out-of-sample extension," *Neurocomputing*, vol. 153, pp. 96–107, Apr. 2015.

[59] Q. Gu and J. Zhou, "Co-clustering on manifolds," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun./Jul. 2009, pp. 359–368.

[60] W. Rudin, *Principles of Mathematical Analysis*, vol. 3. New York, NY, USA: McGraw-Hill, 1964.

[61] L.-J. Li and L. Fei-Fei, "What, where and who? classifying events by scene and object recognition," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–8.

[62] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 413–420.

[63] H. J. Kim and J.-M. Frahm, "Hierarchy of alternating specialists for scene recognition," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 451–467.

[64] E. Zolghadr and B. Furht, "Scene understanding using context-based conditional random field," in *Proc. 24th Conf. Comput. Graph., Vis. Comput. Vis. (WSCG)*, 2016.

[65] A. Deshpande, J. Rock, and D. Forsyth, "Learning large-scale automatic image colorization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 567–575.

[66] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[67] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.

[68] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. Workshop Stat. Learn. Comput. Vis. (ECCV)*, May 2004, pp. 1–2.

[69] 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.

[70] L. Fei, Y. Xu, X. Fang, and J. Yang, "Low rank representation with adaptive distance penalty for semi-supervised subspace classification," *Pattern Recognit.*, vol. 67, pp. 252–262, Jul. 2017.

[71] Z. Ma, F. Nie, Y. Yang, J. R. R. Uijlings, N. Sebe, and A. G. Hauptmann, "Discriminating joint feature analysis for multimedia data understanding," *IEEE Trans. Multimedia*, vol. 14, no. 6, pp. 1662–1672, Dec. 2012.

[72] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2017, pp. 2408–2414.

[73] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

CAIXIA ZHENG was born in Gongzhuling, China. She received the B.S. and M.S. degrees from the College of Computer Science and Information Technology, Northeast Normal University, China, in 2009 and 2012, respectively, and the Ph.D. degree from the School of Mathematic and Statistic, Northeast Normal University, in 2016. From 2013 to 2014, she was a Visiting Ph.D. Student with the University of California, Davis, CA, USA. She is currently a Lecturer with the College of Information Science and Technology, Northeast Normal University. Her research interests include pattern recognition, machine learning, and image/video understanding.

JIANYU CHEN was born in Shuozhou, China. He received the B.S. degree from the College of Information Engineering, Inner Mongolia University of Technology, China, in 2016. He is currently pursuing the M.S. degree with the College of Information Science and Technology, Northeast Normal University. His research interests include pattern recognition, machine learning, and image/video understanding.

JUN KONG was born in Jilin, China. He received the B.S. and M.S. degrees from the Department of Mathematics, Northeast Normal University, China, in 1992 and 1997, respectively, and the Ph.D. degree from the College of Mathematics, Jilin University, in 2001. From 2003 to 2004, he was a Visiting Scholar with Edith Cowan University, Perth, WA, Australia. He is currently a Professor with the College of Information Science and Technology, Northeast Normal University. His research interests include artificial intelligence, digital image processing, pattern recognition, machine learning, biometrics, and information security.

YUGEN YI was born in Pingxiang, China. He received the B.S. degree from the College of Humanities and Sciences, Northeast Normal University, China, in 2009, the M.S. degree from the College of Computer Science and Information Technology, Northeast Normal University, in 2012, and the Ph.D. degree from the School of Mathematics and Statistics, Northeast Normal University, in 2015. He is currently a Lecturer with the School of Software, Jiangxi Normal University. His research interests include artificial intelligence, computer vision, and machine learning.

YINGHUA LU received the B.S. degree in computer science from Jilin University, the M.S. degree from Utsunomiya University, Japan, and the Ph.D. degree in computer science from Jilin University. He is currently the President of the College of Humanities and Sciences, Northeast Normal University, Jilin, China, where he is currently a Professor with the School of Information Science and Technology. His research interests include artificial intelligence, pattern recognition, and machine learning.

JIANZHONG WANG was born in Changchun, China. He received the B.S. degree from Computer School, Jilin University, China, in 2004, the M.S. degree from Computer School, Northeast Normal University, China, in 2007, and the Ph.D. degree from the School of Mathematics and Statistics, Northeast Normal University, in 2010. He is currently an Associate Professor with the College of Information Science and Technology, Northeast Normal University. His research interests include dimensionality reduction and image processing.

CHONG LIU is currently pursuing the B.S. degree with the School of Software, Jiangxi Normal University, Nanchang, China. His research interests include image processing and image classification.