Embedding Regularizer Learning for Multi-View Semi-Supervised Classification

Aiping Huang, *Member, IEEE*, Zheng Wang, Yannan Zheng, Tiesong Zhao, *Senior Member, IEEE*, and Chia-Wen Lin, *Fellow, IEEE*

***Abstract*— Classification remains challenging when confronted with the existence of multi-view data with limited labels. In this paper, we propose an embedding regularizer learning scheme for multi-view semi-supervised classification (ERL-MVSC). The proposed framework integrates diversity, sparsity and consensus to dexterously manipulate multi-view data with limited labels. To encourage diversity, ERL-MVSC recasts a linear regression model to derive view-specific embedding regularizers and automatically determines their weights. This is able to tactfully incorporate complementary information of different views. To ensure sparsity, ERL-MVSC imposes $\ell_{2,1}$-norm on a fused embedding regularizer to exploit the sparse local structure of samples, thereby conveying valuable classification information and enhancing the robustness against noise/outliers. To enhance consensus, ERL-MVSC learns a shared predicted label matrix, which serves as the comment target of multi-view classification. With these techniques, we formulate ERL-MVSC as a joint optimization problem of an embedding regularizer and a predicted label matrix, which can be solved by a coordinate descent method. Extensive experimental results on real-world datasets demonstrate the effectiveness and superiority of the proposed algorithm.**

***Index Terms*— Multi-view learning, multi-view semi-supervised classification, embedding regularizer.**

I. Introduction

MANY computer vision applications attempt to associate visual data with one or more semantic labels. This task can be achieved by classification which can establish an accurate correspondence between perceptual-level visual information and semantic-level linguistic descriptions. Recently, there have been a great number of classification algorithms related to object recognition [1], visual localization [2], image semantic segmentation [3] and retrieval [4]. To improve classification performance, many of them require to train the classifier on a large-scale annotated dataset. Nevertheless, data annotation is time-consuming and laborious, resulting in scarcity of labels in the collected data. The effective combination of unlabeled samples with labeled ones is therefore of critical importance [5]–[10].

Besides the scarcity of labeled samples, another important issue for many visual applications lies in the existence of multi-view data. For example, RGB video cameras, depth cameras and on-body sensors are often equipped together to offer different representations of visual data for pedestrians [11]. These data that describe the same instance with diverse modalities or features are called multi-view data. Each modality or feature representation is referred to as a view. Intuitively, simultaneous analysis of multi-view features facilitates utilizing their complementarity for disambiguation [12]. Previous research efforts [13]–[17] also show the superiority of multi-view feature fusion, which can surpass the performance achieved by using single-view features or simple concatenation of multi-view features.

The limited labels in multi-view data have motivated the research of multi-view semi-supervised classification [18]–[22] to exploit the latent information of both labeled and unlabeled data. A representative multi-view semi-supervised classification model is the co-training [23], which was originally designed for two views. It trains classifiers with the labeled data, and classifies the unlabeled data on each view independently. Then the most confidently predicted samples from one classifier are selected to train the other classifier in each iteration. Since the standard co-training algorithm was proposed, many variants have been devised [24]–[26]. However, co-training and its variants require accurate classification results in each view. Once erroneous information from either of the classifiers is provided, the overall performance will be deteriorated.

Recently, graph-based multi-view semi-supervised classification techniques have attracted increasing attention [27]–[29]. They treat labeled and unlabeled samples as vertices of a graph and propagate label information through edges. The local structure in the sample space is one of the important elements affecting classification performance, and it generally requires graph construction to embody. Therefore, some state-of-the-art

Manuscript received December 24, 2020; revised June 6, 2021 and July 21, 2021; accepted July 24, 2021. Date of current version August 9, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61901119 and Grant 62001116 and in part by the Natural Science Foundation of Fujian Province under Grant 2019J01222 and Grant 2020J01466. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Clinton Fookes. *(Corresponding author: Tiesong Zhao.)*

Aiping Huang, Zheng Wang, and Yannan Zheng are with the Fujian Key Laboratory for Intelligent Processing and Wireless Transmission of Media Information, College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China (e-mail: sxxhap@163.com; N191120078@fzu.edu.cn; N191127059@fzu.edu.cn).

Tiesong Zhao is with the Fujian Key Laboratory for Intelligent Processing and Wireless Transmission of Media Information, College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China, and also with the Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: t.zhao@fzu.edu.cn).

Chia-Wen Lin is with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan, and also with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: cwlin@ee.nthu.edu.tw).

Digital Object Identifier 10.1109/TIP.2021.3101917

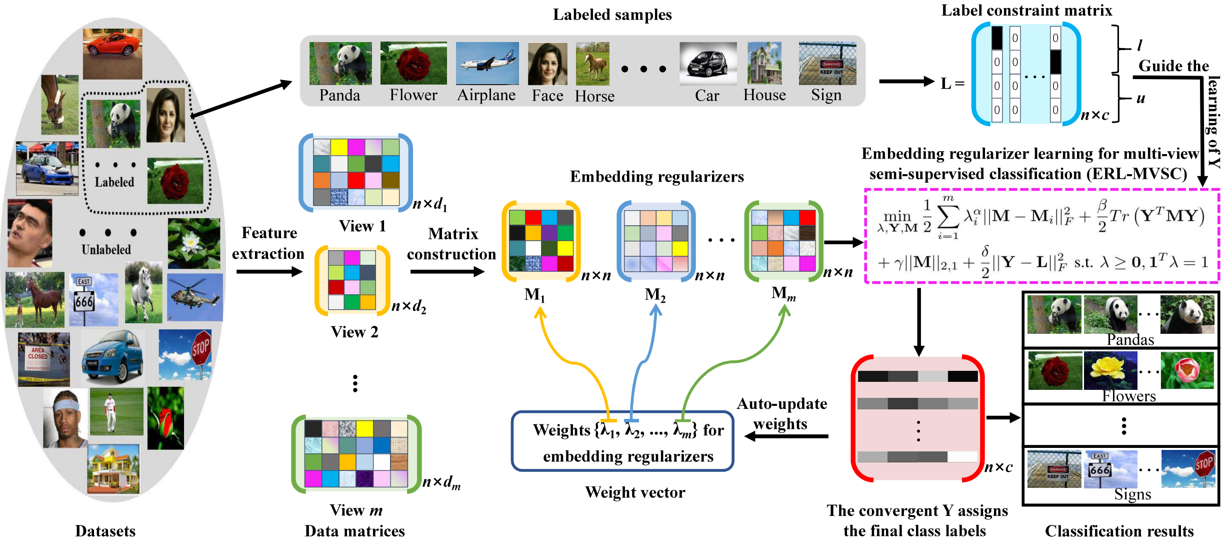

Labeled samples
Panda
Flower
Airplane
Face
Horse
Car
House
Sign
Label constraint matrix
$\mathbf{L} =$
0
$n\times c$
l
u
Guide the learning of Y
Embedding regularizer learning for multi-view semi-supervised classification (ERL-MVSC)
$\min_{\lambda,\mathbf{Y},\mathbf{M}} \frac{1}{2}\sum_{i=1}^{m}\lambda_i^{\alpha}||\mathbf{M}-\mathbf{M}_i||_F^2 + \frac{\beta}{2}Tr\left(\mathbf{Y}^T\mathbf{M}\mathbf{Y}\right)$
$+\gamma||\mathbf{M}||_{2,1} + \frac{\delta}{2}||\mathbf{Y}-\mathbf{L}||_F^2$ s.t. $\lambda \geq \mathbf{0}, \mathbf{1}^T\lambda = 1$
Labeled
Unlabeled
Feature extraction
View 1
$n\times d_1$
View 2
$n\times d_2$
View m
$n\times d_m$
Data matrices
Matrix construction
Embedding regularizers
$\mathbf{M}_1$
$\mathbf{M}_2$
$\mathbf{M}_m$
$n\times n$
Weights $\{\lambda_1, \lambda_2, ..., \lambda_m\}$ for embedding regularizers
Weight vector
Auto-update weights
$n\times c$
The convergent Y assigns the final class labels
Pandas
Flowers
Signs
Classification results
Datasets

Fig. 1. Framework of the proposed method. We firstly extract feature matrices of data from m views and compute view-specific embedding regularizers from them. With a sparsity constraint and label guidance, the predicted label matrix $\mathbf{Y}$ is obtained by the proposed algorithm.

methods coupled with manifold learning have been proposed, including adaptive neighbors based [30] and hyper-Laplacian regularized [31] multi-view semi-supervised classification. In these methods, both the way of construction and the noise in raw data might impair the constructed graph, resulting in unstable classification performance.

In this paper, we propose a brand-new method, named embedding regularizer learning for multi-view semi-supervised classification (ERL-MVSC). As outlined in Fig. 1, the whole procedure includes three aspects. Firstly, a traditional linear regression model is recast to derive view-specific classifiers and regularizer constructors. Secondly, with the embedding regularizers of different views as input, the derived classifier is applied in the multi-view scenario to learn a shared embedding regularizer for label propagation. To convey valuable classification information, the learned embedding regularizer is constrained to be sparse to exploit important local structure information. Finally, to incorporate the label information as learning guidance, a constraint matrix is introduced into our framework. The main contributions of this paper are summarized as follows:

- Embedding regularizer learning derived from a linear regression model: A traditional linear regression model is recast to derive an embedding regularizer learning framework that integrates diversity, sparsity and consensus for multi-view semi-supervised classification.
- Sparse representation for the fused embedding regularizer: The sparse representation is automatically updated with $\ell_{2,1}$-norm minimization. It aims to capture the sparse local structure of multi-view data that conveys useful classification clues, while simultaneously improving the robustness against noise and outliers.
- Mathematical solutions for the proposed formulation: We employ the coordinate descent method to factorize our objective function into several small sub-problems to efficiently attain their solutions. Experimental results validate the superior performance of our solutions.

The rest of this paper is arranged as follows. Section II reviews the related work to this paper. In Section III, the proposed framework with optimization algorithm is developed. Section IV presents extensive experiments and analysis on real-world datasets. Conclusions are drawn in Section V.

II. Related Work

Due to the existence of multi-view data and the scarcity of labeled samples, multi-view semi-supervised classification has become an imperative and challenging research topic. Existing algorithms were mainly developed in two paradigms based on co-training and graph.

The principle of co-training was first introduced in [23] for semi-supervised learning. Since then, several successful variants of co-training have been proposed. For example, the method proposed in [24] embeds an EM algorithm into the co-training procedure for parameter estimation. A Bayesian undirected graphical model was proposed in [25] to clarify the fundamental assumptions of co-training related algorithms. The method in [26] resorts to a co-regularization term to minimize the distinction between the predictor functions of two views. Recently, co-training methods coupled with deep networks have been proposed for semi-supervised classification [32], [33]. These methods utilize two different-view features (RGB and depth in [32]; image and text in [33]) to respectively learn a deep neural network, then apply the learned network in the unlabeled pool to iteratively generate training sets for each other. Generally, the co-training-based approaches address the scarcity of labeled samples with two views' classifiers interacting with each other to augment the training data. However, the process is time-consuming, and the final classification accuracy can be easily degraded by misleading pseudo labels. To address the problems, [34] runs Graph Convolutional Network (GCN) with a weighted combination of Laplacians in each view then aggregates these two GCNs to infer labels.

Graph-based method is another representative paradigm for multi-view semi-supervised classification. Reference [35] proposed a graph-based semi-supervised learning that adaptively fuses different views to learn a shared class indicator matrix for image categorization. Reference [36] utilizes sparse weights to learn an optimally fused graph for label propagation. To handle noise and outlying entries embedded in data, [30] adaptively learns local manifold structures to obtain an optimal graph for classification. Reference [28] proposed a multiple kernel-based framework to simultaneously perform classification and similarity learning. Reference [29] integrates deep constrained matrix factorization and low-rank similarity learning into a unified objective function for classification. Reference [31] utilizes hyper-graph regularization to design a parameter-free semi-supervised learning framework for addressing the data classification in multiple nonlinear subspaces. Reference [37] proposed a new structural regularization term to learn a unified graph that is more suitable for semi-supervised learning. Although many graph-based semi-supervised classification methods have been proposed, their constructed graphs could be easily affected by the way of construction and the quality of data, resulting in unstable performance. Meanwhile, due to the computational costs of graph construction and label propagation, these methods are difficult to scale to large data.

As a result, some other methods such as subspace-based [38] and regression-based methods [39]–[42] have been developed. [38] proposed a tensorized multi-view subspace representation learning, where a low-rank tensor was employed to explore high-order multi-view correlations and a constraint matrix was devised to guide the representation learning. Reference [39] employs a statistical approach and hierarchical regression to infer a reliable classifier for multimedia analysis. Reference [40] regresses to label matrix directly by formulating the objective function as a linear weighted combination of all regression-based loss functions. Reference [41] employs a discriminative regression target and a set of learnable weights to formulate a regression-based framework for multi-view classification. Reference [42] proposed a probabilistic square hinge loss and a power mean incorporation strategy to exploit both consensus and diversity information of multi-view data for semi-supervised classification. This paper also resorts to a regression model but with the following differences: (i) employing embedding regularizers as inputs of the model, (ii) recasting a linear regression model to provide important modules for efficient multi-view semi-supervised classification, and (iii) integrating the derived important modules, diversity, sparsity and consensus into a unified cost function as the classification loss, that learns a predicted label matrix rather than a transformation matrix and a bias vector.

III. Multi-View Embedding Regularizer for Semi-Supervised Classification

In this section, we elaborate the formulation and optimization of ERL-MVSC. To facilitate the understanding of mathematical derivations, the key notations used in this paper are summarized in Table I.

A. Problem Formulation

TABLE I

Summary of Key Notations Used in the Paper

Notations	Explanations
$\mathbf{X}^{(i)} \in \mathbb{R}^{n \times d_i}$	the data matrix of the ith view
$\{\mathbf{X}^{(i)}\}_{i=1}^{m}$	data matrices of m views
$\{\mathbf{M}_i\}_{i=1}^{m}$	embedding regularizers of m views
$\mathbf{M} \in \mathbb{R}^{n \times n}$	the fused embedding regularizer
$\mathbf{Y} \in \mathbb{R}^{n \times c}$	the predicted label matrix
$\mathbf{L} \in \{0,1\}^{n \times c}$	the label constraint matrix
$\lambda \in \mathbb{R}^{m}$	the weight vector
$n/m/c$	the number of samples/views/classes
d_i	the dimensionality of the ith view
l/u	the number of labeled/unlabeled data
α	the smoothing factor
β	the embedding parameter
γ	the regularization parameter
δ	the fitting coefficient

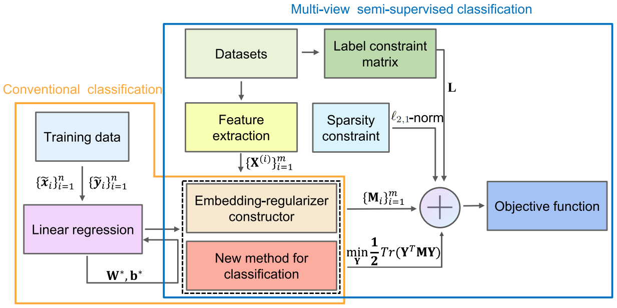

Multi-view semi-supervised classification
Conventional classification
Datasets
Label constraint matrix
L
Feature extraction
Sparsity constraint
$\ell_{2,1}$-norm
Training data
$\{\mathbf{X}^{(i)}\}_{i=1}^{m}$
$\{\widetilde{\mathbf{x}}_i\}_{i=1}^{n}$
$\{\widetilde{\mathbf{y}}_i\}_{i=1}^{n}$
Embedding-regularizer constructor
$\{\mathbf{M}_i\}_{i=1}^{m}$
Objective function
Linear regression
New method for classification
$\min_{\mathbf{Y}} \frac{1}{2} Tr(\mathbf{Y}^T\mathbf{MY})$
$\mathbf{W}^*, \mathbf{b}^*$

Fig. 2. Construction of objective function. We start with conventional classification, recast a widely-used linear regression model (*i.e.*, (1)) to derive two important modules (*i.e.*, (6) and (7)) for multi-view semi-supervised classification. Incorporating label information and $\ell_{2,1}$-norm regularization, an objective function (*i.e.*, (9)) with diversity, sparsity and consensus for multi-view semi-supervised classification is formulated.

In multi-view semi-supervised classification, the local structure information in the sample space is more capable of reflecting the relationships among samples. To this end, lots of schemes coupled with manifold learning have been proposed with graph construction to embed the local structure information. However, a constructed graph can be easily impaired by the ways of construction and the noise embedded in raw data [40]. To avoid this, we attempt to exploit the local structure information from a brand-new viewpoint. As demonstrated in Fig. 2, we recast a widely-used linear regression model to derive an embedding regularizer constructor and a new classifier, then employ these two modules coupled with the sparsity constraint and label guidance to derive a multi-view semi-supervised classification model. The sparsity constraint imposed here is to excavate important local relationships among samples, so as to convey valuable information for classification.

Linear regression models are among the representative methods for classification [43], [44]. Given a set $\{\widetilde{\mathbf{x}}_i\}_{i=1}^{n} \subset \mathbb{R}^d$ of sample points and a destination set $\{\widetilde{\mathbf{y}}_i\}_{i=1}^{n} \subset \mathbb{R}^c$, $\widetilde{\mathbf{y}}_i$ is assigned as $+1/-1$ for two-class problems or a class label vector for multi-class problems. A linear regression model aims to regress each sample to its label vector through computing c transformation vectors and bias constants which are respectively denoted as $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_c] \in \mathbb{R}^{d \times c}$ and $\mathbf{b} = [b_1, b_2, \cdots, b_c]^T \in \mathbb{R}^c$. To avoid over-fitting, a regularization term is added. One of the most widely-used linear regression models is

$$\min_{\mathbf{W},\mathbf{b}} \sum_{i=1}^{n} ||\widetilde{\mathbf{y}}_i - \mathbf{W}^T \widetilde{\mathbf{x}}_i - \mathbf{b}||_2^2 + \theta ||\mathbf{W}||_F^2, \tag{1}$$

where θ is a regularization parameter, $||\cdot||_2$ is the Euclidean norm and $||\cdot||_F$ denotes the Frobenius norm. Denoting $\widetilde{\mathbf{X}} = [\widetilde{\mathbf{x}}_1, \cdots, \widetilde{\mathbf{x}}_n]^T \in \mathbb{R}^{n\times d}$ and $\widetilde{\mathbf{Y}} = [\widetilde{\mathbf{y}}_1, \cdots, \widetilde{\mathbf{y}}_n]^T \in \mathbb{R}^{n\times c}$, the above optimization problem can be rewritten as

$$\min_{\mathbf{W},\mathbf{b}} \frac{1}{2} ||\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\mathbf{W} - \mathbf{1}\mathbf{b}^T||_F^2 + \frac{\theta}{2} ||\mathbf{W}||_F^2, \tag{2}$$

where $\mathbf{1}$ is a n-dimensional column vector with all elements being 1. We know that (2) has the following optimal closed-form solutions:

$$\mathbf{W}^* = (\widetilde{\mathbf{X}}^T \mathbf{H} \widetilde{\mathbf{X}} + \theta \mathbf{I})^{-1} \widetilde{\mathbf{X}}^T \mathbf{H} \widetilde{\mathbf{Y}}, \tag{3}$$

$$\mathbf{b}^* = \frac{1}{n} \sum_{i=1}^{n} \left(\widetilde{\mathbf{y}}_i - \mathbf{W}^T \widetilde{\mathbf{x}}_i \right), \tag{4}$$

where $\mathbf{I}$ is an identity matrix and $\mathbf{H} \in \mathbb{R}^{n\times n}$ is a centering matrix, *i.e.*, $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$. Plugging $\mathbf{W}^*$ and $\mathbf{b}^*$ into (1), the optimal objective function can be represented as

$$\frac{1}{2} \sum_{i=1}^{n} \widetilde{\mathbf{y}}_i^T \widetilde{\mathbf{M}} \widetilde{\mathbf{y}}_i \tag{5}$$

with positive semi-definite matrix

$$\widetilde{\mathbf{M}} = \mathbf{H}(\mathbf{I} - \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^T \mathbf{H} \widetilde{\mathbf{X}} + \theta \mathbf{I})^{-1} \widetilde{\mathbf{X}}^T)\mathbf{H}. \tag{6}$$

Without loss of generality, the objective function of linear regression for classification in (2) can be recast as

$$\min_{\widetilde{\mathbf{Y}}} \frac{1}{2} Tr(\widetilde{\mathbf{Y}}^T \widetilde{\mathbf{M}} \widetilde{\mathbf{Y}}), \tag{7}$$

where the positive semi-definite matrix $\widetilde{\mathbf{M}}$ serves as an embedding regularizer. Obviously, the above formulation is solvable. Hereto, we provide a new method for classification. Given a data matrix, we need to calculate $\widetilde{\mathbf{M}}$ by (6), and optimize (7), then the ultimate class labels are predicted. Next, we will extend this classification model to design an embedding regularizer learning framework for multi-view semi-supervised classification.

In the multi-view setting, one sample is represented by several different views. Suppose $\{\mathbf{X}^{(i)}\}_{i=1}^m$ is a multi-view dataset with m views, where $\mathbf{X}^{(i)} \in \mathbb{R}^{n\times d_i}$. For semi-supervised learning, a small amount of labeled samples are available. We rearrange all the samples and let the first l $(0 < l \ll n)$ ones be labeled. The corresponding set of class label vectors is denoted as $\{l_i\}_{i=1}^l \subset \mathbb{R}^c$. Multi-view semi-supervised classification aims to integrate multi-view features to predict the class labels for unlabeled data with limited label information. In other words, it aims to obtain a predicted label matrix $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n]^T \in \mathbb{R}^{n\times c}$ from the multi-view feature matrices $\{\mathbf{X}^{(i)}\}_{i=1}^m$ and the label information $\{l_i\}_{i=1}^l$. Extending (7) to the multi-view case, ERL-MVSC can be formulated as

$$\min_{\lambda,\mathbf{Y},\mathbf{M}} \frac{1}{2} \sum_{i=1}^{m} \lambda_i^\alpha ||\mathbf{M} - \mathbf{M}_i||_F^2 + \frac{\beta}{2} Tr\left(\mathbf{Y}^T \mathbf{M} \mathbf{Y}\right) + \gamma ||\mathbf{M}||_{2,1} + \frac{\delta}{2} \sum_{i=1}^{l} ||\mathbf{y}_i - l_i||_F^2 \ \text{s.t.}\ \lambda \geq \mathbf{0}, \mathbf{1}^T \lambda = 1, \tag{8}$$

where $\mathbf{M}$ is a shared embedding regularizer to be learned, $\mathbf{M}_i$ is constructed from $\mathbf{X}^{(i)}$ by (6), $\lambda = [\lambda_1, \cdots, \lambda_m]^T$ is a weighted vector to measure the importance of different views, $||\cdot||_{2,1}$ denotes the $\ell_{2,1}$-norm, $\alpha > 1$ is a smoothing factor, and β, γ, δ are three nonnegative hyper-parameters. However, (8) can not be directly optimized with a closed-form solution of $\mathbf{Y}$ due to variable coupling. For the sake of facilitating optimization and balancing the importance of each term in (8), we add a smooth regularization term for unlabeled data, and further formulate ERL-MVSC as

$$\min_{\lambda,\mathbf{Y},\mathbf{M}} \frac{1}{2} \sum_{i=1}^{m} \lambda_i^\alpha ||\mathbf{M} - \mathbf{M}_i||_F^2 + \frac{\beta}{2} Tr\left(\mathbf{Y}^T \mathbf{M} \mathbf{Y}\right) + \gamma ||\mathbf{M}||_{2,1} + \frac{\delta}{2} ||\mathbf{Y} - \mathbf{L}||_F^2 \ \text{s.t.}\ \lambda \geq \mathbf{0}, \mathbf{1}^T \lambda = 1, \tag{9}$$

where $\mathbf{L} = [l_1, l_2, \cdots, l_n]^T \in \{0, 1\}^{n\times c}$ is a label constraint matrix. For labeled data, $\mathbf{L}_{ij} = 1$ if the ith sample belongs to the jth class, and 0 otherwise; While for unlabeled data, l_i $(i > l)$ is set to a zero vector.

There are four terms in (9). The first term ensures to learn an embedding regularizer shared by different views with a minimal fitting error. The second term extends (7) to the multi-view scenario for regressing each sample to its class label. The third term aims to learn a sparse embedding regularizer $\mathbf{M}$. Related research [45] on manifold learning revealed that a sparse matrix characterizing locality relationships conveys valuable information for classification [46]. In this work, $\mathbf{M}$ describes the relevances between samples by multi-view features. We impose a sparsity constraint on $\mathbf{M}$ to remove those relevances that are not discriminative for given samples. The obtained sparse $\mathbf{M}$ encodes important local relationships among samples, conveying valuable information for classification and enhancing the robustness to noise and outliers. Herein, $\ell_{2,1}$-norm rather than ℓ_1-norm is adopted as the sparsity regularizer for the following two reasons. First, different from the flat sparsity of ℓ_1-norm, the $\ell_{2,1}$-norm regularization involves structured sparsity [47]. This type of sparsity can greatly improve the efficacy of sparse learning algorithms through encoding specific structured information [48]. Second, the $\ell_{2,1}$-norm regularized minimization problem can be solved by the method in [49], which is much easier than that with the ℓ_1-norm regularization. The last term is to ensure the smoothness for unlabeled data and to leverage the labeled information to guide the prediction of $\mathbf{Y}$. The weighted vector λ can be regarded as a regularization parameter to the multi-objective optimization problem, where $\lambda \geq \mathbf{0}$ and $\lambda^T \mathbf{1} = 1$.

B. Optimization

In this subsection, we employ the coordinate descent method to solve the above optimization problem. Denote

$\mathcal{J}(\lambda, \mathbf{Y}, \mathbf{M}) \triangleq \frac{1}{2}\sum_{i=1}^{m} \lambda_i^{\alpha} ||\mathbf{M} - \mathbf{M}_i||_F^2 + \frac{\beta}{2} Tr\left(\mathbf{Y}^T \mathbf{M} \mathbf{Y}\right) + \gamma ||\mathbf{M}||_{2,1} + \frac{\delta}{2}||\mathbf{Y} - \mathbf{L}||_F^2$. The objective function for ERL-MVSC can be optimized with the following iterative framework:

$$\begin{aligned} \lambda^{(t+1)} &= \arg\min_{\lambda} \mathcal{J}(\lambda, \mathbf{Y}^{(t)}, \mathbf{M}^{(t)}), \\ \mathbf{Y}^{(t+1)} &= \arg\min_{\mathbf{Y}} \mathcal{J}(\lambda^{(t+1)}, \mathbf{Y}, \mathbf{M}^{(t)}), \\ \mathbf{M}^{(t+1)} &= \arg\min_{\mathbf{M}} \mathcal{J}(\lambda^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{M}). \end{aligned} \tag{10}$$

Update λ when fixing Y and M. Eliminating the constant terms of $\mathcal{J}(\lambda, \mathbf{Y}, \mathbf{M})$, the subproblem of updating λ is written as

$$\min_{\lambda} \frac{1}{2}\sum_{i=1}^{m} \lambda_i^{\alpha} ||\mathbf{M} - \mathbf{M}_i||_F^2 \text{ s.t. } \lambda \geq \mathbf{0}, \mathbf{1}^T\lambda = 1. \tag{11}$$

Utilizing the Lagrange multiplier method, we construct a Lagrangian function as

$$\mathcal{L}(\lambda, \mu) = \frac{1}{2}\sum_{i=1}^{m} \lambda_i^{\alpha} ||\mathbf{M} - \mathbf{M}_i||_F^2 + \mu(1 - \mathbf{1}^T\lambda). \tag{12}$$

Taking the derivative of $\mathcal{L}(\lambda, \mu)$ with respect to λ and μ,

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = \frac{\alpha}{2}\lambda_i^{\alpha-1}||\mathbf{M} - \mathbf{M}_i||_F^2 - \mu, \frac{\partial \mathcal{L}}{\partial \mu} = 1 - \mathbf{1}^T\lambda. \tag{13}$$

Setting $\frac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{0}$ and $\frac{\partial \mathcal{L}}{\partial \mu} = 0$, we have

$$\lambda_i = \frac{(||\mathbf{M} - \mathbf{M}_i||_F^2)^{1/(1-\alpha)}}{\sum_{i=1}^{m}(||\mathbf{M} - \mathbf{M}_i||_F^2)^{1/(1-\alpha)}}. \tag{14}$$

Update M when fixing λ and Y. The subproblem of updating **M** aims to minimize the objective

$$\frac{1}{2}\sum_{i=1}^{m} \lambda_i^{\alpha} ||\mathbf{M} - \mathbf{M}_i||_F^2 + \frac{\beta}{2} Tr\left(\mathbf{Y}^T\mathbf{M}\mathbf{Y}\right) + \gamma ||\mathbf{M}||_{2,1}. \tag{15}$$

According to [49], the above problem can be addressed by solving

$$\begin{aligned} \min_{\mathbf{M}} \mathcal{J}(\lambda, \mathbf{Y}, \mathbf{M}) = \min_{\mathbf{M}} \frac{1}{2}\sum_{i=1}^{m} \lambda_i^{\alpha} ||\mathbf{M} - \mathbf{M}_i||_F^2 \\ + \frac{\beta}{2} Tr\left(\mathbf{Y}^T\mathbf{M}\mathbf{Y}\right) + \gamma\, Tr(\mathbf{M}^T\mathbf{D}\mathbf{M}), \end{aligned} \tag{16}$$

where $\mathbf{D} \in \mathbb{R}^{n\times n}$ is a diagonal matrix, defined by

$$\mathbf{D} = \begin{bmatrix} \frac{1}{2||\mathbf{m}_1||_2} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{2||\mathbf{m}_n||_2} \end{bmatrix} \tag{17}$$

with $\mathbf{M} = [\mathbf{m}_1; \cdots; \mathbf{m}_n]$. Taking the derivative of $\mathcal{J}(\lambda, \mathbf{Y}, \mathbf{M})$ with respect to **M**, we obtain

$$\frac{\partial \mathcal{J}}{\partial \mathbf{M}} = \sum_{i=1}^{m} \lambda_i^{\alpha}(\mathbf{M} - \mathbf{M}_i) + \frac{\beta}{2}\mathbf{Y}\mathbf{Y}^T + 2\gamma\,\mathbf{D}\mathbf{M}. \tag{18}$$

Setting $\frac{\partial \mathcal{J}}{\partial \mathbf{M}} = \mathbf{0}$, we have the closed-form solution

$$\mathbf{M} = \left(\sum_{i=1}^{m} \lambda_i^{\alpha}\mathbf{I} + 2\gamma\,\mathbf{D}\right)^{-1}\left(\sum_{i=1}^{m} \lambda_i^{\alpha}\mathbf{M}_i - \frac{\beta}{2}\mathbf{Y}\mathbf{Y}^T\right). \tag{19}$$

Update Y when fixing λ and M. The subproblem of updating **Y** is formulated as

$$\min_{\mathbf{Y}} \mathcal{J}(\lambda, \mathbf{Y}, \mathbf{M}) = \min_{\mathbf{Y}} \frac{\beta}{2} Tr\left(\mathbf{Y}^T\mathbf{M}\mathbf{Y}\right) + \frac{\delta}{2}||\mathbf{Y} - \mathbf{L}||_F^2. \tag{20}$$

Setting the derivative of $\mathcal{J}(\lambda, \mathbf{Y}, \mathbf{M})$ with respect to **Y** to zero, *i.e.*,

$$\frac{\partial \mathcal{J}}{\partial \mathbf{Y}} = \frac{\beta}{2}(\mathbf{M} + \mathbf{M}^T)\mathbf{Y} + \delta(\mathbf{Y} - \mathbf{L}) = \mathbf{0}, \tag{21}$$

from which it infers the following closed form solution

$$\mathbf{Y} = \left(\frac{\beta}{2\delta}(\mathbf{M} + \mathbf{M}^T) + \mathbf{I}\right)^{-1}\mathbf{L}. \tag{22}$$

Once the predicted label matrix **Y** is attained, the ultimate class label y_i for unlabeled data x_i can be calculated by the following decision function:

$$\begin{aligned} y_i &= \arg\max_{j} \mathbf{Y}_{ij}. \\ \forall i &= l+1, l+2, \cdots, n. \quad \forall j = 1, 2, \cdots, c. \end{aligned} \tag{23}$$

Summarizing the aforementioned analysis of respective optimal solutions to all subproblems, the algorithm for ERL-MVSC is presented in Algorithm 1.

Algorithm 1 The Algorithm of ERL-MVSC

```
Input:
  The data matrix: $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \cdots, \mathbf{X}^{(m)}$.
  The label constraint matrix: $\mathbf{L} \in \{0, 1\}^{n\times c}$.
  The parameters: smoothing factor $\alpha$, hyper-parameters
  $\beta$, $\gamma$ and $\delta$.
Output:
  The class label for unlabeled data.
  The weight vector $\lambda = [\lambda_1, \cdots, \lambda_m]^T$.
Initialization:
  Compute each embedding regularizer $\mathbf{M}_i$ from $\mathbf{X}^{(i)}$
  using (6).
  Initialize $\lambda = [\frac{1}{m}, \cdots, \frac{1}{m}]^T$, $\mathbf{Y} = \mathbf{L}$ and $\mathbf{M} =
  \frac{1}{m}\sum_{i=1}^{m}\mathbf{M}_i$.
Procedure:
  while Not convergent do
    Update $\mathbf{Y}$ using (22);
    Update $\mathbf{D}$ using (17);
    Update $\mathbf{M}$ using (19);
    Update $\lambda$ using (14);
  end while
  Obtain the ultimate class labels by using (23);
  return The class label for unlabeled data and the
  weight vector.
```

ERL-MVSC converges for the following two reasons. Firstly, we decompose the proposed multi-variable optimization problem by the convergent coordinate descent method. Secondly, we alternately get the closed-form solutions of λ, $\mathbf{Y}$ and $\mathbf{M}$ at each iteration. Besides, the experimental evidence on real data also shows a good convergence behavior.

The proposed algorithm also consumes a reasonable computational complexity, which mainly depends on the initialization of $\mathbf{M}_i$ and update of $\mathbf{Y}$ and $\mathbf{M}$. These three parts respectively contribute $\mathcal{O}(d_i n^2 + d_i^2 n + d_i^3)$, $\mathcal{O}(n^3 + n^2 c)$ and $\mathcal{O}(n^2 c)$ to the overall complexity. Because the number c of classification is much lower than d_i and n, the overall computational cost of ERL-MVSC is $\mathcal{O}(n^3 + \sum_{i=1}^{m}(d_i n^2 + d_i^2 n + d_i^3))$.

IV. Experiments

A. Test Datasets

To evaluate the proposed method, the experiments are performed on several real-world benchmark datasets. These datasets cover four different applications, including generic object classification, digit classification, video classification, and news article classification.

ALOI[1] is a collection of color images with small objects, which were taken with various viewing angles, illumination directions, illumination colors, and object orientations. For each image, RGB color histograms, HSV color histograms, color similarity and Haralick texture features are extracted as four views.

Caltech101[2] is a dataset containing 101 categories with 9,144 samples, where six different features are extracted as views, *i.e.*, 48-D Gabor features, 40-D wavelet moments features, 254-D CENTRIST features, 1,984-D HOG features, 512-D GIST features, and 928-D local binary pattern (LBP) features.

HW[3] consists of handwritten numerals from 0 to 9 digit classes with total 2,000 samples. There are six public features are available, including 76-D Fourier coefficients of the character shapes (FOU), 216-D profile correlations (FAC), 64-D Karhunen-Loѐve coefficients (KAR), 240-D pixel averages in 2 × 3 windows (PIX), 47-D Zernike moment (ZER) and 6-D morphological (MOR) features.

MNIST[4] is a handwritten digit collection from which 30,000 samples are selected for testing, along with three views of features produced by IsoProjection with 30 dimensions, linear discriminant analysis with 9 dimensions and neighborhood preserving embedding with 9 dimensions.

YouTube[5] contains approximately 120,000 instances and we select 2,000 samples for test. Each sample is described by six views consisting of both audio features (mfcc, volume stream, and spectrogram stream) and visual features (cuboids histogram, hist motion estimate, and HOG features).

3sources[6] consists of 169 news reported by 3 well-known online news sources: BBC, Reuters and the Guardian which are treated as views. Each new is manually annotated with one of 6 topical labels.

BBCnews[7] is derived from the BBC news corpora [50]. It contains a total of 685 documents with 5 annotated topic labels. In our experiments, we construct new synthetic datasets with 4 views by splitting each document into four related segments with 4,659, 4,633, 4,665 and 4,684 dimensions, respectively.

BBCsports[7] consists of 544 documents from the BBC sport website involving the sports news in 5 topical areas [50]. For each document, two different types of features are extracted as views with 3,183 and 3,203 dimensions, respectively.

The important statistics of the above datasets are summarized in Table II. Several sample images from the selected three image datasets are demonstrated in Fig. 3.

B. Compared Algorithms and Parameter Settings

To evaluate the effectiveness of ERL-MVSC, we compare the method with several state-of-the-art multi-view semi-supervised classification approaches. The brief introduction and parameter setting of each method are presented as follows:

1) **SVM** utilizes different types of kernel functions to project nonlinear separable samples onto a high dimensional space for classification.
2) **AMGL** [27] is a parameter-free multi-view learning method based on spectral clustering and can be extended to semi-supervised classification tasks.
3) **MVAR** [40] is an adaptive regression based multi-view semi-supervised model, where class labels can be directly predicted by non-smooth $\ell_{2,1}$-norm minimization.
4) **MLAN** [30] automatically calculates view weights and simultaneously performs semi-supervised classification and local structure learning.
5) **AWDR** [41] employs a discriminative regression target and a set of learnable weights to formulate a regression-based framework for multi-view classification.
6) **Co-GCN** [34] integrates co-training, spectral graph information and the expressive power of neural network into one unified framework for multi-view semi-supervised learning.
7) **HLR-M^{2}VS** [31] is a hyper-laplacian regularized multi-linear multi-view self-representation model. It can capture both the global and local structures of data for semi-supervised classification.

Among the above methods, the first one is a single-view baseline, while the rest ones are the state-of-the-art methods in multi-view semi-supervised learning. Prior to conduct comparison, there are several algorithmic parameters need to be set. For the proposed ERL-MVSC, we set the smoothing factor $\alpha = 2$, the embedding parameter $\beta = 1$, the regularization parameter $\gamma = 1$, and the fitting coefficient $\delta = 10$. The way

[1] http://aloi.science.uva.nl/

[2] http://www.vision.caltech.edu/Image Datasets/Caltech101/

[3] http://archive.ics.uci.edu/ml/datasets/Multiple+Features

[4] http://yann.lecun.com/exdb/mnist/

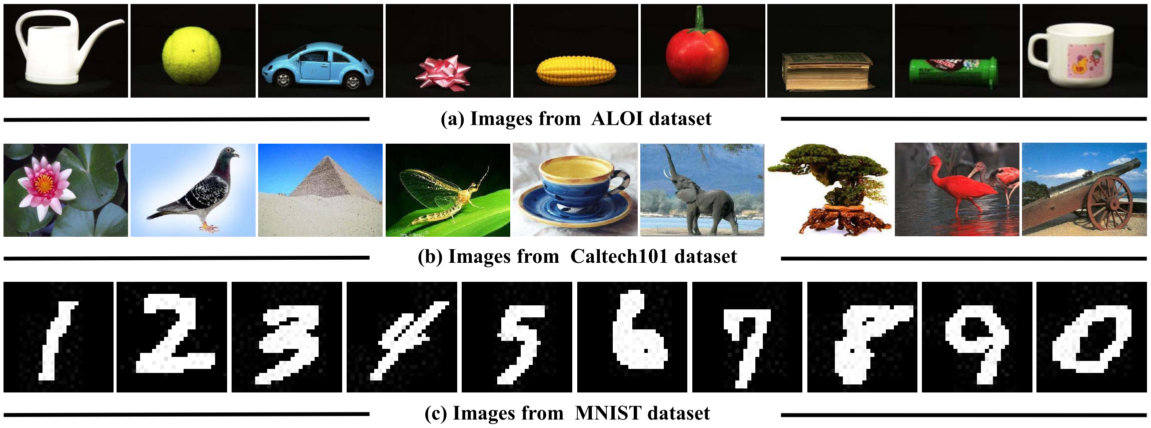
[5] http://archive.ics.uci.edu/ml/datasets/YouTube+Multiview+Video+Games+Dataset

[6] http://mlg.ucd.ie/datasets/3sources.html

[7] http://mlg.ucd.ie/datasets/segment.html

TABLE II
STATISTICAL CHARACTERISTICS OF ALL TESTING DATASETS

Datasets	Samples	Views	Total features	Classes	Data types
ALOI	1079	4	218	10	Object image
Caltech101	9,144	6	3,766	101	Object image
HW	2,000	6	649	10	Digit image
MNIST	30,000	3	48	10	Digit image
YouTube	2,000	6	1,589	10	Video data
3sources	169	3	10,259	6	Text document
BBCnews	685	4	18,641	5	Text document
BBCsports	544	2	6,386	5	Text document


(a) Images from ALOI dataset

(b) Images from Caltech101 dataset

(c) Images from MNIST dataset

Fig. 3. Sample images from three datasets.

to determine these parameters will be expounded in detail in Section IV-E.3. As for the compared algorithms, we directly utilize the source codes provided by the authors. For fair comparison, we adopt the parameters of each method just as reported in the corresponding papers if feasible. For example, the maximum number of iterations for AMGL and MLAN are respectively fixed as 100 and 30, and the number of nearest neighbors are respectively tuned as 5 and 9. As to MVAR, the weight distribution coefficient r is set as 2. For AWDR, the trade-off parameter $\lambda = 1$, the termination parameter $\epsilon = 10^3$, and the maximum number of iterations $t = 200$. Regarding Co-GCN, the maximum number of training iterations is set to be 2500, and the learning rate $\alpha_{\mathbf{W}} = 10^3$ and $\alpha_\pi = 10^2$. Note, HLR-M^{2}VS requires different parameter settings in different datasets. However, due to the addition of new test datasets, we set each parameter as a uniform fixed value, *i.e.*, $\lambda_1 = 0.2$ and $\lambda_2 = 0.4$, respectively. For single-view baseline, we concatenate feature vectors of different views together for the all-view classification setting.

C. Evaluation Metrics

We evaluate the performance of individual algorithm in term of classification accuracy, *i.e.*, the proportion of correctly classified samples. The purpose of semi-supervised learning is to excavate more unlabeled information from limited labeled samples. To present the classification capacity of different algorithms, we set the percentage of labeled samples ranging from 10% to 80% for training. All the experiments are performed on a server with Inter (R) Xeon (R) E5-2680 CPU (2.40GHz) and 256G memory. For each dataset, we execute each algorithm 10 times with randomly selected labeled data, and then record the average classification accuracy and its standard deviation corresponding to each labeling rate.

D. Experimental Results

Tables III∼VI present the classification accuracies with standard deviations of all methods under different labeling rates on eight datasets. For ease of comparison, we highlight the best results in red boldface and the second best in blue boldface. Through the analysis of experiment results, some conclusions are drawn.

1) Superiority of ERL-MVSC over single-view baseline: Not all multi-view methods outperform the single-view baseline. This conversely illustrates that the learning performance cannot be improved if multiple views are not properly integrated. In this respect, ERL-MVSC shows its superiority over the other multi-view classification methods. It can always achieve better classification accuracies than the single-view based method. Fig. 4 presents the comparison of ERL-MVSC with fused multi-view and single-view information, respectively. The results further validate that ERL-MVSC can reasonably fuse different views to exploit the hidden complementary information.
2) Robustness of ERL-MVSC: Different compared methods exhibit respective strengths for different types of

TABLE III

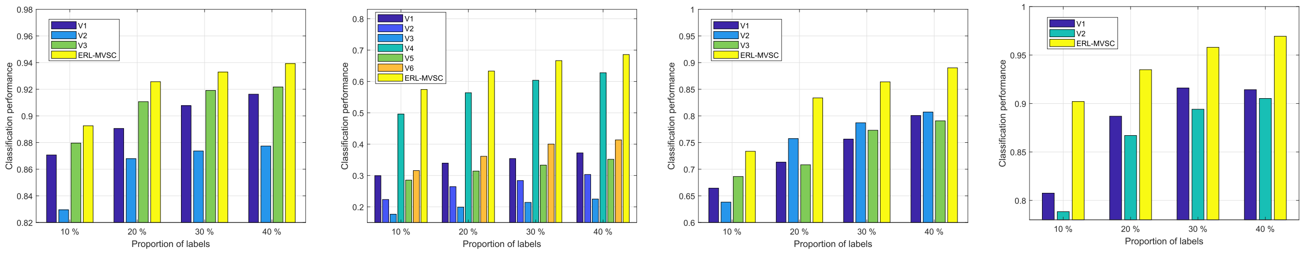
CLASSIFICATION ACCURACY (%) COMPARISON OF DIFFERENT ALGORITHMS ON ALOI AND CALTECH101. THE BEST RESULTS ARE HIGHLIGHTED IN RED BOLDFACE AND THE SECOND BEST ARE MARKED WITH BLUE BOLDFACE. (HIGHER MEANS BETTER)

Datasets	Methods	10%	20%	30%	40%	50%	60%	70%	80%
ALOI	SVM	33.1 (4.7)	37.2 (3.6)	40.1 (2.0)	40.7 (2.4)	41.5 (1.8)	42.9 (1.5)	45.2 (2.4)	44.7 (4.1)
	AMGL	82.4 (3.3)	89.3 (2.3)	94.6 (1.7)	95.4 (0.7)	97.0 (0.7)	98.1 (0.6)	98.1 (0.6)	98.4 (1.0)
	MVAR	67.1 (6.1)	77.8 (8.5)	80.3 (6.9)	88.1 (6.2)	86.2 (5.2)	85.1 (1.3)	88.8 (8.3)	92.5 (3.6)
	MLAN	87.6 (2.6)	91.9 (2.9)	91.9 (1.7)	94.8 (2.8)	96.8 (0.8)	94.2 (4.6)	96.3 (1.2)	97.5 (1.7)
	AWDR	93.5 (1.9)	95.0 (0.7)	95.5 (0.5)	95.8 (0.2)	96.1 (0.3)	96.8 (0.4)	97.3 (1.2)	97.4 (1.0)
	Co-GCN	**96.9 (0.3)**	**97.6 (0.4)**	**97.6 (0.5)**	**97.7 (0.2)**	**98.3 (0.6)**	**98.3 (0.4)**	**98.2 (0.5)**	**98.5 (0.5)**
	HLR-M^{2}VS	89.3 (2.2)	94.2 (0.9)	95.7 (0.8)	96.1 (0.7)	96.4 (0.8)	96.7 (0.9)	97.3 (1.1)	97.3 (0.7)
	ERL-MVSC	**93.6 (1.1)**	**95.7 (0.6)**	**97.2 (1.0)**	**97.8 (0.7)**	**98.4 (0.6)**	**98.5 (0.6)**	**98.8 (0.6)**	**99.1 (0.4)**
Datasets	Methods	10%	20%	30%	40%	50%	60%	70%	80%
Caltech101	SVM	22.6 (0.6)	26.6 (0.2)	28.9 (0.3)	29.8 (0.3)	31.4 (0.3)	32.6 (0.3)	32.8 (0.4)	32.5 (0.2)
	AMGL	24.3 (0.8)	29.5 (1.8)	34.8 (0.5)	37.3 (0.5)	38.3 (0.1)	41.0 (0.0)	42.2 (1.2)	44.1 (0.0)
	MVAR	43.5 (0.3)	47.7 (0.6)	51.9 (0.5)	53.8 (1.0)	53.4 (0.3)	53.4 (1.0)	53.0 (0.8)	53.4 (1.6)
	MLAN	31.4 (0.4)	37.2 (0.2)	38.1 (1.7)	40.4 (0.7)	41.4 (1.2)	44.1 (0.0)	44.2 (0.0)	44.7 (0.0)
	AWDR	45.0 (1.3)	53.1 (0.3)	56.9 (0.2)	59.6 (0.1)	61.4 (0.6)	62.7 (1.5)	63.0 (0.3)	64.0 (0.7)
	Co-GCN	**46.1 (0.6)**	52.6 (0.8)	54.3 (0.4)	55.2 (1.3)	56.1 (0.8)	55.8 (1.1)	55.6 (0.9)	55.9 (0.9)
	HLR-M^{2}VS	45.6 (0.7)	**54.2 (0.8)**	**57.4 (0.9)**	**60.4 (0.6)**	**62.5 (1.1)**	**63.3 (0.8)**	**64.1 (1.0)**	**65.2 (1.9)**
	ERL-MVSC	**46.2 (0.1)**	**56.0 (1.0)**	**60.2 (0.0)**	**61.7 (0.0)**	**63.9 (0.9)**	**66.2 (1.2)**	**66.4 (1.1)**	**67.0 (1.6)**

TABLE IV

CLASSIFICATION ACCURACY (%) COMPARISON OF DIFFERENT ALGORITHMS ON HW AND MNIST. THE BEST RESULTS ARE HIGHLIGHTED IN RED BOLDFACE AND THE SECOND BEST ARE MARKED WITH BLUE BOLDFACE. IN MNIST DATASET, THE RESULTS OF HLR-M^{2}VS ARE NOT PROVIDED DUE TO THE OUT-OF-MEMORY EXCEPTION CAUSED BY LARGE-SCALE SIMILARITY MATRIX. (HIGHER MEANS BETTER)

Datasets	Methods	10%	20%	30%	40%	50%	60%	70%	80%
HW	SVM	31.3 (2.7)	69.6 (1.9)	80.5 (1.0)	84.6 (0.9)	86.6 (0.9)	88.1 (1.1)	89.2 (1.0)	90.0 (1.1)
	AMGL	88.8 (0.7)	92.1 (0.5)	93.5 (0.6)	94.8 (0.7)	95.5 (0.9)	96.5 (0.1)	97.0 (0.8)	97.0 (0.8)
	MVAR	76.2 (0.6)	55.6 (2.7)	53.8 (3.2)	88.8 (0.6)	93.4 (0.9)	95.1 (0.4)	95.8 (0.7)	96.0 (0.7)
	MLAN	**95.7 (0.8)**	**96.6 (0.2)**	**96.7 (0.4)**	**96.9 (0.9)**	**96.8 (0.5)**	**97.1 (0.5)**	**97.0 (0.4)**	**97.5 (1.0)**
	AWDR	89.2 (1.0)	92.3 (1.0)	92.2 (0.8)	92.3 (0.3)	94.0 (0.6)	95.5 (0.3)	95.3 (0.7)	96.4 (0.8)
	Co-GCN	89.1 (1.1)	90.9 (1.3)	92.1 (1.4)	92.0 (1.3)	92.6 (1.3)	93.6 (1.2)	94.1 (1.6)	94.1 (1.1)
	HLR-M^{2}VS	86.4 (1.8)	89.9 (1.6)	93.3 (0.9)	94.0 (0.7)	95.2 (0.5)	95.4 (0.5)	96.3 (0.7)	96.2 (0.7)
	ERL-MVSC	**94.3 (1.1)**	**96.8 (0.6)**	**97.1 (0.4)**	**97.3 (0.4)**	**97.6 (0.4)**	**98.0 (0.4)**	**98.3 (0.5)**	**98.5 (0.6)**
Datasets	Methods	10%	20%	30%	40%	50%	60%	70%	80%
MNIST	SVM	82.9 (0.2)	87.6 (0.2)	**89.9 (0.1)**	**90.9 (0.1)**	**91.7 (0.3)**	**92.2 (0.2)**	**92.9 (0.2)**	**93.1 (0.1)**
	AMGL	32.1 (0.3)	33.4 (0.3)	34.6 (0.1)	35.1 (0.2)	35.7 (0.2)	35.9 (0.5)	36.6 (0.5)	37.2 (0.5)
	MVAR	53.4 (2.1)	53.9 (0.1)	53.1 (0.7)	53.2 (0.8)	54.1 (0.3)	53.4 (0.7)	53.4 (0.5)	53.8 (0.5)
	MLAN	31.3 (0.0)	31.5 (0.4)	32.2 (0.9)	33.0 (0.9)	29.6 (0.6)	27.3 (0.8)	10.0 (0.9)	33.1 (0.7)
	AWDR	77.4 (0.3)	79.3 (0.2)	80.0 (0.1)	80.3 (0.4)	81.1 (0.2)	81.4 (0.1)	81.2 (0.5)	81.9 (0.4)
	Co-GCN	**87.2 (0.2)**	**88.2 (0.9)**	88.8 (1.2)	89.2 (1.3)	89.5 (1.3)	89.9 (1.4)	90.1 (1.4)	91.2 (0.2)
	HLR-M^{2}VS	-	-	-	-	-	-	-	-
	ERL-MVSC	**91.6 (0.1)**	**93.1 (0.2)**	**93.8 (0.2)**	**93.9 (0.1)**	**94.4 (0.2)**	**94.8 (0.2)**	**94.9 (0.3)**	**95.2 (0.3)**



MNIST YouTube 3sources BBCsports

Fig. 4. Comparison of ERL-MVSC with fused multi-view and single-view information, where Vn indicates the n-th view is used.

images. AMGL and MLAN are graph-based multi-view semi-supervised classification methods. They perform well on HW and ALOI, but achieve unsatisfactory performance on BBCsports and 3sources as shown in

TABLE V

CLASSIFICATION ACCURACY (%) COMPARISON OF DIFFERENT ALGORITHMS ON YOUTUBE AND 3SOURCES. THE BEST RESULTS ARE HIGHLIGHTED IN RED BOLDFACE AND THE SECOND BEST ARE MARKED WITH BLUE BOLDFACE. (HIGHER MEANS BETTER)

Datesets	Methods	10%	20%	30%	40%	50%	60%	70%	80%
YouTube	SVM	20.5 (2.4)	26.0 (1.6)	29.8 (1.3)	33.0 (1.5)	35.8 (1.5)	38.8 (1.8)	40.5 (1.9)	42.7 (3.0)
	AMGL	44.2 (0.8)	54.3 (1.4)	60.9 (1.0)	63.8 (0.6)	65.9 (1.2)	68.1 (0.9)	71.0 (1.7)	71.7 (1.5)
	MVAR	37.6 (1.8)	44.8 (2.8)	52.4 (0.9)	54.1 (3.9)	59.6 (2.4)	61.9 (2.1)	63.7 (1.4)	64.8 (3.1)
	MLAN	37.0 (1.3)	43.0 (0.8)	46.2 (1.2)	48.8 (1.9)	49.6 (1.8)	51.6 (1.9)	52.3 (2.4)	53.7 (3.1)
	AWDR	36.1 (2.8)	43.4 (1.1)	51.8 (1.3)	55.8 (1.7)	57.2 (1.9)	58.1 (1.7)	58.4 (2.0)	61.0 (1.9)
	Co-GCN	**58.0 (0.5)**	**65.2 (1.3)**	**70.6 (0.9)**	**71.4 (0.7)**	**72.7 (1.1)**	**74.6 (1.4)**	**75.7 (1.0)**	**76.0 (1.0)**
	HLR-M^{2}VS	32.7 (1.1)	37.2 (0.8)	40.4 (1.4)	42.3 (0.7)	43.9 (1.2)	46.2 (1.9)	48.2 (1.9)	46.9 (1.8)
	ERL-MVSC	**57.4 (1.8)**	**63.3 (1.3)**	**66.6 (1.3)**	**68.6 (0.9)**	**71.2 (1.4)**	**73.3 (1.7)**	**74.0 (1.6)**	**75.0 (2.5)**
Datesets	Methods	10%	20%	30%	40%	50%	60%	70%	80%
3sources	SVM	33.8 (1.4)	38.9 (9.6)	42.1 (8.9)	46.2 (8.8)	51.4 (5.5)	56.9 (5.3)	56.9 (6.9)	60.6 (8.5)
	AMGL	39.9 (7.9)	41.9 (5.9)	45.0 (3.4)	46.8 (3.5)	43.8 (4.7)	48.2 (2.8)	46.2 (7.6)	44.3 (7.1)
	MVAR	46.8 (13.6)	66.2 (8.3)	65.6 (1.0)	75.3 (5.1)	77.2 (3.2)	75.6 (9.8)	78.0 (6.1)	81.8 (7.9)
	MLAN	45.5 (10.2)	55.0 (2.8)	57.2 (3.6)	55.3 (4.5)	60.0 (6.7)	58.6 (4.5)	56.4 (4.3)	56.8 (6.6)
	AWDR	50.1 (7.6)	**78.7 (8.5)**	74.8 (1.5)	82.0 (7.4)	**86.3 (3.0)**	**89.7 (1.5)**	83.0 (3.0)	82.4 (2.9)
	Co-GCN	53.4 (1.6)	75.9 (2.2)	78.1 (1.3)	81.1 (2.0)	84.5 (1.8)	85.3 (3.6)	**90.9 (0.9)**	**92.2 (3.7)**
	HLR-M^{2}VS	**69.1 (7.5)**	75.7 (4.7)	**80.6 (5.1)**	**82.9 (4.5)**	83.4 (4.3)	84.6 (3.1)	85.7 (5.0)	87.9 (5.6)
	ERL-MVSC	**73.4 (0.9)**	**83.4 (2.3)**	**86.4 (3.4)**	**89.0 (2.8)**	**91.5 (1.5)**	**92.4 (1.2)**	**94.0 (1.6)**	**95.3 (2.6)**

TABLE VI

CLASSIFICATION ACCURACY (%) COMPARISON OF DIFFERENT ALGORITHMS ON BBCNEWS AND BBCSPORTS. THE BEST RESULTS ARE HIGHLIGHTED IN RED BOLDFACE AND THE SECOND BEST ARE MARKED WITH BLUE BOLDFACE. (HIGHER MEANS BETTER)

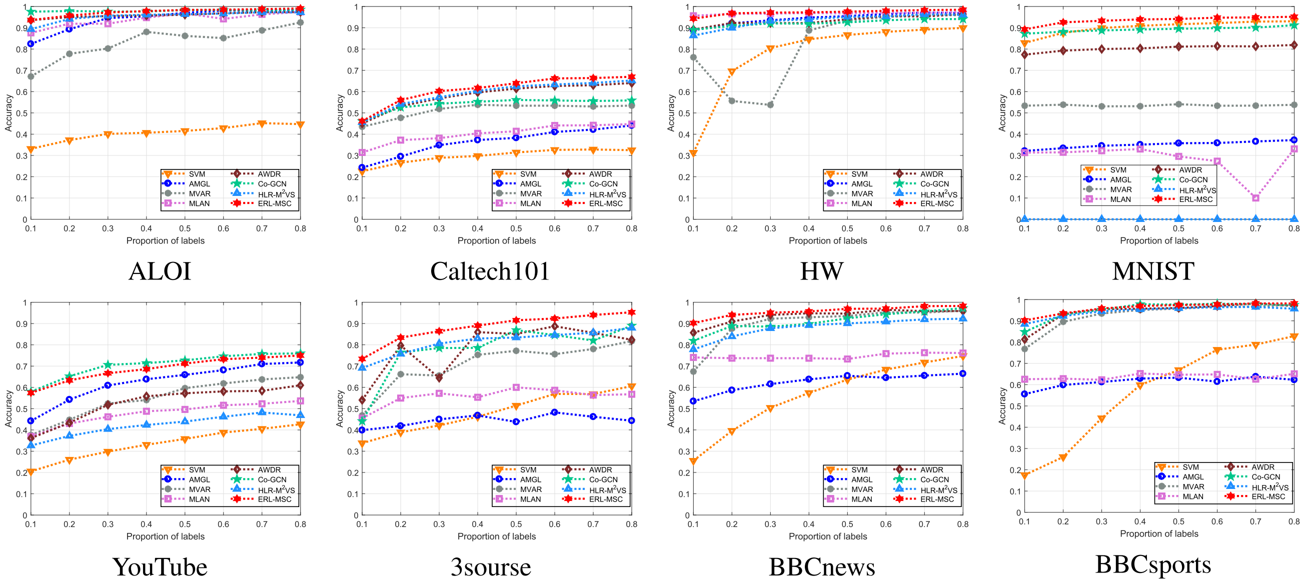
Datesets	Methods	10%	20%	30%	40%	50%	60%	70%	80%
BBCnews	SVM	25.5 (7.8)	39.6 (5.1)	50.4 (4.8)	57.4 (2.2)	63.6 (3.1)	68.4 (3.3)	71.7 (3.2)	74.7 (3.2)
	AMGL	53.5 (2.6)	58.7 (2.8)	61.5 (2.1)	63.8 (2.4)	65.5 (1.2)	64.5 (2.6)	65.5 (2.2)	66.4 (3.1)
	MVAR	67.4 (1.6)	87.6 (2.7)	92.3 (1.4)	92.9 (0.8)	93.5 (1.0)	95.3 (1.2)	95.2 (1.5)	96.1 (1.5)
	MLAN	74.1 (0.9)	73.7 (1.1)	73.7 (1.0)	73.7 (1.0)	73.3 (1.9)	75.9 (1.9)	76.3 (1.1)	76.1 (2.6)
	AWDR	**85.7 (1.4)**	**90.9 (5.7)**	**94.0 (1.5)**	**94.8 (0.6)**	**94.7 (0.6)**	**96.5 (0.4)**	**95.8 (1.4)**	95.9 (1.8)
	Co-GCN	81.9 (1.5)	88.9 (1.3)	88.5 (0.7)	89.8 (1.0)	92.5 (0.6)	94.4 (0.6)	95.7 (0.5)	**97.2 (1.1)**
	HLR-M^{2}VS	77.8 (4.1)	83.9 (2.3)	87.8 (1.6)	89.2 (2.4)	90.1 (1.9)	90.9 (2.2)	91.9 (2.2)	92.3 (2.6)
	ERL-MVSC	**90.2 (1.8)**	**94.0 (1.0)**	**95.1 (1.1)**	**95.7 (1.5)**	**96.9 (1.8)**	**97.0 (2.4)**	**98.2 (1.4)**	**98.3 (2.0)**
Datesets	Methods	10%	20%	30%	40%	50%	60%	70%	80%
BBCsports	SVM	17.5 (6.1)	26.0 (6.1)	44.3 (5.1)	59.8 (2.5)	66.9 (2.4)	76.4 (1.8)	78.8 (4.1)	82.8 (3.5)
	AMGL	55.6 (1.4)	59.9 (1.8)	61.3 (1.2)	62.9 (2.2)	63.3 (1.7)	61.4 (2.4)	63.8 (2.4)	62.3 (4.1)
	MVAR	76.8 (9.6)	89.5 (3.7)	93.4 (2.3)	95.1 (1.2)	95.7 (2.0)	96.6 (1.2)	96.7 (2.3)	97.6 (1.5)
	MLAN	62.6 (2.2)	62.8 (1.0)	62.3 (0.6)	65.2 (2.2)	64.7 (2.5)	64.8 (0.7)	62.6 (1.5)	65.1 (3.3)
	AWDR	81.3 (3.3)	**93.1 (4.0)**	**95.5 (0.8)**	95.7 (0.8)	96.0 (0.4)	96.9 (1.5)	**98.2 (1.2)**	96.9 (2.8)
	Co-GCN	84.2 (1.4)	92.9 (1.1)	95.2 (0.8)	**96.9 (0.6)**	**97.4 (0.6)**	**97.8 (0.6)**	98.1 (0.6)	**98.5 (1.1)**
	HLR-M^{2}VS	**88.5 (4.1)**	92.3 (1.8)	94.5 (1.2)	95.4 (0.9)	95.9 (1.3)	96.4 (1.2)	96.5 (1.9)	95.7 (1.4)
	ERL-MVSC	**90.2 (2.3)**	**93.5 (3.2)**	**95.8 (1.0)**	**97.2 (0.9)**	**97.8 (1.1)**	**98.1 (1.0)**	**98.2 (1.3)**	**98.8 (2.0)**

TABLE VII

OVERALL AVERAGE CLASSIFICATION ACCURACY (%) COMPARISON OF DIFFERENT METHODS ON ALL TEST DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN RED BOLDFACE AND THE SECOND-BEST ONES ARE MARKED WITH BLUE BOLDFACE

Methods	10%	20%	30%	40%	50%	60%	70%	80%
SVM	33.4 (3.2)	44.0 (3.5)	50.7 (2.9)	55.3 (2.4)	58.6 (2.0)	62.0 (2.1)	63.5 (2.5)	65.1 (3.0)
AMGL	52.6 (2.4)	57.4 (2.1)	60.8 (1.3)	62.5 (1.5)	63.1 (1.3)	64.2 (1.3)	65.1 (2.2)	65.2 (2.3)
MVAR	58.6 (6.3)	65.4 (3.7)	67.8 (2.1)	75.1 (2.4)	76.7 (1.9)	77.0 (2.2)	78.1 (2.7)	79.5 (2.6)
MLAN	58.1 (2.3)	61.5 (1.2)	62.3 (1.4)	63.5 (1.7)	64.2 (1.9)	64.2 (1.7)	61.9 (1.3)	65.5 (2.3)
AWDR	70.3 (3.0)	78.4 (1.8)	78.8 (1.0)	82.5 (0.6)	83.2 (1.5)	84.6 (1.1)	84.4 (1.9)	84.5 (1.8)
Co-GCN	**73.6 (0.9)**	**81.6 (1.1)**	**83.2 (0.9)**	**84.0 (1.0)**	**85.8 (1.2)**	**86.1 (1.2)**	**86.2 (1.8)**	**87.4 (1.0)**
HLR-M^{2}VS	69.9 (3.1)	75.3 (1.8)	78.5 (1.7)	80.1 (1.4)	81.1 (1.4)	81.9 (1.4)	82.9 (1.8)	83.1 (1.8)
ERL-MVSC	**79.6 (1.1)**	**84.5 (1.3)**	**86.5 (1.0)**	**87.7 (0.9)**	**88.9 (1.0)**	**89.7 (1.1)**	**90.3 (1.0)**	**90.8 (1.5)**

Tables V and VI. This instability may be due to the fact that the constructed graphs are easily affected by the type and quality of data. HLR-M^{2}VS constructs a hypergraph rather than a traditional pairwise graph on an optimized subspace representation to discover high-order local geometrical structures; Co-GCN employs a graph convolutional network for label propagation to exploit the structural information from different views



ALOI Caltech101 HW MNIST

YouTube 3sourse BBCnews BBCsports

Fig. 5. Performance comparison of various methods as the ratio of labeled data ranges in $\{0.1, 0.2, \cdots, 0.8\}$.

with the expressive power of neural network. Both of them avail to achieve robust performance. MVAR and AWDR are regression-based methods. MVAR can automatically weaken the impact of low-quality classifier and bypass the problem of graph construction by directly regressing to labels, which can improve the stability of classification performance to some extent. AWDR maps multi-view data into a low-dimensional discriminative subspace to enhance the discriminability of the features for subsequent classification, which achieves better and more robust performance than MVAR. The proposed method ERL-MVSC outperforms the state-of-the-arts on seven of the eight datasets. An exception is the YouTube dataset, on which our method is ranked the second best of all algorithms. This may be attributed to the reason that the graph convolutional network is more effective in capturing high-dimensional dense features of YouTube. Nevertheless, from the overall average performance shown in the Table VII, ERL-MVSC outperforms all compared methods, which validates the superiority of ERL-MVSC.

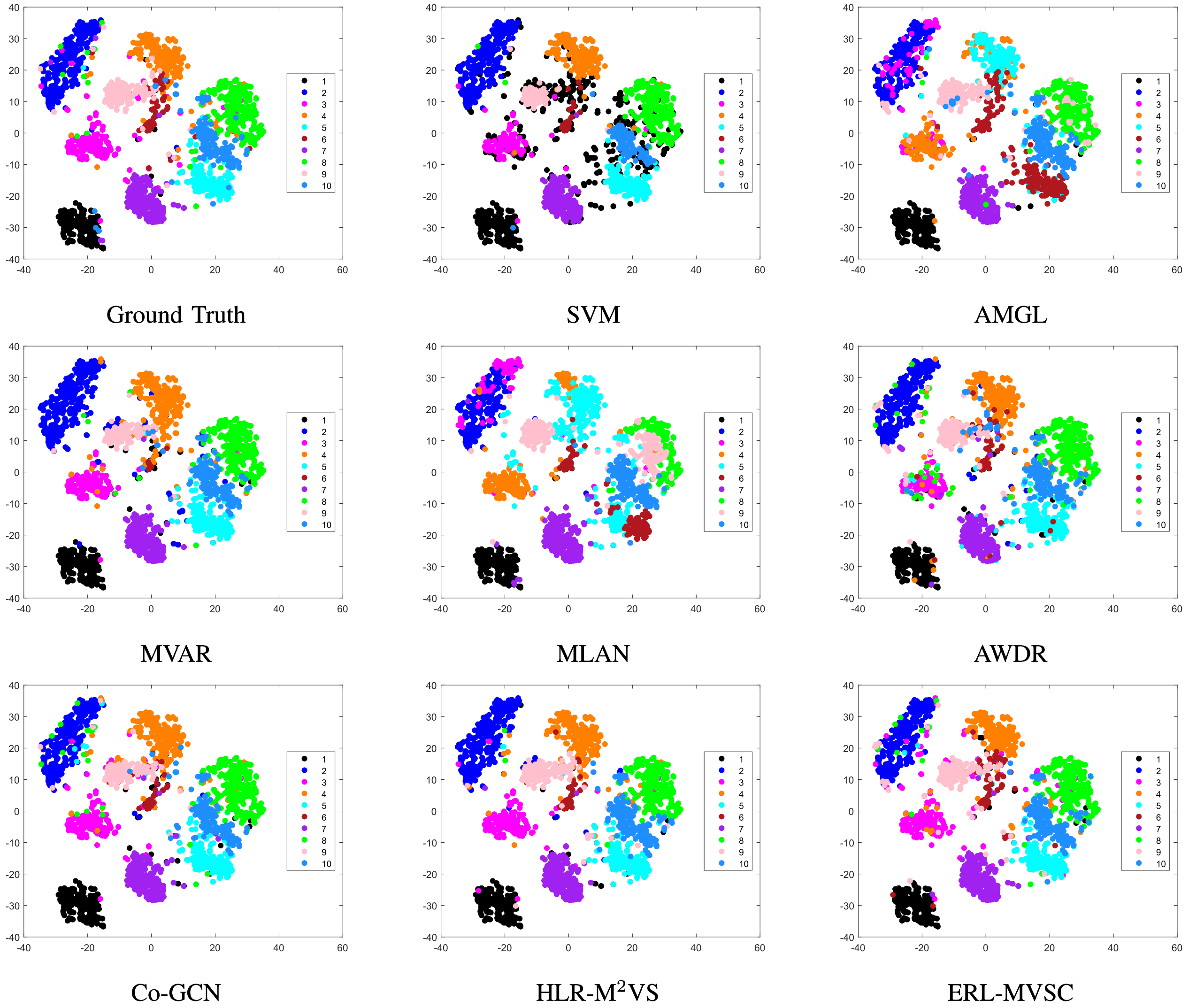
3) Superior label-utilization capability of ERL-MVSC: As shown in Fig. 5, exploiting more training samples does not necessarily bring positive effects. It may hinder their evolution towards higher agreement levels, resulting in unsatisfactory performance. For example, the performance of MVAR on HW tends to decrease when the labeling ratio ≤ 0.3; The performance of MLAN on MNIST presents a descend trend when label rate varies in [0.4, 0.7]; In contrast, the proposed ERL-MVSC still remains stably rising performance with more label information. This indicates that ERL-MVSC can effectively utilize label information and propagate them to unlabeled data.

To intuitively show the classification performance, Fig. 6 presents scatter diagrams of all algorithms with 0.1 label ratio on a subset of MNIST with 2,000 samples. We first concatenate feature vectors of 3 views together, and employ t-distributed stochastic neighbor embedding (t-SNE) [51] to project the original high-dimensional data onto a 2 D space. The mapped 2 D data is then colored with the class labels obtained by different methods and ground truth. In this figure, our ERL-MVSC method assigns more consistent class labels with the ground truth, which further validates the effectiveness and superiority of our model.

E. Model Analysis

1) Runtime Complexity: Table VIII compares the runtime complexities of different algorithms on all test datasets. With 10% labeled samples, we execute each algorithm 10 times and report the average execution time. From Table VIII, Co-GCN and HLR-M^{2}VS consume more time than the other methods due to the parameter training of graph neural network and the computations for hyper-graph construction, respectively. Our ERL-MVSC method performs stably and exhibits a good running speed compared with the other competitors. Considering its superior classification performance, its time consumption is acceptable for real-world applications.

2) Convergence Validation: Fig. 7 shows the convergence behaviors of ERL-MVSC on ALOI, YouTube, 3sources and BBCnews under 10% of label samples. From the figure, the objective function values decrease with increasing iteration number and then reach convergence within 5 iterations. The fast convergence speed suggests the effectiveness of ERL-MVSC and its good scalability in practical application. Fig. 8 shows the evolution of view weights on the other four test datasets. Each view weight is initialized equally and then converge within a limited number of iterations, which also indicates the stable convergence of ERL-MVSC.



Ground Truth

SVM

AMGL

MVAR

MLAN

AWDR

Co-GCN

HLR-M^{2}VS

ERL-MVSC

Fig. 6. Experimental visualizations of data distributions with different methods on a subset of MNIST with 2,000 samples.

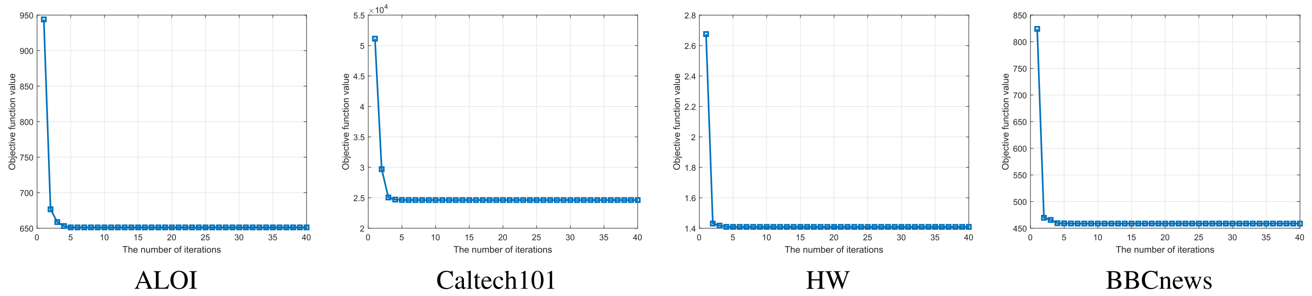
TABLE VIII

RUN-TIME COMPLEXITY COMPARISON OF CLASSIFICATION WITH DIFFERENT ALGORITHMS ON EIGHT DATASETS (IN SECONDS). THE BEST RESULTS ARE HIGHLIGHTED IN RED BOLDFACE AND THE SECOND BEST ARE MARKED WITH BLUE BOLDFACE. IN MNIST DATASET, THE RESULT OF HLR-M^{2}VS IS NOT SHOWN DUE TO ITS OUT-OF-MEMORY EXCEPTIONS. (LOWER MEANS BETTER)

Datasets	SVM	AMGL	MVAR	MLAN	AWDR	Co-GCN	HLR-M^{2}VS	ERL-MVSC
ALOI	13.8	1.2	2.4	2.8	**0.2**	2,676.2	103.6	**0.9**
Caltech101	688.0	**346.6**	521.3	1210.3	**25.9**	2,320.3	78,206.3	400.3
HW	9.3	**6.0**	16.3	79.1	**0.6**	568.2	1,351.9	**6.0**
MNIST	**33.1**	2,115.5	432.7	11454.5	**23.6**	36,012.6	-	2,397.3
YouTube	33.8	**2.9**	27.2	22.1	**4.4**	968.4	412.19	5.8
3sources	0.5	**0.1**	0.4	**0.2**	8.4	29.2	784.4	**0.2**
BBCnews	**1.9**	**0.7**	7.4	2.1	32.5	394.6	66.0	146.5
BBCsports	**0.3**	**0.2**	1.9	0.7	2.4	172.2	21.6	31.0

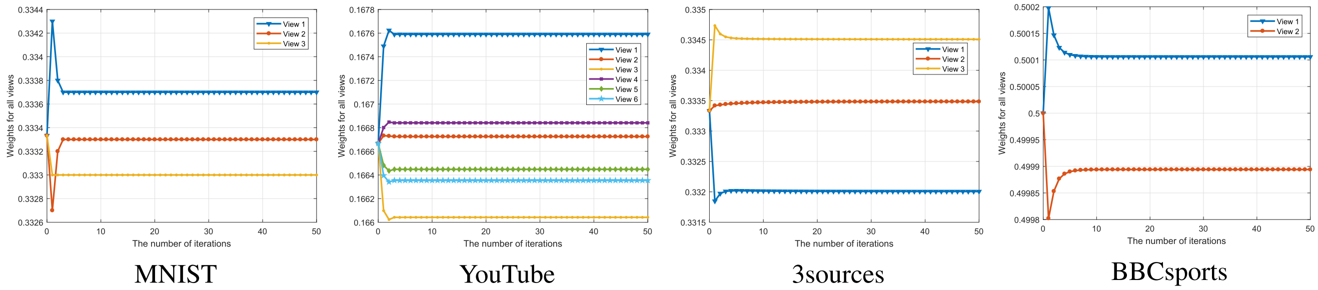
3) Parameter Sensitivity: To investigate the performance variations of proposed method under different settings, parameter sensitivity analysis is conducted with respect to different labeling ratios. There are four parameters in our model, including the smoothing factor α, the embedding parameter β, the regularization parameter γ and the fitting coefficient δ. For each parameter, we find the optimal values on each dataset by grid searching, and then select an acceptable common value to implement multi-view classification in all benchmarks. Herein, we mainly present the parametric sensitivity of β, γ and δ, as shown in Figs. 9 $\sim$ 11.

While keeping α, γ and δ unchanged, the accuracy variation curves are reported in Fig. 9, where the embedding parameter β ranges in $\{10^{-5}, 10^{-4}, \cdots, 10^{2}\}$. The best results

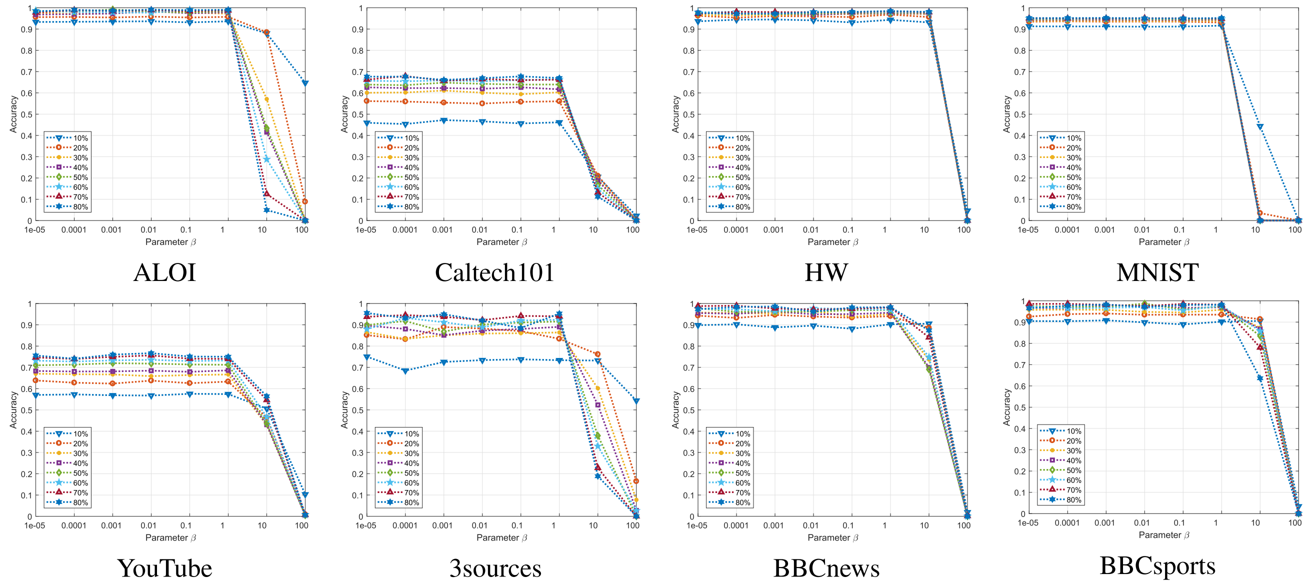


ALOI Caltech101 HW BBCnews

Fig. 7. Convergence curves of ERL-MVSC on four test datasets.



MNIST YouTube 3sources BBCsports



ALOI Caltech101 HW MNIST

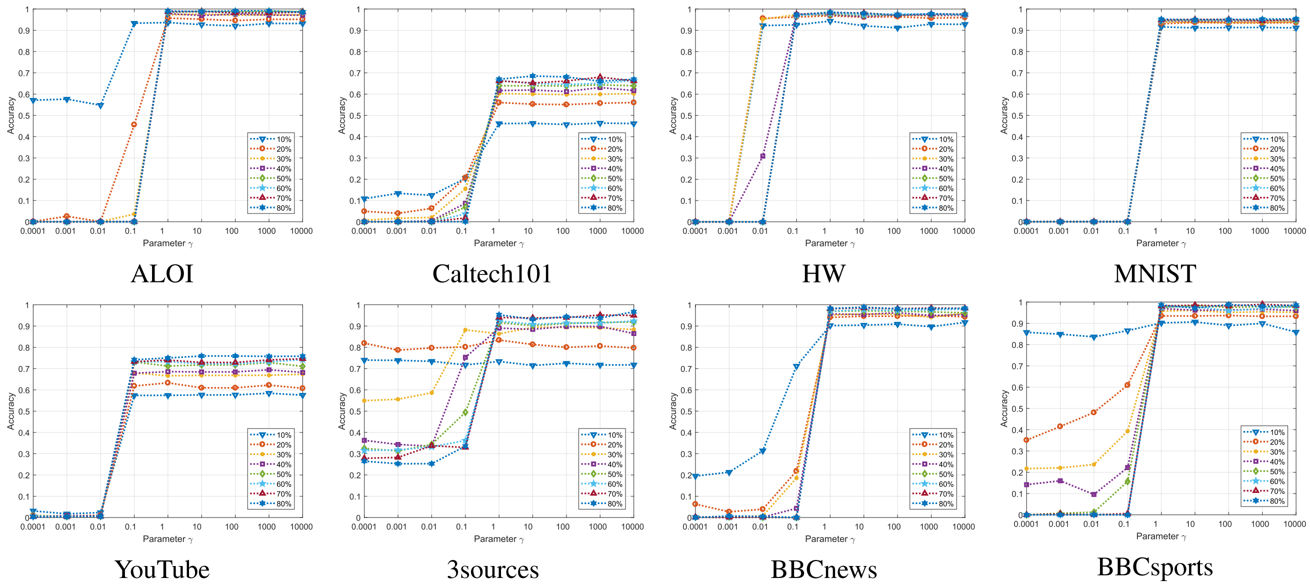
YouTube 3sources BBCnews BBCsports

Fig. 9. Performance of ERL-MVSC with various embedding parameter β ranging in $\{10^{-5}, 10^{-4}, \cdots, 10^{2}\}$.

of ERL-MVSC are attained at $[10^{-5}, 1]$, then the performances sharply decrease as β exceeds 1 in most cases. It may be attributed to the following reason. As $\mathbf{Y} = \left(\frac{\beta}{2\delta}(\mathbf{M}+\mathbf{M}^T)+\mathbf{I}\right)^{-1}\mathbf{L}$, the necessary condition for the existence of $\left(\frac{\beta}{2\delta}(\mathbf{M}+\mathbf{M}^T)+\mathbf{I}\right)^{-1}$ is that the spectral radius of $\frac{\beta}{2\delta}(\mathbf{M}+\mathbf{M}^T)$ is less than 1. A large β will make the spectral radius of $\frac{\beta}{2\delta}(\mathbf{M}+\mathbf{M}^T)$ exceed 1, resulting in unsatisfactory performance.

With fixed α, β and δ, the parametric sensitivity of the proposed ERL-MVSC with respect to γ is analyzed in Fig. 10, where the regularization parameter γ varies in $\{10^{-4}, 10^{-3}, \cdots, 10^{3}, 10^{4}\}$. It shows that the classification performance of ERL-MVSC is relatively stable when γ falls in the range of $[1, 10^{4}]$. However, when γ is very small (close to 0), its classification performance becomes poor. This fact demonstrates the importance of sparsity constraint.

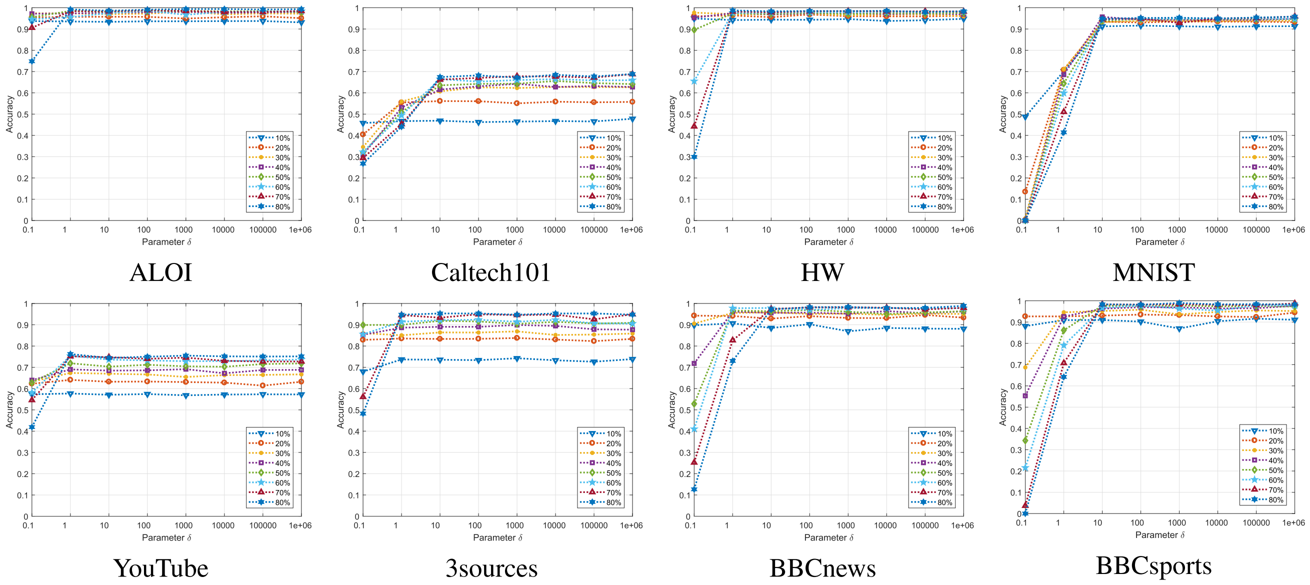
Keeping α, β and γ unchanged, the classification performance of ERL-MVSC is reported in Fig. 11. Herein, the fitting coefficient δ ranges in $\{10^{-1}, 1, \cdots, 10^{5}, 10^{6}\}$. The overall performance is relatively stable when δ varies in the range of $[10^{1}, 10^{6}]$. However, similar to the parameter sensitivity analysis of β, a smaller δ may make the spectral radius



ALOI Caltech101 HW MNIST

YouTube 3sources BBCnews BBCsports

Fig. 10. Performance of ERL-MVSC with various regularization parameter γ ranging in $\{10^{-4}, 10^{-3}, \cdots, 10^3, 10^4\}$.



ALOI Caltech101 HW MNIST

YouTube 3sources BBCnews BBCsports

Fig. 11. Performance of ERL-MVSC with various regularization parameter δ ranging in $\{10^{-1}, 1, \cdots, 10^5, 10^6\}$.

of $\frac{\beta}{2\delta}(\mathbf{M} + \mathbf{M}^T)$ to be greater than 1, resulting in inaccurate label assignment.

V. Conclusion

In this paper, we proposed a multi-view semi-supervised classification algorithm named ERL-MVSC, that exploits diversity, sparsity and consensus information of different views. Extensive experiments performed on eight datasets demonstrate the clear superiority of ERL-MVSC compared to the state-of-the-art algorithms. In the future, a more accurate embedding regularizer for multi-view semi-supervised classification may be learned by deep learning techniques. Particularly, a link between embedding regularizers and graph convolutions can be constructed to handle the problem with severely limited labeled data.

References

[1] D. Wang and S. Zhang, "Unsupervised person re-identification via multi-label classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10978–10987.

[2] X. Li, S. Wang, Y. Zhao, J. Verbeek, and J. Kannala, "Hierarchical scene coordinate classification and regression for visual localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11980–11989.

[3] B. Zhuang, C. Shen, M. Tan, L. Liu, and I. Reid, "Structured binary neural networks for accurate image classification and semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 413–422.

[4] R. Yu, J. Sun, and H. Li, "Second-order spectral transform block for 3D shape classification and retrieval," *IEEE Trans. Image Process.*, vol. 29, pp. 4530–4543, 2020.

[5] P. Li, H. Wang, C. Bohm, and J. Shao, "Online semi-supervised multi-label classification with label compression and local smooth regression," in *Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2020, pp. 1359–1365.

[6] W. Lin, Z. Gao, and B. Li, "Shoestring: Graph-based semi-supervised classification with severely limited labeled data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4173–4181.

[7] C. Zhang, J. Cheng, and Q. Tian, "Unsupervised and semi-supervised image classification with weak semantic consistency," *IEEE Trans. Multimedia*, vol. 21, no. 10, pp. 2482–2491, Oct. 2019.

[8] M. Luo, X. Chang, L. Nie, Y. Yang, A. G. Hauptmann, and Q. Zheng, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 648–660, Feb. 2018.

[9] Y. Zhang, J. Wu, Z. Cai, and P. S. Yu, "Multi-view multi-label learning with sparse feature selection for image annotation," *IEEE Trans. Multimedia*, vol. 22, no. 11, pp. 2844–2857, Nov. 2020.

[10] L. Wang, K. Yang, C. Li, L. Hong, Z. Li, and J. Zhu, "ORDisCo: Effective and efficient usage of incremental unlabeled data for semi-supervised continual learning," 2021, *arXiv:2101.00407*. [Online]. Available: http://arxiv.org/abs/2101.00407

[11] P. Peng, Y. Tian, Y. Wang, J. Li, and T. Huang, "Robust multiple cameras pedestrian detection with multi-view Bayesian network," *Pattern Recognit.*, vol. 48, no. 5, pp. 1760–1772, May 2015.

[12] Z. zhong Lan, L. Bao, S.-I. Yu, W. Liu, and A. G. Hauptmann, "Double fusion for multimedia event detection," in *Proc. Int. Conf. Multimedia Model. (MMM)*, 2012, pp. 173–185.

[13] Z. Hu, F. Nie, R. Wang, and X. Li, "Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding," *Inf. Fusion*, vol. 55, pp. 251–259, Mar. 2020.

[14] J. Wu, Z. Lin, and H. Zha, "Essential tensor learning for multi-view spectral clustering," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5910–5922, Dec. 2019.

[15] C. Zhang *et al.*, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, Jan. 2020.

[16] Z. Tao, H. Liu, H. Fu, and Y. Fu, "Multi-view saliency-guided clustering for image cosegmentation," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4634–4645, Sep. 2019.

[17] A. Huang, T. Zhao, and C.-W. Lin, "Multi-view data fusion oriented clustering via nuclear norm minimization," *IEEE Trans. Image Process.*, vol. 29, pp. 9600–9613, 2020.

[18] Z. Xue, J. Du, D. Du, W. Ren, and S. Lyu, "Deep correlated predictive subspace learning for incomplete multi-view semi-supervised classification," in *Proc. 28th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2019, pp. 4026–4032.

[19] F. Wu *et al.*, "Semi-supervised multi-view individual and sharable feature learning for webpage classification," in *Proc. 28th Int. World Wide Web Conf. (WWW)*, 2019, pp. 3349–3355.

[20] X.-Y. Jing, F. Wu, X. Dong, S. Shan, and S. Chen, "Semi-supervised multi-view correlation feature learning with application to webpage classification," in *Proc. 31st AAAI Conf. Artif. Intell. (AAAI)*, 2017, pp. 1374–1381.

[21] G. Lin, K. Liao, B. Sun, Y. Chen, and F. Zhao, "Dynamic graph fusion label propagation for semi-supervised multi-modality classification," *Pattern Recognit.*, vol. 68, pp. 14–23, Aug. 2017.

[22] S. Wang, Z. Chen, S. Du, and Z. Lin, "Learning deep sparse regularizers with applications to multi-view clustering and semi-supervised classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, May 21, 2021, doi: 10.1109/TPAMI.2021.3082632.

[23] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Comput. Learn. Theory*, 1998, pp. 92–100.

[24] K. Nigam and R. Ghani, "Analyzing the effectiveness and applicability of co-training," in *Proc. 9th Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2000, pp. 86–93.

[25] S. Yu, B. Krishnapuram, R. Rosales, H. Steck, and R. B. Rao, "Bayesian co-training," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2007, pp. 1665–1672.

[26] V. Sindhwani, P. Niyogi, and M. Belkin, "A co-regularization approach to semi-supervised learning with multiple views," in *Proc. 22th Int. Conf. Mach. Learn. (ICML)*, 2005, pp. 74–79.

[27] F. Nie, J. Li, and X. Li, "Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification," in *Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2016, pp. 1881–1887.

[28] Z. Kang, X. Lu, J. Yi, and Z. Xu, "Self-weighted multiple kernel learning for graph-based clustering and semi-supervised classification," in *Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2018, pp. 2312–2318.

[29] Z. Xue, J. Du, D. Du, G. Li, Q. Huang, and S. Lyu, "Deep constrained low-rank subspace learning for multi-view semi-supervised classification," *IEEE Signal Process. Lett.*, vol. 26, no. 8, pp. 1177–1181, Aug. 2019.

[30] F. Nie, G. Cai, J. Li, and X. Li, "Auto-weighted multi-view learning for image clustering and semi-supervised classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1501–1511, Mar. 2018.

[31] Y. Xie, W. Zhang, Y. Qu, L. Dai, and D. Tao, "Hyper-laplacian regularized multilinear multiview self-representations for clustering and semi-supervised learning," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 572–586, Feb. 2020.

[32] Y. Cheng, X. Zhao, K. Huang, and T. Tan, "Semi-supervised learning for RGB-D object recognition," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 3345–3351.

[33] E. M. Ardehaly and A. Culotta, "Co-training for demographic classification using deep learning from label proportions," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 1017–1024.

[34] S. Li, W.-T. Li, and W. Wang, "Co-gcn for multi-view semi-supervised learning," in *Proc. 34th AAAI Conf. Artif. Intell. (AAAI)*, 2020, pp. 4691–4698.

[35] 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.

[36] M. Karasuyama and H. Mamitsuka, "Multiple graph label propagation by sparse integration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 1999–2012, Dec. 2013.

[37] F. Nie, L. Tian, R. Wang, and X. Li, "Multiview semi-supervised learning model for image classification," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 12, pp. 2389–2400, Dec. 2020.

[38] C. Zhang, H. Fu, J. Wang, W. Li, X. Cao, and Q. Hu, "Tensorized multi-view subspace representation learning," *Int. J. Comput. Vis.*, vol. 128, no. 8, pp. 2344–2361, 2020.

[39] 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.

[40] H. Tao, C. Hou, F. Nie, J. Zhu, and D. Yi, "Scalable multi-view semi-supervised classification via adaptive regression," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4283–4296, Sep. 2017.

[41] M. Yang, C. Deng, and F. Nie, "Adaptive-weighting discriminative regression for multi-view classification," *Pattern Recognit.*, vol. 88, pp. 236–245, Apr. 2019.

[42] W. Zhuge, C. Hou, S. Peng, and D. Yi, "Joint consensus and diversity for multi-view semi-supervised classification," *Mach. Learn.*, vol. 109, no. 3, pp. 445–465, Mar. 2020.

[43] D. Wang, F. Nie, and H. Huang, "Large-scale adaptive semi-supervised learning via unified inductive and transductive model," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2014, pp. 482–491.

[44] M. Luo, L. Zhang, F. Nie, X. Chang, B. Qian, and Q. Zheng, "Adaptive semi-supervised learning with discriminative least squares regression," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2421–2427.

[45] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[46] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with L1-graph for image analysis," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 858–866, Apr. 2010.

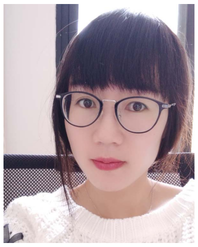
[47] H. Wang, F. Nie, and H. Huang, "Multi-view clustering and feature learning via structured sparsity," in *Proc. 30th Int. Conf. Mach. Learn. (ICML)*, vol. 2013, pp. 352–360.

[48] L.-B. Qiao, B.-F. Zhang, J.-S. Su, and X.-C. Lu, "A systematic review of structured sparse learning," *Frontiers Inf. Technol. Electron. Eng.*, vol. 18, no. 4, pp. 445–463, Apr. 2017.

[49] J. Lu, S. Yi, J. Zhao, Y. Liang, and W. Liu, "Interpretable robust feature selection via joint-norms minimization," in *Proc. 13th Int. Conf. Mach. Learn. Comput.*, Feb. 2021, pp. 1813–1821.

[50] D. Greene and P. Cunningham, "Practical solutions to the problem of diagonal dominance in kernel document clustering," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 377–384.

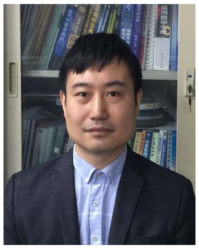
[51] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.



Aiping Huang (Member, IEEE) received the B.S. degree in mathematics and applied mathematics from Putian University, Putian, China, in 2011, the M.S. degree in basic mathematics from Minnan Normal University, Zhangzhou, China, in 2014, and the Ph.D. degree in communication and information system from Fuzhou University, Fuzhou, China, in 2021. She has served as a Lecturer for the School of Information Science and Technology, Xiamen University Tan Kah Kee College, from 2014 to 2018. She is currently a Postdoctoral Scholar with the College of Physics and Information Engineering, Fuzhou University. Her research interests include computer vision, image processing, machine learning, and data mining.

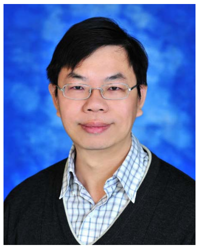
Zheng Wang received the B.S. degree in information management and information system from Chongqing Jiaotong University, Chongqing, China, in 2019. He is currently pursuing the M.S. degree with the College of Physics and Information Engineering, Fuzhou University. His research interests include computer vision, image processing, and video quality assessment.

Yannan Zheng received the B.S. degree in electronic and information engineering from Fuzhou University, Fuzhou, China, in 2018, where he is currently pursuing the M.S. degree with the College of Physics and Information Engineering. His research interests include computer vision and image processing.



Tiesong Zhao (Senior Member, IEEE) received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2006, and the Ph.D. degree in computer science from City University of Hong Kong, Hong Kong, in 2011.

He has served as a Research Associate with the Department of Computer Science, City University of Hong Kong, from 2011 to 2012, a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, from 2012 to 2013, and a Research Scientist with the Ubiquitous Multimedia Laboratory, The State University of New York at Buffalo, from 2014 to 2015. He is currently a Minjiang Distinguished Professor with the College of Physics and Information Engineering, Fuzhou University, China. His research interests include multimedia signal processing, coding, and quality assessment and transmission. Due to his contributions in video coding and transmission, he received the Fujian Science and Technology Award for Young Scholars in 2017. He has been serving as an Associate Editor for *IET Electronics Letters* since 2019.



Chia-Wen Lin (Fellow, IEEE) received the Ph.D. degree in electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2000. He was with the Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan, from 2000 to 2007. Prior to joining academia, he worked with the Information and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu, from 1992 to 2000. He is currently a Professor with the Department of Electrical Engineering, NTHU, and the Institute of Communications Engineering, NTHU. He is also the Deputy Director of the AI Research Center, NTHU. His research interests include image and video processing, computer vision, and video networking. He was a Steering Committee Member of the IEEE TRANSACTIONS ON MULTIMEDIA from 2014 to 2015. He was a Distinguished Lecturer of the IEEE Circuits and Systems Society from 2018 to 2019. He has served as the President for the Chinese Image Processing and Pattern Recognition Association, Taiwan, from 2010 to 2019. His papers won the Best Paper Award of IEEE VCIP 2015, Top 10% Paper Awards of IEEE MMSP 2013, and the Young Investigator Award of VCIP 2005. He received the Young Investigator Award presented by the Ministry of Science and Technology, Taiwan, in 2006. He was the Chair of the Multimedia Systems and Applications Technical Committee of the IEEE Circuits and Systems Society from 2013 to 2015. He has also served as the Technical Program Co-Chair for IEEE ICME 2010. He will be the General Co-Chair of IEEE VCIP 2018 and the Technical Program Co-Chair of IEEE ICIP 2019. He has also served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE MULTIMEDIA, and the *Journal of Visual Communication and Image Representation*.