IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 April 2023, accepted 7 May 2023, date of publication 10 May 2023, date of current version 22 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3274837

RESEARCH ARTICLE

Enhanced Multi-View Subspace Clustering via Twist Tensor Nuclear Norm and Constraint Propagation

WEI YAN[1,2], YU WANG[3], MENGXIN WANG[1], AND JUNJIE YANG[4,5]

[1]Guangzhou Institute of International Finance, Guangzhou University, Guangzhou 510006, China
[2]Zhuhai DeltaFit Technology Company Ltd., Zhuhai 519031, China
[3]Faculty of Finance, City University of Macau, Macau, China
[4]School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[5]Guangdong Key Laboratory of IoT Information Technology, Guangzhou 510006, China

Corresponding author: Junjie Yang (junjieyang@gdut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62003101, and in part by the Natural Science Foundation of Guangdong Province under Grant 2022A1515010181 and Grant 2023A1515011290.

ABSTRACT Multi-view subspace clustering (MVSC) can effectively group multi-view data distributed around several low-dimensional subspaces. Although encouraging results, most existing methods suffer from two typical limitations, resulting in clustering performance degradation. They ignore high-order correlations underlying the multi-view data, leading to degeneration of complementary power; in addition, they rely on much prior knowledge (e.g., pairwise constraints) for clustering enhancement. In this paper, a novel algorithm called Enhanced Multi-view Subspace Clustering (EMVSC) is proposed to address both limitations. EMVSC can effectively exploit high-order correlations and optimally use limited prior knowledge for better clustering performance. Specifically, EMVSC imposes twist tensor nuclear norm on multi-view tensor representation constructed by stacking view-specific self-representations; in addition, EMVSC exploits prior knowledge of pairwise constraints from whole dataset by employing constraint propagation, which propagates limited constraint knowledge from constrained samples to unconstrained samples. To efficiently optimize EMVSC, an extended intact augmented Lagrangian method is derived with good convergence. Experimental results on seven standard multi-view databases demonstrate its efficacy.

INDEX TERMS Multi-view clustering, low-rank tensor representation, tensor singular value decomposition (T-SVD), constrained clustering.

I. INTRODUCTION

Multi-view data are collected from multiple modals or descriptors [1]. For example, a person can be identified by modals: face, fingerprint, iris, and signature; image can be characterized by descriptors: SIFT, HOG, and LBP; and document can be written in languages: english, chinese, and japanese, etc. Multi-view data possess complementary information among views and can boost an algorithm's learning performance. In multi-view learning, multi-view clustering aims to group samples with different views into clusters, in such a way as to organize similar samples together and assign dissimilar ones to different groups. In general, multi-view clustering is superior to single-view clustering since it can well exploit the complementary information in different views [2], [3].

Multi-view subspace clustering (MVSC) is prototype method of multi-view clustering [4]; it can effectively group data points which are drawn from multiple linear sub-spaces. MVSC is derived from the fact that many real-world data in a cluster often approximately lie in a subspace, such as face images of a subject under varying illumination [5] and hand-written digit images with rotations and translations [6]. MVSC methods could be generally cast as the following three steps. One learns view-specific affinity matrices which indicate membership among samples; then an unified affinity

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang.

matrix is constructed, followed by applying the technique of spectral clustering on it. Technically, most methods build on the idea of self-expressiveness, which refers to data points can be represented as a linear combination of others from the same subspace.

Consider the required assumptions on correlation among views, existing MVSC methods can be roughly classified into two categories. One category is pairwise-based. Typically, Brbić and Kopriva [7] extended low-rank sparse subspace clustering model to multi-view clustering model and carried out two strategies (e.g., pairwise-consensus and centroid-consensus) for pursuing multi-view consensus. Abavisani and Patel [8] introduced another sparse and low-rank model which focuses on learning robust common subspace representation of views. Zhu and Peng [9] exploited the sparseness structure and low-rankness structure under deep neural network framework. For further improvement, Luo et al. [10] well considered consistent and specific structures. While empirical success, these methods represented correlation of multi-views in matrix space, resulting in sub-optimal representation. This motivated the emerge of another category (e.g., high-order correlation based) [11]. High-order correlation among views boost clustering performance and can only be exploited by tensor representation [12]. Zhang et al. [13] stacked self-representations into a third-order tensor, followed by imposing an unfolding-based tensor nuclear norm on it for pursuing low-rank structure. Albeit implement easily, it lacks of a clear physical meaning due to the used unfolding-based tensor nuclear norm. In addition, it puts the same weights on all ranks of a tensor is ineffective, resulting in sub-optimal clustering performance [14].

On the other hand, prior knowledge reveals cluster information and can be used to improve MVSC. Building on tensor representation, Xiao et al. [15] integrated prior knowledge (e.g., explicit labels, semantic similarities, and weak-domain cues) to tensor-based self-representations. Another similar idea can be found in [16], Zhang et al. encoded prior knowledge of labels as a constraint matrix and incorporated it into tensor-based MVSC. Recently, Tang et al. [17] achieved constrained tensor representations by exploiting expert knowledge in forms of two types of pairwise constraints: must-link constraints (i.e., samples from the same cluster) and cannot-link constraints (i.e., samples from different clusters). Despite effectiveness, these methods highly depend on the amount of prior information which is limited due to its high cost of collection. In other words, the unlabeled samples are not well considered, which will seriously degenerate their discriminating power, and likewise the clustering performance when little prior knowledge is available [18].

A. MOTIVATION AND CONTRIBUTION OF THIS PAPER

Being aware of the importance of modeling the high-order correlation and practical condition of limited prior knowledge, we propose an enhanced MVSC (EMVSC) method (see Fig. 1 for its illustration). EMVSC can effectively mine high-order correlation and well exploit prior information for clustering enhancement. Specifically, EMVSC resorts to the twist tensor nuclear norm (t-TNN) [19] and constraint propagation (CP) [20]. t-TNN contributes to the exploitation of the high-order correlation, leading to better representation than that obtained by the unfolding-based tensor nuclear norm; CP benefits collection of prior information of whole samples, helping EMVSC to obtain optimal clustering performance under condition of limited prior knowledge. In addition, CP helps EMVSC obtain discriminating view-specific representations. Technically, view-specific representation is constrained by a view-specific pairwise constraint, which allocates large values for data points from the same cluster and assigns smaller values for data points from different cluster. That is, EMVSC pulls data from the same cluster closer and push data from the different cluster further. Our contributions lie in three aspects:

- We propose an enhanced multiview subspace clustering method called EMVSC. Unlike previous methods, EMVSC does not require a large amount of prior info and it can perform accurate clustering with scarce prior knowledge. Technically, EMVSC utilizes constraint propagation for collecting constraint knowledge from whole data set. Specifically, it propagates constraints from constrained samples to unconstrained samples.
- To effectively optimize it, we design an extended intact augmented Lagrangian method with good convergence. In addition, we analysis the model and give its complexity.
- We conduct experiments on seven public multiview data sets. Moreover, we design the ablation study which provides a more detailed and fine-grained evaluation of the proposed techniques and their contributions, helping other researchers understand its strengths and weaknesses.

The rest of this paper is arranged as below. Section II introduces related work. Section III attaches notations and preliminaries. The proposed EMVSC, its optimization algorithm and convergence analysis are presented in Section IV. Section V provides experiments and discussion. Section VI concludes this paper.

II. RELATED WORKS

Low-rank representation (LRR) [22] is a subspace clustering method, assuming data lie within the union of multiple linear sub-spaces. Its objective function is:

$$\begin{aligned} &\min_{Z,E} \lambda\|E\|_{2,1} + \|Z\|_*, \\ &s.t.\ X = XZ + E, \end{aligned} \tag{1}$$

where X is given samples, Z represents low-rank representation, E denotes error matrices, $\|\cdot\|_*$ and $\|\cdot\|_{2,1}$ indicate nuclear norm and $l_{2,1}-$norm, respectively.

When it comes to multi-view clustering, LRR can be directly extended to fulfill this goal [7]. The extended

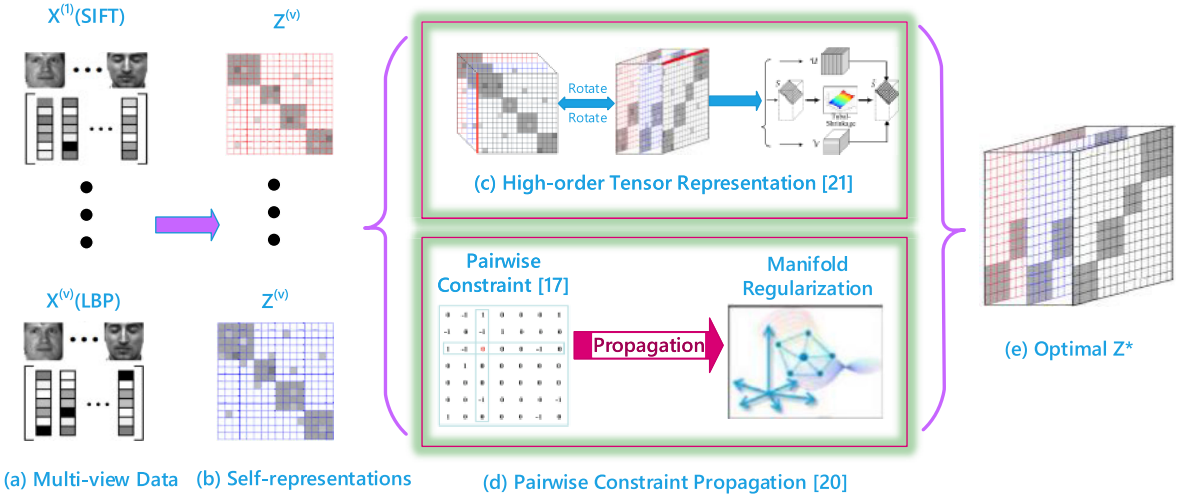

FIGURE 1. Illustration of proposed EMVSC. Given multi-view data, we aim to learning discriminating tensor representation Z^*. Specifically, we first learn view-specific self-representations, following by stacking these representations to constructing a tensor [21], which is restricted by both twist tensor nuclear norm [19] for exploiting complement information among views and view-specific manifold regularization for discriminating power. Note that our regularization is prior knowledge-driven (e.g., pairwise constraint [17]); in addition, given limited constraints we employ constraint propagation [20] for collecting constraints of whole data set.

objective function is:

$$\begin{aligned} &\min_{Z^{(v)},E} \sum_{v}^{V} (\|Z^{(v)}\|_* + \lambda_v \|E^{(v)}\|_{2,1}) \\ &s.t.\ X^{(v)} = X^{(v)} Z^{(v)} + E^{(v)}, \quad v = 1, 2, \ldots, V \\ &\quad E = [E^{(1)}; E^{(2)}; \ldots; E^{(V)}] \end{aligned} \tag{2}$$

where $X^{(v)}$ represents the $v-$th view data, $Z^{(v)}$ denotes the $v-$th view's low-rank representation, V is the number of views, $\{E^{(V)}\}$ is error matrix. Once obtained $\{Z^{(v)}\}_{v=1}^{V}$, the affinity matrix A is calculated by: $A = \frac{1}{V}\sum_{v=1}^{V}(|Z^{(v)}| + |Z^{(v)}|^T)/2$. One can apply the spectral clustering method [23] on A to get the clustering results.

Based on the above-mentioned model, existing methods can be organized into two categories: matrix-based methods and tensor-based methods.

A. MATRIX-BASED METHODS

These methods impose low-rank constraint on coefficient matrix. To this end, there are two effective schemes, i.e., nuclear norm (NN)-based and non-negative matrix factorization (NMF)-based. NN is used as a convex envelope of non-convex rank function for exploiting low-rank property of coefficient matrix. For example, Xia et al. in [24] imposed the NN on a multi-view shared transition probability matrix. NMF-based algorithms is capable of finding meaningful clustering structure for non-negative constraint imposed on coefficient matrix. For example, Liang et al. [25] used NMF for learning discriminant coefficient matrices and incorporated graph structure of data into the new model. Liang et al. [26] also proposed a prior knowledge-driven NMF model, using binary label matrix to constrain multi-view representation.

B. TENSOR-BASED METHODS

While effectiveness, we observe that the matrix-based methods share the same disadvantage that ignore high-order correlation among views. To address this limitation, tensor-based methods have been investigated. For example, Liu et al. [27] proposed a tensor framework based on high-order analogues of the matrix SVD and PCA, in which both heterogeneous and homogeneous information can be used to help clustering. Zhang et al. [13] presented a multi-view subspace clustering method, utilizing low-rank tensor constraint to exploit high-order correlation among views. Xie et al. [21] introduced a new 3-order tensor decomposition framework to pursuing consistency among multiple views. Recently, Wang et al. [28] proposed a tensor method which unifies tensor representation learning and consensus affinity matrix construction. Tang et al. [17] argued that the prior information should be well considered and that can be used to improve clustering performance. Guo et al. [1] devises a new surrogate of tensor rank, namely the tensor logarithmic Schatten-p norm (TLSpN), which is a nearly unbiased approximation of tensor rank and it fully considers the physical difference between singular values by the non-convex and non-linear penalty function, resulting in compact low-rank tensor representation and well exploitation of high-order correlation. Chen et al. [11] introduced a novel unified model termed Low-rank Tensor Based Proximity Learning (LTBPL) for multi-view clustering. Within this framework, multiple affinity representations are stacked in a low-rank tensor constrained by the t-SVD based weighted tensor nuclear norm to recover higher-order correlations among multiple views, and especially the weights information of singular values corresponding to multiple views is explicitly considered by assigning different contributions. Similarly, Guo et al. [14] proposed a new auto-weighted tensor nuclear norm (AWTNN) as

a low-rank tensor approximation for better capturing the high-order correlation. Especially, the AWTNN considers the difference between singular values through adaptive weights splitting during the AWTNN optimization procedure.

III. NOTATIONS AND PRELIMINARIES

Here, we attend to introduce some notations and preliminaries used throughout this paper. For notations, we use calligraphy letter (e.g., $\mathcal{X}$), capital letter (e.g., X), and lowercase letter (e.g., x) to represent tensor, matrix and vector, respectively. A N-way tensor is a multi-linear structure in $\mathbb{R}^{n_1\times n_2\times\ldots\times n_N}$. The Frobenius norm of $\mathcal{X}$ is $\|\mathcal{X}\|_F := (\sum_{i,j,k}|x_{ijk}|^2)^{\frac{1}{2}}$, and the l_1 norm of $\mathcal{X}$ is $\|\mathcal{X}\|_1 := \sum_{i,j,k}|x_{ijk}|$.

As for preliminaries, we begin with introduction of five block-based operators i.e., bcirc, bvec, bvfold, bdiag and bdfold, followed by tensor singular-value decomposition (t-SVD) [29], [30]. These operators and the t-SVD are essentials for comprehensive understanding of t-SVD-based tensor nuclear norm, which will also be described at last.

A. FIVE BLOCK-BASED OPERATORS

For $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$, the block circulant opetator (bcirc) is defined as:

$$bcirc(\mathcal{X}) := \begin{bmatrix} \mathcal{X}^{(1)} & \mathcal{X}^{(n_3)} & \ldots & \mathcal{X}^{(2)} \\ \mathcal{X}^{(2)} & \mathcal{X}^{(1)} & \ldots & \mathcal{X}^{(3)} \\ \vdots & \ddots & \ddots & \vdots \\ \mathcal{X}^{(n_3)} & \mathcal{X}^{(n_3-1)} & \ldots & \mathcal{X}^{(1)} \end{bmatrix}. \tag{3}$$

The block vectorizing operator (bvec) and its opposite operation (bvfold) are:

$$bvec(\mathcal{X}) := \begin{bmatrix} \mathcal{X}^{(1)} \\ \mathcal{X}^{(2)} \\ \vdots \\ \mathcal{X}^{(n_3)} \end{bmatrix}, \ bvfold(bvec(\mathcal{X})) = \mathcal{X}. \tag{4}$$

The block diag operator (bdiag) and its opposite operation (bdfold) are:

$$bdiag(\mathcal{X}) := \begin{bmatrix} \mathcal{X}^{(1)} & & \\ & \ddots & \\ & & \mathcal{X}^{(n_3)} \end{bmatrix}, \ bdfold(bdiag(\mathcal{X})) = \mathcal{X}. \tag{5}$$

B. TENSOR SINGULAR-VALUE DECOMPOSITION (T-SVD)

To understand t-SVD, we need five essential definitions, t-product, tensor transpose, identity tensor, orthogonal tensor, and f-diagonal tensor.

Definition 1 (t-Product): $\mathcal{M}$ is the t-product of $\mathcal{X} * \mathcal{Y}$.

$$\mathcal{M} = \mathcal{X} * \mathcal{Y} = bvfold\{bcirc(\mathcal{X})bvec(\mathcal{Y})\}, \tag{6}$$

t-product is analogous to matrix multiplication. The difference is that t-product replaces the multiplication operation between the elements with the circular convolution between the mode-3 fibers, that is:

$$\mathcal{M}(i,j,:) = \sum_{k=1}^{n_2} \mathcal{X}(i,k,:) \circ \mathcal{Y}(k,j,:), \tag{7}$$

where $\circ$ denotes the circular convolution between two tubes.

Definition 2 (Tensor Transpose): The transpose of $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is $\mathcal{X}^T \in \mathbb{R}^{n_2\times n_1\times n_3}$, transposing each frontal slice of $\mathcal{X}$, and then reversing the order of the transposed frontal slices 2 through n_3.

Definition 3 (Identity Tensor): $\mathcal{I} \in \mathbb{R}^{n_1\times n_1\times n_3}$ is the identity tensor, whose first frontal slice is the $n_1 \times n_1$ identity matrix and all other frontal slices are zero.

Definition 4 (Orthogonal Tensor): A tensor $\mathcal{Q}$ is orthogonal if

$$\mathcal{Q}^T * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^T = \mathcal{I}, \tag{8}$$

where $*$ is the t-product operation.

Definition 5 (f-Diagonal Tensor): A tensor is f-diagonal if each of its frontal slices is diagonal matrix. The t-product of two f-diagonal tensors is also f-diagonal. Specifically, $\mathcal{M} = \mathcal{X} * \mathcal{Y}$ is a f-diagonal tensor, and its diagonal tube fibers are

$$\mathcal{M}(i,i,:) = \mathcal{X}(i,i,:) \circ \mathcal{Y}(i,i,:), \ i = 1,\ldots,min(n_1,n_2). \tag{9}$$

With these definitions, the tensor singular value decomposition (t-SVD) of $\mathcal{X}$ is given by:

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \tag{10}$$

where $\mathcal{U}$ and $\mathcal{V}$ are orthogonal tensors. $\mathcal{S}$ is an f-diagonal tensor. We conclude the t-SVD in algorithm 1.

Algorithm 1 t-SVD [30]

Input: $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$;
Output: $\mathcal{U}, \mathcal{S}, \mathcal{V}$;
1: $\mathcal{X}_f = FFT(\mathcal{X}, [], 3)$;
2: **for** $k = 1 : n_3$ **do**
3: $\quad [U, \sum, V] = SVD(\mathcal{X}_f^{(k)})$;
4: $\quad \mathcal{U}_f^{(k)} = U, \mathcal{S}_f^{(k)} = \sum, \mathcal{V}_f^{(k)} = V$;
5: **end for**
6: $\mathcal{U} = iFFT(\mathcal{U}_f, [], 3), \mathcal{S} = iFFT(\mathcal{S}_f, [], 3), \mathcal{V} = iFFT(\mathcal{V}_f, [], 3)$;
7: **return** $\mathcal{U}, \mathcal{S}, \mathcal{V}$

C. T-SVD INDUCED TENSOR NUCLEAR NORM

Recall that t-SVD enables $\mathcal{X}$ to be written as a finite sum of outer product of matrices:

$$\mathcal{X} = \sum_{i=1}^{min(n_1,n_2)} \mathcal{U}(:,i,:) * \mathcal{S}(i,i,:) * \mathcal{V}(:,i,:)^T. \tag{11}$$

Definition 6 (Tensor Multi-Rank [29], [30], [31], [32]): The multi-rank of $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is a vector $r \in \mathbb{R}^{n_3\times 1}$ with the i-th element equal to the rank of the i-th frontal slice of $\mathcal{X}_f$.

Based on this definition, we can define the t-SVD induced tensor nuclear norm as:

$$\|\mathcal{X}\|_{\circledast} := \sum_{i=1}^{min(n_1,n_2)} \sum_{k=1}^{n_3} |\mathcal{S}_f(i,i,k)|. \tag{12}$$

It is a valid norm and the tightest convex relaxation to ℓ_1 norm of the tensor multi-rank. According to the unitary invariance of matrix nuclear norm, we have:

$$\|bdiag(\mathcal{X}_f)\|_* = \|bdiag(\mathcal{X}_f)\|_* = \|\mathcal{X}\|_{\circledast}. \tag{13}$$

Due to block circulant matrixes can be block diagonalized by using the Fourier transform (FFT), we have:

$$\begin{aligned}\|bdiag(\mathcal{X}_f)\|_* &= \|(F_{n_3} \otimes I_{n_1}) bcirc(\mathcal{X})(F_{n_3}^* \otimes I_{n_2})\|_* \\ &= \|bcirc(\mathcal{X})\|_*,\end{aligned} \tag{14}$$

where $\otimes$ is the Kronecker product, F_n is the Discrete Fourier Transform (DFT) matrix, and I_n is an identity matrix. Finally, we obtain:

$$\|\mathcal{X}\|_{\circledast} = \|bcirc(\mathcal{X})\|_* \tag{15}$$

In the above equation, $bcirc(\mathcal{X})$ preserves the relationship between entries, and $\|bcirc(\mathcal{X})\|_*$ exploits structure information of a tensor by comparing every row and every column of frontal slices over the third dimension.

IV. THE PROPOSED EMVSC

As aforementioned, most previous MVSC methods suffer from two disadvantages. Their accuracy meets the bottleneck when the amount of prior knowledge is limited. This limitation is fatal since it is nontrivial to collect much prior knowledge in practice. We address it through performing constraint propagation, which is a graph-based learning method. Another disadvantage is that they fail to effectively capture the high-order correlationship, resulting in a compromise of complementary power and thus accuracy of clustering. We resolve it by imposing the twist t-SVD-based tensor nuclear norm on tensor representation, which is constructed by stacking view-specific self-representation.

A. CONSTRAINT PROPAGATION (CP)

As for prior knowledge, we focus on pairwise constraints, including must-connect constraint and cannot-connect constraint. The must-connect constraint refers to relation of one pair of data points that belong to the same cluster, and otherwise the cannot-connect constraint. Compared to explicit label information, the pairwise constraints is relatively easy to collect. One can obtain these connect knowledge from three typical sources (e.g., labels of samples, domain knowledge [33] and data annotations [34]). Note that one cannot derive explicit the labels only from pairwise constraints, particularly for data with multi-class.

To collect pairwise constraints, we resort to constraint propagation (CP). if 10% of the whole data set bring constraint information, CP can predict that of the rest 90%. CP achieves this prediction by leveraging the power of assumptions on data sets. The typical assumption is intrinsic geometrical assumption: data points, that are close in the intrinsic geometry of the data distribution, share the same prior knowledge.

CP can be formulated as problem of two-class classification. The class label is either "+1" (must-connect constraint) or "−1" (cannot-connect constraint). Given $X = \{(x_i, l_i)\}_{i=1}^{n} \cup \{x_i\}_{i=n+1}^{N}$, where n is the number of samples with prior information, $l_i \in \{1, \ldots, C\}$, where C denotes the number of classes and $N-n$ is the number of samples without prior information. The must-connect constraints and cannot-connect constraints are $MC = \{(x_i, x_j) : l_i = l_j, , 1 \le i, j \le N\}$ and $CC = \{(x_i, x_j) : l_i \ne l_j, 1 \le i, j \le N\}$, respectively. Given initial matrix of pairwise constraint $P = \{p_{ij}\}_{N\times N}$, which is defined as:

$$p_{ij} = \begin{cases} +1, & \text{if } (x_i, x_j) \in MC \\ -1, & \text{if } (x_i, x_j) \in CC \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

CP aim to obtain propagated result $F = \{f_{ij}\}_{N\times N}$. $|f_{ij}|$ denotes score of (x_i, x_j) on pairwise constraint. More concretely, if $f_{ij} > 0$, f_{ij} is the similarity between x_i and x_j, where the larger the value of f_{ij} is, the more x_i is similar to x_j, thus it is a must-connect. If $f_{ij} < 0$, and the smaller the value of f_{ij} is, the more x_i is dissimilar to x_j, this is known as a cannot-connect.

CP consists of the following six steps.

1) Construct a graph by defining its weight matrix $W = \{w_{ij}\}_{N\times N}$ as follows:

$$w_{ij} = \begin{cases} \dfrac{a(x_i, x_j)}{\sqrt{a(x_i, x_i)}\sqrt{a(x_j, x_j)}}, & \text{if } x_i \in \mathcal{N}_p(x_j) \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

where $\mathcal{N}_p(x_i)$ denotes the p-nearest neighbor set of the data point x_j. $A = a(x_i, x_j)_{N\times N}$ is the kernel matrix defined on the data X. In particular, $a(x_i, x_j) = \exp(-\|x_i - x_j\|^2/t)$, and t denotes the bandwidth parameter. Set $W = (W + W^T)/2$ to ensure W is symmetric.

2) Compute the normalized graph Laplacian matrix $L = I - D^{-1/2}WD^{-1/2}$, where $D = [d_{ii}]_{N\times N}$ is a diagonal matrix with $d_{ii} = \sum_j W_{ij}$.
3) Iterate $F_v(t+1) = \alpha L F_v(t) + (1-\alpha)P$ for the vertical constraint propagation until it converges, where α is a parameter ranging from 0 to 1. The proof of the convergence can be found in [35].
4) Iterate $F_h(t+1) = \alpha F_h(t)L + (1-\alpha)F_v^*$ for the horizontal constraint propagation until it converges, where F_v^* is the limit of $\{F_v(t)\}$.
5) $F^* = F_h^*$ is the final representation of the propagated dual connected constraints, where F_h^* is the limit of $\{F_h(t)\}$.
6) With the above F^* and the original weight matrix W, we construct a new weight matrix $\widetilde{W} = \{\tilde{w}_{ij}\}_{N\times N}$ by using following operation (18), which allots data from the same cluster big weight and data from the different

cluster small weight [20].

$$\tilde{w}_{ij} = \begin{cases} 1 - \left(1 - f_{ij}^*\right)\left(1 - w_{ij}\right), & \text{if } f_{ij}^* \geq 0 \\ \left(1 + f_{ij}^*\right) w_{ij}, & \text{if } f_{ij}^* < 0. \end{cases} \tag{18}$$

Consider the new weight matrix, we have following observations. $\widetilde{W}$ directly encodes the constraint information of the whole data set and indirectly respect the local geometric information. In $\widetilde{W}$, if a pair of data points belong to the same class, the corresponding weight of the edge connecting them is given a much larger value, otherwise the weight is assigned a much smaller value.

Once obtained $\widetilde{W}$, we use it to construct the prior-driven graph Laplacian matrix, which is used as a regularization term to constrain the self-representation learning. That is,

$$\begin{aligned} \phi(Z) &= \frac{1}{2}\sum_{i,j=1}^{N} \|z_i - z_j\|^2 \tilde{w}_{ij} \\ &= tr(ZDZ^T) - \lambda Tr(ZWZ^T) \\ &= tr(ZLZ^T) \end{aligned} \tag{19}$$

where $tr(\cdot)$ is the trace of a matrix and D denotes a diagonal matrix whose entries are column sums of W, $d_{ii} = \sum_j W_{ij}$. Denote $L = D - W$.

If input data x_i and x_j come from the same class and $\tilde{w}_{ij}$ will be large (the definition of W mentioned above). That means, minimizing $\phi(Z)$ tends to have a small Euclidean distance between z_i and z_j, pushing z_i close to z_j. Here, we use the Euclidean distance for measuring the "dissimilarity" between self-representations of two points based on works [7]. If x_i and x_j come from different classes and $\tilde{w}_{ij}$ will be small. Therefore, z_i and z_j will be much farther. In conclude, the minimization of $\phi(Z)$ can produce points with the same label appear together and points from different classes occur much farther to each other in the obtained self-representation.

B. TWIST TENSOR NUCLEAR NORM (T-TNN)

To exploit high-order correlation, one can resort to low-rank tensor learning. Here, tensor learning means to stack self-representations of views into a tensor, and then apply tensor nuclear norm (TNN) on it. Some use the generalized TNN (g-TNN). In fact, the g-TNN is sub-optimal since this unfolding-based tensor nuclear norm suffers from the loss of representation power, and thus leading to subspace clustering performance degradation [19].

Instead, we find twist tensor nuclear norm (t-TNN), containing the twist operation (TO)[1] and the t-SVD based tensor nuclear norm (tSVD-TNN) is optimal. We first execute TO on the stacked tensor representation, obtaining the twist tensor. Twist tensor is vital, and it helps to preserve the relationship between samples [21]. Notice that g-TNN based methods do not involve the TO process, resulting in sub-optimal. Once obtained the twist tensor representation for multi-view, we impose tSVD-TNN on it for exploiting high-order correlation among views.

Both u-TNN and tSVD-TNN are capable of exploring the high-order correlation among views. However, tSVD-TNN is superior to the former due to the following two facts.

1) Tensor rank is computationally intractable. In fact, tSVD-TNN has been proven to be the tightest convex relaxation to the l-norm of tensor multi-rank [30].
2) With the rotation process, each frontal slice of rotated tensor exploits information among views in the Fourier domain. By this way, tSVD-TNN optimally depicts the complicated relationship between views [21].

C. MODEL OF EMVSC

We formulate EMVSC as

$$\begin{aligned} \min_{Z^{(v)},E} \ & \|\mathcal{Z}\|_{\circledast} + \alpha\|E\|_{2,1} + \beta\sum_{v=1}^{V} tr(Z^{(v)}L^{(v)}Z^{(v)^T}) \\ s.t. \ & X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, v = 1, 2, \ldots, V \\ & \mathcal{Z} = \Phi(Z^{(1)}, Z^{(2)}, \ldots, Z^{(V)}) \\ & E = [E^{(1)}; E^{(2)}; \ldots; E^{(V)}], \end{aligned} \tag{20}$$

where $\|\cdot\|_{\circledast}$ is the t-TNN, $\mathcal{Z}$ is tensor constructed by self-representations, $Z^{(v)}$ is the self-representation of view v, $L^{(v)}$ is the refined graph matrix of view v, α and β are parameters, and E is error matrix. We construct E by vertically concatenating the error matrices $\{E^{(V)}\}$. We assume natural corruptions are always sample-specific. This assumption is fulfilled by using the $l_{2,1}$-norm. $\Phi(\cdot)$ is a operation, constructing the 3-mode tensor $\mathcal{Z}$ by merging representations $Z^{(v)}$. This operation is equipped with the following equation:

$$\Phi_{(v)}^{-1}(\mathcal{Z}) = Z^{(v)}, \tag{21}$$

where $\Phi_{(v)}^{-1}(\cdot)$ is the inverse function of $\Phi(\cdot)$ with respect to the v-th frontal slice.

D. OPTIMIZATION ALGORITHM

To effectively solve this constrained problem, we choose the strategy of the inexact augmented Lagrange multiplier (iALM) [36], which reformulates a constrained problem into an unconstrained one through the technique of auxiliary variable. iALM turns complex problem into easy-solving sub-problems.

For reformulation, we replace $Z^{(v)}$ and $\mathcal{Z}$ with auxiliary matrix variables $Q^{(v)}$ and $\mathcal{G}$, respectively. Then, we have the reformulated problem:

$$\begin{aligned} &\mathcal{L}(Z^{(1)}, \ldots, Z^{(V)}; E^{(1)}, \ldots, E^{(V)}; Q^{(1)}, \ldots, Q^{(V)}; \mathcal{G}) \\ &= \alpha\|E\|_{2,1} + \beta\sum_{v=1}^{V} tr(Q^{(v)}L^{(v)}Q^{(v)^T}) \\ &\quad + (\langle B_v, Q^{(v)} - Z^{(v)}\rangle + \frac{\mu_2}{2}\|Q^{(v)} - Z^{(v)}\|_F^2) \end{aligned}$$

[1]The tensor twist can be achieved by using the command "shiftdim" in Matlab.

$$+ \|\mathcal{G}\|_{\overrightarrow{\circledast}} + \sum_{v=1}^{V}(\langle Y_v, X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\rangle + \frac{\mu_1}{2}\|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\|_F^2) + \langle \mathcal{W}, \mathcal{Z} - \mathcal{G}\rangle + \frac{\rho}{2}\|\mathcal{Z} - \mathcal{G}\|_F^2 \quad (22)$$

where Y_v, B_v, and $\mathcal{W}$ are Lagrange multipliers; μ_1, μ_2, and ρ are penalty parameters, which can be adjusted by using adaptive updating strategy,

For optimization, we come to solve the following four subproblems.

$Z^{(v)}$-Subproblem: With fixed E, $Q^{(v)}$, and $\mathcal{G}$, as well as $\Phi_{(v)}^{-1}(\mathcal{W}) = W^{(v)}$ and $\Phi_{(v)}^{-1}(\mathcal{G}) = G^{(v)}$, we have the following subproblem:

$$\min_{Z^{(v)}} \langle Y_v, X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\rangle + \frac{\mu_1}{2}\|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\|_F^2 + \langle B_v, Q^{(v)} - Z^{(v)}\rangle + \frac{\mu_2}{2}\|Q^{(v)} - Z^{(v)}\|_F^2 + \langle W^{(v)}, Z^{(v)} - G^{(v)}\rangle + \frac{\rho}{2}\|Z^{(v)} - G^{(v)}\|_F^2 \quad (23)$$

We set the derivative of the above equation to zero, and obtain the closed-form solution:

$$Z^{(v)*} = (\rho I + \mu_1 X^{(v)^T} X^{(v)})^{-1} \times (X^{(v)^T} Y_v + \mu_1 X^{(v)^T} X^{(v)} - \mu_1 X^{(v)^T} E^{(v)} - W^{(v)} + \rho G^{(v)} + B_v + \mu_2 Q^{(v)}) \quad (24)$$

$E^{(v)}$-Subproblem: With fixed $Z^{(v)}$, we have

$$E^* = \min_E \frac{\alpha}{\mu_1}\|E\|_{2,1} + \frac{1}{2}\|E - D\|_F^2 \quad (25)$$

where D is built as this way of vertically concatenating the matrices $X^{(v)} - X^{(v)}Z^{(v)} + (1/\mu_1)Y_v$ together along column. Its solution is:

$$E_{:,i}^* = \begin{cases} \dfrac{\|D_{:,i}\|_2 - \frac{\alpha}{\mu_1}}{\|D_{:,i}\|_2} D_{:,i}, & \|D_{:,i}\|_2 > \dfrac{\alpha}{\mu_1} \\ 0, & otherwise \end{cases} \quad (26)$$

where $D_{:,i}$ denotes the ith column of the matrix D.

Q-Subproblem: Given $Z^{(v)}$ and $L^{(v)}$, the closed-form solution to this subproblem is:

$$Q^* = \min_Q \beta tr(Q^{(v)}L^{(v)}Q^{(v)^T}) + \langle B_v, Q^{(v)} - Z^{(v)}\rangle + \frac{\mu_2}{2}\|Q^{(v)} - Z^{(v)}\|_F^2 = (\mu_2 Z^{(v)} - B_v)(2\beta L^{(v)} + \mu_2 I)^{-1} \quad (27)$$

$\mathcal{G}$-Subproblem: Given $\{Z^{(v)}\}_{v=1}^V$, we attend to the following subproblem:

$$\mathcal{G}^* = \min_{\mathcal{G}} \|\mathcal{G}\|_{\overrightarrow{\circledast}} + \frac{\rho}{2}\|\mathcal{G} - (\mathcal{Z} + \frac{1}{\rho}\mathcal{W})\|_F^2 \quad (28)$$

To solve it, we need the following Theorem 1.

Theorem 1: For $\tau > 0$, the globally optimal solution to this problem:

$$\min_{\mathcal{G}} \tau\|\mathcal{G}\|_{\overrightarrow{\circledast}} + \frac{1}{2}\|\mathcal{G} - \mathcal{F}\|_F^2 \quad (29)$$

can be calculated by using the tensor tubal-shrinkage operator:

$$\mathcal{G} = \mathcal{C}_{n_3\tau}(\mathcal{F}) = \mathcal{U} * \mathcal{C}_{n_3\tau}(\mathcal{S}) * \mathcal{V}^T \quad (30)$$

where $\mathcal{F} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ and $\mathcal{C}_{n_3\tau}(\mathcal{S}) = \mathcal{S} * \mathcal{J}$, herein, $\mathcal{J}$ is a f-diagonal tensor whose diagonal element in the Fourier domain is $\mathcal{J}_f(i,i,j) = (1 - [n_3\tau/\mathcal{S}_j^{(j)}(i,i)])_+$.
The proof of the above theorem can be found in [21]. For better understanding, we conclude the optimization of $\mathcal{G}$ in algorithm 2.

Algorithm 2 Optimization of $\mathcal{G}$

Input: $\mathcal{F} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$; $\tau > 0$
Output: $\mathcal{G}$;

```
F_f = fft(F, [], 3); τ' = n_3 τ
for k = 1 : n_3 do
    [U_f^k, S_f^k, V_f^k] = SVD(X_f^(k));
    J_f^k = diag{(1 − τ'/S_f^(k)(i,i))_+}, i = 1, ..., min(n_1, n_2);
    S_{f,τ'}^k = S_f^k J_f^k;
    G_f^k = U_f^k S_{f,τ'}^k V_f^k;
end for
G = ifft(G_f, [], 3);
return G
```

The Lagrange multipliers and the penalty parameters could be updated as:

$$Y_v^* = Y_v + \mu_1(X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}), \quad (31)$$
$$B_v^* = Y_v + \mu_2(Q^{(v)} - Z^{(v)}), \quad (32)$$
$$\mathcal{W}^* = \mathcal{W} + \rho(\mathcal{Z} - \mathcal{G}), \quad (33)$$
$$\mu_i = \min(\eta\mu_i, \mu_{max}), i = 1, 2, \quad (34)$$
$$\rho = \min(\eta\rho, \rho_{max}). \quad (35)$$

Finally, this optimization is summarized in algorithm 3.

E. CONVERGENCE ANALYSIS

We extend the inexact ALM (iALM) to solve our problem. We should note that the problem is non-smooth, it would be not easy to theoretically prove the convergence, and it deserves in its own right a proper study of proof, which is of course not the focus of this paper.

In fact, iALM is only proved to be convergence when the number of blocks is smaller than three [36]. However, our objective function has four blocks (i.e., $\{Z^{(v)}\}_{v=1}^V$, $\{E^{(v)}\}_{v=1}^V$, $\{Q^{(v)}\}_{v=1}^V$, and $\mathcal{G}$). Nevertheless, we believe that the convexity of the Lagrange function could guarantee the empirical validity to some extent, verified by pioneer work [21]. We will show this empirical convergence in Section V-E.

TABLE 1. Characteristics of the datasets.

Datasets	Views	Classes	Size	Type
Politicsie	9	7	348	social media
3Sources	3	6	169	news
Extented YaleB	3	38	2414	face image
Prokaryotic	2	4	551	prokaryotic
Flowers	3	68	1360	flower
Scene-15	3	15	4485	scene
MITIndoor	4	67	5360	scene

V. EXPERIMENTS AND DISCUSSION

The effectiveness of proposed EMVSC is evaluated with extensive experiments. First, we go through the datasets for test, metrics for evaluation, and related methods for comparison. Second, we compare EMVSC with the state-of-the-art algorithms and analyze the impact of parameters on clustering performance. Third, we investigate its convergence and complexity. Fourth, we conduct ablation study and investigate different levels of prior information.

A. DATASETS AND METRICS

We test EMVSC on seven datasets, which are about social media, news, face image, prokaryotic species, and scene. The characteristics of these datasets are summarized in Table 1. Note that both Scene-15 and MITIndoor are large-scale data sets. We briefly introduce these datasets as follows.

Algorithm 3 EMVSC

Input: $X^{(1)}, \ldots, X^{(V)}, L^{(1)}, \ldots, L^{(V)}, \alpha, \beta$
Output: Clustering results C
Initialize:$Z^{(v)} = E^{(v)} = Q^{(v)} = Y_v = B_v = 0, i = 1, \ldots, V$;
$\mathcal{G} = \mathcal{W} = 0$;
$\mu_1 = \mu_2 = \rho = 10^{-5}, \eta = 2, \mu_{max} = \rho_{max} = 10^{10}, \varepsilon = 10^{-7}$;
while not converge **do**
  Update $\{Z^{(v)}\}_{v=1}^{V}$ by 24;
  Update E by 26;
  Update $\{Q^{(v)}\}_{v=1}^{V}$ by 27;
  Obtain $\mathcal{Z} = \Phi(Z^{(1)}, Z^{(2)}, \ldots, Z^{(V)})$;
  Update $\mathcal{G}$ via Algorithm 2;
  Update Y_v, B_v, $\mathcal{W}$, μ_1, μ_2, and ρ by 32-35, respectively;
  $(G^{(1)}, \ldots, G^{(V)}) = \Phi^{-1}(\mathcal{G})$;
  Check the convergence conditions: $\|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\|_\infty < \varepsilon$ and $\|Z^{(v)} - G^{(v)}\|_\infty < \varepsilon$;
end while
Obtain affinity matrix $A = \frac{1}{V}\sum_{v=1}^{V} |Z^{(v)}| + |Z^{(v)^T}|$;
Apply spectral clustering on matrix A;
return Clustering result C.

Politicsie[2] consists of 348 Irish politicians and political organizations belong to 7 groups in terms of their affiliation.

3sources[3] is a new multi-view text dataset, collected from three well-known online news sources: BBC, Reuters, and The Guardian. Each story was manually annotated with one or more of the six topical labels: business, entertainment, health, politics, sport, technology.

Extended YaleB[4] has 38 individuals, each of which has 64 near frontal images captured under different illumination. the first ten individuals, with 640 images in total, are used in experiments.

Prokaryotic is about prokaryotic species described with two views [7]. One view is textual data describing prokaryotic species. Another view constitutes the proteome composition, encoded as relative frequencies of amino acids, and the gene repertoire, encoded as presence/absence indicators of gene families in a genome.

Flowers[5] contains 1360 samples in total with 17 flower categories. We use three kinds of features: color, texture, and shape.

Scene-15[6] is about 15 categories, including kitchen, living room, bedroom, etc. It contains a wide range of outdoor and indoor scene environments [37].

MITIndoor [38] contains 67 different categories. We use a subset of it (5360 images in total) for test. We extracted four kinds of image features on this dataset: PHOW [39], PRI-CoLBP [40], CENTRIST [41] and VGG-VD [42].

Following [43], five standard evaluation metrics, including accuracy (ACC), Normalized Mutual Information (NMI), Adjusted Rank (AR), F-score, and Precision, are used for evaluation. Each value varies from 0 to 1. In general, the larger the value, the better the clustering quality, and vice versa. For each experiment, we run 10 trials and report their average results.

B. COMPARISONS WITH STATE-OF-THE-ART ALGORITHMS

To verify effectiveness, we compare our EMVSC with eight state-of-the-art methods, including two matrix-based methods and six tensor-based methods. All tensor-based methods can exploit high-order correlation among views. For comprehensive comparison, we compare our EMVSC with one graph-based tensor method GLTA, and three prior knowledge driven tensor methods, including PMVSR, TMSRL, and CTLR. We briefly introduce these methods as follows.

1) PMLRSSC [7], a pairwise consensus-guided multi-view low-rank sparse subspace clustering algorithm, which exploits pairwise correlation between views.
2) CMLRSSC [7], a centroid-driven multi-view low-rank sparse subspace clustering, which learns common representation of views.
3) t-SVD-MSC [21], a tensor-based multi-view low-rank subspace clustering method, which utilizes high-order correlation among views by using tensor nuclear norm.

[2]http://mlg.ucd.ie/aggregation/
[3]http://mlg.ucd.ie/datasets/3sources.html
[4]http://vision.ucsd.edu/ leekc/ExtYaleDatabase/ExtYaleB.html
[5]http://www.robots.ox.ac.uk/vgg/data/flowers/
[6]http://www-cvr.ai.uiuc.edu/poncegrp/data/

TABLE 2. Clustering results on Politicsie and 3sources with 10% prior knowledge of pairwise constraints.

Dataset	Politicsie					3sources				
Method	ACC	NMI	AR	F-score	Precision	ACC	NMI	AR	F-score	Precision
PMLRSSC	0.542	0.432	0.295	0.453	0.505	0.583	0.630	0.455	0.557	0.635
CMLRSSC	0.525	0.435	0.288	0.434	0.548	0.598	0.636	0.468	0.560	0.648
t-SVD-MSC	0.882	0.821	0.877	0.908	0.915	0.782	0.677	0.656	0.748	0.685
JLMVC	0.896	0.828	0.885	0.912	0.923	0.846	0.728	0.685	0.761	0.839
GLTA	0.905	0.831	0.896	0.887	0.936	0.859	0.743	0.703	0.755	0.847
PMVSR	0.922	0.928	0.915	0.934	0.938	0.851	0.823	0.813	0.788	0.820
TMSRL	0.894	0.869	0.886	0.909	0.924	0.865	0.853	0.826	0.805	0.835
CTRL	0.941	0.939	0.927	0.938	0.949	0.882	0.875	0.841	0.828	0.848
EMVSC	**0.969**	**0.958**	**0.946**	**0.967**	**0.968**	**0.958**	**0.905**	**0.923**	**0.935**	**0.924**

TABLE 3. Clustering results on Extended YaleB and Prokaryotic with 10% prior knowledge of pairwise constraints.

Dataset	Extended YaleB					Prokaryotic				
Method	ACC	NMI	AR	F-score	Precision	ACC	NMI	AR	F-score	Precision
PMLRSSC	0.216	0.187	0.068	0.255	0.141	0.437	0.437	0.398	0.591	0.725
CMLRSSC	0.228	0.206	0.268	0.160	0.148	0.424	0.424	0.384	0.582	0.712
t-SVD-MSC	0.652	0.667	0.500	0.550	0.514	0.523	0.197	0.137	0.486	0.474
JLMVC	0.606	0.628	0.665	0.554	0.545	0.753	0.485	0.505	0.675	0.662
GLTA	0.614	0.631	0.439	0.497	0.473	0.631	0.358	0.285	0.540	0.675
PMVSR	0.869	0.853	0.813	0.875	0.827	0.544	0.425	0.456	0.529	0.502
TMSRL	0.809	0.792	0.803	0.778	0.795	0.538	0.418	0.445	0.502	0.488
CTRL	0.823	0.818	0.835	0.836	0.848	0.565	0.524	0.485	0.569	0.522
EMVSC	**0.905**	**0.912**	**0.906**	**0.918**	**0.902**	**0.794**	**0.528**	**0.519**	**0.748**	**0.797**

4) JLMVC [44], it exploits high-order correlation as well as non-linear structure of views.
5) GLTA [45], it preserves high-order correlation and local graph structure of views.
6) PMVSR [15], it extracts high-order correlation and uses prior knowledge to constrain view-specific representation.
7) TMSRL [16], it mines high-order correlation and utilizes prior knowledge in terms of must-link information.
8) CTLR [17], it exploits high-order correlation and incorporates pairwise constraint into view-specific representation learning.

For our proposed EMVSC, we need to tune two parameters, noisy parameter α and constraint parameter β. We empirically find values in the range of $[0.01, 0.03, \ldots, 0.09]$ and $[0.1, 0.3, \ldots, 0.9]$, respectively. More details about the parameters analysis will be discussed in Section V-D.

For the parameters of compared methods, we keep the same setting as that in their original papers. For PMLRSSC and CMLRSSC, we use the strategy of sum-to-one, that is the sum of low-rank parameter and that of sparsity parameter is equal to one. And both are chosen from [0.1, 0.9]. As for t-SVD-MSC, the trade-off parameters are set within the range [0.1, 2]. As for JLMVC, the low-rank parameter and the consensus parameter are set within [0.001, 0.005, 0.01, 0.05, 0.1, 0.3, 0.5, 0.7] and [0.001, 0.005, 0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 1], respectively. As for GLTA, the number of the nearest neighbors is 5 and the parameter of weight penalty is set as 10. As for PMVSR, the parameter is empirically selected from $[1.1, 1.3, \ldots, 2.3]$. As for CTLR, the only one parameter is selected from $[10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}]$. For graph-based

TABLE 4. Clustering results on Flowers and Scene-15 with 10% prior knowledge of pairwise constraints.

Dataset	Flowers					Scene-15				
Method	ACC	NMI	AR	F-score	Precision	ACC	NMI	AR	F-score	Precision
PMLRSSC	0.517	0.637	0.348	0.451	0.327	0.401	0.425	0.495	0.388	0.332
CMLRSSC	0.525	0.628	0.375	0.485	0.318	0.420	0.441	0.484	0.391	0.304
t-SVD-MSC	0.843	0.907	0.787	0.782	0.764	0.812	0.858	0.775	0.780	0.745
JLMVC	0.913	0.915	0.904	0.905	0.881	0.802	0.822	0.819	0.801	0.829
GLTA	0.865	0.882	0.828	0.802	0.845	0.842	0.865	0.808	0.829	0.816
PMVSR	0.848	0.855	0.798	0.808	0.772	0.857	0.883	0.853	0.832	0.801
TMSRL	0.851	0.829	0.806	0.793	0.819	0.825	0.878	0.846	0.817	0.830
CTRL	0.837	0.806	0.819	0.768	0.782	0.818	0.826	0.819	0.808	0.834
EMVSC	**0.935**	**0.944**	**0.928**	**0.936**	**0.912**	**0.868**	**0.918**	**0.909**	**0.855**	**0.889**

TABLE 5. Clustering results on MITindoor with 10% prior knowledge of pairwise constraints.

Method	ACC	NMI	AR	F-score	Precision
PMLRSSC	0.423	0.552	0.288	0.293	0.289
CMLRSSC	0.405	0.518	0.230	0.244	0.218
t-SVD-MSC	0.722	0.788	0.607	0.608	0.575
JLMVC	0.744	0.704	0.621	0.617	0.648
GLTA	0.752	0.717	0.658	0.639	0.669
PMVSR	0.758	0.761	0.685	0.714	0.738
TMSRL	0.768	0.823	0.667	0.668	0.684
CTRL	0.764	0.837	0.685	0.691	0.645
EMVSC	**0.802**	**0.858**	**0.835**	**0.825**	**0.855**

methods, following [45], the number of k-nearest neighbor is set at 5 in all experiments.

Recall that our methods aims to perform accurate clustering when the available prior information is limited. Consider this motivation, we use the same lowest amount of prior information (10%) as that in [15], [16], and [17] for experiments. In experiments, we used the given label from data set to construct pairwise constraint.

C. RESULTS AND DISCUSSION

From Table 2-5, we draw conclusions listed as follows.

1) The proposed EMVSC consistently achieves the best results over all seven datasets. Take the 3sources dataset as an example, our EMVSC improves 7.6%, 3.0%, 8.2%, 10.7%, 7.6% with respect to five measures over the second-best method CTRL. EMVSC directly uses limited pairwise constraint and indirectly exploit graph structure to obtain pairwise constraint for unconstrained samples.
2) EMVSC is always superior to three state-of-the-art prior-driven methods (i.e., PMVSR, TMSRL and CTRL), which is owing to three reasons. First, EMVSC considers both samples are with prior information and those samples are without prior knowledge, while the compared methods only focus samples which carry constraints; Second, EMVSC utilizes local graph structure of data in an indirect way. Specifically, EMVSC uses graph structure to perform constraint propagation, whose results are then used to constrain view-specific representation learning; Third, EMVSC obtains much discriminating power since it pulls samples from the same cluster closer and pushes samples from the different cluster farther.
3) EMVSC performs better than the state-of-the-art graph-driven tensor method (i.e., GLTA) as well as the kernel-guided tensor method (i.e., JLMVC), indicating the enhancement of using prior knowledge. Specifically, EMVSC use graph structure to exploit pairwise constraint of the whole data set which are used to constrain the view-specific representations. In other words, EMVSC takes advantage of both prior knowledge and graph structure knowledge.
4) EMVSC's performance is better than both matrix-based clustering methods (i.e., PMLRSSC and CMLRSSC), intuitively showing the effectiveness of the twist tensor nuclear norm for utilizing high-order correlationship among views.

D. PARAMETER ANALYSIS

We analyze α with fixed β, and vice verse. The key observation is that EMVSC always performs better than the compared methods, especially the powerful competitor t-SVD-MSC, indicating the stability of EMVSC. For page limit, we show

TABLE 6. Complexity and running time on seven datasets (in seconds).

Method	Complexity	Politicsie	3Sources	YaleB	Prokaryotic	Flowers	Scene-15	MITindoor
PMVSR	$\mathcal{O}(\mathrm{TVN}^3)$	39.89	47.04	96.45	58.66	94.04	2108.12	3820.25
TMSRL	$\mathcal{O}(\mathrm{dN}^2 + \mathrm{N}^3)$	33.54	50.71	81.75	50.08	100.41	1760.37	3532.41
CTRL	$\mathcal{O}(\mathrm{VN}^3)$	33.73	43.89	89.78	46.75	90.74	2052.58	3617.69
EMVSC	$\mathcal{O}(\mathrm{N}^3) + \mathcal{O}(\mathrm{T}(2\mathrm{N}^2\mathrm{Vlog}(\mathrm{N}))) + \mathrm{PN}^2$	50.58	69.48	215.75	129.76	144.67	2294.25	4052.47

[1] T is number of iterations.
[2] V is the number of views.
[3] N is the number of samples.
[4] d is the number of dimension.

TABLE 7. Accuracy comparisons among different variants on seven datasets.

Dataset	Politicsie	3Sources	Extented YaleB	Prokaryotic	Flowers	Scene-15	MITIndoor
Our EMVSC	0.969	0.958	0.905	0.794	0.935	0.868	0.802
w/o twist operation	0.516	0.528	0.284	0.416	0.549	0.473	0.385
w/o constraint propagation	0.948	0.922	0.838	0.597	0.879	0.818	0.725
TLSpNM-MSC[1]	0.927	0.908	0.816	0.545	0.895	0.987	0.959

[1] w/o twist refers to EMVSC does not use the twist operation.
[2] w/o constraint propagation refers to EMVSC does not perform constraint propagation.

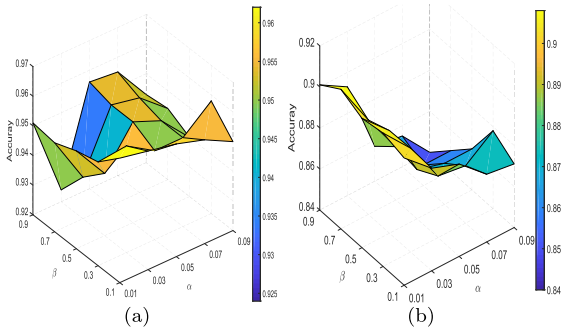

FIGURE 2. Parameters tuning (α and β) in terms of accuracy on two datasets. (a) 3sources (b) Extended YaleB.

accuracy results on two typical datasets (i.e., 3sources and Extended YaleB) in Fig. 2. Specifically, For 3sources, α is best set around 0.05, and β is best chosen around 0.7; For Extended YaleB, α is best selected around 0.03, and β is best adopted around 0.9.

E. EMPIRICAL CONVERGENCE

EMVSC has a good convergence property and converges fast on all data sets. We show the empirical convergence results on four data sets, including 3sources, Extend YaleB, Flowers, and Scene-15 in Fig. 3. Following [21], we define two error terms (i.e., reconstruction error (RE) and match error (ME)).

$$RE = \frac{1}{V}\sum_{v=1}^{V} ||X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}||_{\infty} \tag{36}$$

$$ME = \frac{1}{V}\sum_{v=1}^{V} ||Z^{(v)} - G^{(v)}||_{\infty} \tag{37}$$

Consider complexity and running time, we give results in Table 6.

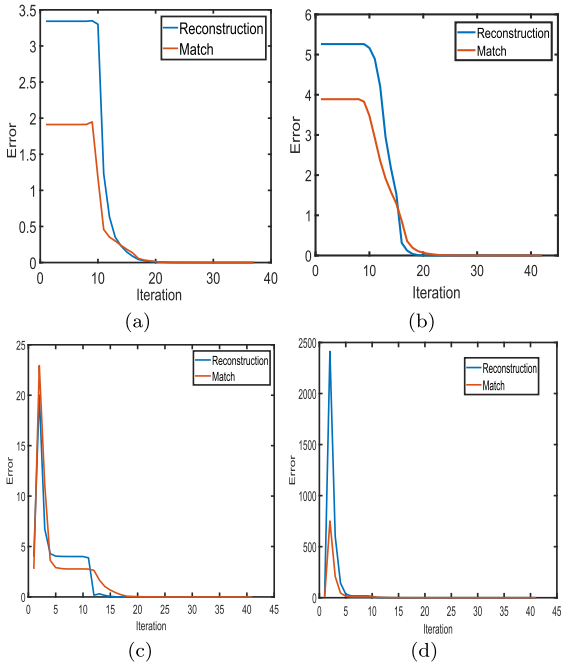

FIGURE 3. Illustration of the convergence curve of EMVSC over four datasets. (a) 3sources (b) Extended YaleB (c) Flowers (d) Scene-15.

F. ABLATION STUDY

Our EMVSC simultaneously exploits the high-order correlationship among view through using the twist tensor nuclear norm and incorporates the prior knowledge of sample in terms of pairwise constraint. We conduct extensive ablation studies to examine the effectiveness of each module of EMVSC. The TLSpNM-MSC [1] is compared for two reasons: (1) it relates to our work and it is the-state-of-art; (2) In contrast to EMVSC, TLSpNM-MSC exploits high-order-correlationship in a nonconvex way. Specifically,

TABLE 8. Accuracy results of Extended YaleB and MITindoor with four levels of prior knowledge.

Dataset	Extended YaleB				MITIndoor			
Method	5%	10%	15%	20%	5%	10%	15%	20%
PMVSR	0.857	0.869	0.876	0.882	0.702	0.758	0.769	0.793
TMSRL	0.727	0.809	0.825	0.839	0.736	0.768	0.787	0.802
CTLR	0.755	0.823	0.847	0.865	0.723	0.764	0.775	0.789
EMVSC	0.877	0.905	0.913	0.928	0.788	0.802	0.829	0.838

TLSpNM-MSC uses a new surrogate of tensor rank, namely the tensor logarithmic Schatten-p norm (TLSpN), which considers the physical difference between singular values by the non-convex and non-linear penalty function.

From Table 7, we observe that our EMVSC obtains best performance in most cases. In complex data sets (e.g., the Scene-15 and the MITIndoor) with highly nonlinear structure, TLSpNM-MSC is superior to others. EMVSC uses the tSVD-TNN which is the best convex surrogate of tensor rank and it is efficient in optimization. However, tSVD-TNN is not the best surrogate in nonlinear relationship mining.

On one hand, our EMVSC is better than ''EMVSC w/o twist operation''. Without tensor twist operation, the self-representations will be destroyed in Fourier domain, resulting in loosing of complementary information among views. On the other hand, EMVSC is better than ''EMVSC w/o constraint propagation''. Obviously, the process of constraint propagation collects pairwise constraint of the whole data set. These constraints improve the discriminating power of the EMVSC.

G. DIFFERENT LEVEL OF PRIOR KNOWLEDGE

To reveal the influence of the level of prior knowledge on these semi-supervised clustering methods, we conduct these experiments with four levels on two data sets: a middle-scale one (e.g., Extended YalB) and a large-scale one (e.g., MITindoor). Results are recorded in Table 8.

From Table 8, we conclude that the clustering performance increases with increasing prior information. We also observe that our EMVSC performs best in all cases, especially when there are limited constraints, which verifies that EMVSC could fully exploit limited prior information through employing constraint propagation.

VI. CONCLUSION

We have presented an enhanced method for multi-view subspace clustering, called EMVSC. EMVSC is based on twist tensor nuclear norm and constraint propagation. EMVSC exploits high-order correlation across views by imposing twist tensor nuclear norm on tensor representation constructed by stacking view-specific self-representations. In addition, EMVSC effectively utilizes limited prior knowledge by performing constraint propagation. Note that EMVSC pulls samples belonging to the same class closer and samples from different class farther, increasing discriminating power. As a result, EMVSC can obtain satisfactory clustering performance when multi-view data are with limited prior knowledge of pairwise constraint. Experiments on seven real-world databases show its efficacy.

ACKNOWLEDGMENT

The authors would like to thank Yuan Xie for providing the codes of t-SVD, and thank to Di Wang for her constructive suggestions of implement of constraint propagation.

REFERENCES

[1] J. Guo, Y. Sun, J. Gao, Y. Hu, and B. Yin, ''Logarithmic Schatten-p norm minimization for tensorial multi-view subspace clustering,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3396–3410, Jun. 2022.
[2] Z. Yang, N. Liang, W. Yan, Z. Li, and S. Xie, ''Uniform distribution non-negative matrix factorization for multiview clustering,'' *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3249–3262, May 2021.
[3] E. M. Al-Sharoa and M. A. Al-Wardat, ''Multi-view robust tensor-based subspace clustering,'' *IEEE Access*, vol. 10, pp. 134292–134306, 2022.
[4] Z. Kang, Z. Lin, X. Zhu, and W. Xu, ''Structured graph learning for scalable subspace clustering: From single view to multiview,'' *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 8976–8986, Sep. 2022.
[5] R. Basri and D. W. Jacobs, ''Lambertian reflectance and linear subspaces,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.
[6] T. Hastie and P. Y. Simard, ''Metrics and models for handwritten character recognition,'' *Stat. Sci.*, vol. 13, no. 1, pp. 54–65, Feb. 1998.
[7] M. Brbić and I. Kopriva, ''Multi-view low-rank sparse subspace clustering,'' *Pattern Recognit.*, vol. 73, pp. 247–258, Jan. 2018.
[8] M. Abavisani and V. M. Patel, ''Multimodal sparse and low-rank subspace clustering,'' *Inf. Fusion*, vol. 39, pp. 168–177, Jan. 2018.
[9] W. Zhu and B. Peng, ''Sparse and low-rank regularized deep subspace clustering,'' *Knowl.-Based Syst.*, vol. 204, Sep. 2020, Art. no. 106199.
[10] S. Luo, C. Zhang, W. Zhang, and X. Cao, ''Consistent and specific multi-view subspace clustering,'' in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, pp. 1–8.
[11] M.-S. Chen, C.-D. Wang, and J.-H. Lai, ''Low-rank tensor based proximity learning for multi-view clustering,'' *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 5, pp. 5076–5090, May 2022.
[12] Y. Chen, X. Xiao, Z. Hua, and Y. Zhou, ''Adaptive transition probability matrix learning for multiview spectral clustering,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4712–4726, Sep. 2021.
[13] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, ''Low-rank tensor constrained multiview subspace clustering,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, May 2015, pp. 1582–1590.
[14] J. Guo, Y. Sun, J. Gao, Y. Hu, and B. Yin, ''Multi-attribute subspace clustering via auto-weighted tensor nuclear norm minimization,'' *IEEE Trans. Image Process.*, vol. 31, pp. 7191–7205, 2022.
[15] X. Xiao, Y. Chen, Y.-J. Gong, and Y. Zhou, ''Prior knowledge regularized multiview self-representation and its applications,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1325–1338, Mar. 2021.
[16] 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.

[17] Y. Tang, Y. Xie, C. Zhang, and W. Zhang, "Constrained tensor representation learning for multi-view semi-supervised subspace clustering," *IEEE Trans. Multimedia*, vol. 24, pp. 3920–3933, 2021.
[18] A. Bibi, A. Alqahtani, and B. Ghanem, "Constrained clustering: General pairwise and cardinality constraints," *IEEE Access*, vol. 11, pp. 5824–5836, 2023.
[19] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, "The twist tensor nuclear norm for video completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2961–2973, Dec. 2016.
[20] Z. Lu and Y. Peng, "Exhaustive and efficient constraint propagation: A graph-based learning approach and its applications," *Int. J. Comput. Vis.*, vol. 103, no. 3, pp. 306–325, Jul. 2013.
[21] Y. Xie, D. Tao, W. Zhang, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, 2018.
[22] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
[23] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 14, Dec. 2001, pp. 849–856.
[24] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 28, no. 1, 2014, pp. 1–11.
[25] N. Liang, Z. Yang, Z. Li, S. Xie, and C.-Y. Su, "Semi-supervised multi-view clustering with graph-regularized partially shared non-negative matrix factorization," *Knowl.-Based Syst.*, vol. 190, Feb. 2020, Art. no. 105185.
[26] N. Liang, Z. Yang, Z. Li, S. Xie, and W. Sun, "Semi-supervised multi-view learning by using label propagation based non-negative matrix factorization," *Knowl.-Based Syst.*, vol. 228, Sep. 2021, Art. no. 107244.
[27] X. Liu, S. Ji, W. Glänzel, and B. De Moor, "Multiview partitioning via tensor methods," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 1056–1069, May 2013.
[28] H. Wang, G. Han, B. Zhang, G. Tao, and H. Cai, "Multi-view learning a decomposable affinity matrix via tensor self-representation on Grassmann manifold," *IEEE Trans. Image Process.*, vol. 30, pp. 8396–8409, 2021.
[29] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, 2011.
[30] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.
[31] O. Semerci, N. Hao, M. E. Kilmer, and E. L. Miller, "Tensor-based formulation and nuclear norm regularization for multienergy computed tomography," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1678–1693, Apr. 2014.
[32] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3842–3849.
[33] Y. Qiu, G. Zhou, Y. Wang, Y. Zhang, and S. Xie, "A generalized graph regularized non-negative tucker decomposition framework for tensor data representation," *IEEE Trans. Cybern.*, vol. 52, no. 1, pp. 594–607, Jan. 2020.
[34] W. Yan, B. Zhang, S. Ma, and Z. Yang, "A novel regularized concept factorization for document clustering," *Knowl.-Based Syst.*, vol. 135, pp. 147–158, Nov. 2017.
[35] Z. Fu, Z. Lu, H. Ip, Y. Peng, and H. Lu, "Symmetric graph regularized constraint propagation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 25, no. 1, 2011, pp. 350–355.
[36] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. NIPS*, 2011, pp. 1–11.
[37] 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.*, vol. 2, Jun. 2005, pp. 524–531.
[38] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 413–420.
[39] 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.
[40] X. Qi, R. Xiao, C.-C. Li, Y. Qiao, J. Guo, and X. Tang, "Pairwise rotation invariant co-occurrence local binary pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2199–2213, Nov. 2014.
[41] 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.
[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[43] Q. Xiao, S. Du, J. Song, Y. Yu, and Y. Huang, "Hyper-Laplacian regularized multi-view subspace clustering with a new weighted tensor nuclear norm," *IEEE Access*, vol. 9, pp. 118851–118860, 2021.
[44] Y. Chen, X. Xiao, and Y. Zhou, "Jointly learning kernel representation tensor and affinity matrix for multi-view clustering," *IEEE Trans. Multimedia*, vol. 22, no. 8, pp. 1985–1997, Aug. 2020.
[45] Y. Chen, X. Xiao, and Y. Zhou, "Multi-view subspace clustering via simultaneously learning the representation tensor and affinity matrix," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107441.

WEI YAN received the M.S. degree from the School of Automation, Guangdong University of Technology (GDUT), Guangzhou, China, in 2016, and the Ph.D. degree from the Faculty of Science and Technology, University of Macau, Macau, China, in 2019.

He is currently a Postdoctoral Fellow with the Guangzhou Institute of International Finance, Guangzhou University, Guangzhou, China. He has published more than eight papers, including TOP journals, such as IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS (TSMC), KNOWLEDGE-BASED SYSTEMS (KBS), and IEEE TRANSACTIONS ON CYBERNETICS (TCYB). His research interests include machine learning, pattern recognition, and signal processing.

YU WANG received the master's degree from the City University of Macau, Macau, China, where he is currently pursuing the Ph.D. degree with the School of Finance. His research interests include pattern recognition, and machine learning techniques with applications on finance.

MENGXIN WANG received the Ph.D. degree from the Renmin University of China, in 2010. He is currently a Professor with the Guangzhou Institute of International Finance, Guangzhou University. His research interests include applied economics, econometric analysis, and machine learning.

JUNJIE YANG received the B.E. degree from the Southwest University of Science and Technology, Mianyang, China, in 2008, and the Ph.D. degree from the School of Automation, Guangdong University of Technology (GDUT), Guangzhou, China, in 2017.

In 2016 and 2019, he was a Research Fellow with the School of Computer, Data and Mathematical Sciences, Western Sydney University, Australia. From 2017 to 2020, he was a Postdoctoral Researcher with the School of Electromechanical Engineering, GDUT. He is currently a Lecturer with the School of Automation, GDUT, and the Key Laboratory of Intelligent Information Processing and System Integration of IoT, Ministry of Education, GDUT. His research interests include blind speech processing and machine learning.