# Spectral Embedding Fusion for Incomplete Multiview Clustering

Jie Chen, *Member, IEEE*, Yingke Chen, Zhu Wang, Haixian Zhang, *Member, IEEE*, and Xi Peng, *Senior Member, IEEE*

***Abstract*— Incomplete multiview clustering (IMVC) aims to reveal the underlying structure of incomplete multiview data by partitioning data samples into clusters. Several graph-based methods exhibit a strong ability to explore high-order information among multiple views using low-rank tensor learning. However, spectral embedding fusion of multiple views is ignored in low-rank tensor learning. In addition, addressing missing instances or features is still an intractable problem for most existing IMVC methods. In this paper, we present a unified spectral embedding tensor learning (USETL) framework that integrates the spectral embedding fusion of multiple similarity graphs and spectral embedding tensor learning for IMVC. To remove redundant information from the original incomplete multiview data, spectral embedding fusion is performed by introducing spectral rotations at two different data levels, i.e., the spectral embedding feature level and the clustering indicator level. The aim of introducing spectral embedding tensor learning is to capture consistent and complementary information by seeking high-order correlations among multiple views. The strategy of removing missing instances is adopted to construct multiple similarity graphs for incomplete multiple views. Consequently, this strategy provides an intuitive and feasible way to construct multiple similarity graphs. Extensive experimental results on multiview datasets demonstrate the effectiveness of the two spectral embedding fusion methods within the USETL framework.**

***Index Terms*— Incomplete multiview clustering, low-rank tensor learning, spectral embedding, spectral rotation.**

## I. Introduction

MULTIVIEW data are usually collected from different sources or described by different types of features. The analysis of multiview data has attracted attention in various applications, e.g., object segmentation [1], [2], information retrieval [3], [4], [5] and scene recognition [6]. Multiview clustering (MVC) aims to categorize data samples into clusters according to a specific similarity rule. It provides an effective way to explore the underlying structure of multiview data.

MVC has been heavily investigated in recent years [5], [7], [8], [9], [10], [11], [12]. Compared with traditional clustering methods for single-view data [13], [14], MVC optimally exploits complementary information among multiple views to improve clustering performance. The analysis of multiview data is essentially the task of fusing information from multiple views. In most MVC methods, the fusion of multiview data information can be performed at three different data levels, i.e., the original data level [5], [15], [16], the feature level [17], [18] and the clustering indicator level [4], [6]. For example, Chen et al. [5] presented a symmetric multiview low-rank representation method to discover the correlations of multiview data. Li et al. [17] presented a consensus graph learning (CGL) method that learns an affinity matrix from multiple spectral embedding graphs. Tang et al. [6] presented a unified one-step MVC framework that incorporates spectral embedding and $k$-means clustering [19] to obtain discrete clustering labels. These algorithms often achieve impressive clustering performance for multiview data at different data levels. Nevertheless, some instances may not be available in one or more views for various reasons; e.g., some of the data may be missing, or there may be corruption in the collecting devices. Such incompleteness often leads to heavy performance deterioration for existing MVC methods. As a result, dealing with incomplete multiview data poses a great challenge for researchers.

A variety of incomplete multiview clustering (IMVC) methods have been proposed for learning an affinity matrix via various similarity constraints [1], [4], [20], [21], [22], [23], [24], [25]. The affinity matrix can be employed to construct a graph of the data samples. From the perspective of spectral graph theory [26], the spectral embedding that is employed for spectral clustering can be calculated from the affinity matrix. For example, Chen et al. [1] presented an augmented sparse representation (ASR) method that learns sparse affinity matrices for individual views under the sparsity constraints. Unfortunately, however, the redundant information contained in the original incomplete multiview data inevitably causes deterioration in the performance of IMVC. Liu et al. [25] presented a regularized IMVC algorithm that imputes incomplete kernel matrices to learn a consensus clustering matrix. However, the results strongly depend on the choice of the

Manuscript received 24 October 2022; revised 14 January 2024 and 5 June 2024; accepted 25 June 2024. Date of publication 4 July 2024; date of current version 9 July 2024. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62176171 and Grant U21B2040, in part by the Fundamental Research Funds for the Central Universities under Grant CJ202303, and in part by Sichuan Science and Technology Planning Project under Grant 24NSFTD0130. The associate editor coordinating the review of this article and approving it for publication was Dr. Nelly Pustelnik. *(Corresponding authors: Xi Peng; Haixian Zhang.)*

Jie Chen, Haixian Zhang, and Xi Peng are with the College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: chenjie2010@scu.edu.cn; zhanghaixian@scu.edu.cn; pengx.gm@gmail.com).

Yingke Chen is with the Department of Computer and Information Sciences, Northumbria University, NE1 8ST Newcastle upon Tyne, U.K. (e-mail: yke.chen@gmail.com).

Zhu Wang is with the Law School, Sichuan University, Chengdu 610065, China (e-mail: wangzhu@scu.edu.cn).

Digital Object Identifier 10.1109/TIP.2024.3420796

base kernel matrices. In addition, a novel multiview matrix factorization model that incorporates sparse regularization and graph embedding has been proposed to investigate the local structures of all incomplete views [21], and a graph-regularized projective consensus representation learning (GPCRL) model has been designed to effectively process new samples and extract information from multiple views [20]. Nevertheless, these models neglect the examination of high-order information across multiple views.

With the goal of exploring the high-order correlations of incomplete multiple views, numerous tensor-based IMVC methods have been proposed [27], [28], [29], [30]. For example, Xie et al. [30] presented a tensor-singular value decomposition (t-SVD)-based multiview subspace clustering (MVSC) method that ensure a consensus among different views by using the tensor nuclear norm (TNN). The t-SVD-based MVSC method has proven effective in practice, even though it is not yet fully supported by convolution algebra theory. Li et al. [28] presented a graph structure refining (GSR) method in which the TNN is introduced to separate the refined graph structure from the bias error. Lv et al. [29] presented a view consistency learning (VCL) algorithm that employs the tensor Schatten $p$-norm to learn a low-dimensional representation for graph learning. These methods apply tensor rotation on the self-representation coefficient tensor to examine the correlations within and across multiple views. Specifically, tensor rotation provides two benefits for MVSC [30]. First, a self-representation coefficient tensor can be preserved in the Fourier domain. Second, each frontal slice of this self-representation coefficient tensor in the Fourier domain incorporates complementary information from multiple views when applying a truncated t-SVD operation. In contrast, the arrangement of the self-representation coefficient is disrupted in Fourier domain when tensor rotation is not considered. This leads to degraded clustering performance since complementary information cannot be effectively captured and propagated across multiple views [30]. These methods exhibit a strong ability to explore high-order information among multiple views with the help of tensors. However, they still suffer from certain limitations. For example, the ability of the GSR method to perform information fusion at the feature level is sensitive to the specific graph structure, and when the VCL method is used to perform information fusion at the original data level, the clustering performance is inevitably affected by redundant information.

In addition to utilizing complementary information among multiple views, another key to IMVC is to mitigate the detrimental effects on clustering performance caused by missing instances in some views. Several existing IMVC methods employ an intuitive strategy of filling in missing instances in incomplete multiview data with zeros or mean values [31]. However, in the case of a relatively large missing data ratio, such a filling strategy could affect the clustering performance since data that contain many such fake instances may not accurately reflect the intrinsic structures of the incomplete multiview data. An alternative strategy is to integrate an indictor matrix for sorting the available instances into graph learning models at the feature level [29], [32]. However, the fake features could still limit the final clustering performance of the models. In addition, some recent work has attempted to recover missing instances or features in the feature space [25], [33]. For example, Wen et al. [33] presented an incomplete multiview tensor spectral clustering method that incorporates missing view recovery and similarity graph learning into a unified framework. However, inferring the accurate features corresponding to the missing instances is an intractable problem if these instances are completely corrupted. Therefore, the development of reasonable strategies for dealing with missing instances in IMVC is worthy of further investigation.

TABLE I
DEFINITIONS OF SYMBOLS

| Symbol | Definition |
|---|---|
| $n$ | Number of samples or features |
| $d$ | Dimension of samples |
| $c$ | Number of clusters |
| $\mathbf{1}_n$ | Column vector filled with all ones |
| $\mathbf{I}_n$ | Identity matrix of size $n \times n$ |
| $\mathbf{X} \in \mathbb{R}^{d \times n}$ | Data matrix containing the $n$ samples |
| $diag(\mathbf{X})$ | Vector containing the $n$ diagonal elements of $\mathbf{X}$ |
| $tr(\mathbf{X})$ | Trace of $\mathbf{X}$ |
| $\|\mathbf{X}\|_1$ | $l_1$-norm of $\mathbf{X}$ |
| $\|\mathbf{X}\|_F$ | Frobenius norm of $\mathbf{X}$ |
| $\|\mathbf{X}\|_*$ | nuclear norm of $\mathbf{X}$ |
| $\|\mathbf{X}\|_{max}$ | maximum value of all the elements in $\mathbf{X}$ |
| $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ | Tensor $\mathcal{T}$ |
| $\|\mathcal{T}\|_*$ | Tensor nuclear norm of tensor $\mathcal{T}$ |
| $\mathbf{Y} \in \{0,1\}^{n \times c}$ | A binary clustering indicator matrix |
| $\mathbf{H} \in \mathbb{R}^{n \times c}$ | A spectral embedding matrix |
| $\mathbf{R} \in \mathbb{R}^{c \times c}$ | A rotation matrix |

To address the above limitations and drawbacks, we present a unified spectral embedding tensor learning (USETL) framework that comprises two spectral embedding fusion methods for IMVC. Unlike existing graph-based IMVC models, we integrate the spectral embedding fusion of multiple similarity graphs and spectral embedding tensor learning into the USETL framework. Specifically, the USETL framework combines two key components: stacking different spectral embedding matrices into a third-order spectral embedding tensor for graph learning and performing spectral embedding fusion of multiple similarity graphs by means of spectral rotation. Spectral embedding fusion is performed by introducing spectral rotation at two different data levels, i.e., the spectral embedding feature level and the clustering indicator level. This is beneficial for removing redundant information from the original incomplete multiview data. Moreover, the introduction of spectral embedding tensor learning enables consistent and complementary information to be captured by exploring the high-order correlations among multiple views. We present an alternating iterative optimization procedure to solve the two optimization problems corresponding to these two levels of spectral embedding fusion. Instead of filling in or recovering missing instances or features, we introduce a strategy of removing missing instances from incomplete multiple views for the construction of multiple similarity graphs. This strategy provides an intuitive and feasible way to construct multiple similarity graphs. In addition, we present the related theoretical analysis of the proposed USETL framework.

Our major contributions are summarized as follows.

- The USETL framework is presented to perform spectral embedding fusion of multiple similarity graphs via two

approaches: different spectral embedding matrices are stacked into a third-order low-rank tensor for graph learning, and spectral embedding fusion of multiple similarity graphs is performed by means of spectral rotation.
- Two spectral embedding fusion schemes are proposed to yield spectral embedding fusion results at two different data levels, i.e., the spectral embedding feature level and the clustering indicator level. An alternating iterative optimization procedure is presented to solve the two optimization problems corresponding to these two spectral embedding fusion schemes.
- Extensive experiments on six incomplete multiview datasets validate the effectiveness of the two spectral embedding fusion methods.

The remainder of this paper is organized as follows. In Section II, we give a brief review of related work. Section III presents the proposed USETL framework. Extensive experiments are conducted to validate the effectiveness of the proposed USETL framework in Section IV. Finally, we provide a conclusion in Section V.

## II. Related Work

In this section, we briefly introduce some related work on the proposed USETL framework. The notation used in this paper is summarized in Table I.

### A. Spectral Embedding Theory

Consider a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d\times n}$ with $n$ samples, where $\mathbf{x}_i \in \mathbb{R}^d$ $(1 \le i \le n)$. The weighted adjacency matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$ is used to represent the similarity among samples. According to spectral graph theory [26], a degree matrix $\mathbf{D} \in \mathbb{R}^{n\times n}$ is defined as $\mathbf{D} = diag[d_1, d_2, \ldots, d_n]$, whose diagonal elements are $d_i = \sum_{j=1}^{n} W_{ij}$. A normalized Laplacian matrix $\mathbf{L} \in \mathbb{R}^{n\times n}$ is also defined as follows:

$$\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}. \tag{1}$$

Suppose that the samples in $\mathbf{X}$ can be partitioned into $c$ clusters. The relaxed objective function for a normalized cut (NCut) [19] can be written as follows:

$$\min_{\mathbf{H}} tr\left(\mathbf{H}\mathbf{L}\mathbf{H}^T\right) \ s.t.\ \mathbf{H}^T\mathbf{H} = \mathbf{I}_c. \tag{2}$$

The optimal solution $\mathbf{H}^*$ consists of the eigenvectors corresponding to the $c$ smallest eigenvalues of $\mathbf{L}$.

The aim of spectral rotation is to find a discrete indicator matrix solution from the real-valued eigenvectors [34], [35]. The objective function of spectral rotation is formulated as follows:

$$\min_{\mathbf{Y},\mathbf{R}} \left\|\mathbf{H}^*\mathbf{R} - \mathbf{Y}\right\|_F^2 \ s.t.\ \mathbf{R}^T\mathbf{R} = \mathbf{I}_c, \mathbf{Y}\mathbf{1}_c = \mathbf{1}_n. \tag{3}$$

In addition, an improved spectral rotation (ISR) method [36] is presented to find a discrete solution of the cluster assignments. The objective function of the ISR method is

$$\min_{\mathbf{Y},\mathbf{R}} \left\|\mathbf{H}^*\mathbf{R} - \mathbf{D}^{1/2}\mathbf{Y}\left(\mathbf{Y}^T\mathbf{D}\mathbf{Y}\right)^{-1/2}\right\|_F^2$$

**Algorithm 1** The $t$-SVT Operator [38]

**Input:** $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and a parameter $\lambda > 0$.
/* $fft$ is a MATLAB command */
Compute $\bar{\mathcal{A}} = fft(\mathcal{A}, [], 3)$;
/* Perform SVT on each frontal slice of $\bar{\mathcal{A}}$ */
**for** $i = 1, \ldots, \frac{\lceil n_3+1\rceil}{2}$ **do**
  $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = svd\left(\bar{\mathbf{A}}^{(i)}\right)$;
  $\bar{\mathbf{W}}^{(i)} = \mathbf{U}\left(\mathbf{S} - \lambda\right)_+ \mathbf{V}^T$;
**end for**
**for** $i = \frac{\lceil n_3+1\rceil}{2} + 1, \ldots, n_3$ **do**
  $\bar{\mathbf{W}}^{(i)} = conj\left(\bar{\mathbf{W}}^{(n_3-i+2)}\right)$;
**end for**
**Output:** $\mathcal{D}_\lambda(\mathcal{A}) = ifft\left(\bar{\mathcal{W}}, [], 3\right)$.

$$s.t.\ \mathbf{R}^T\mathbf{R} = \mathbf{I}_c, \mathbf{Y}\mathbf{1}_c = \mathbf{1}_n. \tag{4}$$

Setting $\mathbf{Y}^* = \mathbf{D}^{1/2}\mathbf{Y}\left(\mathbf{Y}^T\mathbf{D}\mathbf{Y}\right)^{-1/2}$, we have

$$\left(\mathbf{Y}^*\right)^T\mathbf{Y}^* = \mathbf{I}_c. \tag{5}$$

According to Eq. (5), $\mathbf{Y}^*$ provides a better approximation of $\mathbf{H}^*$ than $\mathbf{Y}$ does. Hence, this approach is expected to yield better clustering results [37].

### B. Low-Rank Tensor Learning

For a tensor $\mathcal{Y} \in \mathbb{R}^{n_1\times n_2\times n_3}$, the frontal slice of $\mathcal{Y}$ is represented as $\mathcal{Y}(:,:,i)$ $(1 \le i \le n_3)$. The transpose of the tensor $\mathcal{Y}$, denoted as $\mathcal{Y}^T \in \mathbb{R}^{n_2\times n_1\times n_3}$, is obtained by transposing each frontal slice of $\mathcal{Y}$ and then reversing the order of the transposed frontal slices from the second slice to the $n_3$th slice [30], [39]. The $unfold$ operator [39] converts $\mathcal{Y}$ into a matrix with a size of $n_1n_3 \times n_2$ and $fold$ is its inverse operator, i.e.,

$$unfold(\mathcal{Y}) = \begin{bmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \\ \vdots \\ \mathbf{Y}^{(n_3)} \end{bmatrix}, \quad fold(unfold(\mathcal{Y})) = \mathcal{Y} \tag{6}$$

where $\mathbf{Y}^{(i)} \in \mathbb{R}^{n_1\times n_2}$ compactly denotes the $i$th frontal slice of $\mathcal{Y}$.

The definition of the block circulant matrix $bcirc(\mathcal{Y})$ is

$$bcirc(\mathcal{Y}) = \begin{bmatrix} \mathcal{Y}^{(1)} & \mathcal{Y}^{(n_3)} & \cdots & \mathcal{Y}^{(2)} \\ \mathcal{Y}^{(2)} & \mathcal{Y}^{(1)} & \cdots & \mathcal{Y}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{Y}^{(n_3)} & \mathcal{Y}^{(n_3-1)} & \cdots & \mathcal{Y}^{(1)} \end{bmatrix}. \tag{7}$$

Given two third-order tensors $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2\times n_l\times n_3}$, their $t$-product is defined as

$$\mathcal{E} = \mathcal{A} * \mathcal{B} = fold\left(bcirc(\mathcal{A}) \cdot unfold(\mathcal{B})\right) \tag{8}$$

where $\mathcal{E} \in \mathbb{R}^{n_1\times n_l\times n_3}$ [39]. A tensor is said to be $f$-diagonal if each frontal slice is a diagonal matrix [38], [40]. A third-order tensor $\mathcal{Q} \in \mathbb{R}^{n_1\times n_1\times n_3}$ is orthogonal if it satisfies $\mathcal{Q}^T\mathcal{Q} = \mathcal{Q}\mathcal{Q}^T = \mathcal{I} \in \mathbb{R}^{n_1\times n_1\times n_3}$, where $\mathcal{I}$ is an identity tensor [38].

The $t$-SVD operation imposed on $\mathcal{A}$ can be factorized as follows:

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V} \tag{9}$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an $f$-diagonal tensor [38]. The tube of $\mathcal{S}$ is denoted as $\mathcal{S}(i, j, :)$, where $1 \leq i \leq n_1$ and $1 \leq i \leq n_2$. The tensor tubal rank of $\mathcal{A}$ is defined as the number of nonzero singular tubes of $\mathcal{S}$, i.e.,

$$rank_t(\mathcal{A}) = \#\{i, \mathcal{S}(i, i, 1) \neq 0\} \tag{10}$$

where $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}$ [38], [39]. The TNN of $\mathcal{A}$ is defined as

$$\|\mathcal{A}\|_* = \sum_{i=1}^{r} \mathcal{S}(i, i, 1) \tag{11}$$

where $r = rank_t(\mathcal{A})$ [38], [39].

The discrete Fourier transform of a finite sequence of $n$ complex numbers $z_0, z_1, \ldots, z_{n-1}$ is another sequence of $n$ complex numbers $x_0, x_1, \ldots, x_{n-1}$ [41], which is formulated as follows:

$$x_j = \sum_{k=0}^{n-1} z_k e^{-2\pi i j k/n}, \ (0 \leq j \leq n-1). \tag{12}$$

The inverse discrete Fourier transform [41] is given by the formula shown below:

$$z_j = \frac{1}{n} \sum_{k=0}^{n-1} x_k e^{2\pi i j k/n}. \tag{13}$$

The functions of the discrete Fourier transform and the inverse discrete Fourier transform can be called using the MATLAB[1] commands $fft$ and $ifft$, respectively.

To find a low-tubal-rank approximation of $\mathcal{A}$, the following objective function can be formulated:

$$\min_{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \lambda \|\mathcal{X}\|_* + \frac{1}{2} \|\mathcal{X} - \mathcal{A}\|_F^2. \tag{14}$$

The solution to Eq. (14) can be calculated by means of a proximal operator of the matrix nuclear norm. The tensor-singular-value thresholding ($t$-SVT) operator is such a proximal operator that is given as follows:

$$\mathcal{D}_\lambda(\mathcal{A}) = \mathcal{U} * \mathcal{S}_\lambda * \mathcal{V} \tag{15}$$

where $\mathcal{S}_\lambda = ifft\left(\left(\bar{\mathcal{S}} - \lambda\right)_+, [], 3\right)$, $\bar{\mathcal{S}}$ is a real tensor, $ifft$ is a MATLAB command and $(t)_+ = max(t, 0)$ [38]. The details of the $t$-SVT operator are given in Algorithm 1 [38].

### C. IMVC Techniques

Consider a set of incomplete multiview data $\left\{\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times n}, v = (1, 2, \ldots n_v)\right\}$ with $n$ samples and $c$ clusters, where $\mathbf{X}^{(v)}$ represents the $v$th view of the incomplete multiview data and $d_v$ is the dimensionality of the instances in the $v$th view. Each view $\mathbf{X}^{(v)}$ contains $n$ instances, i.e., $\mathbf{X}^{(v)} = \left[\mathbf{x}_1^{(v)}, \mathbf{x}_2^{(v)}, \ldots, \mathbf{x}_n^{(v)}\right]$. To represent the missing instances in the $v$th view, a diagonal indicator matrix $\mathbf{M}^{(v)} \in \mathbb{R}^{n \times n}$ is defined as follows:

$$M_{ii}^{(v)} = \begin{cases} 1, & \text{instance } \mathbf{x}_i^{(v)} \text{ is in the } v\text{-th view} \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

The existing instances of the incomplete multiview data are represented by $\left\{\mathbf{X}_r^{(v)} \in \mathbb{R}^{d_v \times N_v}\right\}_{v=1}^{n_v}$, where $N_v$ represents the number of existing instances available in the $v$th view. The index matrix $\mathbf{M}_r^{(v)} \in \mathbb{R}^{N_v \times n}$ consists of the rows of $\mathbf{M}^{(v)}$ corresponding to the existing instances.

Recently, numerous tensor-based IMVC methods have been proposed for identifying high-order correlations of incomplete multiple views [22], [28], [29]. For example, one representative tensor-based IMVC model uses a particular tensor norm to characterize the intrinsic structure of the incomplete multiview data: the VCL method employs the tensor Schatten $p$-norm to characterize low-rank structure embeddings in similarity graphs [29]. The self-expressiveness property of the data instances is used to learn the affinity matrix in each view. Specifically, each existing data instance can be represented by a small number of other existing data instances. The objective function of the VCL method is formulated as follows:

$$\begin{aligned} \min_{\mathbf{Z}^{(v)}, \mathbf{E}^{(v)}, \mathbf{H}^{(v)}, \mathbf{F}} & \sum_{v=1}^{n_v} \|\mathcal{Z}\|_*^p + \frac{\lambda_1}{2} \Gamma\left(\mathbf{H}^{(v)}, \mathbf{F}\right) + \sum_{v=1}^{n_v} \lambda_2 \left\|\mathbf{E}^{(v)}\right\|_1 \\ & + \sum_{v=1}^{n_v} \lambda_3 tr\left(\left(\mathbf{H}^{(v)}\right)^T \mathbf{L}^{(v)} \mathbf{H}^{(v)}\right) \\ s.t. \ & \mathbf{X}_r^{(v)} = \mathbf{X}_r^{(v)} \mathbf{Z}^{(v)} + \mathbf{E}^{(v)}, \ \mathbf{Z}^{(v)} \mathbf{1}_n = \mathbf{1}_n, \ 0 \leq \mathbf{Z}^{(v)} \leq 1, \\ & Z_{i,i}^{(v)} = 0, \left(\mathbf{H}^{(v)}\right)^T \mathbf{H}^{(v)} = \mathbf{I}_c, \mathbf{F}^T \mathbf{F} = \mathbf{I}_c \end{aligned} \tag{17}$$

where $\mathbf{E}^{(v)}$ is the noise term, $\|\mathcal{Z}\|_*^p$ is the tensor Schatten $p$-norm of the tensor $\mathcal{Z}$, $\mathbf{Z}^{(v)} \in \mathbb{R}^{N_v \times N_v}$ denotes the self-expressiveness coefficient matrix, the function $\Gamma(\cdot)$ measures the disagreement between a clustering indicator matrix $\mathbf{F} \in \mathbb{R}^{n \times c}$ and a spectral embedding matrix $\mathbf{H}^{(v)} \in \mathbb{R}^{n \times c}$, $\overline{\mathbf{L}}^{(v)} \in \mathbb{R}^{N_v \times N_v}$ is the Laplacian matrix of $\mathbf{Z}^{(v)}$, and $\mathbf{L}^{(v)} = \left(\mathbf{M}_r^{(v)}\right)^T \overline{\mathbf{L}}^{(v)} \mathbf{M}_r^{(v)}$. In Eq. (17), the strategy of filling with zeros is transferred from the original incomplete multiview data to Laplacian graph regularization. Hence, IMVC still faces two major challenges: fusing the information of the incomplete multiple views and dealing with missing instances in the incomplete multiview data.

## III. Spectral Embedding Fusion

In this section, we present a USETL framework that performs spectral embedding fusion for IMVC. The spectral embedding fusion of multiple similarity graphs and spectral embedding tensor learning are integrated into the proposed USETL framework. The spectral embedding fusion of multiple similarity graphs can be performed at two different data levels.

### A. The USETL Framework

The challenge of IMVC is to fuse the information of incomplete multiple views to explore consistent and complementary

[1] https://www.mathworks.com/

information among those incomplete multiple views. In the proposed USETL framework, information fusion of incomplete multiple views can be implemented by capturing the intrinsic structure of the incomplete multiview data. Suppose that a given set of multiple similarity graphs $\left\{\mathbf{W}^{(v)}\right\}_{v=1}^{n_v} \in \mathbb{R}^{n\times n}$ is constructed from incomplete multiple views, where $\mathbf{H}^{(v)}$ and $\mathbf{L}^{(v)}$ are a spectral embedding matrix and a normalized Laplacian matrix, respectively, corresponding to the $v$th view.

To mitigate the detrimental impacts on clustering performance caused by noise, corruption and redundant information in the incomplete multiview data, information fusion is performed in the spectral embedding space. Ideally, the inner product of the spectral embedding matrix $\mathbf{H}^{(v)}$, $\mathbf{H}^{(v)}\left(\mathbf{H}^{(v)}\right)^T$, is a low-rank matrix consisting of $c$ diagonal block submatrices, according to spectral embedding theory [19]. To explore the high-level correlations among the different views, we integrate the spectral embedding matrices into a third-order tensor. Specifically, we stack $\left\{\mathbf{H}^{(v)}\left(\mathbf{H}^{(v)}\right)^T\right\}_{v=1}^{n_v}$ into the third-order tensor $\mathcal{H} \in \mathbb{R}^{n\times n\times n_v}$. The first frontal slices of $\mathcal{H}$ are $n\times n$ low-rank matrices, where each individual frontal slice corresponds to a specific view. To investigate cross-view correlations, we take $\mathcal{T} = \left\{\mathbf{H}^{(v)}\left(\mathbf{H}^{(v)}\right)^T\right\}_{v=1}^{n_v} \in \mathbb{R}^{n\times n_v\times n}$ as a surrogate for $\mathcal{H}$, where first frontal slices of $\mathcal{T}$ are $n \times n_v$ matrices. Thus, the third-order low-rank tensor $\mathcal{T}$ can be considered as a reasonable constraint for spectral embedding tensor learning.

As stated previously, our USETL framework incorporates both spectral embedding fusion of multiple similarity graphs and spectral embedding tensor learning. The general objective function of the USETL framework is formulated as follows:

$$\begin{aligned}\min_{\mathbf{F},\mathbf{R}^{(v)},\mathbf{H}^{(v)}} &\|\mathcal{T}\|_* + \alpha\sum_{v=1}^{n_v} tr\left(\left(\mathbf{H}^{(v)}\right)^T\mathbf{L}^{(v)}\mathbf{H}^{(v)}\right)\\ &+ \beta f\left(\mathbf{F},\mathbf{H}^{(v)},\mathbf{R}^{(v)}\right)\\ s.t.\ &\left(\mathbf{H}^{(v)}\right)^T\mathbf{H}^{(v)} = \mathbf{I}_c,\ \left(\mathbf{R}^{(v)}\right)^T\mathbf{R}^{(v)} = \mathbf{I}_c\end{aligned} \tag{18}$$

where $\alpha$ and $\beta$ are two trade-off parameters, $\mathbf{F}$ is the result of information fusion, $\mathbf{R}^{(v)} \in \mathbb{R}^{c\times c}$ denotes a spectral rotation matrix for the $v$th view, and $f(\cdot)$ represents the information fusion operation.

Within the USETL framework, we propose two different schemes for the spectral embedding fusion of multiple similarity graphs, which are performed at two different data levels, i.e., the spectral embedding feature level and the clustering indicator level. The details of these two schemes are given as follows.

*1) The Spectral Embedding Feature Level:* Inspired by learning the binary clustering indicator matrix in Eq. (3), we exploit spectral rotation to search for a fused clustering indicator matrix at the spectral embedding feature level. Given the individual spectral embedding matrices of multiple views $\left\{\mathbf{H}^{(v)}\right\}_{v=1}^{n_v}$, we perform spectral embedding fusion of multiple similarity graphs by introducing spectral rotation. Mathematically, the operation of spectral embedding fusion at the spectral embedding feature level can be formulated as follows:

$$\begin{aligned}\min_{\mathbf{F},\mathbf{R}^{(v)}} &\sum_{v=1}^{n_v}\left\|\mathbf{F} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2\\ s.t.\ &\left(\mathbf{R}^{(v)}\right)^T\mathbf{R}^{(v)} = \mathbf{I}_c,\ \mathbf{F}^T\mathbf{F} = \mathbf{I}_c\end{aligned} \tag{19}$$

where $\mathbf{F} \in \mathbb{R}^{n\times c}$ represents a fused clustering indicator matrix. The matrix $\mathbf{F}$ is an approximate result obtained by minimizing the difference between itself and each rotated spectral embedding matrix $\mathbf{H}^{(v)}$. The matrix $\mathbf{R}^{(v)}$ establishes rational interactions between $\mathbf{F}$ and $\mathbf{H}^{(v)}$, thereby guaranteeing a reasonable approximation relationship between $\mathbf{F}$ and $\mathbf{H}^{(v)}$. Accordingly, the corresponding expression for spectral embedding fusion can be integrated into Eq. (18) as follows:

$$\begin{aligned}\min_{\mathbf{F},\mathbf{R}^{(v)},\mathbf{H}^{(v)}} &\|\mathcal{T}\|_* + \alpha\sum_{v=1}^{n_v} tr\left(\left(\mathbf{H}^{(v)}\right)^T\mathbf{L}^{(v)}\mathbf{H}^{(v)}\right)\\ &+ \frac{\beta}{2}\sum_{v=1}^{n_v}\left\|\mathbf{F} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2\\ s.t.\ &\left(\mathbf{H}^{(v)}\right)^T\mathbf{H}^{(v)} = \mathbf{I}_c,\ \left(\mathbf{R}^{(v)}\right)^T\mathbf{R}^{(v)} = \mathbf{I}_c,\ \mathbf{F}^T\mathbf{F} = \mathbf{I}_c.\end{aligned} \tag{20}$$

The spectral embedding fusion scheme expressed in Eq. (20) yields the fused clustering indicator $\mathbf{F}$, which captures complementary information from multiple views.

*2) The Clustering Indicator Level:* Instead of learning a fused clustering indicator matrix as in Eq. (19), we can employ spectral rotation to yield a binary clustering indicator matrix for the spectral embedding fusion of multiple similarity graphs at the clustering indicator level. This strategy combines spectral embedding fusion and the subsequent application of the $k$-means algorithm into one step. Combined with Eq. (3), the operation of spectral embedding fusion at the clustering indicator level can be expressed as

$$\begin{aligned}\min_{\mathbf{Y},\mathbf{R}^{(v)}} &\sum_{v=1}^{n_v}\left\|\mathbf{Y}\left(\mathbf{Y}^T\mathbf{Y}\right)^{-1/2} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2\\ s.t.\ &Y_{ij} \in \{0,1\},\ \mathbf{Y}_{i,:}\mathbf{1}_c = 1,\ \left(\mathbf{R}^{(v)}\right)^T\mathbf{R}^{(v)} = \mathbf{I}_c\end{aligned} \tag{21}$$

where $\mathbf{Y} \in \mathbb{R}^{n\times c}$ is an approximate discrete solution representing the clustering indicator matrix, with each individual element having a value of either 0 or 1. In particular, $\mathbf{Y}\left(\mathbf{Y}^T\mathbf{Y}\right)^{-1/2}$ is regarded as a scaled clustering indicator matrix. Accordingly, an alternative objective function can be formulated as follows:

$$\begin{aligned}\min_{\mathbf{Y},\mathbf{R}^{(v)},\mathbf{H}^{(v)}} &\|\mathcal{T}\|_* + \alpha\sum_{v=1}^{n_v} tr\left(\left(\mathbf{H}^{(v)}\right)^T\mathbf{L}^{(v)}\mathbf{H}^{(v)}\right)\\ &+ \frac{\beta}{2}\sum_{v=1}^{n_v}\left\|\mathbf{Y}\left(\mathbf{Y}^T\mathbf{Y}\right)^{-1/2} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2\\ s.t.\ &\left(\mathbf{H}^{(v)}\right)^T\mathbf{H}^{(v)} = \mathbf{I}_c,\ \left(\mathbf{R}^{(v)}\right)^T\mathbf{R}^{(v)} = \mathbf{I}_c,\\ &Y_{ij} \in \{0,1\},\ \mathbf{Y}_{i,:}\mathbf{1}_c = 1.\end{aligned} \tag{22}$$

As observed in Eq. (18), the USETL framework unifies these two schemes to spectral embedding fusion and the search for high-level correlations among different views, which are based on spectral rotation and spectral embedding tensor learning. Spectral embedding fusion plays an essential role in the fusion of information from incomplete multiple views. For incomplete multiview data, the results of spectral embedding fusion among multiple views can be obtained at two different data levels, i.e., the spectral embedding feature level (SEFL) and the clustering indicator level (CIL). Hence, the USETL framework contains two corresponding spectral embedding fusion methods, namely, USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$.

### B. Initialization of $\{\mathbf{L}^{(v)}\}_{v=1}^{n_v}$, $\{\mathbf{H}^{(v)}\}_{v=1}^{n_v}$ and $\mathbf{Y}$

Given a set of incomplete multiview data $\{\mathbf{X}^{(v)}\}_{v=1}^{n_v}$, we need to initialize $\{\mathbf{L}^{(v)}\}_{v=1}^{n_v}$, $\{\mathbf{H}^{(v)}\}_{v=1}^{n_v}$ and $\mathbf{Y}$ for Eqs. (20) and (22). It is crucial to choose a proper strategy for dealing with missing instances. Numerous existing IMVC algorithms perform imputation on missing instances or fill in missing instances with zeros or mean values. In contrast to these strategies for missing instances, the missing instances are directly removed from the incomplete views in our proposed USETL framework.

For each view $\mathbf{X}_r^{(v)}$, we first construct a corresponding similarity matrix $\mathbf{Z}_r^{(v)} \in \mathbb{R}^{N_v \times N_v}$ using the adaptive neighbor graph learning (ANGL) method [42]. The ANGL method exploits the local connectivity among high-dimensional data instances to construct the similarity matrix. Considering two instances $\mathbf{x}_i^{(v)}$ and $\mathbf{x}_j^{(v)}$ in $\mathbf{X}_r^{(v)}$, the objective function of the ANGL method is formulated as follows:

$$\begin{aligned} &\min_{\mathbf{Z}_r^{(v)}} \sum_{i=1}^{N_v}\sum_{j=1}^{N_v} \left\|\mathbf{x}_i^{(v)} - \mathbf{x}_j^{(v)}\right\|_2^2 Z_{ij}^{(v)} + \gamma \left\|\mathbf{Z}_r^{(v)}\right\|_F^2 \\ &s.t.\ Z_{ij}^{(v)} \geq 0,\ (\mathbf{Z}_i)^T \mathbf{1}_n = 1 \end{aligned} \tag{23}$$

where $\gamma$ is a balance parameter and $\mathbf{Z}_i$ is the $i$th column of $\mathbf{Z}_r^{(v)}$. Each element $Z_{ij}^{(v)}$ of $\mathbf{Z}_r^{(v)}$ represents the similarity between the corresponding instances $\mathbf{x}_i^{(v)}$ and $\mathbf{x}_j^{(v)}$. Because $\mathbf{Z}_r^{(v)}$ may not be symmetric, we perform a symmetrization step on it, i.e.,

$$\mathbf{Z}_r^{(v)} \leftarrow \left(\mathbf{Z}_r^{(v)} + \left(\mathbf{Z}_r^{(v)}\right)^T\right)\Big/ 2. \tag{24}$$

Then, to obtain a full similarity matrix $\mathbf{Z}^{(v)} \in \mathbb{R}^{n \times n}$, each element in $\mathbf{Z}_r^{(v)}$ is assigned in accordance with the positions of the existing instances in the corresponding view. The remaining elements in $\mathbf{Z}^{(v)}$ are filled in with zeros. Next, $\mathbf{H}^{(v)}$ and $\mathbf{L}^{(v)}$ are initialized as the spectral embedding and the normalized Laplacian matrix, respectively, of $\mathbf{Z}^{(v)}$ in Eq. (2). Finally, we apply the $k$-means algorithm to the aggregated matrix $\mathbf{H}_{sum}$, which is the sum of $\{\mathbf{H}^{(v)}\}_{v=1}^{n_v}$, i.e., $\mathbf{H}_{sum} = \sum_{v=1}^{n_v} \mathbf{H}^{(v)}$, to derive discrete clustering partitions. These clustering partitions are utilized to initialize the clustering indicator matrix $\mathbf{Y}$.

### C. Optimization

We present an alternating iterative optimization procedure based on the alternating direction method of multipliers (ADMM) framework [43] to solve the optimization problems in Eqs. (20) and (22). For the first objective function, we introduce an auxiliary tensor variable $\mathcal{G} \in \mathbb{R}^{n \times n_v \times n}$ into Eq. (20). Thus, the optimization problem in Eq. (20) can be converted into the following equivalent problem:

$$\begin{aligned} \min_{\mathbf{F},\mathbf{R}^{(v)},\mathbf{H}^{(v)},\mathcal{G}} &\|\mathcal{G}\|_* + \alpha \sum_{v=1}^{n_v} tr\left(\left(\mathbf{H}^{(v)}\right)^T \mathbf{L}^{(v)} \mathbf{H}^{(v)}\right) \\ &+ \frac{\beta}{2} \sum_{v=1}^{n_v} \left\|\mathbf{F} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2 \\ s.t.\ &\left(\mathbf{H}^{(v)}\right)^T \mathbf{H}^{(v)} = \mathbf{I}_c,\ \left(\mathbf{R}^{(v)}\right)^T \mathbf{R}^{(v)} = \mathbf{I}_c, \\ &\mathbf{F}^T\mathbf{F} = \mathbf{I}_c,\ \mathcal{G} = \mathcal{T}. \end{aligned} \tag{25}$$

The augmented Lagrangian function in Eq. (25) is

$$\begin{aligned} \mathcal{L}\left(\mathbf{F}, \mathbf{R}^{(v)}, \mathbf{H}^{(v)}, \mathcal{G}\right) =& \|\mathcal{G}\|_* + \alpha \sum_{v=1}^{n_v} tr\left(\left(\mathbf{H}^{(v)}\right)^T \mathbf{L}^{(v)} \mathbf{H}^{(v)}\right) \\ &+ \frac{\beta}{2} \sum_{v=1}^{n_v} \left\|\mathbf{F} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2 \\ &+ \langle \mathcal{R}, \mathcal{T} - \mathcal{G}\rangle + \frac{\mu}{2}\|\mathcal{T} - \mathcal{G}\|_F^2 \end{aligned} \tag{26}$$

where $\mathcal{R} \in \mathbb{R}^{n \times n_v \times n}$ is a Lagrange multiplier, and $\mu > 0$ is an adaptive penalty parameter. Furthermore, the augmented Lagrangian function of Eq. (26) can be transformed into the following equivalent function:

$$\begin{aligned} \mathcal{L}\left(\mathbf{F}, \mathbf{R}^{(v)}, \mathbf{H}^{(v)}, \mathcal{G}\right) =& \|\mathcal{G}\|_* + \alpha \sum_{v=1}^{n_v} tr\left(\left(\mathbf{H}^{(v)}\right)^T \mathbf{L}^{(v)} \mathbf{H}^{(v)}\right) \\ &+ \frac{\beta}{2} \sum_{v=1}^{n_v} \left\|\mathbf{F} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2 \\ &+ \frac{\mu}{2}\left\|\mathcal{G} - \left(\mathcal{T} + \frac{\mathcal{R}}{\mu}\right)\right\|_F^2. \end{aligned} \tag{27}$$

The optimization problem in Eq. (27) can be partitioned into four subproblems, which correspond to solving for four different variables: $\mathbf{F}$, $\mathbf{R}^{(v)}$, $\{\mathbf{H}^{(v)}\}_{v=1}^{n_v}$ and $\mathcal{G}$. Accordingly, each of these four variables is updated alternately while the other variables are kept fixed, until convergence is reached.

*1) Subproblem of* $\mathbf{F}$*:* When $\mathbf{R}^{(v)}$, $\{\mathbf{H}^{(v)}\}_{v=1}^{n_v}$ and $\mathcal{G}$ are fixed, the problem in Eq. (27) can be rewritten as

$$\min_{\mathbf{F}} \sum_{v=1}^{n_v} \left\|\mathbf{F} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2 \quad s.t. \quad \mathbf{F}^T\mathbf{F} = \mathbf{I}_c. \tag{28}$$

Eq. (28) can be transformed into the following optimization problem:

$$\max_{\mathbf{F}} \mathbf{F}^T \sum_{v=1}^{n_v} \mathbf{H}^{(v)}\mathbf{R}^{(v)} \quad s.t. \quad \mathbf{F}^T\mathbf{F} = \mathbf{I}_c. \tag{29}$$

**Algorithm 2** Updating $\mathbf{H}^{(v)}$ via the GPI Algorithm [44]

**Input:** Data matrices $\mathbf{H}^{(v)}$, $\mathbf{B}^{(v)}$, and $\mathbf{C}^{(v)}$ and a parameter $\lambda_{\max}^{(v)}$.

**initialize:** $t = 1$, $f_0 = 0$, $\varepsilon = 0.1$ and $maxIters = 3$;
**while** not converged **do**
  $\mathbf{P}^{(v)} = \left(\lambda_{\max}^{(v)}\mathbf{I}_n - \mathbf{B}^{(v)}\right)\mathbf{H}^{(v)} + \mathbf{C}^{(v)}$;
  Perform economy SVD on $\mathbf{P}^{(v)}$, i.e., $\mathbf{P}^{(v)} = \mathbf{U}_p\mathbf{\Sigma}_p\mathbf{V}_p^T$;
  $\mathbf{H}^{(v)} = \mathbf{U}_p\mathbf{V}_p^T$;
  $f_t = tr\left(\left(\mathbf{H}^{(v)}\right)^T\mathbf{P}^{(v)}\right)$;
  Check the convergence condition:
  $(f_t - f_{t-1})/f_t \le \varepsilon$;
  **if** $t > maxIters$ and $converged$ **then**
    break;
  **end if**
  $t \leftarrow t + 1$;
**end while**

**Output:** $\mathbf{H}^{(v)}$.

The problem in Eq. (29) has a closed-form solution, namely,

$$\mathbf{F} = \mathbf{U}_m\mathbf{V}_m^T \tag{30}$$

where $\mathbf{M} = \sum_{v=1}^{n_v}\mathbf{H}^{(v)}\mathbf{R}^{(v)}$ and $\mathbf{U}_m$ and $\mathbf{V}_m$ are the left and right parts, respectively, of the economy SVD of the matrix $\mathbf{M}$, i.e., $\mathbf{M} = \mathbf{U}_m\Sigma_m\mathbf{V}_m^T$ [35].

*2) Subproblem of* $\left\{\mathbf{R}^{(v)}\right\}_{v=1}^{n_v}$*:* When $\mathbf{H}^{(v)}$, $\mathbf{F}$ and $\mathcal{G}$ are fixed, the problem in Eq. (27) can be rewritten as

$$\min_{\mathbf{R}^{(v)}}\sum_{v=1}^{n_v}\left\|\mathbf{F} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2 \ s.t.\ \left(\mathbf{R}^{(v)}\right)^T\mathbf{R}^{(v)} = \mathbf{I}_c. \tag{31}$$

The optimization problem in (31) is equivalent to

$$\max_{\mathbf{R}^{(v)}}\left(\mathbf{R}^{(v)}\right)^T\left(\mathbf{H}^{(v)}\right)^T\mathbf{F}\ s.t.\ \left(\mathbf{R}^{(v)}\right)^T\mathbf{R}^{(v)} = \mathbf{I}_c. \tag{32}$$

Similar to solving $\mathbf{F}$ in Eq. (29), the optimal solution of $\mathbf{R}^{(v)}$ in Eq. (32) can be obtained as

$$\mathbf{R}^{(v)} = \mathbf{U}_n\mathbf{V}_n^T \tag{33}$$

where $\mathbf{N} = \left(\mathbf{H}^{(v)}\right)^T\mathbf{F}$ and the economy SVD of the matrix $\mathbf{N}$ is $\mathbf{N} = \mathbf{U}_n\Sigma_n\mathbf{V}_n^T$.

*3) Subproblem of* $\left\{\mathbf{H}^{(v)}\right\}_{v=1}^{n_v}$*:* When $\mathbf{F}$, $\mathbf{R}^{(v)}$ and $\mathcal{G}$ are fixed, the problem in Eq. (27) can be rewritten as

$$\begin{aligned}\min_{\mathbf{H}^{(v)}}\alpha\sum_{v=1}^{n_v}tr\left(\left(\mathbf{H}^{(v)}\right)^T\mathbf{L}^{(v)}\mathbf{H}^{(v)}\right)\\ + \frac{\beta}{2}\sum_{v=1}^{n_v}\left\|\mathbf{F} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2 + \frac{\mu}{2}\left\|\mathcal{G} - \left(\mathcal{T} + \frac{\mathcal{R}}{\mu}\right)\right\|_F^2\\ s.t.\ \left(\mathbf{H}^{(v)}\right)^T\mathbf{H}^{(v)} = \mathbf{I}_c\end{aligned} \tag{34}$$

Setting $\mathcal{A} = \mathcal{G} - \frac{\mathcal{R}}{\mu}$, Eq. (34) becomes

$$\min_{\mathbf{H}^{(v)}}tr\left(\left(\mathbf{H}^{(v)}\right)^T\mathbf{B}^{(v)}\mathbf{H}^{(v)}\right) - tr\left(\left(\mathbf{H}^{(v)}\right)^T\mathbf{C}^{(v)}\right) \tag{35}$$

where $\mathbf{B}^{(v)} = \alpha\mathbf{L}^{(v)} - \frac{\mu}{2}\left(\mathbf{A}^{(v)} + \left(\mathbf{A}^{(v)}\right)^T\right)$ and $\mathbf{C}^{(v)} = \beta\mathbf{F}\left(\mathbf{R}^{(v)}\right)^T$. Furthermore, we have

$$\max_{\mathbf{H}^{(v)}} -tr\left(\left(\mathbf{H}^{(v)}\right)^T\mathbf{B}^{(v)}\mathbf{H}^{(v)}\right) + tr\left(\left(\mathbf{H}^{(v)}\right)^T\mathbf{C}^{(v)}\right) \tag{36}$$

Setting $\mathbf{P}^{(v)} = \left(\lambda_{\max}^{(v)}\mathbf{I}_n - \mathbf{B}^{(v)}\right)\mathbf{H}^{(v)} + \mathbf{C}^{(v)}$, the problem in Eq. (36) can be relaxed to

$$\max_{\mathbf{H}^{(v)}} tr\left(\left(\mathbf{H}^{(v)}\right)^T\mathbf{P}^{(v)}\right)\ s.t.\ \left(\mathbf{H}^{(v)}\right)^T\mathbf{H}^{(v)} = \mathbf{I}_c \tag{37}$$

where $\lambda_{max}^{(v)}$ is the largest eigenvalue of $\mathbf{B}^{(v)}$. The optimal solution to the problem in Eq. (37) can be obtained via the generalized power iteration (GPI) algorithm [44]. Specifically, it can be obtained by iteratively updating $\mathbf{H}^{(v)}$ after an initialization of $\mathbf{H}^{(v)}$ is given. The optimization procedure is outlined in Algorithm 2.

*4) Subproblem of* $\mathcal{G}$*:* When $\mathbf{R}^{(v)}$, $\mathbf{F}$ and $\mathbf{H}^{(v)}$ are fixed, the problem in Eq. (27) can be rewritten as

$$\min_{\mathcal{G}}\|\mathcal{G}\|_* + \frac{\mu}{2}\left\|\mathcal{G} - \left(\mathcal{T} + \frac{\mathcal{R}}{\mu}\right)\right\|_F^2. \tag{38}$$

This problem can be solved by Algorithm 1 [38].

In addition, the updating schemes for the Lagrange multiplier $\mathcal{R}$ and the penalty parameter $\mu$ are formulated as follows:

$$\mathcal{R}^{(v)} \leftarrow \mathcal{R}^{(v)} + \mu\left(\mathcal{G} - \mathcal{T}\right) \tag{39}$$

$$\mu \leftarrow \min(\rho\mu, \mu_{\max}) \tag{40}$$

where $\rho$ and $\mu_{\max}$ are constants.

For the other objective function in Eq. (22), we adopt the same alternating iterative optimization procedure as for the first objective function. The main difference between the two objective functions lies in whether $\mathbf{F}$ or $\mathbf{Y}$ is optimized.

*5) Subproblem of* $\mathbf{Y}$*:* When $\mathbf{R}^{(v)}$ and $\mathbf{H}^{(v)}$ are fixed, the problem in Eq. (22) can be rewritten as

$$\begin{aligned}\min_{\mathbf{Y}}\sum_{v=1}^{n_v}\left\|\mathbf{Y}\left(\mathbf{Y}^T\mathbf{Y}\right)^{-1/2} - \mathbf{H}^{(v)}\mathbf{R}^{(v)}\right\|_F^2\\ s.t.\ Y_{ij} \in \{0,1\},\ \mathbf{Y}_{i,:}\mathbf{1}_c = 1.\end{aligned} \tag{41}$$

The problem in Eq. (41) is equivalent to the following optimization problem:

$$\begin{aligned}\max_{\mathbf{Y}} tr\left(\left(\mathbf{Y}^T\mathbf{Y}\right)^{-1/2}\mathbf{Y}^T\mathbf{Q}\right)\\ s.t.\ Y_{ij} \in \{0,1\},\ \mathbf{Y}_{i,:}\mathbf{1}_c = 1\end{aligned} \tag{42}$$

where $\mathbf{Q} = \sum_{v=1}^{n_v}\mathbf{H}^{(v)}\mathbf{R}^{(v)}$. Eq. (42) can be rewritten as

$$\max_{\mathbf{Y}}\sum_{m=1}^{c}\frac{\sum_{i=1}^{n}y_{im}q_{im}}{\sqrt{\mathbf{y}_m^T\mathbf{y}_m}}\ s.t.\ y_{im} \in \{0,1\},\ \mathbf{Y}_{i,:}\mathbf{1}_c = 1 \tag{43}$$

where $\mathbf{y}_m$ is the $m$th column of $\mathbf{Y}$ and $q_{im}$ denotes the element in the $i$th row and $m$th column of $\mathbf{Q}$. Here, $\mathbf{Y}$ can be sequentially solved for row by row, and each row of $\mathbf{Y}$ has a closed-form solution.

**Algorithm 3** Solving (18) Using the ADMM Framework

**Input:** Data matrices $\left\{\mathbf{H}^{(v)}\right\}_{v=1}^{n_v}$ and $\left\{\mathbf{L}^{(v)}\right\}_{v=1}^{n_v}$ and parameters $\alpha$ and $\beta$.

**initialize:** $\mathcal{G}=\mathcal{R}=0$, $\rho=1.2$, $\mu=10^{-4}$, $\mu_{\max}=10^{6}$, $\varepsilon=10^{-6}$, $t=1$ and $maxIters=100$;
**while** not converged **do**
  Update $\mathbf{F}$ via Eq. (30);
  Update $\left\{\mathbf{R}^{(v)}\right\}_{v=1}^{n_v}$ via Eq. (33);
  Update $\left\{\mathbf{H}^{(v)}\right\}_{v=1}^{n_v}$ via Eq. (37);
  Update $\mathcal{G}$ via Eq. (38);
  Update the Lagrange multiplier $\mathcal{R}$ via Eq. (39);
  Update the parameter $\mu$ via Eq. (40);
  Check the convergence condition: $\|\mathcal{G}-\mathcal{T}\|_{\max}<\varepsilon$;
  **if** $t>maxIters$ or $converged$ **then**
    break;
  **end if**
  $t\leftarrow t+1$;
**end while**

**Output:** $\mathbf{F}$.

Specifically, each row of $\mathbf{Y}$ can be obtained in accordance with an incremental mechanism [6], [36]. Suppose that the optimal solution to Eq. (43) is $\widetilde{\mathbf{Y}}$ in the previous iteration step of the entire optimization procedure. Let $\widetilde{\mathbf{y}}_i$ be the $i$th row of $\widetilde{\mathbf{Y}}$. Then, we can employ an increment of the objective function value that ranges from $\widetilde{y}_{im}=0$ to $\widetilde{y}_{im}=1$ in Eq. (43). The incremental mechanism can be formulated as

$$\rho_{im}=\frac{\sum_{j=1}^{n}\widetilde{y}_{jm}q_{jm}+q_{im}\left(1-\widetilde{y}_{im}\right)}{\sqrt{\widetilde{\mathbf{y}}_m^T\widetilde{\mathbf{y}}_m+\left(1-\widetilde{y}_{im}\right)}}-\frac{\sum_{j=1}^{n}\widetilde{y}_{jm}q_{jm}-q_{im}\widetilde{y}_{im}}{\sqrt{\widetilde{\mathbf{y}}_m^T\widetilde{\mathbf{y}}_m-\widetilde{y}_{im}}}. \tag{44}$$

Let $\widehat{\mathbf{y}}_i$ be the optimal solution for the $i$th row of $\mathbf{Y}$, where the $w$th element is one and the other elements are zeros. The index $w$ can be calculated as

$$w=\arg\max_{1\le m\le c}\rho_{im}. \tag{45}$$

The iterative optimization procedure terminates during iterations when the convergence condition is satisfied, i.e., $\|\mathcal{G}-\mathcal{T}\|_{\max}<\varepsilon$, where $\varepsilon$ is a small constant, e.g., $\varepsilon=1e^{-6}$. According to the above analysis, the optimization procedures are similar for both of the objective functions in Eqs. (20) and (22). For simplicity, we summarize the entire procedure for solving the objective function in Eq. (20) in Algorithm 3.

### D. Theoretical Analysis

*1) Connections Between the Spectral Embedding Fusion Schemes at Two Different Data Levels:* Algorithm 4 summarizes the overall procedures of the proposed USETL framework. In Eq. (18), $f\left(\mathbf{F},\mathbf{H}^{(v)},\mathbf{R}^{(v)}\right)$ represents a spectral embedding fusion step for information fusion. $\mathbf{F}$ and $\mathbf{Y}$ represent the continuous clustering indicator matrix and the discrete clustering indicator matrix in Eqs. (20) and (22), respectively. From the perspective of continuity, it is reasonable to approximate $\mathbf{F}$ as $\mathbf{H}^{(v)}\mathbf{R}^{(v)}$ in Eq. (20) since $\mathbf{F}^T\mathbf{F}=\mathbf{I}_c$ and $\left(\mathbf{H}^{(v)}\mathbf{R}^{(v)}\right)^T\left(\mathbf{H}^{(v)}\mathbf{R}^{(v)}\right)=\mathbf{I}_c$.

**Algorithm 4** Algorithm for the USETL Framework

**Input:** Data matrices $\mathbf{X}=\left\{\mathbf{X}^{(v)}\right\}_{v=1}^{n_v}$, the number of clusters $c$, the number of neighbors $k$, and parameters $\alpha$ and $\beta$.

**for** $v=1$ to $n_v$ **do**
  Compute each individual affinity matrix $\mathbf{Z}^{(v)}$ for $\mathbf{X}^{(v)}$;
  Calculate the normalized Laplacian matrix $\mathbf{L}^{(v)}$ and the spectral embedding $\mathbf{H}^{(v)}$ using (1);
**end for**
Solve (20) using Algorithm 3 and obtain the optimal solution $\mathbf{F}$;
Apply the $k$-means algorithm to $\mathbf{F}$ to obtain $c$ clusters;

**Output:** The $c$ clusters.

However, approximating $\mathbf{Y}$ as $\mathbf{H}^{(v)}\mathbf{R}^{(v)}$ is difficult due to the discrete nature of $\mathbf{Y}$. Fortunately, the scaled clustering indicator matrix $\mathbf{Y}\left(\mathbf{Y}^T\mathbf{Y}\right)^{-1/2}$, denoted by $\mathbf{Y}_s$, can be considered as a good surrogate for $\mathbf{Y}$ because $\left(\mathbf{Y}_s\right)^T\left(\mathbf{Y}_s\right)=\mathbf{I}_c$. The surrogate $\mathbf{Y}_s$ guarantees a better approximation of $\mathbf{H}^{(v)}\mathbf{R}^{(v)}$ in Eq. (22). This indicates that $\mathbf{F}$ in Eq. (20) is a continuous solution for the scaled clustering indicator matrix $\mathbf{Y}_s$ in Eq. (22). Consequently, the two proposed spectral embedding fusion schemes within the USETL framework are closely connected, with the same spectral rotation term, i.e., $\mathbf{H}^{(v)}\mathbf{R}^{(v)}$.

*2) Convergence Analysis:* Algorithm 2 converges to a local maximum of the problem in Eq. (37), as previously proven in the literature [44]. In addition, proving the convergence of the process of solving the optimization problems in Eqs. (20) and (22) using the ADMM framework is still an open issue. Fortunately, a locally optimal solution to each subproblem in Eqs. (20) and (22) can be obtained in each iteration. The convergence condition $\|\mathcal{G}-\mathcal{T}\|_{\max}<\varepsilon$ for these optimization problems is often reached within dozens of iterations.

*3) Computational Complexity Analysis:* There are four subproblems in Algorithm 3. In the subproblem of $\mathbf{F}$, the computational complexity of calculating the SVD of $\mathbf{M}$ is $\mathcal{O}\left(nc^2\right)$. Similarly, the computational complexity of calculating the SVD of $\mathbf{N}$ in the subproblem of $\left\{\mathbf{R}^{(v)}\right\}_{v=1}^{n_v}$ is $\mathcal{O}\left(n_vc^3\right)$. The computational complexities of calculating the SVD of $\mathbf{P}^{(v)}$ and the eigenvalue decomposition of $\mathbf{P}^{(v)}$ in the subproblem of $\left\{\mathbf{H}^{(v)}\right\}_{v=1}^{n_v}$ are $O\left(n_vn^3\right)$ and $\mathcal{O}\left(t_1n_vn^3\right)$, respectively, where $t_1$ represents the number of iterations of Algorithm 2. The computational complexity of updating $\mathcal{G}$ in the subproblem of $\mathcal{G}$ is $\mathcal{O}\left(n_vn^2\log\left(n\right)+n_v^2n^2\right)$. The computational complexity of Algorithm 3 is $\mathcal{O}\left(t_2\left(nc^2+n_vc^3+(t_1+1)n_vn^3+n_vn^2\log\left(n\right)+n_v^2n^2\right)\right)$, where $t_2$ represents the number of iterations of Algorithm 3. Similarly, the computational complexities of initializing $\left\{\mathbf{L}^{(v)}\right\}_{v=1}^{n_v}$ and $\left\{\mathbf{H}^{(v)}\right\}_{v=1}^{n_v}$ in Algorithm 4 are $\mathcal{O}\left(n_vn^2\right)$ and $\mathcal{O}\left(cn_vn^2\right)$, respectively. The final computational complexity of Algorithm 4 is $\mathcal{O}\left(tn^3\right)$ if $n_v\ll n$ and $c\ll n$, where $t=t_1t_2$. In addition, the computational complexity of updating $\mathbf{Y}$ in the subproblem of $\mathbf{Y}$ is $\mathcal{O}\left(nc\right)$. Hence, the final computational complexity of Algorithm 4 remains unchanged when the updating scheme of $\mathbf{F}$ is replaced with that for $\mathbf{Y}$.

*4) Discussion:* An end-to-end framework for graph-based MVC usually merges the three stages of constructing similarity

TABLE II
STATISTICS OF EXPERIMENTAL DATASETS

| Dataset | Clusters | Views | Data samples | Data types |
|---|---|---|---|---|
| MSRC-v1 | 7 | 5 | 210 | Object images |
| Flower17 | 17 | 7 | 1,360 | Object images |
| COIL-20 | 20 | 3 | 1,440 | Object images |
| 100leaves | 100 | 3 | 1,600 | Plant properties |
| Handwritten | 10 | 6 | 2,000 | Digit images |
| Scene-15 | 15 | 3 | 4,485 | Scene images |

matrices, fusing spectral embedding matrices, and cluster partitioning into a single stage [4], [6]. For example, the consensus one-step multiview subspace clustering (COMVSC) method jointly optimizes subspace learning, the fusion of spectral representations and cluster partitioning [4]. Compared with COMVSC, our proposed embedding fusion schemes have two main advantages. On the one hand, the high-order correlations of spectral embeddings among multiple views contain a rich variety of consistent and complementary information. The spectral embedding fusion of multiple similarity graphs is implemented by means of spectral rotation, which is conducive to producing the distinct diagonal block submatrices in $\mathbf{H}^{(v)}\left(\mathbf{H}^{(v)}\right)^T$. On the other hand, the fusion of the spectral embedding matrices is independent of the construction of the similarity matrices. Thus, the computational efficiency of solving the optimization problems can be improved under the USETL framework since the optimization process does not involve solving for the similarity matrices.

## IV. EXPERIMENTS

In this section, we conducted extensive experiments to evaluate the performance of $\text{USETL}_{\mathbf{SEFL}}$ and $\text{USETL}_{\mathbf{CIL}}$. The source code for $\text{USETL}_{\mathbf{SEFL}}$ and $\text{USETL}_{\mathbf{CIL}}$ was implemented in MATLAB 2021b. The MATLAB source code is available online.[2] All of the experiments were conducted on a Windows 10 workstation with an Intel Core i7-10700 CPU and 32 GB of RAM.

### A. Experimental Settings

*1) Datasets:* The two spectral embedding fusion methods were experimentally evaluated on six publicly available multiview datasets. The statistics of the datasets used are summarized in Table II. Descriptions of these multiview datasets are given below.

- **MSRC-v1**[3] **dataset**: This dataset consists of 210 scene recognition images in 7 classes. Each image is represented by five different feature sets.
- **Flower17 dataset** [45]: This dataset contains 1,360 flower images in 17 categories. Each image is represented by seven feature sets.
- **COIL-20 dataset** [46]: This dataset consists of 1,440 images in 20 classes. Each image is represented by three different feature sets.
- **100leaves dataset** [47]: This dataset consists of 1,600 samples of 100 plant species. Shape descriptor, fine scale margin and texture histogram features have been extracted to represent each sample.
- **Handwritten dataset** [48]: This dataset contains 2,000 handwritten images of the digits 0 to 9. Each image is represented by six different feature sets.
- **Scene-15 dataset** [49]: This dataset contains 4,485 scene images in 15 categories from both indoor and outdoor environments. Each image is described by three features.

*2) Comparison Methods:* To validate the superiority of the proposed USETL framework, we compare $\text{USETL}_{\mathbf{SEFL}}$ and $\text{USETL}_{\mathbf{CIL}}$ with several state-of-the-art methods, including CGL [17], unified one-step multiview spectral clustering (UOMSC) [6], high-order correlation-preserved MVC (HCPMC) [22], GPCRL [20], GSR [28] and ASR [1]. Moreover, we consider two baselines for comparison. Specifically, we apply a standard spectral clustering (SC) method, NCut [19], to individual similarity matrices corresponding to each incomplete multiple view and to an accumulated affinity matrix calculated by aggregating all similarity matrices. These two baselines are called as the best single view (BSV) method and the $\text{SC}_{\mathbf{Agg}}$ method. The restrictions on the similarity matrix for these baselines are the same as those for the proposed USETL framework. In CGL and UOMSC, the missing entries in the similarity matrices are filled in with zeros. The source codes of the other competing methods were provided by their authors.

*3) Evaluation Metrics:* Following previous work [14], we employ three standard evaluation metrics to evaluate the clustering performance of all competing algorithms, the clustering accuracy (ACC), the normalized mutual information (NMI) and the $F$-measure. These metrics provide a comprehensive evaluation of all methods considered for comparison by measuring various properties of the clustering results. For each evaluation metric, a higher value indicates better clustering performance in the experiments.

*4) Parameter Settings:* The proposed USETL framework has two parameters, $\alpha$ and $\beta$. The values of both of these parameters were chosen from $\left\{1e^{-3}, 5e^{-3}, 0.01, 0.05, 0.1, 0.5, 1, 5\right\}$ in a grid search strategy. The parameter $k$, which represents the number of nearest neighbors in the ANGL method, was tuned in the range of $\{5, 7, 10, 15\}$ for the COIL-20 and Scene-15 datasets, whereas for the other datasets, we set $k = 5$ in the ANGL method. We repeated each experiment 10 times, and the mean values and standard deviations are reported. For the competing methods in which the $k$-means algorithm is applied in the final step, to ensure fair comparisons, we report the best clustering results obtained by tuning the parameters of these methods. The best and second-best mean values of the clustering results are shown in bold and underlined, respectively.

We first employed all the instances available in all views. Then, a certain percentage of the instances were randomly removed from each view. Specifically, the percentage of missing instances in each view was varied from 10% to 50% in intervals of 20%. In addition, the true number of clusters was assumed to be known for each dataset.

### B. Performance Evaluation

The clustering results of all methods on the six multiview datasets are reported in Table III. The $\text{SC}_{\mathbf{Agg}}$ method achieves

[2] https://codeocean.com/capsule/6355940/tree/v2
[3] https://www.microsoft.com/en-us/research/project/image-understanding/

TABLE III
CLUSTERING RESULTS (MEAN±STD.) OF DIFFERENT METHODS ON SIX MULTIVIEW DATASETS WITH VARIOUS MISSING DATA RATIOS

| Datasets | Methods | ACC | | | | NMI | | | | F-measure | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 10% | 30% | 50% | 0 | 10% | 30% | 50% | 0 | 10% | 30% | 50% |
| MSRC-v1 | BSV | 78.57±0.00 | 64.29±0.00 | 54.29±0.00 | 45.52±0.46 | 68.04±0.00 | 55.13±0.00 | 44.16±0.04 | 34.55±0.59 | 78.19±0.00 | 65.4±0.00 | 60.41±0.01 | 49.06±0.3 |
| | $SC_{Agg}$ | 82.71±1.23 | 81.9±0.00 | 80.48±0.00 | 70.48±0.00 | 72.52±1.56 | 72.72±0.00 | 69.19±0.00 | 58.72±0.00 | 82.73±1.15 | 81.24±0.00 | 79.98±0.00 | 69.27±0.00 |
| | UOMSC | 90.48±0.00 | 69.05±0.00 | 50±0.00 | 32.86±0.00 | 81.85±0.00 | 67.52±0.00 | 40.78±0.00 | 19.3±0.00 | 81.78±0.00 | 70.75±0.00 | 54.65±0.00 | 36.1±0.00 |
| | CGL | **100±0.00** | 90±0.00 | 86.67±0.00 | 36.1±0.01 | **100±0.00** | 84.51±0.00 | 81.63±0.42 | 28.69±0.95 | **100±0.00** | 81.54±0.00 | 78.26±0.38 | 26.35±0.63 |
| | HCPMC | 87.62±0.00 | 86.19±0.00 | 82.86±0.00 | 79.52±0.00 | 78.01±0.00 | 75.96±0.00 | 71.9±0.00 | 64.84±0.00 | 87.56±0.00 | 86.08±0.00 | 82.74±0.00 | 80.18±0.00 |
| | ASR | 91.9±0.00 | 92.86±0.00 | 87.14±0.00 | 80.48±0.00 | 84.75±0.00 | 85.36±0.00 | 77.01±0.00 | 67.18±0.00 | 91.9±0.00 | 92.83±0.00 | 87.35±0.00 | 81.43±0.00 |
| | GPCRL | 89.52±0.00 | 88.57±0.00 | 83.33±0.00 | 81.1±0.84 | 82.25±0.06 | 80.91±0.24 | 73.37±0.00 | 68.74±0.47 | 89.37±0.00 | 88.43±0.07 | 83.58±00.07 | 81.06±0.81 |
| | GSR | 99.05±0.00 | 99±0.15 | 98.57±0.00 | 94.76±0.00 | 98.19±0.00 | 98.08±0.34 | 97.12±0.00 | 92.36±0.00 | 99.05±0.00 | 99±0.16 | 98.55±0.00 | 94.71±0.00 |
| | $USETL_{SEFL}$ | **100±0** | **100±0** | **100±0** | 99.52±0 | **100±0** | **100±0** | **100±0** | 98.92±0 | **100±0** | **100±0** | **100±0** | 99.52±0 |
| | $USETL_{CIL}$ | **100±0** | **100±0** | **100±0** | **100±0** | **100±0** | **100±0** | **100±0** | **100±0** | **100±0** | **100±0** | **100±0** | **100±0** |
| Flower17 | BSV | 42.21±0.75 | 36.68±0.1 | 28.88±0.51 | 22.74±0.28 | 43.14±0.55 | 39.7±0.1 | 31.03±0.27 | 26.04±0.11 | 44.9±0.6 | 41±0.17 | 34.22±0.48 | 27.64±0.19 |
| | $SC_{Agg}$ | 42.5±0.51 | 46.05±0.73 | 44.31±0.27 | 45.46±0.16 | 46.15±0.21 | 48.3±0.4 | 46.77±0.31 | 42.46±0.16 | 46.6±0.25 | 50.22±0.91 | 48.74±0.33 | 48.04±0.4 |
| | UOMSC | 61.03±0.00 | 54.04±0.00 | 38.01±0.00 | 26.91±0.00 | 60.51±0.00 | 52.61±0.00 | 40.48±0.00 | 26.77±0.00 | 43.85±0.00 | 35.92±0.00 | 25.06±0.00 | 15.78±0.00 |
| | CGL | 87.49±0.00 | 63.63±0.02 | 13.09±0.01 | 12.1±0.00 | 88.29±0.04 | 68.1±0.27 | 6.58±0.5 | 51.4±0.14 | 80.88±0.05 | 53.88±0.44 | 6.86±0.18 | 6.61±0.07 |
| | HCPMC | 56±0.36 | 54.88±0.59 | 52.06±0.38 | 46.3±0.1 | 56.11±0.24 | 55.4±0.33 | 52.06±0.44 | 44.94±0.06 | 60.79±0.31 | 59.45±0.45 | 56.46±0.51 | 49.79±0.15 |
| | ASR | 68.27±0.19 | 64.59±0.17 | 62.07±0.26 | 52.17±0.22 | 65.63±0.20 | 61.94±0.16 | 58.8±0.23 | 48.1±0.32 | 71.16±0.17 | 68.17±0.2 | 64.78±0.26 | 55.56±0.19 |
| | GPCRL | 54.62±0.58 | 51.11±0.94 | 50.74±1.3 | 47.43±1.8 | 53.64±0.39 | 51.49±0.65 | 50.55±0.79 | 47.34±0.39 | 57±0.39 | 53.29±0.96 | 53.41±1.57 | 51.07±0.86 |
| | GSR | 92.09±0.27 | 91.42±0.05 | 91.63±0.05 | 91.33±0.05 | 91.22±0.29 | 89.44±0.08 | 89.25±0.06 | 89.13±0.08 | 92.1±0.27 | 91.44±0.05 | 91.63±0.05 | 91.34±0.05 |
| | $USETL_{SEFL}$ | 94.85±0.06 | 94.88±0.08 | 92.9±3.73 | 94.73±0.31 | 93.48±0.06 | 93.24±0.1 | 92.59±1.45 | 93.15±0.28 | 94.84±0.06 | 94.89±0.08 | 93.42±2.66 | 94.73±0.31 |
| | $USETL_{CIL}$ | **95.51±0.00** | **95±0.00** | **95.37±0.00** | **95.96±0.00** | **93.93±0.00** | **93.57±0.00** | **93.69±0.00** | **94.44±0.00** | **95.51±0.00** | **94.99±0.00** | **95.37±0.00** | **95.95±0.00** |
| COIL-20 | BSV | 85.03±0.75 | 84.38±1.03 | 64.22±1.16 | 46.64±0.85 | 92.05±0.44 | 84.14±0.62 | 65.93±0.45 | 51.56±0.35 | 87.39±0.75 | 87.37±1.11 | 69.94±0.87 | 53.03±0.46 |
| | $SC_{Agg}$ | 85.87±1.1 | 86.27±1.22 | 76.47±1.04 | 68.64±2 | 94.31±0.63 | 95.75±0.48 | 85.74±1.02 | 74.55±0.58 | 88.93±0.36 | 89.84±0.74 | 81.17±1.08 | 73.85±1.47 |
| | UOMSC | 80.69±0.00 | 80.42±0.00 | 65.56±0.00 | 37.36±0.00 | 92.68±0.00 | 88.79±0.00 | 79.31±0.00 | 50.93±0.00 | 81.66±0.00 | 79.48±0.00 | 65.11±0.00 | 27.62±0.00 |
| | CGL | 90.15±0.00 | 86.81±0.00 | 59.19±0.01 | 31.61±0.01 | 94.54±0.00 | 92.79±0.00 | 57.73±0.33 | 28.37±0.32 | 89.81±0.01 | 85.23±0.00 | 43.81±0.68 | 17.31±0.28 |
| | HCPMC | 74.01±0.27 | 75.47±0.49 | 72.38±0.05 | 61.42±2.09 | 81.55±0.3 | 81.93±0.31 | 79.93±0.11 | 66.55±1.21 | 75.13±0.15 | 75.37±0.37 | 75.99±0.7 | 67.11±1.84 |
| | ASR | 87.76±0.71 | 92.99±0.00 | 79.08±0.72 | 72.67±1.18 | 95.58±0.43 | 97.13±0.00 | 87.02±0.52 | 75.95±0.63 | 90.96±0.74 | 94.97±0.00 | 82.51±0.00 | 77.2±0.00 |
| | GPCRL | 88.98±0.06 | 84.96±0.04 | 83.97±1.94 | 78.15±2.14 | 91.63±0.19 | 89.69±0.01 | 86.86±1.07 | 82.07±0.79 | 89.03±0.1 | 86.05±0.04 | 84.38±1.08 | 80.4 ±1.21 |
| | GSR | 99.31±0.00 | **99.93±0.00** | 97.48±2.62 | **98.7±0.17** | 99.07±0.00 | **99.88±0.00** | 96.26±1.26 | **98.13±0.23** | 99.31±0.00 | **99.93±0.00** | 97.63±2.14 | **98.7±0.17** |
| | $USETL_{SEFL}$ | 97.81±0.14 | 98.14±0.54 | 96.55±2.72 | 97.58±3.77 | 96.74±0.20 | 97.14±0.71 | 95.88±1.13 | 97.35±2.31 | 97.81±0.14 | 98.14±0.54 | **96.8±1.92** | 97.73±3.27 |
| | $USETL_{CIL}$ | **100±0.00** | 99.03±0.00 | **97.64±0.00** | 96.11±0.00 | **100±0.00** | 98.35±0.00 | **96.49±0.00** | 95.36±0.00 | **100±0.00** | 99.03±0.00 | 97.64±0.00 | 96.11±0.00 |
| 100leaves | BSV | 63.15±1.3 | 56.59±1.48 | 42.49±1.08 | 31.45±0.64 | 82.54±0.25 | 76.57±0.58 | 64.57±0.66 | 56.2±0.4 | 66.14±0.82 | 61.64±1.29 | 51.04±1.02 | 40.9±0.44 |
| | $SC_{Agg}$ | 74.34±1.31 | 70.84±1.13 | 54.94±1.11 | 38.86±0.88 | 83.46±0.44 | 80.65±0.24 | 68.25±0.24 | 56.93±0.44 | 76.25±1.25 | 72.7±0.81 | 55.32±0.86 | 42.6±0.86 |
| | UOMSC | 99.38±0.00 | 78.38±0.00 | 45.75±0.00 | 28.38±0.00 | 99.54±0.00 | 88.87±0.00 | 71.87±0.00 | 59.36±0.00 | 98.73±0.00 | 68.5±0.00 | 53.26±0.00 | 32.7±0.00 |
| | CGL | **100±0.00** | 92.21±0.01 | 74.54±0.01 | 49.86±0.01 | **100±0.00** | 95.78±0.17 | 84.24±0.15 | 69.17±0.34 | **100±0.00** | 87.76±0.09 | 59.82±0.61 | 29.16±0.81 |
| | HCPMC | 77.74±1.68 | 73.47±1.3 | 56.38±0.96 | 39.61±0.79 | 89.15±0.89 | 85.93±0.59 | 73.91±0.38 | 61.92±0.33 | 79.63±1.52 | 75.27±1.25 | 59.08±0.82 | 43.75±0.6 |
| | ASR | 86.41±1.09 | 78.04±0.73 | 56.31±0.81 | 40.52±0.32 | 95.14±0.49 | 89.38±0.42 | 74.34±0.3 | 62.65±0.44 | 88.83±0.92 | 80.99±0.64 | 60.94±0.69 | 46.84±0.47 |
| | GPCRL | 86.66±2.22 | 84.58±1.99 | 81.21±1.76 | 80.3±1.6 | 88.65±0.66 | 85.11±0.6 | 83.79±0.55 | 81.15±0.52 | 83.82±1.79 | 81.65±1.58 | 79.98±1.38 | 78.15±1.37 |
| | GSR | 98.2±0.78 | 97.55±0.37 | 94.71±0.89 | 87.73±0.93 | 99.57±0.13 | 99.37±0.07 | 98.2±0.32 | 95.4±0.29 | 98.75±0.5 | 98.27±0.25 | 96.01±0.57 | 89.74±0.74 |
| | $USETL_{SEFL}$ | 95.54±1.27 | 93.85±0.66 | 91.61±1.63 | 89.84±1.03 | 99.02±0.3 | 98.68±0.16 | 97.56±0.33 | 96.42±0.38 | 96.75±0.95 | 95.53±0.53 | 93.58±1.16 | 91.96±1.01 |
| | $USETL_{CIL}$ | **100±0.00** | **99.94±0.00** | **97.94±0.00** | **96.44±0.00** | **100±0.00** | **99.95±0.00** | **98.33±0.00** | **97.88±0.00** | **100±0.00** | **99.94±0.00** | **97.94±0.00** | **96.74±0.00** |
| Handwritten | BSV | 75.35±0.00 | 74.25±0.02 | 64.55±0.00 | 48.83±0.04 | 74.07±0.00 | 70.58±0.02 | 62.44±0.00 | 46.15±0.07 | 75.12±0.00 | 79.73±0.02 | 74±0.00 | 59.01±0.07 |
| | $SC_{Agg}$ | 79.85±0.00 | 85.66±0.02 | 79.35±0.00 | 73.65±0.02 | 82.62±0.00 | 84.73±0.00 | 81.19±0.00 | 74.27±0.03 | 83.9±0.00 | 87.06±0.01 | 83.6±0.00 | 79.6±0.02 |
| | UOMSC | 98.2±0.00 | 70.95±0.00 | 48.5±0.00 | 21.35±0.00 | 95.88±0.00 | 80.85±0.00 | 50.8±0.00 | 15.89±0.00 | 96.39±0.00 | 71.4±0.00 | 45.34±0.00 | 16.59±0.00 |
| | CGL | 99.55±0.01 | 99.85±0.00 | 32.02±0.00 | 15.78±0.00 | 99.53±1.04 | 99.59±0.00 | 27.71±0.1 | 2.47±0.03 | 99.27±0.02 | 99.7±0.00 | 23.86±0.08 | 10.77±0.02 |
| | HCPMC | 81.17±0.04 | 89.17±0.05 | 84.17±0.09 | 70.31±0.08 | 83.9±0.09 | 83.89±0.04 | 77.87±0.09 | 60.82±0.08 | 84.94±0.04 | 89.17±0.05 | 84.37±0.08 | 74.84±0.07 |
| | ASR | 93.95±0.00 | 92.41±0.13 | 90.85±0.00 | 79.73±0.11 | 88.26±0.00 | 85.75±0.24 | 83.06±0.00 | 71.47±0.01 | 93.98±0.00 | 92.42±0.13 | 91±0.00 | 80.52±0.11 |
| | GPCRL | 88.95±0.19 | 87.78±0.17 | 86.49±0.05 | 82.3±0.09 | 81.53±0.2 | 79.86±0.18 | 78.15±0.09 | 75.69±0.07 | 88.96±0.19 | 87.77±0.17 | 86.57±0.04 | 82.41±0.11 |
| | GSR | 99.85±0.00 | 99.7±0.00 | 99.85±0.00 | 98.9±0.00 | 99.62±0.00 | 99.24±0.00 | 99.59±0.00 | 97.44±0.00 | 99.85±0.00 | 99.7±0.00 | 99.85±0.00 | 98.9±0.00 |
| | $USETL_{SEFL}$ | 99.9±0.00 | 99.9±0.00 | **99.9±0.00** | **99.85±0.00** | 99.76±0.00 | 99.73±0.00 | **99.76±0.00** | **99.59±0.00** | 99.9±0.00 | **99.9±0.00** | **99.9±0.00** | **99.85±0.00** |
| | $USETL_{CIL}$ | **100±0.00** | **99.95±0.00** | **99.9±0.00** | 99.1±0.00 | **100±0.00** | **99.86±0.00** | **99.76±0.00** | 98.04±0.00 | **100±0.00** | **99.95±0.00** | **99.9±0.00** | 99.1±0.00 |
| Scene-15 | BSV | 38.05±0.54 | 34.08±0.11 | 29.16±0.16 | 23.92±0.05 | 38.85±0.07 | 32.08±0.06 | 28.53±0.44 | 29.07±0.03 | 40.49±0.27 | 36.22±0.09 | 33.2±0.23 | 27.5±0.03 |
| | $SC_{Agg}$ | 38.13±0.11 | 40.16±0.04 | 37.87±0.1 | 34.49±0.09 | 39.31±0.09 | 39.24±0.05 | 35.82±0.05 | 28.55±0.04 | 41.18±0.08 | 43.15±0.04 | 40.73±0.08 | 37.06±0.08 |
| | UOMSC | 48.52±0.00 | 32.06±0.00 | 24.44±0.00 | 14.29±0.00 | 52.45±0.00 | 36.08±0.00 | 22.46±0.00 | 11.75±0.00 | 37.34±0.00 | 28.04±0.00 | 18.49±0.00 | 9.96±0.00 |
| | CGL | 43.03±0.00 | 33.47±0.01 | 13.23±0.00 | 9.16±0.00 | 40.95±0.18 | 28.56±0.28 | 6.88±0.14 | 1.06±0.04 | 30.19±0.16 | 22.27±0.28 | 8.95±0.5 | 7.43±0.13 |
| | HCPMC | 37.05±0.07 | 36.36±0.05 | 34.62±0.03 | 29.93±0.29 | 33.95±0.03 | 32.67±0.03 | 27.48±0.03 | 23.46±0.05 | 39.22±0.04 | 38.43±0.03 | 34.95±0.02 | 32.63±0.22 |
| | ASR | 49.71±0.26 | 45.48±1.24 | 34.59±0.72 | 27.48±0.14 | 47.74±0.17 | 39.88±0.71 | 32.6±0.5 | 27.71±0.28 | 50.81±0.27 | 46.44±0.9 | 38.48±0.27 | 31.9±0.22 |
| | GPCRL | 44.05±0.43 | 41.26±0.14 | 35.33±0.68 | 28.62±0.43 | 40.34±0.12 | 37.96±0.13 | 32.9±0.27 | 25.81±0.23 | 45.16±0.34 | 42.36±0.15 | 36.23±0.67 | 29.71±0.31 |
| | GSR | 93.79±0.02 | 94.8±0.07 | 93.44±0.02 | 84.92±1.62 | 91.2±0.03 | 91.18±0.05 | 89.21±0.03 | 84.77±0.34 | 93.78±0.02 | 94.81±0.07 | 93.45±0.02 | 85.52±01.09 |
| | $USETL_{SEFL}$ | **94.17±0.1** | **95.27±0.05** | **96.11±0.03** | **94.48±0.02** | **92.32±0.08** | **93.43±0.06** | **94.25±0.03** | **92.71±0.02** | **94.2±0.1** | **95.3±0.05** | **96.11±0.03** | **94.52±0.02** |
| | $USETL_{CIL}$ | 87.38±0.00 | 89.74±0.00 | 88.49±0.00 | 91.59±0.00 | 90.23±0.00 | 93.03±0.00 | 91.22±0.00 | 90.59±0.00 | 88.32±0.00 | 91.63±0.00 | 89.79±0.00 | 91.69±0.00 |

better clustering results than does the BSV method in most cases. This finding indicates that capturing consistent and complementary information among multiple views is worthy of investigation. On most of the datasets, the $USETL_{CIL}$ method consistently outperforms the other competing methods, including both the state-of-the-art methods and the two baselines mentioned above. For example, the $USETL_{CIL}$ method achieves performance improvements of approximately 0.95%, 1.81% and 0.95% over the second-best method (GSR) in terms of the ACC, NMI, and F-measure, respectively, on the MSRC-v1 dataset with a missing data rate of 0. Similarly, the $USETL_{SEFL}$ method performs much better than the other competing methods on the MSRC-v1, Flower17, Handwritten and Scene-15 datasets. For example, the $USETL_{SEFL}$ method significantly outperforms the other competing methods by at least 2.76%, 2.26% and 2.74% in terms of the ACC, NMI, and F-measure, respectively, on all available instances in the Flower17 dataset. Moreover, we observe the same advantages of the $USETL_{SEFL}$ and $USETL_{CIL}$ methods as the missing data rate gradually increases in most of the datasets.

For example, the USETL$_{\mathbf{CIL}}$ method achieves performance improvements of approximately 3.58%, 3.74% and 4.63% over the second-best method (GSR) in terms of the ACC on the Flower17 dataset with missing data ratios of 10%, 30% and 50%, respectively. These experimental results demonstrate the superiority of the proposed USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods over the other competing methods. In addition, USETL$_{\mathbf{CIL}}$ performs slightly better than USETL$_{\mathbf{SEFL}}$ in most cases. This indicates that the discrete solution in the USETL framework often provides better approximations of rotated spectral embedding matrices than the continuous solution can.

There are three primary reasons for the advantages and effectiveness of the proposed USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods. First, we introduce spectral embedding tensor learning to explore high-order correlations among multiple views. The introduction of a graph learning-based low-rank tensor is beneficial for capturing consistent and complementary information. This also explains why the GSR method achieves encouraging clustering results on all the datasets with a missing rate of 0. In contrast, the clustering performance of the HCPMC method, which also involves low-rank tensor learning, is unsatisfactory due to the lack of graph learning. Unlike these methods, ASR employs discriminative sparse representation learning to capture the local structures of incomplete multiview data. It shows advantages in clustering performance over HCPMC in most cases. Second, the two proposed schemes for the spectral embedding fusion of multiple similarity graphs, which yield spectral embedding fusion results at two different data levels, are implemented by means of spectral rotation. In contrast to the CGL and GSR methods, both spectral embedding graph learning and spectral embedding fusion are integrated into the USETL framework. These two spectral embedding fusion schemes yield superior results for the fusion of multiple views using the USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods. In a one-step manner, USETL$_{\mathbf{CIL}}$ shows better generalization ability than the UOMSC method on IMVC tasks. Third, the strategy of removing missing instances from incomplete multiple views is adopted in the USETL framework. This is a feasible scheme for constructing multiple similarity graphs in the USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods. In most cases, the clustering performance of the ASR and HCPMC methods slowly declines as the missing rate increases from 0 to 50%. In comparison, the clustering performance of the UOMSC and CGL methods decreases dramatically as the missing rate increases.

Table IV shows the average running times of all of the competing state-of-the-art algorithms on the six multiview datasets with different missing data ratios. The computation times of the BSV and SC$_{\mathbf{Agg}}$ methods are not included because they are considered only as baselines for evaluating the clustering performance in terms of the three standard evaluation metrics. The GPCRL method often performs more efficiently than the other methods. In addition, the USETL$_{\mathbf{CIL}}$ method achieves the best clustering performance at a running time comparable to those of GSR, CGL and HCPMC. The running time of USETL$_{\mathbf{SEFL}}$ is approximately the same as that of USETL$_{\mathbf{CIL}}$ for all the datasets. Notably, these methods involve SVD or eigenvalue decomposition on a matrix of size $n \times n$, which is time-consuming. Nevertheless, we observe that the running times of the USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods are lower than those of the GSR and CGL methods on all the datasets.

TABLE IV
COMPUTATION TIMES (IN SECONDS) OF DIFFERENT METHODS ON SIX MULTIVIEW DATASETS WITH VARIOUS MISSING DATA RATIOS

| Datasets | Ratio | UOMSC | CGL | HCPMC | ASR | GPCRL | GSR | USETL$_{\mathbf{SEFL}}$ | USETL$_{\mathbf{CIL}}$ |
|---|---|---|---|---|---|---|---|---|---|
| MSRC-v1 | 0 | 0.35 | 2.76 | 0.88 | 0.69 | **0.12** | 2.67 | 2.69 | 2.73 |
| | 10% | 0.8 | 2.01 | 0.58 | 0.59 | **0.09** | 2.24 | 1.93 | 1.96 |
| | 30% | 0.85 | 2.21 | 0.49 | 0.4 | **0.09** | 1.82 | 1.59 | 1.62 |
| | 50% | 0.71 | 2.03 | 0.61 | 0.24 | **0.08** | 1.89 | 1.51 | 1.26 |
| Flower17 | 0 | 5.56 | 107.9 | 45.53 | 29.55 | **3.32** | 58.17 | 51.93 | 50.72 |
| | 10% | 10.84 | 106.7 | 49.4 | 25.39 | **3.16** | 65.81 | 51.95 | 50.77 |
| | 30% | 8.79 | 106.5 | 51.18 | 18.95 | **2.69** | 63.27 | 60 | 58.26 |
| | 50% | 9.35 | 105.7 | 53.8 | 14.07 | **2.42** | 61.01 | 45.54 | 57.38 |
| COIL-20 | 0 | **5.47** | 54.65 | 42.19 | 297.7 | 26.6 | 35.33 | 20.35 | 30.21 |
| | 10% | **7.97** | 55.83 | 43.55 | 244.2 | 27.77 | 31.09 | 21.98 | 18.32 |
| | 30% | **6.82** | 56.25 | 45.27 | 161.9 | 28.02 | 26.53 | 22.02 | 22.76 |
| | 50% | **7.56** | 56.56 | 47.36 | 61.08 | 29.86 | 30.12 | 26.14 | 29.19 |
| 100leaves | 0 | 3.7 | 187 | 41.33 | 64.63 | **0.93** | 50.63 | 37.73 | 39.73 |
| | 10% | 9.32 | 166.4 | 41.41 | 52.96 | **0.92** | 46.86 | 37.7 | 39.81 |
| | 30% | 16.53 | 174.1 | 42.43 | 35.64 | **0.9** | 43.12 | 38.32 | 35 |
| | 50% | 14.7 | 162 | 43.74 | 28.79 | **0.88** | 45.79 | 39.83 | 40.54 |
| Handwritten | 0 | 13.7 | 239.1 | 99.02 | 50.75 | **2.41** | 178.5 | 116.3 | 163.7 |
| | 10% | 35.72 | 234.6 | 100.2 | 44.16 | **2.11** | 188.9 | 124.8 | 125.3 |
| | 30% | 36.39 | 234.3 | 103.5 | 26.92 | **1.6** | 190.2 | 119 | 121.3 |
| | 50% | 30.4 | 233.3 | 104.8 | 14.12 | **1.37** | 164.3 | 123.7 | 118.1 |
| Scene-15 | 0 | 60.48 | 663.4 | 616.5 | 513.9 | **8.93** | 361.9 | 332.6 | 346.4 |
| | 10% | 131.8 | 662.4 | 618.4 | 391.6 | **8.84** | 370.1 | 350.8 | 288.4 |
| | 30% | 149 | 651 | 621.1 | 200.2 | **7.74** | 399.8 | 317.6 | 344.5 |
| | 50% | 137 | 627.2 | 626.5 | 91.84 | **6.8** | 355.7 | 293.2 | 328.4 |

### C. Empirical Study on the Construction of Multiple Similarity Graphs

The construction of multiple similarity graphs is a preprocessing step for the USETL framework. For this step, we employ the ANGL method, which requires an extra parameter, i.e., the number of nearest neighbors $k$. Hence, we investigated the effect induced by varying the number of nearest neighbors considered in the USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods. The parameters $\alpha$ and $\beta$ were set in accordance with the USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods in Section IV-A.4. The number of nearest neighbors $k$ was chosen from {5, 7, 10, 15}. The ACC was employed to evaluate the clustering quality under different numbers of nearest neighbors.

Figs. 1 and 2 show that the proposed USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods achieve high average ACC values with different numbers of nearest neighbors on the Flower17, 100leaves, and Handwritten datasets. The ACC value remains relatively stable with varying numbers of nearest neighbors as the missing data ratio in each dataset increases from 0 to 30%. However, the ACC is slightly affected when the missing data ratio increases to 50% in a few datasets, e.g., the COIL-2 and Scene-15 datasets. These findings indicate that the performance of the proposed USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods is fairly robust to the choice of the number of nearest neighbors in the ANGL method.

### D. Parameter Sensitivity Analysis

Due to space limitations, we conducted experiments to investigate the parameter sensitivity only on the COIL-20 and

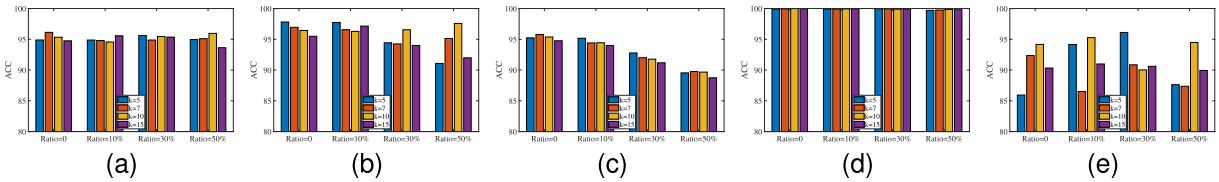



Fig. 1. Comparison of the ACC values of the USETL$_{\mathbf{SEFL}}$ method with different numbers of nearest neighbors $k$ on five datasets: (a) Flower17, (b) COIL-20, (c) 100leaves, (d) Handwritten, and (e) Scene-15.

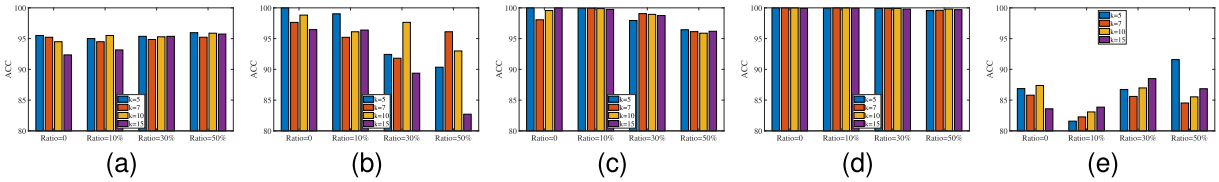



Fig. 2. Comparison of the ACC values of the USETL$_{\mathbf{CIL}}$ method with different numbers of nearest neighbors $k$ on five datasets: (a) Flower17, (b) COIL-20, (c) 100leaves, (d) Handwritten, and (e) Scene-15.

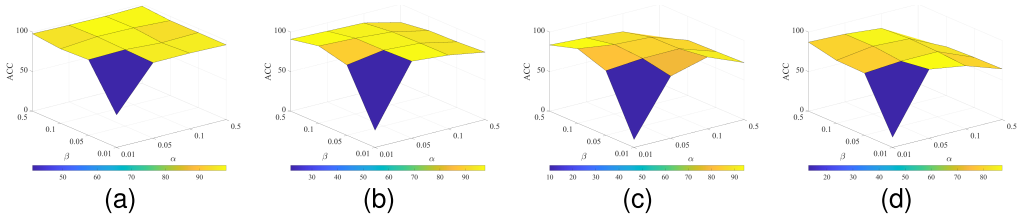



Fig. 3. The ACC of the USETL$_{\mathbf{SEFL}}$ method with different combinations of $\alpha$ and $\beta$ on the COIL-20 dataset for various missing data ratios: (a) $Ratio = 0$, (b) $Ratio = 10\%$, (c) $Ratio = 30\%$, and (d) $Ratio = 50\%$.

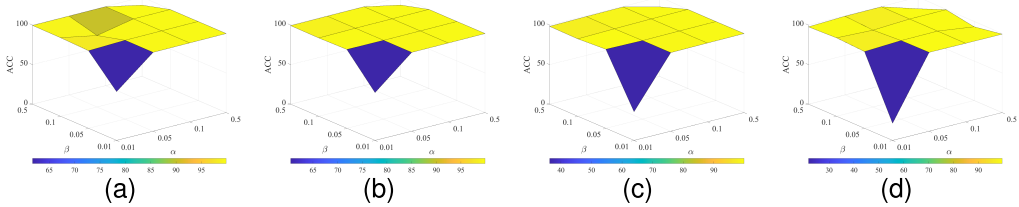



Fig. 4. The ACC of the USETL$_{\mathbf{SEFL}}$ method with different combinations of $\alpha$ and $\beta$ on the Handwritten dataset for various missing data ratios: (a) $Ratio = 0$, (b) $Ratio = 10\%$, (c) $Ratio = 30\%$, and (d) $Ratio = 50\%$.

Handwritten datasets. The parameters $\alpha$ and $\beta$ were chosen from $\{0.01, 0.05, 0.1, 0.5\}$ for USETL$_{\mathbf{SEFL}}$, whereas they were chosen from $\{0.05, 0.1, 0.5, 1\}$ and $\{1e^{-3}, 5e^{-3}, 0.01, 0.05\}$, respectively, for USETL$_{\mathbf{CIL}}$. We also set $k = 5$ in accordance with the information in Section IV-A.4. The ACC values of the USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods are reported with respect to different combinations of $\alpha$ and $\beta$.

Figs. 3-6 show the clustering performance of the USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods in terms of the ACC for different combinations of $\alpha$ and $\beta$ on the COIL-20 and Handwritten datasets. As seen from Figs. 3 and 4, the USETL$_{\mathbf{SEFL}}$ method can achieve stable performance in most combinations when the missing data ratio is 0. However, poor performance may be achieved when both $\alpha$ and $\beta$ are very small, e.g., $\alpha = 0.01$ and $\beta = 0.01$. This is mainly because the corresponding regularization term have very little influence if $\alpha$ and $\beta$ in Eq. (18) are very small. As also shown in Figs. 5 and 6, the clustering performance of the USETL$_{\mathbf{CIL}}$ method fluctuates slightly with different combinations of $\alpha$ and $\beta$. In addition, an increase in the missing data ratio from 0 to 50% only slightly affects the ACC for the USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods. In these experiments, the various specified percentages of instances were randomly removed from each view when constructing multiple similarity graphs. Therefore, the experimental results illustrate the effectiveness of the strategy of constructing multiple similarity graphs for incomplete multiple views.

### *E. Convergence Analysis*

To validate the convergence of the USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods, we computed the values of the convergence condition $\|\mathcal{G} - \mathcal{T}\|_{\max}$ during the iterative process of each method. Figs. 8 and 7 show the convergence curves of the proposed USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$ methods, respectively, on all datasets with different missing data ratios. In most cases, the convergence condition values of the proposed methods slowly decrease before the first 30 iterations and then drop dramatically until convergence. However, we also observe that a few convergence curves fluctuate several times before

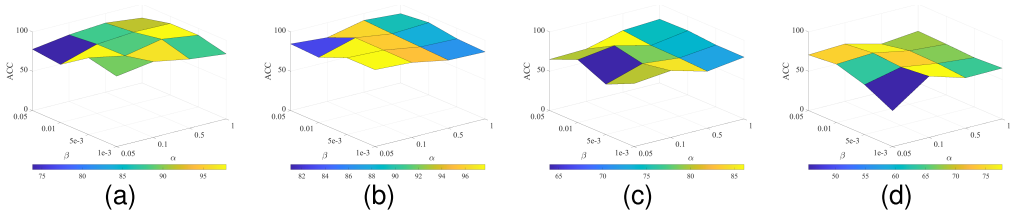



Fig. 5. The ACC of the USETL$_{\mathbf{CIL}}$ method with different combinations of $\alpha$ and $\beta$ on the COIL-20 dataset for various missing data ratios: (a) $Ratio = 0$, (b) $Ratio = 10\%$, (c) $Ratio = 30\%$, and (d) $Ratio = 50\%$.

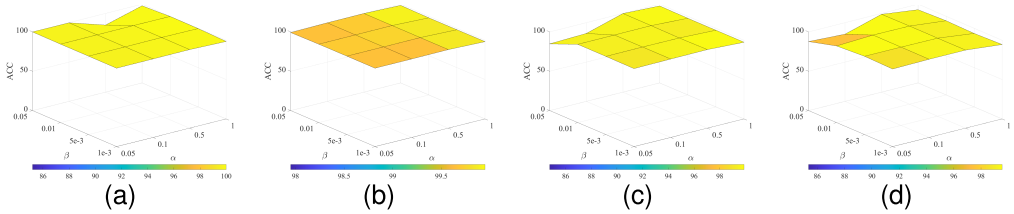



Fig. 6. The ACC of the USETL$_{\mathbf{CIL}}$ method with different combinations of $\alpha$ and $\beta$ on the Handwritten dataset for various missing data ratios: (a) $Ratio = 0$, (b) $Ratio = 10\%$, (c) $Ratio = 30\%$, and (d) $Ratio = 50\%$.

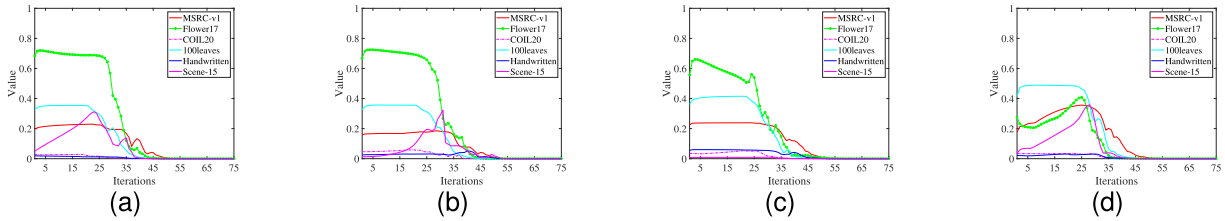



Fig. 7. Convergence results of the USETL$_{\mathbf{SEFL}}$ method on all datasets with various missing data ratios: (a) $Ratio = 0$, (b) $Ratio = 10\%$, (c) $Ratio = 30\%$, and (d) $Ratio = 50\%$.

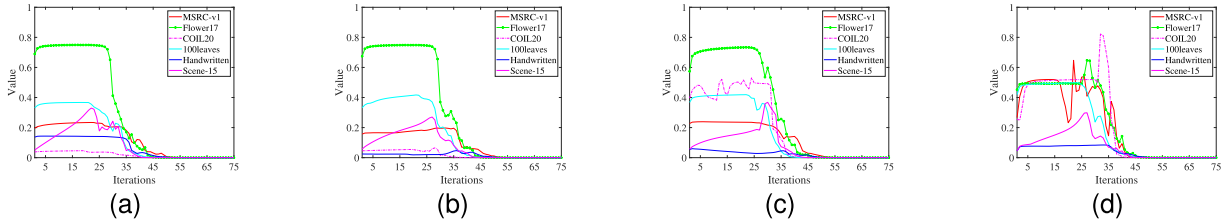



Fig. 8. Convergence results of the USETL$_{\mathbf{CIL}}$ method on all datasets with various missing data ratios: (a) $Ratio = 0$, (b) $Ratio = 10\%$, (c) $Ratio = 30\%$, and (d) $Ratio = 50\%$.

convergence, especially for relatively large missing data ratios, e.g., $ratio = 50\%$. One reason is that the value converges to a local minimum in each step in the proposed methods. Second, the differences between different initializations of $\{\mathbf{H}^{(v)}\}_{v=1}^{n_v}$ become more significant as the missing data ratio increases. Nevertheless, the number of iterations before convergence is always in the dozens for the proposed methods. This demonstrates that the proposed methods exhibit stable convergence behavior.

## V. Conclusion

In this paper, we propose the USETL framework for IMVC, which integrates the spectral embedding fusion of multiple similarity graphs and spectral embedding tensor learning. Within the USETL framework, we exploit spectral rotation to perform spectral embedding fusion at two different data levels, i.e., the spectral embedding feature level and the clustering indicator level, which correspond to two spectral embedding fusion methods, USETL$_{\mathbf{SEFL}}$ and USETL$_{\mathbf{CIL}}$, respectively. An alternating iterative optimization procedure is used to solve the two corresponding objective functions. The redundant information in the original incomplete multiview data is effectively removed when performing spectral embedding fusion at these two data levels. Simultaneously, spectral embedding tensor learning is introduced to find high-order correlations among multiple views. This is conducive to capturing consistent and complementary information among multiple views. Moreover, multiple similarity graphs are constructed using the ANGL method, and in this construction process, the missing instances are removed from the incomplete multiple views. The two proposed spectral embedding fusion methods are robust to the construction of multiple similarity graphs. Extensive experiments conducted on multiview datasets demonstrate the superiority of these two spectral embedding fusion methods over several state-of-the-art approaches.

## REFERENCES


[1] J. Chen, S. Yang, X. Peng, D. Peng, and Z. Wang, "Augmented sparse representation for incomplete multiview clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 3, pp. 4058–4071, Mar. 2024.

[2] 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.

[3] Z. Han, C. Zhang, H. Fu, and J. T. Zhou, "Trusted multi-view classification with dynamic evidential fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2551–2566, Feb. 2023.

[4] P. Zhang et al., "Consensus one-step multi-view subspace clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 10, pp. 4676–4689, Oct. 2022.

[5] J. Chen, S. Yang, H. Mao, and C. Fahy, "Multiview subspace clustering using low-rank representation," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 12364–12378, Nov. 2022.

[6] C. Tang, Z. Li, J. Wang, X. Liu, W. Zhang, and E. Zhu, "Unified one-step multi-view spectral clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 6, pp. 6449–6460, Jun. 2023.

[7] X. Peng, Y. Li, T. Ivor, H. Zhu, J. Lv, and J. T. Zhou, "XAI beyond classification: Interpretable neural clustering," *J. Mach. Learn. Res.*, vol. 23, no. 6, pp. 1–28, Feb. 2022.

[8] Y. Li, M. Yang, D. Peng, T. Li, J. Huang, and X. Peng, "Twin contrastive learning for online clustering," *Int. J. Comput. Vis.*, vol. 130, no. 9, pp. 2205–2221, Sep. 2022.

[9] Q. Wang, Z. Tao, W. Xia, Q. Gao, X. Cao, and L. Jiao, "Adversarial multiview clustering networks with adaptive fusion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 7635–7647, Oct. 2023.

[10] Q. Gao, W. Xia, Z. Wan, D. Xie, and P. Zhang, "Tensor-SVD based graph learning for multi-view subspace clustering," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 3930–3937.

[11] Y. Wang, W. Huang, F. Sun, T. Xu, Y. Rong, and J. Huang, "Deep multimodal fusion by channel exchanging," in *Proc. 34th Adv. Neural. Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2020, pp. 4835–4845.

[12] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, May 2015, pp. 586–594.

[13] Z. Tao, J. Li, H. Fu, Y. Kong, and Y. Fu, "From ensemble clustering to subspace clustering: Cluster structure encoding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 5, pp. 2670–2681, May 2023.

[14] J. Chen, H. Mao, Z. Wang, and X. Zhang, "Low-rank representation with adaptive dictionary learning for subspace clustering," *Knowl.-Based Syst.*, vol. 223, pp. 1–12, Jul. 2021.

[15] Y. Xie et al., "Robust kernelized multiview self-representation for subspace clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 868–881, Feb. 2021.

[16] Y. Chen, S. Wang, C. Peng, Z. Hua, and Y. Zhou, "Generalized nonconvex low-rank tensor approximation for multi-view subspace clustering," *IEEE Trans. Image Process.*, vol. 30, pp. 4022–4035, 2021.

[17] Z. Li, C. Tang, X. Liu, X. Zheng, W. Zhang, and E. Zhu, "Consensus graph learning for multi-view clustering," *IEEE Trans. Multimedia*, vol. 24, pp. 2461–2472, 2022.

[18] K. Zhan, F. Nie, J. Wang, and Y. Yang, "Multiview consensus graph clustering," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1261–1270, Mar. 2019.

[19] J. Shi, J. Malik, and S. Sastry, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 181–214, Aug. 2000.

[20] S. Deng, J. Wen, C. Liu, K. Yan, G. Xu, and Y. Xu, "Projective incomplete multi-view clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Feb. 10, 2023, doi: 10.1109/TNNLS.2023.3242473.

[21] C. Liu, Z. Wu, J. Wen, Y. Xu, and C. Huang, "Localized sparse incomplete multi-view clustering," *IEEE Trans. Multimedia*, vol. 25, pp. 5539–5551, 2022.

[22] Z. Li, C. Tang, X. Zheng, X. Liu, W. Zhang, and E. Zhu, "High-order correlation preserved incomplete multi-view subspace clustering," *IEEE Trans. Image Process.*, vol. 31, pp. 2067–2080, 2022.

[23] J. Wen et al., "A survey on incomplete multiview clustering," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 53, no. 2, pp. 1–14, Aug. 2022.

[24] M. Yang, Y. Li, P. Hu, J. Bai, J. Lv, and X. Peng, "Robust multi-view clustering with incomplete information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 1055–1069, Jan. 2023.

[25] X. Liu et al., "Efficient and effective regularized incomplete multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2634–2646, Aug. 2021.

[26] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[27] J. Chen, H. Mao, D. Peng, C. Zhang, and X. Peng, "Multiview clustering by consensus spectral rotation fusion," *IEEE Trans. Image Process.*, vol. 32, pp. 5153–5166, 2023.

[28] X. Li, M. Chen, C. Wang, and J. Lai, "Refining graph structure for incomplete multi-view clustering," *IEEE Trans. Neural. Netw. Learn. Syst.*, vol. 35, no. 2, pp. 2300–2313, Feb. 2024.

[29] Z. Lv, Q. Gao, X. Zhang, Q. Li, and M. Yang, "View-consistency learning for incomplete multiview clustering," *IEEE Trans. Image Process.*, vol. 31, pp. 4790–4802, 2022.

[30] 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.

[31] M. Hu and S. Chen, "One-pass incomplete multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 3838–3845.

[32] S. Wang et al., "Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jul. 2022, pp. 9776–9785.

[33] J. Wen et al., "Unified tensor framework for incomplete multi-view clustering and missing-view inferring," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 10273–10281.

[34] Y. Pang, J. Xie, F. Nie, and X. Li, "Spectral clustering by joint spectral embedding and spectral rotation," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 247–258, Jan. 2020.

[35] J. Huang, F. Nie, and H. Huang, "Spectral rotation versus $k$-means in spectral clustering," in *Proc. AAAI Conf. Artif. Intell.*, Washington, DC, USA, Jul. 2013, pp. 431–437.

[36] X. Chen, F. Nie, J. Z. Huang, and M. Yang, "Scalable normalized cut with improved spectral rotation," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 1518–1524.

[37] Z. Wang, X. Dai, P. Zhu, R. Wang, X. Li, and F. Nie, "Fast optimization of spectral embedding and improved spectral rotation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 1515–1527, Feb. 2023.

[38] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, Apr. 2020.

[39] 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.

[40] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra its Appl.*, vol. 435, no. 3, pp. 641–658, Aug. 2011.

[41] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[42] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2014, pp. 977–986.

[43] S. Boyd et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[44] F. Nie, R. Zhang, and X. Li, "A generalized power iteration method for solving quadratic problem on the Stiefel manifold," *Sci. China Inf. Sci.*, vol. 60, no. 11, pp. 1–10, Nov. 2017.

[45] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New York, NY, USA, Jun. 2006, pp. 1447–1454.

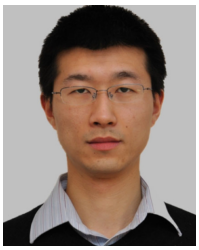
[46] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (COIL-20)," Dept. Comput. Sci., Columbia Univ., New York, NY, USA, Tech. Rep. CUCS-005-96, Feb. 1996.

[47] C. Mallah, J. Cope, and J. Orwell, "Plant leaf classification using probabilistic integration of shape, texture and margin features," *Signal Process. Pattern Recognit. Appl.*, vol. 5, no. 1, pp. 45–54, 2013.

[48] A. Asuncion and D. Newman, "UCI. Machine learning repository," School Inf. Comput. Sci., Univ. California, Oakland CA, USA, 2007. [Online]. Available: http://archive.ics.uci.edu/

[49] F. Li and P. Pietro, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Jun. 2005, pp. 524–531.

**Jie Chen** (Member, IEEE) received the B.Sc. degree in software engineering and the M.Sc. and Ph.D. degrees in computer science from Sichuan University, Chengdu, China, in 2005, 2008, and 2014, respectively. From 2008 to 2009, he was a Software Engineer with Huawei Technologies Company Ltd. He is currently an Associate Professor with the College of Computer Science, Sichuan University. His current research interests include machine learning, big data analysis, and deep neural networks.



**Yingke Chen** received the Ph.D. degree from Aalborg University, Denmark. He did postdoctoral research with Queen's University Belfast, U.K., and Georgia University, USA. He is currently an Associate Professor of computer science with Northumbria University, U.K. His research has led to publications in the *Journal of Artificial Intelligence Research*, AAMAS, AAAI, and IJCAI conferences. His current research interests include intelligent agents, data driven decision making, and their applications in autonomous systems.

**Zhu Wang** received the B.M., L.L.M., and L.L.D. degrees in civil and commercial law from the Renmin University of China, China, in 2003, 2006, and 2009, respectively. He is currently a Professor of law with Sichuan University and the Director of the Key Laboratory of Sichuan Province for AI Empowered Governance in Smart Society (AegisS). His research interests include civil and commercial law and AI in law.

**Haixian Zhang** (Member, IEEE) received the Ph.D. degree in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 2010. She is currently an Associate Professor with the College of Computer Science, Sichuan University, Chengdu. Her current research interests include neural networks and medical intelligence.

**Xi Peng** (Senior Member, IEEE) is currently a Full Professor with the College of Computer Science, Sichuan University. His current research interests include machine intelligence. He has authored more than 50 articles in these areas. He has served as the Area Chair/Senior Program Committee Member for the conferences, such as IJCAI, AAAI, and ICME. He has served as an Associate Editor/Guest Editor for six journals, including IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS and IEEE TRANSACTIONS ON NEURAL NETWORK AND LEARNING SYSTEMS.