Low-Rank Kernel Tensor Learning for Incomplete Multi-View Clustering

Tingting Wu[1,2], **Songhe Feng**[1,2*], **Jiazheng Yuan**[3*]
[1]Tangshan Research Institute, Beijing Jiaotong University, Beijing, China
[2]Key Laboratory of Big Data and Artificial Intelligence in Transportation, Ministry of Education, Beijing, China
[3]College of Science and Technology, Beijing Open University, Beijing, China
{22120438, shfeng}@bjtu.edu.cn, jzyuan@139.com

Abstract

Incomplete Multiple Kernel Clustering algorithms, which aim to learn a common latent representation from preconstructed incomplete multiple kernels from the original data, followed by k-means for clustering. They have attracted intensive attention due to their high computational efficiency. However, our observation reveals that the imputation of these approaches for each kernel ignores the influence of other incomplete kernels. In light of this, we present a novel method called Low-Rank Kernel Tensor Learning for Incomplete Multiple Views Clustering (LRKT-IMVC) to address the above issue. Specifically, LRKT-IMVC first introduces the concept of kernel tensor to explore the inter-view correlations, and then the low-rank kernel tensor constraint is used to further capture the consistency information to impute missing kernel elements, thereby improving the quality of clustering. Moreover, we carefully design an alternative optimization method with promising convergence to solve the resulting optimization problem. The proposed method is compared with recent advances in experiments with different missing ratios on seven well-known datasets, demonstrating its effectiveness and the advantages of the proposed interpolation method.

Introduction

Multi-View Clustering (MVC) has aroused extensive research enthusiasm due to its ability to enhance clustering performance by exploiting consistent and complementary information from different viewpoints (Zhao, Kwok, and Zhang 2009). There are a number of successful multi-view clustering methods that have been proposed and developed over the past decade(Yu et al. 2011; Guo 2013; Li, Jiang, and Zhou 2014; Gao et al. 2015; Tao, Liu, and Fu 2017; Zhang et al. 2018; Chen et al. 2020; Huang et al. 2021; Wu et al. 2021; Huang et al. 2022; Zhang et al. 2022; Wu et al. 2023a; Dong et al. 2023b,a; Duan et al. 2023). Although all methods obtain great clustering performance, most of them assume that all views of the samples are complete. However, in real scenarios, partial views between samples may be missing. There has been considerable application of existing MVC techniques to incomplete views, but this has resulted in significant performance degradation or even complete failure. Thus, Incomplete Multi-View Clustering (IMVC) becomes a challenging problem.

To address the IMVC problem, various excellent approaches have been proposed in literature (Xu, Tao, and Xu 2015; Yin, Wu, and Wang 2015; Tran et al. 2017; Wen et al. 2018, 2019; Zhou, Wang, and Yang 2019; Zhuge et al. 2019; Li, Wan, and He 2021; Liang et al. 2021; Wen et al. 2021a, 2022; Liu et al. 2022; Tang and Liu 2022; Jin et al. 2023; Wu et al. 2023b). The existing methods of incomplete multi-view clustering can be divided into three categories: matrix factorization-based, graph-based, and kernel-based. Matrix factorization-based IMVC methods generally transform the incomplete multi-view data into a unified representation (Shao et al. 2016; Rai et al. 2016; Zhao, Liu, and Fu 2016; Zhao, Ding, and Fu 2017; Xu et al. 2018; Wang et al. 2018; Wen et al. 2021b). The work in (Li, Jiang, and Zhou 2014) first proposes partial multi-view clustering (PVC), which attempts to find the consensus latent representation for all views based on the assumption that samples can be represented similarly across different views. With the development of graph learning, many researchers address the IMVC problems via graph-based information (Zhuang et al. 2012; Li et al. 2015; Wang et al. 2019; Wen et al. 2020; Gao et al. 2020; Li, Wan, and He 2021). This category fuses a consensus graph from multiple views in different perspectives. Wen et al. propose an approach that involves fusing view-specific graphs with adaptive weights to learn a consensus graph (Wen et al. 2020). The third category of IMVC methods is based on kernel learning, which maps all views into kernel space and then imputes incomplete kernels for clustering(Trivedi et al. 2010; Shao, He, and Yu 2015; Guo and Ye 2019; Liu et al. 2020). Liu et al. propose a framework that unifies imputation and clustering.(Liu et al. 2019).

These approaches have achieved great success in IMVC. However, we observe that the consistency information of incomplete views is not adequately exploited. Furthermore, the last category, kernel-based, explodes cluster information over the kernel matrix, while global structures in kernels are not fully exploited. To handle these problems, in this paper, we propose a novel Low-Rank Kernel Tensor Learning for Incomplete Multiple Views Clustering(LRKT-IMVC), which unifies imputation and clustering into one optimization process and incorporates low-rank kernel tensor constraint to capture more consistency information for clus-

*Corresponding authors.

tering. The overall framework of our approach is shown in Fig. 1. First, we introduce an innovative definition for the kernel tensor and establish a formal framework for addressing incomplete multiple kernel clustering using kernel tensor nuclear norm theory. In particular, this represents a pioneering effort within our research community. Instead of considering only the local structure of the kernel matrix as in previous kernel-based methods, we employ circulant algebra for the kernel tensor to explore the high-order correlations underlying multi-view data. Subsequently, we formulate a four-step alternative optimization algorithm with proven convergence properties to effectively address the resulting optimization problem. We then conduct comprehensive experiments on various benchmark datasets to investigate the properties of the proposed LRKT-IMVC. The comprehensive experimental results demonstrate the effectiveness of our proposed algorithm compared to other IMVC competitors. The main contributions of this paper can be summarized as follows:

1) We introduce an innovative concept: the kernel tensor, and develop a versatile IMVC method with the low-rank tensor constraint to impute the missing views. By capturing the low kernel tensor subspace, we address the problem of arbitrary view incompleteness for the first time. LRKT-IMVC, a pioneering work, merges multiple kernel learning with tensor nuclear norm theory in a unified framework for the IMVC problem.

2) This paper presents, for the first time, a novel approach for the comprehensive imputation of absent kernel elements. It explores consistency information among all views through the use of kernel tensor norms. Then, the incomplete kernel matrix is completed by the obtained complementary information.

3) We utilize the auxiliary tensor to address the tough optimization problem. Then, we design a four-step alternative optimization algorithm to effectively and efficiently solve LRKT-IMVC, and discuss its convergence, computational complexity, and potential extensions. Comprehensive experimental results clearly demonstrate the effectiveness and efficiency of our proposed method.

Proposed Method

Preliminary

In this section, we briefly review Multiple Kernel k-means with Incomplete Kernels (MKKM-IK) and Tensor Nuclear Norm Theory.

A. Multiple Kernel k-means with Incomplete Kernels

Given a collection $\{\mathbf{x}_i\}_{i=1}^n \subseteq \mathcal{X}$ with d dimension and n samples, and the pth feature mapping $\phi_p(\cdot) : \mathbf{x} \in \mathcal{X} \mapsto \mathcal{H}_p$, which maps $\mathbf{x}$ into a reproducing kernel Hilbert space $\mathcal{H}_p$ $(1 \leq p \leq m)$. Here, $\phi_{\boldsymbol{\beta}}(\mathbf{x}) = [\beta_1\phi_1(\mathbf{x})^\top, \ldots, \beta_m\phi_m(\mathbf{x})^\top]^\top$ represents the sample $\mathbf{x}$ in the multiple kernel setting, where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_m]^\top$ consists of the coefficients of the m base kernels $\{\kappa_p(\cdot,\cdot)\}_{p=1}^m$, these coefficients are optimized during learning. Based on the above definition, a kernel function can be written as $\kappa_p(\mathbf{x}_i, \mathbf{x}_j) = \phi_{\boldsymbol{\beta}}(\mathbf{x}_i)^\top \phi_{\boldsymbol{\beta}}(\mathbf{x}_j) = \sum_{p=1}^m \beta_p^2 \kappa_p(\mathbf{x}_i, \mathbf{x}_j)$, A kernel matrix $\mathbf{K}_p$ can be obtained by applying the kernel function $\kappa_p(\cdot,\cdot)$ to the sample $\{\mathbf{x}_i\}_{i=1}^n$, Then the fused kernel matrix $\mathbf{K}_{\boldsymbol{\beta}}$ can be expressed as $\mathbf{K}_{\boldsymbol{\beta}} = \sum_{p=1}^m \beta_p^2 \mathbf{K}_p$.

Multiple Kernel K-means with Incomplete Kernels (MKKM-IK) (Liu et al. 2019) combines multiple kernel clustering with incomplete kernels within a unified framework, and it updates each of them alternately. First, it imputes the incomplete kernel based on the clustering results, and then updates the clustering using the imputed kernels. The objective can be defined as

$$\begin{aligned}
&\min_{\{\mathbf{H},\beta,\ \{\mathbf{K}_p\}_{p=1}^m\}} \mathbf{Tr}(\mathbf{K}_{\boldsymbol{\beta}}(\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)),\\
&s.t.\ \mathbf{H} \in \mathbb{R}^{n\times k},\ \mathbf{H}^\top\mathbf{H} = \mathbf{I}_k,\\
&\quad \beta^\top \mathbf{1}_m = 1,\ \beta_p \geq 0,\\
&\quad \mathbf{K}_p(s_p, s_p) = \mathbf{K}_p^{(cc)},\ \mathbf{K}_p \succeq 0,\ \forall p,
\end{aligned} \tag{1}$$

where $s_p (1 \leq p \leq m)$ represents the sample index for which the sample is present in pth kernel and $\mathbf{K}_p^{(cc)}$ denotes the sub-matrix composed of these present samples. $\mathbf{K}_p(s_p, s_p) = \mathbf{K}_p^{(cc)}$ is denoted to guarantee that $\mathbf{K}_p$ maintains the known entries during the process. The optimization process of MKKM-IK additionally takes into account the imputation of incomplete kernels. The optimized details can refer to (Liu et al. 2019).

B. Tensor Nuclear Norm Theory

Let $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ be a tensor, we denote the block circulant matrix $bcirc(\mathcal{X}) \in \mathbb{R}^{n_1n_3\times n_2n_3}$ as

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

Then we define

$$unfold(\mathcal{X}) = \begin{pmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(n_3)} \end{pmatrix},\ fold(unfold(\mathcal{X})) = \mathcal{X}, \tag{3}$$

where the operator $unfold$ maps $\mathcal{X}$ to a matrix whose size is $n_1n_3 \times n_2$ and its inverse operator is $fold$. Based on the above definition, the t-$product$ of two tensors $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2\times n_4\times n_3}$ is concisely defined as

$$\mathcal{S} = \mathcal{A} * \mathcal{B} = fold(bcirc(\mathcal{A})unfold(\mathcal{B})), \tag{4}$$

where $\mathcal{S} \in \mathbb{R}^{n_1\times n_4\times n_3}$, The t-$product$ is analogous to matrix multiplication in the Fourier domain. Then t-SVD is simply written as

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top, \tag{5}$$

where the two tensors $\mathcal{U} \in \mathbb{R}^{n_1\times n_1\times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2\times n_2\times n_3}$ are orthogonal, and $\mathcal{S} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is a f-diagonal tensor. The t-SVD based tensor nuclear norm (t-TNN) is given as

$$\|\mathcal{X}\|_* := \sum_{i=1}^{min(n_1,n_2)} \sum_{k=1}^{n_3} |\mathcal{S}_f(i,i,k)|, \tag{6}$$

where $\mathcal{S}_f = fft(\mathcal{S}, [\,], 3)$ denotes the Fourier transform along the third dimension.

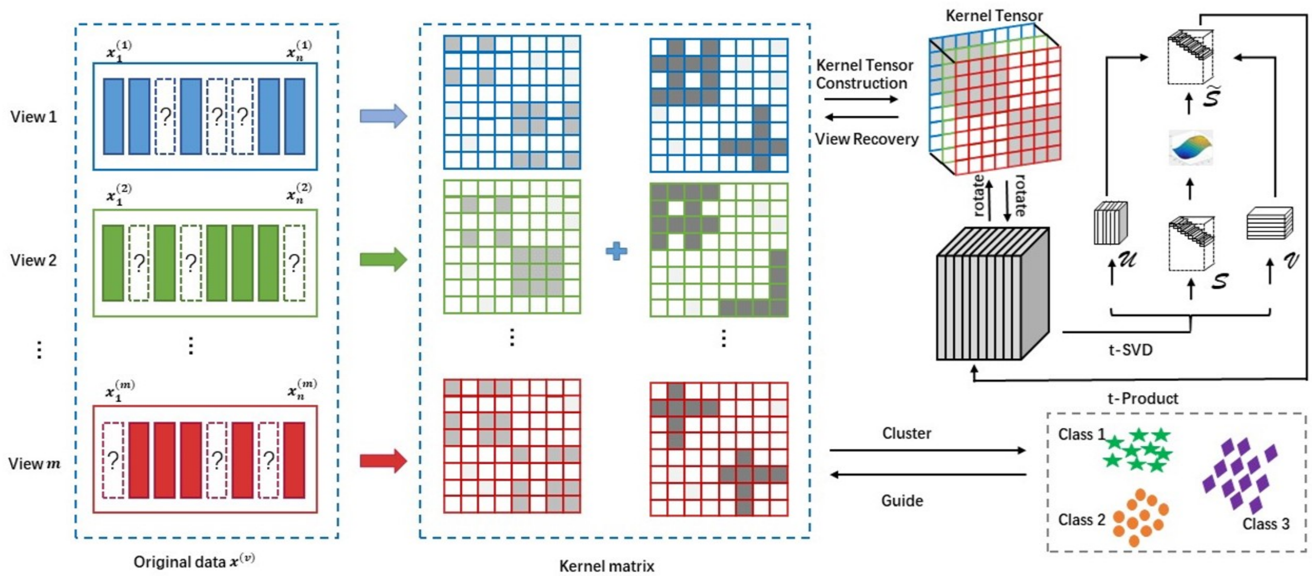


Figure 1: The general framework of LRKT-IMVC. From left to right, incomplete multi-view data are mapped into Hilbert space, forming a kernel tensor via stacked kernel matrices. Then, rotates this kernel tensor uses t-SVD to capture high-order correlations, aims to impute incomplete kernel matrices, and obtains clustering results. At the same time, the clustering result informs and refines the imputation process.

Formulation

Previous clustering approaches based on kernel learning have been widely used to address the problem of incomplete multi-view clustering. It performs clustering and imputation in a unified framework to further improve the clustering property, which has shown its feasibility and availability in various fields. While achieving excellent clustering results, we have identified two drawbacks associated with it: (1) These algorithms do not fully consider that the imputation of each kernel could be performed by another kernel matrix, even if they are missing. (2) It dose not consider the high-order correlations of all kernel matrices, while the global structures in kernels are not fully exploited, resulting in poor imputation and clustering performance.

To overcome the above disadvantages, inspired by tensor nuclear norm theory, we propose a creative clustering algorithm that designs the kernel tensor and then takes advantage of high-order correlations to guide the imputation of each kernel. The definition of the kernel tensor is as follows.

Definition 1. (*Kernel Tensor*) Given a set of kernel matrices $\{\mathbf{K}_p\}_{p=1}^m$, the kernel tensor $\mathcal{K}$ is constructed by stacking the kernel matrices of different views, that is,

$$\mathcal{K} = \Phi(\mathbf{K}_1, \mathbf{K}_2, \cdots, \mathbf{K}_m), \tag{7}$$

where $\Phi(\cdot)$ is a map function that merges kernel matrix $\mathbf{K}_p$ into a 3-mode tensor of size $n \times n \times m$. The property of the kernel tensor is analogous to the tensor, except that each frontal slice of the kernel tensor is a kernel matrix rather than a common matrix.

Based on the above definition of kernel tensor, the knowledge of the tensor can be significantly extended to the kernel tensor, and the consistency information between different views can be effectively captured. Then we propose a unique enhanced incomplete view recovery approach called Low-Rank Kernel Tensor Learning for Incomplete Multiple Views Clustering (LRKT-IMVC). Our method constructs a kernel tensor by stacking kernel matrices obtained from different views, and then rotates the kernel tensor to the size of $n \times m \times n$. We then use t-SVD to capture the high-order correlations underlying multi-view data to recover the missing kernel elements. Incomplete view imputation is achieved by using consistency information. According to the above analysis, we conclude that the incomplete imputation of the proposed LRKT-IMVC naturally introduces a constraint on the kernel tensor nuclear norm similar to the tensor nuclear norm and fully utilizes the information of other incomplete kernels, which leads to significantly improved clustering performance.

In addition, using the kernel tensor constraint not only improves the quality of the imputed incomplete kernels, but also solves the problem with the kernel matrix, which only computes the relationship between two samples and ignores the global structure. LRKT-IMVC constructs the kernel matrices of different views as a kernel tensor $\mathcal{K}$ and tries to use the knowledge of the tensor to explore more information that can better support the performance of clustering. This algorithm captures the high-order correlations of multiple kernels and ensures consensus among multiple kernels from a global perspective, which uses consistency among different kernels to impute missing kernel elements and further improve the quality of clustering.

LRKT-IMVC forms a new example for imputation of incomplete views. Besides, it also adopts clustering and imputation alternate optimization to further achieve great clustering performance. The objective formulation of our proposed LRKT-IMVC is as follows,

$$\begin{aligned}
&\min_{\{\mathbf{H},\boldsymbol{\beta},\{\mathbf{K}_p\}_{p=1}^m\}} \mathbf{Tr}(\mathbf{K}_{\boldsymbol{\beta}}(\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)) + \lambda \|\mathcal{K}\|_*, \\
&s.t.\ \mathbf{H} \in \mathbb{R}^{n\times k},\ \mathbf{H}^\top\mathbf{H} = \mathbf{I}_k,\ \boldsymbol{\beta}^\top\mathbf{1}_m = 1, \\
&\quad \beta_p \geq 0, \mathbf{K}_p(\mathbf{s}_p,\mathbf{s}_p) = \mathbf{K}_p^{(cc)},\ K_p \succeq 0,\ \forall p, \\
&\quad \mathbf{K}_\beta = \sum_{p=1}^m \beta_p^2\mathbf{K}_p,\ \mathcal{K} = \Phi(\mathbf{K}_1,\mathbf{K}_2,\cdots,\mathbf{K}_m),
\end{aligned} \tag{8}$$

where λ is a hyperparameter and $\|\mathcal{K}\|_*$ denotes the kernel tensor nuclear norm, which is similar to Eq. (6). Introduction of the low-rank kernel tensor constraints $\|\mathcal{K}\|_*$ is advantageous to capture the high-order correlations between different kernels.

Optimization

Inspired by the alternating direction method of multipliers (ADMM)(Lin, Liu, and Su 2011), we introduce the auxiliary tensor variable $\mathcal{G}$, the Eq. (8) can be transformed into an unconstrained problem.

$$\begin{aligned}
&\mathcal{L}(\mathbf{H};\mathbf{K}_1,\mathbf{K}_2,\ldots,\mathbf{K}_m;\mathcal{G};\boldsymbol{\beta}) \\
&= \mathbf{Tr}(\sum_{p=1}^m \beta_p^2\mathbf{K}_p(\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)) + \lambda\|\mathcal{G}\|_* \\
&+ \langle \mathcal{W}, \mathcal{K}-\mathcal{G}\rangle + \frac{\rho}{2}\|\mathcal{K}-\mathcal{G}\|_F^2,
\end{aligned} \tag{9}$$

where the tensor $\mathcal{W}$ denotes Lagrange multipliers and ρ is actually the penalty parameter. It seems difficult to jointly optimize $\mathbf{H}$, $\boldsymbol{\beta}$, $\mathcal{G}$ and $\{\mathbf{K}_p\}_{p=1}^m$, not only because of the tensor clear norm of $\mathcal{G}$ but also because the tensor $\mathcal{K}$ is affected on all kernels. We alternately design a simple and efficient algorithm to solve this problem.

- **Update $\mathbf{H}$**: Fixing $\{\mathbf{K}_p\}_{p=1}^m$, $\mathcal{G}$ and $\boldsymbol{\beta}$, the optimization problem of Eq. (9) with respect to $\mathbf{H}$ becomes

$$\begin{aligned}
&\min_{\mathbf{H}} \mathbf{Tr}(\mathbf{K}_{\boldsymbol{\beta}}(\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)), \\
&s.t.\ \mathbf{H}\in\mathbb{R}^{n\times k},\ \mathbf{H}^\top\mathbf{H} = \mathbf{I}_k.
\end{aligned} \tag{10}$$

  The Eq. (10) about optimized $\mathbf{H}$ is a traditional k-means problem that can be solved by existing packages.

- **Update $\mathbf{K}_p$**: Fixing $\mathbf{H}$, $\mathcal{G}$ and $\boldsymbol{\beta}$, the optimization problem of Eq. (9) with respect to $\mathbf{K}_p$ becomes

$$\begin{aligned}
&\min_{\mathbf{K}_p} \mathbf{Tr}(\beta_p^2\mathbf{K}_p(\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)) + \langle \mathbf{W}_p, \mathbf{K}_p - \mathbf{G}_p\rangle \\
&+ \frac{\rho}{2}\|\mathbf{K}_p - \mathbf{G}_p\|_F^2.
\end{aligned} \tag{11}$$

  To solve this subproblem, just setting the derivative of Eq. (11) to zero, $\mathbf{K}_p$ can be obtained by

$$\mathbf{K}_p = \frac{\rho\mathbf{G}_p - \mathbf{W}_p - \beta_p^2(\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)}{\rho}. \tag{12}$$

- **Update $\mathcal{G}$**: Fixing $\mathbf{H}$, $\{\mathbf{K}_p\}_{p=1}^m$ and $\boldsymbol{\beta}$, the optimization in Eq. (9) w.r.t $\mathcal{G}$ is equivalent to the following subproblem:

$$\mathcal{G}^* = \arg\min_{\mathcal{G}} \lambda\|\mathcal{G}\|_* + \frac{\rho}{2}\left\|\mathcal{G} - (\mathcal{K} + \frac{1}{\rho}\mathcal{W})\right\|_F^2. \tag{13}$$

  We can solve Eq. (13) though the follow theorem.

  Theorem 1(Xie et al. 2018): Let $\tau > 0$, and $\mathcal{G} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $\mathcal{F} \in \mathbb{R}^{n_1\times n_2\times n_3}$, the globally optimal solution of the following problem

$$\min_{\mathcal{G}} \tau\|\mathcal{G}\|_* + \frac{1}{2}\|\mathcal{G}-\mathcal{F}\|_F^2, \tag{14}$$

  is given by the tensor tubal-shrinkage operator

$$\mathcal{G} = \mathcal{C}_{n3\tau}(\mathcal{F}) = \mathcal{U} * \mathcal{C}_{n3\tau}(\mathcal{S}) * \mathcal{V}^\top, \tag{15}$$

  noticed that $\mathcal{F} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$ and $\mathcal{C}_{n3\tau} = \mathcal{S} * \mathcal{J}$, where $\mathcal{J} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is a f-diagonal tensor whose diagonal element in the Fourier domain is $\mathcal{J}_f(i,i,j) = (1 - \frac{n3\tau}{\mathcal{S}_f^{(j)}}(i,i))_+$. Furthermore, the Lagrange multiplier $\mathcal{W}$ need to be updated as follows:

$$\mathcal{W}^* = \mathcal{W} + \rho(\mathcal{K} - \mathcal{G}). \tag{16}$$

- **Update $\boldsymbol{\beta}$**: Fixing $\mathbf{H}$, $\{\mathbf{K}_p\}_{p=1}^m$ and $\mathcal{G}$, the optimization problem of Eq. (9) with respect to $\boldsymbol{\beta}$ becomes

$$\begin{aligned}
&\min_{\beta} \sum_{p=1}^m \beta_p^2\mathbf{Tr}(\mathbf{K}_p(\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)), \\
&s.t.\ \boldsymbol{\beta}^\top\mathbf{1}_m = 1,\ \beta_p \geq 0,\ \forall p.
\end{aligned} \tag{17}$$

The variable $\boldsymbol{\beta}$ can be efficiently optimized via solving quadratic programming with linear constraints.

In summary, our four-step algorithm for solving Eq. (9) is described in Algorithm 1, where $obj^{(t)}$ denotes the objective value at the t-th iteration.

Algorithm 1: The proposed Method

Input: $\{\mathbf{K}_p^{(cc)}\}_{p=1}^m$, $\{\mathbf{s}_p\}_{p=1}^m$, λ and ϵ_0;
Initialize: Initialize $\boldsymbol{\beta}^{(0)} = \mathbf{1}_m/m$, $\mathbf{K}_p^{(0)}$ and $t=1, \mathcal{G} = \mathcal{W} = 0$;
1: **while** not converge **do**
2: $\quad\mathbf{K}_{\boldsymbol{\beta}}^{(t)} = \sum_{p=1}^m (\beta_p^{(t-1)})^2\mathbf{K}_p^{(t-1)}$;
3: $\quad$Update $\mathbf{H}^t$ by solving a kernel k-means clustering optimization problem Eq. (10)
4: $\quad$Update $\mathbf{K}_p^{(t)}$ by Eq. (11);
5: $\quad$Update $\mathcal{G}^{(t)}$ via subproblem Eq. (13) ;
6: $\quad$Update $\mathcal{W}^{(t)}$ by Eq. (16);
7: $\quad$Update $\boldsymbol{\beta}^{(t)}$ by Eq. (17);
8: $\quad$Check the convergence conditions: $obj^{(t)} - obj^{(t-1)} \leq \epsilon_0$;
9: **end while**
10: Output $\mathbf{H}$, perform k-means on $\mathbf{H}$ to get final clustering result.

Discussion and Extension

Convergence. In this paper, LRKT-IMVC is theoretically guaranteed to converge to a local optimum. The objective value in Eq. (8) is monotonically decreasing at each iteration when one variable is optimized with the others fixed, and thus it is proved that the whole optimization algorithm converges to a local optimum. To demonstrate this point in practice, we plot the object value curves of our approach, i.e., the number of iterations with missing ratios on partial datasets. As recognized by our experimental results in

Fig. 2, the objective value of LRKT-IMVC decreases monotonically with each iteration and usually converges quickly. In fact, our method achieves convergence in less than 10 iterations under most conditions.

Computational Complexity. According to the optimization procedure in Algorithm 1, the computational complexity of our algorithm at each iteration is $\mathcal{O}(n^3 + 2n^2 m \log(mn) + m^3)$, where n, m and k represent the number of samples, views, and clusters, respectively. Note that $\mathbf{K}_p$ can be calculated in parallel by using auxiliary tensors. In this way, our method should scale well to other kernel methods.

Extension. Our proposed algorithm adopts the low-rank tensor constraint to capture the high-order correlations of incomplete kernels. Therefore, various similarity methods can be used to extend this work and multiple kernel clustering to further improve the interpolation effect. Second, our proposed LRKT-IMVC imputes the missing kernel elements from the global structure so that they can be interpolated more perfectly and compensate for the shortcomings of the kernel-based methods for IMVC.

Experiments

Experimental Settings

We perform experiments on seven widely used multiview benchmark datasets shown in Table 1, including **HW2sources**, **UCI_Digits**, **HW**, **Caltech101-20**, **BDGP_fea**, **CCV** and **Hdigit**, respectively. For each of these datasets, we generate the kernel matrix by applying the well-known Gaussian kernel. The sample sizes in the datasets range from 2000 to ten thousand. In addition, the number of kernels and clusters also show large variation, which guarantees the experiment to better evaluate the property of different clustering algorithms.

Various commonly used imputation methods are compared with our approach, including zero filling (ZF), mean filling (MF), k-nearest-neighbor filling (KNN), and Laplacian filling (LF). The widely used MKKM is applied with these imputed basic kernels. These two-stage methods are termed MKKM+ZF, MKKM+MF, MKKM+KNN and MKKM+LF, respectively. Furthermore, we also seriously compared the method of MKKM-IK (Liu et al. 2019) with different initialization for comprehensive comparison, including MKKM-IK+ZF, MKKM-IK+MF, MKKM-IK+KNN and MKKM-IK-MKC. Except for the comparison with the kernel-based method, we briefly compared various approaches of BSV (Zhao, Liu, and Fu 2016), Concat (Zhao, Liu, and Fu 2016), DAIMC (Hu and Chen 2019a) and OPIMC (Hu and Chen 2019b). For fair comparision, we directly utilized the source codes provided by the corresponding literature.

In this paper, we assume that at least one view is available for a sample. Inspired by the approach in (Zhu et al. 2018; Liu et al. 2018, 2019, 2020), we design the index vectors $\{\mathbf{s}_p\}_{p=1}^{m}$ that list the present samples of pth view to generate incomplete kernels. The missing ratio, defined as ϵ, represents the percentage of samples with incomplete views. Note that the performance of clustering is relatively affected by the parameter ϵ. In this experiment, we compare these algorithms to show the above point. Additionally, ϵ is set to $[0.1 : 0.1 : 0.9]$ on all datasets.

Datasets	Samples	Views	Classes
HW2sources	2000	2	10
UCI_Digits	2000	6	10
HW	2000	6	10
Caltech101-20	2386	6	20
BDGP_fea	2500	2	5
CCV	6773	3	20
Hdigit	10000	2	10

Table 1: Datasets used in our experiments.

For all datasets, the true number of clusters is known in advance and is taken as the true number of classes. We use three widely used criteria to evaluate the clustering performance, namely clustering accuracy (ACC), normalized mutual information (NMI) and purity. To reduce the effect of randomness in k-means, we repeat all experiments 50 times with random initialization and record the best result. Meanwhile, we randomly generate the "incomplete" patterns 10 times and report the statistical results.

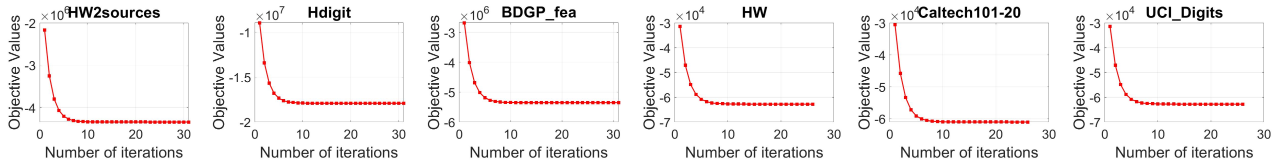
Experimental Results

Overall Clustering Performances. We summarize the results across all missing ratios and report their averages, with the best results marked in Table 2. This table reports the ACC, NMI, and purity of the above algorithms. Meanwhile, Fig. 3 vividly shows the ACC and NMI comparison with the variation of missing ratios on partial data sets. The others are shown in the appendix due to space limitations. We have the following observations.

i) LRKT-IMVC shows the best performance in terms of three metrics against all compared algorithms in most circumstances. We observe that the proposed algorithm significantly outperforms MKKM-IK-ZF, MKKM-IK-MF, MKKM-IK-KNN. For example, consider the results on BDGP_fea. It improves the state-of-the-art by 12.2%, 7.8%, 5.8%, 6.3%, 8.5%, 9.7%, 9.4%, 12.0% and 13.0% respectively, in terms of ACC with the variation of missing ratios in $[0.1 : 0.1 : 0.9]$, indicating the effectiveness of utilizing consistency and complementary information from other kernels to impute the missing kernel elements.

ii) Our algorithm achieves better clustering performance than MKKM-ZF, MKKM-MF, MKKM-LF and MKKM-KNN. We can see the details from Table 2, for instance, it exceeds the best of them by 10.1%, 22.1%, 22.1%, 5.1%, 16.8%, 3.2% and 3.9% in terms of NMI on all benchmark datasets. The gaps for the other criteria are similar. These results well demonstrate the effectiveness of using the tensor constraint in the kernel-based method for IMVC and unifying the imputation of incomplete views and the clustering task into a single optimization procedure.

iii) MKKM-IK-MKC (Liu et al. 2019) (in yellow) is a representative proposed method and demonstrates better performance in various circumstances. Yet its afore-



(a) HW2sources (b) Hdigit (c) BDGP_fea (d) HW (e) Caltech101-20 (f) UCI_Digits

Figure 2: The objective values of LRKT-IMVC with missing ratio 0.1. The curves with other missing ratios and on other datasets are similar and we omit them due to space limitations.

Dataset	Metric (%)	*ZF	*MF	*KNN	*LF	*IK-ZF	*IK-MF	*IK-KNN	*IK-MKC	Ours
HW2sources	ACC	18.0(2.3)	18.0(2.2)	18.0(2.3)	16.9(2.3)	16.6(2.6)	16.5(2.6)	16.6(2.6)	17.9(4.7)	**26.2(8.5)**
	NMI	8.0(2.5)	7.8(2.9)	8.0(2.6)	6.2(3.0)	5.8(3.2)	5.7(3.3)	5.8(3.2)	10.7(8.1)	**18.1(10.2)**
	PUR	18.8(2.4)	18.8(2.4)	18.8(2.4)	17.7(2.6)	17.4(2.9)	17.3(2.9)	17.4(2.9)	19.5(5.7)	**27.1(8.6)**
UCI_Digits	ACC	55.0(8.9)	55.1(8.2)	56.6(9.0)	55.0(8.8)	55.0(8.8)	55.1(8.3)	55.3(9.3)	54.9(8.9)	**77.4(1.8)**
	NMI	46.3(8.8)	47.0(8.4)	43.4(9.1)	46.4(8.8)	46.4(8.7)	47.0(8.4)	45.0(8.9)	46.4(8.6)	**69.1(1.9)**
	PUR	57.0(7.8)	57.5(6.9)	58.0(7.9)	57.1(7.7)	57.1(7.7)	57.5(6.8)	56.9(8.2)	57.0(7.7)	**77.7(1.8)**
HW	ACC	54.8(8.9)	55.1(8.3)	56.6(9.2)	54.7(8.9)	54.7(9.1)	55.2(8.3)	55.1(9.6)	54.9(8.8)	**77.4(1.8)**
	NMI	46.1(8.7)	47.0(8.4)	43.2(9.4)	46.1(8.8)	46.1(8.9)	47.1(8.3)	45.0(9.1)	46.3(8.6)	**69.1(1.9)**
	PUR	55.4(8.7)	55.3(8.2)	56.8(8.9)	55.4(8.8)	55.4(8.9)	55.4(8.2)	55.6(9.3)	55.5(8.7)	**77.5(1.8)**
Caltech101-20	ACC	31.3(4.1)	30.2(4.3)	31.5(4.1)	32.6(4.0)	34.8(2.3)	34.2(2.6)	35.2(2.4)	32.7(4.2)	**35.9(4.2)**
	NMI	39.1(5.3)	38.8(4.9)	39.0(4.9)	41.4(5.0)	43.9(3.3)	44.1(3.0)	44.1(3.3)	41.6(5.1)	**46.5(3.7)**
	PUR	64.1(4.2)	63.6(4.1)	64.1(4.1)	66.2(4.1)	68.0(2.7)	67.9(2.6)	68.1(2.8)	66.3(4.4)	**70.3(3.4)**
BDGP_fea	ACC	32.9(2.1)	33.0(2.9)	35.4(2.7)	33.6(1.6)	34.1(1.7)	34.5(1.5)	33.7(1.7)	37.6(3.1)	**47.0(1.1)**
	NMI	8.8(1.7)	8.6(2.5)	12.0(2.2)	9.9(1.3)	10.0(1.3)	10.9(1.6)	9.4(1.4)	13.9(3.2)	**28.8(3.0)**
	PUR	34.4(2.1)	33.9(2.9)	37.2(2.7)	36.0(1.4)	36.4(1.6)	36.3(1.8)	35.8(1.4)	36.7(3.3)	**50.6(1.5)**
CCV	ACC	14.1(1.1)	13.5(1.1)	13.9(1.0)	14.7(0.8)	14.7(0.8)	14.7(0.7)	14.7(0.8)	16.3(1.4)	**17.0(1.7)**
	NMI	9.0(1.3)	9.0(1.4)	9.2(1.3)	9.3(1.0)	9.4(1.0)	9.6(0.9)	9.4(1.1)	11.5(1.6)	**12.5(1.7)**
	PUR	17.9(1.1)	17.7(1.2)	17.9(1.1)	18.4(0.9)	18.5(0.8)	18.6(0.7)	18.5(1.0)	19.9(1.5)	**20.8(1.8)**
Hdigit	ACC	15.5(4.3)	15.5(4.4)	15.4(4.3)	14.7(3.5)	14.4(3.4)	14.1(3.2)	14.4(3.4)	12.0(0.9)	**19.5(4.7)**
	NMI	4.7(5.2)	4.8(5.3)	4.7(5.2)	3.7(4.1)	3.2(3.7)	3.0(3.8)	3.2(3.7)	0.7(0.8)	**8.7(6.8)**
	PUR	16.0(4.7)	16.0(4.7)	15.9(4.7)	15.2(3.8)	14.9(3.7)	14.6(3.5)	14.9(3.7)	12.3(1.2)	**20.3(4.9)**

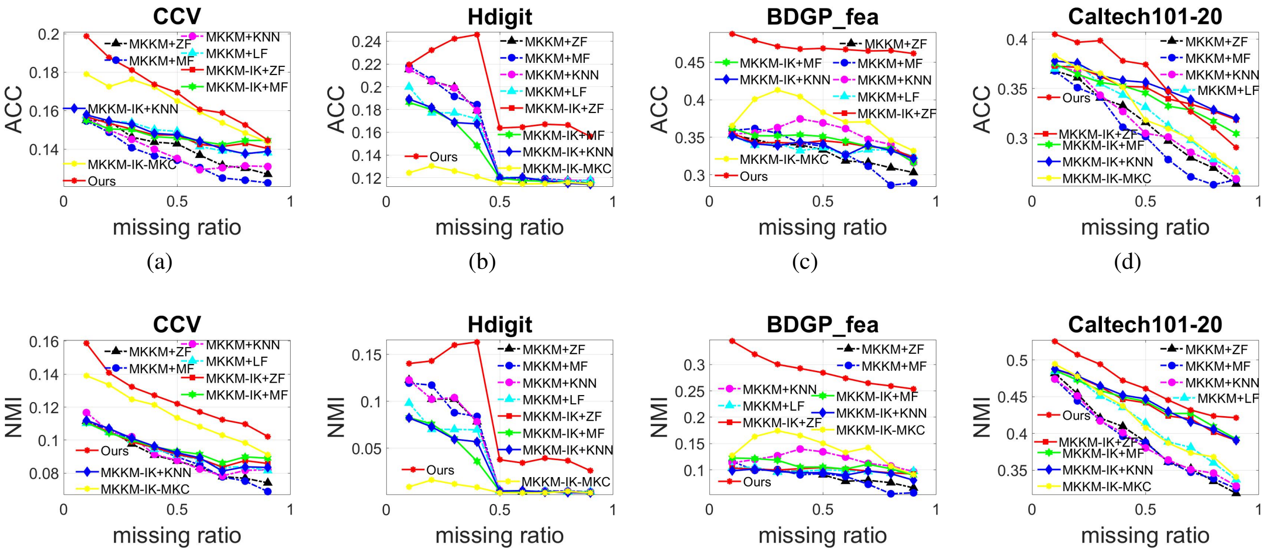
Table 2: Results (mean(std)) of our proposed method and other compared methods on seven datasets. '*' represents 'MKKM-'.

mentioned shortcomings lead to poor performance in many situations. Meanwhile, as we can see from Fig. 3, the proposed LRKT-IMVC dramatically improves MKKM-IK-MKC and achieves excellent clustering performance, it exceeds MKKM-IK-MKC by 9.5%, 10.2%, 11.6%, 12.5%, 4.8%, 5.0%, 5.2%, 5.0% and 4.2% of different missing ratios in Hdigit in terms of ACC. The improvement in other datesets is similar for different missing ratios. These results clearly demonstrate the superiority of capturing high-order correlations in incomplete multiple kernels.

iv) In addition, we make comparisons with various methods that do not rely on multiple kernel learning. As we can see from Fig. 4, when the percentage of present views is high, it changes gently, while the other methods decrease dramatically as the missing ratio increases. From the Fig. 4, the performance of BSV, Concat, DAIMC, and OPIMC is poor, while our algorithm performs almost consistently, especially at high missing ratios, and does not get worse as the number of present views decreases. The result demonstrates the advantages of combining tensor constraint and kernel learning in incomplete multiple views.

In summary, LRKT-IMVC shows superior clustering performance on all datasets compared to other methods, especially in the high missing ratio and can efficiently handle large-scale data sets, validating the effectiveness of our method in the imputation of incomplete kernels.

Parameter Sensitivity. Similar to MKKM-IK-MKC, our method also introduces a parameter λ to balance the quality of clustering and kernel reconstruction. We set λ in a range of $\left[10^{-4}, 10^{1}\right]$ and conduct experiments to study the sensitivity and effect of the parameter on the clustering performance on all datasets. Besides, we set unusual λ in different missing ratio to get the best clustering result. Based on this setting, we surprisingly find that via adjusting the value of λ, we can achieve more excellent clustering performance. As we can see from Fig. 5, our approach obtains more splendid performance with the increase of missing ratio, which seems unreliable in existing kernel-based IMVC methods. This is the power of introducing the kernel tensor constraint in the imputation of incomplete kernels. As the missing ratio increases, the quality of clustering can be improved by changing the value of λ to increase the low-rank kernel ten-

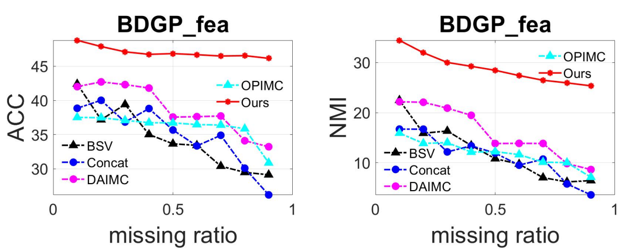


(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3: Comparison of ACC and NMI with variation of missing ratios on partial datasets. The curves with other missing ratios and on other datasets are similar and we omit them due to space limitations.

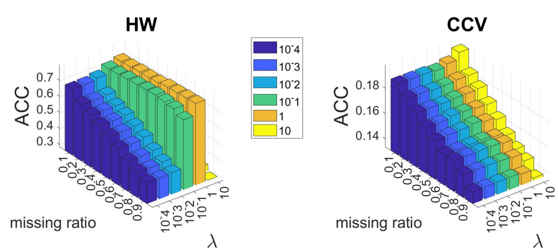
sor constraint. The kernel tensor $\mathcal{K}$ sufficiently excavates the high-order correlations underlying multi-view data and captures more consistency and complementarity information to help impute missing kernel elements. Therefore, the performance of clustering improves as the number of current views decreases. In general, the experimental results show that our approach is somewhat sensitive to the parameter λ.



(a) ACC (b) NMI

Figure 4: ACC and NMI comparison with different missing ratio on BDGP_fea.

Quantitative Study. To further illustrate the effectiveness of LRKT-IMVC solution in missing multiple views especially in kernel-based method, we evaluate the curves of clustering performance on partial datasets with respect to various missing ratio in Fig. 3 and Fig. 4. As we can see, LRKT-IMVC naturally proves the success of unifying kernel tensor nuclear norm and kernel learning into one framework and further improving the quality of the cluster. In addition, the performance of all previous algorithms decreases with the increase of the missing ratio, but we miraculously find that LRKT-IMVC can achieve excellent clustering result even with the decrease of views via adjusting the appropriate hyperparameter. The result is to break our stereotypical thinking in incomplete multi-view and give us new interesting in solving IMVC problem.



(a) HW (b) CCV

Figure 5: The result of LRKT-IMVC with different missing ratio on HW and CCV influenced by parameter λ.

Conclusion

In this paper, we propose a novel algorithm, which stacks the kernels of different views into a kernel tensor and employs the low-rank kernel tensor constraint in the incomplete view interpolation, which enjoys the incomplete kernel imputation and clustering simultaneously. In this way, LRKT-IMVC explores more consistency information among different kernels, which greatly improves the quality of imputation and then guides the result of clustering. To handle the resulting optimization problem, we utilize an auxiliary tensor and design a four-step alternative algorithm with guaranteed convergence. Extensive experiments on benchmark datasets demonstrate the effectiveness of our algorithm.

Acknowledgments

This work was supported by the Beijing Natural Science Foundation (No. 4242046), the Tangshan Municipal Science and Technology Plan Project (No.23130225E), and the National Key Research and Development Project (No. 2018AAA0100300).

References

Chen, M.-S.; Huang, L.; Wang, C.-D.; and Huang, D. 2020. Multi-view clustering in latent embedding space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 3513–3520.

Dong, Z.; Jin, J.; Xiao, Y.; Wang, S.; Zhu, X.; Liu, X.; and Zhu, E. 2023a. Iterative Deep Structural Graph Contrast Clustering for Multiview Raw Data. *IEEE Transactions on Neural Networks and Learning Systems*.

Dong, Z.; Wang, S.; Jin, J.; Liu, X.; and Zhu, E. 2023b. Cross-view Topology Based Consistent and Complementary Information for Deep Multi-view Clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 19440–19451.

Duan, J.; Wang, S.; Zhang, P.; Zhu, E.; Hu, J.; Jin, H.; Liu, Y.; and Dong, Z. 2023. Graph anomaly detection via multi-scale contrastive learning networks with augmented view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 7459–7467.

Gao, H.; Nie, F.; Li, X.; and Huang, H. 2015. Multi-view subspace clustering. In *Proceedings of the IEEE International Conference on Computer Vision*, 4238–4246.

Gao, Q.; Xia, W.; Wan, Z.; Xie, D.; and Zhang, P. 2020. Tensor-SVD based graph learning for multi-view subspace clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 3930–3937.

Guo, J.; and Ye, J. 2019. Anchors bring ease: An embarrassingly simple approach to partial multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 118–125.

Guo, Y. 2013. Convex subspace representation learning from multi-view data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 27, 387–393.

Hu, M.; and Chen, S. 2019a. Doubly aligned incomplete multi-view clustering. *ArXiv Preprint ArXiv:1903.02785*.

Hu, M.; and Chen, S. 2019b. One-pass incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 3838–3845.

Huang, S.; Tsang, I.; Xu, Z.; Lv, J.; and Liu, Q.-H. 2022. Multi-view clustering on topological manifold. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 6944–6951.

Huang, Z.; Ren, Y.; Pu, X.; and He, L. 2021. Non-linear fusion for self-paced multi-view clustering. In *Proceedings of the 29th ACM International Conference on Multimedia*, 3211–3219.

Jin, J.; Wang, S.; Dong, Z.; Liu, X.; and Zhu, E. 2023. Deep Incomplete Multi-view Clustering with Cross-view Partial Sample and Prototype Alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11600–11609.

Li, L.; Wan, Z.; and He, H. 2021. Incomplete multi-view clustering with joint partition and graph learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(1): 589–602.

Li, S.-Y.; Jiang, Y.; and Zhou, Z.-H. 2014. Partial multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28.

Li, Y.; Nie, F.; Huang, H.; and Huang, J. 2015. Large-scale multi-view spectral clustering via bipartite graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29.

Liang, N.; Yang, Z.; Li, L.; Li, Z.; and Xie, S. 2021. Incomplete multiview clustering with cross-view feature transformation. *IEEE Transactions on Artificial Intelligence*, 3(5): 749–762.

Lin, Z.; Liu, R.; and Su, Z. 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. *Advances in Neural Information Processing Systems*, 24.

Liu, W.; Yue, X.; Chen, Y.; and Denoeux, T. 2022. Trusted multi-view deep learning with opinion aggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7585–7593.

Liu, X.; Li, M.; Tang, C.; Xia, J.; Xiong, J.; Liu, L.; Kloft, M.; and Zhu, E. 2020. Efficient and effective regularized incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8): 2634–2646.

Liu, X.; Zhu, X.; Li, M.; Wang, L.; Tang, C.; Yin, J.; Shen, D.; Wang, H.; and Gao, W. 2018. Late fusion incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10): 2410–2423.

Liu, X.; Zhu, X.; Li, M.; Wang, L.; Zhu, E.; Liu, T.; Kloft, M.; Shen, D.; Yin, J.; and Gao, W. 2019. Multiple kernel k k-means with incomplete kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(5): 1191–1204.

Rai, N.; Negi, S.; Chaudhury, S.; and Deshmukh, O. 2016. Partial multi-view clustering using graph regularized NMF. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2192–2197. IEEE.

Shao, W.; He, L.; Lu, C.-t.; and Philip, S. Y. 2016. Online multi-view clustering with incomplete views. In *2016 IEEE International Conference on Big Data (Big Data)*, 1012–1017. IEEE.

Shao, W.; He, L.; and Yu, P. S. 2015. Multiple incomplete views clustering via weighted nonnegative matrix factorization with regularization. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part I*, 318–334. Springer.

Tang, H.; and Liu, Y. 2022. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International Conference on Machine Learning*, 21090–21110. PMLR.

Tao, Z.; Liu, H.; and Fu, Y. 2017. Simultaneous clustering and ensemble. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Tran, L.; Liu, X.; Zhou, J.; and Jin, R. 2017. Missing modalities imputation via cascaded residual autoencoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1405–1414.

Trivedi, A.; Rai, P.; Daumé III, H.; and DuVall, S. L. 2010. Multiview clustering with incomplete views. In *NIPS Workshop*, volume 224, 1–8. Citeseer.

Wang, H.; Zong, L.; Liu, B.; Yang, Y.; and Zhou, W. 2019. Spectral perturbation meets incomplete multi-view data. *ArXiv Preprint ArXiv:1906.00098*.

Wang, Y.; Wu, L.; Lin, X.; and Gao, J. 2018. Multiview spectral clustering via structured low-rank matrix factorization. *IEEE Transactions on Neural Networks and Learning Systems*, 29(10): 4833–4843.

Wen, J.; Yan, K.; Zhang, Z.; Xu, Y.; Wang, J.; Fei, L.; and Zhang, B. 2020. Adaptive graph completion based incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 23: 2493–2504.

Wen, J.; Zhang, Z.; Fei, L.; Zhang, B.; Xu, Y.; Zhang, Z.; and Li, J. 2022. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2): 1136–1149.

Wen, J.; Zhang, Z.; Xu, Y.; Zhang, B.; Fei, L.; and Liu, H. 2019. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 5393–5400.

Wen, J.; Zhang, Z.; Xu, Y.; Zhang, B.; Fei, L.; and Xie, G.-S. 2021a. Cdimc-net: Cognitive deep incomplete multi-view clustering network. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 3230–3236.

Wen, J.; Zhang, Z.; Xu, Y.; and Zhong, Z. 2018. Incomplete multi-view clustering via graph regularized matrix factorization. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 0–0.

Wen, J.; Zhang, Z.; Zhang, Z.; Zhu, L.; Fei, L.; Zhang, B.; and Xu, Y. 2021b. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 10273–10281.

Wu, Y.; Chi, Z.; Wang, Y.; and Feng, S. 2023a. Metagcd: Learning to continually learn in generalized category discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1655–1665.

Wu, Y.; Liang, T.; Feng, S.; Jin, Y.; Lyu, G.; Fei, H.; and Wang, Y. 2023b. MetaZSCIL: A meta-learning approach for generalized zero-shot class incremental learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 10408–10416.

Wu, Y.; Liu, H.; Feng, S.; Jin, Y.; Lyu, G.; and Wu, Z. 2021. GM-MLIC: graph matching based multi-label image classification. *ArXiv Preprint ArXiv:2104.14762*.

Xie, Y.; Tao, D.; Zhang, W.; Liu, Y.; Zhang, L.; and Qu, Y. 2018. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *International Journal of Computer Vision*, 126: 1157–1179.

Xu, C.; Tao, D.; and Xu, C. 2015. Multi-view learning with incomplete views. *IEEE Transactions on Image Processing*, 24(12): 5812–5825.

Xu, N.; Guo, Y.; Zheng, X.; Wang, Q.; and Luo, X. 2018. Partial multi-view subspace clustering. In *Proceedings of the 26th ACM International Conference on Multimedia*, 1794–1801.

Yin, Q.; Wu, S.; and Wang, L. 2015. Incomplete multi-view clustering via subspace learning. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 383–392.

Yu, S.; Tranchevent, L.; Liu, X.; Glanzel, W.; Suykens, J. A.; De Moor, B.; and Moreau, Y. 2011. Optimized data fusion for kernel k-means clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5): 1031–1039.

Zhang, T.; Liu, X.; Zhu, E.; Zhou, S.; and Dong, Z. 2022. Efficient Anchor Learning-based Multi-view Clustering–A Late Fusion Method. In *Proceedings of the 30th ACM International Conference on Multimedia*, 3685–3693.

Zhang, Z.; Liu, L.; Shen, F.; Shen, H. T.; and Shao, L. 2018. Binary multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7): 1774–1782.

Zhao, B.; Kwok, J. T.; and Zhang, C. 2009. Multiple kernel clustering. In *Proceedings of the 2009 SIAM International Conference on Data Mining*, 638–649. SIAM.

Zhao, H.; Ding, Z.; and Fu, Y. 2017. Multi-view clustering via deep matrix factorization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Zhao, H.; Liu, H.; and Fu, Y. 2016. Incomplete multi-modal visual data grouping. In *IJCAI*, 2392–2398.

Zhou, W.; Wang, H.; and Yang, Y. 2019. Consensus graph learning for incomplete multi-view clustering. In *Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14-17, 2019, Proceedings, Part I 23*, 529–540. Springer.

Zhu, X.; Liu, X.; Li, M.; Zhu, E.; Liu, L.; Cai, Z.; Yin, J.; and Gao, W. 2018. Localized incomplete multiple kernel k-means. In *IJCAI*, 3271–3277.

Zhuang, L.; Gao, H.; Lin, Z.; Ma, Y.; Zhang, X.; and Yu, N. 2012. Non-negative low rank and sparse graph for semi-supervised learning. In *2012 IEEE Cnference on Computer Vision and Pattern Recognition*, 2328–2335. IEEE.

Zhuge, W.; Hou, C.; Liu, X.; Tao, H.; and Yi, D. 2019. Simultaneous representation learning and clustering for incomplete multi-view Data. In *IJCAI*, volume 7, 4482–4488.