Distribution Consistency based Fast Anchor Imputation for Incomplete Multi-view Clustering

Xingfeng Li
Nanjing University of Science and Technology
Nanjing, China
lixingfeng@njust.edu.cn

Yinghui Sun*
Nanjing University of Science and Technology
Nanjing, China
yinghuisun@njust.edu.cn

Quansen Sun
Nanjing University of Science and Technology
Nanjing, China
sunquansen@njust.edu.cn

Jia Dai
Beijing Institute of Technology
Beijing, China
daijian1000@163.com

Zhenwen Ren*
Southwest University of Science and Technology
Mianyang, China
Wuzhou University
Wuzhou, China
rzw@njust.edu.cn

ABSTRACT

In practical scenarios, partial missing of multi-view data is very common, such as register information missing from social network analysis, which results in incomplete multi-view clustering (IMVC). How to fill missing data fast and efficiently plays a vital role in improving IMVC, carrying a significant challenge. Existing IMVC methods always use all observed data to fill in missing data, resulting in high complexity and poor imputation quality due to a lack of guidance from consistent distribution. To break the existing limitations, we propose a novel **Distribution Consistency based Fast Anchor Imputation for Incomplete Multi-view Clustering (DCFAI-IMVC)** method. Specifically, to eliminate the interference of redundant and fraudulent features in the original space, incomplete data are first projected into a consensus latent space, where we dynamically learn a small number of anchors to achieve fast and good imputation. Then, we employ global distribution information of the observed embedding representations to further ensure the consistent distribution between the learned anchors and the observed embedding representations. Ultimately, a tensor low-rank constraint is imposed on bipartite graphs to investigate the high-order correlations hidden in data. DCFAI-IMVC enjoys linear complexity in terms of sample number, which gives it great potential to handle large-scale IMVC tasks. By performing extensive experiments, our effectiveness, superiority, and efficiency are all validated on multiple public datasets with recent advances.

*Corresponding author.

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada

ACM ISBN 979-8-4007-0108-5/23/10…$15.00
https://doi.org/10.1145/3581783.3612483

CCS CONCEPTS

• **Theory of computation → Theory and algorithms for application domains**; **Unsupervised learning and clustering**.

KEYWORDS

Incomplete multi-view clustering, distribution consistency, anchor learning, missing data imputation

ACM Reference Format:
Xingfeng Li, Yinghui Sun, Quansen Sun, Jia Dai, and Zhenwen Ren. 2023. Distribution Consistency based Fast Anchor Imputation for Incomplete Multi-view Clustering. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29-November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3581783.3612483

1 INTRODUCTION

Multi-view clustering (MVC) has attracted intensive attention by optimally integrating heterogeneous and homogeneous properties to group unlabeled multi-view data into different clusters [3, 13, 44–46]. By assuming all views are complete, many MVC methods [1, 4, 8, 15, 16, 21, 32, 39, 41] have been presented due to their validity of capturing the paired similarity between samples and views. However, in many practical applications, some view information is unavailable, resulting in incomplete multi-view data. For example, in social network analysis, people may only register in some social networks, which leads to incompleteness of sample information among different social networks. Clustering on such kind of data can be deemed as incomplete multi-view clustering (IMVC). Missing data not simply results in the loss of view information, but also destroys paired similarity, making IMVC become an intractable problem[7, 17].

In recent years, a series of pioneer work has been developed to effectively address this challenge. According to the processing methods of missing data, two types of IMVC methods are mainly included: (1) they **only use the observed view data to perform clustering tasks by discarding missing view data**, so as to avoid the negative impact of missing data. (2) they **fill the**

missing data with all observable view data to boost the clustering performance, simultaneously recovering the missing view information to some extent. For the first category, some scholars [11, 27, 33, 35, 37] employ matrix factorization to synergistically generate a consensus representation across multiple incomplete views. By this way, observed data from different views can be aggregated into a single representation, making it possible to roughly describe the information of multiple incomplete views. Then, consensus representation is fed into k-means to obtain clustering results. Different from sharing consensus representation, another scholars employ different graph learning techniques to synergistically generate a consensus graph across multiple incomplete views [10, 19, 22, 31, 34]. For example, [10, 19, 31] use self-representation subspace learning or adaptive neighborhood graph learning to generate a shared similarity graph for spectral clustering. To this end, observed data of multiple incomplete views can be semantically fused into a consensus graph to roughly reflect intact graph structure of multiple incomplete views. The core idea of these IMVC methods is to discard missing view data, and only use the observed view data to perform clustering tasks. Whereas, this kind of methods reduce sample size to only exchange for formal completeness. In fact, directly discarding missing data can destroy the original structural relation hidden in the incomplete multi-view data [38, 42].

Second, different from ignoring the missing data, a spot of IMVC methods [18, 20, 22, 36, 38] attempt to recover the missing views via all available data, so as to address the sample missing problem. For sample, [36] leverages all observed representations to learn a consensus representation by a unified alignment framework based on matrix factory. Then, consensus representation is reversely used to fill the missing data. To this end, missing data imputation and consensus representation learning iteratively optimize in a mutually reinforcing manner until convergence. To further capture the nonlinear relationship in incomplete data, some scholars employ the predefined incomplete kernel matrices to learn a consensus kennel to simultaneously perform kernel imputation and IMVC. Different from learning a consensus representation or consensus kernel, [18, 19, 38] generate a consensus affinity graph to iteratively impute the missing samples by using familiar graph learning frameworks, such as self-representation learning. Since the graph learning can effectively characterize the similarity structure of sample pair, graph-based methods have achieved promising results in IMVC tasks. Although effective, the above filled IMVC methods select all observed multi-view samples to fill the missing samples, resulting in high computational and space complexity; more importantly, in large-scale tasks, the data in the original feature space usually suffers from the interference of redundant or redundant information, which inevitably reduces the filled quality.

To address the aforementioned issues, we propose a novel **Distribution Consistency based Fast Anchor Imputation for Incomplete Multi-view Clustering (DCFAI-IMVC)** to achieve fast and high-quality imputation by dynamically learning a small number of anchors. As mentioned in Fig. 1, we first project the multi-view data from the original feature space into a latent embedding space, where we assume that a shared view-consistent anchor matrix can be learned with the help of complementarity between different views. Meanwhile, the view-specific bipartite graphs are constructed, since the similarities between each shared anchor point and the embedding representations of different views should exist the diversity. Besides, we also share the cluster center between the observed embedding representations and the learned anchors, where the global distribution information among the observed embedding representations can be captured to synergistically guide a high-quality anchor matrix to effectively impute the missing view data. Finally, to further capture the pair-wise sample correlations and pair-wise view correlations, a tensor low-rank constraint is imposed on the learned bipartite graphs. After performing a well-designed iteration algorithm, an averaged bipartite graph is obtained to perform k-means clustering. Main contributions are as follows.

- Our DCFAI-IMVC proposes to a novel and flexible framework for missing data imputation and IMVC. As far as I know, it dynamically learns a small number of high-quality anchors from incomplete data for the first practice to perform fast and good imputation on large-scale clustering tasks.
- Unlike the existing IMVC methods that employs all the observed data to fill the missing data for clustering in the original feature space, DCFAI-IMVC only requires to learn a spot of anchors to perform fast and good imputation for improving clustering in a latent space. DCFAI-IMVC is a pioneering work to dynamically learn anchors from incomplete data.
- Comprehensive experimental results on eight benchmark datasets verify the advantages of our DCFAI-IMVC compared to existing IMVC competitors.

2 RELATED WORK

In this section, we introduce the existing work most related to our proposed method, including bipartite graph clustering and graph learning for IMVC.

2.1 Bipartite Graph Clustering

Bipartite graph is deemed as a very effective strategy to handle large-scale data by selecting a relative small proportion of representative anchors to establish connection with original samples [2, 14]. This idea of multi-view framework can be traditionally expressed as

$$\min_{\mathbf{Z}_p,\mathbf{Z}} \left\|\mathbf{Y}_p - \mathbf{L}_p\mathbf{Z}_p\right\|_{\mathbf{F}}^2 + \alpha\Omega\left(\mathbf{Z}_p,\mathbf{Z}\right) \text{ s.t. } \mathbf{Z}_p \geq 0, \mathbf{Z}_p^\top\mathbf{1} = \mathbf{1} \tag{1}$$

where $\mathbf{Y}_p \in \mathbb{R}^{d_i\times n}$ and $\mathbf{L}_p \in \mathbb{R}^{d_i\times m}$ represent complete data and its m selected or sampled landmarks corresponding to p-th view. $\Omega(\cdot,\cdot)$ represents certain kind of graph fusion techniques [12, 24]. Eq. (1) can reduce both computational and space complexity since the size of traditional $n\times n$ similarity graph is decreased to $m\times n$ bipartite graph $\mathbf{Z}$. Meanwhile, comparable clustering performance can be also achieved. Based on the high efficiency of this bipartite graph framework, we first try to fill missing views with anchors for further improving the clustering performance in the next section.

2.2 Graph Learning for IMVC

Recently, IMVC has attracted intensive attention since multimodal data collected from real applications tend to be inherently incomplete. For sample, a person cannot be registered on all social networking platforms, resulting in incomplete information about that

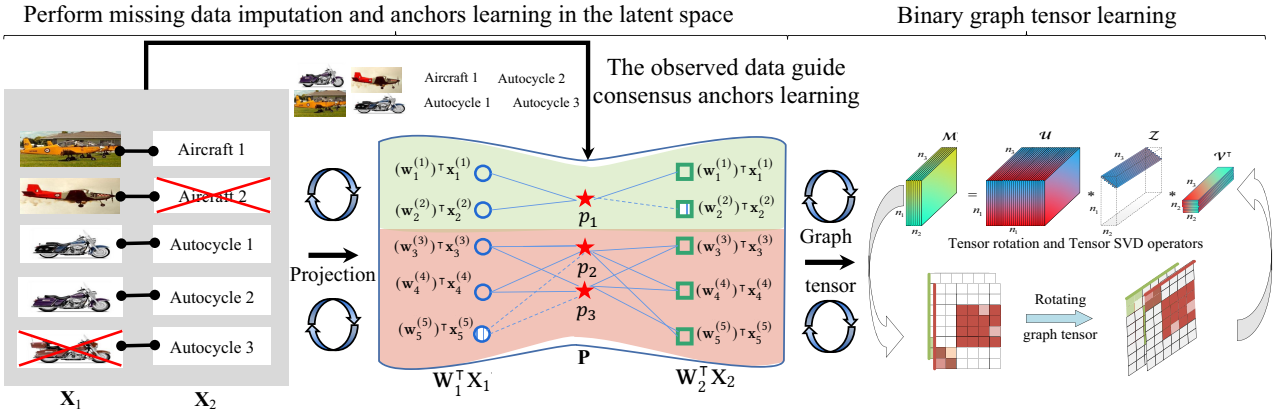


Figure 1: Overview of the proposed DCFAI-IMVC method. Two views are employed for ease of understanding. The red cross and the blue dashed line represent the missing sample and the line connecting the missing sample to the anchor point, respectively.

person on some platforms. Consequently, how to cluster with multi-view partial data becomes a challenging and valuable issue. Fortunately, graph learning has powerful representation ability and can capture the relevant information between data to perform IMVC [5, 6, 43]. For one hand, some graph-based IMVC methods discard missing view data to avoid the negative impact of missing data. Given incomplete multi-view data $\{\mathbf{X}_p \in \mathbb{R}^{d_p\times n}\}_{p=1}^V$ and the indicator matrices $\{\mathbf{G}_p \in \mathbb{R}^{n\times n_p}\}_{p=1}^V$, graph-based IMVC framework can be written as

$$\min_{\mathbf{S}_p} \left\|\mathbf{X}_p\mathbf{G}_p - \mathbf{X}_p\mathbf{G}_p\mathbf{S}_p\right\|_F^2 + \Phi\left(\mathbf{S}_p\right), \text{ s.t. } \mathbf{S}_p \geq 0, \mathbf{S}_p^\top\mathbf{1} = \mathbf{1}, \tag{2}$$

where n_p and $\mathbf{S}_p \in \mathbb{R}^{n_p\times n_p}$ are the number of observed data and subgraph in the p-th view, respectively. $\Phi(\cdot)$ denotes the regularization of $\mathbf{S}_p$. Indicator matrix $\mathbf{G}_p \in \{0,1\}$ can eliminate the missing samples in $\mathbf{X}_p$ to form $\mathbf{X}_p\mathbf{G}_p \in \mathbb{R}^{d_p\times n_p}$, which is all the observed samples of $\mathbf{X}_p$. Then, the complete graph can be sketched with $\mathbf{G}_p\mathbf{S}_p\mathbf{G}_p^\top$ to perform subsequent clustering tasks. By following this scheme, many methods [10, 11, 19, 31, 33] have only employed observed data to perform IMVC. Whereas, discarding missing data directly can ignore the hidden information of the missing views and destroy the original structure of data [38]. For another, different from ignoring the missing parts, a spot of IMVC methods recover the missing views via all observed view data to further explore the hidden information of the missing views. To this end, original data structure information can be inferred to boost the clustering performance. This kind of IMVC framework can be expressed as

$$\min_{\mathbf{S}} \Psi(\{\mathbf{X}_p\}_{p=1}^V, \mathbf{S}) + \Phi\left(\mathbf{S}\right) \tag{3}$$

where Ψ denotes different graph fusion techniques [33, 34]. In Eq. (3), consensus affinity graph $\mathbf{S}$ can be constructed from all the observed samples in $\{\mathbf{X}_p\}_{p=1}^V$. Then, $\mathbf{S}$ reversely fills missing parts of $\{\mathbf{X}_p\}_{p=1}^V$. Different from Eq. (3), recent proposed [18, 38] attempt to use the low-rank tensor constraint to explore high-order relationship of data to improve the quality of imputation. Although these IMVC methods [18, 20, 22, 36, 38] have made some achievements, they select all observed data to fill the missing data, which leads to high time and space complexities. Meanwhile, redundant features or fraudulent features in original feature space can cause the low-quality imputation.

3 FOMULATION

Given d_p dimensions and n samples of complete multi-view data $\{\mathbf{X}_p \in \mathbb{R}^{d_p\times n}\}_{p=1}^V$, we can define V incomplete multi-view data as

$$\mathbf{X}_p = [\mathbf{X}_p^o; \mathbf{X}_p^m], \mathbf{X}_p^o \in \mathbf{R}^{d_p\times n_p}, \mathbf{X}_p^m \in \mathbf{R}^{d_p\times(n-n_p)} \tag{4}$$

where $\mathbf{X}_p^o$ denote the fixed observed data, and $\mathbf{X}_p^m$ denote the missing data required to be imputed. $[\cdot;\cdot]$ is the horizontal concatenation operation, and n_p is the number of the observed samples. To simultaneously perform missing data imputation and IMVC, all the existing IMVC methods employ all the observed data to impute the missing data in the original space, which always leads to low-quality imputation owe to the interference of redundant and fraudulent features or low-quality samples. Meanwhile, redundant samples or features can also cause high time complexity and space complexity. Instead of filling missing data $\mathbf{X}_p^m$ with all observed data, we first attempt to merely learn a small number of anchors from incomplete data, which are used to perform fast and good imputation for $\mathbf{X}_p^m$. This idea can be mathematically expressed as

$$\begin{aligned}&\min_{\substack{\mathbf{X}_p^m,\mathbf{P},\\ \mathbf{W}_p,\mathbf{Z}_p,\boldsymbol{\theta}_p}} \sum_{p=1}^V \theta_p^2\|\mathbf{W}_p^\top\mathbf{X}_p - \mathbf{P}\mathbf{Z}_p\|_F^2 \\ &\text{s.t. } \mathbf{Z}_p \geq 0, \mathbf{Z}_p^\top\mathbf{1} = \mathbf{1}, \mathbf{W}_p^\top\mathbf{W}_p = \mathbf{I}_k, \mathbf{P}^\top\mathbf{P} = \mathbf{I}_m, \boldsymbol{\theta}^\top\mathbf{1} = 1, \boldsymbol{\theta} \geq 0\end{aligned} \tag{5}$$

where p-th projection matrix $\mathbf{W}_p \in \mathbb{R}^{d_p\times k}$ projects the original incomplete data $\mathbf{X}_p$ into a shared latent space. Then, a view-consistent anchor matrix $\mathbf{P}$ and p view-specific bipartite graphs $\mathbf{Z}_p$ can be learned from $\mathbf{X}_p$ corresponding to p views. Orthogonal constraint on $\mathbf{P} \in \mathbb{R}^{k\times m}$ can enhance its discrimination. The view-consistent anchor matrix $\mathbf{P}$ and view-specific bipartite graphs are constructed due to the complementarity and the diversity of different views, respectively. $\boldsymbol{\theta}$ is the weight vector. From Eq. (5), we aim to dynamically learn a view-consistent anchor matrix to essentially represent the whole incomplete multi-view data. However, the learned anchors from Eq. (5) would deviate far from the distribution of the

original real data due to the disturb of missing data, further degrading the subsequent clustering task. To ensure the validity of the learned anchor matrix, as mentioned in Fig. 2, we first use the

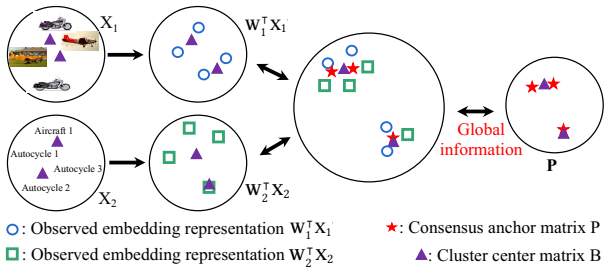


Figure 2: Consensus cluster center of the observed embedding representations guides the learning of anchors.

global distribution information of the observed embedded representations $\mathbf{W}_p^\top \mathbf{X}_p \mathbf{G}_p$ to guide the learning of anchors by sharing the consensus cluster center between the anchors and the observed representations, *i.e.*,

$$\begin{aligned} &\min_{\mathbf{W}_p,\mathbf{P},\mathbf{A}_p,\mathbf{B},\mathbf{F}} \sum_{p=1}^{V} \|\mathbf{W}_p^\top \mathbf{X}_p \mathbf{G}_p - \mathbf{B}\mathbf{A}_p \mathbf{G}_p\|_F^2 + \|\mathbf{P} - \mathbf{B}\mathbf{F}\|_F^2 \\ &\mathbf{W}_p^\top \mathbf{W}_p = \mathbf{I}_k, \mathbf{P}^\top \mathbf{P} = \mathbf{I}_m, \mathbf{A}_p^\top \mathbf{A}_p = \mathbf{I}_c, \mathbf{B}^\top \mathbf{B} = \mathbf{I}_c, \mathbf{F}^\top \mathbf{F} = \mathbf{I}_c \end{aligned} \tag{6}$$

where orthogonal constraints are imposed on $\mathbf{F} \in \mathbb{R}^{c\times n}$ and $\mathbf{A}_p \in \mathbb{R}^{c\times n}$ since $\mathbf{F}$ and $\mathbf{A}_p$ are encode matrices. Orthogonal constraint on $\mathbf{B} \in \mathbb{R}^{k\times c}$ is used to select representative cluster centers.

Meanwhile, a t-SVD based tensor low-rank constraint is imposed on bipartite graphs to deeply investigate the high-order correlations of data, *i.e.*, the pair-wise sample correlations and pair-wise view correlations. By this way, we propose DCFAI-IMVC to learn only a small amount of high-quality anchor points from incomplete data to perform fast and good imputation for large-scale IMVC tasks. This ultimate idea can be mathematically fulfilled as

$$\begin{aligned} &\min_{\substack{\mathbf{X}_p^m,\mathbf{P},\mathbf{B},\mathbf{A}_p,\mathbf{F} \\ \mathbf{W}_p,\mathbf{Z}_p,\theta_p,\mathcal{Z}}} \sum_{p=1}^{V} \theta_p^2 \|\mathbf{W}_p^\top \mathbf{X}_p - \mathbf{P}\mathbf{Z}_p\|_F^2 + \beta \|\mathcal{Z}\|_{\circledast} \\ &+ \alpha(\|\mathbf{W}_p^\top \mathbf{X}_p \mathbf{G}_p - \mathbf{B}\mathbf{A}_p \mathbf{G}_p\|_F^2 + \|\mathbf{P} - \mathbf{B}\mathbf{F}\|_F^2) \\ &\text{s.t. } \mathbf{Z}_p \geq 0, \mathbf{Z}_p^\top \mathbf{1} = \mathbf{1}, \mathbf{W}_p^\top \mathbf{W}_p = \mathbf{I}_k, \mathbf{P}^\top \mathbf{P} = \mathbf{I}_m, \\ &\boldsymbol{\theta}^\top \mathbf{1} = 1, \boldsymbol{\theta} \geq 0, \mathbf{A}_p^\top \mathbf{A}_p = \mathbf{I}_c, \mathbf{B}^\top \mathbf{B} = \mathbf{I}_c, \mathbf{F}^\top \mathbf{F} = \mathbf{I}_c \end{aligned} \tag{7}$$

3.1 Optimization

In this subsection, we can observe that the alternating direction method of multipliers (ADMM) can be used to solve Eq. (7) since it is convex. We first introduce an auxiliary variable $\mathcal{J}$ to make Eq. (7) separable according to the principle of ADMM [25, 30]. Further, we find that $\mathbf{X}_p \mathbf{G}_p \mathbf{G}_p{}^\top = \mathbf{X}_p \otimes \mathbf{H}_p$, where $\mathbf{H}_p = \mathbf{1}_{d_p} \mathbf{g}_p$ and $\mathbf{g}_p = \left[g_{p,1}, \ldots, g_{p,n}\right]^\top$ with $g_{p,j} = \sum_{l=1}^{n_p} \mathbf{G}_{p,j,l}$, which can make $O\left(vn^2\right)$ space complexity reduce to $O(dn)$. Finally, the augmented Lagrangian function of Eq. (7) can be obtained as

$$\begin{aligned} &\min_{\substack{\mathbf{X}_p^m,\mathbf{P},\mathbf{B},\mathbf{A}_p,\mathbf{F} \\ \mathbf{W}_p,\mathbf{Z}_p,\theta_p,\mathcal{J}}} \sum_{p=1}^{V} \theta_p^2 \|\mathbf{W}_p^\top \mathbf{X}_p - \mathbf{P}\mathbf{Z}_p\|_F^2 + \frac{\mu}{2} \|\mathcal{Z} - \mathcal{J} + \frac{\boldsymbol{\mathcal{Y}}}{\mu}\|_F^2 \\ &+ \alpha(\|\mathbf{W}_p^\top \mathbf{X}_p \mathbf{G}_p - \mathbf{B}\mathbf{A}_p \mathbf{G}_p\|_F^2 + \|\mathbf{P} - \mathbf{B}\mathbf{F}\|_F^2) + \beta \|\mathcal{J}\|_{\circledast} \\ &\text{s.t. } \mathbf{Z}_p \geq 0, \mathbf{Z}_p^\top \mathbf{1} = \mathbf{1}, \mathbf{W}_p^\top \mathbf{W}_p = \mathbf{I}_k, \mathbf{P}^\top \mathbf{P} = \mathbf{I}_m, \\ &\boldsymbol{\theta}^\top \mathbf{1} = 1, \boldsymbol{\theta} \geq 0, \mathbf{A}_p^\top \mathbf{A}_p = \mathbf{I}_c, \mathbf{B}^\top \mathbf{B} = \mathbf{I}_c, \mathbf{F}^\top \mathbf{F} = \mathbf{I}_c \end{aligned} \tag{8}$$

which can be solved separately as follows.

► **Step-1 update Z:** Fixing the $\mathcal{J}$, $\mathbf{W}_p$, $\mathbf{X}_p^m$, $\mathbf{P}$, $\boldsymbol{\theta}$, $\mathbf{A}_p$, $\mathbf{B}$, and $\mathbf{F}$, the $\mathbf{Z}$-subproblem can be simplified as

$$\min_{\mathbf{Z}_p} \|\mathbf{Z}_p - \hat{\mathbf{Z}}_p\|_F^2 \text{ s.t. } \mathbf{Z}_p \geq 0, \mathbf{Z}_p^\top \mathbf{1} = \mathbf{1}, \tag{9}$$

where $\hat{\mathbf{Z}}_p = \frac{(\theta_p^2 \mathbf{X}_p^\top \mathbf{W}_p)\mathbf{P} + \frac{\mu}{2}(\mathbf{J}_p^\top - \frac{\mathbf{Y}_p^\top}{\mu})}{(\sum_{p=1}^{V} \theta_p^2 \mathbf{I} + \frac{\mu}{2}\mathbf{I})}$, whose j-th column vector of p-th view is defined as $\hat{\mathbf{Z}}_{p,:,j}$, and its i-th element is $\hat{z}_{p,i,j}$.

$$\min_{\mathbf{z}_j} \left\|\mathbf{z}_j - \hat{\mathbf{z}}_j\right\|_{\mathbf{F}}^2, \text{ s.t. } \mathbf{z}_j^\top \mathbf{1} = 1, \mathbf{z}_{ij} \geq 0 \tag{10}$$

where $\hat{\mathbf{z}}_j = \frac{(\theta_p^2 \mathbf{X}_p^\top \mathbf{W}_p)\mathbf{P} + \frac{\mu}{2}(\mathbf{J}_p^\top - \frac{\mathbf{Y}_p^\top}{\mu})}{(\sum_{p=1}^{V} \theta_p^2 + \frac{\mu}{2})\mathbf{I}}$. Such subproblem can be solved by the following Theorem 1.

Theorem 1. *For any r vectors $\{\hat{\mathbf{z}}\}_{j=1}^r$, a closed-form solution $\mathbf{z}_j^*$ can be achieved as*

$$\mathbf{z}^* = \arg\min_{\mathbf{z}} \|\mathbf{z} - \hat{\mathbf{z}}\|_F^2, \text{ s.t. } \mathbf{z}^\top \mathbf{1} = 1, \mathbf{z} \geq 0 \tag{11}$$

which can be solved by Theorem 2 of [26]. The time complexity of optimizing $\mathbf{Z}$ is $O(nmd)$ with the close-form solution.

► **Step-2 update P:** Fixing the $\mathcal{J}$, $\mathbf{W}_p$, $\mathbf{X}_p^m$, $\mathbf{Z}_p$, $\boldsymbol{\theta}$, $\mathbf{A}_p$, $\mathbf{B}$, and $\mathbf{F}$, the $\mathbf{P}$-subproblem changes to

$$\max_{\mathbf{P}} \mathrm{Tr}(\mathbf{P}^\top \mathbf{C}), \mathbf{P}^\top \mathbf{P} = \mathbf{I}_m, \tag{12}$$

$\mathbf{C} = \sum_{p=1}^{V} \theta_p^2 \mathbf{W}_p^\top \mathbf{X}_p \mathbf{Z}_p^\top + \alpha \mathbf{B}\mathbf{F}$. The optimal solution of optimizing $\mathbf{P}$ can be effectively obtained via singular value decomposition (SVD) [28, 29] on $\mathbf{C}$ with complexity $\boldsymbol{O}(km^2)$.

► **Step-3 update W:** Fixing the $\mathcal{J}$, $\mathbf{P}$, $\mathbf{X}_p^m$, $\mathbf{Z}_p$, $\boldsymbol{\theta}$, $\mathbf{A}_p$, $\mathbf{B}$, and $\mathbf{F}$, the $\mathbf{W}$-subproblem becomes to

$$\max_{\mathbf{W}_p} \mathrm{Tr}(\mathbf{W}_p^\top \mathbf{D}_p), \mathbf{W}_p^\top \mathbf{W}_p = \mathbf{I}_k, \tag{13}$$

where $\mathbf{D}_p = \theta_p^2 \mathbf{W}_p^\top \mathbf{X}_p \mathbf{Z}_p^\top + \beta(\mathbf{X}_p \otimes \mathbf{H}_p)\mathbf{A}_p^\top \mathbf{B}$. Similar to update $\mathbf{P}$, the $\mathbf{W}$-subproblem can be effectively solved via SVD on $\mathbf{D}_p$. The complexity of optimizing $\mathbf{W}$ is $\boldsymbol{O}(Vdk^2)$, where $d = \sum_{p=1}^{V} d_p$.

► **Step-4 update $\mathbf{X}_p^m$:** Fixing the irrelevant variables, and updating $\mathbf{X}_p^m$ can be written as

$$\mathbf{X}_p^m = \mathbf{E}_p(:, \boldsymbol{v}_p^m) \tag{14}$$

where $\mathbf{E}_p(:, \boldsymbol{v}_p^m) = \sum_{p=1}^{V} \theta_p^2 \mathbf{W}_p \mathbf{P} \mathbf{Z}_p(:, \boldsymbol{v}_p^m)$, and $\boldsymbol{v}_p^m$ denotes the missing index of p-th view. The complexity is $\boldsymbol{O}(dkm + km \sum_{p=1}^{V} n_p)$.

► **Step-5 update $\boldsymbol{\theta}$:** Optimizing $\boldsymbol{\theta}$ with the irrelevant variables fixed is equivalent to the following optimization problem

$$\min_{\theta_p} \sum_{p=1}^{V} {\theta_p}^2 \delta_p^2, \text{ s.t. } \boldsymbol{\theta}^\top \mathbf{1} = 1, \boldsymbol{\theta} \geq 0 \tag{15}$$

where $\delta_p = \|\mathbf{W}_p^\top \mathbf{X}_p - \mathbf{P}\mathbf{Z}_p\|_F^2$. According to Cauchy-Schwarz inequality, $\boldsymbol{\theta}$ can be optimally solved via

$$\boldsymbol{\theta} = \frac{\boldsymbol{\varepsilon}}{\sum_{p=1}^{V} \varepsilon_p} \tag{16}$$

where $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_V]$ and $\varepsilon_p = \frac{1}{\delta_p}$. The complexity of optimizing β is $\boldsymbol{O}(nmd)$.

► **Step-6 update B:** Fixing $\mathcal{J}$, $\mathbf{W}_p$, $\mathbf{P}$, $\mathbf{X}_p^m$, $\mathbf{Z}_p$, $\boldsymbol{\theta}$, $\mathbf{A}_p$, and $\mathbf{F}$, $\mathbf{B}$-subproblem of (8) changes to

$$\max_{\mathbf{B}} \mathrm{Tr}(\mathbf{B}^\top \mathbf{T}) \ \text{ s.t. } \mathbf{B}^\top \mathbf{B} = \mathbf{I}_k. \tag{17}$$

where $\mathbf{T} = \mathbf{F}^\top \mathbf{P} + \sum_{p=1}^{V} \mathbf{W}_p^\top (\mathbf{X}_p \otimes \mathbf{H}_p) \mathbf{A}_p^\top$. The optimal solution of optimizing $\mathbf{B}$ can be effectively obtained via singular value decomposition (SVD) on $\mathbf{T}_p$ with complexity $\boldsymbol{O}(dm^2 + nmd)$, where $d = \sum_{p=1}^{V} d_p$.

► **Step-7 update F:** Fixing $\mathcal{J}$, $\mathbf{W}_p$, $\mathbf{P}$, $\mathbf{X}_p^m$, $\mathbf{Z}_p$, $\boldsymbol{\theta}$, $\mathbf{A}_p$, and $\mathbf{B}$, $\mathbf{F}$-subproblem of (8) can be rewritten as

$$\max_{\mathbf{F}} \mathrm{Tr}(\mathbf{F}^\top \mathbf{P} \mathbf{B}^\top) \ \text{ s.t. } \mathbf{F}^\top \mathbf{F} = \mathbf{I}_k, \tag{18}$$

which can be solved with via SVD operator with complexity $\boldsymbol{O}(nc^2)$.

► **Step-8 update A:** Fixing $\mathcal{J}$, $\mathbf{W}_p$, $\mathbf{P}$, $\mathbf{X}_p^m$, $\mathbf{Z}_p$, $\boldsymbol{\theta}$, $\mathbf{B}$, and $\mathbf{F}$, $\mathbf{A}$-subproblem of (8) can be reformulated as

$$\max_{\mathbf{A}_p} \mathrm{Tr}(\mathbf{A}_p^\top ((\mathbf{X}_p \otimes \mathbf{H}_p) \mathbf{W}_p \mathbf{B}^\top)) \ \text{ s.t. } \mathbf{A}_p^\top \mathbf{A}_p = \mathbf{I}_k. \tag{19}$$

Similar to Eq. (13), Eq. (19) can be availably solved via SVD with complexity $\boldsymbol{O}(nc^2)$.

► **Step-9 update $\mathcal{J}$:** Ignoring the irrelevant items *w.r.t.* $\mathcal{J}$, updating $\mathcal{J}$ can be rewritten as

$$\min_{\mathcal{J}} \ \beta \|\mathcal{J}\|_{\circledast} + \frac{\mu}{2} \|\mathcal{J} - (\mathcal{Z} + \frac{\mathcal{Y}}{\mu})\|_F^2 \tag{20}$$

Denoting $\mathcal{M} = \mathcal{Z} + \frac{\mathcal{Y}}{\mu}$, $\mathcal{J}$ can be solved via the tensor tubal-shrinkage of the below Theorem 2 [40].

Theorem 2 *Give two 3-order tensor $\mathcal{J} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with a scalar, the global optimal solution to the following problem*

$$\min_{\mathcal{J}} \rho \|\mathcal{J}\|_{\circledast} + \frac{1}{2} \|\mathcal{J} - \mathcal{M}\|_F^2 \tag{21}$$

is given by the tensor tubal-shrinkage operator.

$$\mathcal{J} = \mathcal{C}_{n_3 \rho}(\mathcal{M}) = \mathcal{U} * \mathcal{C}_{n_3 \rho}(\mathcal{Z}) * \mathcal{V}^\top \tag{22}$$

where $\mathcal{M} = \mathcal{U} * \mathcal{Z} * \mathcal{V}^\top$ and $\mathcal{C}_{n_3 \rho}(\mathcal{Z}) = \mathcal{Z} * \mathcal{Q}$. $\mathcal{Q} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ denotes a f-diagonal tensor and each diagonal element of $\mathcal{Q}$ is defined as $\mathcal{Q}_f(i,i,j) = (1 - \frac{n_3 \rho}{\mathcal{Z}^{(j)}(i,i)})_+$. The complexity of updating $\mathcal{J}$ is $\boldsymbol{O}\left(nm\log(n) + m^2 n\right)$.

Updating ADMM variables are written as

$$\mathcal{Y} = \mathcal{Y} + \mu(\mathcal{Z} - \mathcal{J}), \ \mu = min(\rho\mu, \mu_{max}) \tag{23}$$

where $\mu = 1e^{-4}$ and $\mu_{max} = 10^{10}$, and the complexity is $\boldsymbol{O}(n)$. The whole optimization procedure of Eq. (7) is outlined in Algorithm 1, where convergence criterion is checked by computing the objective value obj^t at the t-th iteration.

Algorithm 1 The algorithm of DCFAI-IMVC

Input: V incomplete multi-view data $\{\mathbf{X}_p\}_{p=1}^V$, number of clusters c, dimension of consensus proxy l, and parameter α.
Initialize $\boldsymbol{\theta}_p = 1/V$, $\mathbf{W}_p = \mathbf{I}_k$, and the others matrices as $\mathbf{0}$.

1: **repeat**
2: Update $\mathbf{Z}_p$ by using Eq. (9);
3: Update $\mathbf{P}$, $\mathbf{W}_p$, $\mathbf{X}_p^m$ and $\boldsymbol{\theta}$ by using Eqs (12)-(15), respectively;
4: Update $\mathbf{B}$, $\mathbf{F}$, $\mathbf{A}_p$, and $\mathcal{J}$ by using Eqs. (17)-(20), respectively;
5: **until** Satisfy $(obj^{(t)} - obj^{(t-1)})/obj^{(t)} \leq 1e-4$.
6: Perform SVD on the averaged bipartite graph $\mathbf{Z}^* = \sum_{p=1}^{V} \mathbf{Z}_p / V$ to obtain left singular value matrix $\mathbf{U}_{\mathbf{Z}^*} \in \mathbb{R}^{m \times n}$.

Output: Perform k-means on $\mathbf{U}_{\mathbf{Z}^*}$ to obtain the clustering results.

Table 1: Complexity analysis of SOTA competitors.

Method	Space Cost	Time Complexity
BSV [23]	Vn^2	$O(n^3)$
MIC [27]	$Vn^2 + nd + nVk + Vdk$	$O(n^3 + n^2dk)$
DAIMC [11]	$Vn^2 + nd + nk + dk$	$O(nd^3 + ndk)$
APMC [9]	$nd + Vnm + nk$	$O(n^3 + nmd + m^3)$
UEAF[36]	$Vn^2 + dn + nVk + dk$	$O(n^3 + dk^2)$
MKKM-IK [22]	$Vn^2 + Vnk$	$O(Vn^3)$
EEIMVC [20]	$Vn^2 + Vnk + Vk^2$	$O(nk^2 + Vk^3)$
FLSD [37]	$Vn^2 + dnk + nk$	$O(nd^2)$
UTF [38]	$Vn^2 + V(n - n_p)d$	$O(Vn^3 + Vn^2log(n) + V^2n^2)$
IMVC-CBG[31]	$mn + (d + m)k$	$O(ndk + nmd + mdk)$
HCP-IMSC[18]	$Vn^2 + V(n - n_p)d + dk$	$O(Vn^3 + V(n - n_p)^3 + kn^2V)$
Ours	$mn + (d + m)k$	$3\boldsymbol{O}(nmd) + \boldsymbol{O}(mn\log(n))$

3.2 Complexity Analysis

From **Step-1** to **Step-9**, the computational complexity of Algorithm 1 is $3\boldsymbol{O}(nmd) + \boldsymbol{O}(km^2) + \boldsymbol{O}(Vdk^2) + \boldsymbol{O}(dkm + km\sum_{p=1}^{V} n_p) + \boldsymbol{O}(dm^2) + \boldsymbol{O}(nc^2) + \boldsymbol{O}(mn\log(n) + m^2 n)$ at each iteration, where n, n_p, c and m denote number of complete samples, missing samples, cluster number and anchors, respectively. k is the dimension of latent space. After obtaining $\mathbf{Z}^*$, it costs $\boldsymbol{O}(nm^2)$ complexity to perform k. For another, space complexity of Algorithm 1 mainly involves four matrices $\mathbf{W}_p \in \mathbb{R}^{d_p \times m}$, $\mathbf{X}_p \in \mathbb{R}^{d_p \times n}$, $\mathbf{P} \in \mathbb{R}^{m \times k}$, and $\mathbf{Z}_p \in \mathbb{R}^{m \times n}$. Therefore, space complexity of DCFAI-IMVC is $(n + k)(d + m)$. Due to $k \ll c$, $k \ll n$, $m \ll n$, and $d \ll n$, Algorithm 1 inherits linear time and space complexities with respect to the number of samples. Table 1 reports the main time and space complexity of all compared methods.

4 EXPERIMENTS

4.1 Datasets and Experimental Setting

Table 2: Detailed information of the used datasets.

Dataset	Sample	Classes	Views	Dimensionality
ORL	400	40	3	4096 /3304 / 6750
NGs	500	5	3	500 / 500 / 500
Handwritten	2000	10	6	240 / 76/ 216/ 47 / 64 / 6
Caltech101-20	2386	20	6	48 / 40 / 254 / 198 / 512 / 928
SUNRGBD	10335	45	2	4096/4096
NUSWIDE	30000	31	5	65/226/145/74/129
Cifar10	50000	10	3	2048 / 512 / 1024
Cifar100	50000	100	3	2048 / 512 / 1024

Following the approach in [31], we generate incomplete datasets by setting missing ratio ψ from 0.1 to 0.9 with step of 0.1, *i.e.*, $\psi = \frac{n_p}{n} \in [0.1 : 0.1 : 0.9]$. Seven widely used benchmark datasets are employed, including: Cifar10, Cifar100, NUSWIDE, SUNRGBD, Handwritten, Caltech101-20, NGs, and ORL. Detailed information of these datasets is provided in Table 2. Concretely, Cifar100, Cifar10, NUSWIDE, SUNRGBD, Handwitten, Caltech101-20, and ORL are the image datasets. NGs is the web page dataset. Note that the number of samples in these datasets ranges from 400 to 50,000. This span is already relatively large in existing IMVC.

Eleven IMVC methods are employed for comparison. These competitors can be divided into two categories, **(1) Deleting Missing Data (DMD):** First category directly deletes missing data, and then the remaining observable data is used for clustering, including **MIC** [27], **APMC** [9], **FLSD** [37], and **IMVC-CBG** [31]. **(2) Filling Missing Data (FMD):** The second category imputes the missing views and clustering, including **BSV** [23] (the optimal value with single view mean imputation), **DAIMC** [11], **UEAF** [36], **IMKKM-IK** [22], **EEIMVC** [20], **UTF** [38], **HCP-IMSC** [18], and our **DCFAI-IMVC**. Note that all the experiments are implemented from their publicly released codes in corresponding papers. Computing platform is an Intel Core i9 CPU and 64GB RAM.

4.2 Experiment Results

The superiority of our method are validated by comparing to 11 state-of-the-art methods on four common metrics [20] (*i.e.*, accuracy (ACC), normalized mutual information (NMI), Purity (PUR) and Fscore (FSC). The average clustering performance and standard deviations (Each experiment is repeated 20 times.) on the all evaluated datasets are reported in Table 3. The best and second best averages are marked in bold and underline, respectively. Meanwhile, Figs. 3-4 present the variation for our method with different missing ratios on all evaluated metrics. "-" denotes the unavailable results of corresponding competitors due to the long execution time or out of space in this paper. From these tables and figures, we can observed that:

Effectiveness. As shown in Table 1 and Figs. 3-4, some IMVC methods are worse than single-view method BSV, while our DCFAI-IMVC not only consistently outperforms the single-view incomplete clustering, but outperforms all the recent proposed IMVC methods. This demonstrates effectiveness of our method in IMVC.

Superiority over DMD based methods. Our method consistently outperforms DMD based methods, *i.e.*, MIC, APMC, and FLSD, and the recent proposed IMVC-CBG. Even compared to IMVC-CBG, the proposed DCFAI-IMVC is consistently superior to IMVC-CBG on the all evaluated datasets. Meanwhile, the performance of our method also falls more slowly than IMVC-CBG with the increasing of the missing ratio even on two large-scale datasets, Cifar10 and Cifar100 (It is more obvious when missing ratio equals to 0.8 or 0.9.). The main reasons are that, firstly, IMVC-CBG directly discards missing samples to destroy the original structural relation hidden in the incomplete multi-view data. Secondly, more importantly, discarding missing samples can greatly affect the objectivity of data and the correctness of results, especially when the missing proportion is relatively large. This proves the superiority of imputation missing data for our DCFAI-IMVC.

Superiority over FMD based methods. Compared to FMD based methods, *i.e.*, DAIMC, UEAF, IMKKM-IK, EEIMVC, UTF, and HCP-IMSC, the proposed DCFAI-IMVC simultaneously have the optimal clustering performance, the lowest time and space complexity as shown in Table 3, Table 1, and Figs. 3-4. UTF and HCP-IMSC have been regarded as very competitive FMD based IMVC methods. By taking their results for example, from Table 3, UTF and HCP-IMSC outperform the optimal DMD based method IMVC-CBG in most cases. But UTF and HCP-IMSC are lower than IMVC-CBG in the ACC of the Caltech-20 dataset. Inversely, DCFAI-IMVC is superior to the most recent methods by a large margin, achieving the SOTA clustering performance. More importantly, Table 1 indicates that most of FMD based methods have $\boldsymbol{O}(n^3)$ computational complexity and $\boldsymbol{O}(n^2)$ space complexity. While DCFAI-IMVC enjoys a linear time and space complexity in large-scale tasks, which is very rare as a FMD based method to simultaneously achieve the SOTA performance, the lowest time complexity, and space complexity.

For another, Fig. 5 (a)-(c) compare the quality of the filled View #1 data for three competitors, *i.e.*, UTF, HCP-IMSC, and our DCFAI-IMVC. We can observe that our DCFAI-IMVC has clearer cluster discriminability and tighter cluster structure than UTF and HCP-IMSC, which indicates that the quality of our filled data is superior to these two competitors. Compared to the original incomplete data on Handwritten in Fig. 6 (a), intuitively, both Fig. 5 (a) and (b) have many wrong imputation values, while almost all of the filled data of our DCFAI-IMVC in Fig. 5 (c) is filled correctly. The main reason is that UTF and HCP-IMSC use all the observed data to fill in the missing data in the original space, rather than learning the high-quality anchors to perform imputation in the latent space. Then, the redundant information or fraud features in observed data leads to poor imputation quality. In addition, Fig. 5 (d)-(f) also show that affinity graph learned by our DCFAI-IMVC has the more compact cluster structure than the graphs of UTF and HCP-IMSC. By this way, the superiority of the proposed method is proved once again.

Validation on the imputation data. Firstly, Fig. 5 have shown the superiority of our DCFAI-IMVC in imputation the missing data compared to the most recent UTF and HCP-IMSC. Furthermore, Fig. 6 has reported the filled results of the incomplete data, the filled data and the learned latent data of our DCFAI-IMVC on View #1 of Handwriten dataset (ψ=0.9). Compare with Fig. 6 (a) and (b), the observed data of black ellipses doesn't change before and after imputation, while red ellipses represents the change from zeros imputation of (a) to the imputation values learned by our DCFAI-IMVC. As can be seen, the filled data (*i.e.*, red ellipses of (b)) by our DCFAI-IMVC is almost exactly right one-to-one correspondence between clusters. In addition, compare (a) with (c), the learned latent data accurately recover the cluster structure of (a), which indicates that the learned latent data can be well used to characterize the data of (a). The above analyses forcefully prove the effectiveness of imputation missing data in latent space rather than original space.

4.3 Parameter Settings and Validity

Algorithm 1 involves three parameters to be set properly, *i.e.*, parameters α, β, and anchor number m. As shown in Fig. 7, by performing grid search on the large-scale Cifar10 dataset, we first vary parameter α and β in $2^{[-5:5]}$ with the fixed $m = 3k$. Then, we tune m

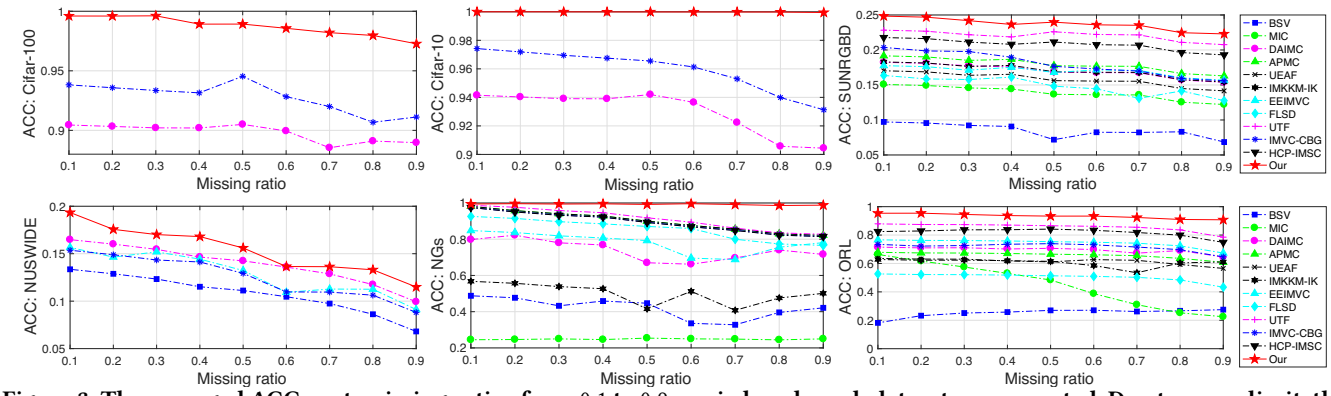


Figure 3: The averaged ACC w.r.t. missing ratios from 0.1 to 0.9 on six benchmark datasets are reported. Due to space limit, the clustering results of other metrics are omitted.

Table 3: Average performance comparison *w.r.t.* different missing ratios. Best/second best results are marked in bold and underline, respectively.

Dataset	Metrics	BSV	MIC	DAIMC	APMC	UEAF	IMKKM-IK	EEIMVC	FLSD	UTF	IMVC-CBG	HCP-IMSC	Our
ORL	ACC	25.18±0.73	44.94 ±1.50	69.45 ±2.16	65.80 ±1.75	61.49 ±2.34	60.41 ±2.73	74.36 ±2.38	50.33 ±1.69	85.57 ±0.16	71.65 ±2.53	81.88 ±0.03	**93.32 ±0.02**
	NMI	48.92±0.75	57.44 ±0.84	82.44 ±0.9	81.10 ±0.66	77.58 ±1.09	79.60 ±1.10	85.93 ±1.16	68.07 ±1.12	92.98 ±0.08	80.61 ±1.46	90.85 ±0.02	**97.24 ±0.01**
	PUR	26.34±0.76	41.17±1.24	73.83 ±1.63	70.18 ±1.32	64.05 ±1.75	61.56 ±2.47	76.82 ±2.03	51.49 ±1.56	88.13 ±0.12	71.65 ±2.02	84.44 ±0.03	**94.52 ±0.01**
	FSC	9.86±0.53	18.80 ±1.02	56.97 ±2.71	50.73 ±2.39	40.23 ±2.58	45.20 ±2.91	64.52 ±2.69	31.91 ±1.84	79.10 ±0.06	47.05 ±3.83	75.42 ±0.01	**91.42 ±0.02**
NGs	ACC	42.37±1.83	24.86±0.26	73.31±9.07	89.89±0.01	89.89±0.03	50.36±0.13	78.37±0.12	85.47±0.03	91.13±0.09	89.26±0.02	89.11±0.00	**99.26±0.00**
	NMI	22.79±1.17	9.28±0.29	58.93±6.28	73.73±0.03	73.73±0.08	33.49±0.11	58.92±0.18	66.48±0.03	75.92±0.15	73.90±0.06	72.90±0.00	**97.59±0.01**
	PUR	44.37±1.31	25.79±0.22	73.31±9.07	89.89±0.01	89.89±0.03	51.23±0.10	78.37±0.12	85.47±0.03	91.13±0.09	89.26±0.02	89.11±0.00	**99.26±0.00**
	FSC	32.97±0.88	33.27±0.13	74.11±7.29	81.29±0.02	81.29±0.06	42.84±0.09	64.00±0.14	72.02±0.03	80.04±0.11	81.23±0.03	80.04±0.00	**98.51±0.00**
Caltech-20	ACC	39.91±0.23	26.69±1.68	45.72±1.93	-	39.71±1.38	32.21±1.66	41.33±1.29	43.45±1.82	49.48±2.09	50.56±1.32	50.05±0.46	**52.62±0.03**
	NMI	25.58±0.83	30.66±1.13	55.70±1.36	-	50.90±0.92	40.09±0.98	54.74±0.61	52.33±0.90	72.31±1.32	52.86±1.61	60.37±0.53	**77.50±0.01**
	PUR	48.70±0.69	55.89±1.13	74.89±0.88	-	70.90±0.74	60.14±1.38	72.28±0.79	73.37±0.53	82.22±1.76	71.04±0.93	76.55±0.45	**86.70±0.01**
	FSC	32.74±0.29	24.46±1.26	40.46±2.19	-	33.34±1.49	29.18±1.65	40.70±1.43	41.50±1.56	43.56±2.62	44.59±2.32	44.81±0.68	**44.82±0.02**
Handwritten	ACC	55.85±1.41	68.96±0.53	90.08±0.52	88.96±0.07	90.18±0.12	88.45±0.15	89.88±0.22	91.14±0.08	95.57±0.98	96.96±0.06	92.56±0.84	**99.27±0.00**
	NMI	52.17±1.55	65.28±0.46	86.40±3.31	85.28±0.05	86.50±0.07	84.83±0.14	86.20±0.16	87.46±0.07	94.54±0.93	95.93±0.02	91.36±1.42	**98.28±0.01**
	PUR	55.85±1.12	68.96±0.23	90.08±1.53	88.96±0.06	90.18±0.10	88.45±0.13	89.88±0.18	91.14±0.09	95.57±1.17	96.96±0.03	92.56±0.96	**99.27±0.00**
	FSC	52.34±1.08	65.45±0.19	86.57±1.49	85.45±0.02	86.67±0.06	84.83±0.09	86.37±0.14	87.63±0.05	94.80±0.76	96.19±0.03	91.09±0.61	**98.54±0.00**
SUNRGBD	ACC	8.49±0.06	13.84±0.48	17.01±0.76	17.93±0.55	15.80±0.39	17.06±0.36	16.97±0.46	14.82±0.39	22.04±1.67	18.05±0.18	20.76±0.06	**23.69±0.01**
	NMI	6.31±0.06	21.02±0.28	21.34±0.35	21.83±0.22	21.21±0.28	20.95±0.28	20.86±0.24	20.86±0.19	35.00±1.32	23.92±0.16	33.72±0.08	**37.65±0.01**
	PUR	13.22±0.18	32.58±0.64	34.90±0.81	34.99±0.39	34.78±0.62	34.51±0.86	34.42±0.37	32.66±0.34	42.55±1.64	33.55±0.28	41.27±0.03	**46.20±0.01**
	FSC	7.18±0.02	9.53±0.18	10.66±0.32	10.98±0.28	9.50±0.14	10.11±0.19	10.02±0.22	11.45±0.01	16.16±1.19	10.51±0.38	14.88±0.05	**16.81±0.01**
NUSWIDEOBJ	ACC	10.76±0.20	-	13.89±0.18	-	-	-	12.86±0.26	-	-	12.56±0.13	-	**15.38±1.21**
	NMI	3.39±0.06	-	11.34±0.39	-	-	-	10.52±0.09	-	-	10.40±0.03	-	**13.17±0.03**
	PUR	14.75±0.08	-	23.69±0.32	-	-	-	21.88±0.23	-	-	21.01±0.15	-	**22.88±0.12**
	FSC	**10.75**±0.05	-	8.62±0.59	-	-	-	8.01±0.06	-	-	7.91±0.03	-	8.28±0.03
Cifar10	ACC	-	-	90.81±0.45	-	-	-	-	-	-	96.19±0.13	-	**99.99±0.00**
	NMI	-	-	90.47±0.55	-	-	-	-	-	-	90.89±0.27	-	**99.95±0.00**
	PUR	-	-	95.81±0.45	-	-	-	-	-	-	96.19±0.26	-	**99.99±0.00**
	FSC	-	-	92.16±0.68	-	-	-	-	-	-	92.67±0.13	-	**99.98±0.00**
Cifar100	ACC	-	-	89.71±1.00	-	-	-	-	-	-	93.09±1.18	-	**98.74±0.00**
	NMI	-	-	98.26±0.16	-	-	-	-	-	-	98.63±0.29	-	**99.72±0.00**
	PUR	-	-	92.60±0.54	-	-	-	-	-	-	94.98±0.82	-	**98.78±0.00**
	FSC	-	-	90.82±0.94	-	-	-	-	-	-	90.87±2.97	-	**98.72±0.00**

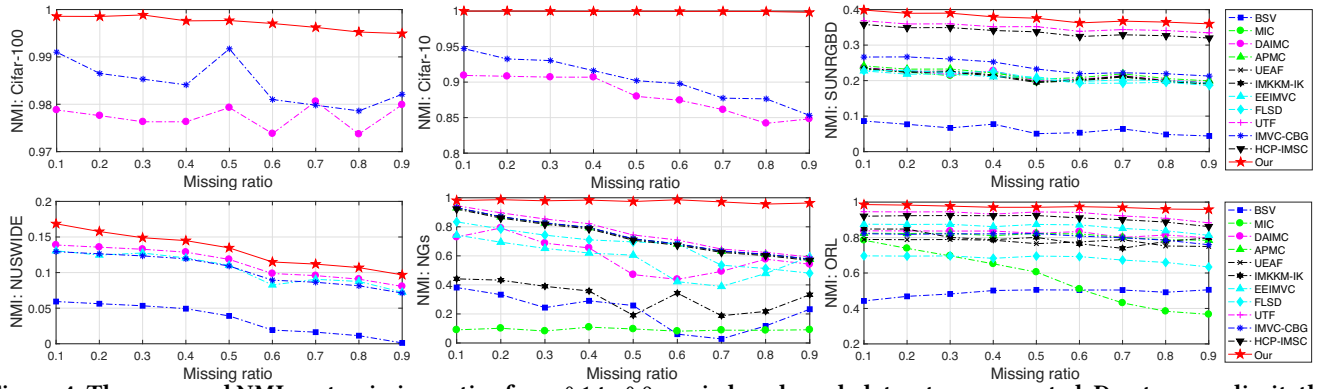
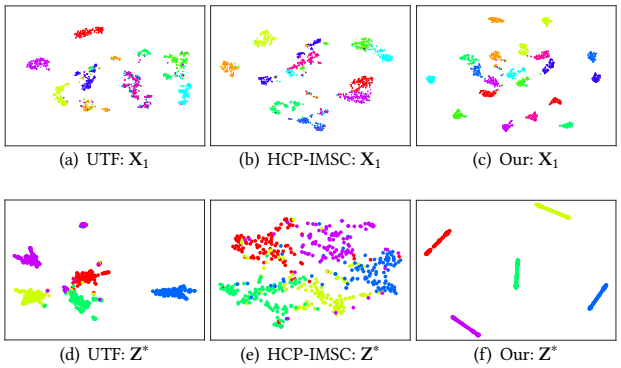


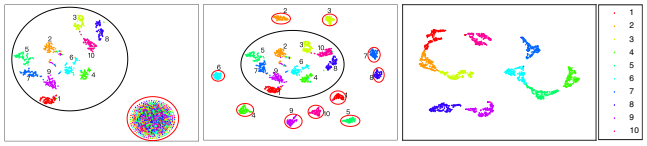
Figure 4: The averaged NMI w.r.t. missing ratios from 0.1 to 0.9 on six benchmark datasets are reported. Due to space limit, the clustering results of other metrics are omitted.



(a) UTF: $\mathbf{X}_1$ (b) HCP-IMSC: $\mathbf{X}_1$ (c) Our: $\mathbf{X}_1$

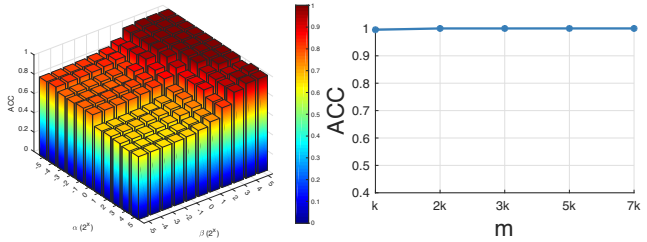
(d) UTF: $\mathbf{Z}^*$ (e) HCP-IMSC: $\mathbf{Z}^*$ (f) Our: $\mathbf{Z}^*$

Figure 5: The filled View #1 data and the learned graphs of UTF, HCP-IMSC, and our DCFAI-IMVC are visualized with t-SNE on Handwritten and NGs datasets (ψ=0.9).



(a) Incomplete data $\mathbf{X}_1$ (b) Filled data $\mathbf{X}_1$ (c) Latent data $\mathbf{W}_1\mathbf{X}_1$

Figure 6: Visualization of incomplete View #1 data, the filled View #1 data and the learned latent View #1 data with t-SNE on Handwriten dataset (ψ=0.9). Note that black and red ellipses in (a) represent the observed data $\mathbf{X}_1^o$ and missing data $\mathbf{X}_1^m$ with zeros imputation. And black and red ellipses in (b) denote the observed data $\mathbf{X}_1^o$ and filled missing data $\mathbf{X}_1^m$.



(a) Cifar10 *w.r.t.* α and β (b) Cifar10 *w.r.t.* m

Figure 7: Cifar10: ACC *w.r.t.* α, β, and m (ψ=0.2).

in $[k, 2k, 3k, 5k, 7k]$ with the fixed $\alpha = 2^{-1}$ and $\beta = 2^4$. Fig. 7 (a) indicates that our DCFAI-IMVC can achieve the satisfying performance in a wide scope for α, β. Thus, Algorithm 1 is insensitive to α and β. And Fig. 7 (b) shows that even a small number of anchors can achieve the excellent and stable clustering performance. This demonstrates that our method only requires a small number of anchors to fill the missing data, achieving the promising performance.

Table 4 reports the running time of all competitors on all benchmark datasets, and demonstrates that our DCFAI-IMVC can achieve very competitive computational efficiency over all evaluated datasets and competitors. Especially on the datasets Cifar10 and Cifar100 of 50,000 samples, our DCFAI-IMVC has the lowest time cost. Meanwhile, many existing IMVC methods fail due to out of space or the long execution time. Additionally, Fig. 8 reports the convergence curve of Cifar10, Cifar100, and ORL, where the objective value is monotonically and fast reducing close to a stable point. Overall, DCFAI-IMVC achieves the lower time cost and space complexity.

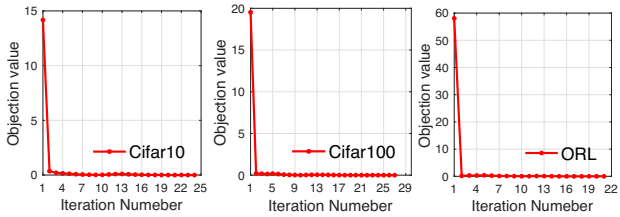


Figure 8: Objection function value over the Cifar10, Cifar100, and ORL datasets (ψ=0.2).

Table 4: Computational cost (In Seconds). Handwritten and NUSWIDEOBJ are respectively simplifies as HW and NUS in this table.

Method	ORL	NGs	Caltech-20	HW	SUNRGBD	NUS	Cifar10	Cifar100
BSV	0.97	0.29	2.32	2.79	2361.05	9624.38	-	-
MIC	418.37	143.94	2847.45	3106.34	6107.6	-	-	-
DAIMC	1205.6	42.61	69.88	78.8	184.06	1667.67	29948.48	26349.92
APMC	0.58	0.61	-	98.23	156.37	-	-	-
UEAF	14.3	2.36	26.63	36.24	89.61	-	-	-
IMKKM-IK	0.58	1.47	126.15	168.92	309.29	-	-	-
EE-R-IMVC	0.73	0.57	3.96	4.31	5.68	1872.44	-	-
FLSD	87.44	3.41	54.04	68.92	130.79	-	-	-
UTF	10.72	6.32	60.38	70.64	209.34	-	-	-
IMVC-CBG	1.63	1.42	5.39	7.24	8.61	22.54	69.18	155.62
HCP-IMSC	8.37	6.32	157.38	80.58	358.34	-	-	-
Our	2.34	3.29	6.43	4.95	10.68	34.49	48.23	86.31

Meanwhile, it also has very competitive clustering performance for IMVC. Therefore, DCFAI-IMVC enjoys more potential to handle large-scale incomplete tasks in practical application.

4.4 Conclusion

In this paper, we propose a novel and highly-efficient IMVC method to fast and good missing data imputation for improving clustering termed DCFAI-IMVC. Different from employing all the original observed data of existing IMVC methods to fill missing data, DCFAI-IMVC can efficiently and flexibly learn a small number of anchors to large-scale missing data imputation. DCFAI-IMVC is the first practice to integrate anchor learning and missing data imputation into a unified model to address large-scale IMVC tasks. Comprehensive experiments and analysis have proved the effectiveness, superiority, and efficiency of our method.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (Grant nos. 61673220 and 62106209), the Base Strengthening Program of National Defense Science and Technology (Grant no. 2022-JCJQ-JJ-0292), the Grant from Guangxi Key Laboratory of Machine Vision and Intelligent Control (Grant no. 2022B07), the Open Research Fund of Anhui Province Key Laboratory of Machine Vision Inspection (Grant no. KLMVI-2023-HIT-08), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant no. KYCX23_0487).

REFERENCES

[1] Mansheng Chen, Tuo Liu, Chang-Dong Wang, Dong Huang, and Jian-Huang Lai. 2022. Adaptively-weighted Integral Space for Fast Multiview Clustering. In *The 30th ACM International Conference on Multimedia*. 3774–3782.

[2] Man-Sheng Chen, Chang-Dong Wang, Dong Huang, Jian-Huang Lai, and Philip S Yu. 2022. Efficient Orthogonal Multi-view Subspace Clustering. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 127–135.

[3] Man-Sheng Chen, Chang-Dong Wang, and Jian-Huang Lai. 2022. Low-rank Tensor Based Proximity Learning for Multi-view Clustering. *IEEE Transactions on Knowledge and Data Engineering* (2022).

[4] Zhe Chen, Xiao-Jun Wu, Tianyang Xu, and Josef Kittler. 2023. Fast Self-guided Multi-view Subspace Clustering. *IEEE Transactions on Image Processing* (2023).

[5] Zhaoliang Chen, Zhihao Wu, Zhenghong Lin, Shiping Wang, Claudia Plant, and Wenzhong Guo. 2023. AGNN: Alternating Graph-Regularized Neural Networks to Alleviate Over-Smoothing. *IEEE Transactions on Neural Networks and Learning Systems* (2023), 1–13. https://doi.org/10.1109/TNNLS.2023.3271623

[6] Zhaoliang Chen, Zhihao Wu, Shiping Wang, and Wenzhong Guo. 2023. Dual Low-Rank Graph Autoencoder for Semantic and Topological Networks. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, Vol. 37. 4191–4198.

[7] Shijie Deng, Jie Wen, Chengliang Liu, Ke Yan, Gehui Xu, and Yong Xu. 2023. Projective Incomplete Multi-View Clustering. *IEEE Transactions on Neural Networks and Learning Systems* (2023).

[8] Zhiqiang Fu, Yao Zhao, Dongxia Chang, Yiming Wang, and Jie Wen. 2022. Latent Low-Rank Representation With Weighted Distance Penalty for Clustering. *IEEE Transactions on Cybernetics* (2022).

[9] Jun Guo and Jiahui Ye. 2019. Anchors bring ease: An embarrassingly simple approach to partial multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 118–125.

[10] Wenjue He, Zheng Zhang, Yongyong Chen, and Jie Wen. 2023. Structured anchor-inferred graph learning for universal incomplete multi-view clustering. *World Wide Web* 26, 1 (2023), 375–399.

[11] Menglei Hu and Songcan Chen. 2019. Doubly aligned incomplete multi-view clustering. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-19* (2019), 2262–2268.

[12] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. 2020. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 4412–4419.

[13] Liang Li, Siwei Wang, Xinwang Liu, En Zhu, Li Shen, Kenli Li, and Keqin Li. 2022. Local Sample-Weighted Multiple Kernel Clustering With Consensus Discriminative Graph. *IEEE Transactions on Neural Networks and Learning Systems* (2022), 1–14.

[14] Liang Li, Junpu Zhang, Siwei Wang, Xinwang Liu, Kenli Li, and Keqin Li. 2023. Multi-View Bipartite Graph Clustering With Coupled Noisy Feature Filter. *IEEE Transactions on Knowledge and Data Engineering* (2023), 1–13.

[15] Xingfeng Li, Zhenwen Ren, Quansen Sun, and Zhi Xu. 2023. Auto-weighted tensor schatten p-norm for robust multi-view graph clustering. *Pattern Recognition* 134 (2023), 109083.

[16] Xingfeng Li, Yinghui Sun, Quansen Sun, and Zhenwen Ren. 2023. Consensus Cluster Center Guided Latent Multi-Kernel Clustering. *IEEE Transactions on Circuits and Systems for Video Technology* 33, 6 (2023), 2864–2876. https://doi.org/10.1109/TCSVT.2022.3229356

[17] Xingfeng Li, Yinghui Sun, Quansen Sun, Zhenwen Ren, and Yuan Sun. 2023. Cross-view Graph Matching guided Anchor Alignment for incomplete multi-view clustering. *Information Fusion* (2023), 101941.

[18] Zhenglai Li, Chang Tang, Xiao Zheng, Xinwang Liu, Wei Zhang, and En Zhu. 2022. High-order correlation preserved incomplete multi-view subspace clustering. *IEEE Transactions on Image Processing* 31 (2022), 2067–2080.

[19] Jiyuan Liu, Xinwang Liu, Yi Zhang, Pei Zhang, Wenxuan Tu, Siwei Wang, Sihang Zhou, Weixuan Liang, Siqi Wang, and Yuexiang Yang. 2021. Self-representation Subspace Clustering for Incomplete Multi-view Data. In *Proceedings of the 29th ACM International Conference on Multimedia*. 2726–2734.

[20] Xinwang Liu, Miaomiao Li, Chang Tang, Jingyuan Xia, Jian Xiong, Li Liu, Marius Kloft, and En Zhu. 2020. Efficient and effective regularized incomplete multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence* 43, 8 (2020), 2634–2646.

[21] Xinwang Liu, Li Liu, Qing Liao, Siwei Wang, Yi Zhang, Wenxuan Tu, Chang Tang, Jiyuan Liu, and En Zhu. 2021. One pass late fusion multi-view clustering. In *International Conference on Machine Learning*. PMLR, 6850–6859.

[22] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, En Zhu, Tongliang Liu, Marius Kloft, Dinggang Shen, Jianping Yin, and Wen Gao. 2019. Multiple kernel k k-means with incomplete kernels. *IEEE transactions on pattern analysis and machine intelligence* 42, 5 (2019), 1191–1204.

[23] Feiping Nie, Jing Li, Xuelong Li, et al. 2017. Self-weighted Multiview Clustering with Multiple Graphs.. In *IJCAI*. 2564–2570.

[24] Feiping Nie, Xiaoqian Wang, Cheng Deng, and Heng Huang. 2017. Learning a structured optimal bipartite graph for co-clustering. In *Advances in Neural Information Processing Systems*. 4129–4138.

[25] Zhenwen Ren, Quansen Sun, and Dong Wei. 2021. Multiple kernel clustering with kernel k-means coupled graph tensor learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 9411–9418.

[26] Zhenwen Ren, Simon X Yang, Quansen Sun, and Tao Wang. 2020. Consensus affinity graph learning for multiple kernel clustering. *IEEE Transactions on Cybernetics* 51, 6 (2020), 3273–3284.

[27] Weixiang Shao, Lifang He, and Philip S Yu. 2015. Multiple incomplete views clustering via weighted nonnegative matrix factorization with L2,1 regularization. In *Joint European conference on machine learning and knowledge discovery in databases*. Springer, 318–334.

[28] Yuan Sun, Dezhong Peng, Haixiao Huang, and Zhenwen Ren. 2022. Feature and semantic views consensus hashing for image set classification. In *Proceedings of the 30th ACM International conference on multimedia*. 2097–2105.

[29] Yuan Sun, Zhenwen Ren, Peng Hu, Dezhong Peng, and Xu Wang. 2023. Hierarchical Consensus Hashing for Cross-Modal Retrieval. *IEEE Transactions on Multimedia* (2023), 1–14. https://doi.org/10.1109/TMM.2023.3272169

[30] Yuan Sun, Xu Wang, Dezhong Peng, Zhenwen Ren, and Xiaobo Shen. 2023. Hierarchical hashing learning for image set classification. *IEEE Transactions on Image Processing* 32 (2023), 1732–1744.

[31] Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu, Jiyuan Liu, Sihang Zhou, and En Zhu. 2022. Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9776–9785.

[32] Siwei Wang, Xinwang Liu, Suyuan Liu, Jiaqi Jin, Wenxuan Tu, Xinzhong Zhu, and En Zhu. 2022. Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences. *arXiv preprint arXiv:2205.15075* (2022).

[33] Jie Wen, Huijie Sun, Lunke Fei, Jinxing Li, Zheng Zhang, and Bob Zhang. 2021. Consensus guided incomplete multi-view spectral clustering. *Neural Networks* 133 (2021), 207–219.

[34] Jie Wen, Yong Xu, and Hong Liu. 2020. Incomplete Multiview Spectral Clustering With Adaptive Graph Learning. *IEEE Transactions on Systems, Man, and Cybernetics* 50, 4 (2020), 1418–1429.

[35] Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang, and Jinxing Li. 2022. A Survey on Incomplete Multiview Clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2022).

[36] Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei, and Hong Liu. 2019. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 5393–5400.

[37] Jie Wen, Zheng Zhang, Zhao Zhang, Lunke Fei, and Meng Wang. 2020. Generalized incomplete multiview clustering with flexible locality structure diffusion. *IEEE transactions on cybernetics* 51, 1 (2020), 101–114.

[38] Jie Wen, Zheng Zhang, Zhao Zhang, Lei Zhu, Lunke Fei, Bob Zhang, and Yong Xu. 2021. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In *Proc. of the 35th AAAI Conference on Artificial Intelligence, Online: AAAI Press*. 10273–10281.

[39] Wei Xia, Quanxue Gao, Qianqian Wang, Xinbo Gao, Chris Ding, and Dacheng Tao. 2022. Tensorized bipartite graph learning for multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[40] Yuan Xie, Dacheng Tao, Wensheng Zhang, Yan Liu, Lei Zhang, and Yanyun Qu. 2018. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *International Journal of Computer Vision* 126, 11 (2018), 1157–1179.

[41] Weiqing Yan, Jindong Xu, Jinglei Liu, Guanghui Yue, and Chang Tang. 2022. Bipartite graph-based discriminative feature learning for multi-view clustering. In *Proceedings of the 30th ACM International Conference on Multimedia*. 3403–3411.

[42] Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. 2022. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 1 (2022), 1055–1069.

[43] Xihong Yang, Yue Liu, Sihang Zhou, Xinwang Liu, and En Zhu. 2022. Mixed Graph Contrastive Network for Semi-Supervised Node Classification. *arXiv preprint arXiv:2206.02796* (2022).

[44] Xihong Yang, Yue Liu, Sihang Zhou, Siwei Wang, Xinwang Liu, and En Zhu. 2022. Contrastive Deep Graph Clustering with Learnable Augmentation. *arXiv preprint arXiv:2212.03559* (2022).

[45] Xihong Yang, Yue Liu, Sihang Zhou, Siwei Wang, Wenxuan Tu, Qun Zheng, Xinwang Liu, Liming Fang, and En Zhu. 2023. Cluster-guided Contrastive Graph Clustering Network. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. 10834–10842.

[46] Junpu Zhang, Liang Li, Siwei Wang, Jiyuan Liu, Yue Liu, Xinwang Liu, and En Zhu. 2022. Multiple Kernel Clustering with Dual Noise Minimization. In *Proceedings of the 30th ACM International Conference on Multimedia*. 3440–3450.