# Feature Space Recovery for Efficient Incomplete Multi-view Clustering

Zhen Long, Ce Zhu, Pierre Comon, Yazhou Ren, Yipeng Liu

# Feature Space Recovery for Efficient Incomplete Multi-view Clustering

Zhen Long, Ce Zhu, *Fellow, IEEE*, Pierre Comon, *Fellow, IEEE*, Yazhou Ren, *Member, IEEE*, Yipeng Liu, *Senior Member, IEEE*

**Abstract**—T-SVD based incomplete multi-view clustering (IMVC) has received wide attention due to its ability to capture high-order correlations. However, t-SVD suffers from rotation sensitivity, failing to fully explore both inter- and intra-view consistencies. Besides, current methods mainly consider inter- or intra-view correlations, ignoring the low-rank information of sample features within views. To address these weaknesses, we first propose a feature space recovery based IMVC (FSR-IMVC) method, where low-rank feature space recovery and low-rank tensor ring based consistency learning are considered into a unified framework. Furthermore, we extend FSR-IMVC by incorporating anchor learning on the latent feature space, resulting in a scalable FSR-IMVC (sFSR-IMVC) approach that is well-suited to large-scale data. In an iterative way, the learned inter- and intra-view correlations will guide the recovery of missing features, while the explored low-rank information from feature spaces will in turn facilitate consistency exploration, eventually achieving outstanding clustering performance. Experimental results show that FSR-IMVC provides a significant improvement over known state-of-the-art algorithms in terms of ACC, NMI and Purity. Compared with FSR-IMVC, sFSR-IMVC performs slightly worse in clustering accuracy, but offers a notable advantage in computational efficiency, particularly for large-scale datasets. The codes of FSR-IMVC and sFSR-IMVC are publicly available at https://github.com/longzhen520/sFSR-IMVC.

**Index Terms**—Incomplete multi-view clustering, Low-rank tensor ring approximation, Feature space recovery, Anchor learning

## 1 Introduction

Advances in information techniques have revolutionized data analysis in real-world applications, allowing for the description of observed data from multiple views. For example, in image processing, images captured from diverse modalities, such as RGB, depth, and infrared, are commonly available for analysis [1], [2]. In social network analysis, interactions between individuals can be derived from multiple sources such as emails, phones, or social media [3], [4]. The availability of multi-view data, which provides consensual and complementary information, has spurred the development of various multi-view learning-based tasks [5]–[9]. The aim of these tasks is to utilize complementary information from different views to enhance the performance of models. Among these tasks, multi-view clustering (MVC) aims to group multi-view data into several clusters by integrating information from different views. It finds applications in various fields including image processing, computer vision, and social network analysis [10]–[14].

In practical scenarios, it is common for certain views to contain missing or incomplete samples due to factors such as technical limitations and privacy concerns during data acquisition or transmission [15]–[18], [18]. Directly applying current MVC methods to such data can result in poor performance, as the hidden information of missing samples is ignored. To address the aforementioned problem, various incomplete multi-view clustering (IMVC) methods, including deep IMVC and incomplete multi-view subspace clustering (IMSC), have been proposed [19]–[22]. Unlike deep IMVC techniques [17], [23], [24], which require big databases and huge computing resources to learn embedded features from incomplete multi-view datasets for clustering, the IMSC methods easily learn the extraction of diverse graphs that effectively indicate the memberships among samples from different views [25]–[30]. For example, Wen et al. [25] proposed a unified framework that simultaneously integrates graph learning and spectral clustering to obtain a consensus representation for IMVC. However, the work by Wen et al. [25] has two limitations: 1) it only considers relationships among observed samples in consensus representation learning, and 2) the learning of each view is performed separately without incorporating inter-view similarity structure, which significantly diminishes the advantages of multi-view data.

Building on this observation, several tensor-based incomplete multi-view subspace clustering methods have been proposed to infer missing samples while performing clustering [26]–[30]. For instance, Xia et al. [27] proposed a tensor completion-based incomplete multi-view clustering (TCIMC) method, which employs tensor Schatten $p$-norm minimization on multi-view self-representations to infer missing instances and explore inter-view similarity. Li et al. [28] proposed a high-order correlation preserved incomplete multi-view subspace clustering (HCP-IMSC) framework, incorporating tensor factorization and hypergraph-induced hyper-Laplacian regularization to explore high-order correlations and recover the missing instances, respectively. Li et al. [29] developed a tensor-based multi-view block-diagonal structure diffusion (TMBSD) for clus-

*Z. Long, C. Zhu, and Y. Liu are with the School of Communication and Information Engineering, and Y. Ren is with the School of Computer Science and Engineering, all at the University of Electronic Science and Technology of China in Chengdu, 611731, China (email: eczhu@uestc.edu.cn; yipengliu@uestc.edu.cn)*
*P. Comon is with Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France.*

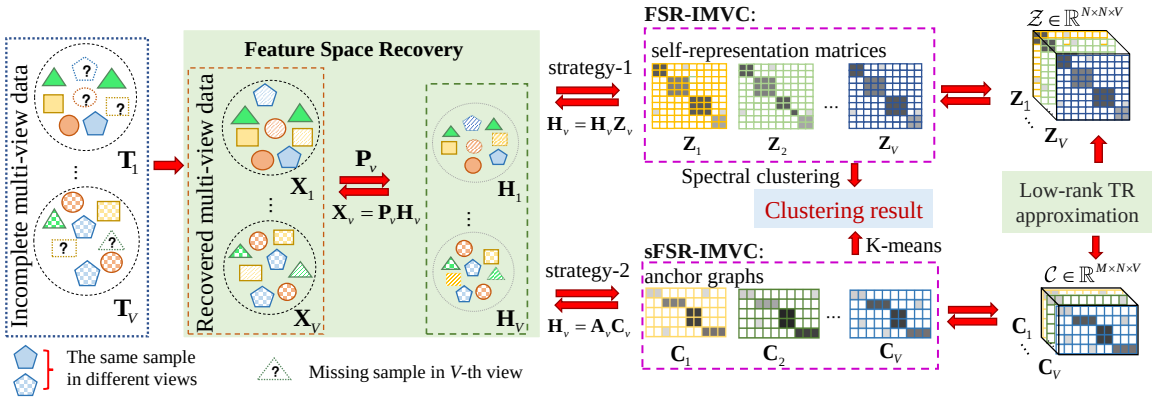



Fig. 1. The pipeline of FSR-IMVC and sFSR-IMVC. Our methods contain two learning objectives, i.e., feature space recovery within each view and consistent representation within and across views. Specifically, the within-view feature space recovery aims at learning the low-rank structure inherent in incomplete multi-view data to mitigate the impact of missing samples. The consistent representation within and across views is learned by the low-rank TR approximation.

tering incomplete multi-view data, which utilizes tensor nuclear norm on the spectral embedding matrices of multi-view data to recover the missing instances and obtain a better consensus representation. Wen et al. [30] proposed an incomplete multi-view tensor spectral clustering with missing-view inferring (IMVTSC-MVI) method, which incorporates Frobenius-norm on the feature space and low-rank tensor constraint on the multi-view self-representations into a unified framework for inferring and clustering. These tensor-based methods all use the so-called tensor singular value decomposition (t-SVD) [31] on the rotated self-representation tensor to explore the high-order correlations among different views.

However, the aforementioned t-SVD based IMVC methods may suffer from the following two limitations:
1)The above works can well explore the correlations across different views, but they suffer from inadequate exploration of intra-view information. Because t-SVD only performs matrix SVDs in the first two modes and linear transformations in the third mode [32], [33].
2) The existing approaches only consider the correlations of samples across or/and within views to infer missing features, ignoring the structural correlations of the features themselves. In fact, the feature spaces within views of multi-view data are highly redundant [34].

Different from t-SVD, tensor ring (TR) decomposition [35] can provide a more flexible and expressive representation to capture the global information [36]–[39]. Therefore, we first consider low-rank TR approximation on the self-representation tensor to capture its global low-rank structure, ensuring that the consistency within each view and across different views is learned well. Besides, concerning the low-rank nature of feature spaces within each view, we integrate the low-rank feature space recovery into the TR-based self-representation learning framework and develop a feature space recovery-based incomplete multi-view clustering framework (FSR-IMVC) as shown in Fig. 1 (Strategy 1) [40]. Finally, the self-representation tensor is used to construct the affinity matrix for the spectral clustering algorithm [41]. FSR-IMVC has shown significant improvements in MVC tasks. Nevertheless, FSR-IMVC requires constructing the membership graph (representation tensor $N \times N \times V$), which involves computing inverses of $V$ matrices of size $N \times N$ with a computational complexity of $O(VN^3)$ per iteration, where $V$ and $N$ denote the numbers of views and samples, respectively. This approach can be inefficient for large-scale data where $N$ becomes large, and such scalability is crucial for real-world applications.

Recently, anchor learning based methods are often employed in large-scale MVC tasks [42]–[45], where an anchor graph of size $M \times N$ is constructed between $M$ anchors and $N$ samples ($M \ll N$) to represent the complex mechanisms of multi-view data, resulting in reduced computational complexity from $O(VN^3)$ to $O(VNM^2)$. Following this, Wang et al. [46] proposed an incomplete large-scale multi-view clustering approach based on the consensus bipartite graph (IMVC-CBG) framework, which integrates anchor selection and anchor graph construction into a unified framework, where all samples share the same anchors and anchor graph to ensure structural consistency across views, enabling fast IMVC tasks. In contrast to view-shared anchors, Liu et al. [47] proposed a fast IMVC with view-independent anchors (FIMVC-VIA) method. It learns individual anchors for each view and constructs a unified anchor graph to tackle large-scale IMVC tasks. However, these anchor learning-based IMVC methods aim at exploring the pairwise correlations between different views, which limits their ability to fully exploit the relationship among views. Besides, they all neglect the feature space information of missing samples.

Motivated by these findings, we propose an improvement of FSR-IMVC with anchor learning, namely scalable FSR-IMVC (sFSR-IMVC), as illustrated in Fig. 1 (Strategy 2). When the number of samples $N$ is large, anchor learning is applied to the latent feature subspaces $\mathbf{H}_v$ to learn the anchor graph $\mathbf{C}_v \in \mathbb{R}^{M \times N}$ for each view $v = 1, \cdots, V$. Furthermore, low-rank TR approximation is utilized to simultaneously explore the inter-view and intra-view information of the tensor $\mathcal{C} = \Omega(\mathbf{C}_1, \cdots, \mathbf{C}_V) \in \mathbb{R}^{M \times N \times V}$, $\Omega$ being an operator which stacks all anchor graphs to a 3-rd order tensor. Similar to FSR-IMVC, the latent feature subspace $\mathbf{H}_v$ simultaneously learns the correlations of inter/intra-view and sample features through the updated $\mathbf{C}_v$ and the low-rank structure of $\mathbf{X}_v$, for recovering the feature space $\mathbf{X}_v$. In turn, the anchor graph tensor $\mathcal{C}$ is adaptively updated

TABLE 1 Comparison of our method with the others mentioned. C #1– #5 denote whether the corresponding approach is scalable, can recover missing samples, can explore high-order correlation across views, can capture the consistency within and across views, and can explore feature space information, respectively.

| Algorithms | Cluster ways | C #1 | C #2 | C #3 | C #4 | C #5 |
|---|---|---|---|---|---|---|
| BSV [20] | K-means | ✓ | × | × | × | × |
| Concat [20] | K-means | ✓ | × | × | × | × |
| OPIMC [19] | Weighted matrix factorization | ✓ | × | × | × | × |
| HCPIMSC [28] | t-SVD-based subspace clustering | × | ✓ | ✓ | × | ✓ |
| TMBSD [29] | t-SVD-based embedding feature learning and K-means | × | ✓ | ✓ | × | × |
| IMVTSC-MVI [30] | t-SVD-based subspace clustering | × | ✓ | ✓ | ✓ | × |
| TCIMC [27] | t-SVD-based subspace clustering | × | × | ✓ | × | × |
| FIMVC-VIA [47] | anchor learning and K-means | ✓ | × | × | × | × |
| IMVC-CBG [46] | anchor learning and K-means | ✓ | × | × | × | × |
| FSR-IMVC [40] | TR-based subspace clustering | × | ✓ | ✓ | ✓ | ✓ |
| sFSR-IMVC | TR-based anchor learning and K-means | ✓ | ✓ | ✓ | ✓ | ✓ |

from $\mathbf{H}_v, v = 1, \cdots, V$ and the low-rank TR approximation to obtain a better consistent representation. Finally, the anchor graph tensor is used to construct the embedded space for the K-means algorithm [48]. Experimental results on small multi-view datasets demonstrate that FSR-IMVC significantly improves the clustering performance compared to state-of-the-art algorithms in terms of ACC, NMI, and Purity. Furthermore, sFSR-IMVC also shows superior performance compared to fast IMVC methods on three large-scale multi-view datasets.

For a clear comparison, we summarize the aforementioned works in Table 1. Compared to them, our work mainly makes contributions in three aspects:

- We first develop an FSR-IMVC model, which integrates low-rank feature space recovery and low-rank TR-based self-representation learning into a unified framework.
  **1).** This approach applies tensor networks for the first time to simultaneously explore memberships across and within views, achieving high-precision clustering performance on small incomplete multi-view data.
  **2).** The FSR-IMVC model represents the first attempt to consider low-rank feature space recovery for IMVC. Experimental results on relatively small multi-view datasets demonstrate its superior performance in comparison to existing methods, as evaluated using three commonly used metrics.
- Based on FSR-IMVC, we extend the approach to deal with large-scale multi-view data and propose an sFSR-IMVC model, where TR-based anchor learning is incorporated to efficiently learn the consistency within and across views and significantly reduce computational complexity. Furthermore, experimental results on large-scale multi-view datasets demonstrate that sFSR-IMVC outperforms the compared fast IMVC methods.

The rest of this paper is organized as follows. Section 2 provides a brief introduction to the notations and related work. Model development, solutions, and complexity analysis are presented in Section 3. Experimental results and a comprehensive analysis are presented in Section 4. Finally, conclusions are drawn in Section 5.

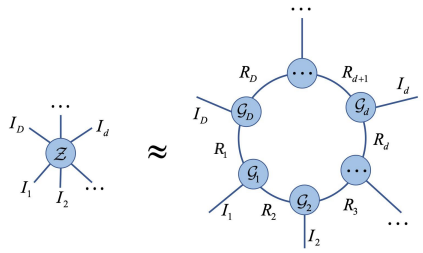


Fig. 2. The graphical illustration for TR decomposition of a $D$th-order tensor.

## 2 Notations and Preliminaries

### 2.1 Notations

In this paper, a scalar, a vector, a matrix, and a tensor are written as $x$, $\mathbf{x}$, $\mathbf{X}$, and $\mathcal{X}$, respectively. For a $D$-th order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$, its $(i_1, \cdots, i_D)$-th element is denoted as $\mathcal{X}(i_1, \cdots, i_D)$. Indices typically range from 1 to their capital version, e.g., $i_d = 1, \cdots, I_d$, $d = 1, \cdots, D$.

**Definition 1** (TR Decomposition [35]). *The goal of TR decomposition is to represent a higher-order tensor by circularly multiplying sequences of 3rd-order tensors. Specifically, the TR decomposition of a $D$th-order tensor, $\mathcal{Z} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$, is defined as*

$$\mathcal{Z}(i_1, i_2, \cdots, i_D) = \mathrm{trace}(\mathcal{G}_1(:, i_1, :) \cdots \mathcal{G}_D(:, i_D, :)),$$

*where $\mathcal{G}_d \in \mathbb{R}^{R_d \times I_d \times R_{d+1}}$, $d = 1, \cdots, D$ are TR core factors, and $\{R_d\}_{d=1}^{D}$ are TR ranks, with $R_{D+1} = R_1$. The TR decomposition can be abbreviated as $\mathcal{Z} = \mathfrak{R}(\mathcal{G}_1, \cdots, \mathcal{G}_D)$. The graphical illustration of TR decomposition is shown in Fig. 2.*

**Lemma 1.** *[49] $\|\mathbf{X}\|_*$ is the tightest convex lower bound of the $\mathrm{rank}(\mathbf{X})$ on the set $\{\mathbf{X} : \|\mathbf{X}\| \leq 1\}$, where $\|\mathbf{X}\|_*$ denotes the nuclear norm of $\mathbf{X}$, which is the sum of its singular values and $\|\mathbf{X}\|$ is the dual norm of the nuclear norm.*

Let $\mathbf{P}$ and $\mathbf{H}$ be arbitrary matrices with compatible sizes such that $\mathbf{X} = \mathbf{PH}$ exists. Then we have the following.

**Lemma 2.** *[50]* $\|\mathbf{X}\|_* = \min_{\mathbf{X}=\mathbf{PH}} \frac{1}{2}(\|\mathbf{P}\|_{\mathrm{F}}^2 + \|\mathbf{H}\|_{\mathrm{F}}^2)$.

**Corollary 1.** *From Lemma 2, and because $\mathrm{trace}(\mathbf{I})$ is constant, we obtain: $\|\mathbf{X}\|_* = \min_{\mathbf{X}=\mathbf{PH}} \frac{1}{2}\|\mathbf{H}\|_{\mathrm{F}}^2$ with $\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}$.*

Therefore, we consider $\min_{\mathbf{X}=\mathbf{PH}} \frac{1}{2}\|\mathbf{H}\|_{\mathrm{F}}^2$ to explore the low-rank information of $\mathbf{X}$ since the optimization problem of rank minimization is NP-hard [51].

### 2.2 Related works

Two well-known frameworks for incomplete multi-view clustering are introduced in this section.

#### 2.2.1 Incomplete multi-view subspace clustering

The subspace-based IMVC framework aims to identify multiple low-dimensional subspaces where samples are expected to lie. Given incomplete multi-view dataset $\mathbf{T}_v \in \mathbb{R}^{D_v \times N}$, $v = 1, \cdots, V$, where $V$ and $N$ are the numbers of views and samples, respectively; and $D_v$ represents the feature dimension in the $v$-th view, the general subspace-based IMVC optimization model can be formulated as:

$$\min_{\{\mathbf{Z}_v, \mathbf{E}_v\}_{v=1}^{V}} \sum_{v=1}^{V} \Psi(\mathbf{E}_v) + \lambda \Phi(\mathbf{Z}_v) \text{ s. t. } \mathbf{T}_v \mathbf{O}_v = \mathbf{T}_v \mathbf{O}_v \mathbf{Z}_v + \mathbf{E}_v, \tag{1}$$

where $\mathbf{O}_v \in \mathbb{R}^{N\times N}$ is a diagonal index matrix of $v$-th view, which is defined as:

$$\mathbf{O}_v(n,n) = \begin{cases} 1 & n\text{-th sample is observed, } n = 1, \cdots, N \\ 0 & \text{otherwise,} \end{cases}$$

$\Phi$ and $\Psi$ depict the regularization term on the self-representation matrices and corrupted noise, respectively. $\lambda$ is a trade-off parameter to balance the effect of the $\Phi$ and $\Psi$ regulations. For instance, Wen et al. [25] consider the nuclear norm and $\ell_1$ norm on $\mathbf{Z}_v$ and $\mathbf{E}_v$ respectively to learn the intra-view information.

Furthermore, benefiting from well-capturing high-order correlations among views, tensor self-representation based IMVC methods have attracted much attention [26]–[30], which is formulated as:

$$\begin{aligned} &\min_{\{\mathbf{E}_v,\mathbf{Z}_v,\mathbf{X}_v\}_{v=1}^V} \Phi(\mathcal{Z}) + \lambda\Psi(\mathbf{E}) \\ &\text{s. t.} \mathbf{X}_v = \mathbf{X}_v\mathbf{Z}_v + \mathbf{E}_v, (\mathbf{X}_v)_{\mathbb{O}_v} = (\mathbf{T}_v)_{\mathbb{O}_v}, v = 1, \cdots, V, \\ &\quad \mathcal{Z} = \Omega(\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_V), \mathbf{E} = [\mathbf{E}_1; \mathbf{E}_2; \cdots; \mathbf{E}_V], \end{aligned} \quad (2)$$

where $\mathbb{O}_v$ is the index set of observed samples within $v$-th view, the term $\Omega(\cdot)$ merges all self-representation matrices $\mathbf{Z}_v (v = 1, \cdots, V)$ to a third-order tensor $\mathcal{Z} \in \mathbb{R}^{N\times N\times V}$, $\Phi$ depicts a low-rank tensor approximation term to explore the relationship of observed samples. Besides, instead of directly learning self-representation on multi-view data, another option for IMVC to improve its clustering performance is to learn self-representation on its latent feature space [52]. In the above tensor self-representation based IMVC methods, t-SVD [53] is a commonly used tensor approximation method to capture high-order correlations among views for IMVC tasks. However, t-SVD fails to simultaneously capture both inter-view and intra-view correlations. Moreover, the above methods neglect the important low-rank property of the feature space, which is crucial for incomplete multi-view clustering.

Inspired by them, we first propose a feature space recovery-based IMVC, as follows [40]:

$$\begin{aligned} &\min_{\{\mathbf{E}_v,\mathbf{Z}_v,\mathbf{P}_v,\mathbf{H}_v,\mathbf{X}_v\}_{v=1}^V} \Phi(\mathcal{Z}) + \sum_{v=1}^V \frac{\gamma}{2}\|\mathbf{H}_v\|_{\mathrm{F}}^2 + \lambda\|\mathbf{E}_v\|_1 \\ &\text{s. t. } \mathbf{X}_v = \mathbf{P}_v\mathbf{H}_v + \mathbf{E}_v^x, \mathbf{H}_v = \mathbf{H}_v\mathbf{Z}_v + \mathbf{E}_v^h, \\ &\quad \mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v = \mathbf{I}_K, (\mathbf{X}_v)_{\mathbb{O}_v} = (\mathbf{T}_v)_{\mathbb{O}_v}, v = 1, \cdots, V. \end{aligned} \quad (3)$$

According to Corollary 1,

$$\min_{\{\mathbf{H}_v,\mathbf{P}_v\}_{v=1}^V} \frac{\gamma}{2}\|\mathbf{H}_v\|_{\mathrm{F}}^2, \text{s. t. } \mathbf{X}_v = \mathbf{P}_v\mathbf{H}_v + \mathbf{E}_v^x, \mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v = \mathbf{I}_K$$

depicts the low-rank property of the recovered feature spaces $\mathbf{X}_v$, $\mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v = \mathbf{I}_K$ means the feature spaces are projected onto discriminating subspaces [54]; $\mathbf{E}_v = [\mathbf{E}_v^x; \mathbf{E}_v^h]$; and $\mathbf{H}_v = \mathbf{H}_v\mathbf{Z}_v + \mathbf{E}_v^h$ means self-representation matrix is learned on the latent feature spaces. Moreover, low-rank TR approximation is considered to explore the similarity structure of the self-representation tensor, e.g., $\mathcal{Z} = \mathfrak{R}(\mathcal{G}_1, \cdots, \mathcal{G}_D)$ denotes $\Phi(\mathcal{Z})$. Benefiting from TR approximation, the inter/intra-view information in $\mathcal{Z}$ can be well captured simultaneously.

#### 2.2.2 Anchor Learning for IMVC

In general, subspace-based methods can better extract geometric information from samples for IMVC. However, subspace learning-based methods often involve computationally expensive operations, such as eigenvalue computation, singular value decomposition, and matrix inverse, which may not be well-suited to large-scale data. To tackle this problem, anchor learning is considered to deal with large-scale datasets for IMVC [42]–[45]. The general framework of an anchor-learning based incomplete multi-view clustering is constructed as follows:

$$\begin{aligned} &\min_{\{\mathbf{C}_v\}_{v=1}^V} \sum_{v=1}^V \|\mathbf{T}_v\mathbf{O}_v - \mathbf{A}_v\mathbf{C}_v\mathbf{O}_v\|_{\mathrm{F}}^2 + \lambda\Phi(\mathbf{C}_v) \\ &\text{s. t. } \mathbf{A}_v^{\mathrm{T}}\mathbf{A}_v = \mathbf{I}_M, \mathbf{C}_v \geq 0, v = 1, \cdots, V \end{aligned} \quad (4)$$

where $\mathbf{A}_v \in \mathbb{R}^{D_v\times M}$ is the learned anchor matrix in the $v$-th view and $M$ is the number of anchors. $\mathbf{C}_v \in \mathbb{R}^{M\times N}$ is the anchor graph, which is used to describe the relationship among samples. Usually, $M \ll N$, which largely reduces the computational complexity from $O(VN^3)$ to $O(VNM^2)$. In addition, $\Phi(\mathbf{C}_v)$ represents the regularization terms. For example, Liu et al. [47] consider view-specific anchor matrix $\mathbf{A}_v$ and consistent anchor graph $\mathbf{C}$ and use Frobenius norm on $\mathbf{C}$ for better clustering performance. Following it, Wang et al. [46] learn anchors and consistent representation in the latent feature space.

## 3 Proposed Method

### 3.1 Model Development

As shown in Fig. 1 (strategy 2), we learn anchors and anchor graphs on the latent feature space $\mathbf{H}_v, v = 1, \cdots, V$ for efficient incomplete multi-view clustering, where feature space recovery and inter/intra-view correlations are learned in a unified framework. The proposed modeling can be formulated as the optimization problem:

$$\begin{aligned} &\min_{\{\mathbf{E}_v,\mathbf{C}_v,\mathbf{P}_v,\mathbf{H}_v,\mathbf{X}_v\}_{v=1}^V} \sum_{v=1}^V \frac{\gamma}{2}\|\mathbf{H}_v\|_{\mathrm{F}}^2 + \lambda\|\mathbf{E}_v\|_1 \\ &\text{s. t. } \mathbf{X}_v = \mathbf{P}_v\mathbf{H}_v + \mathbf{E}_v^x, \mathbf{H}_v = \mathbf{A}_v\mathbf{C}_v + \mathbf{E}_v^h, \mathbf{C}_v \geq 0, \\ &\quad \mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v = \mathbf{I}_K, (\mathbf{X}_v)_{\mathbb{O}_v} = (\mathbf{T}_v)_{\mathbb{O}_v}, v = 1, \cdots, V, \\ &\quad \mathcal{C} = \mathfrak{R}(\mathcal{G}_1, \cdots, \mathcal{G}_D), \mathbf{A}_v^{\mathrm{T}}\mathbf{A}_v = \mathbf{I}_M \end{aligned} \quad (5)$$

where $\mathcal{C} = \Omega(\mathbf{C}_1, \cdots, \mathbf{C}_V)$; $\Omega$ is an operator which stacks all anchor graphs $\mathbf{C}_v, v = 1, \cdots, V$ in a 3-rd order tensor $\mathcal{C} \in \mathbb{R}^{M\times N\times V}$; $\mathbf{A}_v \in \mathbb{R}^{K\times M}$ is the anchor matrix for the $v$-th views, the $\ell_1$ norm is employed on error matrix $\mathbf{E}_v$ to remove sparse noise or outliers; and $\mathbb{O}_v$ is the index set of observed samples for $v$-th view; $\mathcal{C} = \mathfrak{R}(\mathcal{G}_1, \cdots, \mathcal{G}_D)$ represents a low-rank TR approximation of $\mathcal{C}$, which aims to explore the similarity structure within and across views, simultaneously.

According to Fig. 1 (strategy 2), the observed data $\mathbf{T}_v$ can be reconstructed by assuming its feature space is low-rank, denoted as $\mathbf{X}_v = \mathbf{P}_v\mathbf{H}_v$, where $\mathbf{P}_v^T\mathbf{P}_v = \mathbf{I}_K$ and $\|\mathbf{H}_v\|_{\mathrm{F}}$ minimization is applied to enforce the low-rank property of $\mathbf{X}_v$. The anchor graph tensor is then updated adaptively from $\mathbf{H}_v, v = 1, \cdots, V$, and low-rank TR approximation.

$\mathbf{H}_v$ in turn learns from the updated $\mathbf{C}_v$ and the low-rank prior of the feature space to recover $\mathbf{X}_v$, ultimately improving the clustering performance.

### 3.2 Solutions

The above optimization problem can be addressed using an alternating direction of multiplier method (ADMM) framework [55]. To make problem (5) separable, $\mathcal{Y}$ is added as an auxiliary variable as follows:

$$\begin{aligned}&\min_{\{\mathbf{E}_v,\mathbf{C}_v,\mathbf{P}_v,\mathbf{H}_v,\mathbf{X}_v\}_{v=1}^V,\mathcal{Y}} \sum_{v=1}^{V}\frac{\gamma}{2}\|\mathbf{H}_v\|_{\mathrm{F}}^2+\lambda\|\mathbf{E}_v\|_1\\&\text{s. t. } \mathbf{X}_v=\mathbf{P}_v\mathbf{H}_v+\mathbf{E}_v^x, \mathbf{H}_v=\mathbf{A}_v\mathbf{C}_v+\mathbf{E}_v^h, \mathbf{C}_v\geq 0,\\&\mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v=\mathbf{I}_K, (\mathbf{X}_v)_{\mathbb{O}_v}=(\mathbf{T}_v)_{\mathbb{O}_v}, v=1,\cdots,V,\\&\mathcal{C}=\mathcal{Y}, \mathcal{Y}=\mathfrak{R}(\mathcal{G}_1,\cdots,\mathcal{G}_D), \mathbf{A}_v^{\mathrm{T}}\mathbf{A}_v=\mathbf{I}_M.\end{aligned}\tag{6}$$

The corresponding augmented Lagrangian function is formulated as:

$$\begin{aligned}&\mathrm{L}\left(\{\mathbf{C}_v,\mathbf{E}_v,\mathbf{Q}_1^v,\mathbf{Q}_2^v,\mathbf{X}_v,\mathbf{P}_v,\mathbf{H}_v\}_{v=1}^V,\mathcal{Y},\mathcal{Q}_3\right)\\&=\sum_{v=1}^{V}(\frac{\gamma}{2}\|\mathbf{H}_v\|_{\mathrm{F}}^2+\lambda\|\mathbf{E}_v\|_1\\&+\langle\mathbf{Q}_1^v,\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v-\mathbf{E}_v^x\rangle+\langle\mathbf{Q}_2^v,\mathbf{H}_v-\mathbf{A}_v\mathbf{C}_v-\mathbf{E}_v^h\rangle\\&+\frac{\rho_1}{2}\|\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v-\mathbf{E}_v^x\|_{\mathrm{F}}^2+\frac{\rho_2}{2}\|\mathbf{H}_v-\mathbf{A}_v\mathbf{C}_v-\mathbf{E}_v^h\|_{\mathrm{F}}^2)\\&+\langle\mathcal{Q}_3,\mathcal{C}-\mathcal{Y}\rangle+\frac{\rho_3}{2}\|\mathcal{C}-\mathcal{Y}\|_{\mathrm{F}}^2,\end{aligned}\tag{7}$$

under constraints $\mathbf{C}_v\geq 0$, $\mathcal{Y}=\mathfrak{R}(\mathcal{G}_1,\cdots,\mathcal{G}_D)$, $(\mathbf{X}_v)_{\mathbb{O}_v}=(\mathbf{T}_v)_{\mathbb{O}_v}$ and $\mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v=\mathbf{I}_K, \mathbf{A}_v^{\mathrm{T}}\mathbf{A}_v=\mathbf{I}_M, v=1,\cdots,V$, where $\{\mathbf{Q}_1^v,\mathbf{Q}_2^v\}_{v=1}^V$ and $\mathcal{Q}_3$ are Lagrange multipliers and $\rho_1,\rho_2,\rho_3$ are penalty factors. Using an ADMM framework, which alternately updates one variable with others fixed, problem (7) is split into several subproblems.

**Update $\{\mathbf{P}_v\}_{v=1}^V$:** Fixing other variables, the subproblem of $\mathbf{P}_v$ is rewritten as:

$$\max_{\mathbf{P}_v:\mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v=\mathbf{I}_K}\quad \mathrm{trace}(\mathbf{P}_v(\mathbf{H}_v(\mathbf{Q}_1^v+\rho_1\mathbf{X}_v-\rho_1\mathbf{E}_v^x)^{\mathrm{T}})).\tag{8}$$

This subproblem is a well-known orthogonal Procrustes problem. Letting $\mathbf{M}=\mathbf{H}_v(\mathbf{Q}_1^v+\rho_1\mathbf{X}_v-\rho_1\mathbf{E}_v^x)^{\mathrm{T}}$, and $[\mathbf{S},\mathbf{V},\mathbf{D}]=\mathrm{svd}(\mathbf{M})$, the solution of $\mathbf{P}_v$ is $\mathbf{P}_v=\mathbf{D}\mathbf{S}^{\mathrm{T}}$, where svd is the Singular Value Decomposition.

**Update $\{\mathbf{H}_v\}_{v=1}^V$:** Fixing other variables, the subproblem of $\mathbf{H}_v$ can be rewritten as:

$$\begin{aligned}\min_{\mathbf{H}_v}\quad&\frac{\gamma}{2}\|\mathbf{H}_v\|_{\mathrm{F}}^2\\&+\langle\mathbf{Q}_1^v,\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v-\mathbf{E}_v^x\rangle+\langle\mathbf{Q}_2^v,\mathbf{H}_v-\mathbf{A}_v\mathbf{C}_v-\mathbf{E}_v^h\rangle\\&+\frac{\rho_1}{2}\|\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v-\mathbf{E}_v^x\|_{\mathrm{F}}^2+\frac{\rho_2}{2}\|\mathbf{H}_v-\mathbf{A}_v\mathbf{C}_v-\mathbf{E}_v^h\|_{\mathrm{F}}^2).\end{aligned}\tag{9}$$

Differentiating the formula (9) and setting it to zero, we can obtain

$$\mathbf{H}_v=\frac{\mathbf{P}_v^{\mathrm{T}}(\mathbf{Q}_1^v+\rho_1\mathbf{X}_v-\rho_1\mathbf{E}_v^x)+\rho_2(\mathbf{A}_v\mathbf{C}_v+\mathbf{E}_v^h)-\mathbf{Q}_2^v}{\gamma+\rho_1+\rho_2}\tag{10}$$

**Update $\{\mathbf{X}_v\}_{v=1}^V$:** The subproblem of $\mathbf{X}_v$ is:

$$\min_{\mathbf{X}_v}\langle\mathbf{Q}_1^v,\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v-\mathbf{E}_v^x\rangle+\frac{\rho_1}{2}\|\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v-\mathbf{E}_v^x\|_{\mathrm{F}}^2\tag{11}$$

under constraints $(\mathbf{X}_v)_{\mathbb{O}_v}=(\mathbf{T}_v)_{\mathbb{O}_v}$. Similarly, by differentiating the formula (11) and setting it to zero, we can obtain the following:

$$\mathbf{X}_v(:,n)=\begin{cases}\hat{\mathbf{X}}_v(:,n), n\notin\mathbb{O}_v\\\mathbf{T}_v(:,n), n\in\mathbb{O}_v,\end{cases}\tag{12}$$

where $\hat{\mathbf{X}}_v=\mathbf{P}_v\mathbf{H}_v+\mathbf{E}_v^x-(1/\rho_1)\mathbf{Q}_1^v$.

**Update $\{\mathbf{A}_v\}_{v=1}^V$:** The subproblem of solving $\mathbf{A}_v$ is formulated as:

$$\max_{\mathbf{A}_v:\mathbf{A}_v^{\mathrm{T}}\mathbf{A}_v=\mathbf{I}_M}\quad \mathrm{trace}(\mathbf{A}_v(\mathbf{C}_v(\mathbf{Q}_2^v+\rho_2\mathbf{H}_v-\rho_2\mathbf{E}_v^h)^{\mathrm{T}})).\tag{13}$$

Following the same framework of solving $\mathbf{P}_v$, the updating of $\mathbf{A}_v$ is $\mathbf{A}_v=\mathbf{D}\mathbf{S}^{\mathrm{T}}$, where $\mathbf{D}$ and $\mathbf{S}$ are the right and left singular matrices of $\mathbf{M}$, respectively; $\mathbf{M}=\mathbf{C}_v(\mathbf{Q}_2^v+\rho_2\mathbf{H}_v-\rho_2\mathbf{E}_v^h)^{\mathrm{T}}$.

**Update $\{\mathbf{C}_v\}_{v=1}^V$:** By setting the derivative of the objective function (7) w.r.t. $\mathbf{C}_v$ to zero , the solution of $\mathbf{C}_v$ can be obtained via:

$$\mathbf{C}_v=\max(\hat{\mathbf{C}}_v,0),\tag{14}$$

where

$$\hat{\mathbf{C}}_v=\frac{\mathbf{A}_v^{\mathrm{T}}(\mathbf{Q}_2^v+\rho_2\mathbf{H}_v-\rho_2\mathbf{E}_v^h)+\Omega_v^{-1}(\rho_3\mathcal{Y}+\mathcal{Q}_3)}{\rho_2+\rho_3},$$

$\Omega_v^{-1}$ is the inverse operator along $v$-th view, e.g., $\Omega_v^{-1}(\mathcal{C})=\mathbf{C}_v$.

**Update $\{\mathbf{E}_v\}_{v=1}^V$:** The solution of $\mathbf{E}_v$ can be split into two parts and updated by

$$\begin{cases}\mathbf{E}_v^x=\mathrm{sth}(\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v+(1/\rho_1)\mathbf{Q}_1^v,\lambda/\rho_1)\\\mathbf{E}_v^h=\mathrm{sth}(\mathbf{H}_v-\mathbf{A}_v\mathbf{C}_v+(1/\rho_2)\mathbf{Q}_2^v,\lambda/\rho_2),\end{cases}\tag{15}$$

where $\mathrm{sth}(x,\tau)$ is the well-known soft thresholding operator, denoted as: $\mathrm{sth}(x,\tau)=\mathrm{sgn}(x)\max(|x|-\tau,0)$.

**Update $\mathcal{Y}$**: For $\mathcal{Y}$, the problem (7) can be transformed into the following formulation:

$$\begin{aligned}\min_{\mathcal{Y}}\quad&\|\mathcal{Y}-(\mathcal{C}-(1/\rho_3)\mathcal{Q}_3)\|_{\mathrm{F}}^2,\\\text{s. t. }&\mathcal{Y}=\mathfrak{R}(\mathcal{G}_1,\cdots,\mathcal{G}_D).\end{aligned}\tag{16}$$

It can be solved by TR-ALS algorithm in [35] , where the input tensor is $\mathcal{C}-(1/\rho_3)\mathcal{Q}_3$.

**Update Lagrangian multipliers**:

$$\begin{cases}\mathbf{Q}_1^v=\mathbf{Q}_1^v+\rho_1(\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v-\mathbf{E}_v^x), v=1,\cdots,V\\\mathbf{Q}_2^v=\mathbf{Q}_2^v+\rho_2(\mathbf{H}_v-\mathbf{A}_v\mathbf{C}_v-\mathbf{E}_v^h), v=1,\cdots,V\\\mathcal{Q}_3=\mathcal{Q}_3+\rho_3(\mathcal{Y}-\mathcal{C}).\end{cases}\tag{17}$$

Finally, the right singular vectors of $\mathbf{S}=\frac{1}{V}\sum_{v=1}^V\mathbf{C}_v^{\mathrm{T}}$ will be used in the K-means algorithm for the final clustering result [46], [47]. The approach is summarized in Algorithm 1. The convergence condition is reached when $\min(\text{RE-X}, \text{RE-H}, \text{RE-C}, \text{RSE})\leq 10^{-6}$, where RE-X, RE-H, RE-C indicate the maximum variation of recovered feature spaces, latent feature subspaces, anchor graphs between adjacent iterations. For instance, RE-X=$\max_v\|\mathbf{X}_v^t-\mathbf{X}_v^{t-1}\|_\infty$, where $\mathbf{X}_v^t$ is the recovered data of $v$-th view in $t$-th iteration. RSE denotes the maximum relative square error between the recovered feature space $\hat{\mathbf{X}}_v^t$ and the original space $\mathbf{T}_v$ in the observed indices, e.g., RSE $=\max_v\frac{\|\hat{\mathbf{X}}_v^t(\mathbf{o}_v)-\mathbf{T}_v(\mathbf{o}_v)\|_{\mathrm{F}}}{\|\mathbf{T}_v(\mathbf{o}_v)\|_{\mathrm{F}}}$, where $\hat{\mathbf{X}}_v^t$ is the recovered feature space without the constraint $(\mathbf{X}_v)_{\mathbb{O}_v}=(\mathbf{T}_v)_{\mathbb{O}_v}$ in $t$-th iteration; $\mathbf{o}_v$ is the index vector of observed entries of $v$-th view.

**Algorithm 1** Scale feature space recovery based incomplete multi-view clustering (sFSR-IMVC)

**Input:** incomplete multi-view data $\{\mathbf{T}_v\}_{v=1}^V$, the number of anchors $M$,$\gamma$, TR ranks $R$;
**Initialize:** $\mathcal{Y}$=$\mathcal{Q}_3$=0; $\mathbf{P}_v$=**0**, $\mathbf{H}_v$=**0**, $\mathbf{E}_v^x$=**0**, $\mathbf{A}_v$=**0**, $\mathbf{C}_v$=**0**, $\mathbf{E}_v^h$=**0**, $v$=$1,\cdots,V$; $\rho_1$=$\rho_2$=$\rho_3$=$10^{-4}$; $\varepsilon$=$10^{-7}$; $\eta$=1.8; $K$=$B+1$, $B$ is the number of clusters.
**while** not converged **do**
  **for** $v = 1,\cdots,V$ **do**
    Update $\mathbf{P}_v$ via equation (8)
    Update $\mathbf{H}_v$ via equation (9)
    Update $\mathbf{X}_v$ via equation (12)
    Update $\mathbf{A}_v$ via equation (13)
    Update $\mathbf{C}_v$ via equation (14)
    Update $\mathbf{E}_v$ via equation (15)
    Update $\mathbf{Q}_1^v$ and $\mathbf{Q}_2^v$ via equation (17)
  **end for**
  Update $\mathcal{Y}$ via equation (16)
  Update $\mathcal{Q}_3$ via equation (17)
  $\rho_1 = \min\left(\eta\rho_1, 10^{11}\right)$, $\rho_2 = \min\left(\eta\rho_2, 10^{11}\right)$;
**end while**
$\mathbf{S} = \frac{1}{V}\sum_{v=1}^V \mathbf{C}_v^{\mathrm{T}}$;
Apply the K-means algorithm on the right singular vectors of $\mathbf{S}$.
**Output:** Clustering result

TABLE 2 The complexity analysis of FSR-IMVC/ sFSR-IMVC for one iteration, where $D_v$, $N$, and $V$ are the dimension of features in $v$-th view, the number of samples, and views, respectively. Note that the notation '-' indicates no mentioned variable in the method.

| Variables | Storage Complexity | Computational Complexity |
|---|---|---|
| $\mathbf{P}_v$ | $D_vK/D_vK$ | $O(K^2D_v)/O(K^2D_v)$ |
| $\mathbf{H}_v$ | $KN/KN$ | $O(N^3)/O(KD_vN)$ |
| $\mathbf{X}_v$ | $D_vN/D_vN$ | $O(KD_vN)/O(KD_vN)$ |
| $\mathbf{Z}_v$ | $N^2$/- | $O(N^3)/-$ |
| $\mathbf{C}_v$ | -/$MN$ | -/$O(KMN)$ |
| $\mathbf{A}_v$ | -/$KM$ | -/$O(M^2K)$ |
| $\mathbf{E}_v^x$ | $D_vN/D_vN$ | $O(D_vKN)/O(D_vKN)$ |
| $\mathbf{E}_v^h$ | $KN/KN$ | $O(KN^2)/O(KMN)$ |
| $\mathcal{Y}$ | $N^2V/MNV$ | $O(R^2N^2V^2)/O(R^2MNV^2)$ |

### 3.3 Complexity Analysis

Table 2 summarizes the storage and computational complexity of FSR-IMVC [40] and sFSR-IMVC. It can be observed that the main storage complexity of both FSR-IMVC and sFSR-IMVC comes from variable $\mathcal{Y}$. Additionally, the main computational complexity of FSR-IMVC stems from the updates of $\mathbf{Z}_v$ and $\mathbf{H}_v$, requiring $O(N^3)$. Conversely, sFSR-IMVC exhibits a lower computational complexity of $O(N)$, making it scalable for large-scale data.

## 4 Numerical Experiments

### 4.1 Experimental Settings

#### 4.1.1 Multi-view Datasets

Eight well-known multi-view datasets were chosen to evaluate the effectiveness of our method, namely:

TABLE 3 Statistics of different multi-view datasets.

| Datasets | #Sample | #View | #Cluster | #Feature | $(I_1,\cdots,I_4)$ |
|---|---|---|---|---|---|
| Yale | 165 | 3 | 15 | 3304,4096,6750 | (M,11,15,3) |
| Coil20 | 1440 | 3 | 20 | 4096,3304,6750 | (M,40,36,3) |
| BDGP | 2500 | 4 | 5 | 1000,500,250,79 | (M,50,50,4) |
| Scene15 | 4485 | 3 | 15 | 1800, 1180, 1240 | (M,65,69,3) |
| CCV | 6700 | 3 | 20 | 20,20,20 | (M,67,100,3) |
| Caltech-all | 9120 | 5 | 102 | 48,40,254,521,928 | (M,114,80,4) |
| ALOI | 10800 | 4 | 100 | 77,13,64,125 | (M,108,100,4) |
| Reuters | 18750 | 5 | 6 | 21531,24892,34251, 15506,11547 | (M,150,125,5) |

**Yale**[1] contains 165 samples belonging to 15 clusters, with 11 samples per subject. We utilized three views, namely LBP (view1), intensity (view2), and Gabor (view3), as the features. **Coil20** [56] consists of 1400 image samples from 20 clusters, with three views: intensity (view1), LBP (view2), and Gabor (view3). **BDGP** [57] comprises 2500 samples from 5 classes, with four views: lateral (view1), dorsal (view2), ventral (view3), and texture (view4). **Scene15** [58] contains 4485 image samples with 15 scene categories from both indoor and outdoor environments. We used three common image features: GIST (view1), PHOG (view2), and LBP (view3). **Columbia Consumer Video (CCV)** [59] comprises 6773 web videos belonging to 20 semantic categories. It utilizes three well-known audio/visual features, namely SIFT (view1), STIP (view2), and MFCC (view3). We removed the last 73 samples in our experiment, resulting in 6700 samples. **Caltech-all** [60] consists of 9144 samples categorized into 102 classes. It employs five different features, namely Gabor (view1), Wavelet-moments (view2), Cenhist (view3), GIST (view4), and LBP (view5). We removed the last 24 samples in our experiments to obtain 9120 samples. **Amsterdam Library of Object Images (ALOI)** [61] is a collection of 110250 images of 1000 small objects captured under various lighting conditions and rotation angles. Following the approach in [62], we used the first 100 classes as our testing dataset and obtained 10800 samples with four views, including color similarity (view1), haralick features (view2), RGB color histograms (view3), and HSV color histograms (view4). **Retuers** [63] comprises 111740 articles with feature characteristics of documents written in 5 different languages and their translations, covering a standard set of 6 categories. In our experiments, we utilized a subset of Reuters consisting of 18750 samples with five views, including English (view1), French (view2), German (view3), Italian (view4), and Spanish (view5).

The statistical information of the above datasets is summarized in Table 3. The anchor graph tensor of size $M\times N\times V$ is rearranged into a 4th-order with $M\times A\times B\times V$, where $N = AB$ to facilitate low-rank TR approximation. $A$ and $B$ are balanced as closely as possible, as shown in the last column of Table 3. This balancing is crucial for effective low-rank TR approximation, which has proven effective for approximating high-order data [37]–[39]. Furthermore, due to computational limitations, CCV, Caltech-all, ALOI, and Reuters were only applied to the scalable IMVC methods.

#### 4.1.2 Missing Data Construction and Evaluation Metrics

To generate incomplete multi-view data, we randomly remove $P$ samples for each view, ensuring that all samples have at least one view. The missing ratio (MR) is MR = $P/N$

1. http://vision.ucsd.edu/content/yale-face-database

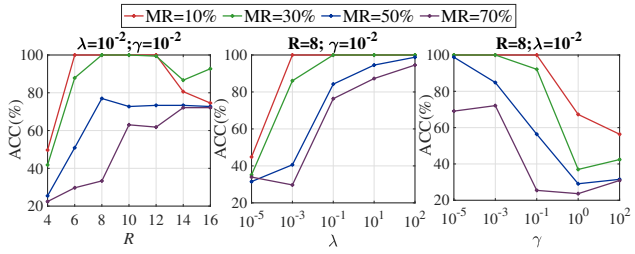



Fig. 3. The change of ACC as parameters $\gamma$, $R$ and $\lambda$ vary in FSR-IMVC on Yale dataset with different MRs.

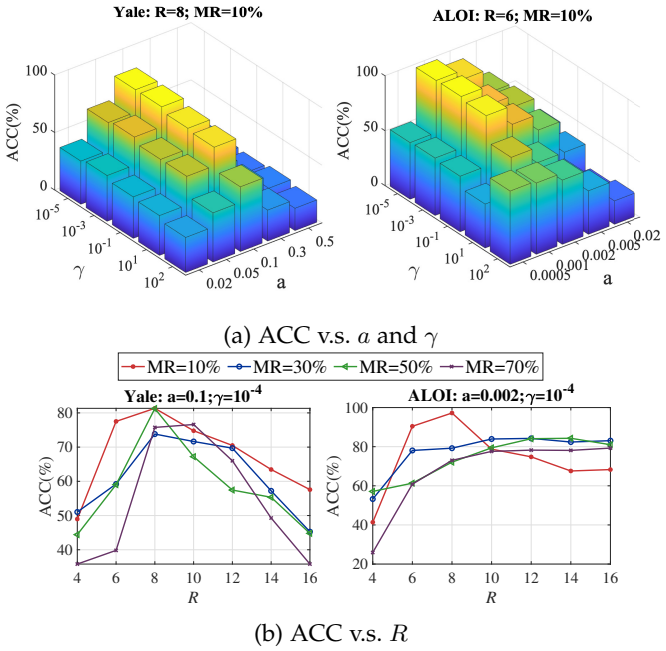
for each view, where $N$ is the total number of samples. MR is varied at values of 10%, 30%, 50%, and 70%. Additionally, normalized mutual information (NMI), accuracy (ACC), purity, and CPU time are used to evaluate the performance in this experiment [20]. Larger values of these metrics, except for CPU time, indicate better clustering performance. Each experiment is repeated 10 times, and the statistical results of these metrics are reported. **Note** that instead of repeating the final clustering method (K-means or spectral clustering) 10 times, we re-randomly generate missing sets while keeping MR unchanged and repeatedly test the performance of the clustering algorithm.

#### 4.1.3 Compared Clustering Algorithms

Nine IMVC methods, including (best single view) BSV, Concat, one-pass incomplete multi-view clustering (OPIMC) (OPIMC) [AAAI, 2019] [19], HCPIMSC [IEEE TIP, 2022] [28], TMBSD [ICME, 2021] [29], IMVTSC-MVI[AAAI, 2021] [30], TCIMC [IEEE TC, 2022] [27], FIMVC-VIA [IEEE TNNLS,2022] [47], and IMVC-CBG [CVPR, 2022] [46] are selected to compare clustering performance. Among them, BSV achieves the best K-means clustering results across all single views, while Concat reports the clustering results obtained by stacking all views [20]. HCPIMSC, TMBSD, IMVTSC-MVI, and TCIMC are tensor-based IMVC methods. IMVC-CBG and FIMVC-VIA are fast IMVC methods, all of which consider anchor learning. In our experiments, each method is tuned to perform optimally on different datasets. All experiments are conducted on a desktop computer with 2.4 GHz Quad-Core Intel Core i5 Processor and 16 GB 2133 MHz LPDDR3 Memory.

#### 4.1.4 Parameter settings

For FSR-IMVC, there are three groups of parameters $\lambda$, $\gamma$ and TR ranks $R_d$=$R, d$=$1, \cdots, 4$ need to be tuned. We tune the parameters via brute force search, where $R$ varies in [4−16], $\gamma$ and $\lambda$ vary in $[10^{-5} - 10^{2}]$, respectively. Fig. 3 reports the change of ACC along $R$, $\lambda$, and $\gamma$ on Yale with different MRs. It can be observed that the performance of FSR-IMVC is related to the MR. As the MR increases, the FSR-IMVC method will require a higher $R$, a smaller $\gamma$, and a larger $\lambda$ to achieve the best clustering performance. This implies that as the MR increases, the performance of low-rank feature space recovery will decrease. Simultaneously, more noise will be introduced in the self-representation learning process, leading to a more complex self-representation tensor. According to this strategy, we choose $R$=8, $\lambda$=100, $\gamma$=$10^{-5}$ for Yale; $R$=6, $\lambda$=10, $\gamma$=$10^{-5}$ for Coil20; $R$=6, $\lambda$=1, $\gamma$=$10^{-5}$ for BDGP; $R$=4, $\lambda$=1, $\gamma$=$10^{-5}$ for Scene15 in the experiments.




(a) ACC v.s. $a$ and $\gamma$

(b) ACC v.s. $R$

Fig. 4. The change of ACC as parameters $\gamma$, $R$ and $a$ vary in sFSR-IMVC on Yale and ALOI datasets.

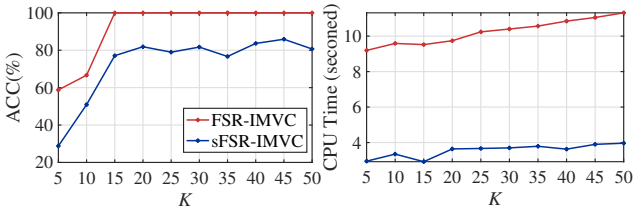



Fig. 5. The clustering performance of FSR-IMVC and sFSR-IMVC on the Yale dataset with respect to $K$ when MR=10%, $R = 8$, $\lambda = 100$ and $\gamma = 10^{-5}$.

Contrary to FSR-IMVC, sFSR-IMVC has groups of parameters: TR ranks ($R_d$, $d$=$1, \cdots, 4$), trade-off parameters ($\lambda$ and $\gamma$), and the number of anchors ($M$=$\lfloor aN \rfloor$, where $\lfloor . \rfloor$ represents the flooring function and $a$ is an anchor ratio). To simplify, we set $\lambda$=1 and $R_d$=$R, d$=$1, \cdots, 4$ for all multi-view datasets. $a$ depends on the dataset size, ranging from [0.02, 0.05, 0.1, 0.3, 0.5 ] for Yale and Coil20, and [0.0005, 0.001, 0.002, 0.005, 0.02] for BDGP, Scene15, CCV, Caltech-all, ALOI, and Reuters. Fig. 4a reports the change of ACC along $\gamma$ and $a$ on Yale and ALOI datasets with 10% MR. From it, the performance on Yale performs well when $a$=0.1 and $\gamma \in [10^{-5} - 10^{-1}]$. Instead, for ALOI, our method performs stably when $\gamma \in [10^{-4} - 10^{-1}]$ and $a$=0.001.

This is because the number of anchors ($M$) needed for optimal performance varies depending on the dataset size ($N$). For example, with $a = 0.001$, 10 anchors ($\lfloor 0.001 \times 10800 \rfloor$ ) are sufficient for representing the relationships among samples in ALOI, while 16 anchors ($\lfloor 0.1 \times 165 \rfloor$) are needed for Yale. In addition, we fix $a$=0.1 and $\gamma$=$10^{-4}$ for Yale and $a$=0.002 and $\gamma$=$10^{-3}$ for ALOI to choose the parameter $R$ in [4−16]. As shown in Fig. 4b, we can observe the clustering performance first increases and then decreases

TABLE 4 Comparison results(ACC(%), average(variance)) using all methods on four small multi-view datasets as MR ranges from 10% to 70%.

| Datasets | MR (%) | BSV | Concat | OPIMC | TMBSD | HCP-IMSC | IMVTSC-MVI | TCIMC | FIMVC-VIA | IMVC-CBG | FSR-IMVC | sFSR-IMVC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Yale | 10.0 | 41.85( 2.12) | 35.03( 9.85) | 38.12( 3.20) | 66.24( 4.33) | 72.00( 0.75) | 70.30( 0.81) | 61.33( 4.31) | 59.39( 4.05) | 50.61( 1.96) | **100.00**( 0.00) | 77.64( 3.55) |
| | 30.0 | 31.65( 1.90) | 21.70( 3.34) | 33.70( 4.45) | 61.94( 4.29) | 71.64( 3.19) | 67.09( 5.09) | 61.03( 5.22) | 56.24( 4.71) | 50.79( 2.47) | **99.58**( 1.34) | 77.95( 4.00) |
| | 50.0 | 23.56( 1.65) | 21.39( 2.89) | 28.48( 2.66) | 60.67( 5.55) | 60.91( 5.04) | 39.15( 2.34) | 55.88( 5.80) | 55.27( 5.17) | 52.73( 3.14) | **97.64**( 3.63) | 71.62( 2.71) |
| | 70.0 | 23.16( 1.62) | 22.06( 3.35) | 28.91( 2.12) | 40.61( 3.48) | 37.94( 3.15) | 30.79( 4.11) | 39.03( 3.58) | 57.58( 3.92) | 51.45( 3.92) | **83.09**( 8.45) | 60.96( 2.27) |
| Coil20 | 10.0 | 53.40( 0.69) | 58.68( 5.47) | 58.24( 3.65) | 80.34( 3.68) | 70.75( 2.01) | **88.03**( 0.20) | 84.55( 0.72) | 71.50( 3.28) | 57.40( 1.80) | 85.85( 4.16) | 83.70( 2.17) |
| | 30.0 | 41.40( 0.95) | 46.41( 4.58) | 54.89( 4.46) | 79.58( 4.32) | 73.87( 0.53) | 86.01( 1.57) | 81.20( 2.64) | 69.39( 2.78) | 57.16( 4.73) | **87.03**( 2.44) | 84.00( 2.50) |
| | 50.0 | 31.39( 0.66) | 32.53( 3.00) | 49.12( 4.83) | 77.59( 3.89) | 63.24( 4.46) | 83.21( 1.33) | 79.47( 2.09) | 71.10( 2.28) | 57.03( 5.36) | **86.33**( 3.81) | 81.95( 2.89) |
| | 70.0 | 30.89( 1.26) | 28.17( 2.27) | 31.59( 2.30) | 76.04( 3.44) | 36.34( 0.98) | 49.40( 2.84) | 41.17( 1.10) | 72.48( 2.04) | 57.49( 6.34) | **83.38**( 3.74) | 59.77( 3.75) |
| BDGP | 10.0 | 49.70( 2.23) | 52.85( 11.23) | 74.78( 8.23) | 43.39( 3.83) | 80.38( 1.72) | 99.62( 0.13) | 25.79( 5.56) | 37.48( 0.29) | 52.38( 6.80) | **99.96**( 0.13) | 96.66( 5.97) |
| | 30.0 | 41.18( 1.80) | 43.97( 12.00) | 64.74( 4.60) | 43.21( 4.04) | 79.89( 6.14) | **99.10**( 0.15) | 34.96( 5.14) | 38.46( 0.58) | 61.98( 0.06) | 98.68( 3.54) | 98.03( 1.89) |
| | 50.0 | 34.10( 0.89) | 31.61( 7.18) | 55.88( 6.50) | 39.41( 4.19) | 61.21( 7.12) | 96.70( 0.74) | 30.82( 3.38) | 39.11( 0.81) | 61.97( 0.10) | **100.00**( 0.01) | 98.39( 1.82) |
| | 70.0 | 29.35( 1.00) | 27.67( 3.23) | 37.46( 2.74) | 37.88( 2.63) | 29.40( 2.58) | 83.92( 2.61) | 23.37( 0.89) | 38.49( 3.34) | 47.57( 4.97) | 96.50( 6.04) | **97.17**( 3.49) |
| Scene15 | 10.0 | 39.81( 0.96) | 38.67( 2.87) | 42.40( 2.24) | 70.93( 4.56) | 48.20( 1.34) | 80.17( 2.34) | 50.38( 3.06) | 55.60( 1.28) | 50.43( 1.63) | 88.44( 0.62) | 68.77( 1.96) |
| | 30.0 | 27.54( 0.98) | 29.35( 3.17) | 35.09( 2.49) | 68.75( 4.40) | 37.87( 1.14) | 78.62( 1.95) | 53.52( 0.97) | 56.11( 0.24) | 50.00( 1.02) | **88.34**( 2.37) | 66.67( 1.59) |
| | 50.0 | 20.57( 0.61) | 21.93( 1.27) | 28.27( 2.99) | 53.45( 2.69) | 27.68( 1.89) | 80.05( 1.19) | 46.31( 2.66) | 56.41( 1.07) | 49.62( 1.25) | **81.07**( 2.80) | 62.70( 2.69) |
| | 70.0 | 17.16( 0.70) | 15.98( 1.02) | 18.82( 1.31) | 29.74( 2.46) | 17.78( 1.56) | 56.41( 2.29) | 26.31( 1.23) | **56.98**( 0.96) | 50.71( 1.47) | 51.45( 5.69) | 33.40( 3.46) |

TABLE 5 Comparison results (ACC (%), average (variance)) using fast IMVC methods on large-scale datasets as MR ranges from 10% to 70%.

| methods | 10 (%) | 30 (%) | 50 (%) | 70 (%) |
|---|---|---|---|---|
| CCV ($R$=10,$\gamma$=$10^{-4}$, a=0.002) | | | | |
| BSV | 16.11( 5.66) | 14.23( 5.01) | 12.98( 4.56) | 11.60( 4.08) |
| Concat | 13.46( 4.75) | 11.92( 4.22) | 10.93( 3.85) | 10.36( 3.66) |
| OPIMC | 15.30( 5.39) | 13.93( 4.97) | 12.15( 4.28) | 10.70( 3.79) |
| FIMVC-VIA | 20.26( 7.12) | 20.66( 7.28) | 19.86( 6.99) | 20.17( 7.10) |
| IMVC-CBG | 17.45( 6.14) | 16.91( 5.95) | 16.50( 5.84) | 18.38( 6.46) |
| sFSR-IMVC | **58.09**( 2.69) | **55.25**( 2.19) | **47.22**( 1.42) | **27.22**( 2.04) |
| Caltech-all ($R$=10,$\gamma$=$10^{-4}$, a=0.002) | | | | |
| BSV | 22.87( 0.29) | 19.21( 0.26) | 16.30( 0.21) | 13.48( 0.21) |
| Concat | 20.89( 0.75) | 16.30( 0.40) | 13.15( 0.49) | 11.08( 0.46) |
| OPIMC | 25.73( 1.02) | 23.48( 1.52) | 20.18( 0.86) | 13.76( 0.77) |
| FIMVC-VIA | 26.96( 0.15) | 27.36( 0.34) | 27.78( 0.18) | 27.80( 0.44) |
| IMVC-CBG | 29.02( 0.60) | 29.10( 0.91) | 28.77( 1.01) | 30.33( 0.84) |
| sFSR-IMVC | **50.14**( 0.47) | **48.67**( 0.47) | **41.75**( 0.66) | **41.88**( 0.59) |
| ALOI ($R$=6,$\gamma$=$10^{-3}$, a=0.001) | | | | |
| BSV | 38.17( 0.27) | 29.83( 0.47) | 21.76( 0.33) | 14.07( 0.17) |
| Concat | 38.20( 1.41) | 27.82( 1.03) | 20.27( 0.37) | 16.05( 0.45) |
| OPIMC | 11.83( 1.50) | 8.45( 1.16) | 6.44( 0.61) | 5.28( 0.58) |
| FIMVC-VIA | 65.03( 0.86) | 65.58( 0.91) | 66.30( 0.01) | **62.80**( 0.95) |
| IMVC-CBG | 38.98( 1.31) | 39.85( 0.91) | 38.63( 1.01) | 39.74( 0.88) |
| sFSR-IMVC | **98.94**( 1.14) | **96.32**( 1.69) | **70.77**( 4.36) | 60.85( 7.97) |
| Reuters ($R$=2,$\gamma$=$10^{-3}$, a=0.0001) | | | | |
| BSV | 42.70( 2.60) | 35.36( 1.66) | 29.36( 0.46) | 27.64( 0.25) |
| Concat | 39.96( 4.83) | 33.86( 4.48) | 32.65( 4.08) | 30.85( 2.07) |
| OPIMC | 45.49( 4.36) | 48.58( 4.46) | 48.11( 4.02) | 44.16( 3.95) |
| FIMVC-VIA | 49.71( 3.17) | 48.80( 2.58) | 48.99( 1.92) | 52.10( 0.52) |
| IMVC-CBG | 38.93( 7.73) | 42.78( 5.15) | 37.65( 4.05) | 37.27( 4.57) |
| sFSR-IMVC | **64.39**( 1.51) | **66.76**( 1.61) | **66.43**( 1.88) | **64.33**( 1.10) |

as $R$ grows up. Besides, the smaller the MR, the smaller $R$ to achieve high ACC. According to this strategy, we choose $R$=8, $\gamma$=$10^{-4}$, $a$=0.1 for Yale; $R$=4, $\gamma$=$10^{-3}$, $a$=0.05 for Coil20; $R$=4, $\gamma$=$10^{-5}$, $a$=0.002 for BDGP; $R$=4, $\gamma$=$10^{-3}$, $a$=0.002 for Scene15; $R$=4, $\gamma$=10, $a$=0.002 for CCV; $R$=10, $\gamma$=$10^{-4}$, $a$=0.002 for Caltech-all; $R$=6, $\gamma$=$10^{-3}$, $a$=0.001 for ALOI; $R$=2, $\gamma$=$10^{-3}$, $a$=0.0001 for Reuters in the experiments.

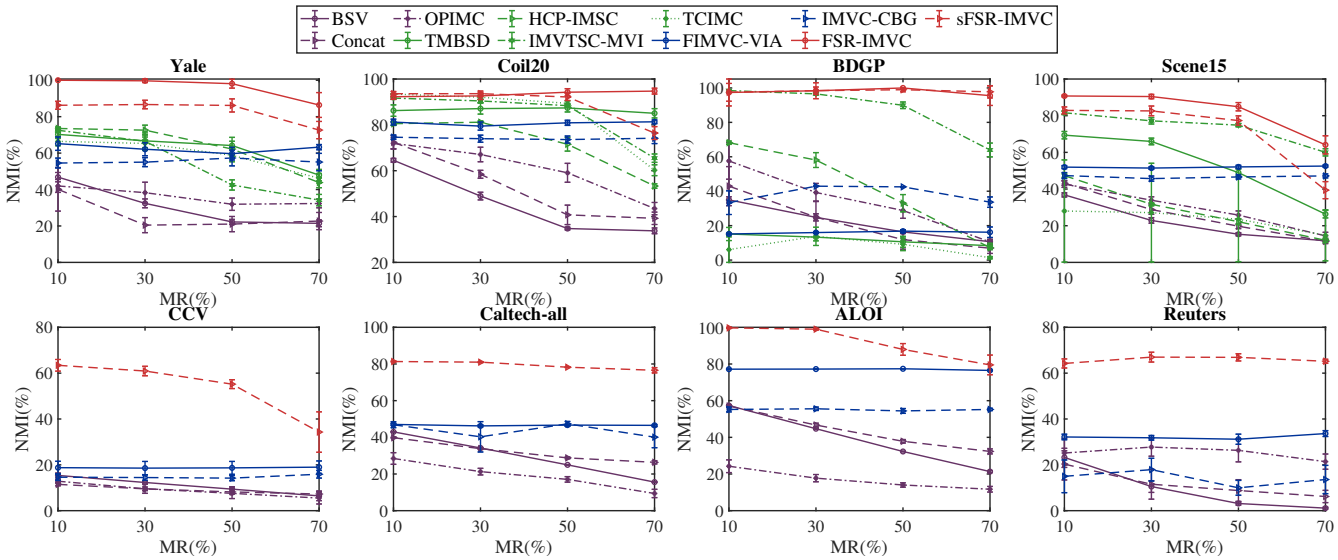
Fig. 5 reports the clustering performance of FSR-IMVC and sFSR-IMVC changes along with the subspace dimension $K$. It can be observed that the ACC first increases as $K$ grows, and performs stable when $K \geq 15$. However, the CPU time increases as K grows. Notably, for the Yale dataset, the number of clusters is 15. Therefore, we set the subspace dimension $K$ as $B+1$ for FSR-IMVC and sFSR-IMVC in the experiments, where $B$ is the number of clusters.

### 4.2 Clustering Performance Analysis

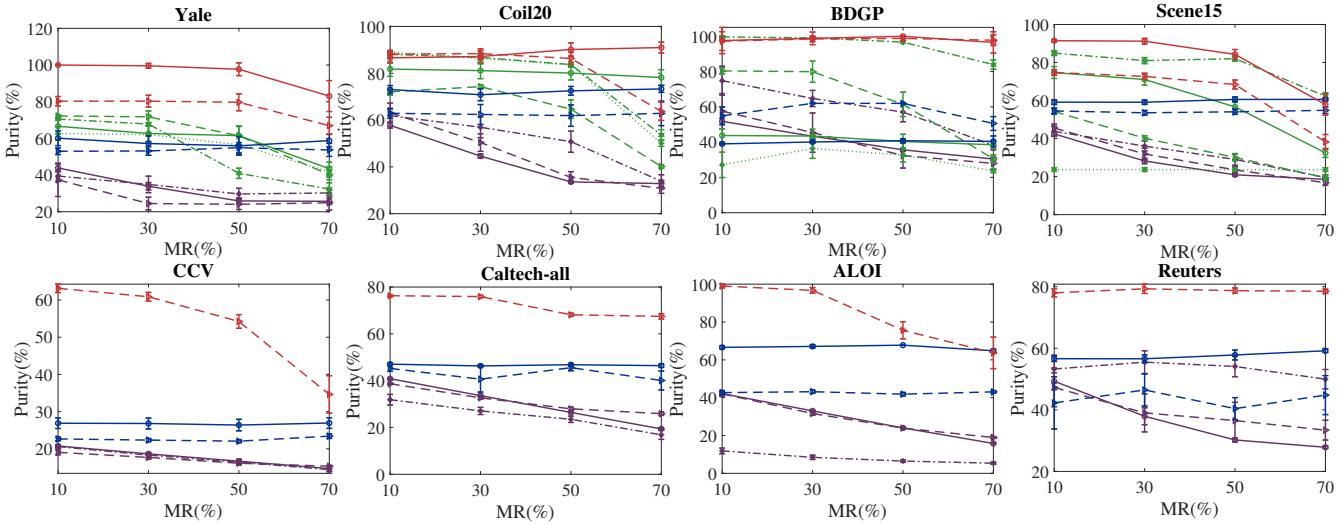
Table 4 presents the clustering performance of all methods on four small multi-view datasets, as measured by ACC, with MR values ranging from 10% to 70%. The best and second best results are written in **bold** and underlined formatting, respectively. In most cases, FSR-IMVC consistently outperforms the state-of-the-art methods on all multi-view datasets and MR values. Specifically, on the Yale dataset, FSR-IMVC demonstrates significant improvements of 28%, 27.94%, and 36.73% compared to HCP-IMSC, respectively, for MR values of 10%, 30%, and 50%. Furthermore, sFSR-IMVC achieves the second-best performance. Notably, in the case of small MRs, sFSR-IMVC exhibits better clustering accuracy than FIMVC-VIA and IMVC-CBG for the anchor-leaning based IMVC methods. For the Coil20, BDGP, and Scene15, tensor-based IMVC methods, including IMVTSC-MVI, FSR-IMVC, and sFSR-IMVC, exhibit higher ACC performance at MR values of 10%, 30%, and 50%.

Table 5 compares performance on large datasets using fast IMVC methods. It can be observed that sFSR-IMVC performs the best in most cases. Specifically, the improvement in clustering performance achieved by sFSR-IMVC on the CCV dataset is notably remarkable. For instance, for MR values of 10% and 30%, the clustering performance of sFSR-IMVC has exhibited notable enhancements of 37.83% and 34.59% respectively, in comparison to FIMVC-VIA. For the Caltech-all, sFSR-IMVC outperforms other methods for all MRs, with improvements of 21.12%, 19.57%, 12.98%, and 11.55%, respectively, compared to IMVC-CBG. For the ALOI, sFSR-IMVC shows a performance advantage of 33.91% and 30.74% over FIMVC-VIA at small MRs such as 10% and 30% , and performs the second best as MR increases. In the case of Reuters, whose size is larger than other large-scale datasets, sFSR-IMVC demonstrates the best performance across all MRs.

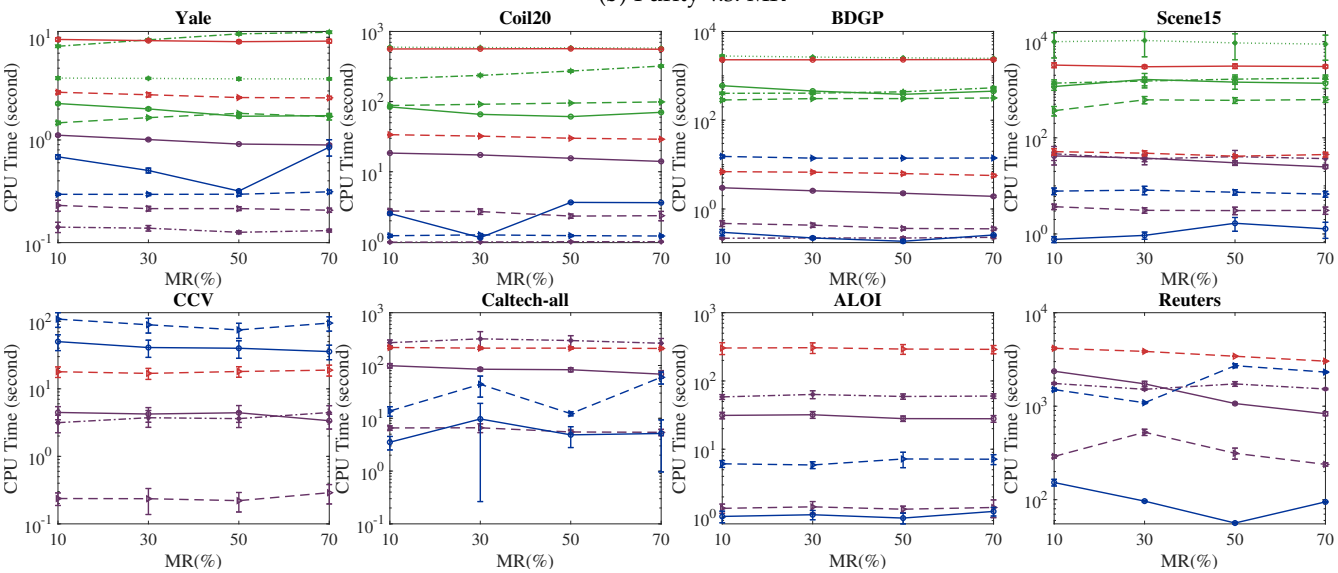
Fig. 6 presents the performance of our methods and several other state-of-the-art algorithms on different datasets in terms of NMI, Purity, and CPU time, as MR ranges from 10% to 70%. It can be observed that FSR-IMVC outperforms other methods in most cases for Yale, COil20, BDGP, and Scene15




(a) NMI v.s. MR




(b) Purity v.s. MR




(c) CPU Time v.s. MR

Fig. 6. Comparison results using different methods and MRs in terms of NMI, Purity, and CPU time.

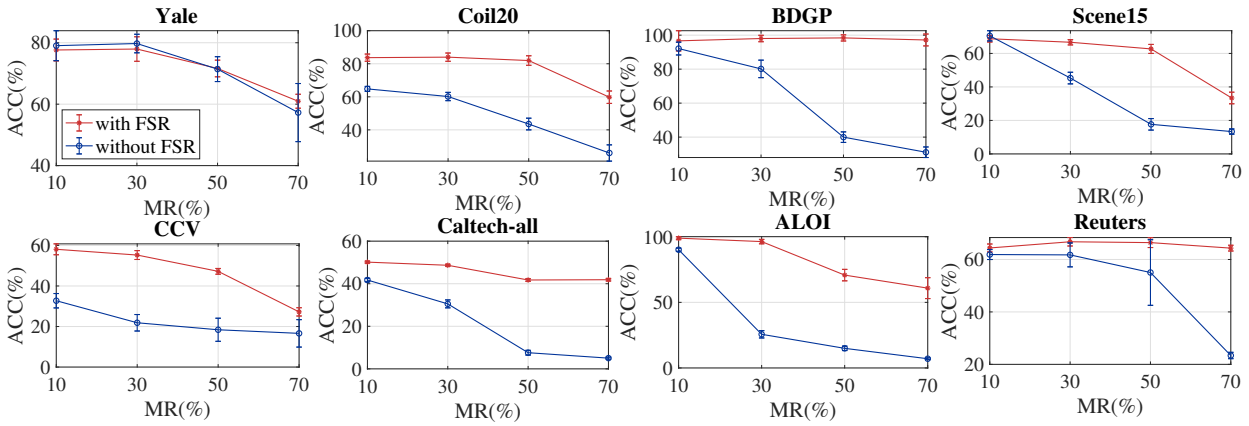



Fig. 7. The clustering performance of sFSR-IMVC on all incomplete multi-view data with/without completion procedure.

TABLE 6 The clustering performance of BSV on incomplete multi-view data/recovered multi-view data (obtained by sFSR-IMVC) varies as the MR changes from 10% to 70%.

| Datasets | ACC | | | | Purity | | | |
|---|---|---|---|---|---|---|---|---|
| | 10 (%) | 30 (%) | 50 (%) | 70 (%) | 10 (%) | 30 (%) | 50 (%) | 70 (%) |
| Yale | 41.85/**49.66** | 31.65/**57.52** | 23.56/**63.05** | 23.16/**60.07** | 43.88/**51.35** | 33.91/**59.25** | 25.90/**64.90** | 25.68/**62.55** |
| Coil20 | 53.40/**64.07** | 41.40/**62.00** | 31.39/**60.17** | 30.89/**56.31** | 57.75/**68.60** | 44.59/**65.45** | 33.58/**62.31** | 32.86/**58.93** |
| BDGP | 49.70/**56.42** | 41.18/**55.95** | 34.10/**60.26** | 29.35/**66.02** | 51.78/**58.76** | 43.36/**57.60** | 35.65/**61.47** | 30.42/**66.53** |
| Scene15 | 39.81/**49.95** | 27.54/**44.81** | 20.57/**39.43** | 17.16/**30.37** | 42.45/**55.02** | 28.22/**48.62** | 20.96/**42.89** | 18.58/**32.51** |
| CCV | 17.59/**23.55** | 16.19/**25.83** | 15.17/**26.71** | 14.09/**25.37** | 18.58/**32.51** | 18.58/**32.51** | 18.58/**32.51** | 18.58/**32.51** |
| Caltech-all | 22.87/**25.55** | 19.21/**27.41** | 16.30/**25.73** | 13.48/**13.92** | 40.85/**44.10** | 33.60/**43.36** | 26.42/**39.94** | 19.40/**20.14** |
| ALOI | 38.17/**42.36** | 29.83/**43.54** | 21.76/**27.90** | 14.07/**15.36** | 42.17/**47.09** | 33.02/**48.81** | 24.01/**31.96** | 15.80/**17.58** |
| Reuters | **42.70**/39.56 | **35.36**/34.37 | 29.36/**31.62** | 27.64/**29.68** | **49.29**/48.22 | 37.91/**43.52** | 30.21/**38.39** | 27.84/**33.53** |

in terms of NMI and Purity. As an extension of FSR-IMVC, the clustering results of sFSR-IMVC are slightly inferior to those of FSR-IMVC. This implies that the similarity matrix constructed from the anchor graph is not as accurate as that from the self-representation matrix. However, sFSR-IMVC significantly reduces the required CPU time compared to FSR-IMVC. Specifically, on Scene15, sFSR-IMVC processes data in only 51.50 seconds, whereas FSR-IMVC takes 3257 seconds, making it nearly 63 times faster. Furthermore, sFSR-IMVC exhibits superior performance on large-scale datasets, particularly on the Caltech-all and Reuters, in terms of NMI and Purity, indicating its effectiveness in large-scale IMVC tasks. Overall, FSR-IMVC and sFSR-IMVC demonstrate significant potential for IMVC tasks.

### 4.3 Model Discussion

#### 4.3.1 Ablation Study

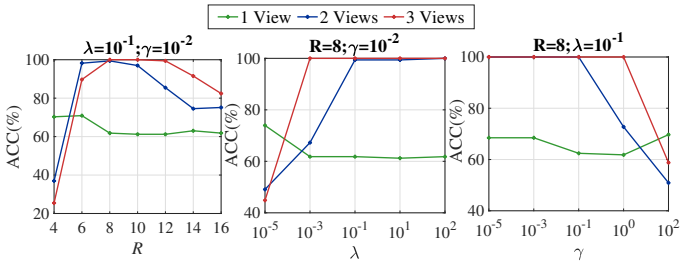
To investigate the usefulness of FSR in sFSR-IMVC, we removed the low-rank completion part in equation (5) and reported the clustering result on all multi-view datasets in Fig. 7. Each parameter was tuned to achieve the best performance. From Fig. 7, two observations can be made:

- Without the FSR component, the proposed method still performs well in the case of MR=10%. For instance, the ACC values on Caltech-all, ALOI, and Reuters are about 42%, 88%, and 60% respectively, which are better than the values obtained by other compared algorithms. It implies that the low-rank TR approximation in sFSR-IMVC can effectively explore the correlations of inter/intra-view, leading to improved clustering performance.
- As MR increases, the ACC of the proposed method decreases significantly without FSR. However, incorporating FSR into the proposed method leads to a significant improvement in terms of ACC, especially when MR=50%, 70%. It implies that FSR can explore more useful information for IMVC from incomplete multi-view data.

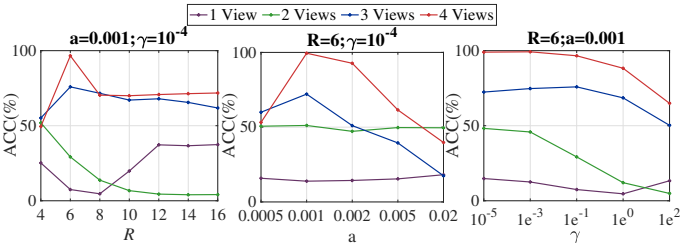
Furthermore, to further validate the benefits of the recovered multi-view data in enhancing clustering, we applied the BSV method to both the incomplete multi-view data and the multi-view data recovered using the sFSR-IMVC. The BSV achieves the best K-means clustering results across all single views. The comparative results are presented in Table 6. It can be observed that, except for the Reuters datasets, the recovered multi-view data demonstrates enhanced clustering performance across all datasets. Notably, when MR is high for the Reuters dataset, the recovered multi-view data also improves the clustering performance.

#### 4.3.2 Impact of the number of views

Fig. 8a illustrates how the clustering performance of FSR-IMVC changes as the number of views varies within the Yale dataset. The Yale dataset consists of three views: LBP, intensity, and Gabor. Here, the term "1 view" refers to the usage of the LBP feature, whereas "2 views" indicates the utilization of both LBP and intensity features. Similarly, Fig. 8a shows the results of sFSR-IMVC on the ALOI datasets. The ALOI dataset consists of four views: color similarity,




(a) ACC v.s. $R$, $\gamma$ and $\lambda$ in FSR-IMVC




(b) ACC v.s. $R$, $\gamma$ and $a$ in sFSR-IMVC

Fig. 8. Comparison between clustering performance of FSR-IMVC and sFSR-IMVC on Yale and ALOI datasets with different numbers of views, respectively.

haralick features, RGB color histograms, and HSV color histograms. Here, the terms "1 view", "2 views", "3 views", and "4 views" respectively refer to the first, the first two, the first three, and the first four features.

The results in the first column of Fig. 8a and Fig. 8b highlight a notable trend: The fewer the number of views, the smaller the TR rank required for achieving optimal clustering performance. This is due to the fact that each view contains both consistent and view-specific information. As the number of views increases, the self-representation tensor/anchor graph tensor becomes more complex. Therefore, in order to capture these correlations, a higher TR rank is required. Furthermore, we can observe that as the number of views increases, our method becomes less sensitive to variations in parameters. For example, when utilizing 3 views, FSR-IMVC achieves the best clustering performance with $\lambda \in [10^{-3} - 10^{2}]$ and $\gamma \in [10^{-5} - 10^{0}]$. In contrast, with 2 views, FSR-IMVC achieves the best clustering outcomes when $\lambda$ falls within the range of $10^{-1}$ to $10^{2}$, and $\gamma$ within the range of $10^{-5}$ to $10^{-1}$.

### 4.3.3 Convergence analysis

Fig. 9 shows the convergence performance of the proposed method on eight multi-view datasets with MR=30%. When $\min(\text{RE-X}, \text{RE-H}, \text{RE-C}, \text{RSE}) \leq 10^{-6}$, our method will stop iterating. It can be observed from the subfigures for RE-X, RE-H, RE-C that the variation of $\mathbf{X}_v$, $\mathbf{H}_v$, $\mathbf{C}_v$ is consistent during the first 20 iterative updates. Meanwhile, the RSE rises during the first few iterations and then stays the same until around the 20th iteration. It implies that in the first about 20 iterations, the inter/intra-view correlation learned with $\mathbf{C}_v$ is passed through $\mathbf{H}_v$ to guide the recovery of missing features in $\mathbf{X}_v$, while $\mathbf{X}_v$ in turn passes the feature space correlation through $\mathbf{H}_v$ to influence $\mathbf{C}_v$'s update, thus benefiting the clustering.

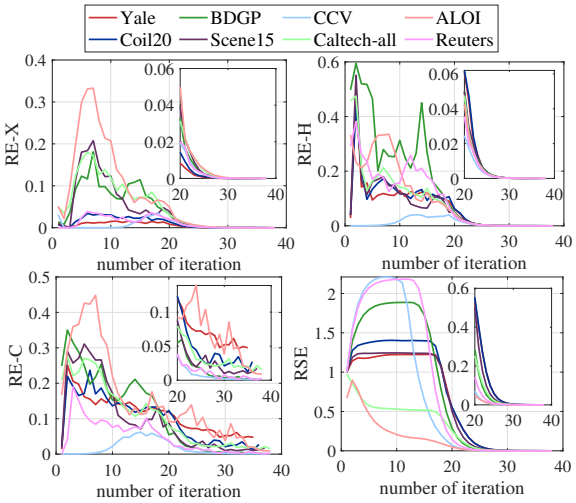



Fig. 9. The convergence performance of sFSR-IMVC on all multi-view datasets with MR=30% in terms of RE-X, RE-H, RE-C and RSE.

## 5 Conclusion

In this paper, we propose an extension to our method FSR-IMVC by incorporating anchor learning, resulting in a more efficient sFSR-IMVC model. The proposed model integrates low-rank matrix learning of the original feature space, anchor learning of the latent feature space, and low-rank TR approximation-based inter/intra-view capturing, within a unified framework for incomplete multi-view clustering. These three processes leverage inter/intra-view and feature space correlations through latent feature spaces and anchor graphs, leading to improved clustering performance. Numerical experiments on five relatively small multi-view datasets with different MRs demonstrate that our FSR-IMVC method outperforms state-of-the-art methods in terms of clustering performance, as measured by ACC, NMI, and Purity. Furthermore, the sFSR-IMVC demonstrates its capability to handle large multi-view datasets effectively and efficiently.

Similar to two-step IMVC methods, the underlying clustering structure extracted from FSR-IMVC/sFSR-IMVC might encounter inaccuracies and could become computationally demanding when applied to large-scale datasets. To address these issues, a promising approach is to combine binary clustering structure learning [64] with tensor-based representation learning from incomplete multi-view data within a unified framework.

## References

[1] F. Wang, J. Pan, S. Xu, and J. Tang, "Learning discriminative cross-modality features for rgb-d saliency detection," *IEEE Trans. Image Process.*, vol. 31, pp. 1285–1297, 2022.

[2] Z. Long, C. Zhu, J. Liu, P. Comon, and Y. Liu, "Trainable subspaces for low rank tensor completion: Model and analysis," *IEEE Trans. Signal Process.*, vol. 70, pp. 2502–2517, 2022.

[3] H. Shuai, C. Shen, D. Yang, Y. Lan, W. Lee, S. Philip, and M. Chen, "A comprehensive study on social network mental disorders detection via online social media mining," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 7, pp. 1212–1225, 2017.

[4] H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, and T. Chua, "Detecting stress based on social interactions in social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1820–1833, 2017.

[5] S. Zheng, X. Cai, C. Ding, F. Nie, and H. Huang, "A closed form solution to multi-view low-rank regression," in *Twenty-Ninth Proc. AAAI Conf. Artif. Intell.*, 2015.

[6] J. Xu, X. Li, W.and Liu, D. Zhang, J. Liu, and J. Han, "Deep embedded complementary and interactive information for multi-view classification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 04, 2020, pp. 6494–6501.

[7] P. Zhang, X. Liu, J. Xiong, S. Zhou, W. Zhao, E. Zhu, and Z. Cai, "Consensus one-step multi-view subspace clustering," *IEEE Trans. Knowl. Data Eng.*, 2020.

[8] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1863–1883, 2018.

[9] G. Chao, S. Sun, and J. Bi, "A survey on multiview clustering," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 146–168, 2021.

[10] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4279–4287.

[11] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 923–931.

[12] J. Xu, Y. Ren, H. Tang, Z. Yang, Y. Pan, L.and Yang, X. Pu, S. Philip, and L. He, "Self-supervised discriminative feature learning for deep multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, 2022.

[13] H. Wang, Y. Yang, and B. Liu, "Gmc: Graph-based multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1116–1129, 2019.

[14] J. Xu, H. Tang, Y. Ren, L. Peng, X. Zhu, and L. He, "Multi-level feature learning for contrastive multi-view clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16 051–16 060.

[15] L. Zhang, Y. Zhao, Z. Zhu, D. Shen, and S. Ji, "Multi-view missing data completion," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 7, pp. 1296–1309, 2018.

[16] Y. Lin, Y. Gou, X. Liu, J. Bai, J. Lv, and X. Peng, "Dual contrastive prediction for incomplete multi-view representation learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022.

[17] J. Xu, C. Li, L. Peng, Y. Ren, X. Shi, H. Shen, and X. Zhu, "Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering," *IEEE Trans. Image Process.*, vol. 32, pp. 1354–1366, 2023.

[18] G. Chao, S. Wang, S. Yang, C. Li, and D. Chu, "Incomplete multi-view clustering with multiple imputation and ensemble clustering," *Applied Intelligence*, vol. 52, no. 13, pp. 14 811–14 821, 2022.

[19] M. Hu and S. Chen, "One-pass incomplete multi-view clustering," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 3838–3845.

[20] J. Wen, Z. Zhang, L. Fei, B. Zhang, Y. Xu, Z. Zhang, and J. Li, "A survey on incomplete multiview clustering," *IEEE Trans. Syst. Man Cybern. Syst.*, 2022.

[21] X. Liu, C. Li, M.and Tang, J. Xia, J. Xiong, L. Liu, M. Kloft, and E. Zhu, "Efficient and effective regularized incomplete multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2634–2646, 2020.

[22] M. Yang, P. Li, Y.and Hu, J. Bai, J. Lv, and X. Peng, "Robust multi-view clustering with incomplete information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 1055–1069, 2022.

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

[24] Y. Lin, Y. Gou, Z. Liu, B. Li, J. Lv, and X. Peng, "Completer: Incomplete multi-view clustering via contrastive prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 11 174–11 183.

[25] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1418–1429, 2018.

[26] J. Liu, S. Teng, W. Zhang, X. Fang, L. Fei, and Z. Zhang, "Incomplete multi-view subspace clustering with low-rank tensor," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* IEEE, 2021, pp. 3180–3184.

[27] W. Xia, Q. Gao, Q. Wang, and X. Gao, "Tensor completion-based incomplete multiview clustering," *IEEE Trans. Cybern.*, 2022.

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

[29] Z. Li, C. Tang, X. Liu, X. Zheng, W. Zhang, and E. Zhu, "Tensor-based multi-view block-diagonal structure diffusion for clustering incomplete multi-view data," in *Proc. IEEE Int. Conf. Multimedia Expo.* IEEE, 2021, pp. 1–6.

[30] J. Wen, Z. Zhang, Z. Zhang, L. Zhu, L. Fei, B. Zhang, and Y. Xu, "Unified tensor framework for incomplete multi-view clustering and missing-view inferring," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 11, 2021, pp. 10 273–10 281.

[31] M. Kilmer, K. Braman, N. Hao, and R. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.

[32] Y. Liu, L. Chen, and C. Zhu, "Improved robust tensor principal component analysis via low-rank core matrix," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1378–1389, 2018.

[33] L. Feng, C. Zhu, Z. Long, J. Liu, and Y. Liu, "Multiplex transformed tensor decomposition for multidimensional image recovery," *IEEE Trans. on Image Process.*, 2023.

[34] T. Zhou, C. Zhang, X. Peng, H. Bhaskar, and J. Yang, "Dual shared-specific multiview subspace clustering," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3517–3530, 2019.

[35] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, "Tensor ring decomposition," *arXiv preprint arXiv:1606.05535*, 2016.

[36] Y. Liu, J. Liu, Z. Long, and C. Zhu, *Tensor Computation for Data Analysis*. Springer, 2022.

[37] Z. Long, C. Zhu, J. Liu, and Y. Liu, "Bayesian low rank tensor ring for image recovery," *IEEE Trans. Image Process.*, vol. 30, pp. 3568–3580, 2021.

[38] W. Wang, V. Aggarwal, and S. Aeron, "Efficient low rank tensor ring completion," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5697–5705.

[39] J. Liu, C. Zhu, and Y. Liu, "Smooth compact tensor ring regression," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 9, pp. 4439–4452, 2022.

[40] Z. Long, C. Zhu, P. Comon, and Y. Liu, "Feature space recovery for incomplete multi-view clustering," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* IEEE, 2023.

[41] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 14, 2001.

[42] W. Xia, Q. Gao, Q. Wang, X. Gao, C. Ding, and D. Tao, "Tensorized bipartite graph learning for multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022.

[43] X. Li, H. Zhang, R. Wang, and F. Nie, "Multiview clustering: A scalable and parameter-free bipartite graph fusion method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 330–344, 2020.

[44] X. Shu, X. Zhang, Q. Gao, M. Yang, R. Wang, and X. Gao, "Self-weighted anchor graph learning for multi-view clustering," *IEEE Trans. Multimedia*, 2022.

[45] S. Wang, X. Liu, X. Zhu, P. Zhang, Y. Zhang, F. Gao, and E. Zhu, "Fast parameter-free multi-view subspace clustering with consensus anchor guidance," *IEEE Trans. Knowl. Data Eng.*, vol. 31, pp. 556–568, 2021.

[46] S. Wang, X. Liu, L. Liu, W. Tu, X. Zhu, J. Liu, S. Zhou, and E. Zhu, "Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 9776–9785.

[47] S. Liu, X. Liu, S. Wang, X. Niu, and E. Zhu, "Fast incomplete multi-view clustering with view-independent anchors," *IEEE Trans. Neural Networks Learn. Syst.*, 2022.

[48] J. Hartigan and M. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the royal statistical society. series c (applied statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[49] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[50] N. Srebro, J. Rennie, and T. Jaakkola, "Maximum-margin matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, ser. NIPS'04. Cambridge, MA, USA: MIT Press, 2004, p. 1329–1336.

[51] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[52] C. Zhang, C. Li, H.and Chen, X. Jia, and C. Chen, "Low-rank tensor regularized views recovery for incomplete multiview clustering," *IEEE Trans. Neural Networks Learn. Syst.*, pp. 1–13, 2022.
[53] S. Wang, Y. Chen, Y. Ce, L. Zhang, and V. Voronin, "Low-rank and sparse tensor representation for multi-view subspace clustering," in *Proc. IEEE Int. Conf. Image Process.* IEEE, 2021, pp. 1534–1538.
[54] J. Xu, J. Han, F. Nie, and X. Li, "Re-weighted discriminatively embedded $k$ -means for multi-view clustering," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 3016–3027, 2017.
[55] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.
[56] S. Nene, S. Nayar, and H. Murase, "Columbia object image library (coil-100)," 1996.
[57] X. Cai, H. Wang, H. Huang, and C. Ding, "Joint stage recognition and anatomical annotation of drosophila gene expression patterns," *Bioinformatics*, vol. 28, no. 12, pp. i16–i24, 2012.
[58] L. F. and P. P., "A bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2. IEEE, 2005, pp. 524–531.
[59] Y. Jiang, G. Ye, S. Chang, D. Ellis, and A. Loui, "Consumer video understanding: A benchmark database and an evaluation of human and machine performance," in *Proc. ACM Con. Multimedia Retrieval*, 2011, pp. 1–8.
[60] L. F., R. F., and P. P., "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," in *Proc. IEEE Int. Conf. Comput. Vis. workshop*. IEEE, 2004, pp. 178–178.
[61] M. Houle, H. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "Can shared-neighbor distances defeat the curse of dimensionality?" in *Scientific and Statistical Database Management: 22nd International Conference, SSDBM 2010, Heidelberg, Germany, June 30–July 2, 2010. Proceedings 22*. Springer, 2010, pp. 482–500.
[62] Y. Liang, D. Huang, C. Wang, and P. Yu, "Multi-view graph learning by joint modeling of consistency and inconsistency," *IEEE Trans. Neural Networks Learn. Syst.*, pp. 1–15, 2022.
[63] M. Amini, N. Usunier, and C. Goutte, "Learning from multiple partially observed views-an application to multilingual text categorization," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, 2009.
[64] Z. Zhang, L. Liu, F. Shen, H. T. Shen, and L. Shao, "Binary multiview clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1774–1782, 2018.