# Incomplete Multiview Clustering Using Discriminative Feature Recovery and Tensorized Matrix Factorization

Xinxin Wang, Yongshan Zhang, *Senior Member, IEEE*, Jie Zhang, and Yicong Zhou, *Senior Member, IEEE*

***Abstract*—Multiview clustering task groups objects using multiple properties, such as RGB images, infrared images, and texture information. However, incomplete multi-view clustering faces significant challenges due to missing views that hinder clustering performance. This paper proposes a Discriminative Feature Recovery and Tensorized Matrix Factorization method (DFRTMF) that effectively recovers missing views, learns low-dimensional discriminative embeddings, and enables direct clustering. DFRTMF addresses high dimensionality through projection learning and enables the output of soft indicators. To improve projection and facilitate the recovery of missing views, we propose an uncorrelated constraint based on the scatter matrix of the recovered complete data, exploring the correlations between observed and missing views. To capture high-order correlations among views, a low-rank tensor constraint based on tensor Schatten p-norm regularization is applied to a third-order tensor composed of soft indicator matrices. DFRTMF adaptively controls the inter-coordination between these factorizations using view weights to optimally explore complementary information. Furthermore, we propose an alternating optimization algorithm based on the Alternating Direction Method of Multipliers to effectively solve the proposed objective function. Extensive experiments across diverse datasets demonstrate the effectiveness of DFRTMF compared to the state-of-the-art methods.**

***Index Terms*—Multi-view clustering, incomplete data, matrix factorization, discriminative feature.**

## I. INTRODUCTION

WITH the fast development of multimedia technologies, one can collect diverse information about the same object from different perspectives, such as color, depth, and textual information. These data that reflect different properties of objects are referred to as multiview data [1], [2], [3], [4]. Compared to single-view data, multi-view data offers greater information gain for multimedia data analysis. As one of foundational data analysis technologies, multi-view clustering has attracted significant attention in the field of multimedia applications. For example, remote sensing image clustering [5], [6], [7], [8], [9], image reranking [10], and medical image segmentation [11].

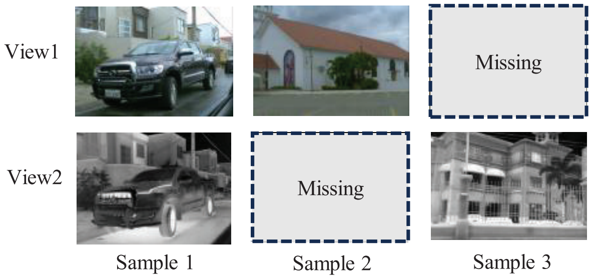


Fig. 1. Incomplete multi-view dataset: This dataset consists of three samples, each with two views. The dashed box means the view is missing.

Most existing multi-view clustering methods assume that all views of objects are available and complete [12], [13], [14], [15], [16]. However, in real-world scenarios, equipment malfunctions and data transmission issues can lead to missing features. As shown in Fig. 1, different visual features in scene recognition, such as RGB and infrared images, represent multiple views, but some scenes may not have all views [17]. Another typical example is in brain activity monitoring, where functional Magnetic Resonance Imaging (fMRI) and Electroencephalography (EEG) are selective. Only certain patients undergo both tests, while others may require only one of the measurements [18]. It is clear that existing multi-view clustering methods struggle to cluster the multi-view data with missing views, as they cannot learn a common clustering partition matrix for all views [19]. Some efforts have been made to adapt learning methods for incomplete multi-view data [20]. For example, [21] proposed a low-rank tensor learning method to recover complete graphs from biased subgraphs corresponding to incomplete views. Reference [22] developed a late fusion model based on initial clustering partition matrices. These incomplete multi-view clustering (IMVC) methods primarily focus on observed views and ignore the underlying semantic information from the missing views. Consequently, when the missing ratio is large, the performance of these methods may become unstable. Recently, researchers have focused on inferring or recovering missing views to enhance various tasks. In the computer

Received 1 November 2024; revised 27 March 2025 and 30 April 2025; accepted 13 May 2025. Date of publication 27 May 2025; date of current version 30 October 2025. This work was supported in part by the Science and Technology Development Fund, Macau, SAR, under Grant 0049/2022/A1 and Grant 0050/2024/AGJ; and in part by the University of Macau under Grant MYRG-GRG2024-00181-FST. This article was recommended by Associate Editor G. Pastuszak. *(Corresponding author: Yicong Zhou.)*

Xinxin Wang, Jie Zhang, and Yicong Zhou are with the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: xinxinwang1024@gmail.com; jiezh1997@gmail.com; yicongzhou@um.edu.mo).

Yongshan Zhang is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: yszhang.cug@gmail.com).

Digital Object Identifier 10.1109/TCSVT.2025.3573927

vision community, 3D Gaussian Splatting models have gained significant traction for generating unseen views, enabling visually realistic renderings that support a range of downstream applications [23], [24]. In machine learning, missing view imputation aims to improve clustering performance [25]. A notable example is incomplete multi-view subspace clustering (IMVSC) [26], where subspace learning captures intrinsic data structure, facilitating infer missing views [27]. For example, [28] introduced a between-view preserving constraint among subspace representation matrices to guide the recovery of missing views. However, most methods struggle with high temporal and spatial complexity levels [29]. For instance, if $n$ denotes the number of data samples, it requires $\mathcal{O}(n^2)$ space for representation matrices and $\mathcal{O}(n^3)$ time for spectral clustering, limiting their widespread applications [30]. Other prevalent IMVC methods for recovering missing views primarily rely on matrix factorization strategies [31]. For example, [32] proposed inferring the original missing data by leveraging the manifold structure of features and a unified embedding process. While these methods provide good interpretability and generalization [33], they still face several limitations: 1) They inadequately address high dimensionality and low separability, which may hinder the model performance; 2) The recovery of missing views may be imprecise due to neglecting explicit correlations with observed views, limiting the adequate refinement of the underlying data distribution; 3) The assumption of a consistent coefficient matrix across views ignores differences among them, potentially leading to overfitting and suboptimal performance.

To simultaneously address the above-mentioned three limitations, we propose a method called Discriminative Feature Recovery and Tensorized Matrix Factorization (DFRTMF). DFRTMF projects the original data into a low dimensional space to address high dimensionality and infers missing features at index positions. Unlike existing uncorrelated constraints based on observed views, it facilitates subspace learning through a global uncorrelated term, enhancing feature separability and improving missing view recovery. Additionally, it considers view discrepancies among coefficient matrices in a tensor to relax the hard consistency constraint. Our main contributions are as follows:

- We propose an incomplete multi-view clustering method that integrates complete discriminative feature recovery and tensorized matrix factorization in a unified model. This model achieves missing view imputation, discriminative subspace learning, and soft indicator generation.
- We propose a dynamic uncorrelated constraint based on the scatter matrices of the recovered complete data. This constraint facilitates subspace learning for intra-view discriminability and enhances missing view recovery by exploring the explicit correlations between observed and missing views.
- We apply a tensor Schatten p-norm to a third-order tensor composed of soft indicator matrices. This approach relaxes the divergence among different indicator matrices and enhances the inter-view spatial low-rank structure.
- We devise an alternating optimization algorithm to effectively solve the proposed objective function. Extensive experiments on diverse datasets demonstrate the effectiveness of our method.

TABLE I

DESCRIPTION OF USED NOTATIONS

| Notation | Explanation |
|---|---|
| $\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ | t-product of two tensors $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ |
| $\boldsymbol{\mathcal{Z}}^{\boldsymbol{T}} \in \mathbb{R}^{n_2 \times n_1 \times n_3}$ | The transpose of a tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ |
| $\|\boldsymbol{\mathcal{Z}}\|_{Ⓢⓟ}^{p}$ | The tensor Schatten p-norm of a tensor $\boldsymbol{\mathcal{Z}}$ |
| $\mathbf{Z}^\top \in \mathbb{R}^{n_2 \times n_1}$ | The transpose of a matrix $\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}$ |
| $\|\mathbf{Z}\|_F$ | The F-norm of a matrix |
| $\operatorname{tr}(\mathbf{Z})$ | The trace of a matrix |
| $\mathbf{X}_v \in \mathbb{R}^{d_v \times n}$ | Original incomplete data with zero padding |
| $\mathbf{H}_v \in \mathbb{R}^{d_v \times n}$ | Recovered complete data |
| $\mathbf{P}_v \in \mathbb{R}^{l \times d_v}$ | Projection matrix |
| $\mathbf{S}_t = \mathbf{H}_v \mathbf{U} \mathbf{H}_v^\top + \gamma \mathbf{I}$ | Global scatter matrix |
| $\mathbf{G}_v \in \mathbb{R}^{l \times k}$ | Orthogonal basis matrix |
| $\mathbf{F}_v \in \mathbb{R}_+^{n \times k}$ | Non-negative soft indicator matrix |
| $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{k \times V \times n}$ | Third-order tensor constructed from $\mathbf{F}_v$ |

Notations throughout this paper are summarized as follows: We use bold lowercase letters $\mathbf{z}$, bold uppercase letters $\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}$, and bold calligraphy letters $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ to represent vectors, matrices, and third-order tensors, respectively. Moreover, $\overline{\boldsymbol{\mathcal{Z}}}$ denotes the discrete Fourier transform (DFT) of $\boldsymbol{\mathcal{Z}}$ along the third dimension, so we have $\overline{\boldsymbol{\mathcal{Z}}} = fft(\boldsymbol{\mathcal{Z}}, [], 3)$ and $\boldsymbol{\mathcal{Z}} = ifft(\overline{\boldsymbol{\mathcal{Z}}}, [], 3)$. The $i$-th frontal slice of $\boldsymbol{\mathcal{Z}}$ and $\overline{\boldsymbol{\mathcal{Z}}}$ are represented as $\boldsymbol{\mathcal{Z}}^{(i)}$ and $\overline{\boldsymbol{\mathcal{Z}}}^{(i)}$, respectively. More notations used in this paper can be found in Table I.

## II. RELATED WORK

### A. Non-Negative Matrix Factorization-Based IMVC Methods

Non-negative Matrix factorization (NMF) [34] aims to find a low-dimensional discriminative coefficient representation based on a compact basis matrix. The resulting matrix product can adequately approximate the original data matrix. Inspired by the significant success of NMF, [35] extended NMF to IMVC by filling in the missing views with average features and performing matrix factorization. Reference [36] introduced weight matrices into the NMF framework to circumvent manual data filling. References [37] and [38] seamlessly integrated matrix factorization and graph regularization techniques to enhance performance. Reference [33] learned consensus projection embeddings to address the imbalanced information among views, while [39] extended NMF to new scenarios where data observations from views are accumulated over time. These methods primarily focus on observed data but fail to address distribution deviations caused by a large missing ratio. References [32] and [40] proposed inferring missing views using a matrix factorization model. However, they ignore the issues of high dimensionality and poor separability and do not consider the spatial structure among views.

### B. Low-Rank Tensor Learning-Based IMVC Methods

Graph-based methods investigate pairwise correlations to capture clustering structures, achieving significant performance [9], [41]. Most methods further explore the high-order spatial correlations among views to enhance performance.

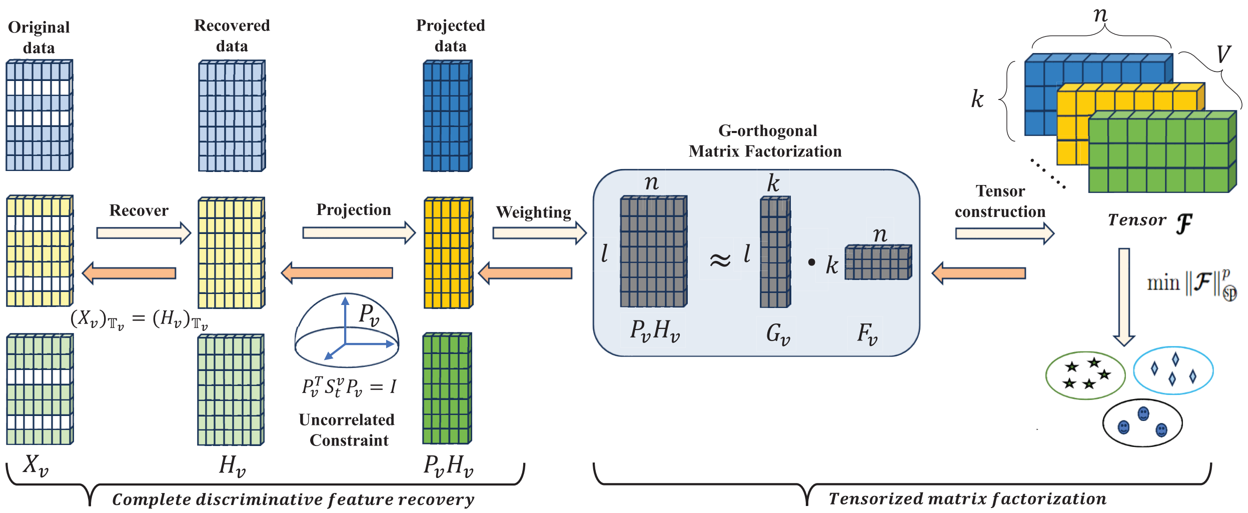


Fig. 2. The proposed DFRTMF model: With data imputation performed at the missing positions, DFRTMF projects sample features into discriminative subspaces, using G-orthogonal matrix factorization to derive intra-view soft indicator matrices. A global uncorrelated constraint enhances intra-view discriminability of low-dimensional embeddings while leveraging explicit correlations between observed and missing views to enhance missing view recovery. A tensor Schatten p-norm is applied to the third-order tensor $\mathcal{F}$ to capture the inter-view correlations. This model adaptively balances the contributions of different views to derive the optimal soft indicator matrices $F_v$, yielding improved clustering results. An alternating optimization algorithm iteratively updates each variable until convergence.

Low-rank tensor learning extends from low-rank matrix learning by integrating multiple matrices as front slices, enabling the exploration of both complementary and consistent information within high-order spatial data. Tensor singular value decomposition (t-SVD) [42] techniques achieve the convex envelope of the tensor average rank, effectively capturing the low-rank structure of a tensor. Inspired by this, many low-rank tensor learning-based IMVC methods have been proposed. Reference [43] employed t-SVD to recover complete sparse graphs from incomplete subspace representations. Reference [44] integrated t-SVD with anchor learning to enhance anchor graphs [17]. Reference [45] applied t-SVD to derive the fused kernel matrix from incomplete kernel matrices. Reference [26], [28], and [46] incorporated tensor learning to impute missing views, achieving promising results. Reference [47] utilized low-rank tensor learning to jointly learn low-rank complete graphs and data. However, these methods ignore the explicit correlations between missing and observed views, which may result in significant deviations in missing view imputation. Recently, a general tensor nuclear norm has been proposed based on t-SVD. This norm is derived through weighted tensor Schatten p-norm minimization (WTSNM) to explicitly incorporate prior knowledge of the singular values of tensor data [48].

*Definition 1 (tensor Schatten p-norm [48]):* Given a third-order tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor Schatten p-norm of $\mathcal{Z}$ is given by

$$\|\mathcal{Z}\|_{\circledS p}^{p} = \left( \sum_{i=1}^{n_3} \left\| \overline{\mathcal{Z}}^{(i)} \right\|_{\circledS p}^{p} \right)^{\frac{1}{p}} = \left( \sum_{i=1}^{n_3} \sum_{j=1}^{\min(n_1,n_2)} \sigma_j \left( \overline{\mathcal{Z}}^{(i)} \right)^{p} \right)^{\frac{1}{p}} \tag{1}$$

where, $\sigma_j\left(\overline{\mathcal{Z}}^{(i)}\right)$ is the $j$-th singular value of matrix $\overline{\mathcal{Z}}$, $0 < p \leq 1$.

## III. Proposed Method

### A. Objective Function

Fig. 2 illustrates the framework of the proposed DFRTMF. DFRTMF focuses on recovering missing views of incomplete multi-view data to enhance clustering performance. The data imputation operation is performed only at the missing positions. To reduce redundant information and maximize feature separability, we perform discriminative projection learning on the recovered complete data, identifying the low-dimensional discriminative subspaces for each view. Next, orthogonal NMF is applied to the complete discriminative embeddings to obtain view-specific soft indicator matrices. Mathematically, we formulate this process as follows:

$$\begin{aligned} \min_{a_v,\mathbf{P}_v,\mathbf{H}_v,\mathbf{G}_v,\mathbf{F}_v} \quad & \sum_{v=1}^{V} a_v^2 \left\|\mathbf{P}_v\mathbf{H}_v - \mathbf{G}_v\mathbf{F}_v^{\top}\right\|_F^2 + \lambda \|\mathbf{P}_v\|_F^2 \\ \text{s.t.} \quad & \mathbf{P}_v\mathbf{S}_t^v\mathbf{P}_v^{\top} = \mathbf{I},\ a_v \geq 0,\ \mathbf{a}^{\top}\mathbf{1} = \mathbf{1}. \\ & \mathbf{G}_v^{\top}\mathbf{G}_v = \mathbf{I},\ (\mathbf{H}_v)_{\mathbb{T}_v} = (\mathbf{X}_v)_{\mathbb{T}_v}, \end{aligned} \tag{2}$$

where $\mathbf{X}_v \in \mathbb{R}^{d_v \times n}$ denotes the original incomplete data for the $v$-th view, with missing views represented by zero padding. $d_v$ indicates the original feature dimension, while $n$ and $V$ represent the total number of samples and views, respectively. We minimize the reconstruction loss to impute the missing views. $\mathbf{H}_v \in \mathbb{R}^{d_v \times n}$ indicates the recovered complete data. $\mathbb{T}_v$ represents the index of available data for the $v$-th view. The constraint $(\mathbf{H}_v)_{\mathbb{T}_v} = (\mathbf{X}_v)_{\mathbb{T}_v}$ ensures that the entries of $\mathbf{H}_v$ are equal to those of $\mathbf{X}_v$ at these positions. The projection matrix for the $v$-th view is $\mathbf{P}_v \in \mathbb{R}^{l \times d_v}$, which maps $\mathbf{H}_v$ to the subspace of $\mathbb{R}^{l \times n}$. This projection addresses the problem of low separability while also serving the purpose of dimensionality reduction. $\mathbf{S}_t^v$ is a scatter matrix, and the constraint $\mathbf{P}_v\mathbf{S}_t^v\mathbf{P}_v^{\top} = \mathbf{I}$ ensures that the subspace data are statistically linearly independent. Additionally, $\mathbf{G}_v \in \mathbb{R}^{l \times k}$ is the basis matrix, and $\mathbf{F}_v \in \mathbb{R}_{+}^{n \times k}$ is the indicator matrix. An orthogonal constraint

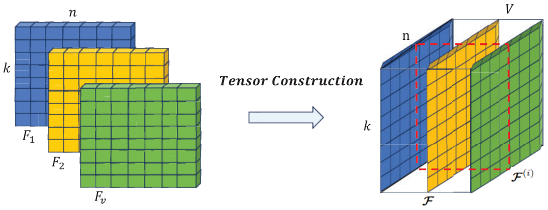


Fig. 3. The third-order tensor construction process.

$\mathbf{G}_v^\top \mathbf{G}_v = \mathbf{I}$ is imposed on $\mathbf{G}_v$ to enhance the discriminability of the bases. The constraint $a_v \geq 0, \mathbf{a}^\top \mathbf{1} = \mathbf{1}$ balances the contributions of different views and allows $a_v^2$ to serve as an adaptive weight of the $v$-th view. $\lambda$ is a trade-off parameter.

Existing methods fail to establish the explicit correlations between observed and missing views, which may lead to significant imputation deviations [32]. To explore the correlations between observed and missing views, we define scatter matrices $\mathbf{S}_t^v$ as $\mathbf{H}_v \mathbf{U} \mathbf{H}_v^\top + \gamma \mathbf{I}$.[1] We set $\mathbf{U}$ simply as the identity matrix in this paper. Instead of computing scatter matrices solely from the original observed data, given by $\overline{\mathbf{S}}_t^v = \mathbf{X}_v \mathbf{X}_v^\top$, the proposed scatter matrices calculate the variances using both observed feature variables and imputed variables. Thus, this global uncorrelated constraint effectively captures the shared subspace between the observed and missing views, thereby facilitating the imputation of missing views.

The ideal indicator matrices $\mathbf{F}_v$ across views should be consistent. Existing methods typically use the Frobenius norm to measure inter-view errors element by element during fusion [37]. However, distinct clustering structures in different views can lead to inconsistencies among these indicator matrices. This reliance on the element-wise measure may result in overfitting and overlook high-order correlations among the views. Therefore, more effective regularization is necessary to encourage the convergence of these indicator matrices. To effectively explore the complementary information among views, we minimize the spatial divergence between different indicator matrices $\mathbf{F}_v$ by imposing a tensor low-rank constraint on the third-order tensor $\mathcal{F}$. Thus, our final objective function is formulated as follows:

$$\begin{aligned} \min_{\Omega} \quad & \sum_{v=1}^{V} a_v^2 \left\| \mathbf{P}_v \mathbf{H}_v - \mathbf{G}_v \mathbf{F}_v^\top \right\|_F^2 + \lambda \|\mathbf{P}_v\|_F^2 + \beta \|\mathcal{F}\|_{\circledS p}^{p} \\ \text{s.t.} \quad & \mathbf{P}_v \mathbf{S}_t^v \mathbf{P}_v^\top = \mathbf{I}, a_v \geq 0, \mathbf{a}^\top \mathbf{1} = \mathbf{1}. \\ & \mathbf{G}_v^\top \mathbf{G}_v = \mathbf{I}, \mathbf{F}_v \geq 0, (\mathbf{H}_v)_{\mathbb{T}_v} = (\mathbf{X}_v)_{\mathbb{T}_v}, \end{aligned} \tag{3}$$

where $\Omega = \{a_v, \mathbf{P}_v, \mathbf{H}_v, \mathbf{G}_v, \mathbf{F}_v, \mathcal{F}\}$ is a set of variables. $\mathcal{F} \in \mathbb{R}^{k \times V \times n}$ is a third-order tensor constructed from the matrices $\mathbf{F}_v$, as illustrated in Fig 3. The $i$-th frontal slice $\mathcal{F}^{(i)}$ represents the relationships between samples and clusters across different views. By imposing a tensor Schatten p-norm minimization constraint on $\mathcal{F}$, we ensure that each $\mathcal{F}^{(i)}$ exhibits a spatial low-rank structure, effectively characterizing the complementary information among different views. Additionally, a non-negative constraint $\mathbf{F}_v \geq 0$ is applied to ensure that $\mathbf{F}_v$ serves as a soft indicator, directly characterizing the cluster structure. The position of the maximum value in each column of $\mathbf{F}_v$ indicates the cluster assignment, yielding cluster labels without any post-processing. The parameter $0 < p \leq 1$ is a manually set, and $\beta$ serves as a balance parameter.

### B. Optimization

Obviously, Eq. (3) is a non-convex optimization problem and it is difficult to directly solve these variables. Inspired by Alternating Direction Method of Multipliers (ADMM), we devise an alternating optimization algorithm to solve this objective function. By introducing an auxiliary variable $\mathcal{J}$, and letting $\mathcal{F} = \mathcal{J}$, the objective function Eq. (3) can be expressed as the following separable augmented Lagrange function:

$$\begin{aligned} \min_{\Omega} \sum_{v=1}^{V} \quad & a_v^2 \left\| \mathbf{P}_v \mathbf{H}_v - \mathbf{G}_v \mathbf{F}_v^\top \right\|_F^2 + \lambda \|\mathbf{P}_v\|_F^2 + \beta \|\mathcal{J}\|_{\circledS p}^{p} \\ & + \frac{\rho}{2} \|\mathcal{F} - \mathcal{J}\|_F^2 + \langle \mathcal{Y}, \mathcal{F} - \mathcal{J} \rangle, \\ \text{s.t.} \quad & \mathbf{P}_v \mathbf{S}_t^v \mathbf{P}_v^\top = \mathbf{I}, a_v \geq 0, \mathbf{a}^\top \mathbf{1} = \mathbf{1}, \\ & \mathbf{G}_v^\top \mathbf{G}_v = \mathbf{I}, \mathbf{F}_v \geq 0, (\mathbf{H}_v)_{\mathbb{T}_v} = (\mathbf{X}_v)_{\mathbb{T}_v}, \end{aligned} \tag{4}$$

where $\Omega = \{a_v, \mathbf{P}_v, \mathbf{H}_v, \mathbf{G}_v, \mathbf{F}_v, \mathcal{F}, \mathcal{J}\}$ are variables to be solved. $\mathcal{Y}$ is the Lagrange multiplier, and $\rho$ is the penalty parameter. Thus, we can solve for each variable individually while fixing the others.

**Update $\mathbf{G}_v$** with fixed $\{a_v, \mathbf{P}_v, \mathbf{H}_v, \mathbf{F}_v, \mathcal{J}\}$, Eq. (4) becomes:

$$\begin{aligned} \min_{\mathbf{G}_v} \sum_{v=1}^{V} \quad & a_v^2 \left\| \mathbf{P}_v \mathbf{H}_v - \mathbf{G}_v \mathbf{F}_v^\top \right\|_F^2, \\ \text{s.t.} \quad & \mathbf{G}_v^\top \mathbf{G}_v = \mathbf{I}, \end{aligned} \tag{5}$$

which can be written as:

$$\begin{aligned} \min_{\mathbf{G}v} \quad & \operatorname{tr}\left( \mathbf{G}_v^\top \mathbf{P}_v \mathbf{H}_v \mathbf{F}_v \right), \\ \text{s.t.} \quad & \mathbf{G}_v^\top \mathbf{G}_v = \mathbf{I}. \end{aligned} \tag{6}$$

To solve Eq. (6), we introduce the following **Theorem 1**.

*Theorem 1: For the following optimization problem, the optimal solution of*

$$\begin{aligned} & \min_{\mathbf{G}} \operatorname{tr}\left( \mathbf{G}^\top \mathbf{B} \right), \\ & \text{s.t. } \mathbf{G}^\top \mathbf{G} = \mathbf{I}, \end{aligned} \tag{7}$$

is $\mathbf{G} = \mathbf{U}[\mathbf{I}, \mathbf{0}]\mathbf{V}^\top$, *where* $\mathbf{U}$ *and* $\mathbf{V}$ *are the left-singular vectors and right-singular vectors produced by singular value decomposition (SVD) on* $\mathbf{B}$, *respectively.*

**Update $\mathbf{F}_v$** with fixed $\{a_v, \mathbf{P}_v, \mathbf{H}_v, \mathbf{G}_v, \mathcal{J}\}$, Eq. (4) becomes:

$$\begin{aligned} \min_{\mathbf{F}_v} \quad & a_v^2 \left\| \mathbf{P}_v \mathbf{H}_v - \mathbf{G}_v \mathbf{F}_v^\top \right\|_F^2 + \frac{\rho}{2} \left\| \mathbf{F}_v - \mathbf{J}_v + \frac{\mathbf{Y}_v}{\rho} \right\|_F^2, \\ \text{s.t.} \quad & \mathbf{F}_v \geq 0. \end{aligned} \tag{8}$$

where $\mathbf{J}_v$ derives from the inverse construction process of $\mathcal{J}$. Thus, minimizing Eq. (8) is equivalent to:

$$\min_{\mathbf{F}_v} \left\| \mathbf{F}_v - \frac{a_v^2 \mathbf{H}_v^\top \mathbf{P}_v^\top \mathbf{G}_v + \frac{\rho}{2} \left( \mathbf{J}_v - \frac{\mathbf{Y}_v}{\rho} \right)}{a_v^2 + \frac{\rho}{2}} \right\|_F^2. \tag{9}$$

---

[1] $\gamma$ is a coefficient to ensure that the scatter matrix is positive definite.

Then, $\mathbf{F}_v$ can be updated directly by using the following update rule:

$$\mathbf{F}_v = \left( \frac{a_v^2 \mathbf{H}_v^\top \mathbf{P}_v^\top \mathbf{G}_v + \frac{\rho}{2}\left(\mathbf{J}_v - \frac{\mathbf{Y}_v}{\rho}\right)}{a_v^2 + \frac{\rho}{2}} \right)_+ . \tag{10}$$

**Update $\mathcal{J}$** with fixed $\{a_v, \mathbf{P}_v, \mathbf{H}_v, \mathbf{G}_v, \mathbf{F}_v\}$, Eq. (4) becomes:

$$\min_{\mathcal{J}} \beta \|\mathcal{J}\|_{\circledS_p}^p + \frac{\rho}{2} \left\| \mathcal{F} - \mathcal{J} + \frac{\mathcal{Y}}{\rho} \right\|_F^2 , \tag{11}$$

which has a closed-formed solution as **Lemma 1** [48]:

*Lemma 1:* Given a third-order tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and its tensor singular value decomposition $\mathcal{Z} = \mathcal{U} * \Sigma * \mathcal{V}^T$. Then the optimal solution of

$$\min_{\mathcal{J}} \beta \|\mathcal{J}\|_{\circledS_p}^p + \frac{1}{2} \|\mathcal{J} - \mathcal{Z}\|_F^2 , \tag{12}$$

is $\mathcal{J}^* = \mathcal{S}_\tau(\mathcal{Z}) = \mathcal{U} * ifft(\Gamma_\tau(\bar{\Sigma})) * \mathcal{V}^T$, where $\Gamma_\tau(\bar{\Sigma})$ is a $f$-diagonal tensor in Fourier domain, whose diagonal elements can be obtained by using the GST algorithm introduced in [48].

By utilizing **Lemma 1**, $\mathcal{J}$ can be updated by using the following update rule:

$$\mathcal{S}_{\frac{\beta}{\rho}} \left( \mathcal{F} + \frac{\mathcal{Y}}{\rho} \right). \tag{13}$$

**Update $\mathbf{P}_v$** with fixed $\{a_v, \mathbf{H}_v, \mathbf{G}_v, \mathbf{F}_v, \mathcal{J}\}$, Eq. (4) becomes:

$$\begin{aligned} &\min_{\mathbf{P}_v} \quad a_v^2 \left\| \mathbf{P}_v \mathbf{H}_v - \mathbf{G}_v \mathbf{F}_v^\top \right\|_F^2 + \lambda \|\mathbf{P}_v\|_F^2 , \\ &\text{s.t. } \mathbf{P}_v \mathbf{S}_t^v \mathbf{P}_v^\top = \mathbf{I}. \end{aligned} \tag{14}$$

Expanding the Eq. (13) to trace norm, we get:

$$\begin{aligned} &\min_{\mathbf{P}_v} \quad a_v^2 \mathrm{tr}\left(\mathbf{P}_v \mathbf{H}_v \mathbf{H}_v^\top \mathbf{P}_v^\top - 2\mathbf{P}_v \mathbf{H}_v \mathbf{F}_v \mathbf{G}_v^\top\right) + \lambda \mathrm{tr}\left(\mathbf{P}_v \mathbf{P}_v^\top\right), \\ &\text{s.t. } \mathbf{P}_v \mathbf{S}_t^v \mathbf{P}_v^\top = \mathbf{I}. \end{aligned} \tag{15}$$

In this paper, we set $\gamma = \frac{\lambda}{a_v^2}$ (introduced in Section III-A) to reduce the parameters of the model. Noting that in Eq. (15), $a_v^2 \mathrm{tr}\left(\mathbf{P}_v \mathbf{H}_v \mathbf{H}_v^\top \mathbf{P}_v^\top\right) + \lambda \mathrm{tr}\left(\mathbf{P}_v \mathbf{P}_v^\top\right)$ is a constant. Thus, minimizing Eq. (15) is equivalent to:

$$\begin{aligned} &\max_{\mathbf{P}_v} \quad \mathrm{tr}\left(\mathbf{P}_v \mathbf{H}_v \mathbf{F}_v \mathbf{G}_v^\top\right), \\ &\text{s.t. } \mathbf{P}_v \mathbf{S}_t^v \mathbf{P}_v^\top = \mathbf{I}. \end{aligned} \tag{16}$$

Since $\mathbf{S}_t^v$ is a positive definite matrix, it can be converted into $\mathbf{S}_t^{v(\frac{1}{2})} \mathbf{S}_t^{v(\frac{1}{2})}$. Thus, Eq. (16) can be rewritten as:

$$\begin{aligned} &\max_{\mathbf{P}_v} \quad \mathrm{tr}\left(\mathbf{P}_v \mathbf{S}_t^{v(\frac{1}{2})} \mathbf{S}_t^{v(-\frac{1}{2})} \mathbf{H}_v \mathbf{F}_v \mathbf{G}_v^\top\right), \\ &\text{s.t. } \mathbf{P}_v \mathbf{S}_t^{v(\frac{1}{2})} \mathbf{S}_t^{v(\frac{1}{2})} \mathbf{P}_v^\top = \mathbf{I}. \end{aligned} \tag{17}$$

Let $\mathbf{Q}_v = \mathbf{P}_v \mathbf{S}_t^{v(\frac{1}{2})}$ and $\mathbf{N}_v = \mathbf{S}_t^{v(-\frac{1}{2})} \mathbf{H}_v \mathbf{F}_v \mathbf{G}_v^\top$, then Eq. (17) can be rewritten as:

$$\max_{\mathbf{Q}_v \mathbf{Q}_v^\top = \mathbf{I}} \quad \mathrm{tr}\left(\mathbf{Q}_v \mathbf{N}_v\right). \tag{18}$$

Supposing $\mathbf{U}_v \Lambda \mathbf{V}_v^\top = svd(\mathbf{N}_v)$ and using **Theorem 1**, we can obtain the solution for $\mathbf{Q}_v$ as follows:

$$\mathbf{Q}_v = \mathbf{V}_v \mathbf{U}_v^\top. \tag{19}$$

Consequently, $\mathbf{P}_v$ can be updated as follows:

$$\mathbf{P}_v = \mathbf{Q}_v \mathbf{S}_t^{v(-\frac{1}{2})} = \mathbf{V}_v \mathbf{U}_v^\top \mathbf{S}_t^{v(-\frac{1}{2})}. \tag{20}$$

**Update $\mathbf{H}_v$** with fixed $\{a_v, \mathbf{P}_v, \mathbf{G}_v, \mathbf{F}_v, \mathcal{J}\}$, Eq. (4) becomes:

$$\begin{aligned} &\min_{\mathbf{H}_v} \quad \left\| \mathbf{P}_v \mathbf{H}_v - \mathbf{G}_v \mathbf{F}_v^\top \right\|_F^2 , \\ &\text{s.t.} \quad (\mathbf{H}_v)_{\mathbb{T}_v} = (\mathbf{X}_v)_{\mathbb{T}_v}. \end{aligned} \tag{21}$$

By differentiating the Eq. (21) and setting the derivative to zero, $\mathbf{H}_v$ can be updated using the following update rule:

$$\mathbf{H}_v(:, i) = \begin{cases} \mathbf{X}_v(:, i), i \in \mathbb{T}_v \\ \mathbf{M}_v(:, i), i \notin \mathbb{T}_v \end{cases}, \tag{22}$$

where $\mathbf{M}_v = \mathbf{P}_v^{-1} \mathbf{G}_v \mathbf{F}_v^\top$.

**Algorithm 1** DFRTMF

```
Input: $V$-view incomplete dataset $\{\mathbf{X}^v\}_{v=1}^V$, the number of
    cluster $k$, and parameters $\lambda$ and $\beta$.
Output: Clustering labels of data points.
Initialize $\mathbf{G}_v = 0$, $\mathbf{F}_v = \mathbf{J}_v = 0$, $\mathbf{P}_v = 0$, $\mathbf{Y} = 0$.
Initialize $\mathbf{H}_v = \mathbf{X}_v$, $a_v = \frac{1}{V}$, $\rho = 10^{-4}$, $\mu = 1.1$
while not converge do
    Update $\mathbf{G}^v$ by Eq. (7);
    Update $\mathbf{F}^v$ by Eq. (10);
    Update $\mathcal{J}$ by solving Eq. (11);
    Update $\mathbf{P}^v$ by solving Eq. (20);
    Update $\mathbf{H}^v$ by solving Eq. (22);
    Update $a_v$ by Eq. (24);
    Update $\mathcal{Y}$ and $\rho$: $\mathcal{Y} = \mathcal{Y} + \rho(\mathcal{F} - \mathcal{J})$, $\rho = \mu\rho$;
end while
Calculate the clustering results by using
    $\mathbf{F} = \sum_{v=1}^V a_v^2 \mathbf{F}_v / \sum_{v=1}^V a_v^2$.
return: Clustering results.
```

**Update $a_v$** with fixed $\{\mathbf{P}_v, \mathbf{H}_v, \mathbf{G}_v, \mathbf{F}_v, \mathcal{J}\}$, Eq. (4) becomes:

$$\begin{aligned} &\min_{a_v} \quad \sum_{v=1}^V a_v^2 \left\| \mathbf{P}_v \mathbf{H}_v - \mathbf{G}_v \mathbf{F}_v^\top \right\|_F^2 , \\ &\text{s.t.} \quad a_v \geq 0, \mathbf{a}^\top \mathbf{1} = \mathbf{1}. \end{aligned} \tag{23}$$

Let $r^v = \left\| \mathbf{P}_v \mathbf{H}_v - \mathbf{G}_v \mathbf{F}_v^\top \right\|_F^2$. According to Cauchy-Schwarz's inequality, we have

$$a_v = \frac{\frac{1}{r^v}}{\sum_{v=1}^V \frac{1}{r^v}}. \tag{24}$$

Finally, the entire optimization procedure for Eq. (4) is listed in Algorithm 1.

### C. Complexity Analysis

*1) Time Complexity:* The computational burden of the proposed DFRTMF involves updating six variables. It takes $\mathcal{O}(lk^2)$ to update $\mathbf{G}^v$ through SVD. To solve $\mathbf{F}_v$, it requires $\mathcal{O}(nd_v lk)$ to perform the threshold operation. For updating $\mathcal{J}$, it computes the 3D FFT and 3D inverse FFT on a $k \times V \times n$ tensor, along with performing $n$ SVDs of $k \times V$ matrices in the Fourier domain. Thus, its computational complexity is $\mathcal{O}(2knVlog(n) + V^2kn)$. As for $\mathbf{P}_v$, it needs $\mathcal{O}(l^2 d_v)$ to

TABLE II
MULTI-VIEW DATASETS IN EXPERIMENTS

| Datasets | Samples | Clusters | Views |
|---|---|---|---|
| MSRCv1 [49] | 210 | 7 | 24, 576, 512,256,254 |
| Caltech101-7 [17] | 1474 | 7 | 48, 40, 254, 1984, 512, 928 |
| BDGP [17] | 2500 | 5 | 1000, 500, 250 |
| CCV [45] | 6773 | 20 | 20, 20, 20 |
| Animal [50] | 11673 | 20 | 2689, 2000, 2001, 2000 |
| FMNIST [19] | 60000 | 10 | 521, 512, 1280 |

perform SVD operation. To solve $\mathbf{H}_v$, it mainly costs $\mathcal{O}(lkn)$ for matrix multiplication. When updating $a_v$, it only costs $\mathcal{O}(1)$. In summary, the overall complexity is $\mathcal{O}((Vlk^2+ndlk+2knVlog(n)+V^2kn+l^2d+Vlkn)t)$, where $t$ denotes the number of iterations and $d=\sum_{v=1}^{V} dv$ is the total data dimension.

*2) Space Complexity:* The major space burden of the proposed DFRTMF consists of matrices $\mathbf{H}_v \in \mathbb{R}^{d_v \times n}$, $\mathbf{F} \in \mathbb{R}^{n \times k}$ and $\mathcal{J} \in \mathbb{R}^{k \times V \times n}$, $\mathbf{G}_v \in \mathbb{R}^{l \times k}$, and $\mathbf{P}_v \in \mathbb{R}^{l \times d_v}$. Thus, the corresponding space complexity is $\mathcal{O}\left(\sum_{v=1}^{V} d_v n + 2kVn + \sum_{v=1}^{V} lk + \sum_{v=1}^{V} ld_v\right)$.

## IV. EXPERIMENTS

### A. Experimental Setup

*1) Benchmark Datasets:* Six widely used multi-view datasets are tested in our experiments. Table II summarizes the detailed information of these datasets.

- **MSRCv1**[2] is a general object database comprising 210 images across 7 categories. Each image is presented by five views of features: the CM feature (view 1), the HOG feature (view 2), the GIST feature (view 3), the LBP feature (view 4), and the CENT feature (view 5).
- **Caltech101**[3] is an object recognition dataset containing 9146 images. The Caltech101-7 subset consists of 1474 images across 7 categories. Each image is presented by six views of features: the Gabor feature (view 1), the WM feature (view 2), the CENTRIST feature (view 3), the HOG feature (view 4), the GIST feature (view 5), and the LBP feature (view 6).
- **BDGP**[4] is a dataset of Drosophila embryos images, comprising three views: lateral image, dorsal, and ventral. It contains 2500 samples across 5 classes. Each view extracts a SIFT feature to construct three perspectives.
- **CCV**[5] is a Columbia Consumer Video dataset. It is composed of 6773 samples selected from 20 categories. Each video is described by three views of features: the SIFT feature (view 1), the STIP feature (view 2), and the MFCC feature (view 3).
- **Animal**[6] is an animal image dataset, which consists of 11673 images representing 20 species. Each image is described by four views of features: the LSS feature (view 1), the SIFT feature (view 2), the RGSIFT feature (view 3), and the SURF feature (view 4).
- **FMNIST**[7] is a grayscale images dataset of clothing items, which consists of 60000 images across 10 different classes. Three visual features are extracted to present each image.

For the FMNIST multi-view datasets, no specifics of each dataset's views are provided by the original authors. To construct the incomplete datasets, we adopt the approach in [17]. Specifically, we generate incomplete datasets with missing ratios (MR) of 0.1, 0.5, and 0.9. For instance, when the MR is 0.1, we randomly select 90% of the samples as complete data and randomly drop partial views from the remaining 10%, ensuring that at least one view is preserved for each sample.

*2) Compared Methods:* To evaluate the effectiveness of our proposed DFRTMF model for incomplete multi-view clustering, we compare it with the following nine state-of-the-art multi-view clustering methods. We list them as follows:

- DAIMC [36] is an extended matrix factorization method, which aligns the distinct basis matrices through regression model.
- UEAF [32] is a matrix factorization approach that not only infers the missing views but also ensures the common local structure on the reconstructed representation across different views.
- FLSD [37] develops a matrix factorization based IMC method, which introduces graph regularizer without extra term and parameter.
- PIMVC [33] projects the original features into low-dimensional representation under uncorrelated constraint, on which local geometric structure regularization is applied.
- HCLS_CGL [20] develops confidence-nearest-neighbor graphs to guide the consensus graph learning.
- AGC-IMC [41] integrates the graph recovering and common representation learning with exact cluster structure.
- IMVTSC-MVI [47] is a subspace clustering framework with consensus graph and tensor rank constraints, which enables the inferring of missing views.
- IMVC-CBG [17] introduces anchor graph to efficiently learn the clustering structures across incomplete views.
- FCMVC-IV [39] proposes a continual learning method based on matrix factorization framework.
- DVIDE [51] is a method based on deep neural networks that extends multi-view contrastive learning.
- CPSPAN [52] is a multiview contrastive learning model with partially-aligned setting.

*3) Implementation Details:* For all datasets, the true number of clusters is known in advance and is taken as the true number of categories. Following [32], [33], we utilize three evaluation metrics: clustering accuracy (ACC), normalized mutual information (NMI), and Purity to assess clustering performance. Higher values indicate better clustering quality. To alleviate the effect of randomness, we repeat all experiments 10 times and report the best result. All experiments are performed on a computer with a 3.5GHz AMD Ryzen 9 3950X CPU and 64 GB RAM, MATLAB2022b (64-bit).

---

[2] https://www.microsoft.com/en-us/research/project
[3] http://www.vision.caltech.edu/Image Datasets/Caltech101/
[4] https://www.fruitfly.org/
[5] https://www.ee.columbia.edu/ln/dvmm/CCV/
[6] https://cvml.ista.ac.at/AwA/
[7] https://github.com/zalandoresearch/fashion-mnist

TABLE III

CLUSTERING PERFORMANCE OF DIFFERENT METHODS ON SIX DATASETS WITH 10%,50%,90% MISSING RATIOS. THE BEST RESULTS ARE HIGHLIGHTED, UNDERLINE MEANS THAT METHOD RANKS SECOND

| Metrics | | ACC(%) | | | NMI(%) | | | Purity(%) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Datasets | Miss Ritios | 0.1 | 0.5 | 0.9 | 0.1 | 0.5 | 0.9 | 0.1 | 0.5 | 0.9 |
| MSRCv1 [49] | DAIMC [36] | 73.43 | 69.24 | 65.33 | 67.85 | 56.69 | 55.08 | 77.14 | 69.33 | 65.81 |
| | UEAF [32] | 83.52 | 50.28 | 39.43 | 71.45 | 40.43 | 27.64 | 83.52 | 51.33 | 41.14 |
| | FLSD [37] | 74.76 | 69.71 | 71.52 | 66.15 | 59.04 | 60.53 | 77.62 | 71.23 | 71.52 |
| | PIMVC [33] | 90.66 | 72.38 | 81.33 | 82.59 | 64.65 | 67.31 | 90.66 | 75.33 | 81.33 |
| | HCLS-CGL [20] | 73.33 | 71.05 | 64.76 | 69.61 | 64.97 | 56.94 | 76.66 | 74.38 | 67.14 |
| | AGC-IMC [41] | 78.09 | 84.28 | 77.90 | 74.69 | 72.54 | 63.18 | 78.57 | 84.28 | 77.90 |
| | IMVTSC-MVI [47] | 82.09 | 85.24 | 83.33 | 75.86 | 77.29 | 73.87 | 82.09 | 85.24 | 83.33 |
| | IMVC-CBG [17] | 66.76 | 56.86 | 53.52 | 56.92 | 46.52 | 42.44 | 66.90 | 58.10 | 54.40 |
| | FCMVC-IV [39] | 65.24 | 57.14 | 52.38 | 52.84 | 40.87 | 36.38 | 65.71 | 57.62 | 54.29 |
| | DIVIDE [51] | 57.52 | 54.95 | 45.05 | 47.35 | 43.00 | 32.51 | 59.52 | 57.14 | 47.33 |
| | CPSPAN [52] | 74.29 | 68.20 | 78.57 | 76.67 | 68.28 | 76.67 | 64.76 | 59.76 | 68.10 |
| | Ours | **98.10** | **98.10** | **96.67** | **96.03** | **95.70** | **93.49** | **98.10** | **98.10** | **96.67** |
| BDGP [17] | DAIMC [36] | 45.55 | 28.59 | 32.13 | 19.21 | 5.95 | 8.48 | 46.17 | 29.21 | 33.08 |
| | UEAF [32] | 50.57 | 45.52 | 38.66 | 26.37 | 22.94 | 16.26 | 51.00 | 45.98 | 38.82 |
| | FLSD [37] | 43.64 | 38.74 | 33.20 | 19.79 | 14.95 | 9.69 | 45.40 | 38.86 | 33.96 |
| | PIMVC [33] | 33.68 | 28.74 | 29.79 | 11.17 | 6.53 | 7.77 | 34.70 | 30.07 | 29.86 |
| | HCLS-CGL [20] | 23.08 | 20.48 | 23.33 | 2.91 | 0.38 | 0.80 | 23.16 | 20.48 | 23.46 |
| | AGC-IMC [41] | 37.40 | 34.80 | 35.08 | 17.72 | 10.44 | 7.38 | 39.36 | 36.88 | 35.08 |
| | IMVTSC-MVI [47] | 30.68 | 23.56 | 27.48 | 6.10 | 1.59 | 3.17 | 30.86 | 23.68 | 28.28 |
| | IMVC-CBG [17] | 44.74 | 34.44 | 32.22 | 25.18 | 11.14 | 10.85 | 45.25 | 34.52 | 33.82 |
| | FCMVC-IV [39] | 31.96 | 34.00 | 38.00 | 10.53 | 10.42 | 12.94 | 33.60 | 35.16 | 39.12 |
| | DIVIDE [51] | 35.01 | 34.40 | 31.79 | 13.58 | 11.97 | 6.73 | 36.82 | 36.28 | 33.84 |
| | CPSPAN [52] | 46.12 | 22.08 | 46.68 | 43.64 | 18.26 | 45.00 | 29.12 | 3.68 | 30.36 |
| | Ours | **98.76** | **99.64** | **98.12** | **95.94** | **98.55** | **94.55** | **98.76** | **99.64** | **98.12** |
| Caltech101_7 [17] | DAIMC [36] | 45.49 | 50.35 | 42.02 | 51.36 | 47.03 | 46.74 | 85.12 | 83.05 | 82.03 |
| | UEAF [32] | 45.65 | 36.16 | 32.32 | 43.25 | 28.05 | 19.44 | 84.92 | 79.44 | 73.74 |
| | FLSD [37] | 55.89 | 56.36 | 55.51 | 6.39 | 5.80 | 2.58 | 58.63 | 58.56 | 56.23 |
| | PIMVC [33] | 67.09 | 66.65 | 67.15 | 56.14 | 53.95 | 50.09 | 87.65 | 86.66 | 87.42 |
| | HCLS-CGL [20] | 69.28 | **77.73** | 66.35 | 55.06 | 62.16 | 49.02 | 86.27 | 88.40 | 85.28 |
| | AGC-IMC [41] | 65.94 | 65.87 | 67.23 | 50.60 | 52.06 | 54.38 | 84.24 | 85.14 | 86.29 |
| | IMVTSC-MVI [47] | 59.45 | 60.23 | 64.09 | 48.97 | 46.30 | 46.62 | 87.53 | 84.57 | 85.10 |
| | IMVC-CBG [17] | 58.57 | 61.14 | 61.15 | 49.25 | 42.72 | 37.66 | 82.45 | 80.98 | 77.77 |
| | FCMVC-IV [39] | 53.12 | 52.99 | 51.83 | 46.74 | 45.81 | 45.04 | 82.77 | 82.84 | 82.56 |
| | DIVIDE [51] | 35.48 | 37.04 | 35.29 | 39.89 | 39.00 | 33.12 | 75.45 | 76.55 | 75.47 |
| | CPSPAN [52] | 55.56 | 54.95 | 55.50 | 57.01 | 54.61 | 48.94 | 87.11 | 86.30 | 84.19 |
| | Ours | **83.85** | 75.37 | **76.12** | **72.01** | **73.38** | **64.72** | **90.98** | **90.71** | **88.13** |

For these baseline methods, we strive to adhere as closely as possible to the recommended parameter settings from the original literature. In our method, the trade-off parameter $\lambda$ is tuned within the range of $\left[1e^{-5}, 1e^{-4} \ldots, 1e^{-1}\right]$. The parameter $\beta$ and the projection dimension $l$ are set to 1 and $k$ by default, where $k$ represents the number of clusters in each dataset. The parameter $p$ is tuned in the range of [0.1 : 0.1 : 1]. Therefore, our model includes only two parameters, allowing for straightforward grid searching to identify the optimal values of $\lambda$ and $p$. Following [32], [33], we use evaluation metrics to identify the optimal parameters.

### B. Performance Comparison

Table III-IV show the clustering performance of the proposed DFRTMF and other state-of-the-art methods across six multi-view datasets with different missing ratios. The best results are highlighted in bold. we can observe that,

1) Our proposed DFRTMF outperforms other state-of-the-art algorithms in terms of ACC, NMI, and Purity metrics in most cases, which demonstrates the effectiveness and superiority of our approach. For example, our approach exceeds the suboptimal approach by 7.44%, 12.86%, and 13.34% in ACC on the MSRCv1 dataset across three missing ratios. Additionally, our method outperforms the deep learning-based methods such as DIVIDE and CPSPAN, demonstrating its promising potential in practical deployment.
2) Our method maintains highly competitive performance at large missing ratios compared to other methods. For example, DFRTMF shows improved performance as the missing ratios increase. when the missing ratio is 0.5, our approach achieves an accuracy of 23.15% on the Animal dataset, while this accuracy increases to 26.24% when the missing ratio is 0.9. This improvement may be due to the low separability of the large-scale Animal dataset, where projection learning enhances the discriminability of the recovered missing views.
3) Simple consistency learning based on element-by-element measurements can extend MVC to IMVC, including methods such as DAIMC, UEAF, FLSD, and FCMVC-IV. However, these methods do not sufficiently explore the complementary information among views. In contrast, our approach utilizes a tensor low-rank constraint to effectively exploit spatial structural information. Taking the BDGP dataset as an example, our

TABLE IV

CLUSTERING PERFORMANCE OF DIFFERENT METHODS ON SIX DATASETS WITH 10%,50%,90% MISSING RATIOS. THE BEST RESULTS ARE HIGHLIGHTED, UNDERLINE MEANS THAT METHOD RANKS SECOND.'-' MEANS OUT OF THE CPU MEMORY

| | Metrics | ACC(%) | | | NMI(%) | | | Purity(%) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Datasets | Miss Ritios | 0.1 | 0.5 | 0.9 | 0.1 | 0.5 | 0.9 | 0.1 | 0.5 | 0.9 |
| CCV [45] | DAIMC [36] | 17.04 | 13.93 | 12.85 | 13.21 | 8.36 | 6.42 | 19.55 | 17.21 | 16.05 |
| | UEAF [32] | 17.51 | 15.05 | 11.71 | 13.91 | 11.26 | 7.31 | 21.97 | 18.92 | 15.79 |
| | FLSD [37] | 15.92 | 15.05 | 13.34 | 12.19 | 11.20 | 9.34 | 20.27 | 19.32 | 17.60 |
| | PIMVC [33] | 23.55 | 21.35 | 18.89 | 22.15 | 18.10 | 14.83 | 28.42 | 25.68 | 23.22 |
| | HCLS-CGL [20] | 18.33 | 18.24 | 17.84 | 13.78 | 13.52 | 12.90 | 22.53 | 22.43 | 21.23 |
| | AGC-IMC [41] | 22.19 | 19.44 | 17.91 | 18.66 | 15.15 | 12.31 | 26.35 | 23.53 | 21.22 |
| | IMVTSC-MVI [47] | 22.49 | 21.84 | 19.17 | 17.87 | 16.15 | 13.00 | 24.37 | 23.98 | 21.70 |
| | IMVC-CBG [17] | 19.29 | 16.34 | 14.69 | 16.20 | 13.30 | 10.56 | 22.82 | 19.61 | 18.39 |
| | FCMVC-IV [39] | 16.65 | 16.33 | 16.76 | 11.70 | 11.17 | 11.59 | 20.49 | 19.92 | 20.51 |
| | DIVIDE [51] | 22.04 | 19.31 | 16.27 | 18.57 | 16.01 | 11.50 | 25.09 | 22.86 | 19.66 |
| | CPSPAN [52] | 17.91 | 15.02 | 21.69 | 18.04 | 14.95 | 21.33 | 16.86 | 13.34 | 20.58 |
| | Ours | **39.82** | **33.40** | **29.57** | **53.54** | **43.94** | **42.73** | **40.69** | **34.48** | **29.65** |
| Animal [50] | DAIMC [36] | 15.03 | 15.05 | 13.85 | 11.94 | 11.18 | 9.20 | 18.53 | 18.49 | 16.71 |
| | UEAF [32] | 18.52 | 17.51 | 14.99 | 16.09 | 14.00 | 11.28 | 22.30 | 21.08 | 17.81 |
| | FLSD [37] | 16.03 | 15.56 | 14.77 | 13.42 | 11.87 | 10.53 | 19.61 | 18.92 | 18.28 |
| | PIMVC [33] | **21.33** | 20.21 | 18.47 | **17.60** | 16.28 | 13.52 | **23.85** | 22.08 | 20.95 |
| | HCLS-CGL [20] | 15.74 | 15.36 | 13.44 | 11.51 | 10.81 | 8.53 | 17.69 | 17.25 | 15.27 |
| | AGC-IMC [41] | 18.87 | 17.94 | 16.29 | 15.47 | 13.78 | 11.29 | 22.35 | 21.16 | 19.49 |
| | IMVTSC-MVI [47] | 12.69 | 11.49 | 10.08 | 5.47 | 3.53 | 3.34 | 13.59 | 12.55 | 12.11 |
| | IMVC-CBG [17] | 16.74 | 15.05 | 14.51 | 12.71 | 11.04 | 9.87 | 20.22 | 18.36 | 17.89 |
| | FCMVC-IV [39] | 15.55 | 14.14 | 12.79 | 11.03 | 8.46 | 6.85 | 19.08 | 17.01 | 15.70 |
| | DIVIDE [51] | 16.10 | 16.33 | 13.81 | 12.14 | 11.87 | 9.55 | 17.00 | 19.69 | 17.00 |
| | CPSPAN [52] | 13.04 | 8.46 | 15.33 | 13.70 | 9.19 | 15.95 | 12.59 | 8.58 | 15.33 |
| | Ours | 19.40 | **23.15** | **26.24** | 14.13 | **17.13** | **23.88** | 21.35 | **24.63** | **29.27** |
| FMNIST [19] | DAIMC [36] | 20.52 | 26.08 | **32.11** | 10.81 | 8.60 | **12.12** | 21.45 | 26.62 | **32.13** |
| | UEAF [32] | - | - | - | - | - | - | - | - | - |
| | FLSD [37] | - | - | - | - | - | - | - | - | - |
| | PIMVC [33] | - | - | - | - | - | - | - | - | - |
| | HCLS-CGL [20] | - | - | - | - | - | - | - | - | - |
| | AGC-IMC [41] | - | - | - | - | - | - | - | - | - |
| | IMVTSC-MVI [47] | - | - | - | - | - | - | - | - | - |
| | IMVC-CBG [17] | 22.76 | 22.33 | 24.00 | 6.58 | 4.83 | 8.26 | 22.83 | 22.33 | 24.33 |
| | FCMVC-IV [39] | 17.87 | 19.85 | 23.58 | 3.06 | 3.83 | 4.92 | 19.00 | 20.16 | 23.81 |
| | DIVIDE [51] | 10.55 | 10.53 | 10.55 | 3.00 | 3.00 | 3.00 | 10.59 | 10.58 | 10.61 |
| | CPSPAN [52] | 13.86 | 14.73 | 17.14 | 2.21 | 1.90 | 8.46 | 15.07 | 15.73 | 18.50 |
| | Ours | **37.32** | **31.92** | 26.49 | **22.52** | **14.35** | 8.48 | **38.74** | **33.05** | 27.37 |

approach significantly enhances clustering results compared to matrix factorization-based methods like UEAF and FLSD, with improvements in ACC, NMI, and Purity exceeding 40%. Compared to graph-based methods such as AGC-IMC and IMVC-CBG, our approach outperforms them approximately by 50% across all metrics.

4) Our method effectively addresses large-scale dataset and demonstrates promising performance. Taking the FMNIST dataset as an example, some methods such as PIMVC, UEAF, and HCLS-CGL fail due to memory limitations. Our method achieves optimal performance in terms of ACC, NMI, and Purity metircs at missing ratios of 10% and 50%. The DAIMC method performs best when the missing ratio is 90%, while our method ranks second. This may be because of DAIMC's consideration of the alignment of basis matrices and feature extraction. This approach aligns well with the underlying structure of this grayscale image dataset, which exhibits low intra-class variance.

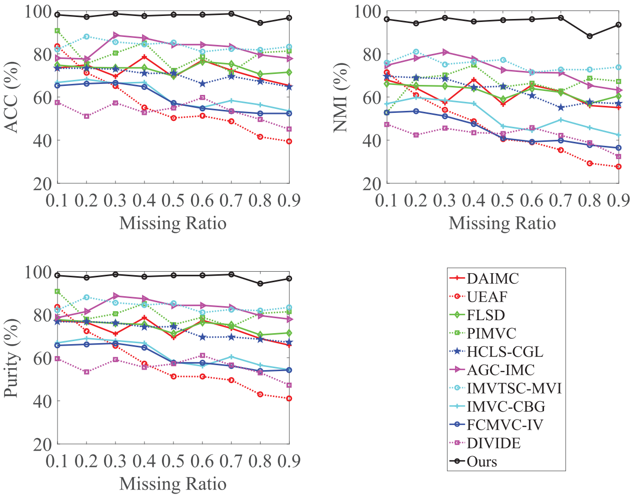


Fig. 4. The clustering results for the ACC, NMI, and Purity metrics on the MSRCv1 dataset with varying missing ratios.

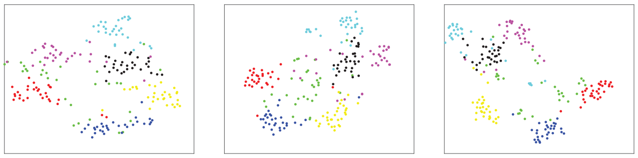
To further evaluate the effectiveness of the proposed DFRTMF model, we vary the missing ratios from 10% to 90% with an interval 10% on the MSRCv1 dataset. The result curves are shown in Fig. 4. We can observe that our method consistently outperforms other methods and demonstrates stable performance across all missing ratio settings.

Fig. 5 demonstrates the distribution of the learned features, using the MSRCv1 dataset with a 10% missing ratio as an

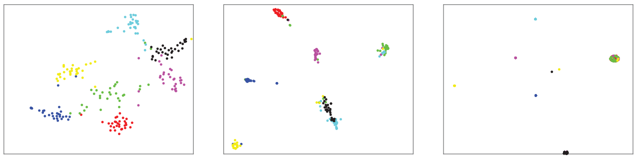
TABLE V

TIME COMPLEXITY AND RUNNING TIME (T/S) COMPARISONS OF DIFFERENT METHODS ON BENCHMARK DATASETS. '-' MEANS OUT OF CPU MEMORY

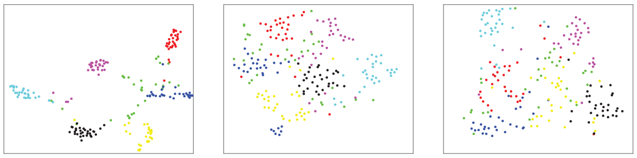
| Methods | Time Complexity | Running Times | | | | | |
|---|---|---|---|---|---|---|---|
| | | MSRCv1 | BDGP | Caltech101-7 | CCV | Animal | FMNIST |
| DAIMC | $\mathcal{O}(Vd^3t + nkdt)$ | 10.12 | 29.55 | 105.36 | 85.64 | 1548.00 | 9766.10 |
| UEAF | $\mathcal{O}(n^3t + dk^2t)$ | 4.32 | 97.29 | 54.65 | 2224.50 | 4264.50 | - |
| FLSD | $\mathcal{O}(n^2dt + dk^2t)$ | 4.13 | 125.92 | 49.33 | 752.19 | 6904.90 | - |
| PIMVC | $\mathcal{O}(n^3 + dk^2t)$ | 0.87 | 1.45 | 1.36 | 6.79 | 27.88 | - |
| HCLS-CGL | $\mathcal{O}(n^2kt)$ | 0.54 | 23.12 | 18.66 | 415.45 | 1060.70 | - |
| AGC-IMC | $\mathcal{O}(Vn^2kt)$ | 50.13 | 2820.64 | 1537.82 | 17313.54 | 29617.01 | - |
| IMVTSC-MVI | $\mathcal{O}((Vn^3 + Vn^2log(n) + V^2n^2)t)$ | 32.20 | 999.03 | 856.24 | 2383.80 | 18329.16 | - |
| IMVC-CBG | $\mathcal{O}(n(km + dk + dm)t + mdkt))$ | 5.07 | 1.01 | 2.51 | 5.19 | 24.91 | 64.39 |
| FCMVC-IV | $\mathcal{O}((Vnk^2 + Vk^3 + dk^2)t)$ | 0.19 | 1.74 | 1.95 | 9.11 | 91.29 | 2422.10 |
| DIVIDE | $\mathcal{O}(nwt)$ | 67.43 | 77.64 | 76.25 | 141.47 | 333.82 | 1005.23 |
| CPSPAN | $\mathcal{O}(nwt)$ | 20.23 | 330.91 | 1525.36 | 859.16 | 530.66 | 16375.24 |
| Ours | $\mathcal{O}(((ndlk + knVlog(n) + V^2kn)t))$ | 10.71 | 33.86 | 104.34 | 11.85 | 570.62 | 634.85 |



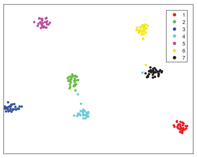
(a) DAIMC [36] (b) UEAF [32] (c) FLSD [37]



(d) PIMVC [33] (e) HCLS-CGL [20] (f) AGC-IMC [41]



(g) IMVTSC-MVI [48] (h) IMVC-CBG [17] (i) FCMVC-IV [39]



(j) Ours

Fig. 5. The t-SNE visualization of ten IMVC methods on MSRCv1 dataset with 10% missing ratios missing.

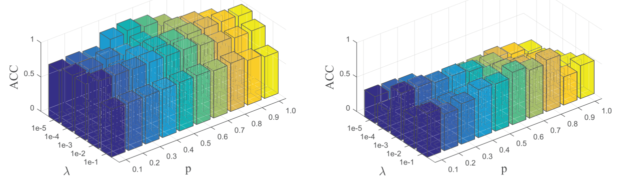
example. Specifically, we use a popular visualization method, t-SNE [53], to display the features learned by the different models. Our method demonstrates a larger inter-class distance and a smaller intra-class distance, ensuring that data points within the same cluster are tightly packed and enhancing the separation between different clusters. The measured intra-class distances are 3.32, 3.30, 2.69, 2.21, 4.25, 3.29, 2.05, 5.13, 5.40, and 0.87, respectively. This indicates that the feature representation learned by our method is more discriminative than those learned by other methods, allowing our model to achieve better clustering results. Furthermore, IMVC methods that enable missing view recovery may show clearer clustering structures than others. This suggests that exploring unobserved views can uncover the intrinsic geometric structure of the data and enhance clustering performance.

### C. Computational Efficiency Analysis

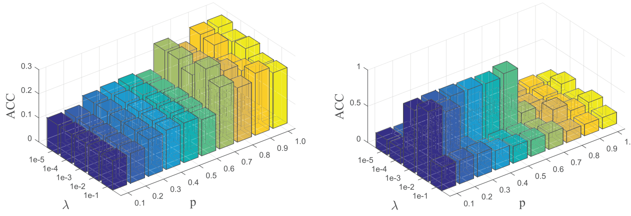
Table V presents the computational complexity and running time comparisons. Here, $d$, $n$, $k$, $V$, $m$, $w$ and $l$ represent the total data dimension, number of samples, number of clusters, number of views, number of anchors, number of network parameters, and projection dimension, respectively. As can be observed, DAIMC, PIMVC, IMVC-CBG, and FCMVC-IV show the linear time complexity for the number of samples, while FLSD, HCLS-CGL,and AGC-IMC exhibit quadratic time complexity. UEAF and IMVTSC-MVI present cubic time complexity for the number of samples. Our method demonstrates efficiency that lies between linear and quadratic time complexity for the number of samples. Additionally, the proposed DFRTMF demonstrates competitive performance in running times. For the large-scale dataset FMNIST, methods with quadratic time complexity fail to execute. Our method takes 634.85 seconds to achieve convergence, demonstrating its efficiency potential.

### D. Parameter Analysis

To evaluate the influence of parameter settings, we conduct corresponding experiments by varying the parameter combinations. Fig. 6 shows the 3-D histograms of ACC value evolution concerning different parameter combinations on the BDGP, Caltech101-7, CCV, and MSRCv1 datasets. The results indicate that parameter settings significantly affect the ACC metric. Specifically, for the BDGP dataset, the performance of our model is close to optimality when $\lambda \in [10^{-4}, 10^{-3}, 10^{-2}]$ and it is not sensitive to $p$. For Caltech101-7 dataset, the model performance is close to optimality when $\lambda \in [10^{-3}, 10^{-2}, 10^{-1}]$ and $p \in [0.1 : 0.1 : 0.8]$. For CCV dataset, the model performance is close to optimality when $p \in [0.7 : 0.1 : 1.0]$, and it is not sensitive to $\lambda$. The MSRCv1 dataset exhibits complex parameter sensitivity. However, our model can still achieve high ACC values when $\lambda \in [10^{-3}, 10^{-2}]$ and $p \in [0.1 : 0.1 : 0.6]$. Since different datasets contains different properties, the optimal range for model parameter selection

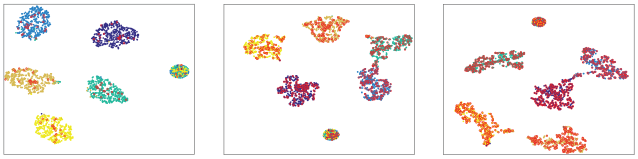


(a) BDGP (b) Caltech-7



(c) CCV (d) MSRCv1

Fig. 6. Sensitivity analysis: ACC values with different $\lambda$ and $p$ on four datasets with missing ratio 90%.



(a) 0.1MR (b) 0.5MR (c) 0.9MR

Fig. 7. The t-SNE visualization of the recovered missing views in the BDGP dataset at three missing ratios. Red circle indicates the recovered features.

varies across these datasets. Nonetheless, it remains feasible to determine the appropriate parameter settings through grid searching. Note that we use a coarse scale for the parameter settings of $\lambda$ during evaluations for simplicity. Finer parameter settings can yield more stable performance. Moreover, our model demonstrates robustness in exploring spatial structure by tuning the parameter $p$.

## *E. The Missing View Recovery*

This subsection assesses the recovery capability of the proposed model for missing views. Taking the BDGP dataset as an example, we apply the t-SNE technique to visualize the recovered discriminative features. As shown in Fig. 7, we select the first view of the BDGP dataset and evaluate different missing ratios: 0.1, 0.5, and 0.9. Red circles indicate the recovered missing views. It is evident that the proposed model effectively recovers the missing views. At a missing ratio of 0.1, the recovered features are compact, and the inter-class distance is 48.86, indicating significant separation between clusters. However, as the missing ratio increases to 0.9, the inter-class distance decreases to 47.67, weakening the separations between clusters. This may be due to limited number of valid samples, which fails to provide sufficient information for effective discriminative subspace learning. To the best of our knowledge, there are no evaluation metrics specifically designed to assess the recovery capability of missing views. It is a common choice to use clustering performance to evaluate the recovery capability. The corresponding accuracy values are 100%, 99.68%, and 97.07% for missing ratios of 0.1, 0.5, and 0.9 on the BDGP dataset, respectively.

TABLE VI

Ablation Study of Our Method on Two Datasets

| Comparison Experiments | | ACC(%) | | |
|---|---|---|---|---|
| Datasets | Methods | 0.1 | 0.5 | 0.9 |
| BDGP | DFRTMF w/o PL | 72.60 | 66.92 | 56.60 |
| | DFRTMF w/o TL | 46.80 | 40.56 | 35.60 |
| | DFRTMF w/o OV | 98.60 | 98.12 | 97.44 |
| | DFR_NMF | **98.76** | **99.64** | **98.12** |
| Caltech7 | DFRTMF w/o PL | 74.08 | 68.59 | 68.18 |
| | DFRTMF w/o TL | 62.14 | 53.32 | 51.43 |
| | DFRTMF w/o OV | 82.84 | 69.95 | 71.23 |
| | DFRTMF | **83.85** | **75.37** | **76.12** |

## *F. Ablation Study*

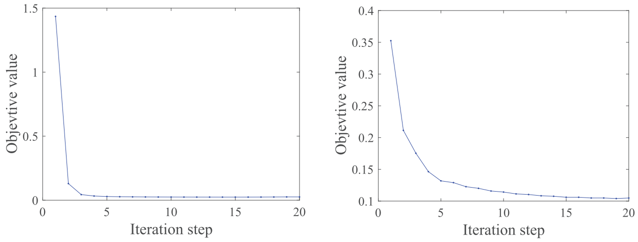
Discriminative projection learning and soft indicator matrix tensor learning are primary contributions of our model. To further assess the effect of these two techniques, we devise three degraded models. “DFRTMF w/o PL” indicates the model removes the discriminative projection learning. “DFRTMF w/o TL” represents the model removes the soft indicator matrix tensor learning. “DFRTMF w/o OV” represents the model that uses scatter matrices solely from observed views. Specifically, The DFRTMF w/o PL model is formulated as follows:

$$\begin{aligned}\min_{\Omega}\quad & \sum_{v=1}^{V} a_v^2 \left\|\mathbf{H}_v - \mathbf{G}_v\mathbf{F}_v^\top\right\|_F^2 + \beta\|\boldsymbol{\mathcal{F}}\|_{\circledS p}^p \\ \text{s.t.}\quad & a_v \geq 0, \mathbf{a}^\top\mathbf{1} = \mathbf{1}. \\ & \mathbf{G}_v^\top\mathbf{G}_v = \mathbf{I}, \mathbf{F}_v \geq 0, (\mathbf{H}_v)_{\mathbb{T}_v} = (\mathbf{X}_v)_{\mathbb{T}_v} \end{aligned} \tag{25}$$

The “DFRTMF w/o TL” model is formulated as follows:

$$\begin{aligned}\min_{\Omega}\quad & \sum_{v=1}^{V} a_v^2 \left\|\mathbf{P}_v\mathbf{H}_v - \mathbf{G}_v\mathbf{F}_v^\top\right\|_F^2 + \lambda\|\mathbf{P}_v\|_F^2 + \beta\left\|\mathbf{F}_v - \hat{\mathbf{F}}\right\|_F^2 \\ \text{s.t.}\quad & \mathbf{P}_v\mathbf{S}_t^v\mathbf{P}_v^\top = \mathbf{I}, a_v \geq 0, \mathbf{a}^\top\mathbf{1} = \mathbf{1}. \\ & \mathbf{G}_v^\top\mathbf{G}_v = \mathbf{I}, \mathbf{F}_v \geq 0, (\mathbf{H}_v)_{\mathbb{T}_v} = (\mathbf{X}_v)_{\mathbb{T}_v}. \end{aligned} \tag{26}$$

As shown in Table VI, DFRTMF outperforms two degraded models on both the BDGP and Caltech101-7 datasets. For example, on the Caltech101-7 dataset with a 10% missing ratio, DFRTMF outperforms DFRTMF w/o PL by 9.77% and exceeds DFRTMF w/o TL by 21.71% in terms of the ACC metric. This demonstrates that discriminative projection learning effectively captures the subspace, enhancing intra-view discriminatibility. Furthermore, low-rank tensor learning applied to soft indicator matrices effectively explores spatial correlations. The Schatten p-norm may ensure that each frontal slice exhibits similar cluster assignments under a relaxed divergence constraint, enhancing the exploration of multi-view complementary information and improving clustering performance. Additionally, DFRTMF outperforms DFRTMF w/o OV by 1.01%, 5.42%, and 4.89% in terms of the ACC metric on the Caltech101-7 dataset under varying missing ratios, demonstrating the effectiveness of the global scatter matrices.



(a) BDGP (b) MSRCv1

Fig. 8. Evolution of the objective value with respective to iteration steps on BDGP and MSRCv1 datasets.

### G. Convergence Study

This subsection verifies the empirical convergence of the proposed method. Our model alternatively updates the variables by minimizing the objective function in Eq. (4). This allows us to calculate the objective value across different iterations. Using BDGP and MSRCv1 as examples, we plotted the objective function values over iterations ranging from 1 to 20, as shown in Fig 8. These experimental results substantiate the convergence of the proposed approach. We computed the convergence times for the MSRCv1 and BDGP datasets, yielding running times of 10.71 seconds and 33.86 seconds, respectively.

## V. CONCLUSION

This paper proposes a novel incomplete multi-view clustering method called DFRTMF, which effectively recovers missing views, learns low-dimensional discriminative embeddings, and facilitates direct clustering. DFRTMF enhances the non-negative matrix factorization through projection learning to produce soft indicators. A global uncorrelated constraint based on the scatter matrix of the complete data is devised, exploring the correlations between observed and missing views. Tensor Schatten p-norm regularization captures high-order correlations among views, while an adaptive weighting strategy balances the contributions of different views. We propose an alternating optimization algorithm based on the ADMM to solve the proposed objective function effectively. Extensive experiments demonstrate the effectiveness of DFRTMF compared to the state-of-the-art methods. Our future work is to develop continual learning methods for scenarios where data observations of new views are accumulated over time.

## REFERENCES

[1] U. Fang, M. Li, J. Li, L. Gao, T. Jia, and Y. Zhang, "A comprehensive survey on multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12350–12368, Dec. 2023.

[2] G. Jiang, J. Peng, H. Wang, Z. Mi, and X. Fu, "Tensorial multi-view clustering via low-rank constrained high-order graph learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 8, pp. 5307–5318, Aug. 2022.

[3] W. Zhu, J. Lu, and J. Zhou, "Structured general and specific multi-view subspace clustering," *Pattern Recognit.*, vol. 93, pp. 392–403, Sep. 2019.

[4] X. Wang, Y. Zhang, and Y. Zhou, "Pseudo-supervision affinity propagation for efficient and scalable multiview clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 1, 2025, doi: 10.1109/TNNLS.2025.3545465.

[5] Y. Zhang, S. Yan, L. Zhang, and B. Du, "Fast projected fuzzy clustering with anchor guidance for multimodal remote sensing imagery," *IEEE Trans. Image Process.*, vol. 33, pp. 4640–4653, 2024.

[6] Y. Zhang, Y. Wang, X. Chen, X. Jiang, and Y. Zhou, "Spectral–spatial feature extraction with dual graph autoencoder for hyperspectral image clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8500–8511, Dec. 2022.

[7] Y. Zhang, J. Qi, X. Wang, Z. Cai, J. Peng, and Y. Zhou, "Tensorial global–local graph self-representation for hyperspectral band selection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 12, pp. 13213–13225, Dec. 2024.

[8] Y. Zhang, X. Wang, X. Jiang, L. Zhang, and B. Du, "Elastic graph fusion subspace clustering for large hyperspectral image," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Feb. 5, 2025, doi: 10.1109/TCSVT.2025.3538926.

[9] X. Wang, Y. Zhang, and Y. Zhou, "Multimodal remote sensing image clustering with multiscale spectral–spatial anchor graphs," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 4405612.

[10] J. Li, C. Xu, W. Yang, C. Sun, J. Xu, and H. Zhang, "Discriminative multi-view privileged information learning for image re-ranking," *IEEE Trans. Image Process.*, vol. 29, pp. 3490–3505, 2020.

[11] L. Hua, Y. Gu, X. Gu, J. Xue, and T. Ni, "A novel brain MRI image segmentation method using an improved multi-view fuzzy c-Means clustering algorithm," *Frontiers Neurosci.*, vol. 15, Mar. 2021, Art. no. 662674.

[12] C. Cui et al., "A novel approach for effective multi-view clustering with information-theoretic perspective," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2023, pp. 44847–44859.

[13] Y. Sun, Y. Qin, Y. Li, D. Peng, X. Peng, and P. Hu, "Robust multi-view clustering with noisy correspondence," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 12, pp. 9150–9162, Dec. 2024.

[14] Y. Gan et al., "Multi-view clustering via multi-stage fusion," *IEEE Trans. Multimedia*, early access, Jan. 27, 2025, doi: 10.1109/TMM.2025.3535360.

[15] W. Zhu, J. Lu, and J. Zhou, "Nonlinear subspace clustering for image clustering," *Pattern Recognit. Lett.*, vol. 107, pp. 131–136, May 2018.

[16] Z. Zhao, T. Wang, H. Xin, R. Wang, and F. Nie, "Multi-view clustering via high-order bipartite graph fusion," *Inf. Fusion*, vol. 113, Jan. 2025, Art. no. 102630.

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

[18] C. Mulert and L. Lemieux, *EEG-fMRI: Physiological Basis, Technique, and Applications*. Cham, Switzerland: Springer, 2023.

[19] X. Wang, Y. Zhang, and Y. Zhou, "Highly efficient rotation-invariant spectral embedding for scalable incomplete multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, 2025, vol. 39, no. 20, pp. 21312–21320.

[20] J. Wen et al., "Highly confident local structure based consensus graph learning for incomplete multi-view clustering," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 15712–15721.

[21] X.-L. Li, M.-S. Chen, C.-D. Wang, and J.-H. Lai, "Refining graph structure for incomplete multi-view clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 2300–2313, Feb. 2022.

[22] Y. Zhang, X. Liu, S. Wang, J. Liu, S. Dai, and E. Zhu, "One-stage incomplete multi-view clustering via late fusion," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 2717–2725.

[23] M. C. Silva, M. Dahaghin, M. Toso, and A. Del Bue, "Contrastive Gaussian clustering: Weakly supervised 3D scene segmentation," 2024, *arXiv:2404.12784*.

[24] U. Jain, A. Mirzaei, and I. Gilitschenski, "GaussianCut: Interactive segmentation via graph cut for 3D Gaussian splatting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 37, 2025, pp. 89184–89212.

[25] Q. Shen, X. Zhang, S. Wang, Y. Li, Y. Liang, and Y. Chen, "Dual completion learning for incomplete multi-view clustering," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 9, no. 1, pp. 455–467, Feb. 2025.

[26] W. Hao, S. Pang, X. Bai, and J. Xue, "Tensor-based incomplete multi-view clustering with low-rank data reconstruction and consistency guidance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 12, pp. 7156–7169, Dec. 2023.

[27] J. Liu et al., "Self-representation subspace clustering for incomplete multi-view data," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 2726–2734.

[28] S. Zhao, J. Wen, L. Fei, and B. Zhang, "Tensorized incomplete multi-view clustering with intrinsic graph completion," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, 2023, pp. 11327–11335.

[29] G.-Y. Zhang, D. Huang, and C.-D. Wang, "Unified and tensorized incomplete multi-view kernel subspace clustering," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 8, no. 2, pp. 1550–1566, Apr. 2024.
[30] S. Liu, X. Liu, S. Wang, X. Niu, and E. Zhu, "Fast incomplete multi-view clustering with view-independent anchors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 6, pp. 1–12, Jun. 2024.
[31] H. Tao, C. Hou, D. Yi, J. Zhu, and D. Hu, "Joint embedding learning and low-rank approximation: A framework for incomplete multiview learning," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1690–1703, Mar. 2021.
[32] J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, and H. Liu, "Unified embedding alignment with missing views inferring for incomplete multi-view clustering," in *Proc. 33rd Conf. Artif. Intell. (AAAI)*, Jan. 2019, pp. 5393–5400.
[33] S. Deng, J. Wen, C. Liu, K. Yan, G. Xu, and Y. Xu, "Projective incomplete multi-view clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 8, pp. 1–13, Aug. 2024.
[34] D. Wang, S. Han, Q. Wang, L. He, Y. Tian, and X. Gao, "Pseudo-label guided collective matrix factorization for multiview clustering," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 8681–8691, Sep. 2022.
[35] H. Zhao, H. Liu, and Y. Fu, "Incomplete multi-modal visual data grouping," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2392–2398.
[36] M. Hu and S. Chen, "Doubly aligned incomplete multi-view clustering," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 2262–2268.
[37] J. Wen, Z. Zhang, Z. Zhang, L. Fei, and M. Wang, "Generalized incomplete multiview clustering with flexible locality structure diffusion," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 101–114, Jan. 2021.
[38] C. Liu, Z. Wu, J. Wen, Y. Xu, and C. Huang, "Localized sparse incomplete multi-view clustering," *IEEE Trans. Multimedia*, vol. 25, pp. 5539–5551, 2022.
[39] X. Wan, B. Xiao, X. Liu, J. Liu, W. Liang, and E. Zhu, "Fast continual multi-view clustering with incomplete views," *IEEE Trans. Image Process.*, vol. 33, pp. 2995–3008, 2024.
[40] J. Yin and S. Sun, "Incomplete multi-view clustering with reconstructed views," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, pp. 2671–2682, Mar. 2023.
[41] J. Wen et al., "Adaptive graph completion based incomplete multi-view clustering," *IEEE Trans. Multimedia*, vol. 23, pp. 2493–2504, 2021.
[42] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.
[43] C. Zhang, H. Li, W. Lv, Z. Huang, Y. Gao, and C. Chen, "Enhanced tensor low-rank and sparse representation recovery for incomplete multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2023, vol. 37, no. 9, pp. 11174–11182.
[44] X. Zhao, Q. Shen, Y. Chen, Y. Liang, J. Chen, and Y. Zhou, "Self-completed bipartite graph learning for fast incomplete multi-view clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2166–2178, Apr. 2024.
[45] T. Wu, S. Feng, and J. Yuan, "Low-rank kernel tensor learning for incomplete multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, 2024, pp. 15952–15960.
[46] Z. Long, C. Zhu, P. Comon, Y. Ren, and Y. Liu, "Feature space recovery for efficient incomplete multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 9, pp. 4664–4677, Sep. 2024.
[47] J. Wen et al., "Unified tensor framework for incomplete multi-view clustering and missing-view inferring," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 10273–10281.
[48] Q. Gao, P. Zhang, W. Xia, D. Xie, X. Gao, and D. Tao, "Enhanced tensor RPCA and its application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 2133–2140, Jun. 2021.
[49] W. Zhuge, C. Hou, X. Liu, H. Tao, and D. Yi, "Simultaneous representation learning and clustering for incomplete multi-view data," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 4482–4488.
[50] J. Wen et al., "A survey on incomplete multiview clustering," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 2, pp. 1136–1149, Feb. 2023.
[51] Y. Lu, Y. Lin, M. Yang, D. Peng, P. Hu, and X. Peng, "Decoupled contrastive multi-view clustering with high-order random walks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, 2024, pp. 14193–14201.
[52] J. Jin, S. Wang, Z. Dong, X. Liu, and E. Zhu, "Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 11600–11609.
[53] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.

**Xinxin Wang** is currently pursuing the Ph.D. degree with the Department of Computer and Information Science, University of Macau. His research interests include multi-view learning and hyperspectral image processing.

**Yongshan Zhang** (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees in computer science from China University of Geosciences, Wuhan, China, in 2014 and 2019, respectively.

She has been a member of the BDSC Laboratory, University of Illinois at Chicago, and the VIP Laboratory, University of Macau. She is currently an Associate Professor with the School of Computer Science, China University of Geosciences. Her current research interests include machine learning and remote sensing image analysis.

**Jie Zhang** received the B.S. degree in electronic information engineering from Wuhan University, Wuhan, China, in 2020, and the M.S. degree in computer science from the University of Macau, Macau, in 2023, where he is currently pursuing the Ph.D. degree. His research interests include hyperspectral image processing and quantum-inspired deep learning.

**Yicong Zhou** (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Tufts University, Medford, MA, USA, in 2010.

He is a Professor with the Department of Computer and Information Science, University of Macau, Macau, China. His research interests include image processing, computer vision, and artificial intelligence.

Dr. Zhou is a fellow of the Society of Photo-Optical Instrumentation Engineers (SPIE) and was recognized as one of "Highly Cited Researchers" in 2020, 2021, 2023, and 2024. He serves as a Senior Area Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and an Associate Editor for IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.