# Feature space recovery for incomplete multi-view clustering

Zhen Long, Ce Zhu, Pierre Comon, Yipeng Liu

# FEATURE SPACE RECOVERY FOR INCOMPLETE MULTI-VIEW CLUSTERING

*Zhen Long*[*], *Ce Zhu*[*], *Pierre Comon*[†], *Yipeng Liu*[*]

[*]University of Electronic Science and Technology of China, Chengdu, 611731, China
[†]Gipsa-Lab, CNRS, Univ. Grenoble Alpes, Grenoble INP, 38000 Grenoble, France

## ABSTRACT

Incomplete multi-view clustering (IMVC), based on imputation and clustering unification, has received wide attention due to its ability to exploit hidden information from missing views. However, current methods mainly consider inter/intra-view correlations, ignoring the structural information of sample features within views. In this paper, we propose a feature space recovery based IMVC method, where low-rank feature space recovery and consensus representation learning of inter/intra-views are considered into a unified framework. Moreover, low-rank tensor ring approximation is used to capture the correlations of self-representation tensor. In an iterative way, the learned inter/intra-view correlations will guide the recovery of missing features, while the explored low-rank information from feature spaces will in turn facilitate self-representation learning, eventually achieving outstanding clustering performance. Experimental results show our method has a very significant improvement over known state-of-the-art algorithms in terms of ACC, NMI and Purity.

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

## 1. INTRODUCTION

Multi-view data which provides consensual and complementary information has sparked much interest in multi-view clustering (MVC) [1–3]. In practice, some views may suffer from missing samples during data acquisition or transmission [4]. The traditional MVC approaches may not perform well enough on incomplete data for some applications due to the ignorance of hidden information of missing samples.

To tackle the above problem, many incomplete multi-view clustering (IMVC) algorithms have been proposed [5,6]. Among them, graph-based IMVC ones have attracted much attention since they can better extract different graphs indicating the memberships among samples. For instance, Wen et al. [7] simultaneously consider graph learning and spectral clustering into one unified framework to learn the consensus representation for IMVC. However, the work in [7] has two limitations: 1) Only relationships between observed samples are involved in consensus representation learning. 2) The learning of each view is separate and does not take inter-view similarity structure, which greatly reduces the advantages of multi-view data. Based on this observation, a series of tensor-based multi-view subspace representation methods have emerged to infer missing samples while clustering [8–12], which all consider tensor singular value decomposition (t-SVD) [13] to explore the high-order correlations among different views.

The aforementioned tensor-based IMVC methods may suffer from the following two limitations: 1) The t-SVD cannot well explore inter/intra-view correlations simultaneously since matrix SVDs are only performed in the first two modes while linear transformations are taken in the third mode. 2) For IMVC, the sample features from missing views are crucial, but the existing approaches only consider the correlations across/within views to infer missing features, ignoring the structural correlations of features. In fact, the feature space within views is highly redundant [14].

In this paper, we consider such redundant information of feature space as a low-rank prior. In addition, tensor ring (TR) decomposition [15–18] is used to capture the high-order relatedness in multi-view data due to its highly expressive and powerful representation ability. Building on them, we propose a feature space recovery based incomplete multi-view clustering framework (FSR-IMVC), as shown in Fig. 1. FSR-IMVC considers low-rank matrix learning based feature space recovery and low-rank TR approximation based self-representations exploration into a unified framework. Specifically, the latent feature subspace $\mathbf{H}_v$ can simultaneously learn the correlations of inter/intra-view and sample features by the updated $\mathbf{Z}_v$ and the low-rank structure of $\mathbf{X}_v$, for recovering feature space $\mathbf{X}_v$. In turn, the self-representation tensor $\mathcal{Z}$ adaptively updates from $\mathbf{H}_v, v = 1, \cdots, V$ and low-rank TR approximation to obtain better consensus representation. Finally, the self-representation tensor is used to construct the affinity matrix for the spectral clustering algorithm.

To validate that the low-rank prior of features can uncover more hidden information about missing samples, we give the feature recovery results in Fig. 2. We can see the facial features of the proposed method are well recovered, which will be beneficial to achieve better clustering results. Exper-

Supported by the National Natural Science Foundation of China (NSFC) under Grant U19A2052, Grant 62020106011 and Grant 62171088.

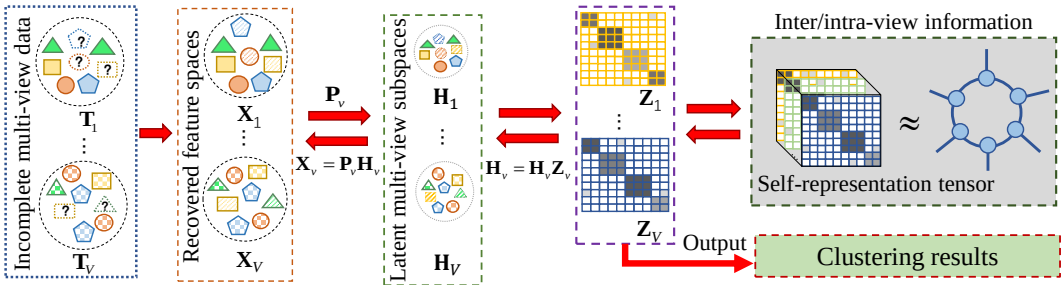



**Fig. 1**: The framework of FSR-IMVC. The observed data $\mathbf{T}_v$ can be firstly reconstructed by assuming its feature space is low-rank, e.g. $\mathbf{X}_v = \mathbf{P}_v\mathbf{H}_v$, where $\mathbf{P}_v^T\mathbf{P}_v = \mathbf{I}$ and $\|\mathbf{H}_v\|_{\mathrm{F}}$ minimization are applied to make $\mathbf{X}_v$ low-rank. Then, the self-representation tensor is adaptively updated from $\mathbf{H}_v, v = 1, \cdots, V$ and low-rank TR approximation. In turn, $\mathbf{H}_v$ learns from the updated $\mathbf{Z}_v$ and low rank prior of feature space to recover $\mathbf{X}_v$, eventually improving the clustering performance.

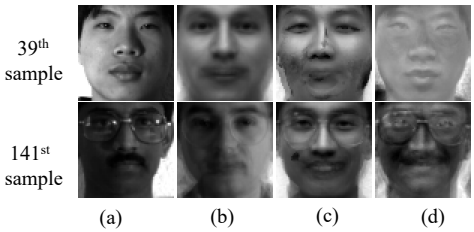



**Fig. 2**: The feature recovery performance for the 2nd view on Yale data with 50% missing ratio;(a) original features;(b)-(d) recovered features by HCPIMSC [10], IMVTSC-MVI [12] and our method, respectively.

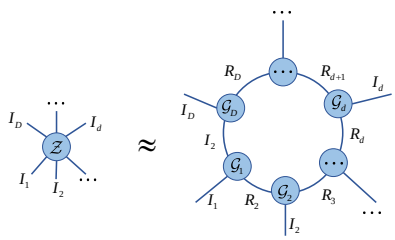


**Fig. 3**: The graphical illustration for TR decomposition.

imental results on four multi-view data demonstrate that the clustering performance of our method has a significant improvement against that of state-of-the-art algorithms in terms of ACC, NMI, and Purity.

## 2. PRELIMINARIES

In this paper, a scalar, a vector, a matrix and a tensor are written as $x$, $\mathbf{x}$, $\mathbf{X}$, and $\mathcal{X}$, respectively. Indices typically range from 1 to their capital version, e.g., $i = 1, \cdots, I$.

**TR Decomposition** [16]. For a $D$th-order tensor $\mathcal{Z} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$, its TR decomposition 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 $\mathcal{Z} = \mathfrak{R}(\mathcal{G}_1, \cdots, \mathcal{G}_D)$ and graphically illustrated in Fig. 3.

**Lemma 1.** *[19]* $\|\mathbf{X}\|_* = \min_{\mathbf{X}=\mathbf{PH}} \frac{1}{2}(\|\mathbf{P}\|_{\mathrm{F}}^2 + \|\mathbf{H}\|_{\mathrm{F}}^2)$. $\|\mathbf{X}\|_*$ *denotes the nuclear norm of* $\mathbf{X}$*, which is the sum of its singular values.*

**Corollary 1.** *From Lemma 1, 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* $\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}$.

## 3. PROPOSED METHOD

### 3.1. Model Development

As shown in Fig. 1, feature space recovery (FSR) and self-representation are learned in a unified framework. According to Corollary 1, the low-rank properties of the recovered feature spaces $\mathbf{X}_v$ can be depicted by

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

$\mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v = \mathbf{I}$ means the feature spaces are projected onto discriminating subspaces [20]. Moreover, low-rank TR approximation is considered to explore the similarity structure of the self-representation tensor, e.g., $\mathcal{Z} = \mathfrak{R}(\mathcal{G}_1, \cdots, \mathcal{G}_D)$. Benefiting from TR approximation, the inter/intra-view information in $\mathcal{Z}$ can be well captured simultaneously. Therefore, for incomplete multi-view data $\mathbf{T}_v \in \mathbb{R}^{D_v \times N}$, $v = 1, \cdots V$, where $V$ and $N$ are the numbers of views and instances, respectively; and $D_v$ represents the feature dimension in the $v$-th view. The proposed modeling can be formulated as the optimization problem:

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

where $\mathbf{E}_v = [\mathbf{E}_v^x; \mathbf{E}_v^h]$; $\mathcal{Z} = \Omega(\mathbf{Z}_1, \cdots, \mathbf{Z}_V)$; $\Omega$ is an operator which stacks all self-representation matrices $\mathbf{Z}_v, v = 1, \cdots, V$ to a 3-th order tensor $\mathcal{Z} \in \mathbb{R}^{N \times N \times V}$; $\ell_1$ is employed on error matrix $\mathbf{E}_v$ to remove sparse noise or outliers; and $\mathbb{O}_v$ is the index set of observed instances for $v$-th view.

### 3.2. Solution

The above optimization problem can be tackled using an ADMM framework [21]. To make problem (1) separable, $\mathcal{Y}$ is added as an auxiliary variable as follows:

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

The corresponding augmented Lagrangian function is formulated as:

$$\begin{aligned}&\mathrm{L}\left(\{\mathbf{Z}_v,\mathbf{E}_v,\mathbf{Q}_1^v,\mathbf{Q}_2^v,\mathbf{X}_v,\mathbf{P}_v,\mathbf{H}_v\}_{v=1}^V,\mathcal{Y},\mathcal{Q}_3\right)\\&=\sum_{v=1}^{V}(\frac{\gamma}{2}\|\mathbf{H}_v\|_{\mathrm{F}}^2+\lambda\|\mathbf{E}_v\|_1\\&+\langle\mathbf{Q}_1^v,\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v-\mathbf{E}_v^x\rangle+\langle\mathbf{Q}_2^v,\mathbf{H}_v-\mathbf{H}_v\mathbf{Z}_v-\mathbf{E}_v^h\rangle\\&+\frac{\rho_1}{2}\|\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v-\mathbf{E}_v^x\|_{\mathrm{F}}^2+\frac{\rho_2}{2}\|\mathbf{H}_v-\mathbf{H}_v\mathbf{Z}_v-\mathbf{E}_v^h\|_{\mathrm{F}}^2)\\&+\langle\mathcal{Q}_3,\mathcal{Z}-\mathcal{Y}\rangle+\frac{\rho_3}{2}\|\mathcal{Z}-\mathcal{Y}\|_{\mathrm{F}}^2,\end{aligned} \tag{3}$$

under constraints $\mathcal{Y}=\mathfrak{R}(\mathcal{G}_1,\cdots,\mathcal{G}_D)$, $(\mathbf{X}_v)_{\mathbb{O}_v}=(\mathbf{T}_v)_{\mathbb{O}_v}$ and $\mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v=\mathbf{I}, v=1,\cdots,V$, 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 (3) is split into several subproblems.

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

$$\begin{aligned}\mathbf{H}_v=&(\mathbf{P}_v^{\mathrm{T}}(\mathbf{Q}_1^v+\rho_1\mathbf{X}_v-\rho_1\mathbf{E}_v^x)+(\rho_2\mathbf{E}_v^h-\mathbf{Q}_2^v)(\mathbf{I}-\mathbf{Z}_v^{\mathrm{T}}))\\&((\gamma+\rho_1)\mathbf{I}+\rho_2(\mathbf{I}-\mathbf{Z}_v)(\mathbf{I}-\mathbf{Z}_v^{\mathrm{T}}))^{-1}\end{aligned} \tag{4}$$

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

$$\max_{\mathbf{P}_v:\mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v=\mathbf{I}} \quad \mathrm{trace}(\mathbf{P}_v(\mathbf{H}_v(\mathbf{Q}_1^v+\rho_1\mathbf{X}_v-\mathbf{E}_v^x)^{\mathrm{T}})). \tag{5}$$

This subproblem is a well-known orthogonal Procrustes problem. Letting $\mathbf{M}=\mathbf{H}_v(\mathbf{Q}_1^v+\rho_1\mathbf{X}_v-\mathbf{E}_v^x)^{\mathrm{T}}$, 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 values decomposition function.

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

$$\mathbf{X}_v(:,n)=\begin{cases}\hat{\mathbf{X}}_v(:,n), n\notin\mathbb{O}_v\\\mathbf{T}_v(:,n), n\in\mathbb{O}_v,\end{cases} \tag{6}$$

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

**Update $\{\mathbf{Z}_v\}_{v=1}^V$:** By setting the derivative of the objective function in (3) concerning $\mathbf{Z}_v$ to zero, the solution of $\mathbf{Z}_v$ can be obtained via:

$$\begin{aligned}\mathbf{Z}_v=&(\rho_3\mathbf{I}+\rho_2\mathbf{H}_v^{\mathrm{T}}\mathbf{H}_v)^{-1}(\mathbf{H}_v^{\mathrm{T}}(\mathbf{Q}_2^v+\rho_2\mathbf{H}_v-\rho_2\mathbf{E}_v^h)\\&+\Omega_v^{-1}(\rho_3\mathcal{Y}+\mathcal{Q}_3)),\end{aligned} \tag{7}$$

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

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

$$\begin{cases}\mathbf{E}_v^x=\mathrm{sth}(\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v+(1/\rho_1)\mathbf{Q}_1^v,\lambda/\rho_1)\\\mathbf{E}_v^h=\mathrm{sth}(\mathbf{H}_v-\mathbf{H}_v\mathbf{Z}_v+(1/\rho_2)\mathbf{Q}_2^v,\lambda/\rho_2),\end{cases} \tag{8}$$

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 (3) can be transferred into the following formulation:

$$\begin{aligned}&\min_{\mathcal{Y}} \quad \|\mathcal{Y}-(\mathcal{Z}-(1/\rho_3)\mathcal{Q}_3)\|_{\mathrm{F}}^2,\\&\text{s. t. } \mathcal{Y}=\mathfrak{R}(\mathcal{G}_1,\cdots,\mathcal{G}_D).\end{aligned} \tag{9}$$

It can be solved by TR-ALS algorithm in [16] , where the input tensor is $\mathcal{Z}-(1/\rho_3)\mathcal{Q}_3$.

**Update Lagrangian multipliers**:

$$\begin{cases}\mathbf{Q}_1^v=\mathbf{Q}_1^v+\rho_1(\mathbf{X}_v-\mathbf{P}_v\mathbf{H}_v-\mathbf{E}_v^x), v=1,\cdots,V\\\mathbf{Q}_2^v=\mathbf{Q}_2^v+\rho_2(\mathbf{H}_v-\mathbf{H}_v\mathbf{Z}_v-\mathbf{E}_v^h), v=1,\cdots,V\\\mathcal{Q}_3=\mathcal{Q}_3+\rho_3(\mathcal{Y}-\mathcal{Z})\end{cases} \tag{10}$$

After finishing updating aforementioned variables, the affinity matrix of multi-view data can be obtained by $\mathbf{S}=\frac{1}{V}\sum_{v=1}^{V}|\mathbf{Z}_v|+|\mathbf{Z}_v^{\mathrm{T}}|$, which will be used in spectral clustering algorithm [22] for the final clustering result.

## 4. EXPERIMENTAL RESULTS

### 4.1. Experimental Settings

**Datasets:** Four well-known multi-view datasets, including Coil20 [23], Yale[1], Caltech7[2], and BDGP[3], are chosen to evaluate the effectiveness of FSR-IMVC, where sample features are selected as suggested in [6, 10, 12]. To construct the incomplete multiview data, we randomly remove $M$ samples for every view under the condition that all samples have at least one view, and the missing ratio (MR, MR=$\{10\%, 30\%, 50\%, 70\%\}$) is defined as MR $= M/N$ for every view, where $N$ is the number of samples.

**Evaluation Metrics**: Normalized mutual information (NMI), accuracy (ACC), and Purity are considered to evaluate the performance in this experiment [6]. Larger values of these metrics indicate better clustering performance.

**Compared Clustering Algorithms:** BSV, Concat, OPIMC [5], HCPIMSC [10], TMBSD [11], IMVTSC-MVI [12], and TCIMC [9], are selected to compare clustering performance. Among them, BSV shows the best K-means clustering results on all single views and Concat reports the clustering results of stacking all views [6]. All Experiments are tested on a

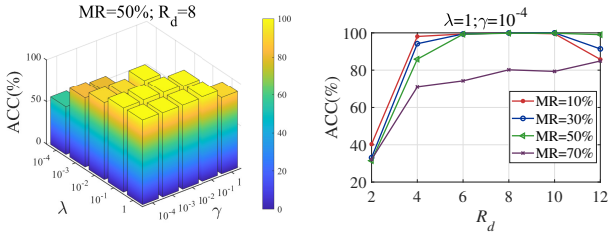
[1]http://vision.ucsd.edu/content/yale-face-database
[2]https://data.caltech.edu/records/mzrjq-6wc02
[3]https://www.fruitfly.org

**Table 1**: The comparison results of ACC(%)/NMI(%)/Purity(%) on different datasets as MR ranges from 10% to 70%.

| Datasets | MR (%) | BSV | Concat | OPIMC | TMBSD | HCPIMSC | IMVTSC-MVI | TCIMC | Ours |
|---|---|---|---|---|---|---|---|---|---|
| Yale | 10.0 | 39.39/ 45.51/ 43.03 | 38.79/ 46.94/ 43.64 | 40.61/ 42.99/ 41.21 | 81.82/ 82.03/ 81.82 | 77.58/ 76.67/ 77.58 | 70.91/ 73.63/ 70.91 | 61.21/ 65.97/ 62.42 | **100.00/100.00/100.00** |
| | 30.0 | 29.70/ 30.02/ 33.33 | 27.88/ 31.07/ 29.70 | 32.73/ 36.27/ 33.33 | 69.09/ 74.47/ 71.52 | 67.88/ 69.88/ 68.48 | 69.70/ 70.27/ 69.70 | 56.97/ 63.13/ 58.79 | **100.00/100.00/100.00** |
| | 50.0 | 27.27/ 26.39/ 30.30 | 21.82/ 18.70/ 24.24 | 24.85/ 22.22/ 26.67 | 60.60/ 65.57/ 61.21 | 49.09/ 52.38/ 50.30 | 68.48/ 66.76/ 69.09 | 55.76/ 58.20/ 55.76 | **99.76/99.69/99.75** |
| | 70.0 | 19.39/ 15.72/ 20.61 | 13.94/ 11.10/ 16.97 | 21.82/ 21.38/ 23.64 | 36.36/ 37.11/ 36.36 | 35.15/ 38.33/ 35.76 | 32.73/ 36.48/ 35.15 | 35.15/ 32.91/ 36.36 | **79.81/84.29/80.42** |
| Coil20 | 10.0 | 57.85/ 66.08/ 59.10 | 59.65/ 75.26/ 64.79 | 49.17/ 68.74/ 52.92 | 89.03/ 93.18/ 89.03 | 68.12/ 78.60/ 71.18 | 86.81/ 90.39/ 86.88 | 77.57/ 91.82/ 84.72 | **89.58/ 93.28/ 89.58** |
| | 30.0 | 42.29/ 48.22/ 44.03 | 48.33/ 58.81/ 51.39 | 44.03/ 56.83/ 46.18 | 81.67/ 88.54/ 81.81 | 68.47/ 77.38/ 72.15 | 82.43/ 89.06/ 84.72 | 75.28/ 85.41/ 81.88 | **92.50/ 94.41/ 92.64** |
| | 50.0 | 33.96/ 36.97/ 34.86 | 30.76/ 40.01/ 36.11 | 39.44/ 54.94/ 43.26 | 75.97/ 85.73/ 77.57 | 54.79/ 64.30/ 57.01 | 83.12/ 88.18/ 83.26 | 41.39/ 59.72/ 53.19 | **87.92/ 93.93/ 90.42** |
| | 70.0 | 20.14/ 19.81/ 22.43 | 22.71/ 24.66/ 24.51 | 23.33/ 27.75/ 23.75 | 77.99/ 86.28/ 78.89 | 33.54/ 36.76/ 34.44 | 51.18/ 56.85/ 51.94 | 31.11/ 38.21/ 39.44 | **86.18/ 91.93/ 88.68** |
| BDGP | 10.0 | 48.40/ 32.61/ 48.40 | 53.28/ 43.56/ 58.48 | 81.48/ 65.30/ 81.48 | 78.00/ 51.43/ 78.00 | 78.92/ 67.49/ 78.92 | 99.68/ 98.68/ 99.68 | 44.36/ 23.21/ 44.88 | **100.00/100.00/100.00** |
| | 30.0 | 45.40/ 27.42/ 46.56 | 36.24/ 17.99/ 36.88 | 79.64/ 54.96/ 79.64 | 54.16/ 23.51/ 54.16 | 76.04/ 57.55/ 76.04 | 98.84/ 95.59/ 98.84 | 29.84/ 11.21/ 33.76 | **100.00/100.00/100.00** |
| | 50.0 | 29.96/ 13.61/ 32.80 | 30.40/ 10.83/ 31.48 | 55.00/ 29.92/ 55.00 | 53.36/ 22.13/ 53.36 | 68.96/ 38.13/ 68.96 | 96.52/ 89.05/ 96.52 | 38.24/ 13.84/ 39.68 | **100.00/100.00/100.00** |
| | 70.0 | 28.24/ 9.30/ 29.08 | 28.00/ 8.37/ 28.04 | 41.76/ 13.00/ 41.76 | 39.16/ 11.88/ 39.20 | 32.12/ 7.98/ 33.32 | **81.72**/ 59.72/ **81.72** | 25.20/ 3.23/ 25.24 | 78.60/ **71.08**/78.60 |
| Caltech7 | 10.0 | 52.72/ 45.98/ 84.56 | 42.11/ 41.58/ 84.35 | 62.04/ 52.95/ 81.84 | 36.46/ 39.75/ 83.54 | 60.68/ 53.66/ 88.91 | **62.31**/ 52.22/ 85.58 | 61.70/ 44.35/ 83.54 | 61.56/ **60.66/ 92.38** |
| | 30.0 | 44.15/ 36.02/ 75.24 | 45.44/ 37.61/ 82.38 | **64.42**/ 45.57/ 83.81 | 36.80/ 33.47/ 80.48 | 50.75/ 47.73/ 86.94 | 50.48/ 48.87/ 85.78 | 62.45/ 48.80/ 83.61 | 64.08/ **62.33/ 91.73** |
| | 50.0 | 53.88/ 30.83/ 71.90 | 45.58/ 32.91/ 73.47 | 60.68/ 35.01/ 80.68 | 35.37/ 24.15/ 76.33 | 56.39/ 43.05/ 85.65 | 52.18/ 45.52/ 85.51 | **63.67**/ 46.40/ 84.29 | 62.82/ **60.69/ 90.71** |
| | 70.0 | 48.91/ 13.22/ 61.36 | 49.80/ 22.83/ 72.04 | 53.40/ 21.30/ 74.01 | 30.41/ 13.52/ 68.37 | 50.00/ 34.36/ 80.48 | 31.84/ 16.37/ 67.55 | 55.99/ 36.77/ 79.73 | **62.48/ 61.10/ 93.06** |

desktop computer with a 3.30GHz Intel(R) Xeon(R)(TM) CPU and 256GB RAM.

**Parameter settings**: In our model, each self-representation tensor is balancely reshaped into a 4th-order tensor to use low-rank TR approximation. Therefore, there are two groups of free parameters with penalty factors $\rho_1$=$\rho_2$=$\rho_3$=$10^{-3}$, one is TR ranks, e.g., $R_d$,$d$=$1,\cdots,4$, and the other is trade-off parameters $\lambda$ and $\gamma$. We fix one group's parameters and tune the other via brute force search. For example, we first fix $R_d = 8$ and tune parameters $\lambda$ and $\gamma$ on Yale dataset, where $\lambda$ and $\gamma$ vary from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$. The clustering results are reported in Fig.4a , where the ACC almost approaches 1 and performs stable when $\lambda$ ranges in $[0.1, 1]$ and $\gamma$ ranges in $[10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]$. In addition, we fix $\lambda$=1 and $\gamma$=$10^{-4}$ to choose the parameter $R_d$ from {2,4,6,8,10,12}. From Fig. 4b, we can observe the clustering performance is good and stable when $R_d$ ranges in $[6, 10]$ with small MRs. According to this strategy, we choose $R_d$=8, $\lambda$=1, $\gamma$=$10^{-4}$ for Yale; $R_d$=4, $\lambda$=$10^{-4}$, $\gamma$=$10^{-3}$ for Coil20; $R_d$=4, $\lambda$=$10^{-3}$,$\gamma$=$10^{-3}$ for BDGP; $R_d$=6, $\lambda$=$10^{-1}$, $\gamma$=$10^{-1}$ for Caltech7 in the following experiments.




(a) The performance on $\lambda$ and $\gamma$ (b) The performance on $R_d$

**Fig. 4**: The change of ACC as parameters $\lambda$, $\gamma$ and $R_d$ vary.

### 4.2. Clustering Performance Analysis

Table 1 shows the clustering performance of different methods on four multi-view datasets with MR ranging from {10%, 30%, 50%, 70%} in terms of ACC, NMI, and Purity, where the best and the second best results are highlighted in **bold** and underlined, respectively. It can be observed that the clustering performance of our method is superior to the state-of-art in terms of NMI and Purity for all multiview datasets and MRs. Especially, the proposed method achieved a 18.18%, 30.30% and 31.28% improvement in terms of ACC on the Yale dataset when MR=10%, 30% and 50%, respectively. When the MR reaches 70%, the clustering accuracy of our algorithm remains at a high level (ACC=79.81%) while the performance of other algorithms is extremely degraded.

### 4.3. Ablation Study

To investigate the usefulness of FSR in the FSR-IMVC, the low-rank completion part in equation (1) is removed and the clustering result on Yale and Caltech7 is reported in Fig. 5. Each one is tuned the best. We can see that ACC decreases significantly without FSR, especially when MR=50%, 70%. This may imply that FSR can explore more useful information for IMVC from incomplete multi-view data.

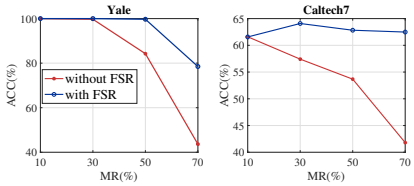



**Fig. 5**: The clustering performance on incomplete multiview data with/without completion procedure.

## 5. CONCLUSION

In this paper, we incorporate low-rank matrix learning of original feature space and low-rank TR approximation-based high-order self-representation learning into a unified framework for incomplete multi-view clustering. The two processes fully exploit inter/intra-view and feature space correlations linked by the latent feature spaces, which ultimately improves the clustering performance. Numerical experiments on four well-known multiview datasets with different MRs show that our method outperforms all state-of-the-art methods on clustering performance.

## 6. REFERENCES


[1] Xia Dong, Danyang Wu, Feiping Nie, Rong Wang, and Xuelong Li, "Dependence-guided multi-view clustering," in *ICASSP*. IEEE, 2021, pp. 3650–3654.

[2] Danyang Wu, Feiping Nie, Rong Wang, and Xuelong Li, "Multi-view clustering via mixed embedding approximation," in *ICASSP*. IEEE, 2020, pp. 3977–3981.

[3] Jie Xu, Yazhou Ren, Guofeng Li, Lili Pan, Ce Zhu, and Zenglin Xu, "Deep embedded multi-view clustering with collaborative training," *Information Sciences*, vol. 573, pp. 279–290, 2021.

[4] Piyush Rai, Anusua Trivedi, Hal Daumé III, and Scott L DuVall, "Multiview clustering with incomplete views," in *Proceedings of the NIPS Workshop on Machine Learning for Social Computing*. Citeseer, 2010.

[5] Menglei Hu and Songcan Chen, "One-pass incomplete multi-view clustering," in *AAAI*, 2019, vol. 33, pp. 3838–3845.

[6] Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang, and Jinxing Li, "A survey on incomplete multiview clustering," *TSMC*, 2022.

[7] Jie Wen, Yong Xu, and Hong Liu, "Incomplete multi-view spectral clustering with adaptive graph learning," *TCYB*, vol. 50, no. 4, pp. 1418–1429, 2018.

[8] Jianlun Liu, Shaohua Teng, Wei Zhang, Xiaozhao Fang, Lunke Fei, and Zhuxiu Zhang, "Incomplete multi-view subspace clustering with low-rank tensor," in *ICASSP*. IEEE, 2021, pp. 3180–3184.

[9] Wei Xia, Quanxue Gao, Qianqian Wang, and Xinbo Gao, "Tensor completion-based incomplete multiview clustering," *TCYB*, 2022.

[10] Zhenglai Li, Chang Tang, Xiao Zheng, Xinwang Liu, Wei Zhang, and En Zhu, "High-order correlation preserved incomplete multi-view subspace clustering," *TIP*, vol. 31, pp. 2067–2080, 2022.

[11] Zhenglai Li, Chang Tang, Xinwang Liu, Xiao Zheng, Wei Zhang, and En Zhu, "Tensor-based multi-view block-diagonal structure diffusion for clustering incomplete multi-view data," in *ICME*. IEEE, 2021, pp. 1–6.

[12] Jie Wen, Zheng Zhang, Zhao Zhang, Lei Zhu, Lunke Fei, Bob Zhang, and Yong Xu, "Unified tensor framework for incomplete multi-view clustering and missing-view inferring," in *AAAI*, 2021, vol. 35, pp. 10273–10281.

[13] Misha E Kilmer, Karen Braman, Ning Hao, and Randy C Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIMAX*, vol. 34, no. 1, pp. 148–172, 2013.

[14] Tao Zhou, Changqing Zhang, Xi Peng, Harish Bhaskar, and Jie Yang, "Dual shared-specific multiview subspace clustering," *TCYB*, vol. 50, no. 8, pp. 3517–3530, 2019.

[15] Yipeng Liu, Jiani Liu, Zhen Long, and Ce Zhu, "Tensor subspace cluster," in *Tensor Computation for Data Analysis*, pp. 219–239. Springer, 2022.

[16] Qibin Zhao, Guoxu Zhou, Shengli Xie, Liqing Zhang, and Andrzej Cichocki, "Tensor ring decomposition," *arXiv preprint arXiv:1606.05535*, 2016.

[17] Zhen Long, Ce Zhu, Jiani Liu, and Yipeng Liu, "Bayesian low rank tensor ring for image recovery," *TIP*, vol. 30, pp. 3568–3580, 2021.

[18] Jiani Liu, Ce Zhu, and Yipeng Liu, "Smooth compact tensor ring regression," *TKDE*, vol. 34, no. 9, pp. 4439–4452, 2022.

[19] Nathan Srebro, Jason D. M. Rennie, and Tommi S. Jaakkola, "Maximum-margin matrix factorization," in *Proceedings of the 17th International Conference on Neural Information Processing Systems*, Cambridge, MA, USA, 2004, NIPS'04, p. 13291336, MIT Press.

[20] Jinglin Xu, Junwei Han, Feiping Nie, and Xuelong Li, "Re-weighted discriminatively embedded $k$ -means for multi-view clustering," *TIP*, vol. 26, no. 6, pp. 3016–3027, 2017.

[21] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., "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.

[22] Ulrike Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[23] Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al., "Columbia object image library (coil-100)," 1996.