# Self-representation Subspace Clustering for Incomplete Multi-view Data

Jiyuan Liu, Xinwang Liu*
{liujiyuan13,xinwangliu}@nudt.edu.cn
National University of Defense Technology
Changsha, Hunan, China

Yi Zhang, Pei Zhang
National University of Defense Technology
Changsha, Hunan, China

Wenxuan Tu, Siwei Wang
National University of Defense Technology
Changsha, Hunan, China

Sihang Zhou, Weixuan Liang
National University of Defense Technology
Changsha, Hunan, China

Siqi Wang
National University of Defense Technology
Changsha, Hunan, China

Yuexiang Yang
National University of Defense Technology
Changsha, Hunan, China

## ABSTRACT

Incomplete multi-view clustering is an important research topic in multimedia where partial data entries of one or more views are missing. Current subspace clustering approaches mostly employ matrix factorization on the observed feature matrices to address this issue. Meanwhile, self-representation technique is left unexplored, since it explicitly relies on full data entries to construct the coefficient matrix, which is contradictory to the incomplete data setting. However, it is widely observed that self-representation subspace method enjoys a better clustering performance over the factorization based one. Therefore, we adapt it to incomplete data by jointly performing data imputation and self-representation learning. To the best of our knowledge, this is the first attempt in incomplete multi-view clustering literature. Besides, the proposed method is carefully compared with current advances in experiment with respect to different missing ratios, verifying its effectiveness.

## CCS CONCEPTS

• **Computing methodologies** → **Cluster analysis**; *Statistical relational learning*; *Spectral methods*; • **Theory of computation** → **Unsupervised learning and clustering**; • **Mathematics of computing** → Nonconvex optimization.

## KEYWORDS

Incomplete data, multi-view clustering, subspace clustering, self-representation learning

**ACM Reference Format:**
Jiyuan Liu, Xinwang Liu, Yi Zhang, Pei Zhang, Wenxuan Tu, Siwei Wang, Sihang Zhou, Weixuan Liang, Siqi Wang, and Yuexiang Yang. 2021. Self-representation Subspace Clustering for Incomplete Multi-view Data. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3474085.3475379

*Corresponding author


ACM ISBN 978-1-4503-8651-7/21/10…$15.00
https://doi.org/10.1145/3474085.3475379

## 1 INTRODUCTION

With the wide spread of multimedia technology in people's life, it is easy and natural to describe an object from multiple modalities. For instance, a person can be identified from its biometric features, such as physical appearance (image), and its social behaviors, including Internet tweets (text), colleague relationships (graph), etc. How to utilize them to build a better model is crucial and valuable in real-world applications. Current multi-view, also known as multi-modal, approaches address this issue by sufficiently exploring their complementary information [13, 14, 20, 21, 32, 33, 39]. However, the collected data are mostly incomplete in practice due to sensor failure or human error. There are three types of data missing in common sense, including feature missing, view missing and mixed missing. Here concerns the view missing setting which is formally defined in Definition 1.1.

*Definition 1.1. Incomplete multi-view data (view missing).* For $n$ given multi-view data entries, $\mathcal{S}_v$ collects indexes of the observed ones in $v$-th view. The $i$-th data entry is complete in $v$-th view if $i \in \mathcal{S}_v$, or totally missing if $i \notin \mathcal{S}_v$. Besides,

$$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \mathcal{S}_V = \{1, 2, \cdots, n\}, \quad \mathcal{S}_1 \cap \mathcal{S}_2 \cap \cdots \mathcal{S}_V \neq \emptyset. \tag{1}$$

To obtain the underlying discriminative information, clustering research community has proposed a large set of methods for incomplete multi-view data. For instance, By constructing incomplete kernels corresponding to each data view, Liu et al. try to learn a complete consensus kernel on which the standard kernel $k$-means technique is applied concurrently to obtain the clustering result [23]. On the contrary, some late-fusion methods firstly employ the basic single-view clustering techniques on each data view to compute corresponding incomplete soft-label matrices, then regularize them towards a complete one [22]. Besides, Cai et al. propose to separately optimize partial clustering results with spectral technique on both individual and shared parts of different data views [4]. The target solution is inferred by simply combining the partial results. At the same time, neural network is adopted in incomplete multi-view learning [25, 37]. Wang et al. map the data into bottleneck representations with independent encoders and integrate the

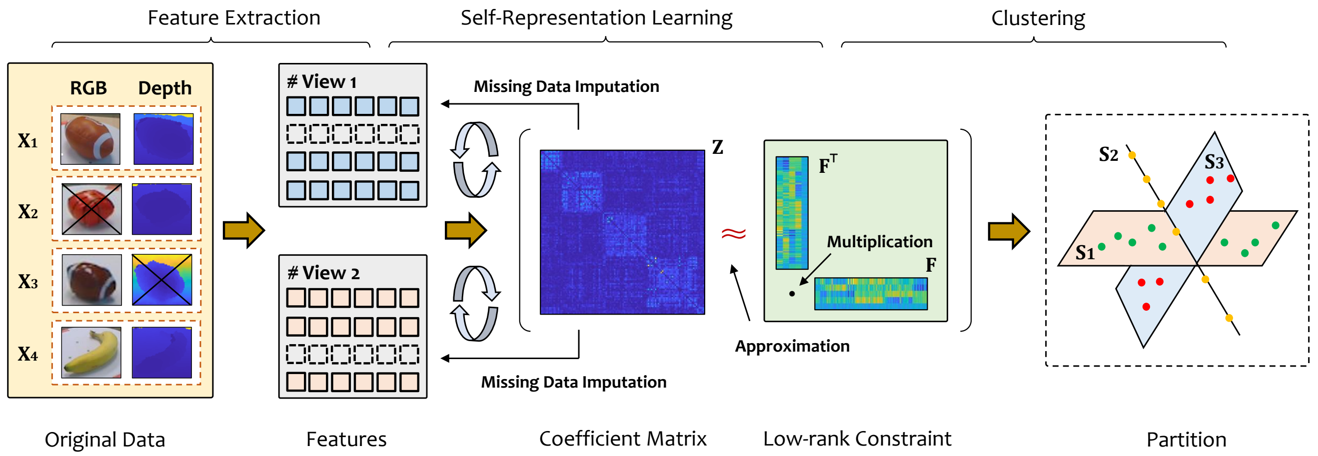


**Figure 1: Overview of the proposed algorithm. For the ease of presentation, two data views are concerned, including RGB and depth images. From left to right, the incomplete multi-view data X are extracted into feature vectors. Then, a consensus coefficient matrix Z is obtained with low-rank regularization. Meanwhile, the self-representation learning and missing data imputation are jointly performed. At last, cluster assignments are generated from the coefficient matrix.**

complementary information by employing t-student distribution [30]. Generative adversarial network (GAN) is further utilized to boost the clustering performance.

Apart from the aforementioned approaches, factorization based subspace methods are well studied in incomplete multi-view setting [11, 12, 17, 26, 28, 29, 35, 40]. For example, Li et al. factorize data of each view to learn the latent subspaces independently but enforce them overlapping at entries observed in all views [17, 40]. One step further, graph laplacian regularization is employed to improve its clustering performance [26]. In addition, Hu et al. establish a consensus basis matrix with the help of $\ell_{2,1}$-norm regularized regression to reduce the influence of missing entries [11]. By regularizing the divergences across all data views, Shao et al. push all learned subspaces towards a common consensus [28]. To ease the computing overhead problem of subspace clustering, Hu et al. propose an one-pass algorithm embedded with regularized matrix factorization and weighted matrix factorization [12]. Besides, Wen et al. impose orthogonal constraint on basis matrices corresponding to each view, at the same time, regularize data reconstruction with binary weight to preserve the local structure [35].

Compared with factorization based methods, self-representation subspace algorithm explicitly relies on all data entries to build corresponding coefficient matrix, resulting in a larger gap of adapting it into incomplete multi-view setting. To the best of our knowledge, few researches have been established to reduce this gap. However, self-representation method, on the other side, benefits from the built coefficient matrix, for it reflects the precise relationships among data entries, making it feasible to achieve a better clustering performance [16]. We also validate this by conducting an ablation study in the experiment part. Therefore, we propose a self-representation subspace clustering algorithm in incomplete multi-view setting (IMSR). Taking two-view data as an instance, its overview can visualized in Fig. 1. Specifically, the incomplete multi-view data in Definition 1.1 are first extracted into corresponding feature matrices. Next, the self-representation subspace clustering is employed to learn the coefficient matrix which is, at the same time, regularized to be low-rank by approximating it with the multiplication of low-dimension matrix. Nevertheless, the obtained coefficient matrix guides to impute missing data as feedback, leading to a cyclical procedure. Once the process converges, we can obtain the accurate relationships among entries and compute the cluster assignments. Overall, the contributions are summarized in the following.

(1) We shed the first light of self-representation subspace clustering in incomplete multi-view data. At the same time, an ablation study is designed in experiment to validate its effectiveness by comparing with the factorization based subspace method.
(2) With regularizing the coefficient matrix to be low rank, an incomplete multi-view clustering algorithm (IMSR) is developed. Meanwhile, we design an alternate strategy to solve the resultant optimization problem, and compare it with current advances with respect to different missing ratios, verifying its superiority.

## 2 RELATED WORK

Given $n$ data entries $\{\mathbf{X}_v \in \mathbb{R}^{d_v \times n}\}_{v=1}^{V}$ drawn from $k$ distributions where $d_v$ refers to feature dimension of the $v$-th view, multi-view clustering aims to group the data by exploring complementary information across $V$ views. Correspondingly, we employ $\{\mathcal{S}_v^{(o)}\}_{v=1}^{V}/\{\mathcal{S}_v^{(m)}\}_{v=1}^{V}$ to indicate index sets of the observed/missing entries and denote $\{[\mathbf{X}_v^{(o)}, \mathbf{X}_v^{(m)}]\}_{v=1}^{V}$ as the incomplete multi-view data specified in Definition 1.1. Note that, $[\cdot,\cdot]/[\cdot;\cdot]$ is the horizontal/vertical concatenation operation. In the following are the closely related subspace clustering reviews in incomplete data setting.

### 2.1 Factorization based Clustering

Non-negative matrix factorization [3] is the most typical and widely used clustering technique in machine learning community. It approximates data $\mathbf{X} \in \mathbb{R}^{d \times n}$ with dot product of the so-called basis matrix $\mathbf{U} \in \mathbb{R}^{d\times k}$ and latent feature matrix $\mathbf{V} \in \mathbb{R}^{n\times k}$ as follows.

$$\min_{\mathbf{U},\mathbf{V}} \|\mathbf{X} - \mathbf{U}\mathbf{V}^\top\|_F^2, \quad s.t. \ \mathbf{U} \geq 0, \ \mathbf{V} \geq 0. \tag{2}$$

Besides, a large volume of researches are conducted in matrix factorization and can be generally summarized as

$$\min_{\mathbf{U},\mathbf{V}} \ell(\mathbf{X} - \mathbf{U}\mathbf{V}^\top) + \lambda\mathcal{R}(\mathbf{U},\mathbf{V}), \quad s.t. \ \Omega(\mathbf{U},\mathbf{V}) \tag{3}$$

where $\ell$ refers to different norms. Meanwhile, $\mathcal{R}$ represents the regularizations on $\mathbf{U}$ and $\mathbf{V}$, including the non-negativity [3], orthogonality [6], etc. $\Omega$ defines feasible regions of the variables.

Since matrix factorization method does not explicitly rely on the relationships between data entries as shown in Eq. (3), it is natural and straightforward to apply in incomplete multi-view data. Li et al. [17] consider the two view consequence by minimizing

$$\begin{aligned}\min_{\mathbf{U}_1,\mathbf{U}_2,\mathbf{V}_1,\mathbf{V}_2} & \left\|[\mathbf{X}_1^{(o)}, \mathbf{X}_1^{(m_1)}] - \mathbf{U}_1[\mathbf{V}^{(o)}; \mathbf{V}_1^{(m_1)}]^\top\right\|_F^2 + \lambda\|\mathbf{V}_1\|_F^2 \\ & + \left\|[\mathbf{X}_2^{(o)}, \mathbf{X}_2^{(m_2)}] - \mathbf{U}_2[\mathbf{V}^{(o)}; \mathbf{V}_2^{(m_2)}]^\top\right\|_F^2 + \lambda\|\mathbf{V}_2\|_F^2 \\ s.t. & \mathbf{U}_1 \geq 0, \ \mathbf{V}_1 \geq 0, \ \mathbf{U}_2 \geq 0, \ \mathbf{V}_2 \geq 0,\end{aligned} \tag{4}$$

in which $\mathbf{V}_v = [\mathbf{V}^{(o)}; \mathbf{V}_v^{(m_v)}]$, and $(m_v)$ specifies the entries only missing in $v$-th view. It can be observed that the latent feature matrices of two views share a consensus part at the observed data entries. Subsequent researches [12, 28, 35] adapt this framework into more than two views as

$$\begin{aligned}\min_{\{\mathbf{U}\}_{v=1}^V, \{\mathbf{V}_v\}_{v=1}^V, \mathbf{V}_*} & \sum_{v=1}^V \beta_v \ell([\mathbf{X}_v^{(o)}, \mathbf{X}_v^{(m)}] - \mathbf{U}_v[\mathbf{V}_v^{(o)}; \mathbf{V}_v^{(m)}]^\top) \\ & + \lambda\mathcal{R}(\{\mathbf{U}_v\}_{v=1}^V, \{\mathbf{V}_v\}_{v=1}^V, \mathbf{V}_*) \\ s.t. & \ \Omega(\{\mathbf{U}_v\}_{v=1}^V, \{\mathbf{V}_v\}_{v=1}^V, \mathbf{V}_*, \{\beta_v\}_{v=1}^V)\end{aligned} \tag{5}$$

where $\mathbf{V}_v = [\mathbf{V}_v^{(o)}; \mathbf{V}_v^{(m)}]$, and $\mathbf{V}_*$ is the target consensus latent feature matrix. $\mathcal{R}$ and $\Omega$ are the regularizations and feasible regions of $\{\mathbf{U}_v\}_{v=1}^V$, $\{\mathbf{V}_v\}_{v=1}^V$, $\mathbf{V}_*$ and $\{\beta_v\}_{v=1}^V$, respectively. Some literatures [11] assume all data views can be factorized into an unique subspace by imposing $\mathbf{V}_v = \mathbf{V}_*, \ \forall v$.

## 2.2 Self-representation Clustering

In single-view setting, self-representation subspace clustering aims to reconstruct each data entry with a linear combination of the others, which can be formulated into

$$\min_{\mathbf{Z}} \ell(\mathbf{X} - \mathbf{X}\mathbf{Z}) + \lambda\mathcal{R}(\mathbf{Z}), \quad s.t. \ \Omega(\mathbf{Z}), \tag{6}$$

in which $\mathbf{Z} \in \mathbb{R}^{n\times n}$ is named coefficient matrix and contains the relationships across all data entries. In literature, $\ell$ and $\mathcal{R}$ are instanced with multiple norms and regularizations, such as $\ell_2$-norm [24], $\ell_1$-norm [7], graph regularization [10], etc. Besides, it is natural to extend self-representation clustering into multi-view setting. By following the framework of factorization based methods in Eq. (5), multi-view self-representation subspace clustering approach [15, 36] for complete data is formulated into

$$\begin{aligned}\min_{\{\mathbf{Z}_v\}_{v=1}^V, \mathbf{Z}_*} & \sum_{v=1}^V \beta_v \ell(\mathbf{X}_v - \mathbf{X}_v\mathbf{Z}_v) + \lambda\mathcal{R}(\{\mathbf{Z}_v\}_{v=1}^V, \mathbf{Z}_*) \\ s.t. & \ \Omega(\{\mathbf{Z}_v\}_{v=1}^V, \mathbf{Z}_*, \{\beta\}_{v=1}^V).\end{aligned} \tag{7}$$

Similarly, $\mathbf{Z}_*$ is the consensus coefficient matrix of all data views. There is common consensus that self-representation subspace clustering method outperforms factorization based one, since it learns more precise relationships across entries in $\mathbf{Z}$. This can be validated by the performance comparisons in literature under complete multi-view setting, such as [16]. However, there leaves a research gap, i.e. how to adapt self-representation subspace clustering into the incomplete multi-view data.

# 3 THE PROPOSED ALGORITHM

In order to fill the gap between self-representation subspace clustering and incomplete multi-view data, we propose the IMSR algorithm. At the very beginning, $\ell_2$-norm and diagonal constraint are employed to instance Eq. (6), since the resultant objective in Eq. (8) can group the highly correlated entries together compared with the others in literature, as claimed in [24].

$$\begin{aligned}\min_{\mathbf{Z}} & \ \|\mathbf{X} - \mathbf{X}\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\|_F^2 \\ s.t. & \ \mathrm{diag}(\mathbf{Z}) = \mathbf{0}.\end{aligned} \tag{8}$$

By following the framework in Eq. (7) under complete multi-view data, Eq. (8) can be transformed into

$$\begin{aligned}\min_{\mathbf{Z}} & \ \frac{1}{V}\sum_{v=1}^V \|\mathbf{X}_v - \mathbf{X}_v\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\|_F^2 \\ s.t. & \ \mathrm{diag}(\mathbf{Z}) = \mathbf{0},\end{aligned} \tag{9}$$

in which different data views are equally weighted and share a consensus coefficient matrix $\mathbf{Z}$. When considering the incompleteness specified in Definition 1.1, we propose to perform missing data imputation and self-representation learning simultaneously. In specific, a consensus coefficient matrix is first computed with zero-filling or random-filling data. Afterwards, the missing entries can be imputed from the obtain coefficient matrix which is updated accordingly. This iterative procedure can be accomplished by the following objective.

$$\begin{aligned}\min_{\{\mathbf{X}_v^{(m)}\}_{v=1}^V, \mathbf{Z}} & \ \frac{1}{V}\sum_{v=1}^V \left\|[\mathbf{X}_v^{(o)}, \mathbf{X}_v^{(m)}] - [\mathbf{X}_v^{(o)}, \mathbf{X}_v^{(m)}]\mathbf{Z}\right\|_F^2 + \lambda\|\mathbf{Z}\|_F^2 \\ s.t. & \ \mathrm{diag}(\mathbf{Z}) = \mathbf{0},\end{aligned} \tag{10}$$

where $\mathbf{X}_v^{(o)}$ and $\mathbf{X}_v^{(m)}$ refer to observed and missing entries of the $v$-th data view.

For a clustering-oriented self-representation subspace learning, the obtained coefficient matrix are supposed to be low rank. Ideally, $\mathbf{Z} \in \Delta$ where

$$\Delta = \{\mathbf{Z} \mid \mathrm{diag}(\mathbf{Z}) = \mathbf{0}, \mathrm{rank}(\mathbf{Z}) = k\}, \tag{11}$$

since the data are drawn from $k$ distributions. However, it makes the objective hard to optimize by explicitly imposing the rank of $\mathbf{Z}$ equivalent to $k$. Instead, we push the coefficient matrix towards $\mathbf{F}^\top\mathbf{F}$ where $\mathbf{F} \in \mathbb{R}^{k\times n}$ is an orthogonal matrix. Finally, the overall objective can be formulated as

$$\begin{aligned}\min_{\{\mathbf{X}_v^{(m)}\}_{v=1}^V,\ \mathbf{Z},\ \mathbf{F}} \quad & \frac{1}{V}\sum_{v=1}^{V}\left\|[\mathbf{X}_v^{(o)},\mathbf{X}_v^{(m)}]-[\mathbf{X}_v^{(o)},\mathbf{X}_v^{(m)}]\mathbf{Z}\right\|_F^2 \\ & +\lambda\|\mathbf{Z}\|_F^2+\gamma\|\mathbf{Z}-\mathbf{F}^\top\mathbf{F}\|_F^2 \\ s.t.\ & \mathrm{diag}(\mathbf{Z})=\mathbf{0},\ \mathbf{F}\mathbf{F}^\top=\mathbf{I}_k,\end{aligned} \tag{12}$$

where $\lambda$ and $\gamma$ are positive trade-off parameters. It can be observed from Eq. (12) that both the observed and missing data participate in the self-representation learning, i.e. the construction of coefficient matrix $\mathbf{Z}$. Nevertheless, the missing data entries $\{\mathbf{X}_v^{(m)}\}_{v=1}^V$ are imputed along with the learning of $\mathbf{Z}$. Once the iterative procedure converges, off-the-shelf spectral clustering technique is applied on corresponding laplacian matrix $\mathbf{L}=(|\mathbf{Z}|+|\mathbf{Z}^\top|)/2$ to compute the cluster assignments.

## 4 OPTIMIZATION

To solve the proposed objective in Eq. (12), we design an alternate strategy [19] in which only one variable is optimized while the others fixed. The overall procedure is summarized in Algorithm 1.

### 4.1 F Optimization

With a fixed $\mathbf{Z}$, Eq. (12) can be reduced into

$$\max_{\mathbf{F}} \ \mathrm{Tr}\left[\mathbf{F}(\mathbf{Z}+\mathbf{Z}^\top)\mathbf{F}^\top\right] \quad s.t.\ \ \mathbf{F}\mathbf{F}^\top=\mathbf{I}_k, \tag{13}$$

which is a standard kernel $k$-means optimization problem and can be efficiently solved by off-the-shelf packages. In specific, Eq. (13) reaches the optimal maximum when $\mathbf{F}$ takes the eigen-vectors of $(\mathbf{Z}+\mathbf{Z}^\top)$ corresponding to the largest $k$ eigen-values.

### 4.2 $\mathbf{X}_v^{(m)}$ Optimization

To fill the gap of self-representation learning in incomplete setting, the proposed algorithm imputes the missing data entries with a fixed coefficient matrix $\mathbf{Z}$. The objective can be transformed into

$$\min_{\mathbf{X}_v^{(m)}} \ \mathrm{Tr}\left([\mathbf{X}_v^{(o)},\mathbf{X}_v^{(m)}]\mathbf{B}[\mathbf{X}_v^{(o)},\mathbf{X}_v^{(m)}]^\top\right), \tag{14}$$

where $\mathbf{B}=\mathbf{I}_n-\mathbf{Z}-\mathbf{Z}^\top+\mathbf{Z}\mathbf{Z}^\top$. By replacing $\mathbf{B}$ with the form of block matrix, Eq. (14) equals to

$$\min_{\mathbf{X}_v^{(m)}} \ \mathrm{Tr}\left([\mathbf{X}_v^{(o)},\mathbf{X}_v^{(m)}]\begin{bmatrix}\mathbf{B}^{(oo)} & \mathbf{B}^{(om)}\\ \mathbf{B}^{(mo)} & \mathbf{B}^{(mm)}\end{bmatrix}[\mathbf{X}_v^{(o)},\mathbf{X}_v^{(m)}]^\top\right), \tag{15}$$

which can be expanded to

$$\min_{\mathbf{X}_v^{(m)}} \ \mathrm{Tr}\left(\mathbf{X}_v^{(m)}\mathbf{B}^{(mm)}\mathbf{X}_v^{(m)\top}+\mathbf{X}_v^{(o)}(\mathbf{B}^{(mo)\top}+\mathbf{B}^{(om)})\mathbf{X}_v^{(m)\top}\right). \tag{16}$$

From the definition of $\mathbf{B}$, it can be observed that

$$\mathbf{B}^{(mm)}=(\mathbf{I}_n-\mathbf{Z})^{(m)\top}(\mathbf{I}_n-\mathbf{Z})^{(m)}, \tag{17}$$

in which $(\mathbf{I}_n-\mathbf{Z})^{(m)}$ refers to the columns of $\mathbf{I}_n-\mathbf{Z}$ corresponding to missing entries in the $v$-th view. Thus, $\mathbf{B}^{(mm)}$ is positive-definite and Eq. (16) achieves the optimal solution when its derivative at zero, resulting in

$$\mathbf{X}_v^{(m)}=-\mathbf{X}_v^{(o)}\mathbf{B}^{(om)}\mathbf{B}^{(mm)-1}. \tag{18}$$

**Algorithm 1** Self-representation subspace clustering for incomplete multi-view data (IMSR)

**Require:** Incomplete multi-view data $\{[\mathbf{X}_v^{(o)},\mathbf{X}_v^{(m)}]\}_{v=1}^V$, the number of clusters $k$, parameter $\lambda$ and parameter $\gamma$;
**Ensure:** consensus reconstruction matrix $\mathbf{Z}$;
Initialize $\mathbf{Z}$;
**while** $(obj^{t-1}-obj^t)/obj^t \le \sigma$ **do**
  Update $\mathbf{F}$ by solving Eq. (13).
  Update $\{\mathbf{X}_v^{(m)}\}_{v=1}^V$ with Eq. (18);
  Update $\mathbf{Z}$ with Eq. (29);
  $t=t+1$;
  Calculate objective value $obj^t$ with Eq. (12);
**end while**
Compute the cluster assignments with $(|\mathbf{Z}|+|\mathbf{Z}^\top|)/2$;

### 4.3 Z Optimization

In the optimization of coefficient matrix $\mathbf{Z}$, the other variables, including $\mathbf{X}_v^{(m)}$ and $\mathbf{F}$, are fixed. For the ease of expression, let $\mathbf{X}_v$ denote the horizontal concatenation of observed and missing entries $[\mathbf{X}_v^{(o)};\mathbf{X}_v^{(m)}]$. Semantically, the objective in Eq. (12) can be decomposed into $n$ optimization problems with the $i$-th one as

$$\begin{aligned}\min_{\mathbf{z}_i} \quad & \frac{1}{V}\sum_{v=1}^{V}\left\|\mathbf{x}_v^i-\mathbf{X}_v\mathbf{z}_i\right\|_F^2+\lambda\|\mathbf{z}_i\|_F^2+\gamma\|\mathbf{z}_i-\mathbf{c}_i\|_F^2 \\ s.t.\ & \mathbf{z}_i(i)=0,\end{aligned} \tag{19}$$

in which $\mathbf{x}_v^i$, $\mathbf{z}_i$ and $\mathbf{c}_i$ are the $i$-th column of $\mathbf{X}_v$, $\mathbf{Z}$ and $\mathbf{C}$ ($\mathbf{C}=\mathbf{F}^\top\mathbf{F}$), respectively. Observing that the $i$-th element of $\mathbf{z}_i$ is compulsively set to zero, Eq. (19) can be further transformed into an unconstrained one that

$$\min_{\hat{\mathbf{z}}_i} \ \frac{1}{V}\sum_{v=1}^{V}\left\|\mathbf{x}_v^i-\hat{\mathbf{X}}_v^i\hat{\mathbf{z}}_i\right\|_F^2+\lambda\|\hat{\mathbf{z}}_i\|_F^2+\gamma\|\hat{\mathbf{z}}_i-\hat{\mathbf{c}}_i\|_F^2, \tag{20}$$

where $\hat{\mathbf{X}}_v^i=[\mathbf{x}_v^1,\cdots,\mathbf{x}_v^{i-1},\mathbf{x}_v^{i+1},\cdots,\mathbf{x}_v^V]$, at the same time, $\hat{\mathbf{z}}_i$ and $\hat{\mathbf{c}}_i$ are produced by removing the $i$-th element of $\mathbf{z}_i$ and $\mathbf{c}_i$. Eq. (20) is able to be reformulated into

$$\min_{\hat{\mathbf{z}}_i} \ \mathrm{Tr}\left[\mathbf{E}_i\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\top-2\Big(\frac{1}{V}\sum_{v=1}^{V}\mathbf{x}_v^{i\top}\hat{\mathbf{X}}_v^i+\gamma\hat{\mathbf{c}}_i^\top\Big)\hat{\mathbf{z}}_i\right], \tag{21}$$

where

$$\mathbf{E}_i=\frac{1}{V}\sum_{v=1}^{V}\hat{\mathbf{X}}_v^{i\top}\hat{\mathbf{X}}_v^i+(\lambda+\gamma)\mathbf{I}_{n-1}. \tag{22}$$

It is easy to prove $\mathbf{E}_i$ is positive-definite, indicating that Eq. (21) is convex and takes the global minimum where its derivative equals to zero. Thus, we can obtain the optimal $\hat{\mathbf{z}}_i^*$ as

$$\hat{\mathbf{z}}_i^*=\mathbf{E}_i^{-1}\Big(\frac{1}{V}\sum_{v=1}^{V}\hat{\mathbf{X}}_v^{i\top}\mathbf{x}_v^i+\gamma\hat{\mathbf{c}}_i\Big). \tag{23}$$

However, it requires high computation complexity to optimize $\mathbf{Z}$ via Eq. (23), due that an inverse operation of $O(n^3)$ is needed to compute each column of $\mathbf{Z}$. Defining

$$\begin{aligned}\mathbf{D} &= \left(\frac{1}{V}\sum_{v=1}^{V}\mathbf{X}_v^\top\mathbf{X}_v + (\lambda+\gamma)\mathbf{I}_n\right)^{-1}, \\ \mathbf{X}_v\mathbf{P} &= [\hat{\mathbf{X}}_v^i, \mathbf{x}_v^i],\end{aligned} \tag{24}$$

in which $\mathbf{P}$ is a permutation matrix that $\mathbf{P}^\top\mathbf{P} = \mathbf{P}\mathbf{P}^\top = \mathbf{I}_n$. In such setting, we can get

$$\left[\mathbf{P}^\top\left(\frac{1}{V}\sum_{v=1}^{V}\mathbf{X}_v^\top\mathbf{X}_v + (\lambda+\gamma)\mathbf{I}_n\right)\mathbf{P}\right]^{-1} = \mathbf{P}^\top\mathbf{D}\mathbf{P}. \tag{25}$$

Meanwhile, the following holds

$$\begin{aligned}&\left[\mathbf{P}^\top\left(\frac{1}{V}\sum_{v=1}^{V}\mathbf{X}_v^\top\mathbf{X}_v + (\lambda+\gamma)\mathbf{I}_n\right)\mathbf{P}\right]^{-1} \\ &= \begin{bmatrix}\frac{1}{V}\sum_{v=1}^{V}\hat{\mathbf{X}}_v^{i\top}\hat{\mathbf{X}}_v^i + (\lambda+\gamma)\mathbf{I}_{n-1} & \frac{1}{V}\sum_{v=1}^{V}\hat{\mathbf{X}}_v^{i\top}\mathbf{x}_v^i \\ \frac{1}{V}\sum_{v=1}^{V}\mathbf{x}_v^{i\top}\hat{\mathbf{X}}_v^i & \frac{1}{V}\sum_{v=1}^{V}\mathbf{x}_v^{i\top}\mathbf{x}_v^i + \lambda + \gamma\end{bmatrix}^{-1} \\ &= \begin{bmatrix}\mathbf{E}_i^{-1} & 0 \\ 0 & 0\end{bmatrix} + \sigma_i\begin{bmatrix}\mathbf{b}_i\mathbf{b}_i^\top & \mathbf{b}_i \\ \mathbf{b}_i^\top & 1\end{bmatrix},\end{aligned} \tag{26}$$

in which

$$\begin{aligned}\mathbf{b}_i &= -\mathbf{E}_i^{-1}\left(\frac{1}{V}\sum_{v=1}^{V}\hat{\mathbf{X}}_v^{i\top}\mathbf{x}_v^i\right) \\ \sigma_i &= \lambda + \gamma + \frac{1}{V}\sum_{v=1}^{V}\mathbf{x}_v^{i\top}\mathbf{x}_v^i - \frac{1}{V}\sum_{v=1}^{V}\mathbf{x}_v^{i\top}(\hat{\mathbf{X}}_v^i\mathbf{E}_i^{-1}\hat{\mathbf{X}}_v^{i\top})\mathbf{x}_v^i.\end{aligned} \tag{27}$$

Combining Eq. (25), (26) and (27), it can be obtained that

$$\mathbf{b}_i = \frac{\hat{\mathbf{d}}_i}{\sigma_i}, \quad \sigma_i = \mathbf{d}_i(i), \tag{28}$$

where $\mathbf{d}_i$ is the $i$-th column of $\mathbf{D}$, $\mathbf{d}_i(i)$ is the $i$-th element of $\mathbf{d}_i$ and $\hat{\mathbf{d}}_i$ is the product of removing $\mathbf{d}_i(i)$ from $\mathbf{d}_i$. Therefore, the optimal solution Eq. (23) can be efficiently computed by

$$\begin{aligned}\hat{\mathbf{z}}_i^* &= -\mathbf{b}_i + \gamma\mathbf{E}_i^{-1}\mathbf{c}_i \\ &= -\mathbf{b}_i + \gamma(\hat{\mathbf{D}}_i - \sigma_i\mathbf{b}_i\mathbf{b}_i^\top)\hat{\mathbf{c}}_i \\ &= -\frac{\hat{\mathbf{d}}_i}{\mathbf{d}_i(i)} + \gamma(\hat{\mathbf{D}}_i - \frac{\hat{\mathbf{d}}_i\hat{\mathbf{d}}_i^\top}{\mathbf{d}_i(i)})\hat{\mathbf{c}}_i \\ &= \gamma\hat{\mathbf{D}}_i\hat{\mathbf{c}}_i - \frac{1+\gamma\hat{\mathbf{d}}_i^\top\hat{\mathbf{c}}_i}{\mathbf{d}_i(i)}\hat{\mathbf{d}}_i,\end{aligned} \tag{29}$$

in which $\hat{\mathbf{D}}_i \in \mathbb{R}^{(n-1)\times(n-1)}$ is the sub-matrix of $\mathbf{D}$ by removing the $i$-th column and row.

## 4.4 Convergence and Complexity

The proposed algorithm is theoretically guaranteed to be convergent. For the ease of expression, we denote the objective in Eq. (12) as

$$\min_{\mathbf{F}, \{\mathbf{X}_v^{(m)}\}_{v=1}^V, \mathbf{Z}} f(\mathbf{F}, \{\mathbf{X}_v^{(m)}\}_{v=1}^V, \mathbf{Z}), \quad s.t. \quad (\mathbf{F}, \mathbf{Z}) \in \Delta \tag{30}$$

Along with the iterative alternate optimization procedure, three variables are separately solved at optimality with the others fixed. With superscript $t$ denoting the optimization at $t$-th iteration, the convergence analysis is provided as follows:

1) $\mathbf{F}$ *optimization.* Given $\{\mathbf{X}_v^{(m)(t)}\}_{v=1}^V$ and $\mathbf{Z}^{(t)}$, $\mathbf{F}^{(t+1)}$ can be obtained, leading to

$$\begin{aligned}&f(\mathbf{F}^{(t)}, \{\mathbf{X}_v^{(m)(t)}\}_{v=1}^V, \mathbf{Z}^{(t)}) \\ &\geq f(\mathbf{F}^{(t+1)}, \{\mathbf{X}_v^{(m)(t)}\}_{v=1}^V, \mathbf{Z}^{(t)}).\end{aligned} \tag{31}$$

2) $\{\mathbf{X}_v^{(m)}\}_{v=1}^V$ *optimization.* Given $\mathbf{F}^{(t+1)}$ and $\mathbf{Z}^{(t)}$, $\{\mathbf{X}_v^{(m)(t+1)}\}_{v=1}^V$ can be obtained, leading to

$$\begin{aligned}&f(\mathbf{F}^{(t+1)}, \{\mathbf{X}_v^{(m)(t)}\}_{v=1}^V, \mathbf{Z}^{(t)}) \\ &\geq f(\mathbf{F}^{(t+1)}, \{\mathbf{X}_v^{(m)(t+1)}\}_{v=1}^V, \mathbf{Z}^{(t)}).\end{aligned} \tag{32}$$

3) $\mathbf{Z}$ *optimization.* Given $\mathbf{F}^{(t+1)}$ and $\{\mathbf{X}_v^{(m)(t+1)}\}_{v=1}^V$, $\mathbf{Z}^{(t+1)}$ can be obtained, leading to

$$\begin{aligned}&f(\mathbf{F}^{(t+1)}, \{\mathbf{X}_v^{(m)(t+1)}\}_{v=1}^V, \mathbf{Z}^{(t)}) \\ &\geq f(\mathbf{F}^{(t+1)}, \{\mathbf{X}_v^{(m)(t+1)}\}_{v=1}^V, \mathbf{Z}^{(t+1)}).\end{aligned} \tag{33}$$

Combining Eq. (31), (32) and (33), we can get

$$\begin{aligned}&f(\mathbf{F}^{(t)}, \{\mathbf{X}_v^{(m)(t)}\}_{v=1}^V, \mathbf{Z}^{(t)}) \\ &\geq f(\mathbf{F}^{(t+1)}, \{\mathbf{X}_v^{(m)(t+1)}\}_{v=1}^V, \mathbf{Z}^{(t+1)}),\end{aligned} \tag{34}$$

illustrating that the objective value monotonically decreases along with iteration. At the same time, it is easy to prove that the objective is lower bounded by 0. Therefore, the proposed algorithm is convergent theoretically.

The computational complexity analysis is also analyzed corresponding to each sub-optimization problem. In specific, eigendecomposition on $(\mathbf{Z} + \mathbf{Z}^\top) \in \mathbb{R}^{n\times n}$ is required to optimize $\mathbf{F}$, resulting in $O(n^3)$ computational complexity. In $\mathbf{X}_v^{(m)}$ optimization, it needs $O(d_v n_o n_m + n_o n_m^2)$ to solve Eq. (18) where $d_v$, $n_o$ and $n_m$ refer to the feature number of $v$-th view, the observed entry number and the missing entry number, respectively. For all views, $O(vd_v n_o n_m + vn^o n_m^2)$ is required. For $Z$ optimization, computation of matrix $\mathbf{D}$ is of $O(n^3)$ in Eq. (24). Nevertheless, $\hat{\mathbf{z}}_i^*$ requires $O(n^2)$ in Eq. (29). Assuming the algorithm converges at $t$-th iteration, the overall complexity is of $O(3tn^3 + tvd_v n_o n_m + tvn_o n_m^2)$, which can be simplified into $O(n^3)$ respect to the entry number.

# 5 EXPERIMENT

At the very beginning, the experiment settings are introduced. Next are the experiment results, including an ablation study to validate the superiority of self-representation subspace clustering over the factorization based one, performance comparisons with current advances in literature and property exploration of the proposed algorithm.

## 5.1 Setting

*5.1.1 Dataset.* In the experiments are employed eight popular multi-view datasets to validate effectiveness and superiority of the proposed algorithm, including *BBCSport*[1] [8], *Yale*[2] [2], *ORL*[3]

[1]http://mlg.ucd.ie/datasets/bbc.html
[2]http://vision.ucsd.edu/content/yale-face-database
[3]https://github.com/GPMVCDummy/GPMVC/blob/

**Table 1: Specifications of the used datasets.**

| Dataset | Number of | | Dimension of View | | | |
|---|---|---|---|---|---|---|
| | Entry | Cluster | 1 | 2 | 3 | 4 |
| BBCSport | 116 | 5 | 1991 | 2063 | 2113 | 2158 |
| Yale | 165 | 15 | 4096 | 3304 | 6750 | - |
| ORL | 400 | 40 | 1024 | 288 | - | - |
| Olympics | 464 | 28 | 4942 | 3097 | - | - |
| Still | 467 | 6 | 200 | 200 | 200 | - |
| BBC | 658 | 5 | 4659 | 4633 | 4665 | 4684 |
| Buaa | 1350 | 150 | 100 | 100 | - | - |
| Leaves | 1600 | 100 | 64 | 64 | 64 | - |

[27], *Olympics*[4] [9], *Still*[5] [5], *BBC*[6] [8], *Buaa*[7] [31] and *Leaves*[8] [1]. Their details are specified in Table 1. To produce incomplete multi-view data in Definition 1.1, $n_m = \lfloor n * r \rfloor$ data entries are arbitrarily removed in the first view with $n$ and $r$ being the entry number and missing ratio. Then, we select another random $n_m$ entries in the second view and remove those which exist in the first view, so as to ensure all data are observed at least in one view. For later views, $n_m$ data entries are randomly removed. In the following experiments, five missing ratios from 0.1 to 0.5 are considered.

*5.1.2 Comparative Method.* The proposed algorithm is thoroughly compared with two baselines and seven advances in literature, including

(1) *LSRs* (single-best baseline) first fills the missing entries with random values. Next, LSR [24] algorithm is performed on each view and the best results are reported.
(2) *LSRc* (concatenated baseline) first fills the missing entries with random values and concatenates all views. Next, LSR [24] algorithm is performed on the concatenated data.
(3) *PVC* [17] factorizes data of each view to learn the latent subspaces independently but enforces them overlapping at entries observed in all views.
(4) *MIC* [28] separately obtains the latent feature matrices of each view and pushes them towards a common consensus with $\ell_{2,1}$ regularization.
(5) *IMG* [40], based on PVC [17], employs the graph laplacian to regularize the latent subspaces of each data view.
(6) *DAIMC* [11] is developed on weighted semi-nonnegative matrix factorization. Specifically, it adopts the given instance alignment information to learn a common latent feature matrix, while constructs a consensus basis matrix with $\ell_{2,1}$-norm to reduce the influence of missing entries.
(7) *AGL* [34] first constructs complete graphs of observed entries corresponding to each view, then extracts their partition information into a consensus representation.
(8) *AWGF* [38] utilizes feature extraction and incomplete graph fusion in a framework. A sparse regularization is employed to boost clustering performance.
(9) *PLR* [18] adopts the weighted semi-nonnegative matrix factorization model to handle incomplete multi-view data. Upon the consensus representation matrix, the locality graph is constructed to regularize the shared feature matrix.

We directly use the codes publicly available on corresponding authors' websites, and grid search their parameters to report the best. In the proposed algorithm, we set $\lambda \in 2.\hat{}[-10, -8, \cdots, 10]$ and $\gamma \in 2.\hat{}[-10, -8, \cdots, 10]$. All algorithms are evaluated with three widely used metrics, i.e. accuracy (ACC), normalized mutual information (NMI) and purity. Moreover, corresponding code is publicly available on Github[9].

## 5.2 Result

*5.2.1 Ablation Study.* In literature, most subspace clustering methods in incomplete multi-view setting are developed on matrix factorization technique, with few ones on self-representation learning. However, the later technique benefits from the built coefficient matrix which keeps the precise relationships among data entries, and, as observed in related researches, achieves better clustering performance. To validate this in experiment, we design an ablation study by comparing the clustering performance of factorization based and self-representation subspace clustering. For the sake of fairness, we compare the original NMF [3] of Eq. (3) and LSR [24] of Eq. (8) in incomplete settings, since corresponding multi-view approaches appear with different regularizations, influencing the performances to different degrees. Similar to LSR, NMF is also run in two settings, including the single best and concatenated, as introduced in Section 5.1.2. For the ease of comparison, we report the performance averages over five missing ratios in Table 2, where the best results are marked in bold. Although LSRs and LSRc achieve worse results on *ORL* dataset, it consistently outperforms NMFs and NMFc on *BBCSport*, *Yale*, *Olympics*, *Still*, *BBC*, *Buaa* and *Leaves* in three metrics. This well validates the aforementioned claim and provides a straightforward basis of our motivation to exploring self-representation subspace clustering in incomplete multi-view setting.

*5.2.2 Performance Comparison.* In order to validate effectiveness and superiority of the proposed algorithm, we carefully compare it with two baselines and seven advances in literature. Table 3 summarizes the results across all missing ratios and reports their averages where the best results are marked in bold and the second best with underline. Note that '-' indicates the algorithm reports errors on corresponding dataset. Meanwhile, Fig. 2 presents the clustering accuracy variation with respect to data missing ratio on eight datasets. The NMI and purity comparisons are shown in *Appendix* due to space limit. In the following are four observations:

(1) The proposed algorithm achieves better clustering performances than two baselines, i.e. LSRs and LSRc, by large margins. In specific, 16.55%, 19.98%, 23.13%, 12.66%, 3.24%, 31.24%, 0.81% and 10.43% ACC, 26.48%, 18.35%, 17.59%, 12.49%, 2.03%, 35.01%, 0.28% and 8.44% NMI and 18.28%, 19.46%, 24.22%, 13.28%, 2.83%, 26.26%, 0.57% and 11.34% purity are observed on eight datasets in average. This illustrates that IMSC integrates the beneficial information of different data views in

[4]http://mlg.ucd.ie/aggregation/
[5]https://www.di.ens.fr/willow/research/stillactions/
[6]http://mlg.ucd.ie/datasets/bbc.html
[7]https://github.com/hdzhao/IMG/
[8]https://github.com/cswanghao/gbs/blob/
[9]https://github.com/liujiyuan13/IMSR-code_release

**Table 2: Average performance comparison between NMF and LSR. The postfix 's' and 'c' refer to the single best and concatenated settings as introduced in Section 5.1.2.**

| Dataset | ACC | | | | NMI | | | | Purity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | NMFs | LSRs | NMFc | LSRc | NMFs | LSRs | NMFc | LSRc | NMFs | LSRs | NMFc | LSRc |
| BBCSport | 39.31 | **60.17** | 30.86 | **38.97** | 16.73 | **42.88** | 6.39 | **17.16** | 47.41 | **68.10** | 37.76 | **47.07** |
| Yale | 36.36 | **50.30** | 27.39 | **42.30** | 41.11 | **52.78** | 31.31 | **44.85** | 36.73 | **51.03** | 28.61 | **43.15** |
| ORL | **41.85** | 41.50 | **41.30** | 32.20 | 59.00 | **60.98** | **58.45** | 53.31 | **44.50** | 43.35 | **44.20** | 34.70 |
| Olympics | 42.63 | **60.52** | 33.28 | **59.53** | 52.16 | **66.45** | 43.08 | **69.67** | 51.51 | **65.82** | 42.59 | **69.09** |
| Still | 25.95 | **29.85** | 24.67 | **27.54** | 5.54 | **9.56** | 4.49 | **6.56** | 28.57 | **32.38** | 26.34 | **29.98** |
| BBC | 33.46 | **57.11** | 25.84 | **54.69** | 5.58 | **35.74** | 0.71 | **33.01** | 38.69 | **62.19** | 33.28 | **58.83** |
| Buaa | 1.93 | **40.50** | 1.93 | **52.61** | 13.07 | **67.57** | 13.07 | **76.78** | 11.70 | **42.06** | 11.70 | **55.13** |
| Leaves | 37.75 | **41.39** | 21.55 | **21.96** | 63.16 | **65.06** | 49.54 | **50.75** | 39.28 | **42.69** | 22.43 | **22.85** |

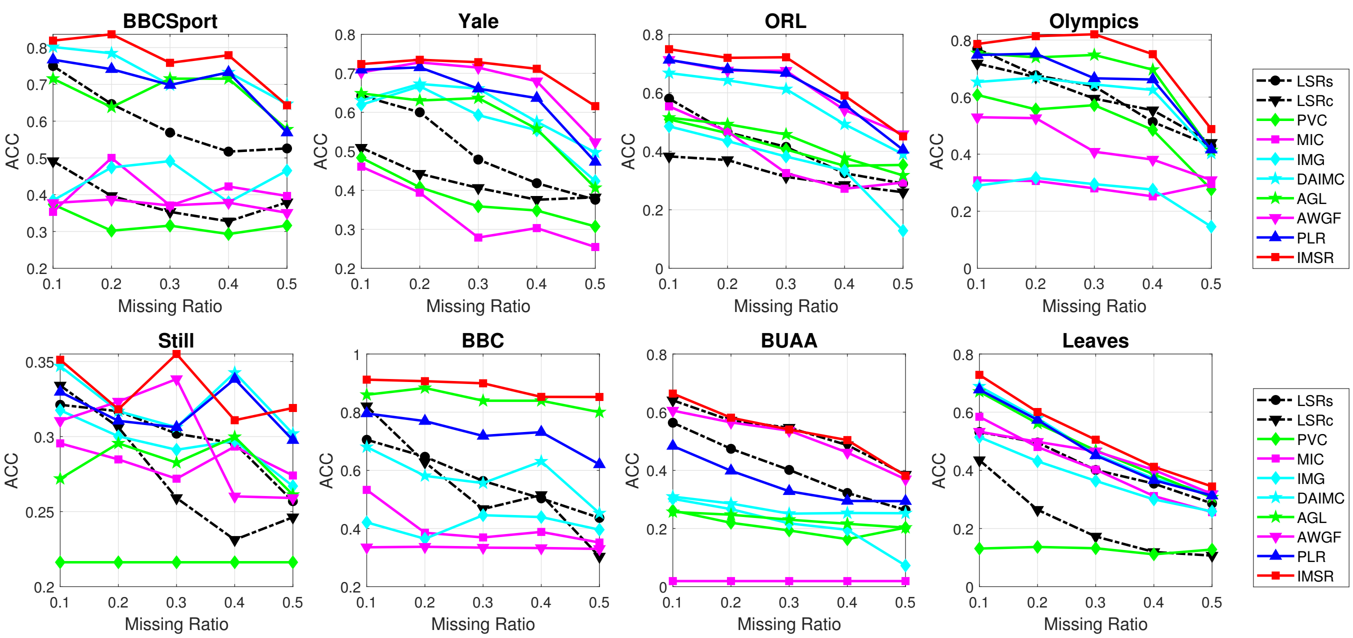


**Figure 2: Performance (ACC) comparison with advances in literature. PVC is not presented on *BBC* due to time out error.**

a more proper way than directly concatenating them into a single feature vector.

(2) The proposed algorithm consistently outperforms seven comparative methods across all missing ratios and datasets. To a closer look at the average performances, it exceeds best results of the others by 3.45%, 3.32%, 3.36%, 6.15%, 0.80%, 4.01%, 0.81% and 3.49% ACC, 8.57%, 1.02%, 1.78%, 5.07%, 1.00%, 4.89%, 0.28% and 1.90% NMI and 5.17%, 3.15%, 1.99%, 7.07%, 0.30%, 4.01%, 0.57% and 3.35% purity, verifying its effectiveness and superiority over the current literatures.

(3) The baselines obtains better results than several comparative methods on specific datasets at a large set of missing ratios. For instance, LSRs exceeds MIC, IMG and AWGF across all missing settings. It is worth to note that comparative methods are developed on matrix factorization, which demonstrates the merits of self-representation learning on incomplete multi-view data again.

(4) PVC and MIC method fails on *Still* and *Buaa*, respectively, by achieving worse results far from the averages of the other algorithms at all missing ratios. Meanwhile, the similar consequence can be observed in AGL on *Leaves*. These observations illustrate the instability of corresponding comparative methods, but conversely demonstrate the superiority of the proposed algorithm.

Overall, the proposed algorithm outperforms the baselines and literature advances consistently, verifying its superiority.

*5.2.3 Parameter Study and Convergence.* In addition, a parameter study is conducted on the proposed method. Specifically, we perform grid search on the two parameters, i.e. $\lambda$ and $\gamma$. Corresponding results on *BBCSport* are shown on the left of Fig. 3. The red color refers to ACC variation of IMSR with respect to $\lambda$ by fixing $\gamma = 2^0$, while blue indicates the one respect to $\gamma$ by fixing $\lambda = 2^1$. Nevertheless, the baselines are marked in black and set to their best

**Table 3: Average performance comparison with advances in literature. The best results are marked in bold and the second with underline. In addition, '-' indicates the algorithm fails on corresponding dataset due to time out error.**

| | Dataset | LSRs[24] | LSRc[24] | PVC[17] | MIC[28] | IMG[40] | DAIMC[11] | AGL[34] | AWGF[38] | PLR[18] | IMSR |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ACC | BBCSport | 60.17 | 38.97 | 32.03 | 40.86 | 43.90 | 73.28 | 67.24 | 37.29 | 70.17 | **76.72** |
| | Yale | 50.30 | 42.30 | 38.10 | 33.82 | 57.13 | 60.73 | 57.58 | 66.96 | 63.88 | **70.28** |
| | ORL | 41.50 | 32.20 | 41.61 | 38.24 | 35.30 | 56.10 | 43.20 | 61.27 | 60.50 | **64.63** |
| | Olympics | 60.52 | 59.53 | 49.95 | 28.84 | 26.47 | 59.87 | 67.03 | 43.09 | 64.87 | **73.17** |
| | Still | 29.85 | 27.54 | 21.63 | 28.39 | 29.46 | 32.29 | 28.22 | 29.83 | 31.65 | **33.09** |
| | BBC | 57.11 | 54.69 | - | 40.55 | 41.34 | 58.01 | 84.44 | 33.40 | 72.70 | **88.45** |
| | Buaa | 40.50 | 52.61 | 20.83 | 1.93 | 21.10 | 27.05 | 23.08 | 50.74 | 35.99 | **53.41** |
| | Leaves | 41.39 | 21.96 | 12.78 | 40.70 | 37.36 | 48.33 | 47.88 | 44.34 | 47.63 | **51.81** |
| NMI | BBCSport | 42.88 | 17.16 | 7.21 | 17.18 | 23.04 | 60.49 | 60.10 | 14.51 | 59.51 | **69.06** |
| | Yale | 52.78 | 44.85 | 42.97 | 37.43 | 60.62 | 64.86 | 61.34 | 70.11 | 66.41 | **71.13** |
| | ORL | 60.98 | 53.31 | 60.90 | 57.86 | 52.55 | 73.73 | 63.80 | 76.79 | 75.83 | **78.57** |
| | Olympics | 66.45 | 69.67 | 62.62 | 41.77 | 35.12 | 72.44 | 74.28 | 55.92 | 77.09 | **82.16** |
| | Still | 9.56 | 6.56 | 0.68 | 8.58 | 10.13 | 10.60 | 7.91 | 8.90 | 10.52 | **11.59** |
| | BBC | 35.74 | 33.01 | - | 20.20 | 18.09 | 40.92 | 65.87 | 3.95 | 53.31 | **70.75** |
| | Buaa | 67.57 | 76.78 | 57.13 | 22.15 | 44.93 | 62.45 | 60.21 | 71.77 | 67.91 | **77.06** |
| | Leaves | 65.06 | 50.75 | 39.41 | 66.52 | 60.75 | 71.59 | 71.32 | 56.40 | 70.61 | **73.49** |
| Purity | BBCSport | 68.10 | 47.07 | 39.23 | 43.45 | 45.10 | 81.21 | 78.28 | 43.19 | 80.00 | **86.38** |
| | Yale | 51.03 | 43.15 | 39.71 | 34.91 | 57.75 | 61.58 | 58.55 | 67.52 | 64.36 | **70.67** |
| | ORL | 43.35 | 34.70 | 44.95 | 40.89 | 40.06 | 59.70 | 45.65 | 65.58 | 63.55 | **67.57** |
| | Olympics | 65.82 | 69.09 | 60.14 | 40.09 | 32.88 | 70.78 | 72.76 | 53.32 | 75.30 | **82.37** |
| | Still | 32.38 | 29.98 | 21.71 | 30.75 | 31.15 | 34.90 | 31.13 | 31.76 | 34.30 | **35.20** |
| | BBC | 62.19 | 58.83 | - | 47.09 | 43.33 | 61.17 | 84.44 | 36.02 | 73.69 | **88.45** |
| | Buaa | 42.06 | 55.13 | 21.99 | 11.70 | 25.11 | 29.11 | 24.53 | 53.69 | 38.04 | **55.70** |
| | Leaves | 42.69 | 22.85 | 14.84 | 43.47 | 40.23 | 50.68 | 50.13 | 46.91 | 49.85 | **54.03** |

values, since they do not share the same parameter setting with our algorithm. We can see that the proposed method keeps relatively stable when $\lambda, \gamma \in 2.\hat{}[-10, -8, \cdots, 0]$. Therefore, they are recommended to set from 0 to 1 in practice. At the same time, we visualize objective values of the proposed algorithm by iterating the alternate procedure 50 times. It is obvious that the objective monotonically decreases along with iteration to a minimum.

## 6 CONCLUSION

Current subspace clustering advances mostly adopt the matrix factorization technique to deal with incomplete multi-view data, leaving self-representation approaches unexplored. However, self-representation subspace clustering is observed to enjoy a better performance than the former one. Therefore, we, to the first attempt, overcome the gap that it build the coefficient matrix by explicitly relying on all data entries which is contradictory to incomplete setting. In specific, the missing data imputation and self-representation learning are elegantly utilized into a cyclical procedure. Additionally, extensive experiments are conducted to validate effectiveness of the proposed algorithm.

## ACKNOWLEDGMENTS

The work is supported by National Key R&D Program of China (No. 2020AAA0107100), National Natural Science Foundation of China (No. 61922088, 61773392, 61872377, 61976196, 62006237 and 62006236), Education Ministry-China Mobile Research Funding (No. MCM20170404), Hunan Provincial Natural Science Foundation (No. 2020JJ5673) and NUDT Research Project (No. ZK20-10).

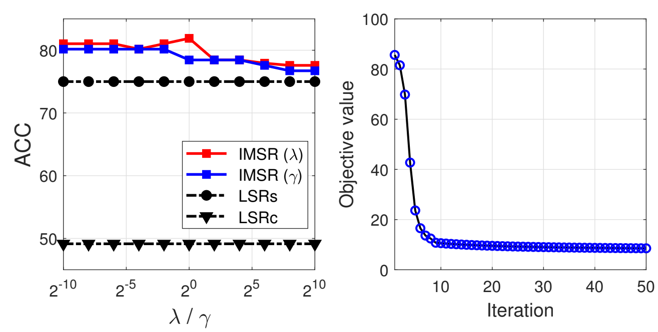


**Figure 3: Parameter study (left) and objective value variation along with iterations (right) on *BBCSport*.**

## REFERENCES

[1] Thibaut Beghin, James S. Cope, Paolo Remagnino, and Sarah Barman. 2010. Shape and Texture Based Plant Leaf Classification. In *Advanced Concepts for Intelligent Vision Systems - 12th International Conference, ACIVS 2010, Sydney, Australia, December 13-16, 2010, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 6475)*. Springer, 345–353. https://doi.org/10.1007/978-3-642-17691-3_32

[2] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman. 1997. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans.*

*Pattern Anal. Mach. Intell.* 19, 7 (1997), 711–720. https://doi.org/10.1109/34.598228

[3] Michael W. Berry, Murray Browne, Amy Nicole Langville, V. Paul Pauca, and Robert J. Plemmons. 2007. Algorithms and applications for approximate nonnegative matrix factorization. *Comput. Stat. Data Anal.* 52, 1 (2007), 155–173. https://doi.org/10.1016/j.csda.2006.11.006

[4] Yang Cai, Yuanyuan Jiao, Wenzhang Zhuge, Hong Tao, and Chenping Hou. 2018. Partial multi-view spectral clustering. *Neurocomputing* 311 (2018), 316–324. https://doi.org/10.1016/j.neucom.2018.05.053

[5] Vincent Delaitre, Ivan Laptev, and Josef Sivic. 2010. Recognizing human actions in still images: a study of bag-of-features and part-based representations. In *British Machine Vision Conference, BMVC 2010, Aberystwyth, UK, August 31 - September 3, 2010. Proceedings.* British Machine Vision Association, 1–11. https://doi.org/10.5244/C.24.97

[6] Chris H. Q. Ding, Tao Li, Wei Peng, and Haesun Park. 2006. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, August 20-23, 2006.* ACM, 126–135. https://doi.org/10.1145/1150402.1150420

[7] Ehsan Elhamifar and René Vidal. 2009. Sparse subspace clustering. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA.* IEEE Computer Society, 2790–2797. https://doi.org/10.1109/CVPR.2009.5206547

[8] Derek Greene and Padraig Cunningham. 2006. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Machine Learning, Proceedings of the Twenty-Third International Conference (ICML 2006), Pittsburgh, Pennsylvania, USA, June 25-29, 2006 (ACM International Conference Proceeding Series, Vol. 148).* ACM, 377–384. https://doi.org/10.1145/1143844.1143892

[9] Derek Greene and Pádraig Cunningham. 2013. Producing a unified graph representation from multiple social network views. In *Web Science 2013 (co-located with ECRC), WebSci '13, Paris, France, May 2-4, 2013.* ACM, 118–121. https://doi.org/10.1145/2464464.2464471

[10] Stephan Günnemann, Ines Färber, Sebastian Raubach, and Thomas Seidl. 2013. Spectral Subspace Clustering for Graphs with Feature Vectors. In *2013 IEEE 13th International Conference on Data Mining, Dallas, TX, USA, December 7-10, 2013.* IEEE Computer Society, 231–240. https://doi.org/10.1109/ICDM.2013.110

[11] Menglei Hu and Songcan Chen. 2018. Doubly Aligned Incomplete Multi-view Clustering. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.* ijcai.org, 2262–2268. https://doi.org/10.24963/ijcai.2018/313

[12] Menglei Hu and Songcan Chen. 2019. One-Pass Incomplete Multi-View Clustering. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019.* AAAI Press, 3838–3845. https://doi.org/10.1609/aaai.v33i01.33013838

[13] Zhao Kang, Zipeng Guo, Shudong Huang, Siying Wang, Wenyu Chen, Yuanzhang Su, and Zenglin Xu. 2019. Multiple Partitions Aligned Clustering. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019.* ijcai.org, 2701–2707. https://doi.org/10.24963/ijcai.2019/375

[14] Zhao Kang, Zhiping Lin, Xiaofeng Zhu, and Wenbo Xu. 2021. Structured Graph Learning for Scalable Subspace Clustering: From Single-view to Multi-view. *CoRR* abs/2102.07943 (2021). arXiv:2102.07943 https://arxiv.org/abs/2102.07943

[15] Zhao Kang, Xinjia Zhao, Chong Peng, Hongyuan Zhu, Joey Tianyi Zhou, Xi Peng, Wenyu Chen, and Zenglin Xu. 2020. Partition level multiview subspace clustering. *Neural Networks* 122 (2020), 279–288. https://doi.org/10.1016/j.neunet.2019.10.010

[16] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. 2020. Large-Scale Multi-View Subspace Clustering in Linear Time. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, New York, NY, USA, February 7-12, 2020.* AAAI Press, 4412–4419. https://aaai.org/ojs/index.php/AAAI/article/view/5867

[17] Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. 2014. Partial Multi-View Clustering. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.* AAAI Press, 1968–1974. http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8241

[18] Huiqiang Lian, Huiying Xu, Siwei Wang, Miaomiao Li, Xinzhong Zhu, and Xinwang Liu. 2021. Partial multiview clustering with locality graph regularization. *International Journal of Intelligent Systems* (2021). https://doi.org/10.1002/int.22409

[19] Jiyuan Liu, Xinwang Liu, Siwei Wang, Sihang Zhou, and Yuexiang Yang. 2021. Hierarchical Multiple Kernel Clustering. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Event, February 2-9, 2021.* AAAI Press, 8671–8679. https://ojs.aaai.org/index.php/AAAI/article/view/17051

[20] Jiyuan Liu, Xinwang Liu, Jian Xiong, Qing Liao, Sihang Zhou, Siwei Wang, and Yuexiang Yang. 2020. Optimal Neighborhood Multiple Kernel Clustering with Adaptive Local Kernels. *IEEE Transactions on Knowledge and Data Engineering* (2020), 1–1. https://doi.org/10.1109/TKDE.2020.3014104

[21] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Xifeng Guo, Marius Kloft, and Liangzhong He. 2021. Multiview Subspace Clustering via Co-Training Robust Data Representation. *IEEE Transactions on Neural Networks and Learning Systems* (2021), 1–13. https://doi.org/10.1109/TNNLS.2021.3069424

[22] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, Chang Tang, Jianping Yin, Dinggang Shen, Huaimin Wang, and Wen Gao. 2019. Late Fusion Incomplete Multi-View Clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 10 (2019), 2410–2423. https://doi.org/10.1109/TPAMI.2018.2879108

[23] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, En Zhu, Tongliang Liu, Marius Kloft, Dinggang Shen, Jianping Yin, and Wen Gao. 2020. Multiple Kernel k-Means with Incomplete Kernels. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 5 (2020), 1191–1204. https://doi.org/10.1109/TPAMI.2019.2892416

[24] Can-Yi Lu, Hai Min, Zhong-Qiu Zhao, Lin Zhu, De-Shuang Huang, and Shuicheng Yan. 2012. Robust and Efficient Subspace Segmentation via Least Squares Regression. In *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VII (Lecture Notes in Computer Science, Vol. 7578).* Springer, 347–360. https://doi.org/10.1007/978-3-642-33786-4_26

[25] Juncheng Lv, Zhao Kang, Xiao Lu, and Zenglin Xu. 2021. Pseudo-Supervised Deep Subspace Clustering. *IEEE Trans. Image Process.* 30 (2021), 5252–5263. https://doi.org/10.1109/TIP.2021.3079800

[26] Nishant Rai, Sumit Negi, Santanu Chaudhury, and Om Deshmukh. 2016. Partial Multi-View Clustering using Graph Regularized NMF. In *23rd International Conference on Pattern Recognition, ICPR 2016, Cancún, Mexico, December 4-8, 2016.* IEEE, 2192–2197. https://doi.org/10.1109/ICPR.2016.7899961

[27] Ferdinand Samaria and Andy Harter. 1994. Parameterisation of a stochastic model for human face identification. In *Proceedings of Second IEEE Workshop on Applications of Computer Vision, WACV 1994, Sarasota, FL, USA, December 5-7, 1994.* IEEE, 138–142. https://doi.org/10.1109/ACV.1994.341300

[28] Weixiang Shao, Lifang He, and Philip S. Yu. 2015. Multiple Incomplete Views Clustering via Weighted Nonnegative Matrix Factorization with $L_2$, 1 Regularization. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 9284).* Springer, 318–334. https://doi.org/10.1007/978-3-319-23528-8_20

[29] Hao Wang, Linlin Zong, Bing Liu, Yan Yang, and Wei Zhou. 2019. Spectral Perturbation Meets Incomplete Multi-view Data. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019.* ijcai.org, 3677–3683. https://doi.org/10.24963/ijcai.2019/510

[30] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. 2021. Generative Partial Multi-View Clustering With Adaptive Fusion and Cycle Consistency. *IEEE Trans. Image Process.* 30 (2021), 1771–1783. https://doi.org/10.1109/TIP.2020.3048626

[31] Xiaokan Wang, Xia Mao, and Catalin-Daniel Caleanu. 2010. Nonlinear Shape-Texture Manifold Learning. *IEICE Trans. Inf. Syst.* 93-D, 7 (2010), 2016–2019. https://doi.org/10.1587/transinf.E93.D.2016

[32] Yang Wang, Xuemin Lin, Lin Wu, Wenjie Zhang, and Qing Zhang. 2014. Exploiting Correlation Consensus: Towards Subspace Clustering for Multi-modal Data. In *Proceedings of the ACM International Conference on Multimedia, MM '14, Orlando, FL, USA, November 03 - 07, 2014.* ACM, 981–984. https://doi.org/10.1145/2647868.2654999

[33] Yang Wang, Lin Wu, Xuemin Lin, and Junbin Gao. 2018. Multiview Spectral Clustering via Structured Low-Rank Matrix Factorization. *IEEE Trans. Neural Networks Learn. Syst.* 29, 10 (2018), 4833–4843. https://doi.org/10.1109/TNNLS.2017.2777489

[34] J. Wen, Y. Xu, and H. Liu. 2020. Incomplete Multiview Spectral Clustering With Adaptive Graph Learning. *IEEE Transactions on Cybernetics* 50, 4 (2020), 1418–1429. https://doi.org/10.1109/TCYB.2018.2884715

[35] Jie Wen, Zheng Zhang, Yong Xu, and Zuofeng Zhong. 2018. Incomplete Multi-view Clustering via Graph Regularized Matrix Factorization. In *Computer Vision - ECCV 2018 Workshops - Munich, Germany, September 8-14, 2018, Proceedings, Part IV (Lecture Notes in Computer Science, Vol. 11132).* Springer, 593–608. https://doi.org/10.1007/978-3-030-11018-5_47

[36] Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu. 2020. Generalized Latent Multi-View Subspace Clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 1 (2020), 86–99. https://doi.org/10.1109/TPAMI.2018.2877660

[37] Changqing Zhang, Zongbo Han, Yajie Cui, Huazhu Fu, Joey Tianyi Zhou, and Qinghua Hu. 2019. CPM-Nets: Cross Partial Multi-View Networks. In *Advances in Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada.* 557–567. https://proceedings.neurips.cc/paper/2019

[38] Pei Zhang, Siwei Wang, Jingtao Hu, Zhen Cheng, Xifeng Guo, En Zhu, and Zhiping Cai. 2020. Adaptive Weighted Graph Fusion Incomplete Multi-View Subspace Clustering. *Sensors* 20, 20 (2020), 5755. https://doi.org/10.3390/s20205755

[39] Zheng Zhang, Li Liu, Fumin Shen, Heng Tao Shen, and Ling Shao. 2019. Binary Multi-View Clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 7 (2019), 1774–1782. https://doi.org/10.1109/TPAMI.2018.2847335

[40] Handong Zhao, Hongfu Liu, and Yun Fu. 2016. Incomplete Multi-Modal Visual Data Grouping. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016.* IJCAI/AAAI Press, 2392–2398. http://www.ijcai.org/Abstract/16/341