RPCA-Induced Graph Tensor Learning for Incomplete Multi-view Inferring and Clustering

Xingfeng Li[1], Yinghui Sun[1], Zhenwen Ren[1,2(✉)], and Quansen Sun[1(✉)]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
{yinghuisun,rzw,sunquansen}@njust.du.cn

[2] Department of National Defence Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China

Abstract. The existing incomplete multi-view graph clustering (IMGC) methods mainly focus on leveraging available samples among different views to explore the weak local structure information, resulting in inexact or unreliable affinity graphs for clustering. More importantly, they fail to exploit the spatial structure information among graphs. To graciously address both, in this paper, we propose a novel IMGC method, claimed as robust principle component analysis (RPCA)-induced graph tensor learning (RPCA-IGTL) for incomplete multi-view inferring and clustering. This model can synchronously perform missing part filling, complete data inferring, and diagonalized graph tensor learning to obtain the more exact and reliable affinity graphs for clustering. Especially, the proposed method first designs a RPCA-induced local manifold learning framework, which bridges complete data inferring in feature space and the diagonalized graph learning in graph semantic space. Both objects can boost each other to fully exploit the underlying local structure information among incomplete view data. Besides, a t-SVD based tensor low-rank constraint introduces to exploit the spatial structure of graph tensor and complementary information of diagonalized affinity graphs. Extensive experiments have demonstrated the effectiveness of our method compared to the previous state-of-the-art methods.

Keywords: Incomplete multi-view inferring and clustering · Robust principle component analysis · Low-rank graph tensor

1 Introduction

Multi-view clustering (MVC) has become an increasingly pervasive research topic due to its powerful ability of handling multi-view data for improving clustering performance during the past few decades [4,19]. The implementation of these methods quite depends on an assumption that views among samples are visible. However, this assumption will not hold anymore owing to the absence of multi-view data, especially in certain practical applications [9,22]. This tremendously limits the aforementioned MVC methods to handle incomplete MVC (IMVC) tasks.

H. Zhang et al. (Eds.): NCAA 2022, CCIS 1637, pp. 85–99, 2022.
https://doi.org/10.1007/978-981-19-6142-7_7

Recently, lots of IMVC methods have been proposed to alleviate the absence problem of data and can be roughly divided into three main categories: kernel based IMVC, non-negative matrix factorization (NMF) based IMVC (NMF-IMVC) [3], and graph based IMVC, also known as incomplete multi-view graph clustering (IMGC) [26,27]. For first category, [16] first proposes to complete kernel Gram matrix to deal with incomplete kernel problem for IMVC. Unfortunately, it is merely appropriate for the incomplete case that at least one base kernel is complete, which greatly prevents its application in practice. To this end, many scholars propose absent multiple kernel learning models for incomplete clustering. Typically, multiple kernel k-means with incomplete kernels clustering (MKK-IKC) [10] proposes to learn the kernel coefficients and the clustering partition matrix in a joint objection to fill the incomplete base kernels and clustering. However, it is great difficult for kernel based IMVC methods to tune the suitable kernel parameters for predefined base kernels.

For second category, most of them directly fuse multiple complementary representations into a low-dimensional consensus representation for clustering [5]. Another methods use mean values or zeros to fill the incomplete parts and then couple a weighted strategy [2,18,21,23]. These methods can directly obtain a consensus representation for reducing the impact of the incomplete samples of different views. Despite its effectiveness, NMF-IMVC methods always obtain an uncompacted representation since they neglect the intrinsic (such as inter-view) structure of multi-view data. For instance, incomplete multi-view clustering with flexible locality structure diffusion (IMVC-FLSD) [23] integrates the consensus representation learning and the objection weighted learning into a unified model to explore the certain paired similarity in different views.

Compared to most NMF-IMVC methods, IMGC methods can better preserve the geometric structure of data, especially for the local geometric structure among data in graph learning process, which has been proved to very significant for clustering [20,22,25]. Incomplete multi-view spectral clustering with adaptive graph learning (IMVSC-AGL) [20] directly uses available multi-view data to generate a consensus representation for clustering. To sum up, the above mentioned IMVC methods always focus on adequately exploring the effective information of available samples to obtain preferable clustering, while lacking of the ability of recovering missing views.

To this end, a unified embedding alignment framework (UEAF) [22] attempts to address IMVC problem by recovering the missing views rather than only leveraging available views data. However, it is seriously sensitive to noise and outliers, resulting in the inferior affinity graph and clustering results.

To address above problems, we propose a novel IMGC method, namely Robust Principle Component Analysis Induced Graph Tensor Learning (RPCA-IGTL), to synergistically recover missing views and infer complete data without noise for improve the IMVC. Particularly, we first extend the traditional RPCA model to manifold regularized RPCA framework for handling the challenging incomplete multi-view recovering problem. And then, we couple the adaptive neighbors graph learning with manifold regularized RPCA framework into a unified objective function. Besides, an enhanced block diagonal constraint and t-SVD based tensor low-rank constraint are simultaneously imposed on the learning of affinity graph, so as to guide the missing data recovering and complete data inferring. In summary, the contributions of this paper mainly include the following:

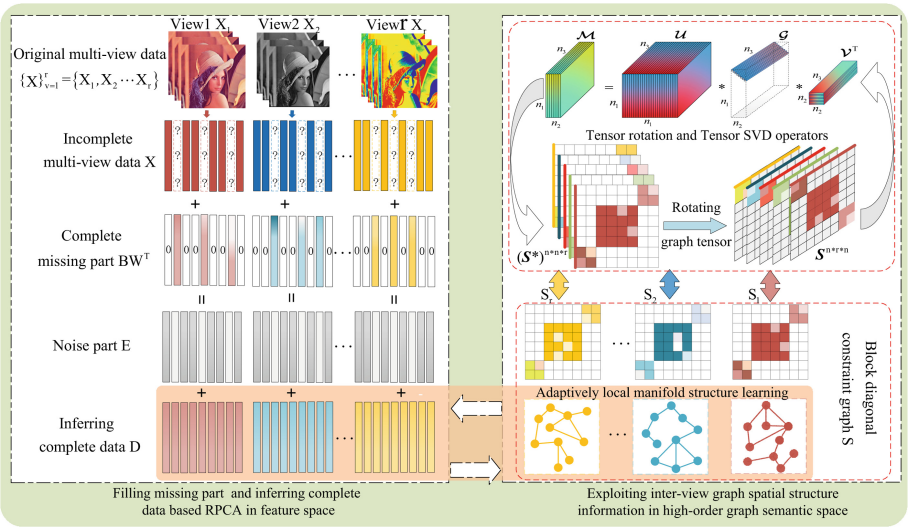


Fig. 1. Framework of the proposed RPCA-IGTL.

- To address the original incomplete and noise data, we design an elegant RPCA-induced manifold learning framework. It integrates missing part completing, low-rank complete data inferring, and local manifold learning into a unified objective function, such that the more exact and reliable graphs can be learned for improving the clustering performance.
- Instead of suboptimal block diagonal regularizer with extra parameter, a more flexible block diagonal constraint with parameter free is considered to precisely control the number of diagonal blocks of graphs in framework, which is greatly significant for clustering. Further, with the t-SVD based tensor low-rank constraint on the learned diagonal graphs, the spatial structure and complementary information among incomplete graphs can be exploited as possible.
- Compared to two base line and several state-of-the-art IMVC and IMGC methods, our method achieves important improvement on various scale datasets.

2 Related Work

2.1 Notation Summary

Through this paper, we denote 3-order tensor, matrix and vector as $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathbf{M}^{n_1 \times n_2}$, $\mathbf{M}$, respectively. $\mathbf{M} \succeq 0$ is the positive semi-definite matrix. The $\boldsymbol{\mathcal{M}}_{i,j,k}$ is the element of $\boldsymbol{\mathcal{M}}$, where the fiber is denoted as $\boldsymbol{\mathcal{M}}(:,j,k)$, $\boldsymbol{\mathcal{M}}(i,:,k)$, and $\boldsymbol{\mathcal{M}}(i,j,:)$, as well as the slice is denoted as horizontal slice $\boldsymbol{\mathcal{M}}(i,:,:)$, lateral slice $\boldsymbol{\mathcal{M}}(:,j,:)$, and frontal slice $\boldsymbol{\mathcal{M}}(:,:,k)$. For convenience, $\boldsymbol{\mathcal{M}}(:,:,k)$ is simplified as $\mathbf{M}^k$ or $\boldsymbol{\mathcal{M}}^k$. $\boldsymbol{\mathcal{M}}_f = \texttt{fft}(\boldsymbol{\mathcal{M}}, [\,], 3)$ and $\boldsymbol{\mathcal{M}} = \texttt{ifft}(\boldsymbol{\mathcal{M}}_f, [\,], 3)$ are the fast Fourier transformation (FFT) and inverse FFT along the third direction of tensor $\boldsymbol{\mathcal{M}}$, respectively.

The tensor singular value decomposition (t-SVD) and t-SVD based tensor nuclear norm (t-TNN) are defined as the following

Definition 2 (t-SVD-Based Tensor Nuclear Norm, t-TNN). $\|\boldsymbol{\mathcal{M}}\|_{\circledast}$ is the t-SVD based tensor nuclear norm of $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, which is denoted as the sum of singular values of all the frontal slices of $\boldsymbol{\mathcal{M}}_f$, *i.e.*,

$$\|\boldsymbol{\mathcal{M}}\|_{\circledast} = \sum_{k=1}^{n_3} \left\| \boldsymbol{\mathcal{M}}_f^k \right\|_* = \sum_{i=1}^{\min(n_1,n_2)} \sum_{k=1}^{n_3} \left| \boldsymbol{\mathcal{S}}_f^k(i,i) \right| \tag{1}$$

where $\boldsymbol{\mathcal{S}}_f^k$ is obtained by t-SVD of frontal slices of $\boldsymbol{\mathcal{M}}_f$, *i.e.*, $\boldsymbol{\mathcal{M}}_f^k = \boldsymbol{\mathcal{U}}_f^k \boldsymbol{\mathcal{S}}_f^k \boldsymbol{\mathcal{V}}_f^{k\mathsf{T}}$. $\boldsymbol{\mathcal{U}}_f^k$ and $\boldsymbol{\mathcal{V}}_f^k$ are the corresponding left and right singular value matrices.

2.2 Adaptive Neighbors Graph Learning (ANGL)

By adaptively assigning an affinity value for each sample as neighbor value of another sample, adaptive neighbors graph learning [14] can be mathematically expressed as

$$\min_{\mathbf{S}^v} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{D}(\mathbf{x}_i^v, \mathbf{x}_j^v) s_{ij}^v \tag{2}$$

$\mathbb{D}$ is the distance criterion between any two samples, and the larger distance indicates the smaller affinity value s_{ij}^v corresponding to v-th view. (2) can effectively learn an affinity graph with local geometrical structure preserving.

2.3 Robust Principle Component Analysis (RPCA)

The model of RPCA takes the form of

$$\min_{\mathbf{D},\mathbf{E}} \|\mathbf{D}\|_* + \alpha\phi(\mathbf{E}) \ \text{ s.t. } \mathbf{X} = \mathbf{D} + \mathbf{E} \tag{3}$$

which has a strong recovery guarantees as the first polynomial-time method [1]. $\mathbf{X}$ in (3) denotes the damaged or noisy samples, and is decomposed as the low-rank sample matrix and sparse error matrix, respectively. $\phi(\cdot)$, α and $\|\cdot\|_*$ denote the certain regularizer of noise, regularization parameter and nuclear norm, respectively. Most algorithms have demonstrated that low-rank representation can well characterize the relationship between data. Low-rank representation in (3) $\mathbf{D}$ can be deemed as the clean data recovered from original data $\mathbf{X}$, which has been extensively used in video surveillance and image denoising [11].

3 Proposed Method

Different from partial multi-view learning [18], incomplete multi-view learning is a significantly challenging research topic. Since the incomplete view data (missing parts) will destroy the original affinity between multi-view data, resulting in deceptive affinity values. Further, the unreliable affinity graphs constructed from these data will cause the terrible clustering results. Thus, we develop a RPCA inferring framework (RPCA-IGTL) to simultaneously infer the missing view data and diagonalized graphs learning for clustering. Intuitively, Fig. 1 has given a clear pipeline of RPCA-IGTL.

3.1 Model of RPCA-IGTL

In complete partial multi-view graph clustering, adaptive neighbors graph learning (ANGL) becomes increasingly prevalent and has achieved the promising results recently [2]. Based on ANGL of (2), multi-view graph clustering (MVGC) can be formulated as

$$\begin{aligned} \min_{\mathbf{S}^v} \sum_{v=1}^{r} \sum_{i,j=1}^{n} \left(\|x_i^v - x_j^v\|_2^2 s_{ij}^v + \gamma {s_{ij}^v}^2 \right) \\ \text{s.t. } \forall i, s_i^\top \mathbf{1} = 1, 0 \le s_i \le 1, \mathtt{rank}(\mathbf{L}_S^v) = n - c \end{aligned} \tag{4}$$

where smaller distance between samples $\mathbf{x}_i^v$ and $\mathbf{x}_j^v$ automatically learns a larger affinity value $\mathbf{s}_{ij}^v$ to construct the affinity graph $\mathbf{S}^v$, and vice versa. γ is a parameter to control the neighbors between samples and r represents the number of view. Here, $\mathbf{L}_S^v = \mathbf{P}^v - (\mathbf{S}^v + (\mathbf{S}^v)^\top)/2$ and c are a Lagrangian matrix and the number of cluster, respectively. The degree matrix $\mathbf{P}^v$ can be computed via $p_{ii}^v = \sum_j s_{ij}^v$. $\mathtt{rank}(\mathbf{L}_S^v) = n - c$ is a widely used Lagrangian rank constraint to pursue a graph with exact connected components for improving clustering performance. Owing to Lagrangian rank constraint and the capacity of local manifold structure, the complete MVGC methods have received lots of attention recently. However, (4) needs all views data to be complete. To address this challenging problem, it urgently requires a way to infer or recover the missing view data. To this end, we consider that RPCA has a strong recovery ability as the first polynomial-time method in (3). Here, one may utilize clean data to substitute $\mathbf{X}$ in (4). Apparently, there is only available samples information without missing samples information in $\mathbf{D}$, such that the missing view information is ignored. Inspired by (3), we first design an inferring framework as mentioned in the left of Fig. 1, and this framework can be mathematically fulfilled as

$$\min_{\mathbf{D},\mathbf{E}} \|\mathbf{D}\|_* + \alpha\phi(\mathbf{E}) \ \text{ s.t. } \mathbf{X} + \mathbf{B}\mathbf{W}^\top = \mathbf{D} + \mathbf{E} \tag{5}$$

where $\mathbf{B}\mathbf{W}^\top$ is the complete missing part, and $\mathbf{W}$ is a index matrix to explicitly enforces the entries to be zeros corresponding to the missing samples from $\mathbf{B}$. $\mathbf{W}$ can be construct via

$$W_{i,j}^v = \begin{cases} 1, & \text{if the } i\text{-th missing sample is } x_j^v \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Not here that we can take full advantage of different incomplete prior information to construct corresponding index matrix $\mathbf{W}$, so as to deal with various of incomplete cases. By extending (5) to multi-view model, and we seamlessly integrate this model and (4) into a unified objection as

$$\begin{aligned} \min_{\substack{\mathbf{B}^v,\mathbf{D}^v,\mathbf{E}^v,\\ \mathbf{S}^v,\mathbf{Z}^v}} \sum_{v=1}^{r} \|\mathbf{D}^v\|_* + \alpha\|\mathbf{E}^v\|_1 + \beta \mathtt{Tr}(\mathbf{D}^v \mathbf{L}_S^v (\mathbf{D}^v)^\top) + \gamma\|\mathbf{S}^v\|_F^2 \\ \text{s.t. } \forall v, \mathbf{X}^v + \mathbf{B}^v(\mathbf{W}^v)^\top = \mathbf{D}^v + \mathbf{E}^v, \mathtt{rank}(\mathbf{L}_S^v) = n - c \end{aligned} \tag{7}$$

where β is hyper-parameter. $\phi(\cdot) = \|\cdot\|_1$ on $\mathbf{E}^v$ is l_1 norm, which can alleviate error in the available samples of $\mathbf{X}^v$ and make the most of useful information in $\mathbf{X}^v$ transfer to $\mathbf{D}^v$. Frobenius norm is imposed on $\mathbf{S}^v$ to avoid trivial solution.

Although (7) has the power of inferring the information of missing views, it still exists the following limits: First, it neglects the complementary information among different views and spatial structure of graphs. Second, 1) n is usually much larger than c, so that it is illogical to find a high-rank $\mathbf{L}_S^v$ by minimizing $\mathtt{rank}(\mathbf{L}_S^v)$; 2) more importantly, $\mathtt{rank}(\mathbf{L}_S^v) = n - c$ is not able to control the targeted number of blocks; 3) the solution of $\mathtt{rank}(\mathbf{L}_S^v) = n - c$ always introduces an extra hyper-parameter. To address these problems, we propose a novel IMGC method, named RPCA-induced graph tensor learning (RPCA-IGTL) for incomplete multi-view inferring and clustering, as follows

$$\begin{aligned} &\min_{\substack{\mathbf{B}^v,\mathbf{D}^v,\mathbf{E}^v,\\ \mathbf{Z}^v,\boldsymbol{\mathcal{S}}}} \sum_{v=1}^{r} \|\mathbf{D}^v\|_* + \alpha\|\mathbf{E}^v\|_1 + \beta\mathtt{Tr}(\mathbf{D}^v\mathbf{L}_S^v{\mathbf{D}^v}^\top) + \gamma\|\boldsymbol{\mathcal{S}}\|_{\circledast} \\ &\text{s.t. } \forall v, \mathbf{X}^v + \mathbf{B}^v(\mathbf{W}^v)^\top = \mathbf{D}^v + \mathbf{E}^v, \mathtt{Tr}(\mathbf{S}^v) = c, 0 \preccurlyeq \mathbf{S}_{ij}^v \preccurlyeq 1 (\mathbf{S}^v)^\top = \mathbf{S}^v \end{aligned} \tag{8}$$

where $\mathtt{Tr}(\cdot)$ is a trace norm. To address the first problem, we stack the r affinity graphs into a tensor $\boldsymbol{\mathcal{S}}^* \in \mathbb{R}^{n\times n\times r}$ to well capture the complementary information among different views and spatial structure information of graphs. And then $\boldsymbol{\mathcal{S}}^*$ is rotated to $\boldsymbol{\mathcal{S}}^{n\times r\times n}$ so as to better exploit inter-view information, which can simultaneously reduce the computational complexity from $\mathcal{O}(rn^2\mathtt{log}(r) + rn^3)$ to $\mathcal{O}(rn^2\mathtt{log}(n) + r^2n^2)$. Further, a t-SVD based tensor nuclear norm is imposed on $\boldsymbol{\mathcal{S}}$ inspired by following two facts: 1) the views of different $\mathbf{S}^v$, originating from the same dataset source, possess some consensus structure information; 2) the learned tensor $\boldsymbol{\mathcal{S}}$ should enjoy the low-rank property since the number of clusters c is always much less than the number of samples n. For another problem, the more exactly block diagonal constraint $\mathtt{Tr}(\mathbf{S}^v) = c, 0 \preccurlyeq \mathbf{S}_{ij}^v \preccurlyeq 1, (\mathbf{S}^v)^\top = \mathbf{S}^v$ not only directly encourages affinity graphs to be block diagonal rather than inducing an extra hyper-parameter to solve $\mathtt{rank}(\mathbf{L}_S^v) = n - c$, but also can control the number of blocks. As mentioned in the right part of Fig. 1, the imposed block diagonal constraint can control the graphs with intra-cluster dense and inter-cluster sparse as the input of tensor, which will vastly benefit to exploiting inter-view graph spatial structure information in high-order graph semantic space.

3.2 Optimization

By introducing an auxiliary variable $\boldsymbol{\mathcal{G}}$ and $\{\mathbf{Z}\}_{v=1}^r$, a seven-step alternating direction method of multipliers (ADMM) is developed to optimize non-convex problem (8) as follows

$$\begin{aligned} \min_{\substack{\mathbf{B}^v,\mathbf{D}^v,\mathbf{E}^v,\\ \mathbf{S}^v,\mathbf{Z}^v,\boldsymbol{\mathcal{G}}}} &\sum_{v=1}^{r} \|\mathbf{D}^v\|_* + \alpha\|\mathbf{E}^v\|_1 + \beta\mathtt{Tr}(\mathbf{Z}^v\mathbf{L}_S^v(\mathbf{Z}^v)^\top) + \gamma\|\boldsymbol{\mathcal{G}}\|_{\circledast} + \frac{\mu}{2}\|\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{S}} + \frac{\boldsymbol{\mathcal{Y}}_3}{\mu}\|_F^2 \\ &+ \frac{\mu}{2}\|\mathbf{X}^v + \mathbf{B}^v(\mathbf{W}^v)^\top - \mathbf{D}^v - \mathbf{E}^v + \frac{\mathbf{Y}_1^v}{\mu}\| + \frac{\mu}{2}\|\mathbf{Z}^v - \mathbf{D}^v + \frac{\mathbf{Y}_2^v}{\mu}\|_F^2 \\ \text{s.t.} \quad &\forall v, \mathbf{Z}^v = \mathbf{D}^v, \boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{S}}, \mathtt{Tr}(\mathbf{S}^v) = c, 0 \preccurlyeq \mathbf{S}_{ij}^v \preccurlyeq 1, (\mathbf{S}^v)^\top = \mathbf{S}^v \end{aligned} \tag{9}$$

where μ is a penalty parameter. $\mathbf{Y}_1^v$, $\mathbf{Y}_2^v$ and $\boldsymbol{\mathcal{Y}}_3^v$ are Lagrangian multipliers.

► **Step 1. D-subproblem:** Fixing the other variables except the recovering $\mathbf{D}$, we update $\mathbf{D}$ via

$$\min_{\mathbf{D}^{(v}}(\|\mathbf{D}^v\|_* + \frac{\mu}{2}\|\mathbf{D}^v - \mathbf{H}^v\|_F^2) \tag{10}$$

where $\mathbf{H}^v = (\mathbf{X}^v + \mathbf{B}^v(\mathbf{W}^v)^\top + \mathbf{Z}^v - \mathbf{E}^v + (\mathbf{Y}_1^v + \mathbf{Y}_2^v)/\mu)/2$, and (10) can obtain a closed-form solution via by the singular value thresholding operator [24].

► **Step 2. Z-subproblem:** By keeping the other variables unchanged except $\mathbf{Z}$, the subproblem of $\mathbf{Z}$ becomes

$$\min_{\mathbf{Z}^v} \&(\beta \mathtt{Tr}(\mathbf{Z}^v \mathbf{L}_S^v (\mathbf{Z}^{(v)})^\top) + \frac{\mu}{2}\|\mathbf{Z}^v - \mathbf{D}^v + \frac{\mathbf{Y}_2^v}{\mu}\|_F^2) \tag{11}$$

By taking the derivative of $\mathbf{Z}$ and setting it to be zeros, we then have $\mathbf{Z}^{(v)^*} = (\mu \mathbf{D}^v - \mathbf{Y}_2^v)(2\beta \mathbf{L}_S^v + \mu \mathbf{I})^{-1}$.

► **Step 3. E-subproblem:** The other variables are fixed except $\mathbf{E}$, the $\mathbf{E}$ subproblem of problem (9) reduces to

$$\min_{\mathbf{E}^v} \&\alpha\|\mathbf{E}^v\|_1 + \frac{\mu}{2}\|\mathbf{E}^v - \mathbf{C}^v\|_F^2 \tag{12}$$

where $\mathbf{C}^v = \mathbf{X}^v + \mathbf{B}^v(\mathbf{W}^v)^\top - \mathbf{D}^v + \frac{\mathbf{Y}_1^v}{\mu}$. Such subproblem can first be written in vector form, and efficiently solved by Lemma 3.2 in [8].

► **Step 4. B-subproblem:** The inferring matrix $\mathbf{B}$ can be obtained by fixing the other variables except $\mathbf{B}$ as follows

$$\min_{\mathbf{B}^v} \frac{\mu}{2}\|\mathbf{X}^v + \mathbf{B}^v(\mathbf{W}^v)^\top - \mathbf{D}^v - \mathbf{E}^v + \frac{\mathbf{Y}_1^v}{\mu}\|_F^2 \tag{13}$$

By setting the first-order derivative of $\mathbf{B}$ to zeros, we can obtain the closed-form solution as $\mathbf{B}^v = -(\mathbf{X}^v - \mathbf{D}^v - \mathbf{E}^v + \mathbf{Y}_1^v/\mu) * \mathbf{W}^v * ((\mathbf{W}^v)^\top \mathbf{W}^v)^{-1}$.

► **Step 5. S-subproblem:** The other variables remain unchanged except $\mathbf{S}$, subproblem of $\mathbf{S}$ can be reformulated as

$$\begin{aligned} &\min_{\mathbf{S}^v} \ \beta \sum_{v=1}^{r} \mathtt{Tr}(\mathbf{Z}^v \mathbf{L}_S^v (\mathbf{Z}^v)^\top) + \frac{\mu}{2}\|\mathbf{G}^v - \mathbf{S}^v + \mathbf{Y}_3^v/\mu\|_F^2 \\ \Rightarrow &\min_{\mathbf{S}^v} \|\mathbf{S}^v - \mathbf{A}^v\| \\ &\text{s.t.} \quad \forall v, \mathtt{Tr}(\mathbf{S}^v) = c, 0 \preccurlyeq \mathbf{S}_{ij}^v \preccurlyeq 1, (\mathbf{S}^v)^\top = \mathbf{S}^v \end{aligned} \tag{14}$$

where $\mathbf{A}^v = \mathbf{G}^v + \mathbf{Y}_3^v/\mu - \frac{\beta}{\mu} * \mathbf{Q}^v$, and $q_{ij}^v = \|z_i^v - z_j^v\|_2^2$. q_{ij}^v is the ij-th element of $\mathbf{Q}^v$. For each $\mathbf{S}^v$, the optimal solution can be obtained by Theorem 1.

Theorem 1 *For a symmetric affinity matrix $\mathbf{S} \in \mathbb{R}^{n\times n}$, the spectral decomposition of $\mathbf{S}$ is denoted as $\mathbf{A} = \mathbf{U}\mathtt{Diag}(\boldsymbol{\delta})\mathbf{U}^\top$. The following problem*

$$\min_{\mathbf{S}} \|\mathbf{S} - \mathbf{A}\|_F^2 \ \textit{s.t.}\ \mathtt{Tr}(\mathbf{S}) = c, \mathbf{S}^\top = \mathbf{S}, \mathbf{0} \preceq \mathbf{S} \preceq \mathbf{1} \tag{15}$$

has optimal solution given by $\mathbf{S}^* = \mathbf{U}\mathtt{Diag}(\boldsymbol{\rho}^*)\mathbf{U}^\top$, where $\boldsymbol{\rho}^*$ is the solution to

$$\min_{\boldsymbol{\rho}} \|\boldsymbol{\rho} - \boldsymbol{\delta}\|_2^2, \ \textit{s.t.}\ 0 \le \boldsymbol{\rho} \le 1, \boldsymbol{\rho}^\top 1 = c. \tag{16}$$

Proof. For two symmetric matrices $\mathbf{S} \in \mathbb{R}^{n\times n}$ and $\mathbf{A} \in \mathbb{R}^{n\times n}$, and let $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_n$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ be the ordered eigenvalues of $\mathbf{S}$ and $\mathbf{A}$, respectively. Due to the fact that $\mathtt{Tr}(\mathbf{S}^\top \mathbf{A}) \leq \sum_{i=1}^{n} \rho_i \sigma_i$ shown in [13], we obtain

$$\begin{aligned} \|\mathbf{S}-\mathbf{A}\|_F^2 &= \mathtt{Tr}\left(\mathbf{S}^\top \mathbf{S}\right) + \mathtt{Tr}\left(\mathbf{A}^\top \mathbf{A}\right) - 2\mathtt{Tr}\left(\mathbf{S}^\top \mathbf{A}\right) \\ &= \sum_{i=1}^{n} \rho_i^2 + \sum_{i=1}^{n} \delta_i^2 - 2\mathtt{Tr}\left(\mathbf{S}^\top \mathbf{A}\right) \\ &\geq \sum_{i=1}^{n} \left(\rho_i^2 + \delta_i^2 - 2\rho_i\delta_i\right) \\ &= \|\rho - \delta\|_2^2 \end{aligned} \tag{17}$$

Note here that the above equality holds when $\mathbf{S}$ admits the spectral decomposition $\mathbf{A} = \mathbf{U}\mathtt{Diag}(\sigma)\mathbf{U}^\top$. Additionally, the constraints $\mathbf{0} \preceq \mathbf{S} \preceq \mathbf{1}, \mathtt{Tr}(\mathbf{S}) = c$ are equicalent to $0 \leq \rho \leq 1, \rho^\top 1 = c$, respectively. Thus, $\mathbf{S}^* = \mathbf{U}\mathtt{Diag}(\sigma)\mathbf{U}^\top$ is optimal to problem (16) with σ^* being optimal to problem (17). After that, we can obtain the final solution, $\mathbf{S}^* = (\mathbf{S}^* + (\mathbf{S}^*)^\top)/2$, to satisfy the constraint $\mathbf{A}^\top = \mathbf{A}$. The proof is completed. ■

Finally, an efficient iterative algorithm in [14] can be employed to solve (17).

► **Step 6. $\mathcal{G}$-subproblem:** The subproblem of $\mathcal{G}$ can be transformed as follows

$$\min_{\mathcal{G}} \gamma\|\mathcal{G}\|_{\circledast} + \frac{\mu}{2}\|\mathcal{G} - \mathcal{S} + \mathcal{Y}_3/\mu\|_F^2 \tag{18}$$

Let $\mathcal{M} = \mathcal{S} - \mathcal{Y}_3/\mu$ and according to the following Theorem 2, we can apply the tensor tubal-shrinkage of $\mathcal{M}$ to solve the problem (18).

Theorem 2 [28] *For a scalar $\tau > 0$ and two three-order tensors $\mathcal{T} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $\mathcal{M} \in \mathbb{R}^{n_1\times n_2\times n_3}$, the global optimal solution of the following problem*

$$\min_{\mathcal{T}} \tau\|\mathcal{T}\|_{\circledast} + \frac{1}{2}\|\mathcal{T} - \mathcal{M}\|_F^2 \tag{19}$$

where $\tau = \frac{\gamma}{\mu}$. And (19) can be computed by tensor tubal-shrinkage operator as follows

$$\mathcal{T} = \mathcal{C}_{n_3\tau}(\mathcal{M}) = \mathcal{U} * \mathcal{C}_{n_3\tau}(\mathcal{G}) * \mathcal{V}^\top, \tag{20}$$

where $\mathcal{M} = \mathcal{U} * \mathcal{G} * \mathcal{V}^\top$ and $\mathcal{C}_{n_3\tau} = \mathcal{G} * \mathcal{Q}$. $\mathcal{Q} \in \mathbb{R}^{n_1\times n_2\times n_3}$ denotes a f-diagonal tensor and each diagonal element of $\mathcal{Q}$ is defined as $\mathcal{Q}_f(i,i,j) = \left(1 - \frac{n_3\tau}{\mathcal{G}(i,i,j)}\right)_+$.

► **Step 7. ADMM variables-subproblem:** We update variables of ADMM via

$$\begin{aligned} \mathbf{Y}_1^v &= \mathbf{Y}_1^v + \mu(\mathbf{X}^v + \mathbf{B}^v(\mathbf{W}^v)^\top - \mathbf{D}^v - \mathbf{E}^v) \\ \mathbf{Y}_2^v &= \mathbf{Y}_2^v + \mu(\mathbf{Z}^v - \mathbf{D}^v) \\ \mathcal{Y}_3 &= \mathcal{Y}_3 + \mu(\mathcal{G} - \mathcal{S}) \\ \mu &= \min(\eta\,\mu, \mu_{\max}) \end{aligned} \tag{21}$$

where both η and $\mu_{\max}$ are the scalars of ADMM. The pseudo-code is depicted as in Algorithm 1, whose convergence condition is $\max\{|\mathrm{obj}^{t+1} - \mathrm{obj}^t|, \|\mathcal{S}^{t+1} - \mathcal{S}^t\|_F^2\} \leq$

Algorithm 1 Algorithm to the proposed IMGC method.

Require: Multiple incomplete data $\{\mathbf{X}^v\}_{v=1}^r$, α, β and γ, $\mu = 0.01$, $\mu_0 = 10^{10}$, $\eta = 2$.
Ensure: Graphs $\mathbf{S}^v$, inferring data $\mathbf{B}^v$, recovering data $\mathbf{D}^v$.
1: Initialize $\{\mathbf{D}\}_{v=1}^r$, $\{\mathbf{B}\}_{v=1}^r$, $\{\mathbf{E}\}_{v=1}^r$, $\{\mathbf{Z}\}_{v=1}^r$, $\{\mathbf{G}\}_{v=1}^r$, $\{\mathbf{Y}_1\}_{v=1}^r$, $\{\mathbf{Y}_2\}_{v=1}^r$, $\{\mathbf{Y}_3\}_{v=1}^r$ to be zeros. $\{\mathbf{S}\}_{v=1}^r$ are initialized by constructing k-nearest neighbor graphs from $\{\mathbf{X}^v\}_{v=1}^r$.
2: **repeat**
3: Update the $\mathbf{D}$ via (10);
4: Update the $\mathbf{Z}$ via (11);
5: Update the $\mathbf{E}$ via (12);
6: Update the $\mathbf{B}$ via (13);
7: Update candidate graphs $\{\mathbf{S}^v\}_{v=1}^r$ via (14);
8: Construct $\boldsymbol{\mathcal{S}}$ via `bvfold` and `rotate` on $\{\mathbf{S}^v\}_{v=1}^r$;
9: Update the tensor $\boldsymbol{\mathcal{A}}$ via (19);
10: Update the ADMM variables via (21);
11: **until** $\max\{|\mathrm{obj}^{t+1} - \mathrm{obj}^t|, \|\boldsymbol{\mathcal{S}}^{t+1} - \boldsymbol{\mathcal{S}}^t\|_F^2\} \leq \epsilon$;
12: **Output** $\widehat{\mathbf{S}} = (\sum_{v=1}^r \mathbf{S}^v)/r$.

Table 1. Benchmark datasets.

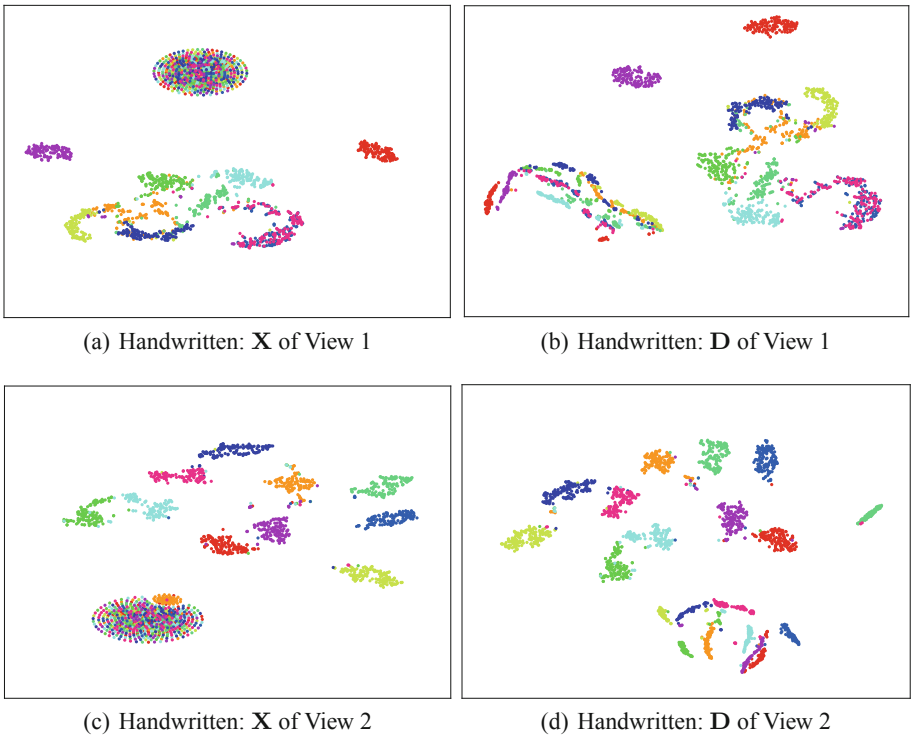
Database	Class	View	Objective	Samples
BBCSport	5	4	Documents	116
Handwritten	10	2	Digit images	2000
Caltech7	7	6	Generic object	1474
MNIST	10	3	Digit images	10000

ϵ. Here, obj, t and $\epsilon = 10^{-5}$ are the objection value of function, iteration number and threshold value, respectively.

Computational Complexity: The computational complexity of (9) involves the following main subproblems, including (10), (11), (12), (13), (15), and (19). And the major computational cost involves (10), (11), (15), and (19). (10) needs to compute matrix SVD with the complexity $\mathcal{O}(rn^2)$ by leveraging an approximation rank technique like PROPACK package [6]. (11) requires the complexity of $\mathcal{O}(rn^3)$ due to the matrix inversion. (15) involves the skinny SVD operation to compute the c largest eigenvalue-eigenvector pair of graphs with $\mathcal{O}(rn^2)$. (19) involves `fft` and `ifft` operators on the $n \times r \times n$ tensor, the complexity is $\mathcal{O}(rn^2 \log(n))$, then it also requires to compute the SVD of each frontal slice with size $n \times r$ in the Fourier domain, thus the complexity is $\mathcal{O}(rn^2 \log(n)) + \mathcal{O}(r^2 n^2)$ in total. Here, we ignore (12), (13) and the variables of ADMM since they are basic matrix operations. The computational complexity of Algorithm 1 can be summarized as $\mathcal{O}(n^3)$ since t and r are far less than n.

4 Experiment

Evaluation and Datasets: As shown in Table 2, 4 popular datasets from various applications, cluster-numbers, view-numbers and sample-numbers are employed to evaluate



(a) Handwritten: $\mathbf{X}$ of View 1 (b) Handwritten: $\mathbf{D}$ of View 1

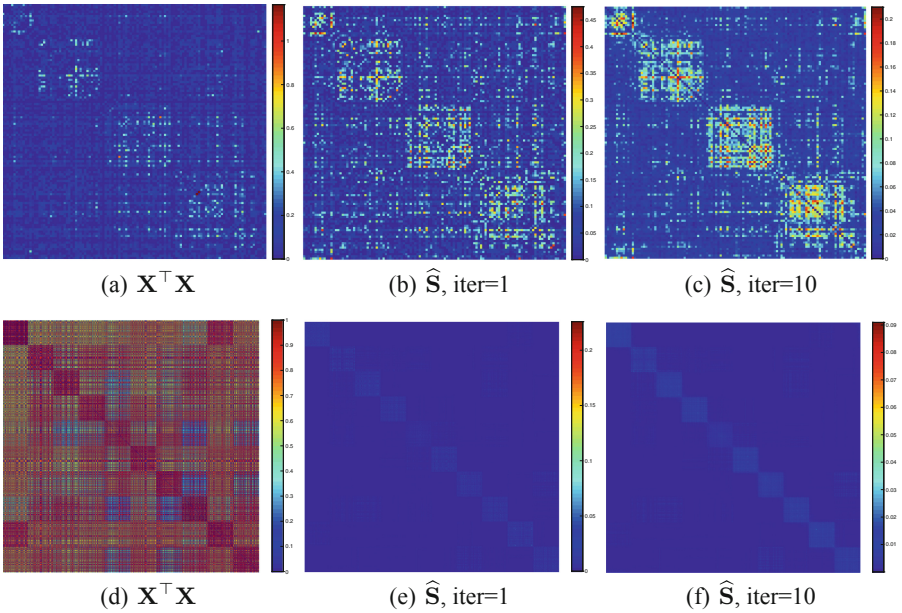
(c) Handwritten: $\mathbf{X}$ of View 2 (d) Handwritten: $\mathbf{D}$ of View 2

Fig. 2. The visualization via t-SNE [12] for original data and inferring complete data with missing rate of 30% on Handwritten dataset.

the compared methods with three criterions: accuracy (ACC), normalized mutual information (NMI), and Purity. Here, the bigger values of criterions indicate a better clustering performance. (1) **Handwritten** contains 2000 samples and averagely distributes in 10 classes, we merely choose pixel average features and Fourier coefficient features to construct two views. (2) **Caltech7** is a prevalent 7-object dataset with 1474 samples, which is consisted of six features corresponding to 6 views, including LBP, GIST, Gabor, wavelet moments, CENTRIST and HOG. (3) **BBCSport** has 116 samples from 5 classes, whose 4 views contains different feature dimensions with 2158, 2113, 2063 and 1991 in this paper. (4) **MNIST** consists of 70000 samples of 28-by-28 pixel size, where we randomly select the 10000 samples, and then following [29] to construct 3 different views.

Compared Methods: We compare our approach with two base line methods, including best single view (BSV) and Concat, as well as several state-of-the-art IMGC methods, including PVC [7], GPMVC [15], IMG [26], MIC [18], OMVC [17], DAIMC [3], UEAF [22], and GIMC-FLSD [23], APGLF [27].

Incomplete Data Construction: Following the incomplete data construction from [22], 10%, 30%, and 50% samples of all evaluated datasets (except MNIST) are randomly



(a) $\mathbf{X}^\top\mathbf{X}$ (b) $\widehat{\mathbf{S}}$, iter=1 (c) $\widehat{\mathbf{S}}$, iter=10

(d) $\mathbf{X}^\top\mathbf{X}$ (e) $\widehat{\mathbf{S}}$, iter=1 (f) $\widehat{\mathbf{S}}$, iter=10

Fig. 3. The graph recovering with missing rate of 50%.

moved for each view to construct the incomplete multi-view datasets with corresponding missing rates, where each sample owns at least one view. Further, 10%, 30%, 50% and 70% samples are randomly selected as the paired samples for MNIST dataset. Then, we remove half of the remaining samples for one view, and follow the previous strategy to remove the next view until the last view.

Inferring Complete Data and Affinity Graph Recovery: Before performing Algorithm 1, the dense clusters of ellipse in Fig. 2 (a) and (c) are caused by the incomplete information of original data $\mathbf{X}$, where incomplete part is filled with 0. After performing our Algorithm 1, the dense clusters of ellipse in Fig. 2 (b) and (d) have a good discriminant, where incomplete information are inferred and filled with the valid information value. This intuitively proves the validity of inferring complete data $\mathbf{D}$.

As mentioned in Fig. 3, we have shown the affinity graph of our method on BBCSport and Handwritten datasets. It is easy to find that: 1) compared to original affinity graph (here, we adopt inner product to produce graph $\mathbf{X}^\top\mathbf{X}$.), the edges of graphs corresponding to the missing part can be gradually completed well with the increasing of iterations; 2) even in missing rate with 50%, both BBCSport and Handwritten have clear block diagonal structure, *i.e.*, blocks of graphs are equivalent to the class of original data, this plays a critical role for clustering.

Experimental Results: Following [23], we repeatedly perform 5 times independent experiments of all the comparison methods, and then the average clustering performance and standard deviation are presented in Table (2). From the experimental results of Table (2), we can obtain the following observations: 1) our RPCA-IGTL

Table 2. Clustering results (*i.e.*, ACC, NMI and Purity) of the comparison methods in terms of five varying scale datasets. The best results are highlighted in bold. For compared methods, most of experimental results are copied from [23].

Dataset	Metrics	MR	BSV	Concat	GPMVC	MIC	DAIMC	OMVC	UEAF	GIMC-FLSD	APGLF	our
BBCSport	ACC	10%	67.35±3.86	69.32 ± 4.56	49.48 ± 6.46	50.51 ± 2.10	62.53 ± 8.16	51.45 ± 5.67	77.13±5.21	79.02±4.47	**91.38±4.21**	88.90±3.26
		30%	54.17±3.91	60.21 ± 3.85	42.89 ± 4.66	47.05 ± 3.96	60.66 ± 9.71	44.28 ± 5.01	85.21±4.61	76.29±1.91	81.03±1.86	**84.48±1.26**
		50%	49.37±2.96	50.82 ± 3.02	40.13 ± 4.77	45.52 ± 1.87	54.51 ± 9.18	49.57 ± 4.37	68.27±4.01	69.71±4.20	75.00±4.16	**78.45±3.81**
	NMI	10%	65.35 ± 3.09	62.86 ± 5.61	28.14 ± 6.93	30.47 ± 2.91	49.24 ± 8.20	39.37 ± 6.46	69.52±3.89	71.81±3.60	79.37±1.66	**81.09±1.83**
		30%	50.39 ± 4.68	40.31 ± 5.28	18.94 ± 4.84	26.30 ± 4.57	46.99 ± 9.23	40.32 ± 5.12	67.21±3.21	64.64±2.54	71.53±2.11	**80.31±3.68**
		50%	36.81 ± 3.95	27.13 ± 5.21	17.82 ± 5.18	24.54 ± 3.19	36.54 ± 9.40	42.65 ± 4.91	54.81±3.08	52.28±5.91	55.27±5.21	**73.07±2.96**
	Purity	10%	72.57 ± 4.98	79.12 ± 4.26	56.83 ± 6.16	56.19 ± 2.13	71.49 ± 7.03	52.56 ± 5.78	87.21±4.92	88.45±2.70	91.38±3.26	**93.97±3.61**
		30%	55.86 ± 4.28	68.51 ± 4.36	47.84 ± 4.16	51.22 ± 4.38	69.71 ± 9.91	54.52 ± 5.24	86.38± 4.21	85.34±1.68	87.07 ±2.52	**93.10±2.63**
		50%	42.21 ± 4.34	50.08 ± 4.37	45.60 ± 4.21	50.41 ± 2.91	62.90 ±4.88	56.47 ± 5.16	76.89±3.76	77.67±5.16	79.31±4.88	**91.38±3.21**
Handwritten	ACC	10%	71.32 ± 4.35	82.12 ± 1.89	79.61 ± 3.96	70.62 ± 3.20	78.23 ± 3.49	73.90 ± 2.63	80.18±2.56	87.27 ± 1.98	88.10 ± 1.72	**99.45 ±2.36**
		30%	55.31 ± 4.21	78.31 ± 1.27	77.15 ± 3.28	63.07 ± 2.59	75.15 ± 2.01	68.19 ± 3.78	77.12 ±2.38	84.29 ± 1.62	87.45±2.11	**98.35 ± 2.34**
		50%	41.43 ± 5.89	66.91 ± 1.14	71.28 ± 2.49	55.17 ± 3.29	64.28 ± 3.92	59.12 ± 3.50	71.26±2.89	77.27 ± 2.28	81.55±2.69	**98.15 ± 1.67**
	NMI	10%	58.85 ± 1.48	78.20 ± 2.90	72.19 ± 2.16	63.27 ± 2.11	64.16 ± 2.17	61.16 ± 3.01	70.13 ±1.58	79.16 ± 1.06	89.04±1.21	**98.53 ± 2.37**
		30%	52.03 ± 1.20	65.34 ± 2.13	69.16 ± 2.31	55.21 ± 1.98	61.16 ± 2.37	58.22 ± 3.38	65.23±1.21	75.27 ± 1.95	85.22±1.31	**96.02 ± 1.08**
		50%	45.52 ± 1.16	56.31 ± 2.66	67.26 ± 3.03	50.15 ± 2.16	52.15 ± 2.11	51.15 ± 2.16	61.31±1.27	69.83 ± 1.21	82.85±1.29	**95.63 ± 1.27**
	Purity	10%	66.15 ± 1.98	83.01 ± 1.56	81.15 ± 2.84	73.11 ± 2.46	78.26 ± 2.17	73.16 ± 2.27	83.72±2.67	87.93 ± 1.62	88.10±1.74	**99.45 ± 0.91**
		30%	58.86 ± 1.35	70.23 ± 2.09	78.22 ± 2.38	64.27 ± 2.27	74.27 ± 2.53	67.27 ± 2.83	80.12± 2.98	83.16 ± 1.15	85.50±1.13	**98.35 ± 1.21**
		50%	52.29 ± 1.69	61.35 ± 2.95	72.27 ± 3.93	58.72 ± 2.27	65.82 ± 3.82	60.27 ± 3.28	73.21±3.52	76.87 ± 2.27	82.85±76.33	**98.15 ± 1.09**
Catech7	ACC	10%	61.32 ± 2.78	45.48 ± 2.67	43.34 ± 3.30	41.77 ± 3.64	42.26 ± 4.03	38.89 ± 2.64	51.63±4.23	48.20±1.55	66.82±1.35	**77.82±1.73**
		30%	45.23 ± 3.15	44.31 ± 1.54	40.81 ± 4.90	40.47 ± 3.36	41.16 ± 3.49	37.77 ± 3.91	43.62±1.02	47.14±1.12	48.14±2.65	**50.14±2.83**
		50%	44.24 ± 1.87	40.61 ± 3.01	34.25 ± 3.93	38.88 ± 4.95	38.35 ± 2.78	36.50 ± 2.79	37.11±3.16	44.08±2.41	46.31±1.88	**49.32±1.94**
	NMI	10%	45.83 ± 1.29	45.71 ± 1.04	29.08 ± 3.00	35.52 ± 1.72	42.71 ± 2.65	27.74 ± 1.57	40.13±2.36	44.84±0.85	50.34±2.06	**60.55±2.13**
		30%	36.25 ± 3.12	41.74 ± 2.45	21.24 ± 7.05	31.10 ± 1.95	40.29 ± 2.44	23.76 ± 3.47	32.21±2.08	42.32±1.05	45.16±3.84	**48.04±2.95**
		50%	28.58 ± 1.56	35.66 ± 2.31	13.41 ± 3.31	26.67 ± 3.29	36.22 ± 2.27	19.87 ± 3.32	25.24±1.51	36.53±1.91	42.65±1.37	**44.82±1.25**
	Purity	10%	84.17 ± 0.94	86.28 ± 0.39	78.41 ± 1.86	80.86 ± 1.12	84.63 ± 1.37	78.95 ± 0.88	82.31±1.87	86.62±0.34	86.58±0.87	**87.11±0.64**
		30%	76.28 ± 1.36	84.28 ± 0.91	74.31 ± 6.28	78.24 ± 1.62	83.79 ± 1.25	76.90 ± 2.19	79.88±2.56	85.63±0.65	85.96±0.97	**86.16±0.87**
		50%	72.19 ± 0.77	79.68 ± 0.84	68.71 ± 3.75	74.99 ± 3.58	82.60 ± 0.96	75.23 ± 2.46	77.61±2.30	83.33±0.96	84.27±1.86	**85.96±1.97**

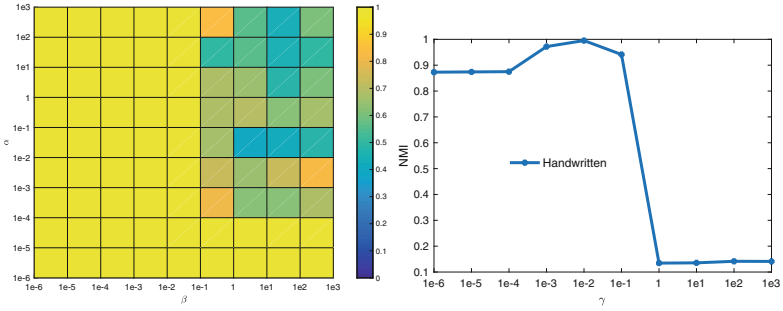
overwhelmingly surpasses the compared methods, especially in Handwritten and BBC-Sports datasets, our method has a larger improvement with the increasing of missing rate, the reason may be that our method can well capture the graph spatial information of graph semantic space; and 2) the clustering performance decreases on all datasets for each method with the increasing of missing rate, and the great decline on BBCSport, and Handwritten datasets for BSV and Concat methods can demonstrate the validity of exploring inter-view information; 3) compared to UEAF and GIMC-FLSD, *w.r.t.* 50% high missing rate, our method improves 34.32% and 25.80% in terms of Purity, respectively. Compared to recent proposed APGLF [27], our method has also achieved the satisfied clustering performance. This profits from the learned inferring complete data and high-quality affinity graphs.

Note here that the datasets of Table (2) are the widely used to deal with the IMVC methods. To better evaluate the performance of the proposed method, we also employ the large scale MNIST to perform incomplete clustering and report the results in Table (3). As can be seen, our method still has a superior clustering performance in large scale dataset.

Parameter Sensitivity Analysis: Our RPCA-IGTL involves three parameters α, β, and γ required to be set properly, which can control the noise term $\mathbf{E}^v$, manifold regularized term $\mathtt{Tr}(\mathbf{D}^v\mathbf{L}_S^v)\mathbf{D}^{v\top}$, and the effect of spatial graph tensor $\boldsymbol{\mathcal{S}}$, respectively. As shown in Fig.(4), it is insensitive for our parameters to work well for a wide range of α, β. Although it seems to be a little sensitive to γ, this indicates effectiveness of high-order spatial structure information. In fact, we always find satisfied clustering performance in a large range (*i.e.*, $\alpha \in [10^3, 10^{-4}]$, $\beta \in [10^{-5}, 10^{-1}]$, and $\gamma \in [10^{-3}, 10^{-1}]$).

Table 3. Experimental results of UEAF, GIMC-FLSD, APGLF and our RPCA-IGTL on the MNIST dataset with 10,000 samples.

Methods	Metrics	Missing rate	UEAF	GIMC-FLSD	APGLF	Proposed
MNIST	ACC	10%	80.28	86.18	89.31	**97.19**
		30%	77.39	83.06	87.69	**95.89**
		50%	69.28	71.68	84.46	**93.61**
		70%	64.72	68.82	80.29	**92.20**
	NMI	10%	69.86	74.25	88.45	**94.64**
		30%	67.71	70.41	86.12	**93.27**
		50%	58.36	61.32	73.96	**91.05**
		70%	49.16	56.26	70.45	**84.31**
	Purity	10%	80.28	86.18	88.46	**97.19**
		30%	78.31	83.06	85.46	**95.89**
		50%	70.03	72.21	74.16	**94.21**
		70%	66.26	69.18	71.79	**92.93**



(a) $\gamma = 1e-2$ and tune α, β. (b) $\alpha = \beta = 1e-2$ and tune γ.

Fig. 4. The NMI in terms of α, β and γ on the Handwritten dataset with a missing rate of 30%.

Convergence Analysis: To demonstrate the convergence of Algorithm (1), we experimentally record the objective values and NMI in terms of BBCSport and Handwritten datasets with missing rate 10%, 30%, 50% at each iteration. As illustrated in Fig. 5 , we find that: 1) the residual curves for different missing rate converge rapidly and consistently till to the stable point, where the objective values of each iteration are calculated via $\max\{|\text{obj}^{t+1} - \text{obj}^{t}|, \|\mathcal{S}^{t+1} - \mathcal{S}^{t}\|_F^2\} \leq \epsilon$; and 2) NMI *w.r.t.* iterator of RPCA-IGTL consistently and gradually increases until objective values become stable. This can prove the fast and stable convergence property of Algorithm (1).

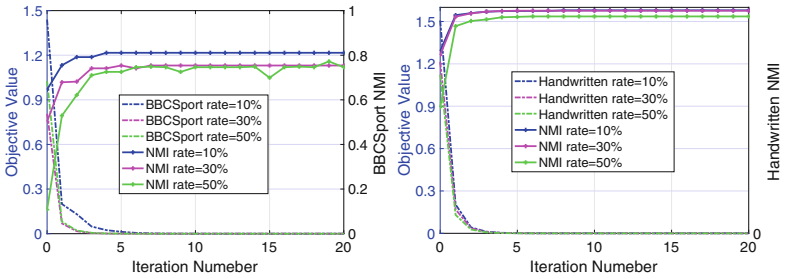


Fig. 5. The convergence curves and NMI of the proposed RPCA-IGTL method on BBCSport dataset in terms of 10%, 30%, and 50% missing rate.

5 Conclusion

Although the existing ANGL-based IMGC methods have the power of handling incomplete multi-view data, while they cannot take full use of the structure information hidden in the incomplete view data. To address these problems, we propose a novel IMGC method, *i.e.*, RPCA-IGTL. It designs an elegant RPCA-induced manifold learning framework and jointly introduces the graph tensor low-rank constrain and enhanced block diagonal constraint. By leverage a mutually reinforcing way, it can fully exploit the manifold structure information of inferring complete data in feature space and the inter-view graph spatial structure information of graph tensor in graph semantic space. Experimental results on several datasets with various scale and missing rate have demonstrated the superiority for inferred complete data and clustering performance.

References

1. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM (JACM) **58**(3), 1–37 (2011)
2. Hu, M., Chen, S.: Incomplete multi-view clustering. Springer International Publishing (2016)
3. Hu, M., Chen, S.: Doubly aligned incomplete multi-view clustering, pp. 2262–2268
4. Huang, S., Tsang, I., Xu, Z., Lv, J.C.: Measuring diversity in graph learning: a unified framework for structured multi-view clustering. In: IEEE TKDE, pp. 1–1 (2021)
5. Wen, J., Zhang, Z., Xu, Y., Zhong, Z.: Incomplete multi-view clustering via graph regularized matrix factorization. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018. LNCS, vol. 11132, pp. 593–608. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11018-5_47
6. Larsen, R.M.: Propack-software for large and sparse svd calculations 2008-2009 (2004). http://sun.stanford.edu/rmunk/PROPACKpp
7. Li, S.Y., Jiang, Y., Zhou, Z.H.: Partial multi-view clustering. In: AAAI, vol. 28 (2014)
8. Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 171–184 (2012)
9. Liu, X., et al.: Efficient and effective regularized incomplete multi-view clustering. IEEE Trans. Pattern Anal. Mach. Intell. **43**, 2634–2646 (2020)
10. Liu, X., et al.: Multiple kernel k k-means with incomplete kernels. IEEE Trans. Pattern Anal. Mach. Intell. **42**(5), 1191–1204 (2019)

11. Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., Yan, S.: Tensor robust principal component analysis with a new tensor nuclear norm. IEEE Trans. Pattern Anal. Mach. Intell. **42**(4), 925–938 (2020)
12. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. J. Mach. Learn. Res. **9**(86), 2579–2605 (2008)
13. Marshall, A.W., Olkin, I., Arnold, B.C.: Inequalities: Theory of Majorization and Its Applications. SSS, Springer, New York (2011). https://doi.org/10.1007/978-0-387-68276-1
14. Nie, F., Wang, X., Jordan, M.I., Huang, H.: The constrained laplacian rank algorithm for graph-based clustering. In: Thirtieth AAAI Conference on Artificial Intelligence, pp. 1969–1976 (2016)
15. Rai, N., Negi, S., Chaudhury, S., Deshmukh, O.: Partial multi-view clustering using graph regularized nmf. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 2192–2197. IEEE (2016)
16. Rai, P., Trivedi, A., Daumé III, H., DuVall, S.L.: Multiview clustering with incomplete views. In: NIPS Workshop, pp. 1–4 (2010)
17. Shao, W., He, L., Lu, C.t., Philip, S.Y.: Online multi-view clustering with incomplete views. In: 2016 IEEE International Conference on Big Data (Big Data), pp. 1012–1017 (2016)
18. Shao, W., He, L., Yu, P.S.: Multiple incomplete views clustering via weighted nonnegative matrix factorization with $L_{2,1}$ regularization. In: Appice, A., Rodrigues, P.P., Santos Costa, V., Soares, C., Gama, J., Jorge, A. (eds.) ECML PKDD 2015. LNCS (LNAI), vol. 9284, pp. 318–334. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23528-8_20
19. Shi, S., Nie, F., Wang, R., Li, X.: Fast multi-view clustering via prototype graph. IEEE Transactions on Knowledge and Data Engineering, pp. 1–1 (2021)
20. Wen, J., Yan, K., Zhang, Z., Xu, Y., Zhang, B.: Adaptive graph completion based incomplete multi-view clustering. IEEE Trans. Multimedia **23**, 2493002504 (2020)
21. Wen, J., Xu, Y., Liu, H.: Incomplete multiview spectral clustering with adaptive graph learning. IEEE Trans. Cybern. **50**(4), 1418–1429 (2018)
22. Wen, J., Zhang, Z., Xu, Y., Zhang, B., Fei, L., Liu, H.: Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 5393–5400 (2019)
23. Wen, J., Zhang, Z., Zhang, Z., Fei, L., Wang, M.: Generalized incomplete multiview clustering with flexible locality structure diffusion. IEEE Trans. Cybern. **51**(1), 101–114 (2020)
24. Zhang, C., Fu, H., Hu, Q., Cao, X., Xie, Y., Tao, D., Xu, D.: Generalized latent multi-view subspace clustering. IEEE Trans. Pattern Anal. Mach. Intell. **42**(1), 86–99 (2018)
25. Zhang, P., et al.: Adaptive weighted graph fusion incomplete multi-view subspace clustering. Sensors **20**(20), 5755 (2020)
26. Zhao, H., Liu, H., Fu, Y.: Incomplete multi-modal visual data grouping. In: IJCAI, pp. 2392–2398 (2016)
27. Zheng, X., Liu, X., Chen, J., Zhu, E.: Adaptive partial graph learning and fusion for incomplete multi-view clustering. Int. J. Intell. Syst. **37**(1), 991–1009 (2022)
28. Zhou, P., Lu, C., Feng, J., Lin, Z., Yan, S.: Tensor low-rank representation for data recovery and clustering. IEEE Trans. Pattern Anal. Mach. Intell. **43**(5), 1718–1732 (2019)
29. Zhou, S., et al.: Multi-view spectral clustering with optimal neighborhood laplacian matrix. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 6965–6972 (2020)