IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 August 2023, accepted 20 September 2023, date of publication 26 September 2023, date of current version 5 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3319564

RESEARCH ARTICLE

Anchor Pseudo-Supervise Large-Scale Incomplete Multi-View Clustering

SONGBAI ZHU[1,2], JIAN DAI[2], GUOLAI YANG[1], AND ZHENWEN REN[3], (Member, IEEE)

[1]School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[2]Southwest Automation Research Institute, China South Industries Group Corporation, Mianyang 621000, China
[3]School of National Defence Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China

Corresponding authors: Guolai Yang (sqspaper2@163.com) and Zhenwen Ren (rzw@njust.edu.cn)

This work was supported in part by the Project of Guangxi Key Laboratory of Machine Vision and Intelligent Control under Grant 2022B07, in part by the Miaozi Project of Sichuan Province under Grant MZGC20230072, and in part by the Natural Science Foundation of Sichuan Province under Grant 2023NSFSC1373.

ABSTRACT In real life, only partial information of samples is available everywhere, this makes Incomplete multi-view clustering (IMVC) becomes a significant research topic to handle data loss situations. Recently, several methods leverage the anchor strategy by selecting fixed anchors to handle the challenging large-scale IMVC. However, all of them ignore the guidance of prior information hidden in the bipartite graph. Therefore, we propose a novel Anchor Pseudo-supervise Large-scale Incomplete Multi-view Clustering (AP-LIMC) method by introducing a prior indicator matrix as a pseudo-supervise anchor learning paradigm. Specifically, the prior indicator matrix is first introduced to control the distribution of anchors in each cluster. Then, an anchor pseudo-supervise learning framework is designed to generate high-quality anchors and a unified bipartite graph with prior indicator supervision. In addition, we design an optimized process with linear computational and extensive experiments on multiple public datasets with recent advances to validate the effectiveness, superiority, and efficiency. For example, on the Stl10 dataset, the performance of the proposed AP-LIMC improved by 23.95%,15.71%,27.39%, and 18.24% in terms of four evaluation metrics, respectively.

INDEX TERMS Incomplete multi-view clustering, anchor learning, tensor, bipartite graph.

I. INTRODUCTION

In recent years, multi-view clustering (MVC) has garnered considerable scholarly and practical attention, driven by the increasing presentation of information in diverse applications [1], [2], [3]. MVC aims to utilize heterogeneous and homogeneous information to partition unlabeled multi-view data into different clusters. While many MVC methods [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17] have been developed to capture the paired similarity between samples and views, these approaches assume that all views are complete. However, the practical applications often present a contrasting scenario, wherein certain view information may be missing, thereby giving rise to instances of incomplete multi-view data. For example, in social network analysis [18], [19], individuals may only be registered on certain social networks, leading to incomplete sample information across different networks. This scenario is known as incomplete multi-view clustering (IMVC), and it poses a significant challenge due to the loss of view information and destruction of paired similarity caused by missing data [20].

In the past few years, researchers have developed effective solutions to address missing data, which can be broadly categorized into three groups. The first group [21], [22], [23], [24], [25] involves using matrix factorization to generate a consensus representation that combines information from multiple incomplete views. This allows data from different views to be aggregated into a single representation, making it possible to describe the information from multiple incomplete views. The resulting consensus representation can then be used as input for clustering algorithms such as k-means to obtain clustering results.

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

For the second group, instead of sharing a consensus representation, kernel-based IMVC methods [26], [27], [28], [29], [30] aim to learn a consensus kernel or consensus partition from incomplete views. This is achieved by uncovering the nonlinear information present in incomplete views. In the third group, some researchers use different graph learning techniques to generate a consensus graph across multiple incomplete views [28], [31], [32], [33], rather than learning a consensus representation or consensus kernel. For instance, in studies such as [33] and [34], self-representation subspace learning or adaptive neighborhood graph learning is used to generate a shared similarity graph for spectral clustering. These methods fuse the observed data of multiple incomplete views into a consensus graph that reflects the intact graph structure of all incomplete views. In this way, they are able to capture graph structure information among incomplete views. Overall, the common idea behind these IMVC methods is to learn a consensus representation, consensus kernel, or consensus graph by exploiting the complementary information present in incomplete views.

Methods that integrate diverse perspectives of information exhibit the capability to address the challenges of IMVC. Nevertheless, their computational intricacy and memory requisites impose limitations on their applicability to substantial tasks, particularly within the domain of graph-based learning methodologies. To address this issue, some methods attempt to use bipartite graphs to achieve highly efficient clustering [31], [35]. For example, one method [35] selects samples with complete views as anchor points, which can connect all instances of each view to construct a bipartite graph for subsequent clustering tasks. However, this idea only works when each cluster contains enough samples with complete views, which is impractical. Other methods use k-means or sampling strategies to choose anchors from observed data, but the fixed selection of anchors from original incomplete data limits the flexibility and quality of anchors.

To address the aforementioned limitations, we propose a novel **Anchor Pseudo-supervise Large-scale Incomplete Multi-view Clustering (AP-LIMC)** for high-quality anchor learning and high-efficient IMVC. Specifically, AP-LIMC first proposes a pseudo-supervised anchor learning framework as mentioned in Fig. 1, which aims to learn multiple view-independent anchors according to the proposed prior indicator matrix. Simultaneously, the prior matrix has the capacity to regulate the distribution of anchors within each cluster, thereby providing guidance for the construction of a unified bipartite graph. In this way, the superior view-independent anchors and a consensus bipartite graph can be obtained simultaneously in a mutually reinforcing manner, such that the complementary among view-independent anchors and underlying consistency structure of among views can be together explored. Afterward, the optimal bipartite graph can be learned to perform k-means for obtaining results. In summary, AP-LIMC enjoys three-fold contributions:

- AP-LIMC proposes a new anchor learning strategy to generate the ideal anchors with priori indicator supervision.
- With the help of dexterously designing anchor learning strategy, AP-LIMC further proposes a novel and effective framework for large-scale IMVC.
- An efficient and effective solver with linear computational and memory complexity is designed to perform extensive experiments on scalable datasets to demonstrate the superiority of AP-LIMC.

TABLE 1. Detailed information of notations.

Notations	Definitions
$\mathbf{X}_p \in \mathbb{R}^{d_p \times n}$	The incomplete data matrix
$\mathbf{Y}_p \in \mathbb{R}^{d_p \times n}$	The complete data matrix
$\mathbf{H}_p \in \mathbb{R}^{n \times n}$	Affinity graph
$\mathbf{S}_p \in \mathbb{R}^{m \times n}$	The p-th bipartite graph
$\mathbf{S} \in \mathbb{R}^{n \times n}$	The consensus bipartite graph
$\mathbf{E}_p \in \mathbb{R}^{n \times n_p}$	Missing index for p-th view
n	The number of samples
d_p	The dimension of samples in p-th view
k	The number of clusters
m	The number of anchors
$\boldsymbol{\theta}$	The weighted vector

II. RELATED WORK

In this section, we provide an overview of the notations used in our proposed method, as well as the related research in the field. This includes a discussion of typical graph-based IMVC methods and bipartite graph clustering. Additionally, Table 1 presents the relevant definitions used in this paper.

A. GRAPH BASED IMVC

IMVC has gained attention due to the incomplete nature of multimodal data in real-world applications. For example, individuals may not be registered on all social networking platforms, leading to incomplete information on some platforms. As a result, clustering with multi-view partial data is a challenging and valuable issue. Graph learning offers powerful representation ability and can capture relevant information between data, making it an effective method for performing IMVC.Given incomplete multi-view data $\{\mathbf{X}_p \in \mathbb{R}^{d_p \times n}\}_{p=1}^{V}$, graph-based IMVC framework can be written as

$$\begin{aligned} &\min_{\mathbf{H}_p} \sum_{p=1}^{V} \boldsymbol{\Psi}(\mathbf{X}_p, \mathbf{H}_p) + \beta \boldsymbol{\Phi}\left(\mathbf{H}_p\right), \\ &\text{s.t. } \mathbf{H}_p \geq 0, \mathbf{H}_p^{\intercal}\mathbf{1} = \mathbf{1}, \end{aligned} \tag{1}$$

where $\mathbf{H}_p \in \mathbb{R}^{n_p \times n_p}$, n_p and β are the subgraph, number of observed data, and hyper-parameter in the p-th view, respectively. $\boldsymbol{\Psi}(\cdot)$ and $\boldsymbol{\Phi}(\cdot)$ represent different graph regularization terms. The subgraphs $\{\mathbf{H}_p\}_{p=1}^{V}$ can be sketched into $n \times n$ complete graphs, which are then fused into a consistent graph for spectral clustering. Many methods [24], [25] have been developed to boost the performance of IMVC via different regularizations on graphs. Some deep methods

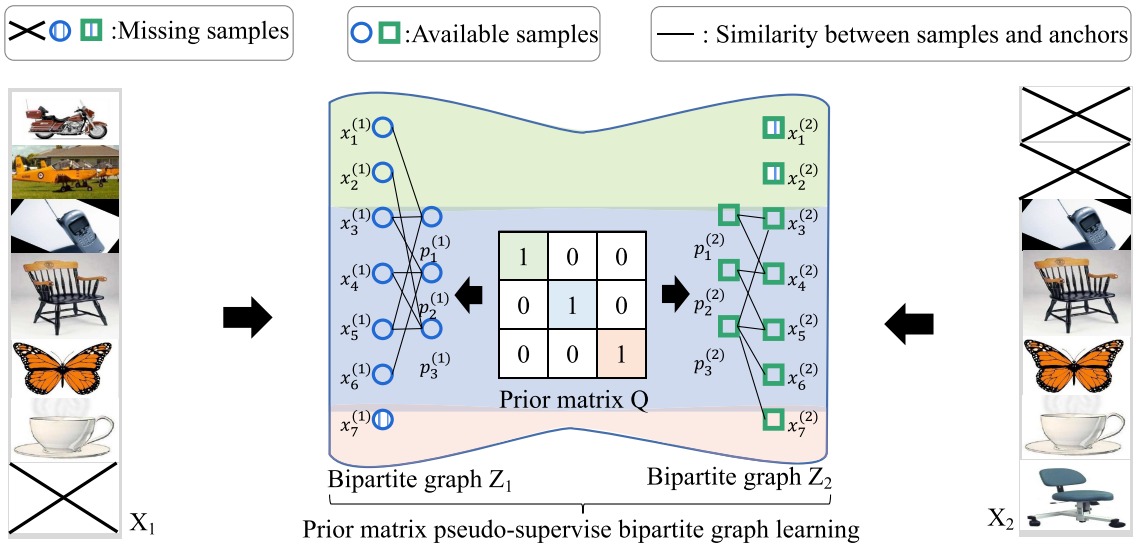


FIGURE 1. Overview of the anchor pseudo-supervise learning of proposed AP-LIMC method. Two views are employed for ease of understanding. From Fig. 1, we develop a prior indicator matrix **Q** artificially to enforce the learned bipartite graph to enjoy the desired k block diagonal structure (this structure of graph plays a vital role in clustering), where k is the cluster number. That is, we set **Q** as the k block diagonal structure, then the prior **Q** could directly make the product of bipartite graph $\mathbf{S}\mathbf{S}^\top$ be also k diagonal blocks. Further, k block diagonal structure $\mathbf{S}\mathbf{S}^\top$ could indirectly make the learned anchors uniform distribution to enhance the representation ability of anchors. The main contributions are that: (1) to obtain the k block diagonal bipartite graph, we design a prior indicator matrix **Q** to directly protect the block diagonal structure of the graph, and indirectly enhance the representation ability of anchors; and (2) we employ the designed **Q** to develop a prior matrix pseudo-supervise bipartite graph learning model for large-scale incomplete multi-view clustering.

[36], [37] employ deep networks to enhance the quality of the graph, which can better handle complex data and are more robust to noise compared to shallow methods. However, existing shallow methods suffer from $\mathcal{O}(n^3)$ computational complexity and $\mathcal{O}(n^2)$ memory complexity, while deep methods are time-consuming and require lots of samples for deep network training.

B. BIPARTITE GRAPH CLUSTERING

Bipartite graphs are frequently employed for managing extensive datasets through the selection of a limited set of representative anchors that establish connections with the original samples. This approach proves notably efficacious in practice. Without loss of general, bipartite graph clustering framework can be mathematically expressed as

$$\begin{aligned} &\min_{\mathbf{S}_p,\mathbf{S}} \sum_{p=1}^{V} \left\|\mathbf{Y}_p - \mathbf{P}_p\mathbf{S}_p\right\|_{\mathbf{F}}^2 + \beta\Omega(\mathbf{S}_p) \\ &\text{s.t. } \mathbf{S}_p \geq 0, \mathbf{S}_p^\top\mathbf{1} = \mathbf{1} \end{aligned} \tag{2}$$

where $\mathbf{Y}_p \in \mathbb{R}^{d_i\times n}$ and $\mathbf{P}_p \in \mathbb{R}^{d_i\times m}$ represent complete data and its m selected or sampled landmarks corresponding to p-th view. $\Omega(\cdot,\cdot)$ represents certain regularization terms [38], [39]. By reducing the size of the traditional $n \times n$ similarity graph to a $m \times n$ bipartite graph $\mathbf{S}_p$. Eq. (3) can effectively reduce both computational and memory complexity while achieving comparable clustering performance. However, this equation is not suitable for the IMVC method due to the poor quality of learned anchors caused by the random absence of samples from different views. To elaborate, varying perspectives might opt for distinct anchor quantities within the identical cluster, or they might even exclude all anchors from a particular cluster. This discrepancy arises owing to the influence of missing samples. To address this issue, [35] proposes selecting data points with complete views as anchors, but this method requires all clusters to have enough anchors with complete views to be selected, which is not always feasible in real-world applications.

III. THE INNOVATION FORMULA

A. PRIOR INDICATOR GUIDED ANCHOR LEARNING

In order to ensure that the anchor points can be selected uniformly anchors from different missing samples to better represent individual samples, this paper first proposes to introduce the predefined indication matrix $\mathbf{Q} \in \mathbb{R}^{m\times m}$ to guide the learning of individual anchors. The predefined **Q** pseudo-supervises those m anchors that are uniformly learned from each cluster, rather than selecting data points with complete views as anchors. Thus, in this paper, we claim this learning process as an anchor pseudo-supervised learning scheme. By integrating Eq. (2), this idea is mathematically fulfilled as

$$\begin{aligned} &\min_{\boldsymbol{\theta},\{\mathbf{P}_p\}_{p=1}^v,\mathbf{S}} \sum_{p=1}^{v} \theta_p^2 \left\|\mathbf{X}_p\mathbf{E}_p - \mathbf{P}_p\mathbf{S}_p\mathbf{E}_p\right\|_{\mathbf{F}}^2 - \beta\mathtt{Tr}(\mathbf{S}_p\mathbf{S}_p^\top\mathbf{Q}) \\ &\text{s.t. } \mathbf{S}_p \geq 0, \mathbf{S}_p^\top\mathbf{1} = \mathbf{1}, \mathbf{P}_p^\top\mathbf{P}_p = \mathbf{I} \end{aligned} \tag{3}$$

where $\boldsymbol{\theta}$ is the weight vector, and $\mathbf{E}_p$ denotes the index matrix for n_p observed samples corresponding to p-th view. $\mathbf{E}_p$ is

defined as

$$\mathbf{E}_{p,i,j}=\begin{cases}1, & \text{if the entry } e_{p,i,j}=i,\ \forall j=1,2,\ldots,n_p\\ 0, & \text{otherwise.}\end{cases} \tag{4}$$

where $\mathbf{e}_p$ is the indicator vector of n_p observed samples to report the sorted index, p, i, j denotes i-th column and j-th low of p-th view. $\{\mathbf{X}_p\mathbf{E}_p \in \mathbb{R}^{d_p\times n_p}\}_{p=1}^{V}$ are defined as complete data matrices, and $\mathbf{X}_p\mathbf{E}_p$ is sorted observed samples of $\mathbf{X}_p$. Note that the prior indication matrix $\mathbf{Q}$ can be set as follows: (1) The number of diagonal blocks of $\mathbf{Q}$ is equal to the number of sample clusters; and (2) Each block is a $m_i * m_i$ square matrix, where m_i is the number of samples in i-th cluster. Specifically, $\mathbf{Q}$ is defined as follows

$$\mathbf{Q}_{i,j}=\begin{cases}1, & \text{if } i\text{th and } j\text{th anchors in the same cluster}\\ 0, & \text{otherwise.}\end{cases} \tag{5}$$

To enhance the equilibrium of the learned anchors, we strive for an equal number of anchors for each cluster, which, in turn, bolsters the representation capacity of the anchors. Generally, the prior indicator matrix $\mathbf{Q}$ takes the form of a block diagonal matrix with k blocks, where all entries within each block are set to 1, while the non-diagonal entries remain 0. In essence, the specification of $\mathbf{Q}$ governs the count of anchor points chosen within each cluster. Through the execution of Eq. (4), the selection count of anchors in each view becomes consistent, ensuring uniformity across clusters.

B. SEUDO-SUPERVISE UNIFIED BIPARTITE GRAPH LEARNING

Additionally, drawing inspiration from the intrinsic subspace structure often shared among multi-view data, we capitalize on the anchors to jointly and directly acquire a consensus bipartite graph $\mathbf{S}$ as follows:

$$\begin{aligned}\min_{\boldsymbol{\theta},\{\mathbf{P}_p\}_{p=1}^{v},\mathbf{S}} &\sum_{p=1}^{v}\theta_p^2\left\|\mathbf{X}_p\mathbf{E}_p-\mathbf{P}_p\mathbf{S}\mathbf{E}_p\right\|_{\mathbf{F}}^2-\beta\mathtt{Tr}(\mathbf{S}\mathbf{S}^{\intercal}\mathbf{Q}),\\ &\text{s.t. } \mathbf{S}\geq 0,\mathbf{S}^{\intercal}\mathbf{1}=\mathbf{1},\mathbf{P}_p^{\intercal}\mathbf{P}_p=\mathbf{I}\end{aligned} \tag{6}$$

With the help of indicator matrices $\mathbf{E}_p \in \mathbb{R}^{n\times n_p}$, even if i-th sample of p-th view is missing, the affinity values between anchors and i-th sample are not still affected (*i.e.*, $\mathbf{x}_{p,i,:}=\mathbf{0}\rightarrow\mathbf{z}_i=\mathbf{0}$). Thus, $\mathbf{z}_i$ should be learned by borrowing the affinities from the other complete view. This idea can be illustrated by collaborative learning as shown in Fig. 2. Intuitively, $\mathbf{x}_7^{(1)}=\theta_2^2\mathbf{x}_7^{(2)}+\theta_3^2\mathbf{x}_7^{(3)}$, where $(\cdot)_i^{(v)}$ and $(\cdot)_i^v$ denotes the i-th sample of v-th view in Fig. 2 for convenience. In this way, a complete subspace structure matrix $\mathbf{S}$ can be collaboratively learned from the incomplete multi-view data. More importantly, the prior indication matrix $\mathbf{Q}$ can encourage individual anchors in different views to evenly distribute in each cluster. That is, the prior matrix can make individual bipartite graphs to better achieve semantic consistency due to this guidance of even distribution of anchors for each view. This can maximumly help incomplete multi-view data to exploit the underlying subspace structure of bipartite graph $\mathbf{S}$.

C. OPTIMIZATION

According to the augmented Lagrange multiplier method [18], [40], we first introduce an auxiliary variable $\mathbf{D}$ to make Eq. (6) separable, and then rewrite Eq. (6) as the following

$$\begin{aligned}\min_{\boldsymbol{\theta},\{\mathbf{P}_p\}_{p=1}^{v},\mathbf{S}} &\sum_{p=1}^{v}\theta_p^2\left\|\mathbf{X}_p\mathbf{E}_p-\mathbf{P}_p\mathbf{S}\mathbf{E}_p\right\|_{\mathbf{F}}^2-\beta\mathtt{Tr}(\mathbf{S}\mathbf{D}^{\intercal}\mathbf{Q})\\ &+\frac{\mu}{2}\|\mathbf{S}-\mathbf{D}+\frac{\mathbf{J}}{\mu}\|_F^2\\ &\text{s.t. } \mathbf{S}\geq 0,\mathbf{S}^{\intercal}\mathbf{1}=\mathbf{1},\mathbf{D}\geq 0,\mathbf{D}^{\intercal}\mathbf{1}=\mathbf{1},\mathbf{P}_p^{\intercal}\mathbf{P}_p=\mathbf{I}\end{aligned} \tag{7}$$

which can be solved according to the following steps:

Step-1 update S: With fixed the irrelevant variables of $\mathbf{S}$, Eq. (7) becomes

$$\begin{aligned}\min_{\mathbf{S}} &\sum_{p=1}^{V}\theta_p^2\mathtt{Tr}(\mathbf{S}^{\top}\mathbf{S}(\mathbf{E}_p\mathbf{E}_p^{\top}+\frac{\mu}{2\theta_p^2}\mathbf{I}_n)-2\mathbf{X}_p^{\top}\mathbf{P}_p\mathbf{S}\mathbf{E}_p\mathbf{E}_p^{\top})\\ &-\beta\mathtt{Tr}(\mathbf{D}^{\intercal}\mathbf{Q}+\mu(\mathbf{D}-\frac{\mathbf{J}}{\mu}))\mathbf{S}\\ &\text{s.t. } \mathbf{S}\geq 0,\mathbf{S}^{\intercal}\mathbf{1}=\mathbf{1}\end{aligned} \tag{8}$$

where $\mathbf{X}_p\mathbf{E}_p\mathbf{E}_p{}^{\intercal}=\mathbf{X}_p\otimes\mathbf{G}_p$, where $\mathbf{G}_p=\mathbf{1}_{d_p}\mathbf{e}_p$ and $\mathbf{e}_p=\left[e_{p,1},\ldots,e_{p,n}\right]^{\intercal}$ with $e_{p,j}=\sum_{l=1}^{n_p}\mathbf{E}_{p,j,l}$.

Considering $\mathbf{X}_p\mathbf{E}_p\mathbf{E}_p^{\top}=\mathbf{X}_p\otimes\mathbf{G}_p$, we can solve Eq. (8) via the following vector form

$$\min_{\mathbf{z}_i}\|\mathbf{z}_i-\mathbf{b}_i\|_{\mathbf{F}}^2,\ \text{s.t. } \mathbf{z}_i^{\intercal}\mathbf{1}=1,\mathbf{z}_{ij}\geq 0 \tag{9}$$

where $\mathbf{b}_i^{\intercal}=\frac{\mathbf{G}_{p,1,j}\mathbf{X}_{p,:,i}^{\intercal}\mathbf{P}_p+\beta\mathbf{M}_{:,j}}{\sum_{p=1}^{V}\theta_p^2\mathbf{G}_{p,1,j}+\frac{\mu}{2}}$, and $\mathbf{M}_1=\beta\mathbf{D}^{\intercal}\mathbf{Q}^{\intercal}+\mu(\mathbf{D}^{\intercal}-\frac{\mathbf{J}^{\intercal}}{\mu})$. Then, **Theorem 1** can be used to solve this subproblem with $\mathcal{O}(nmd)$ time complexity.

Theorem 1: Given r vectors $\{\mathbf{b}\}_{j=1}^{r}$, its closed-form solution $\mathbf{z}_j^*$ can be solved by

$$\mathbf{z}^*=\arg\min_{\mathbf{z}}\|\mathbf{z}-\mathbf{b}\|_F^2,\ \text{s.t. } \mathbf{z}^{\intercal}\mathbf{1}=1,\mathbf{z}\geq 0 \tag{10}$$

which can be proved by **Theorem 2** of [9].

Step-2 update D: Fixing $\mathbf{S}$, $\mathbf{P}$, $\mathbf{J}$, and $\boldsymbol{\theta}$, the $\mathbf{D}$-subproblem changes to

$$\begin{aligned}\min_{\mathbf{D}} &-\beta\mathtt{Tr}(\mathbf{S}\mathbf{D}^{\intercal}\mathbf{Q}^{\intercal})+\frac{\mu}{2}\|\mathbf{S}-\mathbf{D}+\frac{\mathbf{J}}{\mu}\|_F^2\\ &\text{s.t. } \mathbf{D}\geq 0,\mathbf{D}^{\intercal}\mathbf{1}=\mathbf{1}\end{aligned} \tag{11}$$

Similar to the optimization of $\mathbf{S}$-subproblem, Eq. (11) can be effectively solved via **Theorem 1** with $\mathcal{O}(nmd)$ time complexity.

Step-3 update P: Fixing the irrelevant variables with $\mathbf{P}$, and then optimizing $\mathbf{P}$ becomes to

$$\mathtt{Tr}(\mathbf{P}_p^{\intercal}\mathbf{T}_p),\mathbf{P}_p^{\intercal}\mathbf{P}_p=\mathbf{I}_m, \tag{12}$$

where $\mathbf{T}_p=(\mathbf{X}_p\otimes\mathbf{G_p})\mathbf{S}^{\intercal}$. The optimal solution of optimizing $\mathbf{P}$ can be effectively obtained via singular value

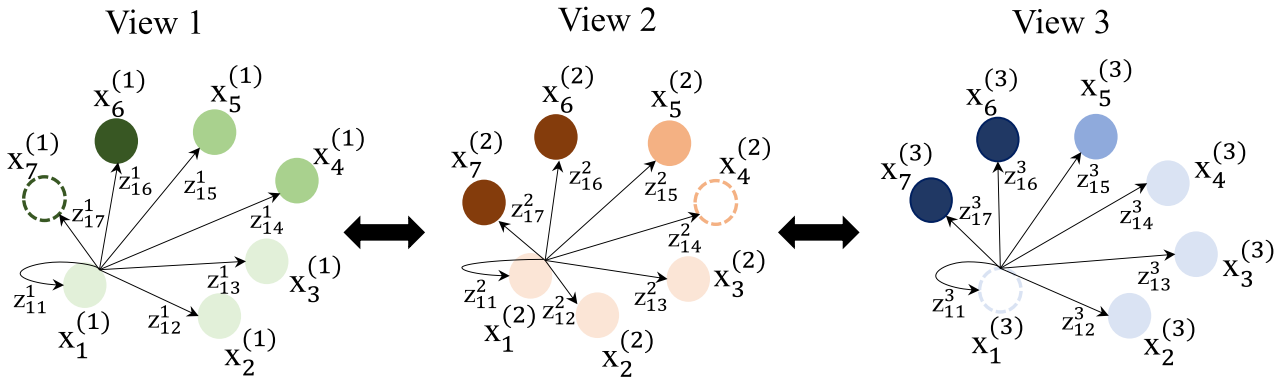


FIGURE 2. Illustration of collaborative filling for incomplete affinities. The dashed lines denote the missing affinities, which require to be filled by the combination of weights for other views.

decomposition (SVD) on $\mathbf{T}_p$ with complexity $\mathbf{W}$ is $\mathcal{O}(dm^2 + nmd)$, where $d = \sum_{p=1}^{V} d_p$.

Step-4 update $\boldsymbol{\theta}$: Optimizing γ with the irrelevant variables fixed is equivalent to the following optimization problem

$$\min_{\theta_p} \sum_{p=1}^{V} {\theta_p}^2 \delta_p, \text{ s.t. } \boldsymbol{\theta}^\top \mathbf{1} = 1, \boldsymbol{\theta} \geq 0 \tag{13}$$

in which $\delta_p = \|\mathbf{X}_p\mathbf{E}_p - \mathbf{P}_p\mathbf{S}\mathbf{E}_p\|_\mathbf{F}^2$. By using Cauchy-Schwarz inequality, the weight vector $\boldsymbol{\theta}$ is obtained via

$$\boldsymbol{\theta} = \frac{\boldsymbol{\varepsilon}}{\sum_{p=1}^{V} \varepsilon_p} \tag{14}$$

where $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_V]$ and $\varepsilon_p = \frac{1}{\delta_p}$. Then, we can optimize γ with $\mathcal{O}(nmd)$ complexity.

Step-5 ADMM variables: Fixing the irrelevant variables, and updating $\mathbf{J}$ can be written as

$$\begin{aligned} \mathbf{J} &= \mathbf{J} + \mu(\mathbf{S} - \mathbf{D}) \\ \mu &= min(\rho\mu, \mu_{max}) \end{aligned} \tag{15}$$

where the involved two ADMM variables μ and μ_{max} are respectively set as $1e^{-4}$ and 10^{10}. Algorithm 1 reports the optimization procedure of our AP-LIMC model. The objective function value obj^t of Algorithm 1 is also reported to control the convergence criterion in the t-th iteration.

Algorithm 1 AP-LIMC Algorithm

Input: V incomplete data $\{\mathbf{X}_p\}_{p=1}^{V}$, parameter α, number of clusters k and anchor number m.

1: Initialize $\boldsymbol{\theta} = 1/\sqrt{V}$, $\mathbf{P}_p = \mathbf{I}$, others matrices as $\mathbf{0}$. $\mu = 1e^{-4}$, $\mu_{max} = 10^{10}$, $\rho = 2$.

Output: Consensus bipartite graph $\mathbf{S}$.

2: **repeat**
3: Update $\mathbf{S}$ via Eq. (8);
4: Update $\mathbf{D}$ via Eq. (11);
5: Update $\mathbf{P}_p$ via (12);
6: Update $\boldsymbol{\theta}$ via Eq. (13);
7: Update ADMM variables via Eq. (14);
8: **until** Satisfy $(obj^{(t)} - obj^{(t-1)})/obj^{(t)} \leq 1\ e-4$.
9: Perform k-means on $\mathbf{S}$.

D. COMPUTATIONAL COMPLEXITY

The main computational complexity of Algorithm 1 is involved from **Step-1** to **Step-4**, *i.e.*, $2\mathcal{O}(nmd) + \mathcal{O}(nm) + \mathcal{O}(dm^2 + nmd) + \mathcal{O}(nmd)$ at each iteration, where n and m denote number of complete samples and anchors, respectively. After obtaining $\mathbf{S}$, it costs $\mathcal{O}(nm^2)$ complexity to perform k-means. Consequently, Algorithm 1 involves linear computation complexity to n. Table 2 reports the detailed computational and memory complexity of all compared methods.

E. MEMORY COMPLEXITY

Memory complexity of Algorithm 1 mainly involves four matrices $\mathbf{G}_p \in \mathbb{R}^{d_p \times n}$, $\mathbf{P}_p \in \mathbb{R}^{d_p \times m}$, and $\mathbf{S}_p \in \mathbb{R}^{m \times n}$. Therefore, memory complexity of AP-LIMC is $(n+k)(d+m)$, which can be further approximated as n when dealing with large-scale data, *i.e.*, $k \ll n$, $m \ll n$, and $d \ll n$.

IV. EXPERIMENTS

A. EXPERIMENT SETTINGS

1) FUNDAMENTAL DATASETS

Six commonly used datasets are leveraged to evaluate the proposed method, including: Cifar10, Cifar100,[1] STL10,[2] WebKB,[3] Caltech101-7,[4] and NGs.[5] Detailed information of these datasets is provided in Table 3. Concretely, Cifar100, Cifar10, STL10, and Caltech101-7 are the image datasets. NGs and WebKB are the web page and document datasets, respectively. Note that the number of samples in these datasets ranges from 500 to 50,000. This span is already relatively large in existing IMVC.

Following [41], for each view of all datasets mentioned above, we randomly remove parts of instances to construct the incomplete multi-view datasets for performing the

[1] http://www.cs.toronto.edu/ kriz/cifar.html
[2] https://cs.stanford.edu/ acoates/stl10/
[3] http://www.cs.umd.edu/ sen/lbc-proj/LBC.html
[4] http://www.vision.caltech.edu/Image Datasets/Caltech101/
[5] https://lig-membres.imag.fr/grimal/data.html

TABLE 2. Detailed information of the used datasets.

Method	Memory Cost	Time Complexity
Ours	$mn + (d + m)k$	$\mathcal{O}(nm + 2nmd + mdk)$
IMVC-CBG [41]	$mn + (d + m)k$	$\mathcal{O}(ndk + nmd + mdk)$
FLSD [24]	$vn^2 + dnk + nk$	$\mathcal{O}\left(nd^2\right)$
UEAF [25]	$vn^2 + dn + nvk + dk$	$\mathcal{O}\left(n^3 + dk^2\right)$
MKKM-IK [28]	$vn^2 + vnk$	$\mathcal{O}\left(vn^3\right)$
EEIMVC [29]	$vn^2 + vnk + vk^2$	$\mathcal{O}\left(nk^2 + vk^3\right)$
MIC [21]	$vn^2 + nd + nvk + vdk$	$\mathcal{O}\left(n^3 + n^2dk\right)$
DAIMC [22]	$vn^2 + nd + nk + dk$	$\mathcal{O}\left(nd^3 + ndk\right)$
APMC [35]	$nd + vnm + nk$	$\mathcal{O}\left(n^3 + nmd + m^3\right)$
BSV [42]	vn^2	$\mathcal{O}\left(n^3\right)$

incomplete multi-view clustering. But these incomplete data should satisfy the principle that each sample requires to be presented at least one view, where incomplete multi-view datasets are generated by denoting the missing ratio as [0.1, 0.9] in intervals of 0.1 [33], [41].

2) EVALUATION METRIC AND COMPARISON METHOD

We compare our method with nine competitors on four common metrics [29], including ACC, NMI, PUR, and Fscore. Specifically, ACC is a measure of how many instances are correctly classified. It is calculated by dividing the number of correctly classified instances by the total number of instances and multiplying by 100%. NMI is a measure of the similarity between two sets of classifications. It is calculated by comparing the joint probability distribution of the two sets of classifications with the product of their individual probability distributions. NMI ranges from 0 to 1, where 0 indicates no overlap between the two sets and 1 indicates perfect overlap. PUR is a measure that combines precision and recall into a single metric. Precision measures the proportion of correctly classified instances among all instances classified as positive, while recall measures the proportion of correctly classified instances among all actual positive instances. PUR is calculated by plotting precision against the recall and calculating the area under the curve. A perfect score is 1, indicating perfect precision and recall, while a score of 0 indicates random performance. These comparisons include:

- **BSV** [43] applies individual spectral clustering to each view, imputing mean in incomplete segments and presenting optimal single-view results.
- **MIC** [21] acquires latent representations from individual views and optimizes a shared clustering representation.
- **IMKKM-IK** [28] is an incomplete multi-kernel approach that addresses missing components, applying kernel k-means.
- **DAIMC** introduces view-specific weight matrices, tackling missing views and aligning basis matrices.
- **APMC** [35] suggests using samples across views as anchors and performing spectral clustering for a conclusive outcome.
- **UEAF** [25] simultaneously recovers absent views and learns a unified clustering representation.
- **EEIMVC** [29] generates reduced-dimensional base feature matrices, imputes, and optimizes a consensus matrix within a unified framework.
- **FLSD** [24] learns distinct latent representations per view, aiming for shared clustering based on semantic consistency.
- **IMVC-CBG** [41] presents a scalable anchor graph framework to address IMC for the first time.

3) EXPERIMENTAL CONDITIONS AND SETTINGS

All codes of competitors are collected from their public home page or Git Hub, and the implementation of their experiments is according to their settings in the corresponding papers. We repeat 20 times for each experiment, and then the average results and their standard deviations are all reported, where the best results marked in bold. i9 CPU and 32GB RAM are employed to construct the computing platform of 2016 Matlab.

B. NUMERICAL EXPERIMENT RESULTS

Table 4 reports the averaged clustering results among all missing ratios on the three less than 10,000 samples of benchmark datasets, where the bold values represent the best-averaged results. Meanwhile, Fig. 3 demonstrates the variation of ACC and NMI on the three more than 10,000 samples of datasets compared with two competitors. To further demonstrate the ability to handle large-scale IMVC, Table 5 and Fig. 4 also report the averaged clustering performance across all missing ratios and standard deviations on the three more than 10,000 samples of datasets. Note that only three methods, *i.e.*, DAIMC, IMVC-CBG, and our AP-LIMC can be used to estimate the performance on Cifar10 and Cifar100 datasets. The other comparisons fail due to their high memory and time complexity caused by too large a

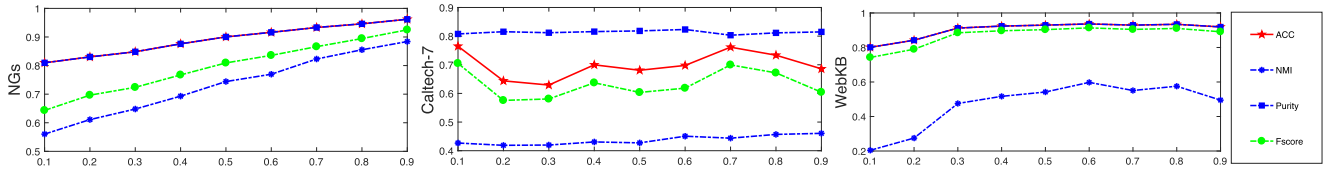


FIGURE 3. The averaged ACC and NMI on the three evaluated datasets, and their size is less than 10,000 samples.

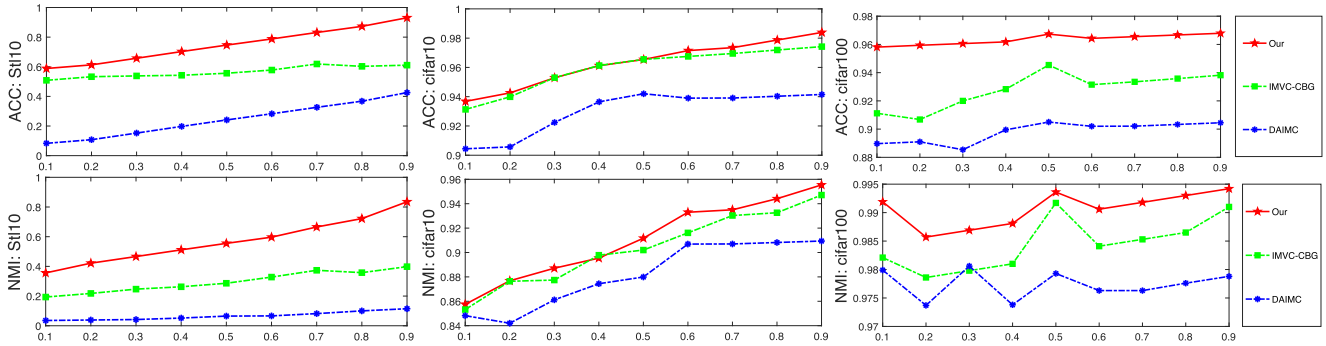
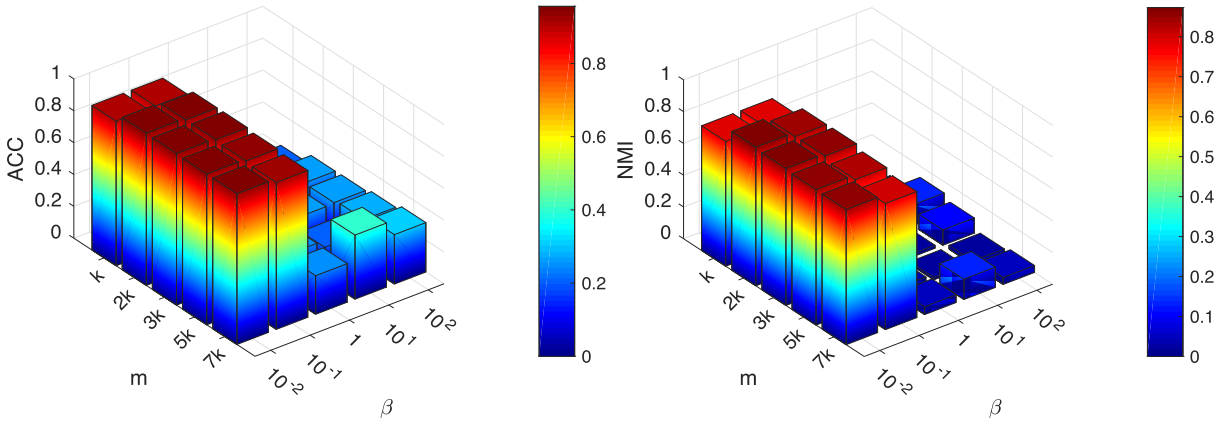


FIGURE 4. The averaged ACC and NMI on the three evaluated datasets, and their size is more than 10,000 samples.

TABLE 3. Details of the evaluated datasets. Sample, Classes, Views, and Dimensionality mean the sample number n, class number k, view number V, and dimensionality d of each dataset, respectively.

Dataset	Sample	Classes	Views	Dimensionality
Cifar10	50,000	10	3	2,048 / 512 / 1,024
STL10	13,000	10	3	2,048 / 512 / 1,024
WebKB	1,051	2	2	1,840 / 3,000
Caltech101-7	1,474	7	6	48 / 40 / 254 / 198 / 512 / 928
Cifar100	50,000	100	3	2,048 / 512 /1,024
NGs	500	5	3	500 / 500 / 500

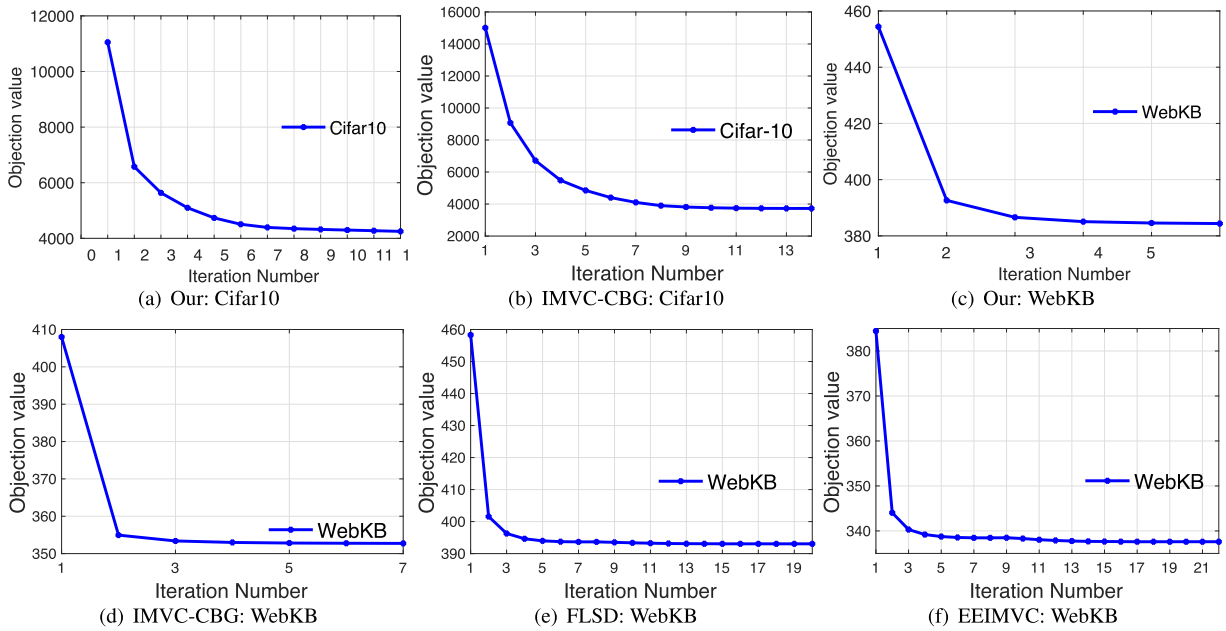


(a) NGs:ACC (b) NGs:NMI

FIGURE 5. ACC and NMI *w.r.t.* m and β on the NGs datasets (missing ratio=0.2).

number of samples as shown in Table 2. From these tables and figures, we can observe that:

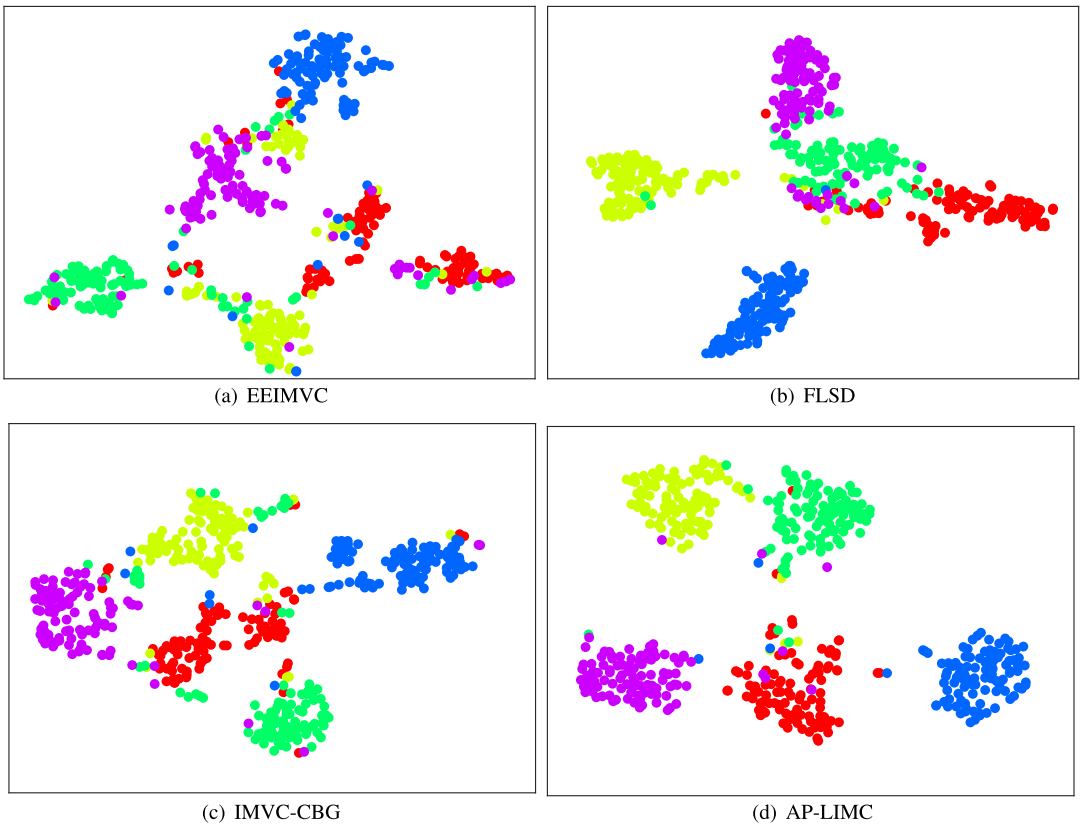
- Most incomplete multi-view methods are better than single-view methods with mean filling. However, some of the IMVC methods are even lower than the single-view method in some cases. As shown in Table 4, our method not only consistently and largely outperforms the single-view incomplete clustering, but outperforms the existing multi-view methods. Taking the WebKB dataset as an example, many IMVC



(a) Our: Cifar10 (b) IMVC-CBG: Cifar10 (c) Our: WebKB

(d) IMVC-CBG: WebKB (e) FLSD: WebKB (f) EEIMVC: WebKB

FIGURE 6. Objection value over the Cifar10 and WebKB datasets (missing ratio=0.2).



(a) EEIMVC (b) FLSD

(c) IMVC-CBG (d) AP-LIMC

FIGURE 7. Illustrations of the clustering result of IMVC-CBG and our AP-LIMC with T-SNE method on the NGs dataset (missing ratio=0.2).

methods, such as matrix-factory based MIC, kernel based IMKKM-IK, and graph based FLSD only achieve several NMI. While AP-LIMC can enjoy the average 91.07% NMI stably. This demonstrates the superiority of AP-LIMC.

- In most circumstances, AP-LIMC is superior to the most recent methods by a large margin, especially in handling more than 10,000 samples of incomplete datasets in Table 5, AP-LIMC consistently outperforms DAIMC and IMVC-CBG. This indicates that

TABLE 4. Average clustering results for various of missing ratios on three datasets of less than 10,000 samples. The best performance is marked in bold. '-' means out of the CPU memory. Supplementary materials report the more detailed results.

Dataset	Metrics	BSV	MIC	DAIMC	APMC	UEAF	IMKKM-IK	EEIMVC	FLSD	IMVC-CBG	Our
NGs	Fscore	33.64±0.99	32.9±0.19	64.83±7.49	80.28±0.02	**80.55±0.06**	42.28±0.09	63.75±0.14	71.8±0.05	79.49±0.03	79.57±0.00
	PUR	45.36±1.67	21.3±0.42	75.56±9.27	89.41±0.01	**89.49±0.03**	50.57±0.10	77.84±0.12	84.19±0.03	88.95±0.02	89.14±0.00
	NMI	22.31±1.12	2.37±0.49	59.97±6.48	73.36±0.03	**73.60±0.08**	32.76±0.11	57.23±0.18	64.22±0.04	73.05±0.06	73.21±0.00
	ACC	44.25±1.89	20.92±0.46	74.53±9.27	89.41±0.01	**89.49±0.03**	49.87±0.13	77.84±0.12	84.19±0.03	88.95±0.02	89.14±0.00
Caltech101-7	Fscore	55.24±0.08	37.74±1.35	49.44±2.93	-	44.48±0.19	43.60±1.66	44.65±0.34	55.63±0.03	59.28±0.98	**63.29±0.05**
	PUR	62.37±0.86	71.58±1.24	81.12±1.08	-	80.91±0.10	77.52±0.57	80.60±0.25	80.03±0.27	80.96±0.08	**81.35±0.01**
	NMI	18.34±0.51	24.8±1.13	**45.45±2.12**	-	39.98±0.22	31.26±0.53	42.96±0.33	37.20±0.35	43.58±0.58	43.37±0.01
	ACC	53.26±0.01	37.33±1.97	42.10±3.72	-	40.65±0.34	43.64±2.02	40.10±0.67	55.08±0.22	60.40±0.88	**69.96±0.04**
WebKB	Fscore	58.33±0.32	61.10±0.61	-	84.05±0.03	78.51±0.73	60.21±1.86	62.75±2.14	79.40±0.00	83.01±0.01	**87.08±0.00**
	PUR	73.08±0.61	78.12±0.00	-	90.17±0.02	82.91±0.09	76.34±0.00	78.12±0.12	78.72±1.15	84.52±0.02	**90.32±0.00**
	NMI	1.62±1.65	3.29±0.29	-	47.88±0.10	23.52±0.97	3.21±3.64	3.57±0.56	4.56±4.14	37.27±0.10	**47.02±0.00**
	ACC	55.23±2.17	63.88±0.76	-	85.36±0.03	82.64±0.67	60.61±0.02	61.67±3.49	78.21±0.00	84.52±0.02	**90.32±0.00**

TABLE 5. Average clustering results for various of missing ratios on three datasets of more than 10,000 samples. Meanwhile, time complexity is also reported. The best performance is marked in bold.

Dataset	Metrics	DAIMC 2019	IMVC-CBG 2022	Our
Stl10	Fscore	11.23±1.32	35.83±0.03	**59.78±0.00**
	PUR	20.51±1.78	56.54±0.02	**72.25±0.00**
	NMI	6.62±1.37	29.60±0.03	**56.99±0.00**
	ACC	24.27±1.62	56.54±0.02	**74.78±0.00**
Cifar10	Fscore	89.38±0.67	92.60±0.26	**93.28±0.00**
	PUR	93.20±0.86	95.62±0.23	**96.30±0.00**
	NMI	88.19±1.22	90.05±0.26	**91.07±0.00**
	ACC	93.01±1.01	95.62±0.19	**96.30±0.00**
Cifar100	Fscore	88.45±0.68	90.43±0.69	**95.59±0.01**
	PUR	92.60±0.99	94.67±0.82	**97.37±0.01**
	NMI	97.74±1.56	98.84±0.55	**99.07±0.02**
	ACC	89.81±1.61	92.47±0.46	**96.35±0.01**
Dataset	Time cost	DAIMC 2019	IMVC-CBG 2022	Our
Stl10	Seconds	680.21	30.62	6.79
Cifar10	Seconds	3105.63	106.31	22.82
Cifar100	Seconds	28365.78	121.93	67.59

AP-LIMC has a strong capacity to handle large-scale IMVC.

- Compared to APMC, which selects anchors by using the sample points with complete views, both IMVC-CBG and AP-LIMC can learn the anchors instead of the fixed anchors to significantly improve performance for APMC across all missing ratios. The results of Table 4, Table 5, and Fig. 4 demonstrate that IMVC-CBG and AP-LIMC have better clustering performance. This proves the advantages of learning anchors dynamically.
- Compared to the second best method IMVC-CBG, although both IMVC-CBG and AP-LIMC can learn anchors from incomplete data, IMVC-CBG aims to explore the view-consistence relationship between multi-view data. Moreover, it lacks guidance from a pr indicator matrix. While AP-LIMC learns view-specific anchors with the guidance of the prior matrix, which greatly improves the quality of anchors and simultaneously better explores the complementary information between different views. As mentioned in Table 4, Table 5, and Fig. 4, AP-LIMC outperforms IMVC-CBG *w.r.t.* all evaluated metrics and missing ratios in most cases. Furthermore, AP-LIMC shows substantially shorter running over the large-scale datasets, *i.e.*, Cifar10 and Cifar100. All the above analyses indicate the effectiveness and the efficiency of AP-LIMC to cluster large-scale IMVC.
- For large-scale datasets (such as Stl10, Cifar10, and Cifar100), our method achieves satisfactory results. For example, on the Stl10 dataset, the performance of the proposed AP-LIMC improved by 23.95%,15.71%,27.39%, and 18.24% in terms of four evaluation metrics, respectively. On the one hand, many methods are difficult to adapt to large-scale data due to their high computational complexity. On the other hand, large-scale data can learn more semantic information, thus improving the performance of our method.

C. PARAMETER ANALYSIS

Two parameters, *i.e.*, anchor number m and parameter β are analyzed for Algorithm 1. Specifically, we vary m in $[k, 2k, 3k, 5k, 7k]$ and tune β in $[10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}]$. As shown in Fig. 5, by performing a grid search on the NGs dataset, our AP-LIMC can obtain satisfying clustering

results in a wide scope for m and β. It can be seen that we only require a few anchors to get encouraging clustering performance.

D. EFFICIENCY AND CONVERGENCE

Table 5 reports average clustering results for various of missing ratios on three datasets of more than 10,000 samples. Meanwhile, it also displays the average times of DAIMC, IMVC-CBG, and the proposed method across more than 10,000 samples of benchmark datasets. As can be seen, compared to both DAIMC and IMVC-CBG, our method consistently outperforms them on three datasets of more than 10,000 samples, *i.e.*, enjoying the optimal clustering performance while having minimum time consumption.

In addition, Fig. 6 reports the convergence of IMVC-CBG [41], FLSD [32], EEIMVC [29], and our method on the two datasets, *i.e.*, Cifar10 and WebKB in the missing ratio of 0.2. To be specific, Fig. 6 gives the objective value of four Algorithms with iterations. It can be seen that our objective value of Algorithm 1 decreases monotonically with iteration, indicating the favorable convergence of the proposed AP-LIMC. In addition, our AP-LIMC can achieve faster convergence than the three competitors on the two evaluated datasets. This indicates the effectiveness of our AP-LIMC in the convergence.

E. VISUALIZATION RESULTS

The learned ultimate bipartite graphs of IMVC-CBG [41], FLSD [32], EEIMVC [29], and our method on the NGs dataset are plotted in Fig. 7. In Fig. 7, each point of different colors represents a cluster of the NGs dataset, with the same color indicating the same clusters of samples. Obviously, our AP-LIMC enjoys clearer cluster discrimination and a more compact cluster structure. More importantly, it can be seen that our AP-LIMC is able to learn a higher-quality bipartite graph than the three state-of-the-art competitors.

V. CONCLUSION

In this paper, we first propose an anchor pseudo-supervise learning technique to guide the learning of anchor matrix. By employing the anchor pseudo-supervise learning technique, a novel incomplete multi-view clustering method, named AP-LIMC, is proposed to simultaneously perform dynamic anchor learning and IMVC. Different from selecting sample points with complete views as the anchors, AP-LIMC learns anchors dynamically from incomplete views with the guidance of the prior indicator matrix. Meanwhile, AP-LIMC has linear memory and time complexity, demonstrating its potential to deal with large-scale tasks. Extensive experiments on six commonly used datasets with four metrics indicate the effectiveness and efficiency of our AP-LIMC. For AP-LIMC, the prior indication matrix is fixed, which may limit the flexibility of anchors to some extent. In the future, we will explore dynamic optimization of the indication matrix to further improve the representation of anchors.

REFERENCES

[1] C. Zhang, H. Li, W. Lv, Z. Huang, Y. Gao, and C. Chen, "Enhanced tensor low-rank and sparse representation recovery for incomplete multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2023, vol. 37, no. 9, pp. 11174–11182.

[2] J. You, Z. Ren, F. R. Yu, and X. You, "One-stage shifted Laplacian refining for multiple kernel clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 4, 2023, doi: 10.1109/TNNLS.2023.3262590.

[3] J. You, Z. Ren, X. You, H. Li, and Y. Yao, "Priori anchor labels supervised scalable multi-view bipartite graph clustering," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2023, vol. 37, no. 9, pp. 10972–10979.

[4] Y. He and U. K. Yusof, "Self-weighted graph-based framework for multi-view clustering," *IEEE Access*, vol. 11, pp. 30197–30207, 2023.

[5] Z. Ren and Q. Sun, "Simultaneous global and local graph structure preserving for multiple kernel clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 1839–1851, May 2021.

[6] C. Zhang, H. Li, C. Chen, X. Jia, and C. Chen, "Low-rank tensor regularized views recovery for incomplete multiview clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 30, 2022, doi: 10.1109/TNNLS.2022.3232538.

[7] J. Liu, X. Liu, Y. Yang, S. Wang, and S. Zhou, "Hierarchical multiple kernel clustering," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 8671–8679.

[8] X. Liu, L. Liu, Q. Liao, S. Wang, Y. Zhang, W. Tu, C. Tang, J. Liu, and E. Zhu, "One pass late fusion multi-view clustering," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 6850–6859.

[9] Z. Ren, S. X. Yang, Q. Sun, and T. Wang, "Consensus affinity graph learning for multiple kernel clustering," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3273–3284, Jun. 2021.

[10] Z. Ren, Q. Sun, and D. Wei, "Multiple kernel clustering with kernel K-means coupled graph tensor learning," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 11, pp. 9411–9418.

[11] Z. Kang, W. Zhou, Z. Zhao, J. Shao, M. Han, and Z. Xu, "Large-scale multi-view subspace clustering in linear time," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 4, pp. 4412–4419.

[12] C. Zhang, H. Li, C. Chen, Y. Qian, and X. Zhou, "Enhanced group sparse regularized nonconvex regression for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2438–2452, May 2022.

[13] J. You, Y. Hou, Z. Ren, X. You, J. Dai, and Y. Yao, "Cluster center consistency guided sampling learning for multiple kernel clustering," *Inf. Sci.*, vol. 606, pp. 410–422, Aug. 2022.

[14] X. Li, Z. Ren, Q. Sun, and Z. Xu, "Auto-weighted tensor schatten p-norm for robust multi-view graph clustering," *Pattern Recognit.*, vol. 134, Feb. 2023, Art. no. 109083.

[15] X. Li, Y. Sun, Q. Sun, and Z. Ren, "Consensus cluster center guided latent multi-kernel clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 6, pp. 2864–2876, Jun. 2023.

[16] Y. Mei, Z. Ren, B. Wu, T. Yang, and Y. Shao, "Robust kernelized multiview clustering based on high-order similarity learning," *IEEE Access*, vol. 10, pp. 54221–54234, 2022.

[17] J. Wang, Y. Liu, and W. Ye, "FMvC: Fast multi-view clustering," *IEEE Access*, vol. 11, pp. 12808–12820, 2023.

[18] Y. Sun, X. Wang, D. Peng, Z. Ren, and X. Shen, "Hierarchical hashing learning for image set classification," *IEEE Trans. Image Process.*, vol. 32, pp. 1732–1744, 2023.

[19] Y. Sun, D. Peng, H. Huang, and Z. Ren, "Feature and semantic views consensus hashing for image set classification," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 2097–2105.

[20] X. Li, Q. Sun, Z. Ren, and Y. Sun, "Dynamic incomplete multi-view imputing and clustering," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 3412–3420.

[21] W. Shao, L. He, and P. S. Yu, "Multiple incomplete views clustering via weighted nonnegative matrix factorization with $L_{2,1}$ regularization," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2015, pp. 318–334.

[22] M. Hu and S. Chen, "Doubly aligned incomplete multi-view clustering," in *Proc. 13th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2019, pp. 2262–2268.
[23] J. Wen, H. Sun, L. Fei, J. Li, Z. Zhang, and B. Zhang, "Consensus guided incomplete multi-view spectral clustering," *Neural Netw.*, vol. 133, pp. 207–219, Jan. 2021.
[24] J. Wen, Z. Zhang, Z. Zhang, L. Fei, and M. Wang, "Generalized incomplete multiview clustering with flexible locality structure diffusion," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 101–114, Jan. 2020.
[25] J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, and H. Liu, "Unified embedding alignment with missing views inferring for incomplete multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 5393–5400.
[26] X. Liu, X. Zhu, M. Li, L. Wang, C. Tang, J. Yin, D. Shen, H. Wang, and W. Gao, "Late fusion incomplete multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2410–2423, Oct. 2018.
[27] X. Zhu, X. Liu, M. Li, E. Zhu, L. Liu, Z. Cai, J. Yin, and W. Gao, "Localized incomplete multiple kernel K-means," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 3271–3277.
[28] X. Liu, X. Zhu, M. Li, L. Wang, E. Zhu, T. Liu, M. Kloft, D. Shen, J. Yin, and W. Gao, "Multiple kernel kk-means with incomplete kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1191–1204, May 2020.
[29] X. Liu, M. Li, C. Tang, J. Xia, J. Xiong, L. Liu, M. Kloft, and E. Zhu, "Efficient and effective regularized incomplete multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2634–2646, Aug. 2020.
[30] Y. Zhang, X. Liu, S. Wang, J. Liu, S. Dai, and E. Zhu, "One-stage incomplete multi-view clustering via late fusion," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 2717–2725.
[31] W. He, Z. Zhang, Y. Chen, and J. Wen, "Structured anchor-inferred graph learning for universal incomplete multi-view clustering," *World Wide Web*, vol. 26, pp. 1–25, Mar. 2022.
[32] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1418–1429, Apr. 2020.
[33] J. Wen, Z. Zhang, Z. Zhang, L. Zhu, L. Fei, B. Zhang, and Y. Xu, "Unified tensor framework for incomplete multi-view clustering and missing-view inferring," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 11, pp. 10273–10281.
[34] J. Liu, X. Liu, Y. Zhang, P. Zhang, W. Tu, S. Wang, S. Zhou, W. Liang, S. Wang, and Y. Yang, "Self-representation subspace clustering for incomplete multi-view data," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 2726–2734.
[35] J. Guo and J. Ye, "Anchors bring ease: An embarrassingly simple approach to partial multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, Jul. 2019, vol. 33, no. 1, pp. 118–125.
[36] J. Wen, Z. Zhang, Z. Zhang, Z. Wu, L. Fei, Y. Xu, and B. Zhang, "DIMC-net: Deep incomplete multi-view clustering network," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 3753–3761.
[37] Y. Lin, Y. Gou, Z. Liu, B. Li, J. Lv, and X. Peng, "COMPLETER: Incomplete multi-view clustering via contrastive prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 11174–11183.
[38] W. Xia, Q. Gao, Q. Wang, X. Gao, C. Ding, and D. Tao, "Tensorized bipartite graph learning for multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 5187–5202, Apr. 2022.
[39] Z. Ren, H. Lei, Q. Sun, and C. Yang, "Simultaneous learning coefficient matrix and affinity graph for multiple kernel clustering," *Inf. Sci.*, vol. 547, pp. 289–306, Feb. 2021.
[40] Y. Sun, Z. Ren, P. Hu, D. Peng, and X. Wang, "Hierarchical consensus hashing for cross-modal retrieval," *IEEE Trans. Multimedia*, early access, May 5, 2023, doi: 10.1109/TMM.2023.3272169.
[41] S. Wang, X. Liu, L. Liu, W. Tu, X. Zhu, J. Liu, S. Zhou, and E. Zhu, "Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jul. 2022, pp. 9776–9785.
[42] F. Nie, J. Li, and X. Li, "Self-weighted multiview clustering with multiple graphs," in *Proc. IJCAI*, 2017, pp. 2564–2570.
[43] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 14, 2001, pp. 1–8.

SONGBAI ZHU received the M.Sc. degree from the Southwest Automation Research Institute, Mianyang, China. He is currently pursuing the Ph.D. degree with the School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, China. He is a Senior Engineer Professor with the Southwest Automation Research Institute. His research interests include multi-view learning and data analysis.

JIAN DAI received the B.Sc. and M.Sc. degrees from the Southwest University of Science and Technology (SWUST), Mianyang, China, in 2013 and 2018, respectively, and the Ph.D. degree from the Beijing Institute of Technology, Beijing, China, in 2023. He has published more than ten peer-reviewed articles. His research interests include machine learning and computer vision.

GUOLAI YANG received the Ph.D. degree in artillery, automatic weapons, and ammunition engineering from the Nanjing University of Science and Technology. He is currently a Professor with the School of Mechanical Engineering, Nanjing University of Science and Technology, and serves as a Supervisor for Ph.D. students. He undertakes many national scientific research projects and has a wide research scope, including in-depth research on artillery design simulation and tank marching analysis.

ZHENWEN REN (Member, IEEE) received the Ph.D. degree from the Nanjing University of Science and Technology, China, in 2021. He is currently a Professor with the Southwest University of Science and Technology, China. He has published more than 70 peer-reviewed papers, including those in highly regarded journals and conferences, such as IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, and IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. His research interests include computer vision, machine learning, deep learning, and industrial software.