Incomplete Multi-view Deep Clustering with Data Imputation and Alignment

Jiyuan Liu[1], **Xinwang Liu**[1*], **Xinhang Wan**[1], **Ke Liang**[1], **Weixuan Liang**[1]
Sihang Zhou[1], **Huijun Wu**[1] **and Kehua Guo**[2]
[1] National University of Defense Technology, Changsha, Hunan, China. 410072.
[2] Central South University, Changsha, Hunan, China. 410083.
liujiyuan13@nudt.edu.cn

Abstract

Incomplete multi-view deep clustering is an emerging research hot-pot to incorporate data information of multiple sources or modalities when parts of them are missing. Most of existing approaches encode the available data observations into multiple view-specific latent representations and subsequently integrate them for the next clustering task. However, they ignore that the latent representations are unique to a fixed set of data samples in all views. Meanwhile, the pair-wise similarities of missing data observations are also failed to utilize in latent representation learning sufficiently, leading to unsatisfactory clustering performance. To address these issues, we propose an incomplete multi-view deep clustering method with data imputation and alignment. Assuming that each data sample corresponds to a same latent representation among all views, it projects the latent representations into feature spaces with neural networks. As a result, not only the available data observations are reconstructed, but also the missing ones can be imputed accordingly. Moreover, a linear alignment measurement of linear complexity is defined to compute the pair-wise similarities of all data observations, especially including those of the missing. By executing the above two procedures iteratively, the discriminative latent representations can be learned and used to group the data into categories with off-the-shelf clustering algorithms. In experiment, the proposed method is validated on a set of benchmark datasets and achieves state-of-the-art performances.

1 Introduction

With the development of electronical device and information technology, the data observations are widely accumulated from different sources and modalities, collectively referred to as multi-view data in academic communities. For instance, when making advertisement recommendation to Internet users, the content provider would collect their personal details from multiple aspects in advance, such as living habit, shopping record, etc. [1] In the diagnosis and treatment of Alzheimer's disease, multiple types of data are collected and analyzed along with a set of medical examinations, including blood test, Cerebrospinal Fluid (CSF) examination, Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), etc. [2] Therefore, how to integrate the multi-view data effectively and efficiently is a critical problem. In this background, multi-view clustering can explore the consensus and complementary information among different data views and improve the performance over single-view clustering by large margins, catching a large volume of attentions from researchers [3].

*Corresponding author

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

Since most of the existing multi-view clustering are derived from single-view algorithms, they can be grouped into five categories accordingly, i.e., multiple kernel clustering [4, 5, 6, 5], multi-view subspace clustering [7, 8, 9], multi-view spectral clustering [10, 11], multi-view matrix factorization [12, 13] and multi-view deep clustering [14, 15]. Insides, multi-view deep clustering is emerging in recent years along with the rapid development of neural network techniques. Among the advances, Cui et al. propose to extract high-level features from data observations of each view with multiple view-specific fully connected auto-encoders, then design the dual contrasting losses to maximize the distance between different clusters and enhance the compactness within clusters [16]. Yu et al. employ graph convolutional encoder on each data view [17]. Subsequently, the contrastive learning technique and block diagonalization constraint are adopted to guide representation matrix learning, while representation learning is utilized into clustering result generation. Similarly, Chen et al. learn view-invariant data representations with auto-encoders and compute results by contrasting the cluster assignments among different views [18]. Usually, multi-view deep clustering approaches achieve better performance than the others due to the superior capability of neural networks to generate high-quality latent representations on data observations.

Due to sensor failure or restricted conditions, the collected multi-view data are often incomplete where one or more views are missing completely. In such setting, the researchers propose a number of incomplete multi-view deep clustering approaches to make the most of available data observations rather than discarding the data samples with missing views [19, 20]. Mostly, they first encode the available data observations into multiple view-specific latent representations and subsequently integrate them by designing different losses on them. For instance, Chao et al. adopt the contrastive loss to maximize representation similarities of the same data sample in different views while minimize those of different data samples [21]. At the same time, Lin et al. modify the contrastive loss to maximize mutual information across all data views, promoting the learning of informative and consistent latent representations [22].

Although the existing methods achieves satisfactory performance, they ignore the fact that the latent representations of a fixed set of data samples are unique and invariant to different views. Meanwhile, they only concentrate on the integration of available data observations but overlook the pair-wise similarity information among missing data observations, limiting the further improvement of clustering performance. To address these issues, we propose a novel Incomplete Multi-view Deep Clustering with Data Imputation and Alignment (IMDC-DIA). Specifically, it assumes that all views share a same latent representation with respect to a certain data sample. Contrary to the mainstream methods of encoding data observations into latent representations, the proposed approach projects the unique latent representations to feature space of each view with multiple independent neural networks. As a consequence, not only the available data observations are reconstructed, but also the missing ones can be imputed accordingly. Moreover, a linear alignment measurement is defined and analyzed accordingly. Then, we adopt it to align the latent representations with the data observations of each view, so as to utilize the pair-wise similarities of all data observations, especially including those of the missing. By executing the above data imputation and alignment coherently and iteratively, the discriminative latent representations can be learned and used to group the data into categories with off-the-shelf clustering algorithms. To validate effectiveness of the proposed IMDC-DIA method, we conduct extensive experiments on a set of benchmark datasets and compare it with the most recent advances in literature. Corresponding results show it achieves state-of-the-art clustering performance in almost all settings, well validating its effectiveness and superiority.

2 Related work

2.1 Multi-view deep clustering

As a representative of multi-view clustering, multi-view deep clustering is one of the most effective technique to group data samples with integrating their multi-view information. Typically, it achieves better performances than the other multi-view clustering methods, since the adopted neural networks can extract higher-quality latent representations on the data observations of each view. Among them, the Multiple Layer Perceptron (MLP) network is the most widely used in literature, such as those in [16, 23]. Besides, the other types of neural networks are also employed to further improve the quanlity of latent representations. For example, Yu et al. generate the view-specific latent representations with Graph Convolutional Network (GCN) encoder on each view, successfully capturing the structural data information along with learning attribute features [17]. Observing the poor scalability of MLP,

Zhu et al. transform the feature vector into image and use Convolutional Neural Networks (CNN) subsequently, achieving satisfactory clustering performance [24]. Yang et al. adopt the Transformer structure by utilizing self-attention mechanism, enabling the encoder to better retain complementary information across different views and remove the exclusive information [25].

Apart from the neural network structure, plenty of constraints and loss functions are developed in existing researches. Insides, contrastive loss is the most popular one of them. Commonly, it regards the latent representations of a data sample in different views as positive pairs while the rest as negative pairs, and maximizes the similarities of the former while minimizes those of the latter [25, 23]. At the same time, Cui et al. integrate the pseudo-labels in training stage and consider the latent representations of data samples in a same temporary cluster as positive pair, improving the clustering performance effectively [16]. In addition, a large number of researches employ the self-representation loss derivate from the self-representation constraint in classical multi-view clustering [24, 26]. It assumes each latent representation to be a linear sum of the others and minimizes their differences. Nevertheless, some of the other popular constraints and loss functions would be the adversarial similarity constraint [27], the discriminative constraint [15] and the self-supervised loss [17].

2.2 Incomplete multi-view deep clustering

To deal with incomplete data in real-world scenarios, incomplete multi-view deep clustering is emerging to be a promising approach and explored in recent literature. Some of them only rely on the available data observations to compute the clustering results [28, 29]. They usually encode the available data observations into latent representations with multiple neural networks with each in one data view. On the basis, Wen et al. propose a graph embedding strategy to simultaneously capture the high-level features and local structure of each data view, while a self-paced strategy is also utilized to select the most confident samples in model training, reducing the negative influence of outliers [29]. Differently, Xu et al. learn latent representations by incorporating the Mixture-of-Gaussians prior information to enhance their clustering-friendly structure and develop a Product-of-Experts approach to efficiently aggregate them [30]. Xu et al. utilize all view-specific latent representations into a consensus one with an adaptive feature projection module, avoiding to impute the missing data [31]. Then, the correlated common cluster information is explored by maximizing its mutual information, while the distribution alignment is achieved by minimizing its mean discrepancy coherently.

Different from the above approaches, the other incomplete multi-view deep clustering methods try to recover the missing data along with the latent representation learning and data clustering [21, 32, 33]. For instance, Wang et al. learn the low-dimensional latent representations with view-specific encoder networks and explicitly generate the missing data with generative adversarial networks at the same time [33]. Lin et al. unify the cross-view consistency learning and data recovery techniques by maximizing the mutual information of different views and minimizing the conditional entropy through dual prediction, respectively [22, 34]. Liu et al. propose a two-stage autoencoder network with recurrent graph reconstruction mechanism to extract high-level latent representations and recover the missing data synchronously [35].

3 Methodology

In this section, the proposed IMDC-DIA method is introduced, where its framework is visualized in Fig. 1. Briefly, it aims to learn the high-quality latent representations which are subsequently fed to off-the-shelf clustering algorithms so as to group the data accurately. To accomplish this, the latent representations are assumed to be unique to all views. On the basis, three main components are concerned, i.e., data reconstruction, imputation and alignment. On the left of Fig. 1, the proposed IMDC-DIA method projects the unique latent representations to feature space of each view with multiple independent neural networks. Thereby, not only the available data observations are reconstructed (data reconstruction), but also the missing ones can be imputed and completed (data imputation) accordingly. In data alignment, a linear alignment measurement is defined and, as shown on the right of Fig. 1, is adopted to align the unique data representations with the completed data observations of each view.

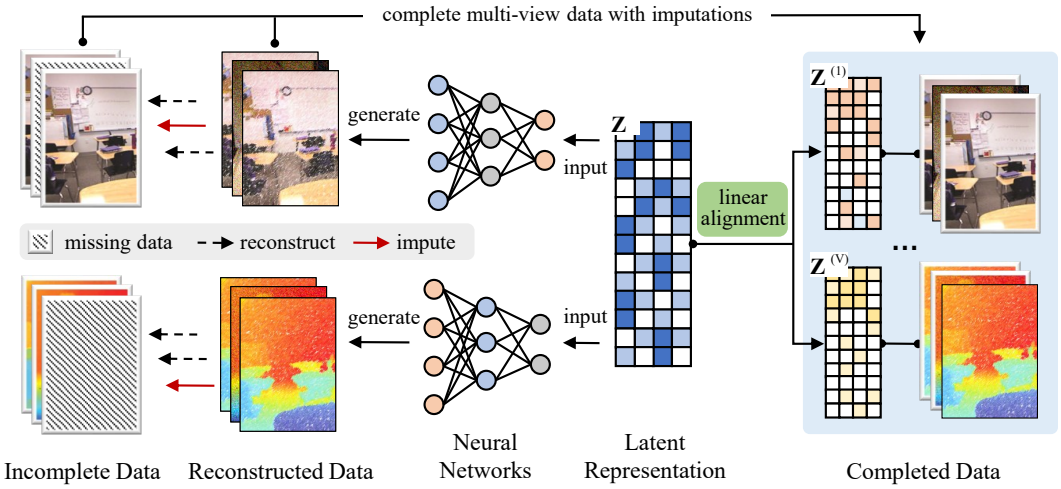

Figure 1: Framework of the proposed IMDC-DIA method. For the ease of expression, only two data views are presented, but arbitrary number of views are supported. Also, any types of neural networks, apart from the fully-connected, can be employed, only if they project the latent representations into data feature space.

3.1 Formulation

Denoting N, V and D_v as the number of data samples, the number of views and the feature dimension of v-th view, the multi-view data observations can be specified to $\{\mathbf{x}_i^{(v)}\}_{i,v=1}^{N,V}$ in which $\mathbf{x}_i^{(v)} \in \mathbb{R}^{D_v}$. Corresponding availability indicators are $\{\mathbf{a}_i\}_{i=1}^{N}$ where $\mathbf{a}_i$ is a discrete vector in $\{0,1\}^V$. Here, the data observation of v-th view $\mathbf{x}_i^{(v)}$ is available if the v-th element $\mathbf{a}_{i,v}$ is 1, while missing if 0. In such setting, incomplete multi-view deep clustering methods aim to group the data into clusters with the available observations where the cluster number K is known in advance.

3.1.1 Data reconstruction and imputation

Instead of encoding the available data observations into multiple view-specific latent representations, the proposed IMDC-DIA method assumes that the data latent structure is intrinsic and corresponding latent representations should be unique to all data views. Denoting the latent representations to be $\{\mathbf{z}_i\}_{i=1}^{N}$ where $\mathbf{z}_i \in \mathbb{R}^D$, it proposes to reconstruct the multi-view data observations by projecting the latent representations into feature spaces with V independent neural networks as

$$\hat{\mathbf{x}}_i^{(v)} = g_{\Theta_v}(\mathbf{z}_i), \tag{1}$$

where Θ_v represents parameters of the v-th neural network. It is worth to note that the latent representation $\mathbf{z}_i$ is not fixed but a variable which can be optimized in model training. Correspondingly, we can obtain the reconstruction loss by minimizing the differences between the generated data and the available data observations via

$$L_r = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{V_i}\sum_{v=1}^{V}\mathbf{a}_{i,v}\|\hat{\mathbf{x}}_i^{(v)} - \mathbf{x}_i^{(v)}\|_2^2, \tag{2}$$

in which V_i refers to the number of available views of i-th data sample. Nevertheless, it can be observed in aforementioned reconstruction process that not only the available data observations are reconstructed, but also the missing ones are imputed accordingly. Therefore, the multi-view data can be completed by filling the missing with those imputed in Eq. (1). As a result, the complete data can be formulated as

$$\bar{\mathbf{x}}_i^{(v)} = \begin{cases} \mathbf{x}_i^{(v)} & \text{if} \quad \mathbf{a}_{i,v} = 1, \\ \hat{\mathbf{x}}_i^{(v)} & \text{if} \quad \mathbf{a}_{i,v} = 0. \end{cases} \tag{3}$$

3.1.2 Data alignment

On the basis of the complete multi-view data of Eq. (3), the proposed IMDC-DIA method also incorporate the pair-wise similarities of all data observations, especially those of the missing, to regularize the latent representations. Primarily, the *Linear Alignment* of two matrices is defined in Definition 1.

Definition 1 (Linear Alignment) *Given two arbitrary matrices* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *of size* $n \times d_1$ *and* $n \times d_2$ *respectively, the linear alignment is computed as*

$$LA = \frac{L_F^2(\mathbf{X}_1^\top \mathbf{X}_2)}{L_F(\mathbf{X}_1^\top \mathbf{X}_1) \cdot L_F(\mathbf{X}_2^\top \mathbf{X}_2)}, \tag{4}$$

in which $L_F(\cdot)$ *denotes the Frobenius norm of matrix. Corresponding computation complexity is linear to* n*.*

Denoting vector $\mathbf{x}_i$ to the i-th row of an arbitrary matrix $\mathbf{X}$, the **pair-wise linear similarity** between its i-th and j-th rows is measured by their dot-products that

$$k_{i,j} = \mathbf{x}_i \mathbf{x}_j^\top, \quad w.r.t.\ i, j \in \{1, 2, \cdots, N\} \tag{5}$$

With considering two arbitrary matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ coherently, their pair-wise linear similarities should be

$$k_{i,j}^1 = \mathbf{x}_i^1 \mathbf{x}_j^{1\top} \quad \text{and} \quad k_{i,j}^2 = \mathbf{x}_i^2 \mathbf{x}_j^{2\top} \tag{6}$$

On this basis, the similarity **consistency** of these two matrices can be measured by the sum of their dot-products as

$$c = \frac{\sum_{i,j=1}^N k_{i,j}^1 k_{i,j}^2}{\sqrt{\sum_{i,j=1}^N k_{i,j}^1 k_{i,j}^1} \sqrt{\sum_{i,j=1}^N k_{i,j}^2 k_{i,j}^2}} \tag{7}$$

where the denominator is a scaler term to ensure the obtained consistency in range $[-1, 1]$.

Theorem 1 *The linear alignment of two arbitrary matrices measures the consistency of the pair-wise linear similarities of their rows.*

Proof 1 *The consistency of Eq. (7) can be transformed to*

$$\begin{aligned} c &= \frac{\sum_{i,j=1}^N k_{i,j}^1 k_{i,j}^2}{\sqrt{\sum_{i,j=1}^N k_{i,j}^1 k_{i,j}^1} \sqrt{\sum_{i,j=1}^N k_{i,j}^2 k_{i,j}^2}} = \frac{\sum_{i,j=1}^N k_{i,j}^1 k_{j,i}^2}{\sqrt{\sum_{i,j=1}^N k_{i,j}^1 k_{i,j}^1} \sqrt{\sum_{i,j=1}^N k_{i,j}^2 k_{i,j}^2}} \\ &= \frac{< \mathbf{K}_1, \mathbf{K}_2 >_F}{\sqrt{< \mathbf{K}_1, \mathbf{K}_1 >_F < \mathbf{K}_2, \mathbf{K}_2 >_F}} = \frac{\mathrm{Tr}[(\mathbf{X}_1 \mathbf{X}_1^\top)(\mathbf{X}_2 \mathbf{X}_2^\top)]}{\sqrt{\mathrm{Tr}[(\mathbf{X}_1 \mathbf{X}_1^\top)(\mathbf{X}_1 \mathbf{X}_1^\top)] \cdot \mathrm{Tr}[(\mathbf{X}_2 \mathbf{X}_2^\top)(\mathbf{X}_2 \mathbf{X}_2^\top)]}} \\ &= \frac{\mathrm{Tr}[(\mathbf{X}_1^\top \mathbf{X}_2)^\top (\mathbf{X}_1^\top \mathbf{X}_2)]}{\sqrt{\mathrm{Tr}[(\mathbf{X}_1^\top \mathbf{X}_1)(\mathbf{X}_1^\top \mathbf{X}_1)] \cdot \mathrm{Tr}[(\mathbf{X}_2^\top \mathbf{X}_2)(\mathbf{X}_2^\top \mathbf{X}_2)]}} = \frac{\|\mathbf{X}_1^\top \mathbf{X}_2\|_F^2}{\|\mathbf{X}_1^\top \mathbf{X}_1\|_F \cdot \|\mathbf{X}_2^\top \mathbf{X}_2\|_F} = LA, \end{aligned} \tag{8}$$

in which the second equation holds for

$$k_{j,i}^2 = \mathbf{x}_{2,j} \mathbf{x}_{2,i}^\top = \mathbf{x}_{2,i} \mathbf{x}_{2,j}^\top = k_{i,j}^2, \tag{9}$$

while the third holds for

$$\begin{aligned} \mathbf{K}_1 &= \begin{bmatrix} \mathbf{x}_{1,1}\mathbf{x}_{1,1}^\top & \mathbf{x}_{1,1}\mathbf{x}_{1,2}^\top & \cdots & \mathbf{x}_{1,1}\mathbf{x}_{1,N}^\top \\ \mathbf{x}_{1,2}\mathbf{x}_{1,1}^\top & \mathbf{x}_{1,2}\mathbf{x}_{1,2}^\top & \cdots & \mathbf{x}_{1,2}\mathbf{x}_{1,N}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{1,N}\mathbf{x}_{1,1}^\top & \mathbf{x}_{1,N}\mathbf{x}_{1,2}^\top & \cdots & \mathbf{x}_{1,N}\mathbf{x}_{1,N}^\top \end{bmatrix} = \begin{bmatrix} k_{1,1}^1 & k_{1,2}^1 & \cdots & k_{1,3}^1 \\ k_{2,2}^1 & k_{2,2}^1 & \cdots & k_{2,3}^1 \\ \cdots & \cdots & \ddots & \cdots \\ k_{N,1}^1 & k_{N,2}^1 & \cdots & k_{N,N}^1 \end{bmatrix} \text{ and} \\ \mathbf{K}_2 &= \begin{bmatrix} \mathbf{x}_{2,1}\mathbf{x}_{2,1}^\top & \mathbf{x}_{2,1}\mathbf{x}_{2,2}^\top & \cdots & \mathbf{x}_{2,1}\mathbf{x}_{2,N}^\top \\ \mathbf{x}_{2,2}\mathbf{x}_{2,1}^\top & \mathbf{x}_{2,2}\mathbf{x}_{2,2}^\top & \cdots & \mathbf{x}_{2,2}\mathbf{x}_{2,N}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{2,N}\mathbf{x}_{2,1}^\top & \mathbf{x}_{2,N}\mathbf{x}_{2,2}^\top & \cdots & \mathbf{x}_{2,N}\mathbf{x}_{2,N}^\top \end{bmatrix} = \begin{bmatrix} k_{1,1}^2 & k_{1,2}^2 & \cdots & k_{1,3}^2 \\ k_{2,2}^2 & k_{2,2}^2 & \cdots & k_{2,3}^2 \\ \cdots & \cdots & \ddots & \cdots \\ k_{N,1}^2 & k_{N,2}^2 & \cdots & k_{N,N}^2 \end{bmatrix} \end{aligned} \tag{10}$$

This completes the proof.

According to Theorem 1, the linear alignment of two arbitrary matrices measures the consistency of the pair-wise linear similarities of their rows. To consider the pair-wise similarities of all data observations, especially including those of the missing ones, we propose to maximize the linear alignment between the latent representations and the completed data observations of v-th view, i.e., $\max LA(\mathbf{Z}, \bar{\mathbf{X}}^{(v)})$ Therefore, corresponding loss can be formulated to

$$L_a^{(v)} = -LA(\mathbf{Z}, \bar{\mathbf{X}}^{(v)}) = -\frac{L_F^2(\mathbf{Z}^\top \bar{\mathbf{X}}^{(v)})}{L_F(\mathbf{Z}^\top \mathbf{Z}) \cdot L_F(\bar{\mathbf{X}}^{(v)\top} \bar{\mathbf{X}}^{(v)})}, \tag{11}$$

in which $\mathbf{Z}$ and $\bar{\mathbf{X}}^{(v)}$ are the matrix form of $\{\mathbf{z}_i\}_{i=1}^N$ and $\{\bar{\mathbf{x}}_i^{(v)}\}_{i=1}^N$, respectively. Furthermore, to consider the data quality of each view, an additional set of weight parameters $\{w_v\}_{v=1}^V$ are introduced to balance their effect on latent representations. As a consequence, the overall data alignment loss can be written to

$$L_a = \sum_{v=1}^{V} w_v L_a^{(v)}, \quad s.t. \sum_{v=1}^{V} w_v^2 = 1. \tag{12}$$

It is worth to note that $\{w_v\}_{v=1}^V$ are variables to be optimized in model learning rather than hyper-parameters specified in advance.

3.1.3 Loss function

As seen in the framework of Fig. 1, the proposed IMDC-DIA method is composed of three main components, including the data reconstruction, imputation and alignment. By utilizing the reconstruction loss of Eq. (1), missing data imputation of Eq. (3) and data alignment loss of Eq. (12), the overall loss function is written to

$$L = L_r + \beta L_a, \quad s.t. \sum_{v=1}^{V} w_v^2 = 1 \tag{13}$$

in which β is a trade-off parameter and should be specified before model training.

3.2 Optimization

According to the overall loss of Eq. (13), there are three variables to optimize in model training, including the neural network parameter $\{\Theta_v\}_{v=1}^V$, the unique latent representation $\{\mathbf{z}_i\}_{i=1}^N$ and weight parameter $\{w_v\}_{v=1}^V$. Each of them can be optimized in the following.

Optimization of neural network parameter $\{\Theta_v\}_{v=1}^V$. Same to most of deep learning methods, the neural network parameter can be optimized with gradient descent strategy. In experiment, we adopt the popular Adaptive Moment Estimation (Adam) optimizer.

Optimization of latent representation $\{\mathbf{z}_i\}_{i=1}^N$. Different from the existing multi-view clustering methods, such as [13], the proposed IMDC-DIA method is difficult to find the close-form solution of data latent representation. Therefore, the gradient descent strategy with Adam optimizer is adopted in its optimization.

Optimization of weight parameter $\{w_v\}_{v=1}^V$. With fixing the others, $L_a^{(v)}$ is given and minimizing the overall loss of Eq. (13) equals to

$$\max_{\{w_v\}_{v=1}^V} \sum_{v=1}^{V} w_v(-L_a^{(v)}), \quad s.t. \sum_{v=1}^{V} w_v^2 = 1. \tag{14}$$

According to o Cauchy–Schwarz inequality [36],

$$\sum_{v=1}^{V} w_v(-L_a^{(v)}) \leq \sqrt{\left(\sum_{v=1}^{V} w_v^2\right)\left(\sum_{v=1}^{V} L_a^{(v)2}\right)} = \sqrt{\sum_{v=1}^{V} L_a^{(v)2}}, \tag{15}$$

where the equality holds for when

$$\frac{w_1}{-L_a^{(1)}} = \frac{w_2}{-L_a^{(2)}} = \cdots = \frac{w_V}{-L_a^{(V)}}. \tag{16}$$

Unifying the constraint $\sum_{v=1}^{V} w_v^2 = 1$ in Eq. (14), the solution can be computed to

$$w_v^* = -L_a^{(v)} / \sqrt{\sum_{v'=1}^{V} L_a^{(v')2}}. \tag{17}$$

Overall, the aforementioned three parameters are optimized alternately and the latent representations can be achieved until the convergence or maximal epoch. Next, k-means algorithm is applied on the computed latent representations $\{\mathbf{z}_i\}_{i=1}^N$ to group multi-view data into categories. Furthermore, the pseudo-code of the optimization procedure is summarized in Alg. 1.

Algorithm 1 Incomplete Multi-view Deep Clustering with Data Imputation and Alignment

Input: incomplete multi-view data $\{\mathbf{x}_i^{(v)}\}_{i,v=1}^{N,V}$ with availability indicator $\{\mathbf{a}_i\}_{i=1}^N$, cluster number k
Output: data labels

initialize latent representation $\{\mathbf{z}_i\}_{i=1}^N$ and network parameters $\{\Theta_v\}_{v=1}^V$;
$t = 0$;
while $t < epochs$ **do**
  # forward
  compute L_r with Eq. (2);
  complete multi-view data with Eq. (3);
  compute $\{L_a^{(v)}\}_{v=1}^V$ with Eq. (11);
  update the view weights $\{w_v\}_{v=1}^V$ with Eq. (17);
  compute the overall loss L with Eq. (13);
  # back propagation
  update latent representation $\{\mathbf{z}_i\}_{i=1}^N$ and network parameters $\{\Theta_v\}_{v=1}^V$;
  $t = t + 1$;
end while
compute the data labels with k-means on latent representations $\{\mathbf{z}_i\}_{i=1}^N$;

3.3 Computation complexity

Specifically, the data reconstruction and imputation of Eq. (1), (2) and (3) only require projecting the latent representations into feature spaces with parameterized neural networks, hence introducing a $\mathcal{O}(N)$ complexity. Also, corresponding updates on latent representations and network parameters are of $\mathcal{O}(N)$ complexity. Besides, the complexity of data alignment loss $L_a^{(v)}$ in Eq. (12) is mainly on the computation of $L_F^2(\mathbf{Z}^\top \bar{\mathbf{X}}^{(v)})$, $L_F(\mathbf{Z}^\top \mathbf{Z})$ and $L_F(\bar{\mathbf{X}}^{(v)\top} \bar{\mathbf{X}}^{(v)})$ which are all of $\mathcal{O}(N)$ complexity. Nevertheless, the optimization of weight parameter $\{w_v\}_{v=1}^V$ via Eq. (17) is of $\mathcal{O}(V)$ complexity. In summary, the overall complexity of the proposed IMDC-DIA method is linear to the number of data samples, i.e., $\mathcal{O}(N)$.

4 Experiment

4.1 Experiment setting

To validate the proposed IMDC-DIA method, we conduct extensive experiments on four benchmark datasets, including HandWritten[2] [37], Caltech5V[3] [38], Flower17[4] [39] and MSRCV1[5] [40]. On the basis, we generate incomplete datasets by following the common strategy [41] in literature. Specifically, assuming the missing ratio to be m, m percent of data samples are selected to remove at least one views randomly. In the following experiments, m is set in $\{0.1, 0.3, 0.5, 0.7.0.9\}$. Meanwhile, five recent incomplete multi-view deep clustering approaches are considered in comparison, including DITA-IMVC [42], DSIMVC [41], DCP [34], DVIMVC [30] and CPSPAN [43]. Note that, we

[2] `https://archive.ics.uci.edu/ml/datasets/Multiple+Features`
[3] `https://data.caltech.edu/records/mzrjq-6wc02`
[4] `https://www.robots.ox.ac.uk/~vgg/data/flowers/17`
[5] `https://github.com/youweiliang/Multi-view_Clustering`

Table 1: Performance comparison between the proposed IMDC-DIA and recent incomplete deep multi-view clustering approaches. The *avg.* column refers to performance averages of all missing ratios. Note that, the best results are marked in bold, while the second-best with underline.

Metric		ACC						NMI					
Missing ratio		0.1	0.3	0.5	0.7	0.9	avg.	0.1	0.3	0.5	0.7	0.9	avg.
HandWritten	DITA-IMVC	75.48	78.92	81.37	81.02	55.00	74.36	75.81	78.05	77.62	75.54	52.16	71.84
	DSIMVC	79.55	80.73	78.83	77.48	50.87	73.49	78.57	77.76	74.93	71.97	48.53	70.35
	DCP	81.95	75.73	77.23	71.77	13.07	63.95	84.37	78.95	79.22	74.39	0.93	63.57
	DVIMC	86.42	83.05	45.48	25.20	18.92	51.81	87.64	84.78	56.58	35.43	17.96	56.48
	CPSPAN	90.42	91.25	91.08	90.27	86.73	89.95	83.84	84.37	84.01	83.55	82.62	83.68
	IMDC-DIA	**96.37**	**93.68**	**91.80**	**90.65**	**87.93**	**92.09**	**91.80**	**93.68**	**91.80**	**90.65**	**87.93**	**91.17**
Caltech5V	DITA-IMVC	79.10	75.76	68.02	58.90	40.29	64.41	67.56	64.78	58.84	52.40	29.29	54.57
	DSIMVC	76.64	73.14	66.36	57.38	44.95	63.69	68.02	63.11	56.71	49.23	33.99	54.21
	DCP	44.50	46.95	46.45	44.67	16.98	39.91	45.22	52.31	50.69	45.07	0.54	38.77
	DVIMC	**88.64**	81.36	**84.93**	80.86	68.81	80.92	80.52	73.38	76.27	73.56	62.40	73.23
	CPSPAN	83.29	81.10	77.21	76.79	79.02	79.48	73.81	71.30	68.73	68.43	69.26	70.31
	IMDC-DIA	86.05	**85.05**	83.33	**85.02**	**81.29**	**84.15**	**86.05**	**85.05**	**83.33**	**85.02**	**81.29**	**84.15**
Flower17	DITA-IMVC	18.75	17.52	13.77	13.16	12.84	15.21	18.15	16.60	10.23	8.65	8.66	12.46
	DSIMVC	18.63	16.86	13.50	13.36	13.36	15.14	17.51	14.78	9.29	9.27	9.22	12.01
	DCP	26.15	25.49	22.21	15.05	10.78	19.94	26.18	25.94	23.47	15.37	3.69	18.93
	DVIMC	36.69	31.23	26.84	23.41	21.69	27.97	35.43	29.82	24.30	20.93	20.08	26.11
	CPSPAN	36.27	39.34	31.15	37.35	36.00	36.02	37.13	39.65	30.88	38.15	35.97	36.36
	IMDC-DIA	**52.03**	**46.47**	**44.22**	**41.54**	**39.14**	**44.68**	**52.03**	**46.47**	**44.22**	**41.54**	**39.14**	**44.68**
MSRCV1	DITA-IMVC	76.03	74.60	70.95	66.83	55.87	68.86	67.62	63.02	60.61	54.96	43.32	57.91
	DSIMVC	78.57	76.98	70.48	71.59	64.76	72.48	68.37	65.81	62.26	59.19	51.78	61.48
	DCP	17.94	20.00	22.06	22.22	22.70	20.98	7.11	7.54	7.05	4.95	4.62	6.25
	DVIMC	74.60	61.43	56.67	43.17	38.25	54.82	66.68	56.40	54.91	42.02	30.08	50.02
	CPSPAN	85.08	86.51	85.40	84.29	77.62	83.78	75.86	77.25	75.51	74.09	69.06	74.35
	IMDC-DIA	**91.27**	**92.22**	**86.03**	**85.08**	**83.02**	**87.52**	**91.27**	**92.22**	**86.03**	**85.08**	**83.02**	**87.52**

directly adopt the public codes available on the authors' websites without modification. To ensure the comparison fairness, we grid-search their parameters and report the best. So does the proposed IMDC-DIA[6] method with setting its only parameter β in $\{0.01, 0.1, 1, 10, 100\}$. By following the literature, four most common metrics are adopted to measure the clustering performance, including accuracy (ACC), normalized mutual information (NMI), purity (PUR) and adjusted rand index (ARI). Nevertheless, all methods are executed multiple times and their averages are reported to remove randomness effect.

4.2 Effectiveness

In order to validate effectiveness of the proposed IMDC-DIA method, we compare it with the recent advances on incomplete deep multi-view clustering in literature. Corresponding results can be found in Table 1. It is obvious that the proposed IMDC-DIA outperforms the recent advances in almost all missing ratios on all datasets. Concretely, it achieves the improvements of 5.95%, 2.43%, 0.72%, 0.38%, 1.20% in accuracy and 4.16%, 8.90%, 7.79%, 7.10%, 5.31% in NMI on HandWritten; 15.34%, 7.13%, 13.07%, 4.19%, 3.14% in accuracy and 14.90%, 6.82%, 13.34%, 3.39%, 3.17% in NMI on Flower17; 6.19%, 5.71%, 0.63%, 0.79%, 5.40% in accuracy and 15.41%, 14.97%, 10.52%, 10.99%, 13.96% in NMI, respectively. Although slight decreases in 0.1 and 0.3 missing ratios on Caltech5V are observed, i.e., 2.59% and 1.60%, the proposed IMDC-DIA obtains better performances on the other missing ratios consistently. Nevertheless, with averaging the results in all missing ratios, we can see that it improves the accuracy by 2.14%, 2.27%, 3.14%, 3.74% and the NMI by 7.49%, 10.92%, 8.32%, 13.17% on the four benchmark datasets. To be summarized, the aforementioned observations

[6]The code is available at `https://github.com/liujiyuan13/IMDC-DIA-code_release`.

Table 2: Performance comparison between the proposed IMDC-DIA method and its variants of modifying and removing the data imputation and alignment components. The *avg.* column refers to performance averages of all missing ratios, while the *avg.* row reports the performances in the setting of completing missing data with average values. Note that, the best results are marked in bold, while the second-best with underline.

Metric			ACC						NMI					
Missing ratio			0.1	0.3	0.5	0.7	0.9	avg.	0.1	0.3	0.5	0.7	0.9	avg.
HandWritten	w/o imp.	avg.	75.58	66.17	56.47	50.63	41.78	58.13	75.30	64.92	55.95	51.55	42.18	57.98
		zero	83.85	68.87	61.33	48.23	44.72	61.40	76.90	64.85	56.41	49.60	42.58	58.07
		rand	79.98	69.73	58.37	55.05	44.33	61.49	78.33	68.31	57.06	52.81	42.45	59.79
		none	96.13	82.55	87.82	78.70	55.83	80.21	91.37	82.39	82.03	74.80	49.31	75.98
	w/o align.		93.62	86.65	90.57	81.27	82.78	86.98	87.78	81.89	81.82	76.88	73.44	80.36
	IMDC-DIA		**96.37**	**93.68**	**91.80**	**90.65**	**87.93**	**92.09**	**91.80**	**86.68**	**83.31**	**82.05**	**77.39**	**84.25**
Caltech5V	w/o imp.	avg.	74.14	73.19	63.86	52.07	46.98	62.05	72.26	66.11	56.68	47.16	39.89	56.42
		zero	80.74	73.60	67.38	61.81	52.12	67.13	74.06	63.71	57.33	52.74	43.37	58.24
		rand	78.76	68.31	62.05	58.10	47.95	63.03	72.70	60.89	53.63	50.37	37.17	54.95
		none	**87.98**	82.14	81.71	76.31	61.74	77.98	**78.57**	72.57	70.35	62.81	46.61	66.18
	w/o align.		80.21	81.90	80.07	83.50	**81.60**	81.46	73.42	72.84	71.04	72.08	67.51	71.38
	IMDC-DIA		86.05	**85.05**	**83.33**	**85.02**	81.29	**84.15**	76.39	**75.28**	**72.22**	**72.28**	**68.14**	**72.86**
Flower17	w/o imp.	avg.	47.23	39.80	32.70	29.78	25.02	34.91	46.18	38.56	33.21	27.93	24.29	34.03
		zero	45.15	36.86	31.81	27.38	24.63	33.17	44.51	34.58	29.05	25.27	22.77	31.24
		rand	45.64	35.20	30.15	25.98	23.75	32.14	45.17	33.54	27.65	23.60	20.75	30.14
		none	42.01	35.93	25.64	13.11	17.23	26.78	42.74	34.80	22.81	7.58	11.43	23.87
	w/o align.		38.28	36.18	35.54	30.81	28.50	33.86	35.90	35.13	31.87	27.53	24.44	30.97
	IMDC-DIA		**51.13**	**45.00**	**42.89**	**40.69**	**37.06**	**43.35**	**47.15**	**45.02**	**41.33**	**38.42**	**34.81**	**41.35**
MRSCV1	w/o imp.	avg.	72.70	57.46	51.27	42.86	36.67	52.19	64.86	51.51	43.94	38.28	30.14	45.75
		zero	68.10	62.86	56.83	45.56	42.38	55.14	59.26	46.69	46.66	36.53	30.58	43.94
		rand	66.35	55.08	48.89	44.60	43.17	51.62	59.30	44.20	40.67	32.61	31.49	41.65
		none	78.89	62.06	58.25	31.90	39.84	54.19	75.10	58.82	50.66	17.53	23.40	45.10
	w/o align.		85.56	85.24	73.65	67.30	60.95	74.54	78.09	76.37	60.14	52.43	48.95	63.20
	IMDC-DIA		**91.27**	**92.22**	**86.03**	**85.08**	**83.02**	**87.52**	**84.37**	**85.63**	**77.84**	**74.21**	**70.47**	**78.51**

well illustrate the effectiveness of the proposed IMDC-DIA method and its superiority over the recent advances in literature.

4.3 Ablation study

To demonstrate the rationality of our motivations and effectiveness of the proposed techniques, we further conduct an thorough ablation study in the following. Two settings are designed by modifying and removing the data imputation and alignment components. Specifically,

1. The *avg.*, *zero* and *rand* of *w/o imp.* substitute the data imputation into completing the missing data with averages of available data observations, zeros and random values, respectively. Meanwhile, the *none* of *w/o imp.* refers to only using the available data observations in data alignment module rather than completing the missing.
2. *w/o align.* removes the data alignment module, which can be easily implemented by setting parameter β to 0.

By comparing results of the proposed IMDC-DIA method and its *w/o imp.* settings, the former achieves better clustering performances in almost missing ratios on all benchmark datasets. Although slight decreases in missing ratio 0.1 on Caltech5V are observed, i.e., 1.93% accuracy and 2.18% NMI, the proposed IMDC-DIA obtains better performances on the other missing ratios, i.e., 2.91%, 1.62%, 8.71%, 19.55% accuracy and 2.71%, 1.87%, 9.47% 21.53% NMI. In average, it improve the accuracy by 11.88%, 6.17%, 8.44%, 32.38% and the NMI by 8.27%, 6.68%, 7.32%, 32.76%. These observations well validate effectiveness of the proposed data imputation module. In addition, we

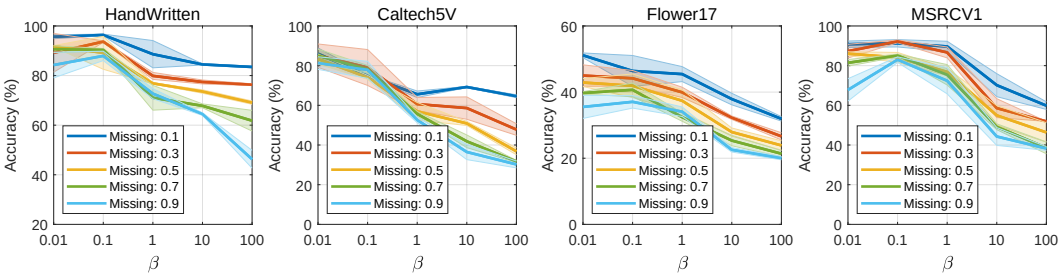

Figure 2: Accuracy variation with respect to parameter β in different missing ratios on the four benchmark datasets. The shaded area represents the accuracy variance correspondingly.

find the proposed IMDC-DIA exhibits larger and larger performance improvements when increasing the missing ratio. Also, the results in *none* setting are better than those in *avg.*, *zero* and *rand* settings in most cases. The two observations indicate that complementing the missing data without data-irrelevant values would introduce extra bias hence decreasing the performance.

By comparing results of the proposed IMDC-DIA method and its *w/o align.* settings, similar observations can be obtained. Specifically, 0.31% accuracy decrease is presented in missing ratio 0.9 on Caltech5V, while different degrees of improvements are observed in the other missing ratios, i.e., 5.84%, 3.15%, 3.26%, 1.52% accuracy and 2.97%, 2.44%, 1.18%, 0.20% NMI. Also, with averaging the results in all missing ratios, the proposed IMDC-DIA method achieves better accuracies by 5.11%, 2.69%, 9.49% 12.98% and NMIs by 3.89%, 1.48%, 10.38%, 15.31%, respectively. These well demonstrate that the defined linear alignment is conducive to the performance improvement.

Globally, the proposed IMDC-DIA method achieves better clustering performances than its *none* of *w/o imp.* setting, i.e., 11.88%, 6.17%, 16.57%, 33.33% accuracies and 8.27%, 6.68%, 17.48%, 33.41% NMIs. This proves that incorporating the data similarities among missing data observations can largely enhance the learning of latent representations, hence obtaining better clustering results.

4.4 Parameter analysis

In the proposed IMDC-DIA method, the only parameter is the trade-off β between the data reconstruction loss and data alignment loss. To investigate its effect on performance, we collect the clustering accuracies when β set in range $\{0.01, 0.1, 1, 10, 100\}$ and present their average curves and corresponding variances in Fig. 2. In general, the clustering accuracy decreases with a larger β in almost all missing ratios on all benchmark datasets. Specifically, it can be seen that the best accuracies are obtained mostly when setting β to 0.01 and sometimes when setting β to 0.1. Therefore, we recommend to set it from 0.01 to 0.1 in first priority.

5 Conclusion

In literature, existing incomplete multi-view deep clustering approaches overlook the fact that the latent representations should be unique for a specific set of data samples. Also, they fail to utilize the pair-wise similarities of missing data observations sufficiently. Therefore, this paper proposes an incomplete multi-view deep clustering method with data imputation and alignment by assuming each data sample corresponds to a same latent representation among all views. Insides, a novel linear alignment measurement of linear complexity is defined and incorporated to integrate the pair-wise similarities of all data observations, including those of the missing ones. Nevertheless, an alternate optimization strategy is proposed to learn high-quality latent representations for the next clustering task. In experiment, the proposed method is compared with recent advances, while its motivations and method design are validated. Also, its parameter effect is explored empirically.

Although the proposed method achieves state-of-the-art results compared to recent advances, it may be partially limited by lacking sufficient constraints on the latent representations, preventing from further improvement of the clustering performance. In the future work, we will explore more feasible constraints. Meanwhile, the evaluation on missing data imputation is also a promising research direction on the basis of this work.

Acknowledgments and Disclosure of Funding

This work is supported by the National Natural Science Foundation of China (No. 62306324, 62376279, U24A20333, 62325604, 62276271, 62421002), the Science and Technology Innovation Program of Hunan Province (No. 2024RC3128) and the National University of Defense Technology Research Foundation (No. ZK24-30).

References

[1] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th International Conference on World Wide Web, WWW 2015, Florence, Italy, May 18-22, 2015*, pages 278–288. ACM, 2015.

[2] Xiaobo Zhang, Yan Yang, Tianrui Li, Yiling Zhang, Hao Wang, and Hamido Fujita. CMC: A consensus multi-view clustering model for predicting alzheimer's disease progression. *Comput. Methods Programs Biomed.*, 199:105895, 2021.

[3] Jiyuan Liu, Xinwang Liu, Siqi Wang, Xinhang Wan, Dongsheng Li, Kai Lu, and Kunlun He. Communication-efficient federated multi-view clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2025.

[4] Zhenwen Ren, Simon X. Yang, Quansen Sun, and Tao Wang. Consensus affinity graph learning for multiple kernel clustering. *IEEE Trans. Cybern.*, 51(6):3273–3284, 2021.

[5] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Qing Liao, and Yuanqing Xia. Contrastive multi-view kernel learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(8):9552–9566, 2023.

[6] Jiyuan Liu, Xinwang Liu, Jian Xiong, Qing Liao, Sihang Zhou, Siwei Wang, and Yuexiang Yang. Optimal neighborhood multiple kernel clustering with adaptive local kernels. *IEEE Trans. Knowl. Data Eng.*, 34(6):2872–2885, 2022.

[7] Tao Zhou, Changqing Zhang, Xi Peng, Harish Bhaskar, and Jie Yang. Dual shared-specific multiview subspace clustering. *IEEE Trans. Cybern.*, 50(8):3517–3530, 2020.

[8] Shudong Huang, Hongjie Wu, Yazhou Ren, Ivor W. Tsang, Zenglin Xu, Wentao Feng, and Jiancheng Lv. Multi-view subspace clustering on topological manifold. In *Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

[9] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Xifeng Guo, Marius Kloft, and Liangzhong He. Multiview subspace clustering via co-training robust data representation. *IEEE Trans. Neural Networks Learn. Syst.*, 33(10):5177–5189, 2022.

[10] Jie Wen, Yong Xu, and Hong Liu. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE Trans. Cybern.*, 50(4):1418–1429, 2020.

[11] Jie Wen, Ke Yan, Zheng Zhang, Yong Xu, Junqian Wang, Lunke Fei, and Bob Zhang. Adaptive graph completion based incomplete multi-view clustering. *IEEE Trans. Multim.*, 23:2493–2504, 2021.

[12] Sheng Huang, Yunhe Zhang, Lele Fu, and Shiping Wang. Learnable multi-view matrix factorization with graph embedding and flexible loss. *IEEE Trans. Multim.*, 25:3259–3272, 2023.

[13] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Li Liu, Siqi Wang, Weixuan Liang, and Jiangyong Shi. One-pass multi-view clustering for large-scale data. In *IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pages 12324–12333. IEEE, 2021.

[14] Zongmo Huang, Yazhou Ren, Xiaorong Pu, Shudong Huang, Zenglin Xu, and Lifang He. Self-supervised graph attention networks for deep weighted multi-view clustering. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 7936–7943. AAAI Press, 2023.

[15] Qianqian Wang, Jiafeng Cheng, Quanxue Gao, Guoshuai Zhao, and Licheng Jiao. Deep multi-view subspace clustering with unified and discriminative learning. *IEEE Trans. Multim.*, 23:3483–3493, 2021.

[16] Jinrong Cui, Yuting Li, Han Huang, and Jie Wen. Dual contrast-driven deep multi-view clustering. *IEEE Trans. Image Process.*, 33:4753–4764, 2024.

[17] Xuejiao Yu, Yi Jiang, Guoqing Chao, and Dianhui Chu. Deep contrastive multi-view subspace clustering with representation and cluster interactive learning. *IEEE Trans. Knowl. Data Eng.*, 37(1):188–199, 2025.

[18] Jie Chen, Hua Mao, Wai Lok Woo, and Xi Peng. Deep multiview clustering by contrasting cluster assignments. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pages 16706–16715. IEEE, 2023.

[19] Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China*, pages 3911–3919. ijcai.org, 2023.

[20] Yiding Lu, Yijie Lin, Mouxing Yang, Dezhong Peng, Peng Hu, and Xi Peng. Decoupled contrastive multi-view clustering with high-order random walks. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, February 20-27, 2024, Vancouver, Canada*, pages 14193–14201. AAAI Press, 2024.

[21] Guoqing Chao, Yi Jiang, and Dianhui Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, February 20-27, 2024, Vancouver, Canada*, pages 11221–11229. AAAI Press, 2024.

[22] Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. COMPLETER: incomplete multi-view clustering via contrastive prediction. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pages 11174–11183. Computer Vision Foundation / IEEE, 2021.

[23] Shizhe Hu, Chengkun Zhang, Guoliang Zou, Zhengzheng Lou, and Yangdong Ye. Deep multiview clustering by pseudo-label guided contrastive learning and dual correlation learning. *IEEE Trans. Neural Networks Learn. Syst.*, 36(2):3646–3658, 2025.

[24] Pengfei Zhu, Xinjie Yao, Yu Wang, Binyuan Hui, Dawei Du, and Qinghua Hu. Multiview deep subspace clustering networks. *IEEE Transactions on Cybernetics*, 54(7):4280–4293, 2024.

[25] Zhiyuan Yang, Changming Zhu, and Zishi Li. Deep contrastive multi-view clustering with doubly enhanced commonality. *Multim. Syst.*, 30(4):196, 2024.

[26] Chenhang Cui, Yazhou Ren, Jingyu Pu, Xiaorong Pu, and Lifang He. Deep multi-view subspace clustering with anchor graph. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI '23, 2023.

[27] Jin Chen, Aiping Huang, Wei Gao, Yuzhen Niu, and Tiesong Zhao. Joint shared-and-specific information for deep multi-view clustering. *IEEE Trans. Circuits Syst. Video Technol.*, 33(12):7224–7235, 2023.

[28] Jie Xu, Chao Li, Yazhou Ren, Liang Peng, Yujie Mo, Xiaoshuang Shi, and Xiaofeng Zhu. Deep incomplete multi-view clustering via mining cluster complementarity. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Virtual Event, February 22 - March 1, 2022*, pages 8761–8769. AAAI Press, 2022.

[29] Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei, and Guo-Sen Xie. Cdimc-net: Cognitive deep incomplete multi-view clustering network. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 3230–3236. ijcai.org, 2020.

[30] Gehui Xu, Jie Wen, Chengliang Liu, Bing Hu, Yicheng Liu, Lunke Fei, and Wei Wang. Deep variational incomplete multi-view clustering: Exploring shared clustering structures. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, February 20-27, 2024, Vancouver, Canada*, pages 16147–16155. AAAI Press, 2024.

[31] Jie Xu, Chao Li, Liang Peng, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. *IEEE Trans. Image Process.*, 32:1354–1366, 2023.

[32] Zhangqi Jiang, Tingjin Luo, and Xinyan Liang. Deep incomplete multi-view learning network with insufficient label information. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, February 20-27, 2024, Vancouver, Canada*, pages 12919–12927. AAAI Press, 2024.

[33] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Trans. Image Process.*, 30:1771–1783, 2021.

[34] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(4):4447–4461, 2023.

[35] Chengliang Liu, Jie Wen, Zhihao Wu, Xiaoling Luo, Chao Huang, and Yong Xu. Information recovery-driven deep incomplete multiview clustering network. *IEEE Trans. Neural Networks Learn. Syst.*, 35(11):15442–15452, 2024.

[36] Tadashi Tokieda. A viscosity proof of the cauchy-schwarz inequality. *Am. Math. Mon.*, 122(8):781, 2015.

[37] Martijn van Breukelen, Robert P. W. Duin, David M. J. Tax, and J. E. den Hartog. Handwritten digit recognition by combined classifiers. *Kybernetika*, 34(4):381–386, 1998.

[38] Li Fei-Fei, Robert Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Comput. Vis. Image Underst.*, 106(1):59–70, 2007.

[39] Maria-Elena Nilsback and Andrew Zisserman. A visual vocabulary for flower classification. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), 17-22 June 2006, New York, NY, USA*, pages 1447–1454. IEEE Computer Society, 2006.

[40] John M. Winn and Nebojsa Jojic. LOCUS: learning object classes with unsupervised segmentation. In *10th IEEE International Conference on Computer Vision (ICCV 2005), 17-20 October 2005, Beijing, China*, pages 756–763. IEEE Computer Society, 2005.

[41] Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 21090–21110. PMLR, 2022.

[42] Weiqing Yan, Kanglong Liu, Wujie Zhou, and Chang Tang. Deep incomplete multi-view clustering via dynamic imputation and triple alignment with dual optimization. *IEEE Trans. Circuits Syst. Video Technol.*, 35(4):3250–3261, 2025.

[43] Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pages 11600–11609. IEEE, 2023.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: This paper focuses on the reseach on incomplete multi-view deep clustering algorithm. It propose a novel incomplete multi-view deep clustering with data imputation and alignment, effectively improving the clustering performance on incomplete multi-view data.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [NA]

   Justification: This paper has no obvious limitations.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Theorem 1 is proved in Proof 1.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The datasets are publicly avaiable and the proposed method is clearly described with equations.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The code is released at `https://github.com/liujiyuan13/IMDC-DIA-code_release`. Meanwhile, the used datasets are popular in literature and available publicly.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Since all data samples are fed at once as for clustering algorithm, there is no need to split them into training and test sets. Also, a parameter study is conducted.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: The error bars are reported in Fig. 2. In addition, Table 1 and 2 only reports the averages while ignore the variances due to space limit. They can be provided if required.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: They are reported in Section D.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This is doubly checked.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: This paper propose an incomplete multi-view deep clustering method and there is no obvious negative societal impact of the work performed.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Please see Section B.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: We do not use LLM at all.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

A Convergence analysis

In the optimization, the proposed IMDC-DIA method aims to minimize the loss L of Eq. (13). Insides, three variables are optimized alternately, including the neural network parameter $\{\Theta_v\}_{v=1}^V$, the unique latent representation $\{\mathbf{z}_i\}_{i=1}^N$ and weight parameter $\{w_v\}_{v=1}^V$. With a little abuse of notation, let the loss be represented by

$$L = \ell(\{\Theta_v\}_{v=1}^V, \{\mathbf{z}_i\}_{i=1}^N, \{w_v\}_{v=1}^V). \tag{18}$$

Taking the loss at t epoch to be

$$L^{(t)} = \ell(\{\Theta_v^{(t)}\}_{v=1}^V, \{\mathbf{z}_i^{(t)}\}_{i=1}^N, \{w_v^{(t)}\}_{v=1}^V), \tag{19}$$

By fixing $\{\mathbf{z}_i\}_{i=1}^N$ and $\{w_v\}_{v=1}^V$ at $\{\mathbf{z}_i^{(t)}\}_{i=1}^N$ and $\{w_v^{(t)}\}_{v=1}^V$, optimizing $\{\Theta_v\}_{v=1}^V$ at $t+1$ epoch induces to

$$\ell(\{\Theta_v^{(t)}\}_{v=1}^V, \{\mathbf{z}_i^{(t)}\}_{i=1}^N, \{w_v^{(t)}\}_{v=1}^V) \leq \ell(\{\Theta_v^{(t+1)}\}_{v=1}^V, \{\mathbf{z}_i^{(t)}\}_{i=1}^N, \{w_v^{(t)}\}_{v=1}^V). \tag{20}$$

So do the optimizations of $\{\mathbf{z}_i\}_{i=1}^N$ and $\{w_v\}_{v=1}^V$, resulting in

$$\begin{aligned} &\ell(\{\Theta_v^{(t)}\}_{v=1}^V, \{\mathbf{z}_i^{(t)}\}_{i=1}^N, \{w_v^{(t)}\}_{v=1}^V) \leq \ell(\{\Theta_v^{(t+1)}\}_{v=1}^V, \{\mathbf{z}_i^{(t)}\}_{i=1}^N, \{w_v^{(t)}\}_{v=1}^V) \\ &\leq \ell(\{\Theta_v^{(t+1)}\}_{v=1}^V, \{\mathbf{z}_i^{(t+1)}\}_{i=1}^N, \{w_v^{(t)}\}_{v=1}^V) \leq \ell(\{\Theta_v^{(t+1)}\}_{v=1}^V, \{\mathbf{z}_i^{(t+1)}\}_{i=1}^N, \{w_v^{(t+1)}\}_{v=1}^V), \end{aligned} \tag{21}$$

which can be simplified to

$$L^{(t)} \leq L^{(t+1)}. \tag{22}$$

Nevertheless,

$$L = L_r + \beta L_a \geq 0 + (-1) = -1. \tag{23}$$

To be summarized, Eq. (22) indicates that the loss monotonically decreases, while Eq. (23) illustrates that the loss is lower bounded. Therefore, the optimization algorithm is convergent.

B Dataset detail

The specifics of the used benchmark datasets are presented in Table 3. Also, they are briefly introduced as follows:

1. HandWritten[7] [37]. It consists of 2,000 digital images of ten classes from 0 to 9 in which there are 200 images in each class. Here, six features are used, i.e., 216-D profile correlations, 76-D Fourier coefficients of the character shapes, 64-D Karhunen-Love coefficients, 6-D morphological features, 240-D pixel averages in 2×3 windows and 47-D Zernike moments.
2. Caltech5V[8] [38]. It collects 1400 object images belonging to 7 classes. I experiment, five well-known conventional feature descriptors are adopted, including 48-D wavelet moments, 40-D CENTRIST, 254-D LBP, 512-D GIST and 928-D HOG, respectively.
3. Flower17[9] [39]. It collects flower images of 17 categories with 80 images for each class. Corresponding features are 5376-D Color Histogram, 512-D GIST, 5376-D HOG (2×2), 5376-D HOG (3×3), 1239-D LBP, 5376-D SIFT and 5376-D SSIM features.
4. MSRCV1[10] [40]. It collects 210 images of seven categories with each category provides 30 images. For each image, six features are extracted, including 1302-D CENTRIST, 48-D Color Moment, 512-D GIST, 100-D HOG, 256-D LBP and 210 SIFT, respectively.

For HandWritten, Flower17 and MSRCV1, the creators share them publicly but do not issue licenses, while Caltech5V is issued the *Creative Commons Attribution 4.0 International (CC BY 4.0)* license.

C Incomplete multi-view data setting

The incomplete multi-view data are generated by following the common strategy [41] in literature. Specifically, assuming the missing ratio to be m, m percent of data samples are selected to remove at least one views randomly. In the following experiments, m is set in $\{0.1, 0.3, 0.5, 0.7.0.9\}$. Meanwhile, the corresponding psuedo-code is outlined in Alg. 2.

[7] `https://archive.ics.uci.edu/ml/datasets/Multiple+Features`

[8] `https://data.caltech.edu/records/mzrjq-6wc02`

[9] `https://www.robots.ox.ac.uk/~vgg/data/flowers/17`

[10] `https://github.com/youweiliang/Multi-view_Clustering`

Table 3: Details of the used benchmark datasets. Note that, in brackets are the feature dimensions of each data view.

Dataset	Samples	Clusters	Number of Views (Dimensions)
HandWritten	2000	10	6 (216/76/64/6/240/47)
Caltech5V	1400	7	5 (48/40/254/512/928)
Flower17	1360	17	7 (512/5376/5376/1239/5376/5376)
MSRCV1	210	7	6 (1032/48/512/100/256/210)

Algorithm 2 Generation of Incomplete Multi-view Data

Input: incomplete multi-view data $\{\mathbf{x}_i^{(v)}\}_{i,v=1}^{N,V}$, missing ratio m

Output: availability indicator $\{\mathbf{a}_i\}_{i=1}^{N}$

obtain the numbers of data samples and views, i.e. N and V;
random sample incomplete data index $\mathcal{S}$ with missing ratio m;
initialize availability indicator $\mathbf{a}_i = \{1, 2, \cdots, V\}$ w.r.t. $i \in \{1, 2, \cdots, N\}$;
for $p \in \mathcal{S}$ **do**
    random sample missing index $\mathbf{m}_p \subset \{1, 2, \cdots, V\}$ and $\mathbf{m}_p = \emptyset$;
    set availability indicator $\mathbf{a}_p = \{1, 2, \cdots, V\} - \mathbf{m}_p$;
end for

D More experiment settings

In the experiments, we adopt fully-connected neural networks on all datasets where different numbers of neurons are adopted according to different feature dimensions, as shown in Table 4. Meanwhile, the unique data latent representation and parameters of neural netowrks are both optimized with gradient descent strategy with Adam optimizer whose learning rate is set to 0.001. Additionally, the codes of the proposed method and recent advances in comparison are executed on a server with 40 *Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz* CPUs and 2 *NVIDIA GeForce RTX 3090* GPUs. The software environment includes *Python 3.10.13* and *PyTorch 1.12.1* optimized with *CUDA 11.4*.

E Additional result

Apart from the ACC and NMI results in Table 1 and 2, we also provide the experiment results in PUR and ARI metrics in Table 5 and 6, respectively. It can be observed that the PUR and ARI results follow similar trends with those in ACC and NMI. Nevertheless, to analyze the training process more deeply, we also record the loss value and corresponding performances at each training epoch, which are visualized in Fig. 3. It is obvious that the loss value continuously decreases and finally converges to the minimum, while the ACC and NMI increase to the top and fluctuate around their top values. These observations well illustrate the effectiveness of loss design in Eq. (13) and the proposed IMDC-DIA method.

Table 4: The neuron numbers of the proposed IMDC-DIA method specific to different datasets.

	HandWritten	Caltech5V	Flower17	MSRCV1
1st-view	(60, 64, 216)	(21, 16, 40)	(34, 512, 5376)	(21, 256, 1302)
2nd-view	(60, 64,76)	(21, 64, 254)	(34, 256, 512)	(21, 32, 48)
3rd-view	(60, 32, 64)	(21, 128, 928)	(34, 512, 5376)	(21, 64, 512)
4th-view	(60, 16, 6)	(21, 128, 512)	(34, 512, 5376)	(21, 32, 100)
5th-view	(60, 64, 240)	(21, 256, 1984)	(34, 256, 1239)	(21, 64, 256)
6th-view	(40, 32, 47)	-	(34, 512, 5376)	(21, 64, 210)
7th-view	-	-	(34, 512, 5376)	-

Table 5: Performance (PUR and ARI) comparison between the proposed IMDC-DIA and recent incomplete deep multi-view clustering approaches. The *avg.* column refers to performance averages of all missing ratios. Note that, the best results are marked in bold, while the second-best with underline.

Metric		PUR						ARI					
Missing ratio		0.1	0.3	0.5	0.7	0.9	avg.	0.1	0.3	0.5	0.7	0.9	avg.
HandWritten	DITA-IMVC	77.85	80.90	81.38	81.02	55.00	75.23	66.61	70.78	70.87	68.67	35.92	62.57
	DSIMVC	81.13	80.73	79.88	77.50	51.18	74.08	71.69	70.19	67.27	63.91	32.36	61.08
	DCP	83.97	78.08	79.40	73.40	13.42	65.65	75.32	61.27	62.61	53.62	0.02	50.57
	DVIMC	86.53	83.12	45.53	25.27	19.45	51.98	82.38	77.91	35.65	14.20	5.74	43.18
	CPSPAN	90.42	91.25	91.08	90.27	87.82	90.17	80.83	82.12	81.69	80.46	78.02	80.62
	IMDC-DIA	**96.37**	**93.68**	**91.80**	**90.65**	**87.93**	**92.09**	**92.09**	**93.68**	**91.80**	**90.65**	**87.93**	**91.23**
Caltech5V	DITA-IMVC	79.10	75.76	68.02	60.64	41.31	64.97	61.06	57.26	48.75	40.62	18.10	45.16
	DSIMVC	76.64	73.14	66.36	58.57	46.31	64.20	60.87	54.83	47.05	37.67	22.88	44.66
	DCP	49.24	54.31	51.52	48.40	17.29	44.15	23.18	27.54	28.11	20.46	-0.06	19.85
	DVIMC	**88.64**	81.36	**84.93**	81.38	68.83	81.03	77.33	67.12	71.87	68.58	52.92	67.56
	CPSPAN	83.29	81.10	78.14	77.76	80.05	80.07	69.41	65.71	62.80	61.27	63.60	64.56
	IMDC-DIA	86.05	**85.05**	83.33	**85.02**	**81.29**	**84.15**	**86.05**	**85.05**	**83.33**	**85.02**	**81.29**	**84.15**
Flower17	DITA-IMVC	20.61	19.53	15.10	13.80	13.41	16.49	5.63	4.81	2.20	2.10	1.96	3.34
	DSIMVC	20.07	18.58	14.34	14.00	13.82	16.16	5.38	4.30	1.98	2.06	2.48	3.24
	DCP	27.11	27.18	23.87	15.76	11.10	21.00	10.07	8.02	3.86	1.76	0.03	4.75
	DVIMC	37.84	32.57	27.87	24.53	22.97	29.16	18.88	14.32	10.93	8.47	7.71	12.06
	CPSPAN	38.11	40.86	32.55	39.68	37.23	37.69	17.47	19.36	13.19	19.06	16.89	17.19
	IMDC-DIA	**52.03**	**46.47**	**44.22**	**41.54**	**39.14**	**44.68**	**52.03**	**46.47**	**44.22**	**41.54**	**39.14**	**44.68**
MSRCV1	DITA-IMVC	76.83	74.60	70.95	66.83	56.67	69.18	58.42	54.43	49.80	43.22	29.20	47.01
	DSIMVC	78.57	76.98	70.79	71.59	65.40	72.67	59.93	57.68	51.54	50.56	41.58	52.26
	DCP	18.10	20.63	22.54	22.54	23.17	21.40	-0.01	0.07	0.24	-0.16	-0.21	-0.01
	DVIMC	74.60	61.43	56.83	43.81	38.25	54.98	57.67	45.83	41.15	27.52	19.05	38.24
	CPSPAN	85.08	86.51	85.40	84.29	78.10	83.88	70.63	72.54	70.95	69.02	61.63	68.95
	IMDC-DIA	**91.27**	**92.22**	**86.03**	**85.08**	**83.02**	**87.52**	**91.27**	**92.22**	**86.03**	**85.08**	**83.02**	**87.52**

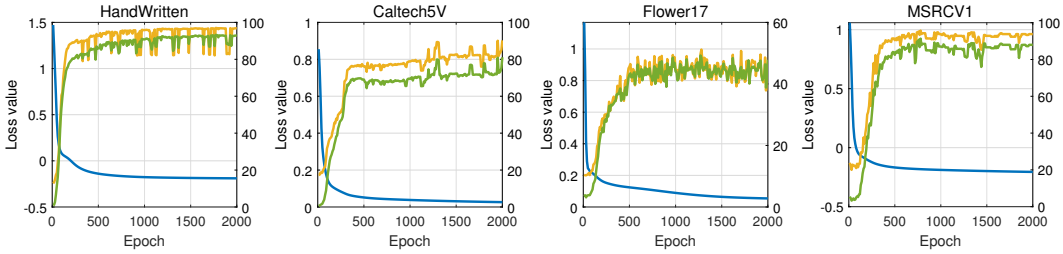

Figure 3: Loss and performance variation at each epoch on the four benchmark datasets. The blue curve represents loss value, while the yellow and green curves are accuracy and NMI, respectively.

Table 6: Performance (PUR and ARI) comparison between the proposed IMDC-DIA method and its variants of modifying and removing the data imputation and alignment components. The *avg.* column refers to performance averages of all missing ratios, while the *avg.* row reports the performances in the setting of completing missing data with average values. Note that, the best results are marked in bold, while the second-best with underline.

Metric			PUR						ARI					
Missing ratio			0.1	0.3	0.5	0.7	0.9	avg.	0.1	0.3	0.5	0.7	0.9	avg.
HandWritten	w/o imp.	avg	78.12	68.13	59.27	53.78	44.48	60.76	62.70	44.29	24.49	18.14	11.49	32.22
		zero	83.98	70.98	62.73	51.35	46.63	63.14	71.39	47.75	32.75	21.58	15.48	37.79
		rand	80.18	71.13	59.95	55.62	46.58	62.69	70.79	55.61	32.89	25.63	15.86	40.16
		none	96.13	84.73	89.28	79.67	56.68	81.30	91.59	77.60	80.11	69.16	37.18	71.13
	w/o align.		93.62	88.18	90.57	83.77	84.12	88.05	86.73	78.23	80.55	71.71	69.92	77.43
	IMDC-DIA		**96.37**	**93.68**	**91.80**	**90.65**	**87.93**	**92.09**	**92.09**	**86.37**	**82.60**	**80.53**	**75.42**	**83.40**
Caltech5V	w/o imp.	avg	78.50	74.48	64.81	54.62	48.86	64.25	60.29	46.02	30.32	17.49	12.95	33.41
		zero	82.17	73.60	67.88	61.81	53.00	67.69	65.43	50.59	38.09	27.57	16.36	39.61
		rand	79.24	68.67	62.43	58.93	48.24	63.50	67.77	51.12	37.88	28.01	17.16	40.39
		none	**87.98**	82.45	81.71	76.31	61.93	78.08	**76.63**	68.91	68.49	58.79	38.25	62.22
	w/o align.		80.81	82.19	80.93	83.50	81.60	81.80	68.21	68.29	67.07	68.89	65.25	67.54
	IMDC-DIA		86.05	**85.05**	**83.33**	**85.02**	**81.29**	**84.15**	73.54	**72.05**	**69.00**	**71.06**	**65.78**	**70.29**
Flower17	w/o imp.	avg	48.70	41.72	34.07	30.59	26.20	36.25	25.93	14.78	11.40	8.11	5.40	13.12
		zero	45.81	37.55	32.70	27.97	25.42	33.89	27.61	19.97	13.97	9.85	7.61	15.80
		rand	46.37	35.71	30.96	26.79	24.46	32.86	**28.77**	18.25	12.66	9.27	6.30	15.05
		none	42.92	36.84	26.62	13.85	18.14	27.67	22.89	18.09	9.91	0.68	2.98	10.91
	w/o align.		39.46	37.60	36.57	32.28	29.71	35.12	18.55	18.39	16.34	13.59	11.55	15.68
	IMDC-DIA		**52.03**	**46.47**	**44.22**	**41.54**	**39.14**	**44.68**	27.04	**25.11**	**21.99**	**20.89**	**17.73**	**22.55**
MSRCV1	w/o imp.	avg	72.86	59.52	52.22	44.60	38.10	53.46	52.89	27.30	20.49	10.75	7.80	23.85
		zero	69.05	63.33	57.30	46.51	43.81	56.00	43.29	33.62	25.93	14.12	11.75	25.74
		rand	67.46	56.83	51.59	46.03	44.92	53.37	48.75	31.53	26.17	15.75	12.57	26.96
		none	79.68	62.38	62.22	33.97	41.43	55.94	66.16	46.72	39.52	8.04	14.72	35.03
	w/o align.		85.56	85.24	73.65	67.62	62.86	74.98	72.51	70.51	52.78	43.35	39.98	55.83
	IMDC-DIA		**91.27**	**92.22**	**86.03**	**85.08**	**83.02**	**87.52**	**80.92**	**82.50**	**70.95**	**69.28**	**64.59**	**73.65**