

ARTICLE

Deep Incomplete Multi-View Clustering Based on Subspace Learning

Jiao Wang^{1,*}, Tingting Song^{2,*} and Yunhui Zhou¹

¹School of Information and Control Engineering, Southwest University of Science and Technology, Mianyang, China

²Health Electronics Center, Institute of Microelectronics of Chinese Academy of Sciences, Beijing, China

*Corresponding Authors: Jiao Wang, Email: jiaowang@swust.edu.cn; Tingting Song, Email: songtingting@ime.ac.cn

Received: 22 June 2026; Accepted: 30 July 2026; Published: 22 October 2026

ABSTRACT: In practical scenarios, multi-view data often contains missing entries caused by complicated data collection and transmission procedures, posing great challenges to clustering analysis. Existing incomplete multi-view clustering methods have two obvious limitations: (1) Imputation-based methods inevitably generate inaccurate information during data recovery; (2) Imputation-free methods struggle to balance cross-view consistency and complementarity. To tackle the above issues, this paper proposes a deep incomplete multi-view clustering approach based on subspace learning, which integrates latent feature extraction, K-nearest neighbor-based feature imputation, cross-view contrastive alignment, and attention-driven fusion within a unified deep learning framework. Specifically, we first leverage autoencoders to extract view-specific latent features. Inspired by the observation that different views of the same sample should share similar latent representations, we impute missing latent features by leveraging cross-view correlations, using both view-specific complete features and common complete features through the K-nearest neighbor strategy. Then, we construct a fully-connected neural network layer to learn the self-expression matrix for each view. To minimize the impact of imputation noise on learning the self-expression matrix for clustering, these self-expression matrices are enforced to align using contrastive learning to capture cross-view consistency, and then fused based on a channel attention mechanism for pursuing cross-view complementarity. Finally, a high-quality cross-view consensus self-expression matrix is obtained for final clustering segmentation. Comprehensive evaluations on six benchmark datasets under various missing rates demonstrate the effectiveness and superiority of our method compared with state-of-the-art methods.

KEYWORDS: Deep incomplete multi-view clustering; subspace learning; missing feature imputation; contrastive alignment; attentive fusion

1 Introduction

Multi-view data widely exists in practical applications and is acquired via diverse sensors, media equipment, cameras, feature extraction modules. As a dominant unsupervised learning paradigm, multi-view clustering seeks to explore the intrinsic structure underlying multi-view data, attracting growing research across machine learning and computer vision [1]. However, multi-view data often exhibit missing entries caused by unstable sensors or transmission process, known as incomplete multi-view data. These data severely restricts the applicability of existing clustering methods. We can neither arbitrarily eliminate samples containing missing entries nor simply impute them with specific values (such as 0 or 1), as they tend to compromise clustering performance by either discarding critical information or introducing harmful noise. To address these challenges, numerous incomplete multi-view clustering (IMVC) methods have

been developed, which can be roughly classified into multi-kernel learning-based, matrix decomposition-based, graph learning-based, subspace learning-based, and deep learning-based approaches. Among these approaches, deep IMVC (DIMVC) has drawn widespread attention owing to the powerful capability in effectively modeling complex high-dimensional nonlinear structures. Current DIMVC methods can be primarily divided into two paradigms: imputation-based methods and imputation-free methods.

Imputation-based methods utilize data recovery techniques to complete missing data before proceeding with clustering. A common technique often employs generative adversarial networks (GAN) to reconstruct missing data or generate shared latent representations through adversarial learning [2–4]. Other methods are based on the data structure to fill in missing data. For example, a cross-view completion framework employing soft clustering assignments and global cluster centroids to reconstruct the latent features of the missing view [5]. To enhance imputation quality, a dual-mechanism approach combining K Nearest Neighbors (KNN)-based view completion with a view-category interactive Transformer is proposed to exploit complementary information between multi-view representations and categorical semantics [6]. A diffusion model is also introduced for predicting missing data, coupled with data augmentation to maintain generation quality under high missing rates [7]. While imputation-based methods provide pragmatic solutions for incomplete multi-view data, their reconstruction fidelity critically depends on the completeness of observable views. These methods are inherently susceptible to introducing approximation noise and degrading the efficiency during the completion process, particularly under high view-missing rates. This sensitivity may propagate completion deviations that ultimately degrade clustering performance. Imputation-free methods mitigate the negative influence brought by missing entries on clustering results by aligning across-view consistent latent representations. The primary technique adopts contrastive learning based on a contrastive prediction framework to pursue consistency across views [8]. To reduce the negative impacts brought by false negative samples, the work in [9] constructs a noise-robust contrastive loss. The work integrates a graph convolutional network (GCN) for handling missing data and combines it with instance-level contrastive learning to enhance feature representations [10]. A triple cross-view alignment based on contrastive learning with prototype, feature, and clustering assignment is proposed to preserve consistency among multiple views [11]. Other methods utilize intrinsic information within incomplete multi-view data to supervise the cross-view consistency. Embedded features of views are projected into a high-dimensional weighted space for exploiting inter-view complementarity, which is subsequently leveraged as high-confidence supervision for cross-view consistency mining [12]. An adaptive feature projection is employed to map the available data into a common space while achieving inter-view alignment through mean discrepancy minimization [13]. High-quality pseudo-labels are generated based on complete-view subset data for guidance, aligning feature representations with cluster assignments through local and global propagation networks [14]. Although these imputation-free methods can reduce the noise or bias caused by the data completion process, they tend to neglect the critical influence of view-specific discrimination on learning consistent representation. This leads to suboptimal utilization of cross-view consensus information and compromises clustering performance.

Moreover, existing DIMVC methods still have limitations in inherent interpretability and rigorous theoretical foundation. To address these limitations, researchers have attempted to integrate subspace learning, which is theoretically grounded in linear algebra and statistics. Although deep subspace clustering methods for complete multi-view data have been widely researched [15–17], they fail to directly handle incomplete data because they cannot construct self-expression matrices among multi-view samples with missing values. To the best of our knowledge, only one self-supervised subspace learning approach has been proposed for IMVC [18]. However, this method ignores the discriminative contributions of different views, as well as the impact of missing view data on learning the common subspace representation. To overcome these challenges, we propose a novel DIMVC based on subspace learning (DIMVCSL) method. Our

method pioneers a unified framework to harness the inter-view consistency and cross-view complementarity, combined with fast missing completion. Specifically, we employ multiple deep autoencoders to extract latent feature representations of views. To avoid the time-consuming overhead of data imputation, we adopt the training-free KNN method and instance-driven mechanism to complete the missing latent features for views. In contrast to traditional imputation methods that rely on distribution estimation or statistical prediction, we perform view-specific missing feature recovery by exploiting the topological relationships among the pairwise view feature instances. Then, we utilize multiple fully-connected layers to learn the self-expression coefficient matrices from the completed multi-view features. To avoid the interference caused by imputation noise and explore inter-view consistent semantic information across views, we adopt contrastive learning to align these matrices. Subsequently, these matrices are fused with discriminative weights by exploring the higher-order relations guided by a channel attention mechanism. Finally, spectral clustering is conducted over the fused self-expression coefficient matrix to yield the results. The main contributions are listed as follows.

- A novel unified end-to-end framework of deep incomplete multi-view clustering based on subspace learning is proposed, named DIMVCSL. In contrast to existing disjoint or loosely coupled pipelines, DIMVCSL seamlessly integrates fast latent-space missing feature imputation, self-expression matrix learning, consistency-aware contrastive alignment, and complementarity-driven attentive fusion into a joint learning paradigm.
- Our method adaptively completes missing data in the latent space via topological propagation among cross-view neighbors. This is further reinforced by contrastive alignment on self-expression matrices, which not only effectively mitigates the imputation noise but also can explore cross-view semantic consistency, thereby ensuring the reliability of subsequent subspace clustering.
- We design a channel-attentive fusion module to learn a consensus self-expression matrix for exploring cross-view discriminative complementarity. Comprehensive evaluations on six datasets under various missing rates indicate that DIMVCSL achieves superior performance compared with state-of-the-art IMVC methods.

Compared with existing DIMVC methods, our DIMVCSL exhibits three key architectural innovations: (1) Cross-view topological imputation in latent space: Our method performs imputation in the dynamically optimized latent space, where cross-view semantic consistency is progressively enhanced during training. This enables the imputation quality to improve in tandem with representation learning, rather than being fixed by initial feature quality. (2) Self-expression-level contrastive alignment: While most contrastive IMVC methods align latent features or cluster assignments, DIMVCSL aligns self-expression coefficient matrices that directly encode sample-sample relationships. This design bridges the gap between representation learning and the ultimate clustering objective, as spectral clustering operates on the affinity matrix derived from these coefficients. (3) Unified subspace-oriented framework: To the best of our knowledge, DIMVCSL is the first end-to-end deep subspace clustering framework for incomplete multi-view data that jointly optimizes imputation, self-expression learning, cross-view alignment, and attentive fusion, providing a theoretically grounded solution with inherent interpretability from subspace learning.

2 Related Work

2.1 Incomplete Multi-View Clustering

Driven by the fast advancement of multi-modal sensing and data collection techniques, IMVC has become a popular research topic within the machine learning community. Existing IMVC methods that adopt traditional machine learning techniques can be generally grouped into non-negative matrix factorization (NMF)-based, kernel trick-based, graph learning-based, and subspace learning-based methods [19].

(1) IMVC based on NMF. The central principle of such methods is to mine a consensus representation for two or more incomplete views using NMF, thereby addressing the missing view problem for subsequent clustering tasks [20,21]. Subsequently, to further enhance the clustering performance, $\ell_{2,1}$ -norm and graph regularization are respectively integrated with semi-NMF into a joint framework for IMVC [22]. To employ discriminative learning on well-completion and poorly completion, weighted NMF methods are proposed for IMVC [23]. However, most NMF-based IMVC methods still suffer from limited representation learning capability, as they fail to fully capture the deep nonlinear correlations among multi-view data. (2) IMVC based on kernel trick. These methods usually transform incomplete multiple views into a high-dimensional kernel space to exploit nonlinear correlations, and then complete the missing kernel information using the available view data. For example, zero values are adopted to complete the incomplete kernels for view data [24]. The incomplete base matrix is imputed by the learned consensus clustering matrix [25]. LRKT-IMVC introduces a kernel tensor regularized by low-rank restrictions to explore the consistent information for completing the missing kernel values [26]. Nevertheless, kernel-based methods often rely on manual kernel selection, which lacks adaptability, and the high computational complexity of kernel operations limits their scalability to large-scale datasets. (3) IMVC based on graph learning. Such methods concentrate on learning a unified graph that preserves the intrinsic geometric structure of view data, since graph-based representation can characterize the similarity and correlation among data samples [27,28]. In [29], a unified framework is proposed to impute missing values and learn a similarity graph. A self-supervised consensus graph completion has been introduced to estimate missing values [30]. To alleviate the heavy computational requirement, consensus bipartite graph learning has been proposed in [31,32]. However, graph-based IMVC methods are often sensitive to the quality of view missing completion. (4) IMVC based on subspace learning. As one of the most widely studied directions in IMVC, subspace learning-based methods aim to extract low-dimensional latent representations of incomplete views through dimensionality reduction, projection matrix learning, or feature selection, and then construct a consensus self-expression coefficient matrix that exploit multi-view inherent consistency and complementarity [33]. Missing data imputation and self-expression coefficient learning are jointly performed in [34]. A reconstruction term is introduced to impute missing entries and then learns the consistent self-expression coefficient matrix from the recovered views [35]. RTSL seeks to learn a low-rank coefficient tensor by imposing a nuclear norm to explore the correlation among views based on matrix factorization to recover missing entries [36]. SLRTOP jointly learns an orthogonal projection and a representation tensor to construct a more informative affinity matrix, effectively alleviating redundancy and noise in tensor-based multi-view subspace learning and achieving superior performance [37]. Despite their promising performance, existing subspace learning-based IMVC methods fail to explore the discriminative cross-view complementarity, resulting in suboptimal clustering results under high missing rates.

2.2 Deep Incomplete Multi-View Clustering

Due to the limitations of traditional machine learning-based IMVC in handling high-dimensional incomplete multi-view data, DIMVC methods have garnered widespread attention and become a research hotspot in this field. The fundamental deep learning frameworks employed in DIMVC methods include deep autoencoders (DAE), GAN, and graph neural networks (GNN), which are elaborated as follows. (1) DAE-based IMVC method. As one of the widely used deep frameworks, the low-dimensional latent features of data can be extracted by optimizing the reconstruction loss of DAE [12,38,39]. However, the presence of missing values may render the latent features inadequate to support high-quality clustering. Consequently, missing data completion or consistency constraints are typically incorporated into this framework. Completion methods can be categorized as prediction-based imputation [8,40], neighbor-based imputation [39,41], and

prototype-based imputation [42,43]. Specifically, prediction-based methods complete missing samples or features across available views by minimizing the conditional entropy. Neighbor-based methods impute missing entries by leveraging the nearest neighbor relationships from fully observed instances. Prototype-based methods complete missing data by integrating the sample features from observed views with prototype representations from the missing views. More recently, a sample-score-select strategy is proposed to pick one high-quality prototype per batch for contrastive alignment, showing that prototype-level alignment alone ensures cross-view consistency without complex imputation or complete paired data [44]. Instead of explicit imputation, consistency constraints enforce latent feature consistency across views in a shared latent space. Contrastive learning is an effective approach for exploiting multi-view consistency, which achieves this goal through boosting the matching degree of latent embeddings from positive sample pairs and suppresses the similarity measurement among negative sample pairs [8,9,11]. Self-supervised information is also adopted to obtain cross-view consistency between available and incomplete entries [12]. Another imputation-free method projects latent embeddings from incomplete views into a shared space as similar as possible. For example, minimizing mean discrepancy is used to align the similarity of multi-view incomplete views within the common space [13]. Additionally, several methods combine completion with consistency constraints, such as diffusion contrastive generation that integrates diffusion-based missing data recovery and contrastive alignment [45]. (2) GAN-based IMVC method. GAN-based methods utilize the generator to impute missing data and discriminators to force the generated data to approximate the real distribution. Through this adversarial training, a cross-view consistent shared embedding space is learned for clustering [3,4]. The work in [46] imputes missing representations via cross-sample attention and enforces semantic consistency and fairness through cluster-level contrastive learning and adversarial disentanglement. GAN is combined with optimal transport and fusion for learning discriminative latent graph embedding [47]. A GAN-based model of integrating semantic discrimination and semantic invariant learning is proposed in [48]. (3) GNN-based IMVC method. GNN learns missing data via a message propagation mechanism, which is further integrated with instance-level contrastive learning to enhance feature representations [10]. The work in [5] imputes missing features under the graph convolutional network (GCN) framework, guided by soft cluster assignments and global cluster centroids. Additionally, a pseudo-label propagation strategy based on GCN is proposed to improve clustering performance [14].

3 Proposed Method

This section begins with a detailed notations of our DIMVCSL method, whose network architecture is visualized in Fig. 1. We then present its four core modules with implementation details. Finally, the overall network objective function and its optimization are provided.

3.1 Notation

Throughout the paper, let $\{\mathbf{X}^v \in \mathbb{R}^{N \times d^v}\}_{v=1}^V$ denote incomplete multi-view data, which contains N samples and V views with different dimensional size of d^v . To denote the status (missing/available) of samples, we introduce an indicator matrix $\mathbf{M} \in \{0, 1\}^{N \times V}$, where $\mathbf{M}_{i,j} = 0$ means the missing data of the i -th sample in j -th view, oppositely, $\mathbf{M}_{i,j} = 1$ means that the available data of the i -th sample in j -th view. For clarity, we summarize the primary notation definitions in Table 1.

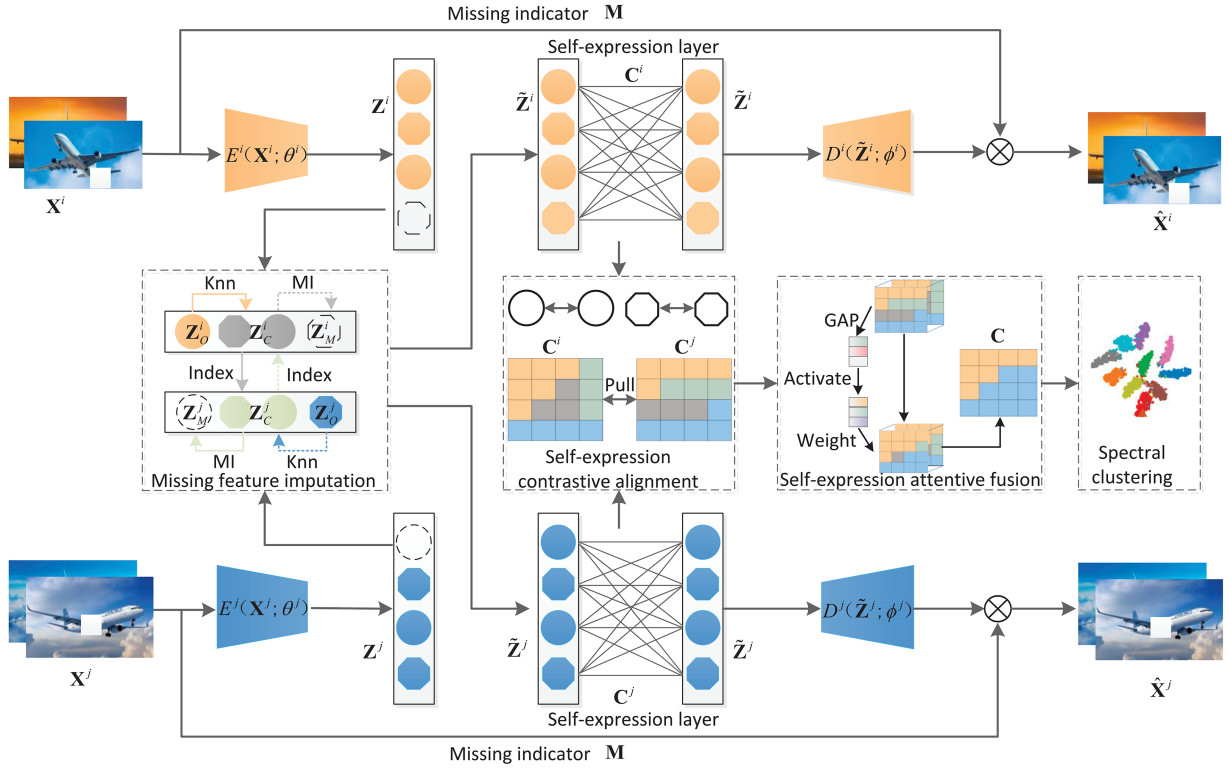


Figure 1: The network architecture of the proposed DIMVCSL method.

Table 1: Summary of notations.

Notation	Definition	Notation	Definition
N	Number of samples	$\mathbf{M} \in \mathbb{R}^{N \times V}$	Missing indicator matrix of multi-view data
V	Number of views	$\mathbf{X}^v \in \mathbb{R}^{N \times d^v}$	Incomplete multi-view data of the v -th view
K	Number of nearest neighbors	$\hat{\mathbf{X}}^v \in \mathbb{R}^{N \times d^v}$	Reconstructed data of the v -th view
G	Number of clusters	$\mathbf{Z}^v \in \mathbb{R}^{N \times d}$	Latent features of the v -th view
d^v	Dimension of the samples in v -th view	$\tilde{\mathbf{Z}}^v \in \mathbb{R}^{N \times d}$	Completed latent features of the v -th view
d	Dimension of deepest latent features	$\mathbf{C}^v \in \mathbb{R}^{N \times N}$	Self-expression matrix of the v -th view
$\mathbf{Z}_C^v$	Common complete features in the v -th view	$\mathbf{C} \in \mathbb{R}^{N \times N}$	Consensus self-expression matrix
$\mathbf{Z}_O^v$	Complete features existing only in the v -th view	$\mathbf{W} \in \mathbb{R}^{N \times N}$	Affinity matrix
$\mathbf{Z}_M^v$	Missing features of the v -th view	$\mathbf{Q} \in \mathbb{R}^{N \times G}$	Cluster indicator matrix

3.2 Encoder-Decoder Reconstruction

As illustrated in Fig. 1, our framework employs an encoder-decoder pipeline for each view, which is implemented by fully connected neural networks for simplicity. The encoders aim to extract view-specific latent features from the original incomplete data, such that these features also remain missing entries. The decoders are on target to reconstruct the original incomplete data. Specifically, since the original multi-view data are incomplete, we first construct a missing indicator matrix $\mathbf{M}$ to characterize their sample status. For arbitrary v -th view, the original view $\mathbf{X}^v$ works as the input of the encoder $E^v(\mathbf{X}^v; \theta^v)$, where the v denotes the parameters of encoder. Thus, the latent features can be present as $\mathbf{Z}^v = E^v(\mathbf{X}^v; \theta^v)$. Then, we complete the incomplete features $\mathbf{Z}^v$ in the latent space, and the implementation details can be found

in Section 3.3. The completed latent features of arbitrary v -th view denote as $\tilde{\mathbf{Z}}^v$. They work as the input of the decoder $D^v(\tilde{\mathbf{Z}}^v; \phi^v)$, where the ϕ^v is the parameters of the decoder. Subsequently, the output of decoder $\tilde{\mathbf{X}}^v$ and missing indicator matrix $\mathbf{M}$ denoted as $\hat{\mathbf{X}}^v = \tilde{\mathbf{X}}^v \mathbf{m}^v$. Where $\mathbf{m}^v$ is the column vector of $\mathbf{M}$, which represents the sample status of v -th view. Finally, through minimizing the reconstruction loss between pipeline ($\mathbf{X}^v$), the latent features $\mathbf{Z}^v$ of each view can be extracted. To simplify parameter configuration and achieve a unified network, we design the encoders and decoders with an identical number of neurons in both the hidden and output layers. Thus, the deepest view-specific latent features have the same dimension, denoted as $\mathbf{Z}^v = [\mathbf{z}_1^v, \mathbf{z}_2^v, \dots, \mathbf{z}_N^v] \in \mathbb{R}^{N \times d}$. The multi-view reconstruction loss can be formulated as:

$$\mathcal{L}_{rec} = \sum_{v=1}^V \|\mathbf{X}^v - \hat{\mathbf{X}}^v\|_F^2 \quad (1)$$

3.3 Missing Feature Imputation

To capture the similarity relationships across samples of the views, it is essential to complete the missing latent features. Inspired by the neighbor-based completion [41], we design a similar missing feature imputation module based on the KNN method in the latent space, aiming to achieve efficient completion. The motivation for performing KNN imputation in the latent space, rather than the original feature space, is grounded in two key observations: (1) Semantic compactness: Through encoder-decoder reconstruction optimization, the latent space progressively organizes semantically similar samples into local neighborhoods, making KNN relationships more reliable than in the original high-dimensional space; (2) Collaboratively optimize with representation learning: Unlike fixed original-space imputation, latent-space imputation dynamically updates as the encoder-decoder parameters are optimized, creating a virtuous cycle where improved representations lead to better imputation. In this module, the latent features of arbitrary view $\mathbf{Z}^v \in \mathbb{R}^{N \times d}$ are divided into three categories: view-specific complete features existing only in the v -th view, missing features of the v -th view, common complete features across views in the v -th view. They are denoted as $\mathbf{Z}_O^v \in \mathbb{R}^{N_O \times d}$, $\mathbf{Z}_M^v \in \mathbb{R}^{N_M \times d}$, and $\mathbf{Z}_C^v \in \mathbb{R}^{N_C \times d}$, respectively, where $N = N_O + N_M + N_C$. We utilize the cross-view correlation of view-specific complete features and common complete features to impute the missing features.

Specifically, we first identify multi-view three categories latent features based on the missing indicator matrix $\mathbf{M}$ and latent features $\{\mathbf{Z}^v\}_{v=1}^V$. Taking arbitrary two views for example, we construct the common complete features space $\mathbf{Z}_C^i$ and $\mathbf{Z}_C^j$ in both view i and view j , which are extracted by the view-specific encoder. View-specific complete features existing only in view i and j , denoted as $\mathbf{Z}_O^i$ and $\mathbf{Z}_O^j$, respectively, are also obtained. Subsequently, we find the index set of the similar K-nearest neighbors within the common complete features space of another view for the view of missing features, which preserves the topological relationships among views. For the arbitrary missing feature $\forall \mathbf{z}_M^i \in \mathbf{Z}_M^i$ in i -th view, we utilize the $\mathbf{z}_O^j$ with the same topological relationships to identify the index set of the KNN in $\mathbf{Z}_C^j$. Mathematically, this process can be formulated as:

$$\mathcal{N}(\mathbf{z}_O^i) = \left\{ \mathbf{z}_C^j \in \mathbf{Z}_C^j \mid \text{rank}(\|\mathbf{z}_O^i - \mathbf{z}_C^j\|) \leq K \right\} \quad (2)$$

where $\mathcal{N}(\mathbf{z}_O^i)$ denotes the index set of the K-nearest neighbors of $\mathbf{z}_O^i$ and $\text{rank}(\cdot)$ is a sorting function and $\|\cdot\|$ represents the distance relationships and $\mathbf{z}_O^i \in \mathbf{Z}_O^i$, respectively.

Then, according to the consistency topological relationships among the different views, we utilize these KNN index set $\mathcal{N}(\mathbf{z}_O^i)$ to learn the neighboring features from the common complete features $\mathbf{Z}_C^j$, and the

mean values of these features are used to impute the missing features $\mathbf{z}_M^i$, which can be formulated as:

$$\mathbf{z}_M^i = \frac{1}{K} \sum_{k=1}^K \mathbf{z}_C^i \left(\mathcal{N}(\mathbf{z}_O^i) \right) \quad (3)$$

where K is the number of nearest neighbors.

After missing features imputation, the complete latent features $\tilde{\mathbf{Z}}^i$ are formed by concatenating $\mathbf{Z}_O^i$, $\mathbf{Z}_C^i$, and $\mathbf{Z}_M^i$. The missing features imputation module leverages the topological relationships among the sample space of views with-out requiring any training parameters, thereby maintaining training stability and efficiency. Moreover, it can dynamically complete the missing features and automatically update them as the latent features, extracted by the encoder and decoder pipeline, are optimized during forward propagation.

3.4 Self-Expression Contrastive Alignment

Following completion of the missing features, a fully-connected self-expression layer is inserted between the completed features and decoder to learn the representation correlations among samples of each view. The network parameters of the layer are defined as the self-expression coefficient matrix $\mathbf{C}^v \in \mathbb{R}^{N \times N}$, which is a subspace-oriented representation. the view-specific coefficient matrix is dynamically acquired by minimizing the representation error between the input and output of each self-expression layer. To capture multi-view diversity and discrimination, multiple self-expression layers are employed across all views. The loss of self-expression of all views is expressed as

$$\mathcal{L}_{sel} = \sum_{v=1}^V \|\tilde{\mathbf{Z}}^v - \mathbf{C}^v \tilde{\mathbf{Z}}^v\|_F^2 + \sum_{v=1}^V \|\mathbf{C}^v\|_F^2 \quad \text{s.t.} \quad \text{diag}(\mathbf{C}^v) = \mathbf{0} \quad (4)$$

where the Frobenius norm acts as a regularization loss that strengthens intra-subspace connections through the grouping effect, thereby facilitating subsequent clustering, and together with constraint $c_{ii}^v = 0$, avoids trivial solutions.

Due to the inevitable introduction of noise during the process of completing missing features and the interference from view-private meaningless information, the consistency of the cross-view self-representation relationship is compromised. To address this problem, a contrastive alignment module is designed for achieving the consistency of $\{\mathbf{C}^v\}_{v=1}^V$ for the self-expression layers of all views. The alignment of self-expression coefficient matrices, rather than feature-level or assignment-level contrastive learning, is primarily based on the two considerations: (1) In subspace clustering, the self-expression matrix $\mathbf{C}^v$ directly encodes the pairwise sample relationships that determine the affinity matrix $\mathbf{W} = 1/2 (|\mathbf{C}| + |\mathbf{C}|^T)$ used for spectral clustering. Therefore, enforcing consistency at the self-expression level directly optimizes the representation that drives the final clustering decision, establishing a principled connection between representation learning and clustering. (2) However, feature-level alignment may propagate imputation noise, while self-expression matrices, being higher-level abstractions, are inherently more robust to low-level perturbations. Unlike cluster assignment alignment that forces identical hard assignments, self-expression alignment allows soft, view-specific relationship patterns while encouraging cross-view consistency, better preserving view complementarity.

Specifically, to begin with, we categorize positive and negative sample pairs based on multi-view self-expression coefficient matrices. Coefficients belonging to the same sample from separate views serve as positive pairs; in contrast, any coefficients associated with different samples, either inside one view or among various views, are taken as negative pairs. Taking the coefficient of one sample c_n^v for example, where $n \in [1, \dots, N]$ and $v \in [1, \dots, V]$, the number of positive pairs and negative pairs is $V - 1$ and $V(N - 1)$,

respectively. The self-expression contrastive alignment module aims to boost the similarity degree of positive pairs and reduce the similarity value of negative pairs simultaneously. To this end, the coefficients among samples of the same category can be enhanced, while those among samples of different categories can be effectively suppressed. Consequently, the cross-view consistency can be fully captured, meanwhile the interference caused by view-private features and imputation noise is eliminated, thereby yielding more discriminative representations for clustering and segmentation. We adopt contrastive loss to optimize this module for realizing the above target. Given the i -th and j -th view, the self-expression contrastive loss is represented as follows:

$$\ell_n^{(i,j)} = -\log \frac{\exp(s(c_n^i, c_n^j)/\tau)}{\sum_{m=1}^N [\exp(s(c_n^i, c_m^i)/\tau) + \exp(s(c_n^i, c_m^j)/\tau)]} \quad (5)$$

where τ is a temperature parameter and n and m represent the samples in views and $s(\cdot)$ denotes the similarity of coefficient positive pairs and negative pairs, respectively. In this module, cosine distance is employed to measure the similarity:

$$s(c_n^i, c_n^j) = \frac{\langle c_n^i, c_n^j \rangle}{\|c_n^i\| \|c_n^j\|} \quad (6)$$

For a dataset with V views and N samples, the accumulated self-expression contrastive loss is denoted as:

$$\mathcal{L}_{con} = \frac{1}{2N} \sum_{i,j=1, i \neq j}^V \sum_{n=1}^N \ell_n^{(i,j)} \quad (7)$$

3.5 Self-Expression Attentive Fusion

Following the alignment of the cross-view consistency of self-expression relationships, this paper uses channel attention mechanism to fuse them to acquire a consensus matrix [49], thereby exploring cross-view diversity and complementarity. Firstly, we stack all the self-expression coefficient matrices $\{C^v\}_{v=1}^V$ as a three-order tensor $\mathcal{C} = \Psi C^1, C^2, \dots, C^V$ ($\mathcal{C} \in \mathbb{R}^{N \times N \times V}$). Then, the high-order relationships within the tensor are extracted through a network, which is implemented by a Global Average Pooling (GAP) operation and a Multi-Layer Perceptron (MLP). GAP is employed to learn the global description factor $g \in \mathbb{R}^V$ of each matrix in the tensor $\mathcal{C}$ for exploring global information, which can be formulated as

$$g = \frac{1}{N \times N} \sum_{i=1}^N \sum_{j=1}^N \mathcal{C}(i, j) \quad (8)$$

Subsequently, the global description factor g is excited into different weights by a two-layer bias-free MLP. The first layer aims to compress the dimension of g , and the second one recovers the dimension. In this manner, the importance of each self-expression matrix is learned. The obtained weights are then employed to scale the tensor $\mathcal{C}$.

$$\tilde{\mathcal{C}} = \mathcal{C} \cdot \sigma(\theta_2 \delta(\theta_1 g)) \quad (9)$$

where $\theta_1 \in \mathbb{R}^{V \times \frac{V}{r}}$ and $\theta_2 \in \mathbb{R}^{\frac{V}{r} \times V}$ denote the learned parameters of the two fully connected layers. $\delta(\cdot)$ and $\sigma(\cdot)$ correspond to their respective activation functions. In general, we adopt ReLU for $\delta(\cdot)$ and Sigmoid for $\sigma(\cdot)$.

Finally, the scaled tensor $\tilde{\mathcal{C}}$ is fused by a convolution operation to obtain the consensus matrix $\mathbf{C} = k * \tilde{\mathcal{C}}$, where k is the convolutional kernel size.

3.6 Clustering Module

After obtaining the consensus self-expression coefficient matrix $\mathbf{C}$ of all views, spectral clustering is employed to generate the clustering results. We first compute the affinity matrix by $\mathbf{W} = 1/2 (|\mathbf{C}| + |\mathbf{C}|^T)$. The Laplacian matrix is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is a diagonal matrix with diagonal elements given by $\mathbf{d}_{ii} = \sum_j w_{ij}$. Then, we can obtain the clustering results by optimizing the following objective function:

$$\min_{\mathbf{Q}} \text{tr}(\mathbf{Q}^T \mathbf{L} \mathbf{Q}) \quad \text{s.t. } \mathbf{Q}^T \mathbf{Q} = \mathbf{I} \quad (10)$$

where $\mathbf{Q} \in \mathbb{R}^{N \times G}$ denotes the cluster indicator matrix, which represents the N samples are categorized into G clusters. The constraint $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ ensures orthogonality of these eigenvectors and prevents trivial solutions.

3.7 Optimization and Objective Function

The training procedure of our method is divided into two phases: pre-training and fine-tuning. During pre-training, the multiple encoder-decoder pipeline without missing imputation and self-expression layers is used for extracting latent features, which are optimized by reconstruction loss. This stage aims to initialize parameters of the view-specific autoencoder, as well as to obtain the initial latent features. The objective function is defined as follows:

$$\min_{\{\mathbf{Z}^v, \theta_e^v, \phi_d^v\}_{v=1}^V} \mathcal{L}_1 = \sum_{v=1}^V \|\mathbf{X}^v - \hat{\mathbf{X}}^v\|_F^2 \quad (11)$$

During fine-tuning, we optimize the overall network by integrating reconstruction loss with missing feature imputation (Eq. (1)), multi-view self-expression loss (Eq. (4)), regularization loss ($\|\mathbf{C}\|_F^2$), and self-expression contrastive loss (Eq. (7)) to obtain cross-view consensus matrix $\mathbf{C}$. The objective function of DIMVCSL becomes:

$$\begin{aligned} \min_{\mathbf{C}} \mathcal{L}_2 &= \mathcal{L}_{rec} + \lambda_1 (\mathcal{L}_{sel} + \|\mathbf{C}\|_F^2) + \lambda_2 \mathcal{L}_{con} \\ &= \sum_{v=1}^V \|\mathbf{X}^v - \hat{\mathbf{X}}^v\|_F^2 + \lambda_1 \cdot \left(\sum_{v=1}^V \|\tilde{\mathbf{Z}}^v - \mathbf{C}^v \tilde{\mathbf{Z}}^v\|_F^2 + \sum_{v=1}^V \|\mathbf{C}^v\|_F^2 + \|\mathbf{C}\|_F^2 \right) \\ &\quad + \lambda_2 \cdot \left(-\frac{1}{2N} \sum_{i,j=1, i \neq j}^V \sum_{n=1}^N \log \frac{\exp(s(c_n^i, c_n^j)/\tau)}{\sum_{m=1}^N [\exp(s(c_n^i, c_m^i)/\tau) + \exp(s(c_n^i, c_m^j)/\tau)]} \right) \\ \text{s.t. } &\text{diag}(\mathbf{C}^v) = \mathbf{0}, \text{diag}(\mathbf{C}) = \mathbf{0} \end{aligned} \quad (12)$$

where λ_1 and λ_2 are the trade-off parameters.

Finally, spectral clustering is applied to $\mathbf{W} = \frac{1}{2} (|\mathbf{C}| + |\mathbf{C}|^T)$ for acquiring clustering results. According to the entire methodology of our proposed DIMVCSL, the optimization process is listed in Algorithm 1.

Algorithm 1: The optimization process of DIMVCSL

Input: Multi-view incomplete data $\{\mathbf{X}^v \in \mathbb{R}^{N \times d^v}\}_{v=1}^V$, trade-off parameters λ_1 and λ_2 , maximum number of iterations for pre-training and fine-tuning stages T1 and T2, respectively.

(Continued)

Algorithm 1 (continued)

```

1: While  $t_1++ < T_1$  do
2: Train the network including only view-specific encoder and decoder to initialize the parameters
    $\{\theta^v\}_{v=1}^V$  and  $\{\phi^v\}_{v=1}^V$  by optimizing Eq. (11).
3: Load the pre-training model.
4: End while
5: While  $t_2++ < T_2$  do
6: Train the overall network to obtain the cross-view consensus self-expression coefficient matrix  $\mathbf{C}$  by
   optimizing Eq. (12).
7: End while
8: Perform spectral clustering on  $\mathbf{W} = \frac{1}{2}(|\mathbf{C}| + |\mathbf{C}|^T)$  to obtain clustering results  $\mathbf{Q}$  by minimizing Eq. (10).
Output: The clustering results.

```

4 Experiments

This section conducts a set of experiments to verify the effectiveness of our DIMVCSL method. We first compare the clustering performance of DIMVCSL against multiple SOTA IMVC methods to prove its outstanding performance across different missing rate. After that, ablation studies are carried out to quantify the effect of each core module, including missing feature imputation, self-expression contrastive alignment, and attentive fusion. Finally, parameter selection, convergence and stability analysis experiments are performed to further verify the stability and effectiveness of DIMVCSL.

4.1 Datasets

Six widely used multi-view datasets covering different data types and scales, are utilized to evaluate the clustering performance of DIMVCSL. They are briefly summarized as follows.

- MSRC-V1 [50] contains 210 images across 7 categories, consisting of trees, buildings, airplanes, cows, faces, cars, and bicycles, which include two types of feature views (GIST and HOG).
- LandUse-21 [51] is a dataset consisting of 2100 satellite images categorized into 21 categories, with PHOG and LBP as two views.
- Scene-15 [52] dataset describes 4485 scene images belonging to 15 categories, for example, office, kitchen, living room, and mountain. Following the comparison methods, GIST and LBP are utilized as distinct views.
- Synthetic3d [53] is a composite 3D dataset comprising 3 categories. Different from real-world datasets, the samples in this dataset are generated by computer graphics technology or other simulation methods.
- CUB [54] selects 600 samples belonging to 10 categories from the Caltech-UCSD Birds dataset, including deep visual features and text features.
- HandWritten [55] includes 2000 samples categorized into 10 categories corresponding to digits from 0 to 9.

Following the existing comparison methods, we randomly remove a proportion of samples (i.e., missing rates of 0.1, 0.3, 0.5, and 0.7) of multiple views to construct the incomplete data. For instance, given a missing rate of 0.3, 30% of the samples in each view are randomly discarded, with the constraint that the same sample cannot be missing simultaneously across different views. The missing indicator matrix $\mathbf{M}$ denotes the binary indicator of the missing samples.

4.2 Comparison Methods

We conduct comprehensive comparisons between our DIMVCSL algorithm and the following SOTA IMVC baselines.

- Completer [8] adopts contrastive learning based on a contrastive prediction framework for completing the missing data.
- DCP [40] provides a unified framework based on contrastive learning that focuses on both cross-view consistency exploration and missing data completion.
- ProImp [42] leverages dual-attention layers and dual-contrastive learning to learn prototypes as well as the correlations between samples and prototypes, so as to reconstruct incomplete samples.
- APADC [13] is an imputation-free method that projects the available view data into a common feature space, achieving cluster exploration and distribution alignment through maximizing mutual information and minimizing mean discrepancy, respectively.
- GIMVC [56] utilizes existing feature and graph information for clustering by two modules of feature learning and graph learning.
- ICMCV [10] addresses missing entries by combining multi-view consist relation transfer with graph convolutional networks, and integrates attention fusion, high-confidence guiding, and contrastive learning to explore instance complementarity and consistency.
- NBIMVC [41] leverages the neighboring features to complete missing values, and subsequently aligns cross-view features and semantic consistency for clustering.
- DCG [44] proposes a diffusion contrastive generation method by integrating a diffusion model to complete missing data with contrastive alignment.

Following the convention, ACCuracy (ACC), Normalized Mutual Information (NMI), and Adjusted Rand Index (ARI) are employed to evaluate the clustering performance of DIMVCSL in comparison with the aforementioned methods. Accordingly, all the metrics with higher values represent superior clustering performance. To mitigate the impact of experimental randomness, we repeat all experiments on the six datasets three times and record the averaged results for comparison.

4.3 Network Implementation

The proposed DIMVCSL is coded in Python under the PyTorch framework, optimized with the Adam optimizer, and trained on an NVIDIA GeForce RTX 4090 GPU. For all datasets, we utilize three-layer fully connected networks to construct the encoders and decoders. During pre-training, the encoder for each view is structured as $\mathbf{X}^v - \mathbf{F}_{512} - \mathbf{Z}_{256}^v$ and the corresponding decoder is formulated as $\mathbf{Z}_{256}^v - \mathbf{F}_{512} - \hat{\mathbf{X}}^v$. During fine-tuning, the encoder maintains the same architecture, while the decoder is modified to $(\mathbf{C}\tilde{\mathbf{Z}})_{256}^v - \mathbf{F}_{512} - \hat{\mathbf{X}}^v$. Here, 512 and 256 denote the feature dimensions for each view, respectively. The number of K in the missing feature imputation module and the τ in self-expression contrastive alignment module are empirically fixed at 5 and 0.1 across all datasets. In the self-expression attentive fusion module, the convolutional kernel k is assigned as 3×3 to effectively explore cross-view complementary information. The trade-off parameters λ_1 and λ_2 are selected from $\{0.01, 0.01, 0.1, 1, 10, 100\}$. Throughout all the experiments, we fix the pre-training epochs at 2000 and adopt 0.001 as the learning rate, while the fine-tuning epochs vary across different datasets.

4.4 Experiments Results and Analysis

We comprehensively evaluate DIMVCSL against SOTA methods on six multi-view datasets, using three metrics and under various missing rates. Table 2 reports the average clustering results over three random

runs. In comparison with the baseline methods, we draw the following observations and analyses: Our proposed DIMVCSL method outperforms existing state-of-the-art IMVC methods across most missing rate settings on all datasets. Compared to suboptimal clustering results, DIMVCSL averagely improves the performance by about 0.53% on MSRC-V1, 7.88% on LandUse-21, 20.47% on Synthetic3d, 14.79% on CUB, and 20.22% on HandWritten dataset, respectively. Especially, DIMVCSL yields remarkable clustering performance improvements on the Synthetic3d, CUB, and HandWritten datasets, demonstrating its superior clustering capability. Taking Synthetic3d as an example, in comparison with the suboptimal results, our DIMVCSL method improves the ACC, NMI, and ARI by 4.57%, 15.67%, 14.59%, 6.45%, 23.13%, 20.65%, 9.02%, 33.01%, 29.22%, 10.75%, 42.85%, and 35.81% with the missing rates varying from 0.1 to 0.7. The significant performance improvement demonstrates the effectiveness of our proposed method, which can be attributed to three core designs: (1) The missing features imputation module leverages mean features among the similar sample space based on topological relationships to dynamically complete the missing features, enabling efficient imputation while minimizing the introduction of highly noisy information. (2) Compared with conventional clustering pipelines based on k-means clustering over the latent features, the proposed self-expression contrastive alignment module can explore the representation relationship among samples for the clustering tasks, which is more consistent with the intrinsic clustering objective of grouping similar samples into the same cluster. Furthermore, contrastive alignment of self-expression representations effectively suppresses the interference from view-specific information and imputation noise, thus enhancing cross-view consistency. (3) Finally, the self-expression attentive fusion module can effectively capture and integrate cross-view complementary information, further boosting the final clustering performance.

Table 2: Clustering results comparison on different datasets under various missing rate. The best results are highlighted in bold and the second in underline.

Dataset	Missing Rates	0.1			0.3			0.5			0.7		
	Evaluation Metrics	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
MSRC-V1	Completer (2021)	66.48	54.91	41.67	61.05	51.47	35.79	55.71	47.62	28.02	55.14	47.59	26.34
	DCP (2022)	56.95	45.48	35.55	48.19	35.44	24.00	35.71	24.69	11.07	30.57	21.30	6.40
	ProImp (2023)	68.57	56.96	47.13	62.86	54.91	42.49	60.76	50.65	38.14	62.57	51.62	39.59
	APADC (2023)	71.43	61.73	50.92	72.38	62.11	51.57	70.95	60.10	49.95	66.51	54.66	43.62
	GIMVC (2024)	<u>81.62</u>	72.77	65.88	<u>81.24</u>	72.54	65.71	76.14	<u>72.52</u>	<u>63.80</u>	74.86	<u>67.82</u>	<u>58.51</u>
	ICMCV (2024)	89.14	<u>79.99</u>	77.24	83.62	<u>74.21</u>	69.79	78.95	69.80	62.33	58.95	51.41	38.61
	NBIMVC (2025)	56.67	45.99	34.99	51.90	45.55	31.50	57.62	43.32	31.44	50.48	41.09	27.48
	DCG (2025)	75.24	62.36	54.77	69.05	57.04	45.91	45.24	34.18	21.52	54.76	41.88	27.85
	DIMVCSL (Ours)	77.46	85.65	<u>75.39</u>	72.38	81.71	<u>67.09</u>	<u>78.41</u>	80.55	69.76	<u>71.43</u>	74.14	59.72
LandUse-21	Completer (2021)	26.93	32.50	13.76	25.86	31.52	11.22	21.90	27.89	11.28	23.92	28.70	10.34
	DCP (2022)	28.00	32.02	14.61	26.96	30.37	13.77	27.03	29.63	13.33	24.82	27.25	11.24
	ProImp (2023)	23.94	27.43	10.47	23.62	27.53	10.23	22.51	25.78	9.55	23.60	26.22	10.59
	APADC (2023)	19.92	24.25	7.72	17.17	23.55	4.76	17.46	23.09	4.50	17.91	21.79	4.73
	GIMVC (2024)	25.83	32.37	11.36	24.20	29.96	9.79	22.62	26.66	8.09	20.07	24.05	5.98
	ICMCV (2024)	26.63	30.68	13.76	25.69	29.55	12.72	25.52	27.24	11.08	23.31	24.71	9.19
	NBIMVC (2025)	<u>29.95</u>	<u>35.33</u>	14.20	<u>29.90</u>	<u>34.21</u>	<u>13.95</u>	<u>27.95</u>	<u>32.88</u>	12.22	<u>27.81</u>	31.87	<u>12.63</u>
	DCG (2025)	27.52	31.36	<u>14.57</u>	27.33	32.09	14.47	25.76	29.07	<u>13.17</u>	25.14	27.23	11.85
	DIMVCSL (Ours)	30.78	43.97	9.86	31.38	40.79	12.59	30.30	42.72	10.35	34.24	41.86	17.92
Scene-15	Completer (2021)	40.80	43.23	23.47	40.26	43.28	24.17	38.93	41.89	22.66	37.27	40.00	21.43
	DCP (2022)	38.92	42.70	22.70	38.35	42.12	21.98	39.19	41.90	23.00	37.62	39.67	20.96
	ProImp (2023)	42.55	44.53	<u>26.52</u>	41.16	42.69	24.90	<u>42.61</u>	<u>43.18</u>	<u>25.56</u>	41.57	41.42	24.25
	APADC (2023)	<u>43.59</u>	<u>45.04</u>	26.28	42.82	43.59	24.99	40.59	42.39	23.97	<u>39.68</u>	<u>42.05</u>	<u>23.89</u>
	GIMVC (2024)	39.82	43.93	25.34	39.40	13.13	24.49	38.85	41.52	23.24	36.78	39.30	21.39
	ICMCV (2024)	38.55	36.44	21.65	37.27	34.80	20.46	31.13	27.82	14.83	25.83	24.56	11.73
	NBIMVC (2025)	43.30	44.25	25.50	<u>44.19</u>	<u>44.53</u>	<u>26.51</u>	43.32	43.48	25.59	37.50	40.23	22.37

(Continued)

Table 2 (continued)

Dataset	Missing Rates	0.1			0.3			0.5			0.7		
	Evaluation Metrics	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
Synthetic3d	DCG (2025)	37.03	33.87	19.95	32.42	29.63	16.62	32.06	28.64	15.97	30.01	25.50	13.45
	DIMVCSL (Ours)	45.17	46.65	27.11	49.14	58.01	38.71	35.41	35.12	17.79	37.28	43.51	22.16
	Completer (2021)	90.93	71.24	75.48	88.27	64.89	69.19	86.17	60.68	64.50	68.90	45.71	43.91
	DCP (2022)	91.46	71.47	76.46	87.47	64.22	67.55	67.56	85.73	58.98	81.50	52.75	53.39
	ProImp (2023)	87.70	64.55	68.11	83.97	57.30	60.05	82.73	53.63	57.33	79.04	46.56	49.77
	APADC (2023)	85.73	59.07	62.98	86.47	60.84	64.90	84.38	58.48	61.85	80.53	49.70	53.42
	GIMVC (2024)	<u>92.17</u>	<u>73.45</u>	<u>78.13</u>	<u>89.50</u>	<u>66.64</u>	<u>71.67</u>	88.17	64.23	68.45	83.17	54.05	57.75
	ICMCV (2024)	85.60	59.45	63.15	87.33	61.86	66.79	84.23	55.40	59.68	70.70	41.72	40.35
	NBIMVC (2025)	73.17	36.20	36.57	63.83	25.24	24.65	63.40	19.68	21.04	61.67	18.21	18.21
CUB	DCG (2025)	88.00	64.49	68.68	88.67	65.96	70.06	<u>88.67</u>	<u>64.61</u>	<u>69.87</u>	<u>86.83</u>	<u>60.06</u>	<u>65.53</u>
	DIMVCSL (Ours)	96.39	84.96	89.53	95.28	82.06	86.47	96.67	85.94	90.29	96.17	85.80	89.00
	Completer (2021)	55.40	60.61	44.25	48.90	62.81	41.51	47.23	58.64	39.47	46.97	58.00	37.25
	DCP (2022)	62.27	66.74	47.35	56.30	65.17	44.41	38.30	43.49	16.23	29.50	25.76	6.76
	ProImp (2023)	77.23	72.15	61.46	78.63	71.08	61.30	72.23	65.28	53.11	64.57	61.61	48.35
	APADC (2023)	69.61	67.12	52.52	67.94	67.74	52.56	67.00	63.54	48.74	66.72	61.11	46.29
	GIMVC (2024)	71.70	72.61	58.82	66.72	68.42	54.72	65.85	64.62	49.92	59.98	59.13	40.86
	ICMCV (2024)	30.80	31.46	15.63	44.00	39.75	23.63	43.30	39.00	22.71	38.43	36.31	20.41
	NBIMVC (2025)	<u>84.67</u>	<u>77.51</u>	<u>70.14</u>	<u>83.50</u>	<u>75.14</u>	<u>68.13</u>	<u>77.50</u>	68.84	<u>59.27</u>	<u>74.67</u>	<u>66.79</u>	<u>56.62</u>
HandWritten	DCG (2025)	82.33	<u>77.70</u>	69.21	77.17	71.35	59.85	75.50	<u>72.21</u>	59.12	74.67	<u>70.19</u>	56.41
	DIMVCSL (Ours)	94.67	91.52	89.16	85.42	82.85	73.36	85.35	82.34	73.84	84.09	82.04	73.24
	Completer (2021)	79.30	82.01	72.39	82.55	82.00	73.35	77.15	77.40	62.79	80.03	77.61	69.92
	DCP (2022)	71.33	77.10	59.53	73.54	73.64	54.40	74.25	73.29	55.03	63.23	65.43	37.60
	ProImp (2023)	82.22	80.00	72.23	79.05	79.18	71.49	81.47	77.86	71.14	80.03	76.24	69.65
	APADC (2023)	69.38	73.95	53.46	72.10	76.88	60.58	67.73	75.32	58.95	62.45	68.14	51.37
	GIMVC (2024)	90.30	82.08	79.96	85.24	76.52	72.92	81.21	70.89	66.16	74.00	66.28	59.57
	ICMCV (2024)	83.50	81.78	75.50	82.22	79.88	72.70	74.95	71.83	63.05	73.03	70.44	60.40
	NBIMVC (2025)	<u>91.05</u>	<u>82.75</u>	<u>81.21</u>	<u>89.35</u>	80.11	<u>78.09</u>	<u>88.35</u>	<u>78.64</u>	<u>76.34</u>	<u>84.70</u>	73.77	69.80
HandWritten	DCG (2025)	80.75	81.31	73.57	83.00	<u>82.71</u>	76.83	75.00	76.15	67.16	83.00	76.99	<u>70.63</u>
	DIMVCSL (Ours)	99.15	97.91	98.12	99.02	97.55	97.83	99.52	98.70	98.93	98.20	95.57	96.03

To further verify the superiority of the DIMVCSL method, Fig. 2 illustrates the NMI metric changing curves of all competing methods when missing rates span 0.1 to 0.7. From the changing trend of clustering results, the following observations can be summarized:

- (1) Without loss of generality, the metric results of the competing methods almost decline with the growth of missing rate. This indicates that the reduction of available information in views will cause the clustering performance to deteriorate. In contrast, compared with the baseline methods, the performance degradation of DIMVCSL is significantly smoother, and it still maintains highly competitive performance even under a high missing rate. For instance, at a missing rate of 0.7, the NMI metric achieved by DIMVCSL surpasses the suboptimal results by 9.26%, 31.34%, 3.47%, 42.85%, 20.78%, and 23.14% on the six datasets, respectively, further demonstrating the robustness and stability of our proposed method.
- (2) However, it can be seen that the clustering metric at a higher missing rate outperforms that at a lower missing rate in some cases. This phenomenon has also been reported in other methods [43]. Through in-depth analysis, this observation is reasonably attributed to the inconsistency of data quality among different views. In the cases, the missing samples may correspond to low-quality or noisy instances, while the remaining samples retain higherquality information. Benefiting from our proposed missing

features imputation module, the missing values achieves high-quality completion of missing data from the reserved samples.

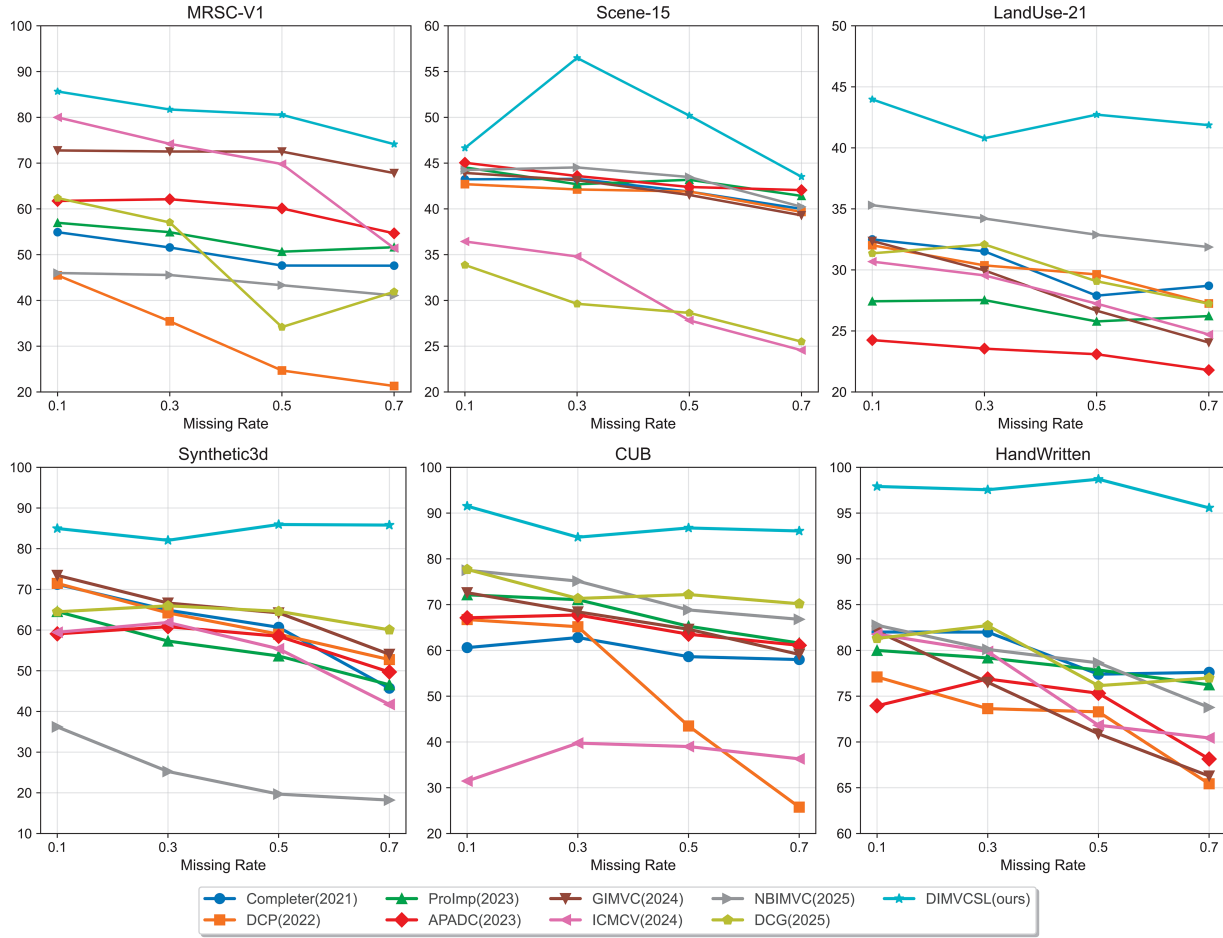


Figure 2: The changing trend of NMI metric with different missing rates on the six datasets.

4.5 Ablation Experiments

This section performs ablation studies to evaluate the effectiveness of each module in DIMVCSL, namely the KNN-based missing imputation (KMI), the self-expression contrastive alignment (SCA), and the self-expression attentive fusion (SAF). The ablation experimental results on the CUB dataset varying r missing rate of 0.1, 0.3, 0.5, and 0.7 are summarized in Table 3. Here, (A) denotes the basic model optimized only by reconstruction loss and self-expression learning, in which the average of multiple self-expression coefficient matrices serves as the cross-view representation. To evaluate the contribution of each module, KMI, SCA, and SAF are incorporated separately or jointly into the basic model. As shown in the clustering results of (B), (C), and (D), we can see that each module yields a clear performance improvement. Specifically, at a missing rate of 0.1, the performance improvement by combining SAF is greater than that of the other two modules. At missing rates from 0.3 to 0.7, KMI contributes the most significant improvement boost, demonstrating its effectiveness in imputing missing values across different views. From the results of (E), (F), and (G), we find that the pairwise integration of any two modules also effectively enhances clustering performance. When the three modules are jointly integrated into the basic model as in (H), DIMVCSL achieves the optimal clustering performance.

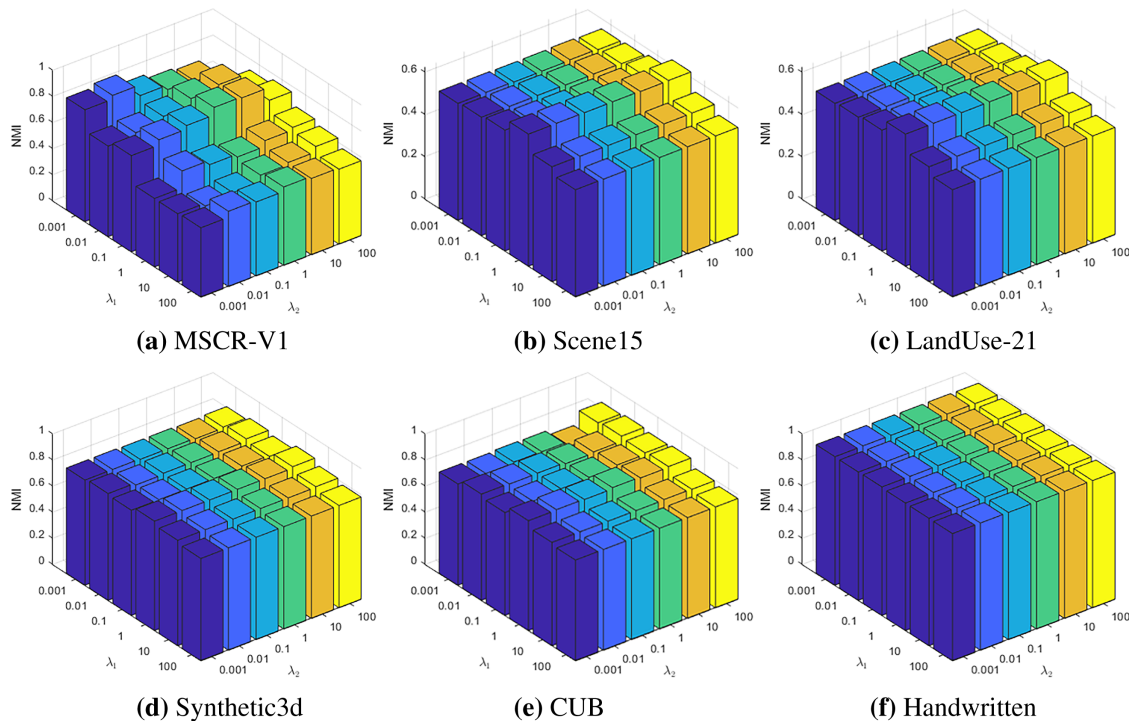
Table 3: Clustering results of ablation experiments on CUB dataset. The best results are highlighted in bold and “✓” signifies active components.

Model	KMI	SCA	SAF	0.1			0.3			0.5			0.7		
				ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
(A)				76.87	81.39	68.46	74.78	80.15	66.98	70.28	78.92	64.32	69.57	78.11	63.53
(B)	✓			77.61	82.10	69.38	78.94	82.69	71.59	77.83	81.95	69.97	76.80	81.39	69.08
(C)		✓		76.72	81.46	68.56	74.78	80.16	67.13	71.08	79.14	65.03	70.26	78.25	64.16
(D)			✓	94.89	91.84	89.44	79.50	76.31	65.98	77.36	74.89	63.86	77.67	76.40	65.54
(E)	✓	✓		77.67	82.15	69.45	79.24	81.67	71.25	78.10	81.78	70.76	77.78	82.34	70.68
(F)	✓		✓	95.22	92.51	90.20	80.39	79.76	69.48	79.81	79.97	69.49	80.63	81.26	70.88
(G)		✓	✓	94.94	91.97	89.56	79.28	75.87	65.91	77.81	74.86	64.46	78.87	76.38	66.39
(H)	✓	✓	✓	95.28	92.56	90.29	83.22	81.09	71.51	80.39	82.50	72.36	83.56	84.78	75.89

4.6 Parameter Sensitivity Analysis

(1) Trade-off parameters

The objective loss of DIMVCSL consists of two trade-off parameters, λ_1 and λ_2 , which are designed to balance the contributions of different modules in DIMVCSL. To evaluate their impact on clustering performance, λ_1 and λ_2 are tuned over a wide range of $\{0.001, 0.01, 0.1, 1, 10, 100\}$ by a grid search strategy. Fig. 3 depicts the NMI performance metric across the six datasets under the varying parameter settings. We can find that the clustering results remain relatively stable with respect to λ_1 and λ_2 within the search range. Specifically, the NMI metric consistently maintains a high level without severe fluctuations, indicating that DIMVCSL is insensitive to small or moderate changes in the trade-off parameters. Such low parameter sensitivity demonstrates the strong robustness and reduces the practical difficulty of precise parameter tuning. Across all six datasets, the parameter sensitivity patterns are highly consistent, which further validates the generalization and robustness of DIMVCSL.

**Figure 3:** Sensitivity analysis with respect to the trade-off parameters λ_1 and λ_2 .

(2) Temperature parameter

We analyze the sensitivity of the contrastive temperature parameter τ on the Synthetic3d and CUB datasets, where τ is tuned from 0.1 to 1.0. As shown in Fig. 4, the ACC, NMI and ARI of DIMVCSL remain relatively stable across the whole range, proving the strong robustness of our method with respect to variations of τ . Specifically, an overly small τ results in over-separated subspace representation, whereas too large τ blurs the boundary between positive and negative samples. According to the experimental results on two benchmarks, we adopt a moderate temperature value $\tau = 0.1$ as the default configuration in our experiments.

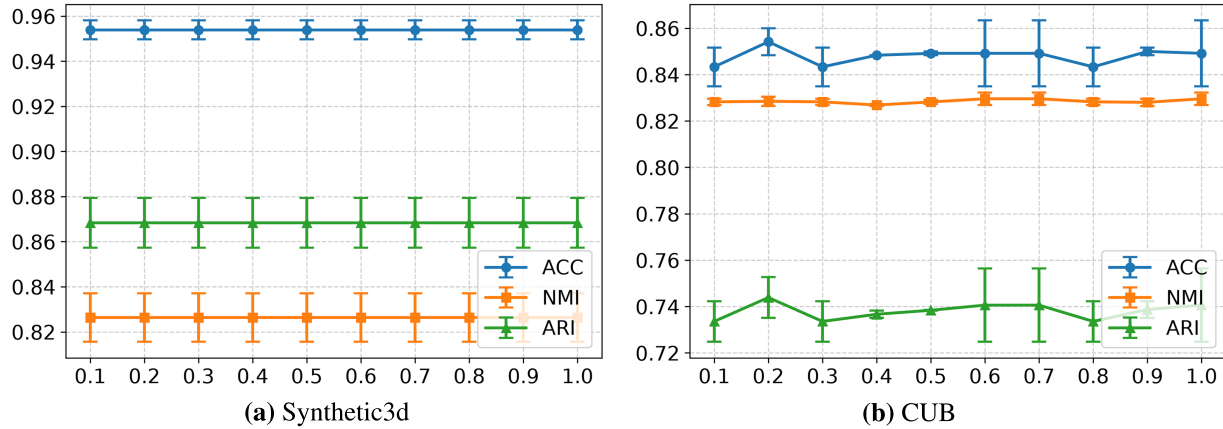


Figure 4: Sensitivity analysis with respect to the temperature parameter τ .

(3) Number of Neighbors

DIMVCSL leverages KNN to impute missing view features. We conduct sensitivity experiments on Synthetic3d and CUB to study how neighbor number k affects clustering, tuning k from 2 to 10 and evaluating ACC, NMI and ARI, which illustrated in Fig. 5. On Synthetic3d, metrics stay consistently high with minor fluctuations. For the complex CUB dataset, performance declines initially and then rises. A small k yields noisy, distorted imputed representations due to insufficient neighbors. Moderately increasing k improves feature recovery, while an excessively large k brings irrelevant samples and over-smooths features. Accordingly, we adopt $k = 5$ as the default setting.

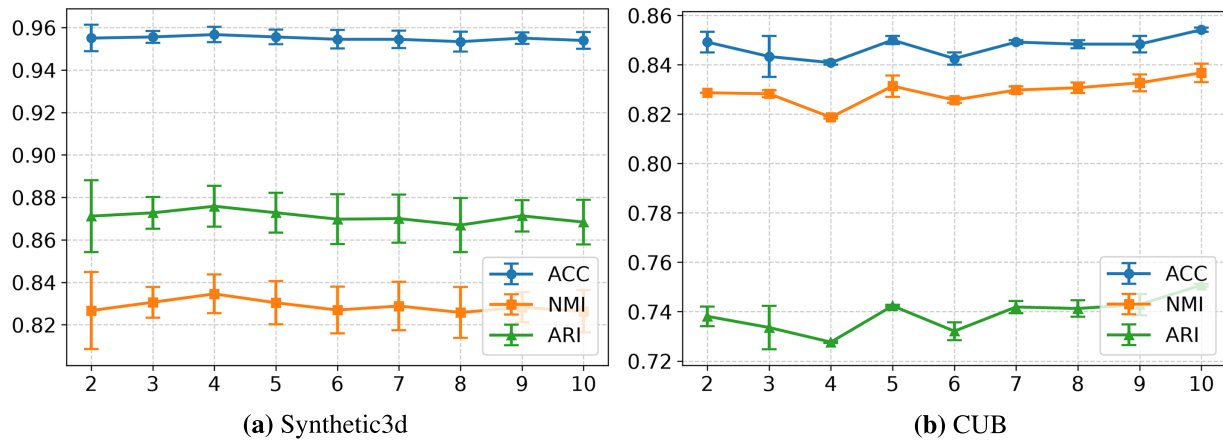


Figure 5: Sensitivity analysis with respect to the number of neighbors.

4.7 Convergence and Stability Analysis

This subsection investigates the convergence and stability characteristics of DIMVCSL through visualizing its training dynamics. Fig. 6 depicts the evolution of performance metrics and the logarithmic loss with training iterations on the CUB dataset under varying missing rates. In the first 100 training iterations, the objective loss drops sharply and approaches convergence, whereas the clustering metrics rise rapidly and eventually stabilize. Notably, the NMI metric exhibits the highest stability among the three performance metrics. This indicates that DIMVCSL can quickly capture the essential patterns of multi-view data and optimize the objective. After some iterations, the curves of ACC, NMI, ARI, and logarithmic loss tend to flatten with only minor fluctuations. This demonstrates that DIMVCSL converges to a stable optimal solution without significant oscillations or performance degradation. Even at the high missing rate of 0.7, the convergence trend remains consistent. The clustering metrics still achieve high values, and the loss decreases steadily, proving that DIMVCSL can effectively handle incomplete multi-view data.

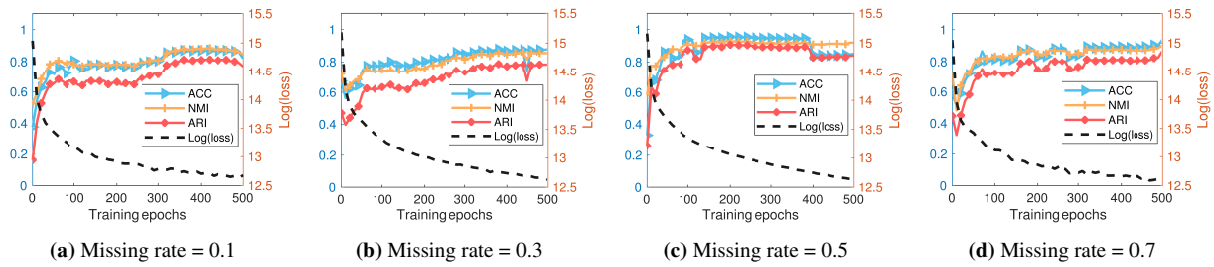


Figure 6: The results and loss curves of DIMVCSL on CUB dataset under varying missing rates.

Furthermore, Fig. 7 presents the corresponding training dynamics on the six datasets at the fixed missing rate of 0.5. As illustrated in these figures, the convergence and stability patterns are highly consistent across the six distinct datasets, confirming that DIMVCSL has strong generalization ability and stable performance on diverse multi-view clustering tasks.

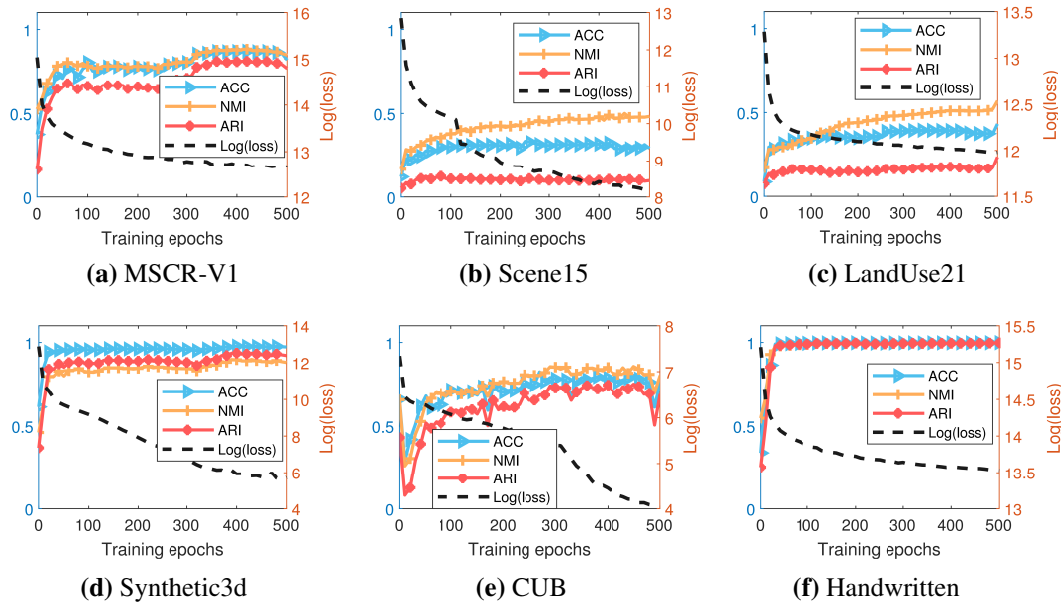


Figure 7: Clustering results and loss curves on the six datasets under the missing rate of 0.5.

5 Conclusion

The paper proposes a deep subspace clustering model for incomplete multi-view data with missing values. This model is a unified end-to-end framework that unifies KNN-based missing feature imputation, cross-view contrastive consistency alignment, and attentive-driven complementarity fusion. This integrated design enables the efficient recovery of missing latent features through cross-view neighbors in a self-supervised manner. More importantly, by enforcing view-invariant structural similarity with contrastive alignment, our model effectively suppresses the interference of imputation noise. Additionally, the proposed attentive fusion in the self-expression space can exploit cross-view complementary information. Comprehensive experimental results across six datasets confirm that DIMVCSL outperforms state-of-the-art IMVC methods under most missing rates, thereby demonstrating its effectiveness and generality. Despite its strong performance, DIMVCSL has several limitations. First, KNN-based imputation relies on sufficient common complete samples and may degrade under structured missing patterns. Second, imputation quality is noise-sensitive, as errors propagate through neighbor-based recovery. Finally, the pairwise computation of the self-expression matrix limits scalability to large datasets. Future research will tackle these issues by exploring distribution-level imputation (e.g., optimal transport), robust confidence-weighted recovery, and scalable anchor-based or hashing methods.

Acknowledgement: Not applicable.

Funding Statement: This research is supported by the Xizang Science and Technology Program (Grant No. XZ202501ZY0060), and the Scientific Research Fund of Southwest University of Science and Technology (Grant No. 25zx7162).

Author Contributions: Jiao Wang: methodology, software, funding acquisition, writing—original draft. Tingting Song: supervision, writing—review & editing. Yunhui Zhou: validation, writing—review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wen J, Zhang Z, Fei L, Zhang B, Xu Y, Zhang Z, et al. A survey on incomplete multiview clustering. *IEEE Trans Syst Man Cybern Syst.* 2022;53(2):1136–49. doi:10.1109/TSMC.2022.3192635.
2. Xu C, Guan Z, Zhao W, Wu H, Niu Y, Ling B. Adversarial incomplete multi-view clustering. In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*; 2019 Aug 10–16; Macao, China. p. 3933–9. doi:10.24963/ijcai.2019/546.
3. Zhang C, Cui Y, Han Z, Zhou JT, Fu H, Hu Q. Deep partial multi-view learning. *IEEE Trans Pattern Anal Mach Intell.* 2020;44(5):2402–15. doi:10.1109/tpami.2020.3037734.
4. Xu C, Liu H, Guan Z, Wu X, Tan J, Ling B. Adversarial incomplete multiview subspace clustering networks. *IEEE Trans Cybern.* 2022;52(10):10490–503. doi:10.1109/tcyb.2021.3062830.
5. Pu J, Cui C, Chen X, Ren Y, Pu X, Hao Z, et al. Adaptive feature imputation with latent graph for deep incomplete multi-view clustering. *Proc AAAI Conf Artif Intell.* 2024;38(13):14633–41. doi:10.1609/aaai.v38i13.29380.
6. Ou S, Xue Z, Li Y, Liang M, Cai Y, Wu J. View-category interactive sharing transformer for incomplete multi-view multi-label learning. In: *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2024 Jun 16–22; Seattle, WA, USA. p. 27457–66. doi:10.1109/cvpr52733.2024.02593.

7. Wen J, Deng S, Wong W, Chao G, Huang C, Fei L, et al. Diffusion-based missing-view generation with the application on incomplete multi-view clustering. In: Proceedings of the 41st International Conference on Machine Learning; 2024 Jul 21–24; Vienna, Austria. p. 52762–78.
8. Lin Y, Gou Y, Liu Z, Li B, Lv J, Peng X. COMPLETER: incomplete multi-view clustering via contrastive prediction. In: Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN, USA. p. 11169–78. doi:10.1109/cvpr46437.2021.01102.
9. Yang M, Li Y, Hu P, Bai J, Lv J, Peng X. Robust multi-view clustering with incomplete information. IEEE Trans Pattern Anal Mach Intell. 2023;45(1):1055–69. doi:10.1109/tpami.2022.3155499.
10. Chao G, Jiang Y, Chu D. Incomplete contrastive multi-view clustering with high-confidence guiding. Proc AAAI Conf Artif Intell. 2024;38(10):11221–9. doi:10.1609/aaai.v38i10.29000.
11. Yan W, Liu K, Zhou W, Tang C. Deep incomplete multi-view clustering via dynamic imputation and triple alignment with dual optimization. IEEE Trans Circuits Syst Video Technol. 2025;35(4):3250–61. doi:10.1109/TCSVT.2024.3508785.
12. Xu J, Li C, Ren Y, Peng L, Mo Y, Shi X, et al. Deep incomplete multi-view clustering via mining cluster complementarity. Proc AAAI Conf Artif Intell. 2022;36(8):8761–9. doi:10.1609/aaai.v36i8.20856.
13. Xu J, Li C, Peng L, Ren Y, Shi X, Shen HT, et al. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. IEEE Trans Image Process. 2023;32:1354–66. doi:10.1109/TIP.2023.3243521.
14. Feng C, Li A, Xu H, Yang H, Liu X. Deep incomplete multiview clustering via local and global pseudo-label propagation. IEEE Trans Neural Netw Learn Syst. 2025;36(5):8684–98. doi:10.1109/TNNLS.2024.3411294.
15. Wang J, Wu B, Ren Z, Zhang H, Zhou Y. Multi-scale deep multi-view subspace clustering with self-weighting fusion and structure preserving. Expert Syst Appl. 2023;213(6):119031. doi:10.1016/j.eswa.2022.119031.
16. Wang J, Wu B, Ren Z, Zhou Y. Decomposed deep multi-view subspace clustering with self-labeling supervision. Inf Sci. 2024;653:119798. doi:10.1016/j.ins.2023.119798.
17. Wang J, Zhou F, Ren Z. Dynamic anchor-based deep multi-view subspace clustering network. Neurocomputing. 2025;639(1):130290. doi:10.1016/j.neucom.2025.130290.
18. Feng W, Sheng G, Wang Q, Gao Q, Tao Z, Dong B. Partial multi-view clustering via self-supervised network. Proc AAAI Conf Artif Intell. 2024;38(11):11988–95. doi:10.1609/aaai.v38i11.29086.
19. Aghayengejeh NP, Balafar MA, Tanha J, Selver MA. Beyond predefined clusters: a comprehensive review of clustering methods for unknown numbers of clusters. IEEE Trans Knowl Data Eng. 2026;38(7):4121–38. doi:10.1109/tkde.2026.3680286.
20. Li SY, Jiang Y, Zhou ZH. Partial multi-view clustering. Proc AAAI Conf Artif Intell. 2014;28(1):1968–74. doi:10.1609/aaai.v28i1.8973.
21. Rai N, Negi S, Chaudhury S, Deshmukh O. Partial multi-view clustering using graph regularized NMF. In: Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR); 2016 Dec 4–8; Cancun, Mexico. p. 2192–7. doi:10.1109/ICPR.2016.7899961.
22. Hu M, Chen S. Doubly aligned incomplete multi-view clustering. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence; 2018 Jul 13–19; Stockholm, Sweden. p. 2262–8. doi:10.24963/ijcai.2018/313.
23. Huang Y, Shen Z, Cai Y, Yi X, Wang D, Lv F, et al. C²IMUFS: complementary and consensus learning-based incomplete multi-view unsupervised feature selection. IEEE Trans Knowl Data Eng. 2023;35(10):10681–94. doi:10.1109/tkde.2023.3266595.
24. Zhu X, Liu X, Li M, Zhu E, Liu L, Cai Z, et al. Localized incomplete multiple kernel k-means. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence; 2018 Jul 13–19; Stockholm, Sweden. p. 3271–7. doi:10.24963/ijcai.2018/454.
25. Liu X, Li M, Tang C, Xia J, Xiong J, Liu L, et al. Efficient and effective regularized incomplete multi-view clustering. IEEE Trans Pattern Anal Mach Intell. 2021;43(8):2634–46. doi:10.1109/tpami.2020.2974828.
26. Wu T, Feng S, Yuan J. Low-rank kernel tensor learning for incomplete multi-view clustering. Proc AAAI Conf Artif Intell. 2024;38(14):15952–60. doi:10.1609/aaai.v38i14.29526.

27. Wen J, Zhang Z, Xu Y, Zhang B, Fei L, Liu H. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. *Proc AAAI Conf Artif Intell.* 2019;33(1):5393–400. doi:10.1609/aaai.v33i01.33015393.
28. Sun L, Wen J, Liu C, Fei L, Li L. Balance guided incomplete multi-view spectral clustering. *Neural Netw.* 2023;166(1):260–72. doi:10.1016/j.neunet.2023.07.022.
29. Wen J, Zhang Z, Zhang Z, Zhu L, Fei L, Zhang B, et al. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. *Proc AAAI Conf Artif Intell.* 2021;35(11):10273–81. doi:10.1609/aaai.v35i11.17231.
30. Liu C, Wu S, Li R, Jiang D, Wong HS. Self-supervised graph completion for incomplete multi-view clustering. *IEEE Trans Knowl Data Eng.* 2023;35(9):9394–406. doi:10.1109/tkde.2023.3238416.
31. Wen Y, Wang S, Liang K, Liang W, Wan X, Liu X, et al. Scalable incomplete multi-view clustering with structure alignment. In: *Proceedings of the 31st ACM International Conference on Multimedia*; 2023 Oct 29–Nov 2; Ottawa, ON, Canada. p. 3031–40. doi:10.1145/3581783.3611981.
32. Du L, Shi Y, Chen Y, Zhou P, Qian Y. Fast and scalable incomplete multi-view clustering with duality optimal graph filtering. In: *Proceedings of the 32nd ACM International Conference on Multimedia*; 2024 Oct 28–Nov 1; Melbourne, VIC, Australia. p. 8893–902. doi:10.1145/3664647.3681346.
33. Tang J, Yi Q, Fu S, Tian Y. Incomplete multi-view learning: review, analysis, and prospects. *Appl Soft Comput.* 2024;153(17):111278. doi:10.1016/j.asoc.2024.111278.
34. Liu J, Liu X, Zhang Y, Zhang P, Tu W, Wang S, et al. Self-representation subspace clustering for incomplete multi-view data. In: *Proceedings of the 29th ACM International Conference on Multimedia*; 2021 Oct 20–24; Chengdu, China. p. 2726–34. doi:10.1145/3474085.3475379.
35. Han X, Ren Z, Zou C, You X. Incomplete multi-view subspace clustering based on missing-sample recovering and structural information learning. *Expert Syst Appl.* 2022;208(1):118165. doi:10.1016/j.eswa.2022.118165.
36. Liang C, Wang D, Zhang H, Zhang S, Guo F. Robust tensor subspace learning for incomplete multi-view clustering. *IEEE Trans Knowl Data Eng.* 2024;36(11):6934–48. doi:10.1109/tkde.2024.3399707.
37. Li J, Zhang X, Wang J, Sun H. Simultaneously learning representation tensor and orthogonal projection for multi-view subspace clustering. *Neurocomputing.* 2026;698(4):134255. doi:10.1016/j.neucom.2026.134255.
38. Wen J, Zhang Z, Zhang Z, Wu Z, Fei L, Xu Y, et al. DIMC-net: deep incomplete multi-view clustering network. In: *Proceedings of the 28th ACM International Conference on Multimedia*; 2020 Oct 12–16; Online. p. 3753–61. doi:10.1145/3394171.3413807.
39. Jin J, Wang S, Dong Z, Liu X, Zhu E. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In: *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2023 Jun 17–24; Vancouver, BC, Canada. p. 11600–9. doi:10.1109/CVPR52729.2023.01116.
40. Lin Y, Gou Y, Liu X, Bai J, Lv J, Peng X. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Trans Pattern Anal Mach Intell.* 2022;45(4):4447–61. doi:10.1109/tpami.2022.3197238.
41. Yan W, Zhu J, Zhou Y, Chen J, Cheng H, Yue K, et al. Neighbor-based completion for addressing incomplete multiview clustering. *IEEE Trans Neural Netw Learning Syst.* 2025;36(8):15374–84. doi:10.1109/tnnls.2025.3540437.
42. Li H, Li Y, Yang M, Hu P, Peng D, Peng X. Incomplete multi-view clustering via prototype-based imputation. In: *Proceedings of the 32nd International Joint Conference on Artificial Intelligence*; 2023 Aug 19–25; Macau, China. p. 3911–9. doi:10.24963/ijcai.2023/435.
43. Yuan H, Sun Y, Zhou F, Wen J, Yuan S, You X, et al. Prototype matching learning for incomplete multi-view clustering. *IEEE Trans Image Process.* 2025;34:828–41. doi:10.1109/tip.2025.3529378.
44. Liang Z, Yuan Y, Ma S, Wang Z. Rethinking prototype alignment for incomplete multi-view clustering. *IEEE Trans Consumer Electron.* 2026;72(2):3306–17. doi:10.1109/tce.2026.3679542.
45. Zhang Y, Lin Y, Yan W, Yao L, Wan X, Li G, et al. Incomplete multi-view clustering via diffusion contrastive generation. *Proc AAAI Conf Artif Intell.* 2025;39(21):22650–8. doi:10.1609/aaai.v39i21.34424.
46. Wang Q, Xu H, Feng W, Gao Q. Adversarial fair incomplete multi-view clustering. *Proc AAAI Conf Artif Intell.* 2026;40(12):10011–9. doi:10.1609/aaai.v40i12.37967.

47. Lin R, Du S, Wang S, Guo W. Consistent graph embedding network with optimal transport for incomplete multi-view clustering. *Inf Sci.* 2023;647(2):119418. doi:10.1016/j.ins.2023.119418.
48. Zeng P, Yang M, Lu Y, Zhang C, Hu P, Peng X. Semantic invariant multi-view clustering with fully incomplete information. *IEEE Trans Pattern Anal Mach Intell.* 2024;46(4):2139–50. doi:10.1109/tpami.2023.3332967.
49. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 7132–41. doi:10.1109/CVPR.2018.00745.
50. Winn J, Jojic N. LOCUS: learning object classes with unsupervised segmentation. In: *Proceedings of the 10th IEEE International Conference on Computer Vision*; 2005 Oct 17–21; Beijing, China. p. 756–63. doi:10.1109/ICCV.2005.148.
51. Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification. In: *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*; 2010 Nov 2–5; San Jose, CA, USA. p. 270–9. doi:10.1145/1869790.1869829.
52. Fei-Fei L, Perona P. A Bayesian hierarchical model for learning natural scene categories. In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; 2005 Jun 20–25; San Diego, CA, USA. p. 524–31. doi:10.1109/CVPR.2005.16.
53. Kumar A, Rai P, Daume H. Co-regularized multi-view spectral clustering. *Adv Neural Inf Process Syst.* 2011;24:1–9.
54. Wah C, Branson S, Welinder P, Perona P, Belongie S. The caltech-UCSD birds-200-2011 dataset. Pasadena, CA, USA: California Institute of Technology; 2011. Report No.: CNS-TR-2011-001.
55. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1989;1(4):541–51. doi:10.1162/neco.1989.1.4.541.
56. Bai S, Zheng Q, Ren X, Zhu J. Graph-guided imputation-free incomplete multi-view clustering. *Expert Syst Appl.* 2024;258(2):125165. doi:10.1016/j.eswa.2024.125165.