# SEMANTIC VIEW EVOLUTION FOR INCOMPLETE MULTI-VIEW CLUSTERING

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Incomplete multi-view clustering faces challenges from missing views and cross-view semantic redundancy. Existing methods typically perform information fusion within a fixed view space, making it difficult to continuously explore complementary semantics. To address this issue, we propose a Semantic View Evolution framework for Incomplete Multi-View Clustering (SVE_IMVC). SVE_IMVC first learns initial semantic representations through cross-view mutual information learning, and then introduces view expansion and view contraction to explore complementary semantics and suppress redundant representations, respectively, enabling the semantic view space to evolve dynamically. Based on the evolved global semantic representation, missing views are further recovered through semantic-guided completion, and the completed representations are used for final clustering. Experimental results demonstrate that SVE_IMVC effectively improves clustering performance on incomplete multi-view data.

## 1 INTRODUCTION

Multi-view clustering exploits complementary information across views to uncover underlying clustering structures (Yang & Wang, 2018; Chen et al., 2022). However, real-world multi-view data are often incomplete due to acquisition, transmission, or storage limitations (Wen et al., 2022; 2019; Hu & Chen, 2019). Such missingness disrupts cross-view relationships and complementarity, making it challenging to learn consistent and discriminative representations. This has motivated increasing efforts to exploit latent semantics for robust incomplete multi-view clustering.

Recent research on semantic learning for incomplete multi-view clustering can be broadly categorized into three directions. The first direction focuses on common semantic learning through latent representation fusion, where the latent representations of different views are aligned, aggregated, or jointly optimized to construct a unified semantic representation (Zhao et al., 2018; Ding et al., 2025; Wen et al., 2019). The second direction explores common semantic learning through structural relation fusion, which models cross-view relationships from the perspectives of graph structures, tensor structures, prototype relationships, or manifold structures, thereby capturing shared semantics beyond the original feature space (Wen et al., 2021; Li et al., 2023b; Zhao et al., 2023). The third direction emphasizes common semantic learning through complementary information mining, aiming to exploit the differences among views while reducing redundant or noisy information, so as to obtain more informative and discriminative common semantics (Xu et al., 2022; Chao et al., 2024; Wang et al., 2024b). These studies have demonstrated the effectiveness of semantic learning in improving representation quality and clustering performance for incomplete multi-view data.

Existing IMVC methods generally learn from a predefined set of observed views, treating the view space as fixed throughout the learning process rather than allowing it to evolve with the learned semantics. Recent approaches attempt to expand the semantic space by constructing additional latent or structural views. For example, CLR-IMVC (Guo & Lu, 2024a) introduces the learned latent representation as an additional view for subsequent high-order correlation modeling, while TGBSR (Guo & Lu, 2024b) constructs the clustering structure as a new view and incorporates it into tensor learning; more recently, GRETT (Guo et al., 2025) further treats the common latent representation as an additional view and jointly exploits it with the recovered views for topological tensor modeling. These studies demonstrate the effectiveness of enriching the available semantic space through the



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

construction of additional representations or views, but primarily focus on introducing new semantic representations without explicitly suppressing views whose semantics have become redundant.

Conversely, view selection and redundancy suppression methods focus on filtering unreliable or repetitive views to reduce the interference of irrelevant information. For example, MICA (Wang et al., 2025) identifies reliable views through adaptive cross-view relation learning to alleviate the impact of unreliable views on missing information recovery, while USETL (Chen et al., 2024) reduces redundancy in incomplete multi-view data through spectral embedding fusion; more recently, RNMSCC (Mao et al., 2025) further identifies and removes redundant and noisy views to improve the robustness of multi-view clustering. These studies demonstrate the effectiveness of selecting reliable information and suppressing redundant semantics from the existing views, but they mainly focus on compressing the available semantic space and cannot actively discover new complementary semantics beyond the existing representations. Therefore, existing approaches lack a unified mechanism that can dynamically coordinate semantic expansion and redundancy suppression, allowing the view space to continuously evolve according to the current semantic state.

To address the aforementioned issues, we propose Semantic View Evolution (SVE), which shifts incomplete multi-view clustering from learning in a fixed view space to learning through an evolving semantic view space. SVE alternates between View Expansion, which introduces complementary semantics, and View Contraction, which suppresses semantics already absorbed by the global representation. This evolution progressively refines a global semantic prototype that integrates informative semantics while reducing redundancy. The evolved prototype then guides missing-view completion, followed by multi-view fusion for final clustering. Our main contributions are summarized as follows:

- We propose a novel self-evolving view semantic framework for incomplete multi-view clustering, which transforms the conventional fixed set of multi-view representations into a dynamically evolving semantic view space through complementary semantic expansion and redundancy contraction.
- We develop a complementarity-driven expansion–contraction mechanism, which extracts orthogonal semantic increments from complementary views while dynamically suppressing semantics absorbed by the global prototype.
- We introduce an evolution-guided missing-view completion method, which transfers the global semantics accumulated during view evolution to missing-view reconstruction for semantically consistent completion.

## 2 Related Work

### 2.1 Common Semantic Learning via Latent Representation Fusion

Incomplete multi-view clustering commonly learns a common representation by fusing or aligning latent representations from individual views. IMC-MCL (Yin et al., 2025b) exploits multi-layer features and cross-view semantic constraints, while HLSCG-IMC (Shang et al., 2023) jointly learns latent representations and consensus structures. URRL-IMVC (Teng et al., 2024) adopts attention-based multi-view fusion, and SOI-Net (Wu et al., 2025) incorporates structural information into a unified representation space. HafGNN (Yan et al., 2025b) aggregates multi-view latent information for missing-view recovery, while LEPLM (Yin et al., 2023) and CLIMC (Liu et al., 2021) learn unified representations through latent alignment and consensus learning, respectively.

Overall, these methods mainly integrate existing latent information but lack explicit mechanisms for discovering non-redundant complementary semantics beyond the original views.

### 2.2 Common Semantic Learning via Structural Relation Fusion

Beyond latent representation fusion, some incomplete multi-view clustering methods explore common semantics through graph, tensor, prototype, and manifold structures. ETLSRR (Zhang et al., 2023) recovers cross-view similarity relationships using tensor low-rank and sparse representation, while TGBSR (Guo & Lu, 2024b) and TTGL (Zhang & He, 2023) model high-order cross-view structures with tensor frameworks. PGFormer (Du et al., 2025) exploits semantic prototypes and

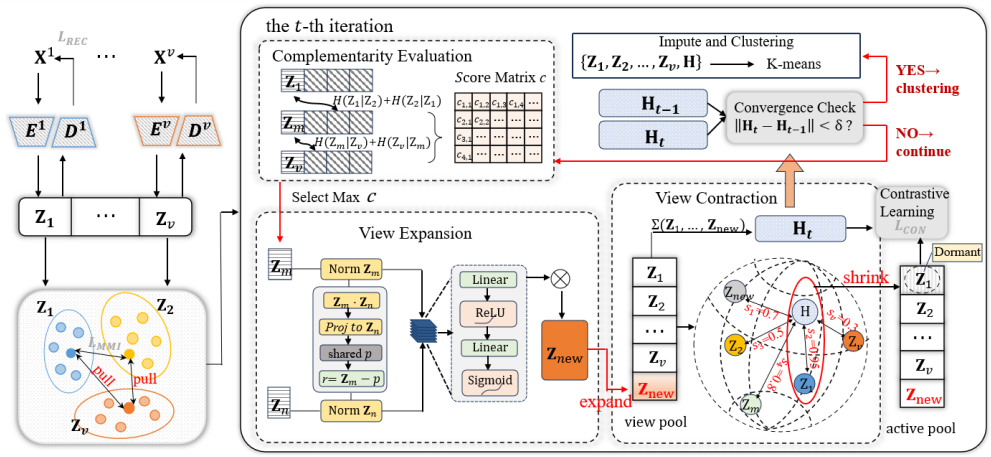



Figure 1: Overview of the proposed SVE_IMVC framework. Given incomplete multi-view data, view-specific latent representations are first learned through autoencoders and cross-view semantic learning. The semantic view set is then dynamically evolved via alternating View Expansion and View Contraction: complementary view pairs are identified to generate new semantic views from non-redundant semantic increments, while views whose information has been sufficiently absorbed by the global semantic prototype are deactivated to suppress redundancy. This evolution process iteratively updates the global semantic prototype until convergence. Finally, the evolved global semantics guide missing-view completion, and the completed multi-view representations are fused for final clustering.

prototype relations, whereas UGMCA (Ma et al., 2026) integrates low-rank tensors and adaptive graphs on the Grassmann manifold. VA-TBGIMC (Han et al., 2024) and TIMVC-IGC (Zhao et al., 2023) capture high-order cross-view correlations through tensorized graph structures, while GSRIMC (Li et al., 2022) refines multi-view graphs to enhance shared structural information.

Overall, these methods exploit structural relationships to capture global semantics beyond direct feature fusion, but mainly refine existing structures and lack explicit mechanisms for discovering and incorporating new complementary semantic components.

### 2.3 Common Semantic Learning via Complementary Information Mining

Different views contain both shared and complementary information, making it important to exploit view-specific differences while maintaining cross-view consistency. Recent IMVC methods jointly model consistency and complementarity. CCCL-IMVC (Xing et al., 2024) and DICMC (Xu et al., 2022) exploit complementary information through consensus and clustering complementarity learning, respectively. ARDGL (Yu & Sun, 2024) reduces redundant information through dynamic dimensionality reduction and weighted graph learning, while IMVC-HG (Chen et al., 2025) captures high-order cross-view relationships using hypergraph and tensor constraints. ToRES (Yu et al., 2024) models complementary and consensus features with view-specific and cross-view prototypes, while CLR-IMVC (Guo & Lu, 2024a) and MKLRRL-IMSC (Li et al., 2024) exploit complementary information through latent-space completion and low-redundancy representation learning.

However, most methods treat the view space as relatively fixed and mainly derive representations from predefined views or suppress redundancy through selection and weighting. A unified mechanism that dynamically discovers complementary semantics while suppressing redundant information remains underexplored.

# 3 METHOD

## 3.1 NOTATION

In this paper, $\left\{\mathbf{X}^v \in \mathbb{R}^{N\times d_v}\right\}_{v=1}^{V}$ represents a multi-view dataset with $V$ views, where $\mathbf{X}^v = \{\mathbf{x}_1^v, ..., \mathbf{x}_i^v, ..., \mathbf{x}_N^v\}$, $d_v$ represents the feature dimension of the $v$-th view, and $\mathbf{x}_l^v$ denotes the $l$-th sample in the $v$-th view. Additionally, we define an indicator matrix $\mathbf{A} \in \{0,1\}^{N\times V}$, whose elements are defined as follows:

$$\mathbf{A}_{iv} = \begin{cases} 1, & \text{if the } i\text{-th sample is available in the } v\text{-th view} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

In this matrix, each column represents the availability status of all instances in a specific view.

To capture view-specific structural characteristics, we employ independent autoencoders for each view. Specifically, an encoder $E^v$ maps the input $\mathbf{X}^v$ to a latent representation $\mathbf{Z}^v$, which is subsequently reconstructed as $\hat{\mathbf{X}}^v$ by a decoder $D^v$. The model is optimized via a Mean Squared Error (MSE) reconstruction loss to ensure $\mathbf{Z}^v$ effectively preserves the essential information of each view.

The overall reconstruction loss is defined as:

$$\mathcal{L}_{\text{REC}} = \sum_{v=1}^{V} \left\|\mathbf{X}^v - D^v\left(E^v\left(\mathbf{X}^v\right)\right)\right\|^2. \tag{2}$$

## 3.2 VIEW-SPECIFIC REPRESENTATION LEARNING

Let $\mathbf{z}_i^v = (z_{i,1}^v, z_{i,2}^v, \ldots, z_{i,D}^v) \in \mathbb{R}^D$ denote the latent representation of the $i$-th sample in the $v$-th view, where $v \in \{1, \ldots, V\}$. To capture cross-view feature dependencies, for each view pair $(v_1, v_2)$, we construct a cross-view dependency matrix $\mathbf{G}^{(v_1,v_2)} \in \mathbb{R}^{D\times D}$, whose $(m,n)$-th entry is defined as $G_{m,n}^{(v_1,v_2)} = \frac{1}{N}\sum_{i=1}^{N} z_{i,m}^{v_1} z_{i,n}^{v_2}$. The corresponding marginal dependencies are obtained by summing the matrix along its rows and columns, respectively.

The dependency matrix is symmetrized and normalized to form a joint distribution, with its row and column sums providing the marginal distributions. Based on the joint and marginal feature dependencies, the cross-view consistency between the view pair $(v_1, v_2)$ is modeled by the following mutual information loss:

$$\mathcal{L}_{\text{MI}}^{(v_1,v_2)} = -\sum_{m=1}^{D}\sum_{n=1}^{D} G_{m,n}^{(v_1,v_2)} \log \frac{G_{m,n}^{(v_1,v_2)}}{\left(\sum_{n'=1}^{D} G_{m,n'}^{(v_1,v_2)}\right)\left(\sum_{m'=1}^{D} G_{m',n}^{(v_1,v_2)}\right)}. \tag{3}$$

The overall cross-view consistency loss is obtained by averaging the pairwise losses over all view pairs:

$$\mathcal{L}_{\text{MMI}} = \frac{2}{V(V-1)} \sum_{1\le v_1 < v_2 \le V} \mathcal{L}_{\text{MI}}^{(v_1,v_2)}. \tag{4}$$

## 3.3 DYNAMIC EVOLUTION OF THE SEMANTIC VIEW

### 3.3.1 OBJECTIVE FORMULATION

Let the initial semantic view set be defined as

$$\mathcal{P}_0 = \{\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_V\}. \tag{5}$$

At the $t$-th iteration, the semantic view set is formulated as $\mathcal{P}_t = \{\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{M_t}\}$. where the corresponding global semantic prototype is denoted by $\mathbf{H}^{(t)}$. We aim to obtain a global semantic prototype that can maximally absorb effective information from the semantic view set while avoiding interference from redundant semantics. Accordingly, the semantic learning process can be formulated as

$$\mathbf{H}^* = \arg\max_{\mathbf{H}} \left(\mathcal{G}(\mathbf{H}, \mathcal{P}) - \mathcal{R}(\mathbf{H}, \mathcal{P})\right), \tag{6}$$

where $\mathcal{G}(\cdot)$ denotes the complementary semantic gain, and $\mathcal{R}(\cdot)$ represents semantic redundancy. Since complementary semantics do not explicitly exist in the original view set, directly optimizing Eq. (6) is challenging. Therefore, we dynamically evolve the semantic view set through *View Expansion* and *View Contraction*, while utilizing the global semantic representation to constrain the evolution process.

### 3.3.2 VIEW EXPANSION

To increase the capacity of complementary information in the semantic space, we introduce a *View Expansion* operator. Given any two semantic representations $\mathbf{Z}_i$ and $\mathbf{Z}_j$ in the current view set $\mathcal{P}_t$, we measure their complementarity using bidirectional conditional entropy:

$$S_{ij} = \mathcal{H}(\mathbf{Z}_i \mid \mathbf{Z}_j) + \mathcal{H}(\mathbf{Z}_j \mid \mathbf{Z}_i). \tag{7}$$

Here, $\mathcal{H}(\mathbf{Z}_i \mid \mathbf{Z}_j)$ denotes the conditional entropy of $\mathbf{Z}_i$ given $\mathbf{Z}_j$, and $\mathcal{H}(\mathbf{Z}_j \mid \mathbf{Z}_i)$ is defined analogously. where a larger conditional entropy indicates that the two views contain richer unique semantic information. Therefore, at each iteration, we select the view pair with the highest complementarity:

$$(i^*, j^*) = \arg\max_{i,j} S_{ij}. \tag{8}$$

For the most complementary view pair $(\mathbf{Z}_{i^*}, \mathbf{Z}_{j^*})$, we further perform orthogonal local fusion in the latent space. To avoid repeatedly accumulating shared semantics, only the non-overlapping semantic structures between the two views are involved in the fusion. Specifically, we first separate from $\mathbf{Z}_{j^*}$ the semantic component that is orthogonal to $\mathbf{Z}_{i^*}$:

$$\mathbf{R}_{j^*} = \mathbf{Z}_{j^*} - \frac{\langle \mathbf{Z}_{j^*}, \mathbf{Z}_{i^*} \rangle}{\|\mathbf{Z}_{i^*}\|_2^2} \mathbf{Z}_{i^*}, \tag{9}$$

where $\mathbf{R}_{j^*}$ represents the component of $\mathbf{Z}_{j^*}$ lying in the orthogonal complement of $\mathbf{Z}_{i^*}$.

Based on this orthogonal component, we employ a gating mechanism to adaptively regulate the fusion ratio of the orthogonal semantics. The filtered orthogonal semantic component $\mathbf{R}_{j^*}$ is then locally fused with $\mathbf{Z}_{i^*}$ to generate a new semantic view:

$$\mathbf{Z}_{\text{new}} = \mathbf{Z}_{i^*} + \mathcal{F}_{\Theta_2}\left(\text{ReLU}\left(\mathcal{F}_{\Theta_1}\left([\mathbf{Z}_{i^*}; \mathbf{R}_{j^*}]\right)\right)\right) \odot \mathbf{R}_{j^*}, \tag{10}$$

where $\mathcal{F}_{\Theta_1}(\cdot)$ and $\mathcal{F}_{\Theta_2}(\cdot)$ denote parameterized nonlinear transformations, with $\Theta_1$ and $\Theta_2$ representing their corresponding learnable parameters. This process performs information interaction only between locally complementary view pairs to avoid redundant shared semantics. Finally, the generated semantic view is added to the current view set:

$$\mathcal{P}_t \leftarrow \mathcal{P}_t \cup \{\mathbf{Z}_{\text{new}}\}. \tag{11}$$

By continuously performing *View Expansion*, higher-order complementary semantics beyond the original view set can be progressively explored, thereby continuously enhancing the complementary semantic gain term $\mathcal{G}(\mathbf{H}, \mathcal{P})$ in the objective function defined in Eq. (6).

### 3.3.3 VIEW CONTRACTION

However, as *View Expansion* proceeds, the number of views in the semantic view set continuously increases, gradually leading to redundant semantic representations. To address this issue, we further introduce *View Contraction* to dynamically compress the semantic space.

Given the current active view set $\mathcal{A}_t \subseteq \mathcal{P}_t$, the global semantic representation is defined as

$$\mathbf{H}^{(t)} = \frac{1}{|\mathcal{A}_t|} \sum_{\mathbf{Z}_v \in \mathcal{A}_t} \mathbf{Z}_v. \tag{12}$$

For each active view $\mathbf{Z}_v$, we calculate its similarity to the current global semantic representation:

$$s_v = \text{sim}\left(\mathbf{Z}_v, \mathbf{H}^{(t)}\right). \tag{13}$$

Here, $\eta$ is a predefined similarity threshold for measuring the degree of information absorption. If $s_v > \eta$, the view is considered sufficiently absorbed by the global semantic representation and is therefore placed in a dormant state and removed from the active view set:

$$\mathcal{A}_t = \left\{ \mathbf{Z}_v \in \mathcal{P}_t \mid \mathrm{sim}\left(\mathbf{Z}_v, \mathbf{H}^{(t)}\right) \leq \eta \right\}. \tag{14}$$

By performing *View Contraction*, redundant representations in the semantic space are continuously compressed, thereby reducing the semantic redundancy term $\mathcal{R}(\mathbf{H}, \mathcal{P})$ in the objective function defined in Eq. (6).

#### 3.3.4 Semantic Representation Evolution

Through the alternating processes of *View Expansion* and *View Contraction*, the semantic view set continuously evolves. After the $t$-th iteration, the global semantic prototype $\mathbf{H}^{(t)}$ is constructed from the active view set $\mathcal{A}_t$ and used as a semantic anchor to constrain the active views, thereby maintaining semantic consistency and stabilizing the evolution process:

$$\mathcal{L}_{\mathrm{CON}} = \frac{1}{|\mathcal{A}_t|} \sum_{\mathbf{Z}_v \in \mathcal{A}_t} \left\| \mathbf{Z}_v - \mathbf{H}^{(t)} \right\|_2^2. \tag{15}$$

As the iterations proceed, the global semantic representation typically becomes more stable as the semantic view set evolves. Therefore, we use the change between the global semantic representations obtained from two consecutive iterations as the convergence criterion:

$$\Delta_t = \left\| \mathbf{H}^{(t+1)} - \mathbf{H}^{(t)} \right\|_2^2. \tag{16}$$

Here, $\epsilon$ is a predefined convergence threshold ($\epsilon = 10^{-5}$). When $\Delta_t < \epsilon$, the semantic gain obtained from further view evolution becomes sufficiently small, indicating that the global semantic representation has reached a stable state. The view evolution process is therefore terminated. Let $T$ denote the iteration at which this convergence condition is first satisfied. The final global semantic representation is then given by

$$\mathbf{H}^* = \mathbf{H}^{(T)}. \tag{17}$$

Through the above process, Eq. (6) is transformed into an iterative optimization process based on the dynamic evolution of the semantic view set. *View Expansion* and *View Contraction* are used to explore complementary semantics and suppress semantic redundancy, respectively, while the semantic prototype alignment constraint maintains the consistency and stability of the semantic space. The procedure for obtaining $\mathbf{H}^*$ is provided in Appendix 1.

### 3.4 View Completion

After the dynamic evolution of the semantic view set, the global semantic prototype $\mathbf{H}^*$ integrates the shared and complementary semantics from multiple views and can serve as a global representation of the complete semantic space. For samples with missing views, relying solely on the available views makes it difficult to recover complete semantics. Therefore, $\mathbf{H}^*$ is further utilized to guide the representation recovery of missing views.

Specifically, for a missing view $v$, we first construct a local context representation from the remaining available semantic representations:

$$\mathbf{C}_v = \Psi\left(\{\mathbf{Z}_i \mid i \neq v\}\right), \tag{18}$$

where $\Psi(\cdot)$ denotes a context aggregation operator that integrates semantic information from the available views.

Subsequently, the local contextual semantics are jointly modeled with the global semantic prototype to generate the latent representation corresponding to the missing view:

$$\hat{\mathbf{Z}}_v = \Phi\left(\mathbf{C}_v, \mathbf{H}^*\right), \tag{19}$$

where $\Phi(\cdot)$ denotes a global-semantic-guided completion mapping, which utilizes the global semantic prototype to supplement the local contextual semantics and thereby recover the representation of the missing view.

The above procedure is applied to all missing views, resulting in a complete set of latent representations:

$$\mathcal{Z} = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_V\}. \tag{20}$$

Finally, the complete latent representations are fused and fed into the clustering module to obtain the final clustering results.

### 3.5 Objective Function and Optimization Algorithm

Therefore, the overall objective function is formulated as:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{REC}} + \lambda_2 \mathcal{L}_{\text{MMI}} + \lambda_3 \mathcal{L}_{\text{CON}}. \tag{21}$$

The overall training procedure of SVE_IMVC is summarized in Appendix 2.

## 4 EXPERIMENTS

### 4.1 Experimental Setup

We conduct experiments on six widely used datasets. Hdigit (Yang et al., 2021) contains 10,000 samples from 10 classes, while 100leaves (Zheng et al., 2022) contains 1,600 samples from 100 classes. MNIST_USPS (Peng et al., 2019) contains 5,000 samples from 10 classes and provides two views per sample. Handwritten (LeCun et al., 1989)contains 2,000 samples from 10 classes with six views per sample. MSRC-v1 (Winn & Jojic, 2005) contains 210 samples from 7 classes, while Fashion (Zheng et al., 2018) contains 10,000 samples from 10 classes with three views per sample.

We compare the proposed method with 13 advanced incomplete multi-view clustering approaches, which can be broadly categorized into three groups. Completion and alignment based methods include PETCAT (He et al., 2026), DIMVC-HIA (Du et al., 2026), HSACC (Ding et al., 2025), DITA-IMVC (Yan et al., 2025a), DCG (Zhang et al., 2025), IMC-MCL (Yin et al., 2025a), MICA (Wang et al., 2025), and IMVC-IE (Huang et al., 2024), which focus on recovering missing views through prototype alignment, contrastive learning, feature completion, or generative modeling. Representation and relation learning methods include DIVIDE (Lu et al., 2024), MRL-CAL (Wang et al., 2024a), and APADC (Xu et al., 2023), which learn reliable multi-view representations or cross-view relations through contrastive, adversarial, or mutual-information-based objectives. Fusion and prototype based methods include MCAC (Zhang & Zhu, 2023) and ProImp (Li et al., 2023a), which exploit adaptive fusion, prototype modeling, and contrastive learning to preserve shared and view-specific information.

### 4.2 Experimental Setting

All experiments are conducted based on the PyTorch 2.3.1 framework on a computing platform equipped with an NVIDIA RTX 4070 GPU and 32 GB of memory. The proposed method adopts a fully connected autoencoder architecture for all datasets, where the encoder and decoder are configured as Input–1024–1024–1024–Output and its symmetric structure, respectively. All layers use the ReLU activation function. During training, each dataset is trained for $E$ epochs (e.g., $E = 100$), with the learning rate set to 0.0001 and the batch size set to 256. Experiments are performed under multiple missing rate settings, and the results under different missing rates are reported in the main text.

### 4.3 Experimental Results and Analysis

Table 1 presents the clustering comparison results under different missing rates, with the best and second-best results highlighted in bold and underlined, respectively. Overall, the proposed method achieves superior performance across multiple datasets and missing rates.

Table 1: Clustering performance (%) on six datasets under different missing rate settings.

| Ratio | Method | Hdigit | | | 100leaves | | | MNIST_USPS | | | Handwritten | | | MSRC-v1 | | | Fashion | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ACC | NMI | ARI | ACC | NMI | ARI | ACC | NMI | ARI | ACC | NMI | ARI | ACC | NMI | ARI | ACC | NMI | ARI |
| 0.1 | PETCAT | 73.43 | 57.13 | 51.51 | 44.75 | 69.57 | 27.42 | 64.62 | 53.94 | 44.98 | 78.35 | 65.20 | 58.67 | 43.33 | 26.56 | 17.61 | 63.91 | 58.20 | 44.80 |
| | DIMVC-HIA | 92.35 | 81.41 | 74.70 | 16.90 | 46.00 | 3.17 | 98.94 | 96.97 | 97.66 | 85.65 | 79.25 | 74.05 | 45.00 | 36.69 | 21.70 | 97.78 | 94.76 | 95.20 |
| | HSACC | 98.41 | 97.03 | 96.82 | 73.35 | 90.20 | 67.45 | 99.00 | 97.33 | 97.81 | 82.04 | 84.39 | 73.04 | 75.33 | 68.22 | 58.12 | 97.44 | 95.23 | 95.01 |
| | DITA-IMVC | 87.47 | 83.22 | 79.32 | 33.32 | 65.97 | 21.08 | 94.94 | 92.75 | 91.77 | 77.89 | 74.20 | 65.73 | 74.47 | 71.95 | 63.06 | 80.66 | 82.12 | 72.51 |
| | DCG | 99.00 | 96.00 | 97.00 | 23.00 | 60.00 | 14.00 | 98.00 | 95.00 | 96.00 | 82.75 | 82.63 | 74.88 | 46.00 | 38.00 | 22.00 | 90.04 | 82.25 | 79.99 |
| | IMC-MCL | 99.00 | 97.05 | 97.61 | 25.38 | 65.70 | 19.30 | 81.72 | 90.35 | 80.67 | 73.33 | 66.28 | 56.94 | 41.24 | 31.91 | 18.14 | 66.11 | 77.00 | 59.98 |
| | MICA | 98.63 | 96.19 | 97.18 | 10.06 | 33.34 | 1.33 | 98.26 | 96.22 | 96.36 | 29.25 | 41.30 | 17.32 | 26.00 | 8.71 | 3.32 | 88.16 | 85.11 | 82.73 |
| | DIVIDE | 98.45 | 95.66 | 96.59 | 74.84 | 86.81 | 63.90 | 93.26 | 90.02 | 88.96 | 87.35 | 79.16 | 75.24 | 60.86 | 50.88 | 39.65 | 92.80 | 87.35 | 85.73 |
| | MRL_CAL | 12.63 | 0.73 | 0.28 | 14.75 | 52.75 | 7.92 | 98.38 | 95.76 | 96.44 | 46.65 | 53.79 | 36.52 | 28.10 | 16.56 | 6.67 | 82.29 | 81.39 | 75.62 |
| | IMVC-IE | 90.80 | 78.87 | 80.68 | 18.31 | 52.95 | 11.31 | 91.38 | 81.43 | 82.01 | 79.25 | 72.82 | 66.98 | 40.95 | 33.13 | 18.58 | 60.17 | 43.81 | 37.08 |
| | APADC | 79.20 | 72.90 | 65.90 | 52.70 | 79.30 | 43.50 | 97.70 | 94.40 | 94.80 | 81.67 | 83.80 | 73.62 | 35.00 | 21.60 | 12.30 | 68.70 | 76.90 | 61.50 |
| | MCAC | 29.57 | 21.85 | 11.92 | 42.60 | 71.13 | 29.45 | 97.21 | 95.91 | 95.67 | 48.87 | 42.78 | 28.89 | 44.36 | 40.97 | 23.47 | 53.59 | 57.51 | 39.01 |
| | ProImp | 97.49 | 94.38 | 94.40 | 70.86 | 84.61 | 62.19 | 98.56 | 96.29 | 96.83 | 83.30 | 81.10 | 75.81 | 45.09 | 34.78 | 21.53 | 77.11 | 74.70 | 65.44 |
| | **Ours** | **98.90** | **97.29** | **97.63** | **81.17** | **90.94** | **73.23** | **99.06** | **97.53** | **97.93** | **95.86** | **92.18** | **91.84** | **76.19** | **72.09** | **65.66** | **98.23** | **96.10** | **96.19** |
| 0.3 | PETCAT | 60.59 | 50.13 | 39.29 | 39.00 | 63.86 | 16.23 | 58.16 | 46.90 | 35.86 | 58.42 | 50.68 | 40.53 | 35.24 | 18.86 | 9.64 | 59.21 | 47.80 | 36.00 |
| | DIMVC-HIA | 91.67 | 80.73 | 82.39 | 14.31 | 42.81 | 3.48 | 97.58 | 93.86 | 94.72 | 80.20 | 75.90 | 68.72 | 43.50 | 33.88 | 21.91 | 94.56 | 88.67 | 88.61 |
| | HSACC | 97.49 | 96.68 | 96.77 | 72.08 | 86.25 | 65.44 | 98.36 | 96.11 | 96.45 | 79.91 | 83.60 | 71.58 | 72.95 | 64.75 | 52.78 | 96.85 | 93.34 | 94.08 |
| | DITA-IMVC | 84.88 | 80.02 | 76.10 | 31.41 | 62.87 | 18.29 | 94.94 | 90.75 | 90.10 | 75.69 | 73.74 | 64.51 | 64.10 | 54.85 | 42.18 | 78.06 | 78.62 | 68.08 |
| | DCG | 97.00 | 93.00 | 94.00 | 22.00 | 56.00 | 11.00 | 97.00 | 92.00 | 93.00 | 82.70 | 80.54 | 73.96 | 50.00 | 39.00 | 25.00 | 93.13 | 86.99 | 86.00 |
| | IMC-MCL | 98.56 | 95.85 | 96.82 | 24.30 | 62.40 | 16.51 | 82.92 | 88.40 | 80.57 | 72.84 | 65.36 | 55.99 | 43.80 | 37.55 | 21.30 | 67.03 | 79.26 | 62.14 |
| | MICA | 69.49 | 82.65 | 70.13 | 9.88 | 34.26 | 2.15 | 69.00 | 79.53 | 66.87 | 31.00 | 46.41 | 22.71 | 24.00 | 10.20 | 3.57 | 88.53 | 89.63 | 83.62 |
| | DIVIDE | 97.40 | 92.98 | 94.32 | 75.30 | 86.50 | 62.61 | 94.18 | 87.25 | 87.62 | 87.98 | 79.46 | 76.21 | 55.43 | 45.36 | 33.02 | 90.28 | 85.01 | 90.70 |
| | MRL_CAL | 12.30 | 0.69 | 0.23 | 12.88 | 50.63 | 7.31 | 96.72 | 91.53 | 92.86 | 50.45 | 54.72 | 37.35 | 25.71 | 16.03 | 5.31 | 91.31 | 86.86 | 83.61 |
| | IMVC-IE | 83.88 | 67.21 | 67.65 | 14.37 | 43.75 | 7.33 | 80.94 | 64.10 | 63.06 | 73.70 | 61.18 | 56.22 | 40.12 | 30.95 | 14.28 | 84.00 | 68.95 | 68.42 |
| | APADC | 84.00 | 81.60 | 74.60 | 38.80 | 70.50 | 27.80 | 96.40 | 92.10 | 92.20 | 81.65 | 83.75 | 70.68 | 30.50 | 24.70 | 8.30 | 76.10 | 80.80 | 65.70 |
| | MCAC | 29.49 | 20.02 | 11.40 | 35.26 | 64.88 | 20.58 | 96.33 | 93.13 | 93.41 | 52.44 | 44.91 | 31.90 | 43.01 | 34.17 | 20.72 | 67.96 | 61.90 | 53.95 |
| | ProImp | 93.17 | 87.46 | 84.81 | 51.54 | 77.34 | 40.29 | 96.43 | 91.67 | 92.26 | 79.81 | 76.65 | 70.49 | 41.34 | 34.64 | 19.20 | 81.33 | 79.92 | 72.73 |
| | **Ours** | **99.10** | **97.50** | **98.03** | **78.21** | **89.41** | **69.05** | **98.96** | **97.09** | **97.71** | **96.76** | **92.89** | **92.95** | **78.29** | **69.89** | **63.71** | **97.50** | **94.81** | **94.67** |
| 0.5 | PETCAT | 52.44 | 43.14 | 24.89 | 38.87 | 62.22 | 9.37 | 45.56 | 38.48 | 19.22 | 47.90 | 45.27 | 31.95 | 29.05 | 14.66 | 4.08 | 48.59 | 42.93 | 23.20 |
| | DIMVC-HIA | 91.22 | 80.16 | 81.50 | 8.81 | 26.55 | 0.39 | 96.14 | 90.42 | 91.63 | 84.55 | 78.29 | 70.60 | 40.00 | 27.39 | 17.75 | 89.06 | 79.98 | 78.28 |
| | HSACC | 96.01 | 94.72 | 94.88 | 68.08 | 80.47 | 56.05 | 97.26 | 94.84 | 95.74 | 80.22 | 84.02 | 72.31 | 71.38 | 63.63 | 52.48 | 95.02 | 90.86 | 89.23 |
| | DITA-IMVC | 85.35 | 77.80 | 74.25 | 34.98 | 64.94 | 21.14 | 95.49 | 89.59 | 90.31 | 83.94 | 80.90 | 75.03 | 61.90 | 51.92 | 38.72 | 78.93 | 76.87 | 68.16 |
| | DCG | 95.00 | 89.00 | 90.00 | 15.00 | 45.00 | 6.00 | 84.00 | 78.00 | 72.00 | 80.80 | 76.21 | 70.45 | 48.00 | 36.00 | 21.00 | 90.04 | 82.25 | 79.99 |
| | IMC-MCL | 97.35 | 92.84 | 94.21 | 24.16 | 60.40 | 15.92 | 80.82 | 85.36 | 77.32 | 69.95 | 62.61 | 52.38 | 41.24 | 31.91 | 18.14 | 66.11 | 77.00 | 59.98 |
| | MICA | 58.86 | 72.52 | 61.93 | 9.25 | 29.25 | 1.55 | 58.68 | 73.73 | 59.71 | 29.25 | 38.12 | 16.99 | 25.00 | 11.37 | 4.74 | 88.16 | 85.11 | 82.73 |
| | DIVIDE | 95.39 | 88.50 | 90.05 | 69.45 | 83.29 | 56.33 | 89.33 | 81.82 | 80.49 | 86.85 | 77.81 | 73.65 | 54.76 | 43.83 | 31.61 | 86.69 | 79.03 | 75.33 |
| | MRL_CAL | 12.62 | 0.82 | 0.28 | 12.12 | 49.35 | 7.03 | 91.58 | 81.52 | 82.02 | 48.00 | 54.52 | 40.87 | 30.48 | 14.87 | 5.76 | 82.29 | 81.39 | 75.62 |
| | IMVC-IE | 72.22 | 50.44 | 47.97 | 11.25 | 32.60 | 3.02 | 70.80 | 49.44 | 46.50 | 64.00 | 50.52 | 42.72 | 38.57 | 22.27 | 11.45 | 60.17 | 43.81 | 37.08 |
| | APADC | 91.50 | 81.70 | 82.00 | 30.50 | 65.00 | 18.80 | 94.30 | 88.10 | 87.50 | 78.37 | 79.86 | 68.21 | 28.70 | 18.90 | 6.80 | 68.70 | 76.90 | 61.50 |
| | MCAC | 24.08 | 14.75 | 6.93 | 27.59 | 58.43 | 12.14 | 95.29 | 90.71 | 90.33 | 40.92 | 37.34 | 22.04 | 35.61 | 29.74 | 13.88 | 53.59 | 57.51 | 39.01 |
| | ProImp | 86.05 | 78.45 | 68.86 | 39.86 | 69.78 | 26.56 | 94.43 | 87.60 | 88.10 | 76.32 | 69.21 | 62.24 | 43.34 | 34.58 | 20.93 | 77.11 | 74.70 | 65.44 |
| | **Ours** | **98.36** | **95.64** | **96.42** | **70.66** | **84.83** | **57.45** | **98.22** | **95.17** | **96.09** | **95.82** | **91.42** | **90.83** | **78.29** | **68.09** | **60.77** | **96.22** | **91.99** | **91.99** |

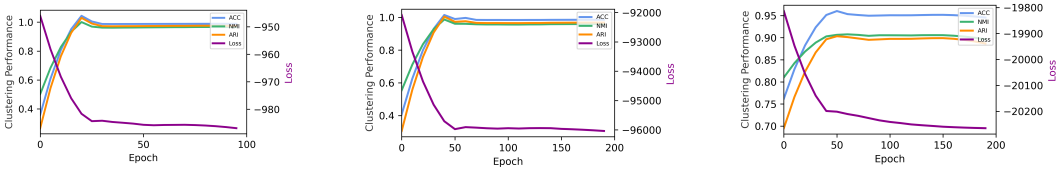
Specifically, when the missing rate is 0.1, our method consistently outperforms the second-best method HSACC on MNIST_USPS in terms of ACC, NMI, and ARI. On 100leaves, our method improves ACC by approximately 6.3 percentage points over DIVIDE. On the Handwritten and MSRC-v1 datasets, our method also achieves better results across all evaluation metrics, with an 8.51 percentage-point improvement in ACC over DIVIDE on Handwritten. As the missing rate increases to 0.3, the proposed method maintains its advantage, improving ACC by 8.78 percentage points over DIVIDE on Handwritten and by 5.34 percentage points over HSACC on MSRC-v1. At a missing rate of 0.5, the method remains competitive across multiple datasets. On Handwritten, the ACC, NMI, and ARI improvements reach 8.97%, 7.40%, and 15.80%, respectively, while the corresponding improvements on MSRC-v1 are 6.91%, 4.46%, and 8.29%. Although the performance of all methods generally decreases as the missing rate increases, the proposed method maintains relatively stable clustering performance across different missing rates, demonstrating its robustness to increasing view incompleteness.

### 4.4 Ablation Study

To evaluate the effectiveness of each loss function, we conduct ablation experiments with different loss combinations on four datasets, i.e., Hdigit, 100leaves, Handwritten, and MSRC-v1. The evaluated components include the single-view reconstruction loss $\mathcal{L}_{\mathrm{REC}}$, mutual information contrastive loss $\mathcal{L}_{\mathrm{MMI}}$, and global semantic consistency loss $\mathcal{L}_{\mathrm{CON}}$. The results are reported in Table 2. The results show that, starting from $\mathcal{L}_{\mathrm{REC}}$ alone (Item-1), introducing either $\mathcal{L}_{\mathrm{MMI}}$ (Item-2) or $\mathcal{L}_{\mathrm{CON}}$

Table 2: Loss Ablation Experiments on Four Data Sets with the Missing Rate of 0.5

| | Loss Components | | | Hdigit | | | 100leaves | | | Handwritten | | | MSRC-v1 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mathcal{L}_{\text{REC}}$ | $\mathcal{L}_{\text{MMI}}$ | $\mathcal{L}_{\text{CON}}$ | ACC | NMI | ARI | ACC | NMI | ARI | ACC | NMI | ARI | ACC | NMI | ARI |
| Item-1 | ✓ | | | 44.56 | 48.84 | 34.53 | 44.49 | 73.22 | 33.27 | 64.72 | 67.53 | 54.40 | 28.29 | 21.55 | 2.29 |
| Item-2 | | ✓ | | 96.01 | 93.27 | 92.45 | 61.19 | 72.89 | 46.40 | 81.00 | 84.88 | 72.61 | 62.38 | 56.30 | 41.99 |
| Item-3 | | | ✓ | 57.18 | 59.60 | 48.27 | 40.81 | 71.32 | 28.73 | 69.30 | 78.05 | 66.29 | 76.10 | 66.55 | 59.10 |
| Item-4 | ✓ | ✓ | | 98.06 | 94.78 | 95.16 | 69.39 | 84.79 | 55.08 | 90.64 | 85.21 | 78.58 | 63.90 | 56.24 | 43.07 |
| Item-5 | | ✓ | ✓ | 97.94 | 94.20 | 95.60 | 70.36 | 84.82 | 56.73 | 83.23 | 85.30 | 74.43 | 70.29 | 62.96 | 51.14 |
| Item-6 | ✓ | | ✓ | 56.56 | 55.10 | 42.60 | 41.65 | 71.57 | 24.63 | 64.57 | 68.08 | 58.57 | 72.48 | 65.71 | 57.09 |
| **Item-7** | ✓ | ✓ | ✓ | **98.36** | **95.64** | **96.42** | **70.66** | **84.83** | **57.45** | **95.82** | **91.42** | **90.83** | **78.29** | **68.09** | **60.77** |




(a) Hdigit (b) MNIST_USPS (c) Fashion

Figure 2: Clustering metrics (ACC, NMI, ARI) and loss curves over training epochs on the Hdigit, MNIST_USPS, and Fashion datasets.

(Item-3) substantially improves clustering performance on all four datasets. For example, on Hdigit, the ACC of Item-2 increases from 44.56% to 96.01%. Further combining $\mathcal{L}_{\text{REC}}$ and $\mathcal{L}_{\text{MMI}}$ (Item-4) leads to additional improvements, with ACC reaching 98.66% on Hdigit and 69.39% on 100leaves, indicating the complementary effects of reconstruction and cross-view mutual information learning. Adding $\mathcal{L}_{\text{CON}}$ to form the complete three-loss collaborative optimization framework (Item-7) further achieves the best performance on multiple datasets. On Handwritten and MSRC-v1, Item-7 achieves ACC values of 95.82% and 78.29%, respectively, improving by 5.18 and 14.39 percentage points over Item-4. This indicates that the global semantic consistency constraint further promotes the collaborative optimization of multi-view representations. The synergistic collaboration among the three loss functions jointly enhances the clustering discriminability of the learned multi-view representations.

### 4.5 Convergence Analysis

To verify the convergence of the proposed method, experiments were conducted on several benchmark datasets, including Hdigit, MNIST_USPS, and Fashion. The optimization process of the model was evaluated by observing the variations of the total loss and clustering metrics (ACC, NMI, and ARI) over training epochs. As illustrated in Figure 2(a)–(c), the total loss on all datasets decreases rapidly and gradually converges as training proceeds, demonstrating that SVE_IMVC exhibits good convergence performance and training stability. Meanwhile, the clustering metrics, including ACC, NMI, and ARI, consistently improve during training, indicating that the model is capable of learning more reliable and discriminative latent representations through iterative optimization.

## 5 Conclusion

In this paper, we proposed **SVE_IMVC**, a semantic view evolution framework for incomplete multi-view clustering. SVE_IMVC dynamically explores complementary semantics and suppresses redundant information through semantic view evolution, and further utilizes the evolved global semantic representation for missing-view recovery. Experimental results on multiple benchmark datasets demonstrate that SVE_IMVC achieves good clustering performance under different missing rates. Despite its effectiveness, the global semantic representation in SVE_IMVC is mainly obtained by aggregating active semantic views, leaving room for further improvement when substantial semantic differences exist among views. Future work will explore more adaptive global semantic modeling methods.

### AI USE STATEMENT

In this work, we used generative AI tools to assist with translation between Chinese and English, improve the readability and clarity of the manuscript, and refine the presentation of technical descriptions. We also used generative AI tools for code debugging assistance. All AI-assisted text and debugging suggestions were carefully reviewed and verified by the authors. The authors take full responsibility for the final content of this work, including all text, claims, experimental results, and code.

### ETHICS STATEMENT

This work does not involve human subjects, personally identifiable information, or sensitive personal data. All experiments are conducted on publicly available benchmark datasets. To the best of our knowledge, this work does not raise specific ethical, privacy, security, or legal concerns.

### REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the proposed framework, optimization objectives, training procedure, experimental settings, and dataset configurations in the main paper and supplementary materials. An anonymous version of the source code is provided as supplementary material for verification and reproducibility. The complete implementation and associated resources will be released upon acceptance.

## REFERENCES

Guoqing Chao, Yi Jiang, and Dianhui Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 11221–11229, 2024.

Jie Chen, Yingke Chen, Zhu Wang, Haixian Zhang, and Xi Peng. Spectral embedding fusion for incomplete multiview clustering. *IEEE Transactions on Image Processing*, 33:4116–4130, 2024. doi: 10.1109/TIP.2024.3420796.

Jin Chen, Huafu Xu, Jingjing Xue, Quanxue Gao, Cheng Deng, and Ziyu Lv. Incomplete multi-view clustering based on hypergraph. *Information Fusion*, 117:102804, 2025.

Man-Sheng Chen, Jia-Qi Lin, Xiang-Long Li, Bao-Yu Liu, Chang-Dong Wang, Dong Huang, and Jian-Huang Lai. Representation learning in multi-view clustering: A literature review. *Data Science and Engineering*, 7(3):225–241, 2022.

Xiaojian Ding, Lin Zhao, Xian Li, and Xiaoying Zhu. Incomplete multi-view clustering via hierarchical semantic alignment and cooperative completion. *Advances in Neural Information Processing Systems*, 38:52175–52195, 2025.

Yiming Du, Yao Wang, Ziyu Wang, Rui Ning, and Lusi Li. Pgformer: A prototype-graph transformer for incomplete multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

Yiming Du, Ziyu Wang, Jian Li, Rui Ning, and Lusi Li. Deep incomplete multi-view clustering via hierarchical imputation and alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 20941–20949, 2026.

Xiaoxing Guo and Gui-Fu Lu. Comprehensive consensus representation learning for incomplete multiview subspace clustering. *Information Sciences*, 678:120935, 2024a.

XiaoXing Guo and Gui-Fu Lu. Tensor-based global block-diagonal structure radiation for incomplete multiview clustering. *Expert Systems with Applications*, 255:124788, 2024b.

Xiaoxing Guo, Gui-Fu Lu, Yangfan Du, and Yaozu Kan. Generative views recovery and error-guided topological tensor network for incomplete multi-view clustering. *Pattern Recognition*, pp. 112750, 2025.

Xuemei Han, Fei Zhou, Zhenwen Ren, Xueyuan Wang, and Xiaojian You. View-specific anchors coupled tensorial bipartite graph learning for incomplete multi-view clustering. *Information Sciences*, 664:120335, 2024.

Wen-Jue He, Chenzi Liu, Zhongliang Luo, and Zheng Zhang. Prototype-tailored cross-view alignment for partially-aligned incomplete multi-view clustering. *Pattern Recognition*, pp. 114233, 2026.

Menglei Hu and Songcan Chen. One-pass incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 3838–3845, 2019.

Binqiang Huang, Zhijie Huang, Shoujie Lan, Qinghai Zheng, and Yuanlong Yu. Incomplete multi-view clustering via inference and evaluation. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8180–8184. IEEE, 2024.

Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

Ao Li, Cong Feng, Yuan Cheng, Yingtao Zhang, and Hailu Yang. Incomplete multiview subspace clustering based on multiple kernel low-redundant representation learning. *Information Fusion*, 103:102086, 2024.

Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. *arXiv preprint arXiv:2301.11045*, 2023a.

Lusi Li, Zhiqiang Wan, and Haibo He. Incomplete multi-view clustering with joint partition and graph learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):589–602, 2023b. doi: 10.1109/TKDE.2021.3082470.

Xiang-Long Li, Man-Sheng Chen, Chang-Dong Wang, and Jian-Huang Lai. Refining graph structure for incomplete multi-view clustering. *IEEE transactions on neural networks and learning systems*, 35(2):2300–2313, 2022.

Jianlun Liu, Shaohua Teng, Lunke Fei, Wei Zhang, Xiaozhao Fang, Zhuxiu Zhang, and Naiqi Wu. A novel consensus learning approach to incomplete multi-view clustering. *Pattern Recognition*, 115:107890, 2021.

Yiding Lu, Yijie Lin, Mouxing Yang, Dezhong Peng, Peng Hu, and Xi Peng. Decoupled contrastive multi-view clustering with high-order random walks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 14193–14201, 2024.

Xuanlong Ma, Yanhong She, Shixun Lin, Fenfang Xie, Xuhang Chen, Ting Shu, and Guo Zhong. Unified grassmann manifold-based completion and alignment for incomplete multi-view clustering. *IEEE Transactions on Consumer Electronics*, 2026.

Zhan-Wang Mao, Lu Sun, and Youlong Wu. Robust multi-view subspace clustering with missing data by aligning nonlinear manifolds. *Pattern Recognition*, 161:111280, 2025.

Xi Peng, Zhenyu Huang, Jiancheng Lv, Hongyuan Zhu, and Joey Tianyi Zhou. Comic: Multi-view clustering without parameter selection. In *International conference on machine learning*, pp. 5092–5101. PMLR, 2019.

Mingchao Shang, Cheng Liang, Jiawei Luo, and Huaxiang Zhang. Incomplete multi-view clustering by simultaneously learning robust representations and optimal graph structures. *Information Sciences*, 640:119038, 2023.

Ge Teng, Ting Mao, Chen Shen, Xiang Tian, Xuesong Liu, Yaowu Chen, and Jieping Ye. Urrl-imvc: Unified and robust representation learning for incomplete multi-view clustering. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2888–2899, 2024.

Haiyue Wang, Wensheng Zhang, and Xiaoke Ma. Contrastive and adversarial regularized multi-level representation learning for incomplete multi-view clustering. *Neural Networks*, 172:106102, 2024a.

Ziyu Wang, Lusi Li, Xin Ning, Wenkai Tan, Yongxin Liu, and Houbing Song. Incomplete multi-view clustering via structure exploration and missing-view inference. *Information Fusion*, 103: 102123, 2024b.

Ziyu Wang, Yiming Du, Yao Wang, Rui Ning, and Lusi Li. Deep incomplete multi-view clustering via multi-level imputation and contrastive alignment. *Neural Networks*, 181:106851, 2025.

Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei, and Hong Liu. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 5393–5400, 2019.

Jie Wen, Zheng Zhang, Zhao Zhang, Lei Zhu, Lunke Fei, Bob Zhang, and Yong Xu. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 10273–10281, 2021.

Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang, and Jinxing Li. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2):1136–1149, 2022.

John Winn and Nebojsa Jojic. Locus: Learning object classes with unsupervised segmentation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pp. 756–763. IEEE, 2005.

Chunming Wu, Zihan Fang, Zhiling Cai, Shide Du, Wei Huang, Jinbo Li, and Shiping Wang. Soinet: Structural optimization-inspired interpretable network for incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 2025.

Lei Xing, Yawen Song, Badong Chen, Changyuan Yu, and Jing Qin. Incomplete multi-view clustering via correntropy and complement consensus learning. *IEEE Transactions on Multimedia*, 26: 8063–8076, 2024.

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 *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 8761–8769, 2022.

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 Transactions on Image Processing*, 32:1354–1366, 2023. doi: 10.1109/TIP.2023.3243521.

Weiqing Yan, Kanglong Liu, Wujie Zhou, and Chang Tang. Deep incomplete multi-view clustering via dynamic imputation and triple alignment with dual optimization. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(4):3250–3261, 2025a. doi: 10.1109/TCSVT.2024.3508785.

Xueming Yan, Ziqi Wang, and Yaochu Jin. Heterogeneity-aware federated graph neural networks for incomplete multi-view clustering. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 10(1):369–381, 2025b.

Yan Yang and Hao Wang. Multi-view clustering: A survey. *Big data mining and analytics*, 1(2): 83–107, 2018.

Zuyuan Yang, Naiyao Liang, Wei Yan, Zhenni Li, and Shengli Xie. Uniform distribution non-negative matrix factorization for multiview clustering. *IEEE Transactions on Cybernetics*, 51(6): 3249–3262, 2021. doi: 10.1109/TCYB.2020.2984552.

Jun Yin, Pei Wang, Shiliang Sun, and Zhonglong Zheng. Incomplete multi-view clustering via multi-level contrastive learning. *IEEE Transactions on Knowledge and Data Engineering*, 37(8): 4716–4727, 2025a. doi: 10.1109/TKDE.2025.3568795.

Jun Yin, Pei Wang, Shiliang Sun, and Zhonglong Zheng. Incomplete multi-view clustering via multi-level contrastive learning. *IEEE Transactions on Knowledge and Data Engineering*, 37(8): 4716–4727, 2025b.

Ming Yin, Xiaohua Liu, Liuyang Wang, and Guoliang He. Learning latent embedding via weighted projection matrix alignment for incomplete multi-view clustering. *Information Sciences*, 634: 244–258, 2023.

Shengju Yu, Zhibin Dong, Siwei Wang, Xinhang Wan, Yue Liu, Weixuan Liang, Pei Zhang, Wenxuan Tu, and Xinwang Liu. Towards resource-friendly, extensible and stable incomplete multi-view clustering. In *Forty-first international conference on machine learning*, 2024.

Yaosong Yu and Dongpu Sun. Incomplete multi-view clustering based on dynamic dimensionality reduction weighted graph learning. *IEEE Access*, 12:19087–19099, 2024.

Chao Zhang, Huaxiong Li, Wei Lv, Zizheng Huang, Yang Gao, and Chunlin Chen. Enhanced tensor low-rank and sparse representation recovery for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11174–11182, 2023.

Yanhao Zhang and Changming Zhu. Incomplete multi-view clustering via attention-based contrast learning. *International Journal of Machine Learning and Cybernetics*, 14(12):4101–4117, 2023.

Yuanyang Zhang, Yijie Lin, Weiqing Yan, Li Yao, Xinhang Wan, Guangyuan Li, Chao Zhang, Guanzhou Ke, and Jie Xu. Incomplete multi-view clustering via diffusion contrastive generation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 22650–22658, 2025.

Zheng Zhang and Wen-Jue He. Tensorized topological graph learning for generalized incomplete multi-view clustering. *Information Fusion*, 100:101914, 2023.

Liang Zhao, Zhikui Chen, Yi Yang, Z Jane Wang, and Victor CM Leung. Incomplete multi-view clustering via deep semantic mapping. *Neurocomputing*, 275:1053–1062, 2018.

Shuping Zhao, Jie Wen, Lunke Fei, and Bob Zhang. Tensorized incomplete multi-view clustering with intrinsic graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11327–11335, 2023.

Qinghai Zheng, Jihua Zhu, and Zhongyu Li. Collaborative unsupervised multi-view representation learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(7):4202–4210, 2022. doi: 10.1109/TCSVT.2021.3127007.

Shuai Zheng, Fan Yang, M Hadi Kiapour, and Robinson Piramuthu. Modanet: A large-scale street fashion dataset with polygon annotations. In *Proceedings of the 26th ACM international conference on Multimedia*, pp. 1670–1678, 2018.

# A Appendix

# B Algorithmic Details

The following algorithms provide the detailed procedures for obtaining the global semantic representation through dynamic semantic view set evolution in SVE_IMVC (Algorithm 1) and training the overall SVE_IMVC framework (Algorithm 2), respectively.

# C Theoretical Analysis

## C.1 Non-Redundancy of Orthogonal Local Fusion

To further theoretically analyze the non-redundancy of the proposed orthogonal local fusion mechanism, we conduct the following analysis on the complementary semantic extraction process to show

**Algorithm 1** Dynamic Semantic View Set Evolution for Generating the Common Representation $\mathbf{H}$

**Input:** Initial semantic view set $\mathcal{P}_0$, thresholds $\epsilon, \eta$
**Output:** Final common representation $\mathbf{H}^*$
Initialize $t = 0$
**repeat**
  Compute complementarity scores $S_{ij}$ via Eq. (7)
  Select complementary view pair $(i^*, j^*)$ via Eq. (8)
  Compute orthogonal component $\mathbf{R}_{j^*}$ via Eq. (9)
  Generate new semantic view $\mathbf{Z}_{\text{new}}$
  Expand semantic view set $\mathcal{P}_t$ via Eq. (11)
  Compute global semantic representation $\mathbf{H}^{(t)}$ via Eq. (12)
  Compute view–global similarity $s_v$ via Eq. (13)
  Update active view set $\mathcal{A}_t$ via Eq. (14)
  Compute updated representation $\mathbf{H}^{(t+1)}$
  Compute $\Delta_t$ via Eq. (16)
  $t \leftarrow t + 1$
**until** $\Delta_t < \epsilon$
Set $T = t$
Obtain final representation $\mathbf{H}^* = \mathbf{H}^{(T)}$

**Algorithm 2** Semantic View Evolution for Incomplete Multi-View Clustering

**Input:** Incomplete multi-view data $\{\mathbf{X}^v\}_{v=1}^V$, indicator matrix $\mathbf{A}$, epochs $E$
**Parameters:** Trade-off coefficients $\lambda_1$, $\lambda_2$, $\lambda_3$
**Output:** Clustering results
**for** $epoch = 1$ to $E$ **do**
  Learn latent representations $\mathbf{Z}^v$ via Eq. (2)
  Maximize cross-view mutual information via Eq. (4)
  Obtain global representation $\mathbf{H}^*$ via Algorithm 1
  Construct local context $\mathbf{C}_v$ via Eq. (18)
  Complete missing views $\hat{\mathbf{Z}}_v$ via Eq. (19)
  Optimize the overall objective via Eq. (21)
**end for**
Fuse complete representations $\mathcal{Z}$
Perform $k$-means on $\mathcal{Z}$

that the extracted candidate complementary information originates from the orthogonal complement of the current semantic direction.

**Proposition 1.** Let $\mathbf{R}_{j^*}$ be defined as in Eq. (9). Then, $\mathbf{R}_{j^*}$ is orthogonal to $\mathbf{Z}_{i^*}$, i.e.,

$$\langle \mathbf{Z}_{i^*}, \mathbf{R}_{j^*} \rangle = 0. \tag{22}$$

Therefore, $\mathbf{R}_{j^*}$ does not contain the projection component along the direction of $\mathbf{Z}_{i^*}$, and its semantic information lies in the orthogonal complement of $\mathbf{Z}_{i^*}$.

*Proof.* According to Eq. (9), we have

$$\mathbf{R}_{j^*} = \mathbf{Z}_{j^*} - \frac{\langle \mathbf{Z}_{j^*}, \mathbf{Z}_{i^*} \rangle}{\|\mathbf{Z}_{i^*}\|_2^2} \mathbf{Z}_{i^*}. \tag{23}$$

Taking the inner product of both sides with $\mathbf{Z}_{i^*}$ yields

$$\begin{aligned} \langle \mathbf{Z}_{i^*}, \mathbf{R}_{j^*} \rangle &= \langle \mathbf{Z}_{i^*}, \mathbf{Z}_{j^*} \rangle - \frac{\langle \mathbf{Z}_{j^*}, \mathbf{Z}_{i^*} \rangle}{\|\mathbf{Z}_{i^*}\|_2^2} \langle \mathbf{Z}_{i^*}, \mathbf{Z}_{i^*} \rangle \\ &= \langle \mathbf{Z}_{i^*}, \mathbf{Z}_{j^*} \rangle - \langle \mathbf{Z}_{j^*}, \mathbf{Z}_{i^*} \rangle \\ &= 0. \end{aligned} \tag{24}$$

Hence,

$$\mathbf{R}_{j^*} \in \operatorname{span}(\mathbf{Z}_{i^*})^{\perp}. \tag{25}$$

Furthermore, $\mathbf{Z}_{j^*}$ can be decomposed as

$$\mathbf{Z}_{j^*} = \frac{\langle \mathbf{Z}_{j^*}, \mathbf{Z}_{i^*} \rangle}{\|\mathbf{Z}_{i^*}\|_2^2} \mathbf{Z}_{i^*} + \mathbf{R}_{j^*}. \tag{26}$$

The first term represents the shared semantic component along the direction of $\mathbf{Z}_{i^*}$, while $\mathbf{R}_{j^*}$ represents the remaining orthogonal complementary information. Since only $\mathbf{R}_{j^*}$ is introduced into the fusion process, the shared semantic component is not directly reintroduced, thereby reducing redundant semantic accumulation during view evolution. □

### C.2 Theoretical Analysis of Global Semantic Evolution

To further analyze the semantic evolution of the proposed dynamic semantic view set, we consider the alternating View Expansion and View Contraction process. The analysis shows that, under the stated conditions, View Expansion introduces new complementary semantic directions, while View Contraction removes only redundant representations. Consequently, the effective semantic information represented by the active semantic views is preserved or increased throughout the evolution process.

**Proposition 2.** Let $\mathcal{A}_t$ denote the active semantic view set at the $t$-th iteration, and let the corresponding global semantic representation be defined as

$$\mathbf{H}^{(t)} = \frac{1}{|\mathcal{A}_t|} \sum_{\mathbf{Z}_v \in \mathcal{A}_t} \mathbf{Z}_v. \tag{27}$$

Define the effective semantic subspace represented by the active semantic views as

$$\mathcal{S}_t = \operatorname{span}\left\{ \mathbf{Z}_v \mid \mathbf{Z}_v \in \mathcal{A}_t \right\}, \tag{28}$$

and define its effective semantic dimensionality as

$$\mathcal{I}^{(t)} = \dim(\mathcal{S}_t). \tag{29}$$

Assume that (i) View Expansion introduces a non-zero complementary component that is not contained in the current semantic subspace, and (ii) View Contraction removes only redundant semantic views whose removal does not reduce the span of the remaining active semantic views. Then, the effective semantic dimensionality of the active semantic representations is non-decreasing throughout the semantic view evolution, i.e.,

$$\mathcal{I}^{(t+1)} \geq \mathcal{I}^{(t)}. \tag{30}$$

Moreover, the inequality is strict when View Expansion introduces a new independent semantic direction.

*Proof.* At the $t$-th iteration, View Expansion selects a complementary view pair and extracts the orthogonal semantic component $\mathbf{R}_{j^*}$ according to Eq. (9). According to Proposition 1,

$$\langle \mathbf{Z}_{i^*}, \mathbf{R}_{j^*} \rangle = 0. \tag{31}$$

Thus, $\mathbf{R}_{j^*}$ contains no component along the direction of $\mathbf{Z}_{i^*}$. Suppose that the extracted component satisfies

$$\mathbf{R}_{j^*} \neq \mathbf{0}, \qquad \mathbf{R}_{j^*} \notin \mathcal{S}_t. \tag{32}$$

Since $\mathbf{Z}_{i^*} \in \mathcal{S}_t$, the generated semantic view

$$\mathbf{Z}_{i^*} + \mathbf{R}_{j^*}$$

introduces a new direction outside the current semantic subspace. Therefore, the semantic subspace immediately after View Expansion is

$$\begin{aligned} \mathcal{S}_t^+ &= \operatorname{span}\left(\mathcal{S}_t, \mathbf{Z}_{i^*} + \mathbf{R}_{j^*}\right) \\ &= \operatorname{span}\left(\mathcal{S}_t, \mathbf{R}_{j^*}\right). \end{aligned} \tag{33}$$

Because $\mathbf{R}_{j^*} \notin \mathcal{S}_t$, it follows that

$$\dim(\mathcal{S}_t^+) = \dim(\mathcal{S}_t) + 1. \tag{34}$$

Table 3: Loss Ablation Experiments on Two Data Sets with the Missing Rate of 0.5

| | Loss Components | | | Fashion | | | MNIST_USPS | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\mathcal{L}_{\mathrm{REC}}$ | $\mathcal{L}_{\mathrm{MMI}}$ | $\mathcal{L}_{\mathrm{CON}}$ | ACC | NMI | ARI | ACC | NMI | ARI |
| Item-1 | ✓ | | | 54.95 | 65.96 | 46.67 | 32.08 | 29.11 | 27.93 |
| Item-2 | | ✓ | | 94.08 | 87.35 | 86.23 | 94.23 | 89.60 | 88.28 |
| Item-3 | | | ✓ | 59.98 | 73.30 | 55.63 | 60.55 | 62.34 | 59.01 |
| Item-4 | ✓ | ✓ | | 96.15 | 89.77 | 90.31 | 95.17 | 91.26 | 90.83 |
| Item-5 | | ✓ | ✓ | 95.43 | 90.18 | 90.01 | 95.03 | 92.52 | 92.79 |
| Item-6 | ✓ | | ✓ | 58.51 | 72.01 | 57.64 | 59.68 | 64.70 | 64.62 |
| **Item-7** | ✓ | ✓ | ✓ | **96.22** | **91.99** | **91.99** | **98.22** | **95.17** | **96.09** |

After View Expansion, View Contraction removes redundant semantic views. We assume that a view is removed only when its removal does not reduce the span of the remaining active semantic views. Hence, if $\mathcal{S}_{t+1}$ denotes the semantic subspace after View Contraction, then

$$\mathcal{S}_t^+ \subseteq \mathcal{S}_{t+1}, \tag{35}$$

and consequently

$$\dim(\mathcal{S}_{t+1}) \geq \dim(\mathcal{S}_t^+). \tag{36}$$

Combining the effects of View Expansion and View Contraction gives

$$\mathcal{I}^{(t+1)} = \dim(\mathcal{S}_{t+1}) \geq \dim(\mathcal{S}_t^+) \geq \dim(\mathcal{S}_t) = \mathcal{I}^{(t)}. \tag{37}$$

Furthermore, when $\mathbf{R}_{j^*} \notin \mathcal{S}_t$ and the newly introduced direction is preserved after View Contraction, we have

$$\mathcal{I}^{(t+1)} > \mathcal{I}^{(t)}. \tag{38}$$

Therefore, under the stated assumptions, the semantic view evolution process preserves the semantic subspace already represented by the active views while incorporating new complementary semantic directions. This establishes that the effective semantic dimensionality associated with the global semantic representation is non-decreasing throughout the evolution process. □

# D Additional Experimental Analysis

## D.1 Ablation Study

To analyze the specific contributions of each loss component, experiments are conducted on the Fashion and MNIST_USPS datasets with a missing rate of 0.5, and the results are reported in Table 3. The results show that using only $\mathcal{L}_{\mathrm{REC}}$ yields relatively low overall performance, indicating that reconstruction alone is insufficient to learn highly discriminative semantic representations. After introducing $\mathcal{L}_{\mathrm{MMI}}$, the ACC, NMI, and ARI scores on both datasets are substantially improved, demonstrating that cross-view mutual information learning effectively enhances semantic consistency among different views. Further combining two or three loss components generally improves the clustering performance. In particular, Item-7, which incorporates all three loss components, achieves the highest overall performance on both datasets, demonstrating the complementary roles of the three loss components in representation reconstruction, cross-view semantic consistency, and global semantic regularization.

## D.2 Parameter Analysis

In this section, we conduct a sensitivity analysis of the key hyperparameters in the overall loss function of the proposed method. Specifically, the weight of the reconstruction loss is fixed as $\lambda_1 = 0.1$, while $\lambda_2$ and $\lambda_3$ are jointly selected from $\{0.01, 0.1, 1.0, 10.0, 100.0\}$. The experimental results are shown in Figure 3.

Overall, when $\lambda_2$ takes moderate or relatively large values and $\lambda_3$ lies within an appropriate range, the proposed method generally achieves better performance, indicating that a proper balance between cross-view shared semantic learning and global semantic consistency is important. Specifically, a small $\lambda_2$ makes it difficult to sufficiently exploit cross-view shared information. As $\lambda_2$

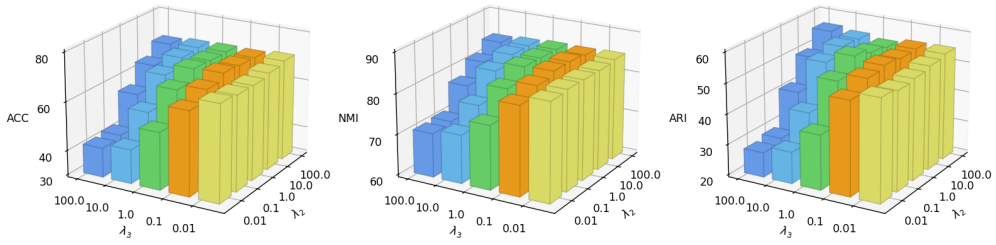
increases, the model can learn more stable representations, providing a better foundation for subsequent semantic fusion. However, the performance improvement gradually becomes marginal as $\lambda_2$ is further increased. On the other hand, $\mathcal{L}_{\mathrm{CON}}$ takes the dynamically formed global semantic representation as an anchor and guides the current active representations toward the global semantic space. An appropriate $\lambda_3$ enhances semantic consistency among different representations and facilitates the convergence of the fusion process. When $\lambda_3$ is too small, the guidance of the global semantics is relatively limited, whereas an excessively large value may overemphasize global consistency and affect the preservation of complementary information. Therefore, a moderate $\lambda_3$ can achieve a better balance between global semantic coordination and the preservation of local complementary information.

The experimental results on the two datasets further demonstrate that a reasonable joint adjustment of $\lambda_2$ and $\lambda_3$ helps maintain an effective balance and promotes the stable evolution of the elastic semantic view set. The global semantics progressively accumulated during the evolution process are further transferred to missing-view completion, thereby establishing a unified framework that connects dynamic semantic evolution with incomplete multi-view clustering.

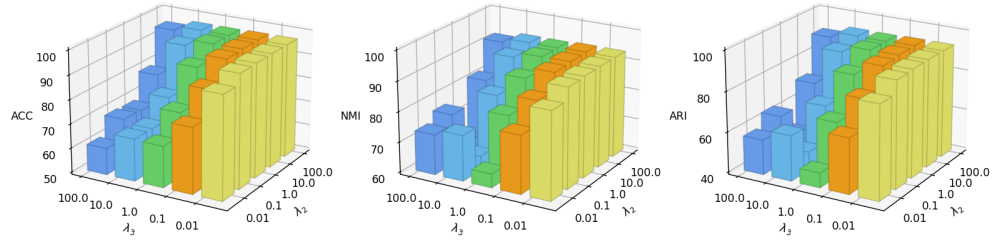
### D.3 VISUALIZATION

As shown in Figure 4, we conduct *t*-SNE visualization experiments on six benchmark datasets. Specifically, subfigures (a)–(d), (e)–(h), (i)–(l), (m)–(p), (q)–(t), and (u)–(x) illustrate the clustering distributions of the learned common representations on the Handwritten, MSRC-v1, Hdigit, 100leaves, MNIST_USPS, and Fashion datasets, respectively, at different training epochs.

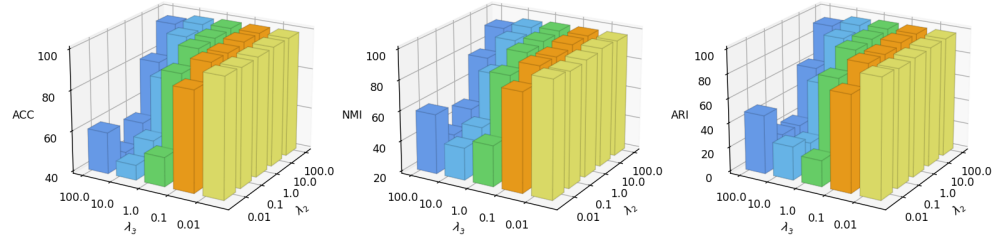
From the visualization results, it can be observed that the learned representations gradually exhibit clearer and more compact cluster structures as training progresses. The boundaries between different categories become increasingly distinguishable, while samples from the same category are aggregated more closely in the latent space. These results indicate that the proposed SVE_IMVC framework progressively learns more discriminative semantic representations during training, thereby improving the clustering quality across different incomplete multi-view datasets.



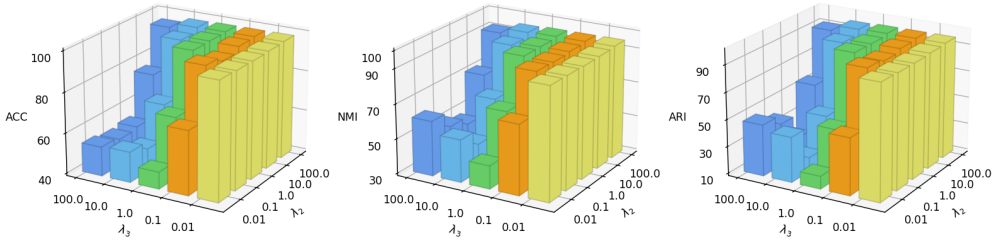

(a) 100leaves




(b) Fashion

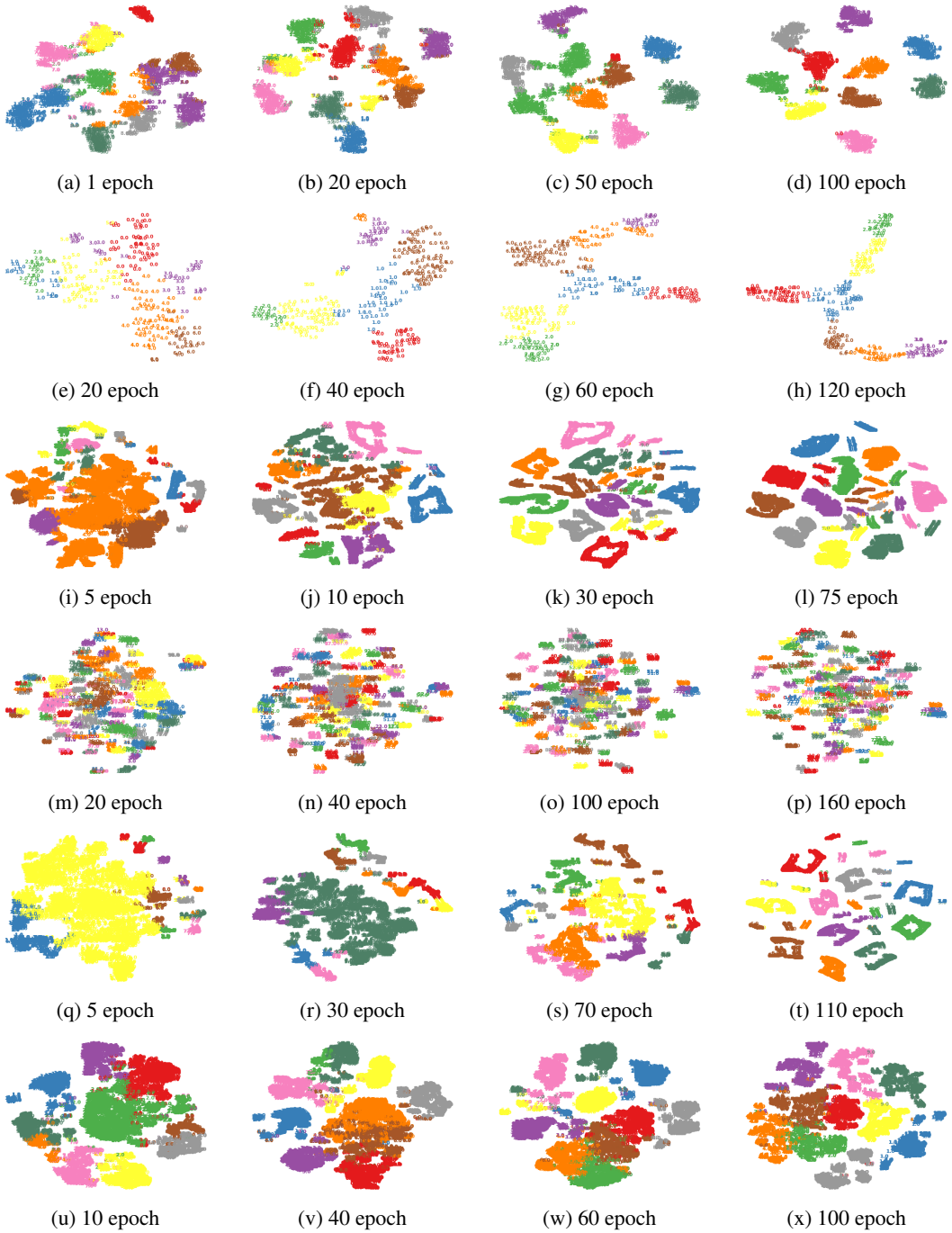



(c) Hdigit




(d) MNIST_USPS

Figure 3: Parameter sensitivity analysis on the 100leaves, Fashion, Hdigit, and MNIST_USPS datasets.



(a) 1 epoch (b) 20 epoch (c) 50 epoch (d) 100 epoch

(e) 20 epoch (f) 40 epoch (g) 60 epoch (h) 120 epoch

(i) 5 epoch (j) 10 epoch (k) 30 epoch (l) 75 epoch

(m) 20 epoch (n) 40 epoch (o) 100 epoch (p) 160 epoch

(q) 5 epoch (r) 30 epoch (s) 70 epoch (t) 110 epoch

(u) 10 epoch (v) 40 epoch (w) 60 epoch (x) 100 epoch

Figure 4: t-SNE visualizations of clustering results on six datasets with increasing training epochs. (a)–(d): Handwritten; (e)–(h): MSRC-v1; (i)–(l): Hdigit; (m)–(p): 100leaves; (q)–(t): MNIST_USPS; (u)–(x): Fashion.