SEPRO: SEMANTIC EVIDENCE PROPAGATION FOR INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete Multi-View Clustering (IMVC) aims to learn coherent representations from data with missing views. However, observed views may also be noisy, biased, or incomplete in semantics, making existing reconstruction- and consistency-based methods vulnerable to unreliable information. To address this issue, we propose SEPro, a semantic evidence propagation framework for IMVC. SEPro represents each view with explicit semantic evidence, uncertainty, and confidence, and propagates this evidence through prototype-mediated high-order relations guided by view availability, semantic affinity, and confidence. The learned evidential representations also support uncertainty-aware inference for missing views. By jointly integrating evidence extraction, propagation, and inference, SEPro learns robust consensus representations for incomplete multi-view clustering. Extensive experiments demonstrate its effectiveness and robustness under varying missing-view conditions.

1 INTRODUCTION

In real-world applications, multi-view observations are often incomplete, and the available views may also vary substantially in their semantic quality. Incomplete Multi-View clustering (IMVC) aims to learn coherent representations from such partially observed data. However, an observed view is not necessarily a reliable source of semantic information: it may contain noise, bias, or insufficient semantic cues. Therefore, the challenge of IMVC goes beyond recovering missing views and lies in determining how semantic information from incomplete and uncertain observations should be represented, propagated, and integrated.

In recent years, deep IMVC methods have mainly explored two directions. One line explicitly recovers missing views through reconstruction, generative models, or probabilistic inference (Wen et al., 2024; Zhang et al., 2025; 2026; Wang et al., 2024; Xu et al., 2024). These methods improve information completeness by generating plausible representations for unavailable views, but the inferred information can be unreliable when the observed views are noisy or semantically insufficient. Another line avoids explicit imputation and learns common representations through contrastive learning, graph propagation, attention mechanisms, or prototype-based alignment (Lin et al., 2023b; 2021; Chao et al., 2025; Yuan et al., 2024; 2025; Teng et al., 2024). Although these methods effectively promote cross-view consistency, they often treat the available views as equally informative or propagate their information without explicitly modeling the uncertainty associated with the evidence. As a result, unreliable observations may be propagated together with informative semantics, which can distort the learned consensus representation.

More fundamentally, existing IMVC methods typically regard semantic fusion and missing-view recovery as separate procedures. Semantic fusion focuses on discovering common structures across observed views, while missing-view recovery aims to predict unavailable representations. Such a separation overlooks an important fact: the semantic information used for fusion and the information used for missing-view inference originate from the same incomplete observations and should therefore be governed by a common mechanism. Moreover, conventional pairwise relations mainly describe interactions between individual samples, which are insufficient to explicitly capture the joint dependencies among samples, views, and latent semantic categories. This motivates a unified question: *how can semantic evidence from incomplete and uncertain views be effectively propagated into a shared semantic space while controlling the influence of unreliable information?*

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.

To address this issue, we propose **SEPro**, a semantic evidence propagation framework for incomplete multi-view clustering (as shown in Figure 1). The central idea of SEPro is to regard each observed view as a source of semantic evidence rather than a deterministic and equally reliable observation. Specifically, SEPro first transforms view-specific latent representations into evidential representations that jointly characterize semantic information, uncertainty, and confidence. These evidential representations are then propagated through a prototype-mediated high-order relational structure. The propagation strength jointly considers view availability, semantic affinity, and evidence confidence, allowing informative evidence to contribute more strongly while suppressing unreliable information. In parallel, the learned evidential representations are used to perform uncertainty-aware inference for missing views, while the propagation mechanism learns a robust consensus semantic space from the available evidence. In this way, evidence extraction, semantic propagation, and missing-view inference form a unified framework for learning robust consensus representations.

The main contributions of this paper are summarized as follows:

- We formulate incomplete multi-view clustering from the perspective of **semantic evidence propagation**. Instead of treating available views as equally reliable deterministic observations, SEPro explicitly represents view-specific semantics together with their uncertainty and confidence, providing a unified basis for reasoning under incomplete observations.
- We develop an **uncertainty-guided semantic evidence propagation** mechanism based on prototype-mediated high-order relations. The propagation weights jointly account for view availability, semantic affinity, and evidence confidence, enabling reliable evidence to be effectively integrated while reducing the influence of uncertain observations.
- We introduce an **evidence-consistent semantic inference** strategy for missing views. The missing representations are inferred from the learned semantic evidence and adaptively calibrated according to predictive uncertainty and confidence, thereby reducing the accumulation of unreliable completion information.

2 Related Work

2.1 Incomplete Multi-View Clustering

Incomplete multi-view clustering has been extensively studied to learn coherent representations from data with missing views. Existing approaches can be broadly divided into completion-based and imputation-free paradigms. Completion-based methods explicitly recover missing information from the observed views. Early generative approaches employ conditional diffusion models to generate missing views and capture cross-view semantic dependencies (Wen et al., 2024; Zhang et al., 2025; 2026). Other methods exploit adversarial learning, variational inference, or multi-level imputation to improve the fidelity and semantic consistency of recovered representations (Wang et al., 2024; Xu et al., 2024; Wang et al., 2025; Yin et al., 2025). Structural priors and cross-view reasoning have also been introduced to improve recovery under highly incomplete settings, such as graph-guided generation and cross-view contextual inference (Li et al., 2025; Gao et al., 2025). These methods make missing-view recovery more flexible, but the inferred representations may accumulate errors when the available views are noisy or semantically insufficient.

In contrast, imputation-free methods avoid explicitly reconstructing the missing views and instead seek a shared semantic space from the available observations. Contrastive learning has been widely adopted to improve cross-view consistency at the feature, semantic, and assignment levels (Lin et al., 2023b; 2021; Yin et al., 2025). Graph-based propagation and attention mechanisms further exploit sample relationships to transfer information across incomplete views (Feng et al., 2025; Chao et al., 2025; Teng et al., 2024). Prototype-based methods reduce the sparsity caused by missing observations by establishing category-level semantic correspondences (Yuan et al., 2024; 2025), while recent studies also explore knowledge distillation, consensus semantic learning, and adaptive information fusion (Dai et al., 2025; Cai et al., 2025). Despite these advances, most existing methods either emphasize information recovery or enforce cross-view consistency. They generally treat the available representations as homogeneous information sources and lack an explicit mechanism for distinguishing reliable semantic evidence from uncertain observations during information propagation.

2.2 Semantic Structure Modeling and Evidential Learning

Beyond missing-view recovery, structural modeling plays an important role in multi-view semantic fusion. Conventional graph-based methods construct pairwise relations between instances and perform graph propagation to exploit local or global geometric structures (Xue et al., 2024; Chao et al., 2025). High-order relational modeling has subsequently been introduced to capture richer dependencies beyond pairwise interactions. For example, high-order random walks and tensor-based formulations have been explored to model complex cross-view structures (Lu et al., 2024; Wu & Lu, 2026; Wang et al., 2026). Prototype learning provides another effective mechanism for semantic abstraction by representing category-level information with compact prototypes and using them to establish cross-view correspondences (Yuan et al., 2024; 2025). More recent approaches combine graphs with optimal transport, adaptive feature fusion, or structural optimization to improve robustness under heterogeneous missing patterns (Lin et al., 2023a; Xue et al., 2022; Pu et al., 2024; Wu et al., 2026). However, these approaches mainly focus on constructing stronger semantic relationships, while the reliability of the information being propagated is not explicitly incorporated into the relational mechanism.

Evidential learning offers a principled way to characterize semantic predictions together with uncertainty and confidence. Recent evidential multi-view clustering methods introduce view- or feature-aware evidential representations to handle ambiguity and conflicting information across modalities. However, such methods mainly focus on uncertainty-aware clustering or view weighting rather than incomplete-view semantic propagation. SEPro is different in that uncertainty and confidence are not treated as auxiliary statistics. Instead, they directly regulate the propagation of semantic evidence through prototype-mediated high-order relations and are further reused for missing-view semantic inference. Thus, SEPro couples semantic representation, evidence propagation, and missing-view inference within a unified framework.

3 Method

3.1 Problem Formulation and Overview

Let $\{\mathbf{X}^v\}_{v=1}^V$ denote a multi-view dataset with V views, where $\mathbf{X}^v \in \mathbb{R}^{N\times d_v}$. Due to incomplete observations, the availability of the v-th view for sample i is indicated by $\mathbf{M}_i^v \in \{0,1\}$, where $\mathbf{M}_i^v = 1$ indicates that sample i is observed in view v. The goal of incomplete multi-view clustering is to learn a robust consensus representation and infer the cluster assignments from partially observed multi-view data.

Different views may provide semantic information with different observation quality. Therefore, instead of treating each available view as an equally reliable deterministic representation, we regard each view-specific representation as a source of semantic evidence. For sample i in view v, the semantic evidence is represented as

$$\mathcal{E}_i^v = \left(\boldsymbol{\mu}_i^v, \sigma_i^{2(v)}, u_i^v\right), \tag{1}$$

where $\boldsymbol{\mu}_i^v$ denotes the semantic representation, $\sigma_i^{2(v)}$ measures the associated uncertainty, and u_i^v denotes the confidence of the evidence.

Based on this formulation, SEPro propagates semantic evidence through prototype-mediated high-order relations. View-specific evidence is propagated according to view availability, semantic affinity, and confidence to learn a consensus representation, while evidential modeling enables uncertainty-aware missing-view inference. These components are jointly optimized, with the learned evidential encoders used for missing-view inference.

3.2 View-specific Semantic Evidence Encoding

3.2.1 View-specific Representation Learning

We employ an independent autoencoder for each view to learn view-specific latent representations. Specifically, an encoder $f_\theta^{(v)}(\cdot)$ maps the input $\mathbf{X}^v$ into the latent space:

$$\mathbf{Z}^v = f_\theta^{(v)}(\mathbf{X}^v), \tag{2}$$

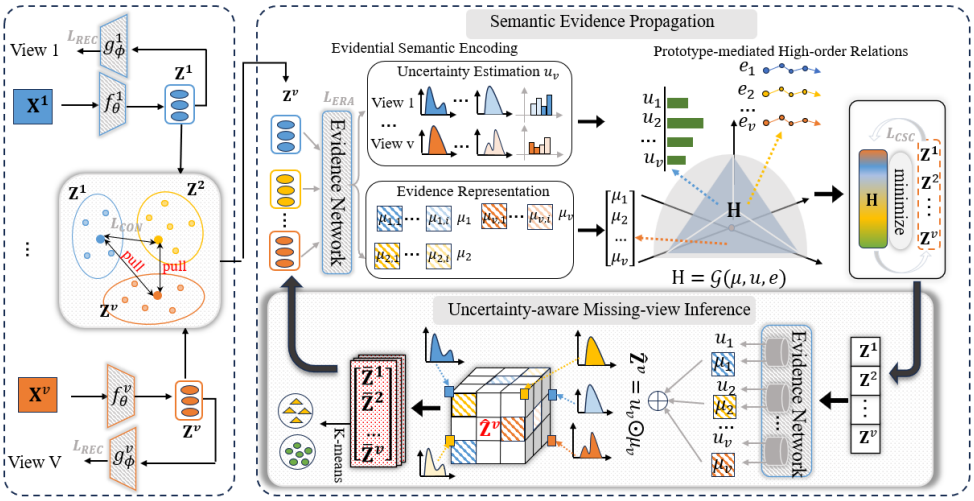


Figure 1: Overview of the proposed SEPro framework. SEPro encodes each available view into semantic evidence with semantic representations, uncertainty, and confidence. The evidence is propagated through prototype-mediated high-order relations to learn a consensus representation and support uncertainty-aware missing-view inference.

and a decoder $g_\phi^{(v)}(\cdot)$ reconstructs the original input as $\widehat{\mathbf{X}}^v = g_\phi^{(v)}(\mathbf{Z}^v)$. The reconstruction objective is defined as

$$\mathcal{L}_{\mathrm{REC}} = \sum_{v=1}^{V} \left\| \mathbf{X}^v - g_\phi^{(v)}\left(f_\theta^{(v)}(\mathbf{X}^v)\right) \right\|_F^2 . \tag{3}$$

To encourage cross-view semantic consistency, we further employ the pairwise mutual-information-based constraint used in the original representation learning process. For two views v_1 and v_2, we construct the cross-view feature interaction by computing $z_{i,m}^{v_1} z_{i,n}^{v_2}$ for each pair of feature dimensions $m, n \in \{1, \ldots, D\}$. Aggregating these interactions over all samples yields the cross-view joint feature matrix $\mathbf{K}^{(v_1,v_2)} \in \mathbb{R}^{D\times D}$, whose (m,n)-th entry is given by $K_{m,n}^{(v_1,v_2)} = \frac{1}{N}\sum_{i=1}^{N} z_{i,m}^{v_1} z_{i,n}^{v_2}$. The corresponding marginal distributions are obtained by summing $\mathbf{K}^{(v_1,v_2)}$ over the second and first dimensions, respectively, i.e., $K_m^{(v_1)} = \sum_{n=1}^{D} K_{m,n}^{(v_1,v_2)}$ and $K_n^{(v_2)} = \sum_{m=1}^{D} K_{m,n}^{(v_1,v_2)}$. The pairwise cross-view consistency loss is defined as

$$\mathcal{L}_{\mathrm{P}}^{(v_1,v_2)} = -\sum_{m=1}^{D}\sum_{n=1}^{D} K_{m,n}^{(v_1,v_2)} \log\left(\frac{K_{m,n}^{(v_1,v_2)}}{K_m^{(v_1)} K_n^{(v_2)}}\right). \tag{4}$$

Finally, the overall cross-view consistency loss is obtained by averaging over all unordered view pairs:

$$\mathcal{L}_{\mathrm{CON}} = \frac{2}{V(V-1)} \sum_{1\le v_1 < v_2 \le V} \mathcal{L}_{\mathrm{P}}^{(v_1,v_2)}. \tag{5}$$

3.2.2 Evidential Semantic Encoding

To characterize the uncertainty associated with view-specific semantic information, we introduce a view-specific evidential encoder. First, an observed-view context is constructed as

$$\widetilde{\mathbf{z}}_i = \frac{\sum_{v=1}^{V} \mathbf{M}_i^v \mathbf{z}_i^v}{\sum_{v=1}^{V} \mathbf{M}_i^v + \epsilon}, \tag{6}$$

where ϵ is a small constant used to avoid division by zero.

The context representation is then fed into the evidential encoder $E_v(\cdot)$ to obtain the semantic evidence of view v:

$$\left(\boldsymbol{\mu}_i^v, \log \sigma_i^{2(v)}, u_i^v\right) = E_v(\widetilde{\mathbf{z}}_i). \tag{7}$$

Here, $\boldsymbol{\mu}_i^v \in \mathbb{R}^d$ denotes the semantic evidence representation, $\sigma_i^{2(v)} = \exp(\log \sigma_i^{2(v)})$ represents the uncertainty, and u_i^v denotes the confidence score.

We model the view-specific latent representation using an isotropic Gaussian distribution:

$$\mathbf{z}_i^v \sim \mathcal{N}\left(\boldsymbol{\mu}_i^v, \sigma_i^{2(v)}\mathbf{I}\right). \tag{8}$$

The evidential parameters are learned by minimizing the corresponding negative log-likelihood:

$$\mathcal{L}_{\mathrm{ERA}}^v = \frac{1}{2}\left(\frac{\|\mathbf{z}_i^v - \boldsymbol{\mu}_i^v\|_2^2}{\sigma_i^{2(v)}} + \log \sigma_i^{2(v)}\right). \tag{9}$$

Aggregating the objective over all views yields

$$\mathcal{L}_{\mathrm{ERA}} = \frac{1}{2V}\sum_{v=1}^{V}\left(\frac{\|\mathbf{z}_i^v - \boldsymbol{\mu}_i^v\|_2^2}{\sigma_i^{2(v)}} + \log \sigma_i^{2(v)}\right). \tag{10}$$

The learned evidential representations and confidence scores are denoted by

$$\mathcal{Z}_e = \left\{\boldsymbol{\mu}^1, \boldsymbol{\mu}^2, \ldots, \boldsymbol{\mu}^V\right\}, \qquad U = \left\{u^1, u^2, \ldots, u^V\right\}. \tag{11}$$

These quantities provide the semantic evidence and confidence information required for subsequent evidence propagation.

3.3 Semantic Evidence Propagation

SEPro propagates view-specific semantic evidence into a shared semantic space through prototype-mediated high-order relations. Unlike conventional pairwise fusion, the propagation process jointly considers sample identity, view identity, semantic category, and evidence confidence.

3.3.1 Prototype-mediated High-order Relations

Let $C = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_K\}$ denote a set of K semantic prototypes. Each prototype acts as an intermediate semantic carrier connecting heterogeneous sample-view evidence.

For the semantic evidence representation $\boldsymbol{\mu}_i^v$, its distance to prototype k is defined as

$$d_{ivk} = \|\boldsymbol{\mu}_i^v - \mathbf{c}_k\|_2. \tag{12}$$

The corresponding semantic affinity is computed by

$$a_{ivk} = \frac{\exp(-d_{ivk}/\tau)}{\sum_{k'=1}^{K}\exp(-d_{ivk'}/\tau)}, \tag{13}$$

where τ is the temperature parameter.

We then define the semantic evidence propagation weight as

$$w_{ivk} = \mathbf{M}_i^v u_i^v a_{ivk}. \tag{14}$$

The propagation weight jointly considers three factors: view availability, semantic affinity, and evidence confidence. Specifically, unavailable views make no contribution through $\mathbf{M}_i^v$, while low-confidence evidence receives a reduced propagation strength through u_i^v.

The resulting tensor $\mathbf{W} \in \mathbb{R}^{N\times V\times K}$ defines a prototype-mediated high-order relational structure among samples, views, and semantic prototypes.

3.3.2 CONFIDENCE-GUIDED EVIDENCE AGGREGATION

The semantic prototypes aggregate evidence from multiple sample-view pairs according to the propagation weights. Specifically, the k-th semantic prototype is obtained by solving the weighted evidence aggregation objective:

$$\min_{\mathbf{e}_k} \sum_{i=1}^{N} \sum_{v=1}^{V} w_{ivk} \left\| \boldsymbol{\mu}_i^v - \mathbf{e}_k \right\|_2^2. \tag{15}$$

The closed-form solution is

$$\mathbf{e}_k = \frac{\sum_{i=1}^{N} \sum_{v=1}^{V} w_{ivk} \boldsymbol{\mu}_i^v}{\sum_{i=1}^{N} \sum_{v=1}^{V} w_{ivk}}. \tag{16}$$

Therefore, each semantic prototype summarizes a set of semantically related evidence while assigning greater influence to observations with higher confidence.

For sample i, the aggregated association with prototype k is defined as

$$w_{ik} = \sum_{v=1}^{V} w_{ivk}, \qquad \widetilde{w}_{ik} = \frac{w_{ik}}{\sum_{k'=1}^{K} w_{ik'}}. \tag{17}$$

The consensus semantic representation is then obtained by propagating evidence from the semantic prototypes:

$$\mathbf{h}_i = \sum_{k=1}^{K} \widetilde{w}_{ik} \mathbf{e}_k. \tag{18}$$

Collecting all samples gives

$$\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N\}. \tag{19}$$

Through this process, semantic evidence from heterogeneous views is propagated to a shared prototype-mediated semantic space, producing a robust consensus representation. The detailed algorithms are provided in Appendix B.

3.3.3 CONSENSUS SEMANTIC REGULARIZATION

To preserve the semantic information of individual views while aligning them with the propagated consensus space, we introduce a cross-space semantic regularization.

For the v-th view, the contrastive loss between $\mathbf{Z}^v$ and $\mathbf{H}$ is defined as

$$\mathcal{L}_{\mathrm{CSC}}^v = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{\exp\left(\mathrm{sim}\left(\mathbf{z}_i^v, \mathbf{h}_i\right)\right)}{\sum_{\substack{j=1 \\ j \neq i}}^{N} \exp\left(\mathrm{sim}\left(\mathbf{z}_i^v, \mathbf{h}_j\right)\right)}. \tag{20}$$

Aggregating over all views yields

$$\mathcal{L}_{\mathrm{CSC}} = \frac{1}{V} \sum_{v=1}^{V} \mathcal{L}_{\mathrm{CSC}}^v. \tag{21}$$

This regularization encourages the propagated consensus representation to preserve the semantics of individual views while reducing cross-view semantic discrepancies.

3.4 UNCERTAINTY-AWARE MISSING-VIEW INFERENCE

For a sample with missing views, SEPro infers the unavailable representations from the semantic information contained in the remaining observed views. Specifically, we first construct an observed-view context as defined in Eq. (6). For a missing view v, the observed context is fed into the view-specific evidential encoder to infer its semantic evidence, as defined in Eq. (7).

The predicted semantic representation is calibrated according to its uncertainty. Specifically, we define the uncertainty suppression as

$$\widetilde{\boldsymbol{\mu}}_i^v = \frac{\boldsymbol{\mu}_i^v}{1 + \sigma_i^{2(v)}}. \tag{22}$$

We further transform the confidence score into an adaptive weighting factor $\alpha_i^v = \mathrm{sigmoid}(u_i^v)$. The completed representation for the missing view is then given by $\widehat{\mathbf{z}}_i^v = \alpha_i^v \widetilde{\boldsymbol{\mu}}_i^v$. For an observed view, the original representation is retained. Therefore, the final representation of each view is

$$\overline{\mathbf{z}}_i^v = \mathbf{M}_i^v \mathbf{z}_i^v + (1 - \mathbf{M}_i^v)\, \widehat{\mathbf{z}}_i^v. \tag{23}$$

After this inference process, all views obtain a unified latent representation set $\{\overline{\mathbf{z}}_i^v\}_{v=1}^V$, which is used for downstream clustering.

3.5 Joint Optimization

The above components are jointly optimized through the following objective:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\mathrm{REC}} + \lambda_2 \mathcal{L}_{\mathrm{CON}} + \lambda_3 \mathcal{L}_{\mathrm{CSC}} + \lambda_4 \mathcal{L}_{\mathrm{ERA}}, \tag{24}$$

where λ_1, λ_2, λ_3, and λ_4 are trade-off coefficients. Specifically, $\mathcal{L}_{\mathrm{REC}}$ preserves view-specific information, $\mathcal{L}_{\mathrm{CON}}$ promotes cross-view semantic consistency, $\mathcal{L}_{\mathrm{CSC}}$ aligns the view-specific representations with the propagated consensus semantic space, and $\mathcal{L}_{\mathrm{ERA}}$ learns uncertainty-aware semantic evidence. The overall training procedure of SEPro is summarized in Appendix B.

4 Experiments

4.1 Experimental Setup

We evaluate the proposed method on six representative datasets: Hdigit (Yang et al., 2021) (10,000 samples, 10 categories), 100leaves (Zheng et al., 2022) (1,600 samples, 100 categories), MNIST_USPS (Peng et al., 2019) (5,000 samples, 10 categories, 2 views), Handwritten (LeCun et al., 1989) (2,000 samples, 10 categories, 6 views), Caltech-5V (Fei-Fei et al., 2004) (1,400 samples, 7 categories), and MSRC-v1 (Winn & Jojic, 2005) (210 samples, 7 categories).

We compare our method with 12 state-of-the-art incomplete multi-view clustering methods, which can be broadly categorized into three groups. Completion-based methods include **DIMVC-HIA**(Du et al., 2026), **HSACC**(Ding et al., 2025), **DITA-IMVC**(Yan et al., 2025), **IMC-MCL**(Yin et al., 2025), **MICA**(Wang et al., 2025), **IMVC-IE**(Huang et al., 2024), and **ProImp**(Li et al., 2023), which mainly focus on recovering missing views through various completion or prediction strategies. Representation learning-based methods include **DCG**(Zhang et al., 2025), **DIVIDE**(Lu et al., 2024), **MRL_CAL**(Wang et al., 2024), and **MCAC**(Zhang & Zhu, 2023), which enhance clustering through contrastive, adversarial, or high-order representation learning. Common-space learning methods include **APADC**(Xu et al., 2023), which learns a shared representation through adaptive projection and mutual information maximization.

4.2 Experimental Setting

All experiments are conducted using PyTorch 2.3.1 on an NVIDIA RTX 4070 GPU with 32 GB memory. We use a fully connected autoencoder with an Input–1024–1024–1024–Output encoder and its symmetric decoder, with ReLU activations. Each view is modeled by an evidential MLP with a $D \to 512 \to 256$ structure, where the first layer uses Batch Normalization and ReLU, followed by ReLU in the second layer. The model is trained for E epochs (e.g., 100), with a learning rate of 0.0001 and a batch size of 256 under multiple missing-rate settings.

4.3 Experimental Results and Analysis

Table 1 reports the clustering performance of different methods under varying missing rates. Overall, SEPro achieves superior performance on most datasets and evaluation metrics, maintaining clear advantages as the missing rate increases. At a missing rate of 0.1, SEPro shows notable improvements

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

Ratio	Method	Hdigit			100leaves			MNIST_USPS			Handwritten			Caltech-5V			MSRC-v1		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
0.1	DIMVC-HIA	92.35	81.41	83.66	16.90	46.00	3.17	98.94	96.97	97.66	85.65	79.25	74.05	76.14	62.25	58.32	45.00	36.69	21.70
	HSACC	99.35	98.36	98.28	73.35	90.20	67.45	99.00	97.33	97.81	82.04	84.39	73.04	53.73	57.70	30.33	75.33	68.22	58.12
	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	75.12	74.42	65.39	74.47	72.75	63.06
	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	52.00	41.00	33.00	46.00	38.00	22.00
	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	55.86	48.33	37.23	41.24	31.91	18.14
	MICA	98.63	96.19	97.18	10.06	33.34	1.33	98.26	96.22	96.36	29.25	40.30	17.32	48.13	44.10	32.22	26.00	8.17	3.32
	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	59.93	53.97	44.04	60.86	50.88	39.65
	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	72.79	58.54	52.93	28.10	16.56	6.67
	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	56.29	37.82	30.87	40.95	33.13	18.58
	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	48.10	41.40	32.00	35.00	21.60	12.30
	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	52.75	55.30	45.93	44.36	40.97	23.47
	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	55.28	45.84	36.89	45.09	34.78	21.53
	SEPro	**99.39**	**98.56**	**98.69**	**80.10**	**92.10**	**73.51**	**99.50**	**97.93**	**98.11**	**95.47**	**91.13**	**90.37**	**77.49**	**75.21**	**67.22**	**77.52**	**69.04**	**67.22**
0.3	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	70.64	59.09	54.40	43.50	33.88	21.91
	HSACC	99.09	97.36	97.17	72.08	86.25	65.44	98.06	95.11	95.85	79.91	83.60	71.58	58.51	64.65	38.66	72.95	64.75	52.78
	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	70.21	61.19	52.03	64.10	54.85	42.18
	DCG	97.00	92.00	94.00	22.00	56.00	11.00	97.00	92.00	93.00	82.70	80.54	73.96	58.00	49.00	38.00	50.00	39.00	25.00
	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	55.11	47.44	36.63	43.80	37.55	21.30
	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	47.89	44.22	33.27	24.00	10.20	3.57
	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	58.74	50.83	40.43	55.43	45.36	33.02
	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	68.29	59.81	58.71	25.71	16.03	5.31
	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	46.21	29.63	21.87	40.12	30.95	14.28
	APADC	84.00	81.00	74.60	38.80	70.50	27.80	96.40	92.10	92.20	81.65	83.75	70.68	49.30	33.20	43.90	30.50	24.70	8.30
	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	61.88	56.72	45.88	43.01	34.17	20.72
	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	49.46	39.40	29.79	41.34	34.64	19.20
	SEPro	**99.15**	**97.75**	**98.14**	**77.46**	**89.79**	**68.26**	**98.28**	**96.10**	**96.29**	**95.54**	**91.41**	**90.38**	**77.77**	**73.08**	**65.22**	**74.76**	**68.29**	**58.07**
0.5	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	65.71	49.81	45.64	40.00	27.39	17.75
	HSACC	99.01	97.00	97.81	68.08	85.71	57.49	97.26	94.84	95.74	80.22	84.02	72.31	58.49	67.11	44.61	71.38	63.63	52.48
	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	75.81	65.44	59.16	61.90	51.92	38.72
	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	56.00	40.00	34.00	48.00	36.00	21.00
	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	55.86	46.80	36.59	41.24	31.91	18.14
	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	41.88	37.85	24.37	25.00	11.37	4.74
	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	60.94	50.57	41.08	54.76	43.83	31.61
	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	68.36	52.79	47.58	30.48	14.87	5.76
	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	37.86	20.36	15.10	38.57	22.27	11.45
	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	55.50	43.10	32.40	28.70	18.90	6.80
	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	39.31	28.90	19.04	35.61	29.74	13.88
	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	47.91	34.91	25.98	43.34	34.58	20.93
	SEPro	**99.10**	**97.26**	**98.02**	**71.52**	**86.67**	**60.99**	**98.06**	**95.00**	**95.77**	**94.87**	**92.41**	**91.30**	**77.54**	**74.39**	**67.34**	**73.33**	**64.91**	**53.90**

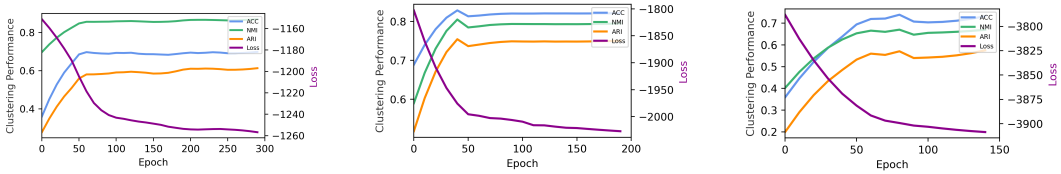
across different types of multi-view datasets. For example, its ACC on 100leaves and Handwritten is 5.26 and 8.12 percentage points higher than the second-best method, respectively. Similar improvements are also observed on Caltech-5V and MSRC-v1, indicating that semantic evidence propagation can effectively exploit complementary information across different views. As the missing rate increases, SEPro maintains a stable performance advantage. At a missing rate of 0.3, SEPro achieves an NMI of 73.08% on Caltech-5V, outperforming the second-best method, HSACC, by 8.43 percentage points. When the missing rate increases to 0.5, SEPro still achieves strong clustering performance across the evaluated datasets. In particular, on Handwritten, its ACC, NMI, and ARI are 8.02, 8.39, and 17.65 percentage points higher than the second-best results, respectively. These results demonstrate that SEPro can effectively handle different levels of view missingness while maintaining good clustering performance as missing information increases.

4.4 Ablation Study

To evaluate the contribution of each learning component in SEPro, we conduct ablation studies on four representative datasets: Handwritten, Caltech-5V, MSRC-v1, and Hdigit. We examine the effects of $\mathcal{L}_{\mathrm{REC}}$, $\mathcal{L}_{\mathrm{CON}}$, $\mathcal{L}_{\mathrm{CSC}}$, and $\mathcal{L}_{\mathrm{ERA}}$, with the results reported in Table 2. The results demonstrate the complementary effects of the proposed objectives. Using $\mathcal{L}_{\mathrm{CSC}}$ alone (M-3) already achieves competitive performance, with ACCs of 66.03%, 58.50%, and 69.52% on Handwritten, Caltech-5V, and MSRC-v1, respectively. Adding $\mathcal{L}_{\mathrm{CON}}$ (M-8) further improves the results, e.g., the ACC on Handwritten increases from 66.03% to 77.31%. Incorporating $\mathcal{L}_{\mathrm{ERA}}$ (M-13) brings additional gains, achieving ACCs of 87.03%, 74.21%, 70.19%, and 90.31% on the four datasets, respectively, indicating the benefit of modeling semantic uncertainty and confidence. Finally, the com-

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

	Loss Components				Handwritten			Caltech-5V			MSRC-v1			Hdigit		
	$\mathcal{L}_{\text{REC}}$	$\mathcal{L}_{\text{CON}}$	$\mathcal{L}_{\text{CSC}}$	$\mathcal{L}_{\text{ERA}}$	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
M-1	✓				18.27	11.67	7.22	41.41	34.22	24.67	28.76	21.22	2.55	21.33	24.47	13.61
M-2		✓			65.26	67.25	54.91	55.04	59.47	33.74	64.38	56.78	45.11	79.71	78.40	74.17
M-3			✓		66.03	70.42	56.67	58.50	43.15	41.63	69.52	62.42	52.46	55.13	71.51	55.15
M-4				✓	36.52	48.63	29.28	24.46	35.13	33.80	21.13	46.76	41.57	27.93	23.57	14.43
M-5			✓	✓	61.09	69.01	54.53	61.07	53.15	42.95	70.48	62.58	52.41	50.48	68.50	50.60
M-6	✓		✓		73.15	74.92	63.21	61.66	52.31	41.85	60.57	52.20	40.03	48.04	65.40	50.60
M-7	✓			✓	59.22	63.20	49.05	49.06	49.17	37.54	56.19	45.28	31.96	28.39	26.18	16.31
M-8		✓	✓		77.31	84.81	71.94	60.44	60.53	40.92	69.88	64.91	47.80	82.02	84.55	74.10
M-9	✓	✓			76.37	74.98	64.18	61.97	67.02	42.22	69.52	60.86	49.80	53.70	67.88	48.38
M-10		✓		✓	68.03	80.52	66.47	73.96	70.17	60.30	62.38	57.61	44.95	94.07	94.18	92.18
M-11	✓		✓	✓	67.01	68.32	55.85	62.96	59.85	49.35	53.90	46.48	33.88	34.81	46.19	28.43
M-12	✓	✓	✓		76.52	77.79	64.99	63.03	62.80	45.39	58.57	53.90	38.83	79.98	81.43	72.64
M-13		✓	✓	✓	87.03	89.83	84.12	74.21	72.68	62.78	70.19	61.56	50.03	90.31	91.10	87.18
M-14	✓	✓		✓	82.07	85.19	77.95	75.16	68.73	60.50	59.90	53.85	40.11	86.78	89.34	83.60
SEPro	✓	✓	✓	✓	**94.87**	**92.41**	**91.30**	**77.54**	**74.39**	**67.34**	**73.33**	**64.91**	**53.90**	**99.10**	**97.26**	**98.02**



(a) 100leaves (b) Caltech-5V (c) MSRC-v1

Figure 2: Clustering metrics (ACC, NMI, ARI) and loss curves over training epochs on the 100leaves, Caltech-5V, and MSRC-v1 datasets.

plete model consistently achieves the highest performance, with ACCs of 94.87%, 77.54%, 73.33%, and 99.10%, respectively. These results confirm that semantic propagation, cross-view consistency, and evidential modeling provide complementary improvements, with semantic propagation being particularly beneficial when multiple views contain richer cross-view semantic interactions.

4.5 Convergence Analysis

To verify the convergence of the proposed method, we conduct experiments on 100leaves, Caltech-5V, and MSRC-v1 by monitoring the total loss and clustering metrics (ACC, NMI, and ARI) over training epochs. As shown in Figure 2(a)–(c), the total loss decreases rapidly and gradually converges, while the clustering metrics consistently improve during training. These results indicate that SEPro achieves stable optimization and progressively learns more reliable and discriminative representations, leading to stable clustering performance.

5 Conclusion

In this paper, we proposed **SEPro**, a semantic evidence propagation framework for incomplete multi-view clustering. SEPro models view-specific semantic evidence with uncertainty and confidence, and propagates it through prototype-mediated high-order relations to learn a robust consensus representation. An uncertainty-aware inference strategy is further employed for missing-view recovery. Experimental results on multiple benchmark datasets demonstrate the effectiveness and robustness of SEPro under different missing rates. Despite its effectiveness, SEPro still depends on the quality of the learned semantic prototypes and uncertainty estimates. Future work will explore adaptive prototype learning and more complex missing-view settings.

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

Yilong Cai, Qianyu Shu, Zhengchun Zhou, and Hua Meng. Incomplete multi-view clustering based on information fusion with self-supervised learning. *Information Fusion*, 117:102849, 2025.

Guoqing Chao, Kaixin Xu, Xijiong Xie, and Yongyong Chen. Global graph propagation with hierarchical information transfer for incomplete contrastive multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 15713–15721, 2025.

Yuzhuo Dai, Jiaqi Jin, Zhibin Dong, Siwei Wang, Xinwang Liu, En Zhu, Xihong Yang, Xinbiao Gan, and Yu Feng. Imputation-free and alignment-free: Incomplete multi-view clustering driven by consensus semantic learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 5071–5081, 2025.

Xiaojian Ding, Lin Zhao, Xian Li, and Xiaoying Zhu. Incomplete multi-view clustering via hierarchical semantic alignment and cooperative completion. *arXiv preprint arXiv:2510.13887*, 2025.

Yiming Du, Ziyu Wang, Jian Li, Rui Ning, and Lusi Li. Deep incomplete multi-view clustering via hierarchical imputation and alignment. *arXiv preprint arXiv:2601.09051*, 2026.

Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *2004 conference on computer vision and pattern recognition workshop*, pp. 178–178. IEEE, 2004.

Cong Feng, Ao Li, Haoyue Xu, Hailu Yang, and Xinwang Liu. Deep incomplete multiview clustering via local and global pseudo-label propagation. *IEEE Transactions on Neural Networks and Learning Systems*, 36(5):8684–8698, 2025. doi: 10.1109/TNNLS.2024.3411294.

Hang Gao, Cheng Liu, Hongming Sun, Gaoyang Li, Ying Li, You Zhou, and Wei Du. Incomplete multi-view clustering with cross-view generation via pre-trained transformer. *Pattern Recognition*, pp. 112166, 2025.

Binqiang Huang, Zhijie Huang, Shoujie Lan, Qinghai Zheng, and Yuanlong Yu. Incomplete multiview 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.

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*, 2023.

Xiaocui Li, Guoliang Li, Xinyu Zhang, Yangtao Wang, Qingyu Shi, and Wei Liang. Incomplete multi-view clustering via local reasoning and correlation analysis. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, pp. 484–492, 2025.

Renjie Lin, Shide Du, Shiping Wang, and Wenzhong Guo. Consistent graph embedding network with optimal transport for incomplete multi-view clustering. *Information Sciences*, 647:119418, 2023a.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11174–11183, 2021.

Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4447–4461, 2023b. doi: 10.1109/TPAMI.2022.3197238.

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.

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.

Jingyu Pu, Chenhang Cui, Xinyue Chen, Yazhou Ren, Xiaorong Pu, Zhifeng Hao, Philip S Yu, and Lifang He. Adaptive feature imputation with latent graph for deep incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 14633–14641, 2024.

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, 2024.

Xinxin Wang, Yongshan Zhang, Xiaochen Yuan, and Yicong Zhou. Cross-view anchor graph learning and factorization for incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 26570–26578, 2026.

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, Shijie Deng, Waikeung Wong, Guoqing Chao, Chao Huang, Lunke Fei, and Yong Xu. Diffusion-based missing-view generation with the application on incomplete multi-view clustering. In *Forty-first international conference on machine learning*, 2024.

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*, 28:196–208, 2026. doi: 10.1109/TMM.2025.3623519.

Shanshan Wu and Gui-Fu Lu. Incomplete multi-view clustering with cross-graph diversity and improved nonconvex tensor regularizer. *Expert Systems with Applications*, 297:129535, 2026.

Gehui Xu, Jie Wen, Chengliang Liu, Bing Hu, Yicheng Liu, Lunke Fei, and Wei Wang. Deep variational incomplete multi-view clustering: Exploring shared clustering structures. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 16147–16155, 2024.

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.

Zhe Xue, Junping Du, Hai Zhou, Zhongchao Guan, Yunfei Long, Yu Zang, and Meiyu Liang. Robust diversified graph contrastive network for incomplete multi-view clustering. In *Proceedings of the 30th ACM international conference on multimedia*, pp. 3936–3944, 2022.

Zhe Xue, Yawen Li, Zhongchao Guan, Wenling Li, Meiyu Liang, and Hai Zhou. Robust multi-graph contrastive network for incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 27: 2747–2759, 2024.

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, 2025. doi: 10.1109/TCSVT.2024.3508785.

Zuyuan Yang, Naiyao Liang, Wei Yan, Zhenni Li, and Shengli Xie. Uniform distribution nonnegative 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, 2025. doi: 10.1109/TKDE.2025.3568795.

Honglin Yuan, Shiyun Lai, Xingfeng Li, Jian Dai, Yuan Sun, and Zhenwen Ren. Robust prototype completion for incomplete multi-view clustering. In *Proceedings of the 32nd ACM international conference on multimedia*, pp. 10402–10411, 2024.

Honglin Yuan, Yuan Sun, Fei Zhou, Jing Wen, Shihua Yuan, Xiaojian You, and Zhenwen Ren. Prototype matching learning for incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 34:828–841, 2025. doi: 10.1109/TIP.2025.3529378.

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.

Yuanyang Zhang, Yijie Lin, Xinhang Wan, Jie Xu, Li Yao, Weiqing Yan, and Chang Tang. Structure-aware conditional diffusion generation for incomplete multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 38(5):3168–3181, 2026. doi: 10.1109/TKDE.2026.3673320.

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.

A Appendix

B Algorithmic Details

The following algorithms provide the detailed procedures for obtaining the consensus representation $\mathbf{H}$ through semantic evidence propagation (Algorithm 1) and training the overall SEPro framework (Algorithm 2), respectively.

Algorithm 1 Semantic Evidence Propagation

Input: Multi-view latent representations $\{\mathbf{Z}^v\}_{v=1}^V$, semantic prototypes $C = \{\mathbf{c}_k\}_{k=1}^K$, indicator matrix $\mathbf{M}$, evidential encoders $E_v(\cdot)$
Output: Consensus representation $\mathbf{H}$
Construct the observed-view context $\widetilde{\mathbf{z}}_i$
Encode view-specific semantic evidence $(\boldsymbol{\mu}_i^v, \sigma_i^{2(v)}, u_i^v)$
Compute prototype affinities a_{ivk}
Compute evidence propagation weights $w_{ivk} = M_i^v u_i^v a_{ivk}$
Aggregate semantic evidence into prototypes $\mathbf{e}_k$
Compute normalized sample-prototype weights $\widetilde{w}_{ik}$
Obtain consensus representations $\mathbf{h}_i$
Return $\mathbf{H} = \{\mathbf{h}_i\}_{i=1}^N$

Algorithm 2 SEPro: Semantic Evidence Propagation for Incomplete Multi-View Clustering

Input: Incomplete multi-view dataset $\{\mathbf{X}^v\}_{v=1}^V$, indicator matrix $\mathbf{M}$, number of epochs E
Parameters: $\lambda_1, \lambda_2, \lambda_3, \lambda_4$
Output: Clustering results
for $epoch = 1$ to E **do**
  Learn view-specific representations via $\mathcal{L}_{\mathrm{REC}}$
  Optimize cross-view semantic consistency via $\mathcal{L}_{\mathrm{CON}}$
  Encode view-specific semantic evidence $(\boldsymbol{\mu}_i^v, \sigma_i^{2(v)}, u_i^v)$
  Propagate semantic evidence through Algorithm 1
  Obtain consensus representation $\mathbf{H}$
  Apply consensus semantic regularization $\mathcal{L}_{\mathrm{CSC}}$
  Infer missing-view evidence from observed-view context
  Calibrate missing-view representations according to uncertainty and confidence
  Construct the unified representation $\{\overline{\mathbf{z}}_i^v\}_{v=1}^V$
  Jointly optimize all objectives via Eq. (24)
end for
Perform k-means clustering on the unified representations

C Theoretical Analysis

This section provides a concise theoretical characterization of the three core mechanisms in SEPro: confidence-weighted semantic aggregation, confidence-guided evidence propagation, and uncertainty-aware missing-view inference. We focus on theoretical properties that are directly connected to the proposed semantic evidence propagation and missing-view inference mechanisms.

C.1 Optimality of Semantic Evidence Aggregation

Given the propagated sample-prototype weights, the consensus representation in Eq. (18) is obtained by a weighted aggregation of semantic prototypes. The following proposition shows that this aggregation has a direct optimization interpretation.

Proposition 1. For a given sample i, assume that $w_{ik} \geq 0$ for all $k \in \{1, \ldots, K\}$ and $\sum_{k=1}^K w_{ik} > 0$. Then the consensus representation

$$h_i = \sum_{k=1}^{K} \widetilde{w}_{ik} e_k, \qquad \widetilde{w}_{ik} = \frac{w_{ik}}{\sum_{k'=1}^K w_{ik'}} \tag{25}$$

is the unique global minimizer of the weighted semantic aggregation objective

$$h_i^* = \arg\min_{h_i} \sum_{k=1}^{K} w_{ik} \|h_i - e_k\|_2^2. \tag{26}$$

Proof. Consider the objective

$$J_i(h_i) = \sum_{k=1}^{K} w_{ik} \|h_i - e_k\|_2^2. \tag{25}$$

Expanding the quadratic terms gives

$$J_i(h_i) = \left(\sum_{k=1}^{K} w_{ik} \right) \|h_i\|_2^2 - 2h_i^\top \left(\sum_{k=1}^{K} w_{ik} e_k \right) + C, \tag{26}$$

where C is independent of h_i. Therefore,

$$\nabla_{h_i} J_i(h_i) = 2 \left(\sum_{k=1}^{K} w_{ik} \right) h_i - 2 \sum_{k=1}^{K} w_{ik} e_k. \tag{27}$$

Setting the gradient to zero yields

$$h_i^* = \frac{\sum_{k=1}^{K} w_{ik} e_k}{\sum_{k=1}^{K} w_{ik}} = \sum_{k=1}^{K} \widetilde{w}_{ik} e_k. \tag{28}$$

Since $\sum_{k=1}^{K} w_{ik} > 0$, the objective is strictly convex in h_i, and thus the stationary point is the unique global minimizer. □

This result shows that the consensus representation in Eq. (18) is not an arbitrary weighted combination of prototypes, but the unique confidence-aware semantic center induced by the propagated evidence.

C.2 Confidence Monotonicity in Evidence Propagation

The core of semantic evidence propagation is to regulate the contribution of each observation according to its availability, semantic affinity, and confidence. Recall from Eq. (14) that

$$w_{ivk} = M_i^v u_i^v a_{ivk}, \tag{27}$$

where M_i^v denotes the availability of view v for sample i, u_i^v denotes the confidence of the corresponding semantic evidence, and a_{ivk} denotes its semantic affinity to prototype k.

Proposition 2. For fixed M_i^v and a_{ivk}, the propagation weight w_{ivk} is monotonically non-decreasing with respect to the confidence score u_i^v. Specifically,

$$\frac{\partial w_{ivk}}{\partial u_i^v} = M_i^v a_{ivk} \geq 0. \tag{28}$$

Moreover, when $M_i^v = 0$, we have $w_{ivk} = 0$ for all k.

Proof. The result follows directly from Eq. (27). Since $M_i^v \in \{0, 1\}$ and $a_{ivk} \geq 0$, differentiating w_{ivk} with respect to u_i^v gives

$$\frac{\partial w_{ivk}}{\partial u_i^v} = M_i^v a_{ivk} \geq 0. \tag{29}$$

When $M_i^v = 0$, Eq. (27) immediately gives $w_{ivk} = 0$. □

Proposition 2 formalizes the intended reliability control of SEPro: higher-confidence observations receive stronger propagation weights, whereas unavailable views contribute no semantic evidence to the propagation process.

C.3 Uncertainty-Aware Missing-View Inference

For a missing view, SEPro calibrates the predicted semantic evidence according to its predictive uncertainty and confidence. Specifically, the uncertainty-suppressed semantic representation and the confidence modulation factor are defined as

$$\widetilde{\mu}_i^v = \frac{\mu_i^v}{1 + \sigma_i^{2(v)}}, \qquad \alpha_i^v = \mathrm{sigmoid}(u_i^v), \tag{29}$$

and the completed representation is given by

$$\widehat{z}_i^v = \alpha_i^v \widetilde{\mu}_i^v. \tag{30}$$

Proposition 3. For fixed μ_i^v, the magnitude of the uncertainty-suppressed semantic evidence is monotonically non-increasing with respect to the predictive uncertainty $\sigma_i^{2(v)}$:

$$\frac{\partial \left\|\widetilde{\mu}_i^v\right\|_2}{\partial \sigma_i^{2(v)}} = -\frac{\left\|\mu_i^v\right\|_2}{\left(1+\sigma_i^{2(v)}\right)^2} \leq 0. \tag{31}$$

Proof. From Eq. (29),

$$\left\|\widetilde{\mu}_i^v\right\|_2 = \frac{\left\|\mu_i^v\right\|_2}{1+\sigma_i^{2(v)}}. \tag{30}$$

Differentiating with respect to $\sigma_i^{2(v)}$ gives

$$\frac{\partial \left\|\widetilde{\mu}_i^v\right\|_2}{\partial \sigma_i^{2(v)}} = -\frac{\left\|\mu_i^v\right\|_2}{\left(1+\sigma_i^{2(v)}\right)^2} \leq 0, \tag{31}$$

which proves the claim. □

Therefore, increasing predictive uncertainty can only reduce the magnitude of the semantic evidence used for missing-view completion. Together with the confidence modulation factor α_i^v, the inference rule in Eq. (30) explicitly suppresses uncertain predictions while retaining stronger contributions from confident semantic evidence.

These three propositions characterize the essential theoretical properties of SEPro: the consensus representation is an optimal confidence-weighted semantic aggregation, confidence directly regulates evidence propagation, and predictive uncertainty suppresses unreliable missing-view evidence.

D Additional Experimental Analysis

D.1 Experimental Results under High Missing Rates

To verify the effectiveness of the proposed method, we conducted clustering performance comparisons with 12 state-of-the-art algorithms on six benchmark datasets under a high missing rate setting of 0.7. The experimental results, as shown in Table 1, demonstrate that the proposed SEPro achieves the best overall performance across three evaluation metrics, exhibiting superior capability in high-missing multi-view clustering scenarios. Specifically, except for a marginally lower performance on the NMI metric of the MNIST_USPS dataset (91.35% versus 91.55% achieved by the best-performing HSACC), SEPro consistently attains leading results across almost all metrics on most datasets. In particular, on the Caltech-5V and Handwritten datasets, which are characterized by complex structures and large semantic discrepancies, the proposed method yields significant performance improvements. For instance, on the challenging Caltech-5V dataset, SEPro achieves ACC and NMI scores of 77.16% and 76.43%, respectively, surpassing all competing methods by more than 10%.

These results indicate that the proposed approach can effectively alleviate the issue of incomplete information caused by high missing rates. In summary, the experimental results fully validate the effectiveness of the proposed cross-view consistency modeling, hypergraph-based higher-order semantic fusion, and uncertainty-aware collaborative optimization mechanisms. Moreover, they further demonstrate that SEPro exhibits stronger robustness and generalization ability in complex missing-data scenarios.

D.2 Ablation Study

As shown in Table 4, we conduct ablation studies on the MNIST_USPS and 100leaves datasets under a missing rate of 0.5 to evaluate the contribution of each component. Specifically, after introducing

Table 3: Clustering performance (%) on six datasets under the missing rate setting of 0.7.

Ratio	Method	Hdigit			100leaves			MNIST_USPS			Handwritten			Caltech-5V			MSRC-v1		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
0.7	DIMVC-HIA	75.40	62.68	58.38	11.75	33.34	0.35	93.52	85.43	86.17	78.75	75.10	67.69	53.14	35.96	27.81	33.50	21.90	8.41
	HSACC	97.46	93.75	94.58	61.68	80.02	40.88	96.16	**91.55**	91.98	90.03	87.22	80.79	61.93	66.21	46.12	53.91	45.80	34.14
	DITA-IMVC	88.06	78.62	76.32	26.56	57.60	13.28	93.84	86.54	86.90	72.24	69.71	59.06	66.60	54.80	46.11	51.90	43.28	29.26
	DCG	92.00	81.00	83.00	14.00	42.00	5.00	78.00	76.00	70.00	79.75	74.00	67.54	55.00	39.00	32.00	44.00	35.00	21.00
	IMC-MCL	95.12	88.05	89.50	25.94	57.41	13.33	83.28	82.91	77.99	60.44	56.13	43.36	56.50	45.08	36.08	41.24	31.91	18.14
	MICA	61.21	61.35	49.69	9.94	32.63	2.14	47.46	64.74	49.27	28.50	40.36	19.19	42.58	38.39	25.52	24.50	6.49	2.43
	DIVIDE	91.28	80.55	81.70	59.59	77.10	43.85	87.47	75.49	74.58	85.43	75.47	71.63	61.04	48.19	40.04	48.57	39.16	25.44
	MRL_CAL	12.59	0.92	0.38	9.94	44.78	4.66	54.94	55.05	45.69	47.30	48.21	34.69	57.43	48.40	35.34	27.62	14.36	5.77
	IMVC-IE	66.33	42.67	39.73	4.69	15.55	0.38	64.88	43.40	38.70	65.20	60.42	50.07	20.57	4.88	2.39	21.43	3.94	1.36
	APADC	86.30	72.50	72.20	28.70	63.10	15.90	89.90	81.40	78.50	57.03	65.26	50.90	54.60	41.40	32.50	27.70	21.30	4.20
	MCAC	24.78	16.48	7.55	27.96	58.19	11.35	88.49	82.18	80.80	56.20	50.08	38.04	36.83	32.71	19.22	31.42	26.27	10.09
	ProImp	82.97	73.36	63.36	34.27	65.00	19.27	88.48	78.55	76.09	74.16	66.76	59.08	51.69	37.63	30.16	46.67	36.07	22.88
	SEPro	**97.98**	**94.35**	**95.57**	**63.24**	**80.32**	**43.90**	**96.35**	91.35	**92.12**	**91.52**	**89.17**	**86.27**	**77.16**	**76.43**	**69.19**	**55.33**	**51.15**	**35.23**

Table 4: Loss Ablation Experiments on Two Data Sets with the Missing Rate of 0.5

	Loss Components				MNIST_USPS			100leaves		
	$\mathcal{L}_{\text{REC}}$	$\mathcal{L}_{\text{CON}}$	$\mathcal{L}_{\text{CSC}}$	$\mathcal{L}_{\text{ERA}}$	ACC	NMI	ARI	ACC	NMI	ARI
M-1	✓				22.72	25.99	15.39	31.39	67.99	23.95
M-2		✓			88.81	86.78	80.31	59.19	72.89	46.40
M-3			✓		46.04	59.86	42.41	32.35	67.88	21.38
M-4				✓	25.20	28.77	15.64	19.40	53.21	9.40
M-5			✓	✓	44.14	61.09	42.61	29.71	66.87	20.01
M-6	✓		✓		57.25	67.32	50.00	33.69	69.05	24.51
M-7	✓			✓	33.27	30.43	19.27	34.99	70.16	30.93
M-8		✓	✓		90.38	88.14	79.50	67.53	83.36	57.46
M-9	✓	✓			77.09	83.21	73.67	69.96	84.70	52.84
M-10		✓		✓	64.41	74.77	58.95	51.71	78.37	39.95
M-11	✓		✓	✓	41.47	50.56	32.59	30.79	68.11	21.48
M-12	✓	✓	✓		83.16	84.35	75.04	41.59	74.47	32.12
M-13		✓	✓	✓	95.22	93.82	92.91	67.74	85.63	58.02
M-14	✓	✓		✓	95.46	93.61	92.62	57.23	75.49	47.70
SEPro	✓	✓	✓	✓	**98.06**	**95.00**	**95.77**	**71.52**	**86.67**	**60.99**

the contrastive learning component $\mathcal{L}_{\text{CON}}$, the model performance is significantly improved. This demonstrates that by maximizing cross-view mutual information, the proposed method can effectively enhance the consistency and complementarity of shared information across views, thereby providing a more stable and reliable semantic representation for subsequent high-level semantic modeling and view completion.

Furthermore, when the hypergraph-based high-order semantic fusion module $\mathcal{L}_{\text{CSC}}$ (M-8) is incorporated, consistent performance gains are observed on both the MNIST_USPS and 100leaves datasets compared with the variant M-2 without this component. This result indicates that the proposed high-order semantic fusion module can effectively model complex inter-view structural relationships and capture latent global consistency, thereby significantly enhancing the discriminative power of the learned representations.

Finally, with the introduction of the uncertainty-aware view completion module $\mathcal{L}_{\text{ERA}}$, the complete model (SEPro) achieves the best performance on both datasets. This confirms that the proposed uncertainty-aware completion strategy can effectively recover critical semantic information and, in combination with the cross-view alignment and high-order semantic fusion modules, form a synergistic optimization process that further improves the overall robustness and clustering performance of the model.

D.3 Parameter Analysis

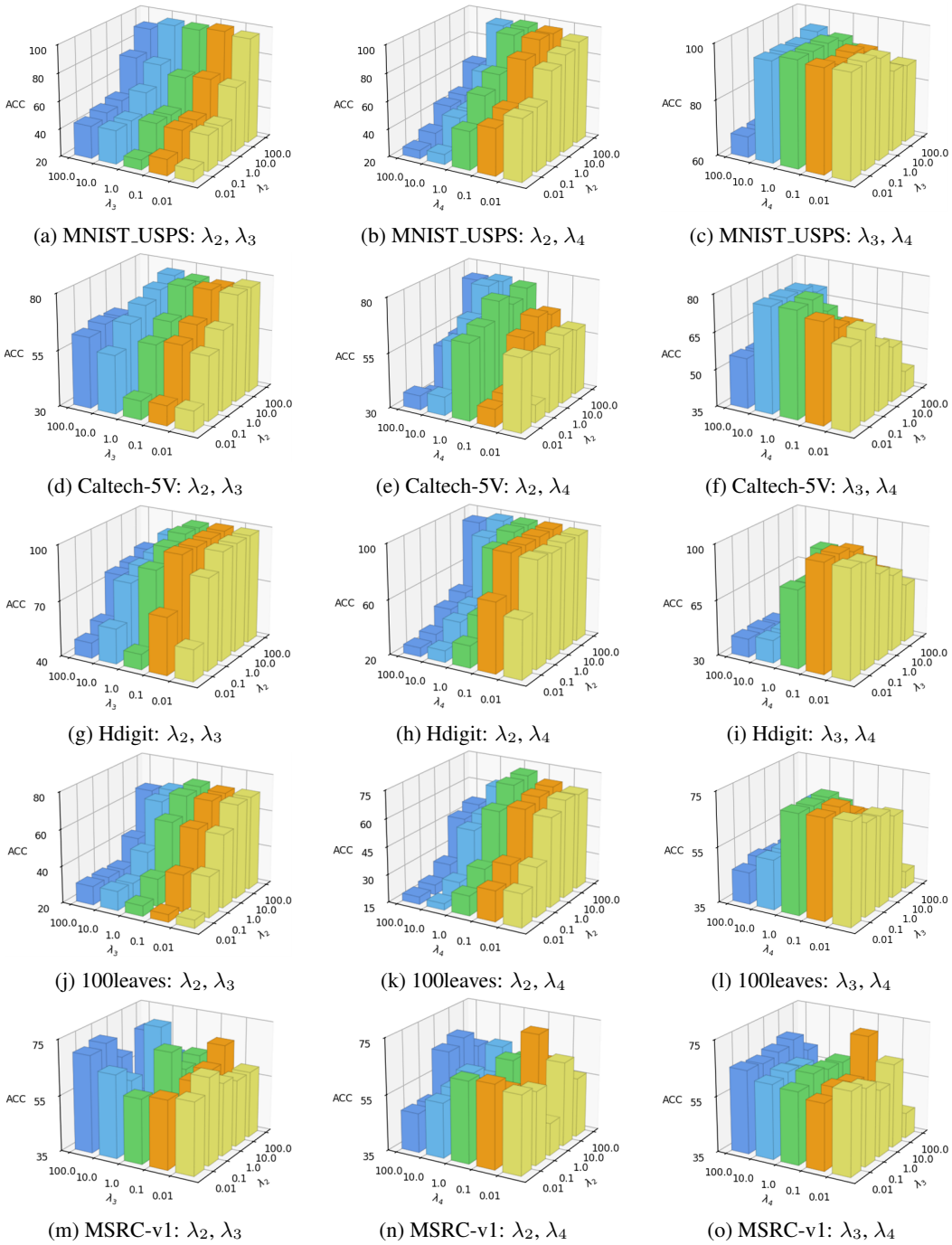
To further investigate the influence of the loss-function hyperparameters on the performance of the proposed method, we conducted parameter sensitivity experiments on five additional datasets,

including MNIST_USPS, Caltech-5V, Hdigit, 100leaves, and MSRC-v1. Following the same experimental settings as described in the main text, the reconstruction loss weight λ_1 was fixed at 0.1, while λ_2, λ_3, and λ_4 were varied in pairwise combinations within the range $\{0.01, 0.1, 1, 10, 100\}$ to perform sensitivity analysis. The experimental results are illustrated in Figure 3. First, λ_2 and λ_3 exhibit a significant synergistic effect. The model achieves the best and most stable performance when both parameters are set within moderate ranges, whereas excessively large or small values lead to noticeable performance degradation. This observation indicates that a proper balance between cross-view consistency learning and hypergraph-based high-order semantic modeling is essential for fully exploiting the advantages of the two-stage contrastive learning framework. Furthermore, the effectiveness of the uncertainty-guided view completion module relies heavily on high-quality representation learning. Only when λ_2 and λ_3 are assigned appropriate values to ensure reliable representation quality can adjusting λ_4 to a moderate range (e.g., 0.1–1.0) further improve clustering performance. By jointly optimizing λ_2, λ_3, and λ_4, the proposed method is able to effectively integrate cross-view consistency learning, hypergraph high-order semantic fusion, and uncertainty-aware view completion into a unified framework, thereby learning more robust and discriminative clustering representations under highly incomplete-view scenarios.

D.4 VISUALIZATION

As shown in Figure 4, we conduct *t*-SNE visualization experiments on six benchmark datasets to further investigate the evolution of the learned common representations during training. Specifically, subfigures (a)–(d) illustrate the clustering distributions on the MNIST-USPS dataset at different training epochs; subfigures (e)–(h) present the results on the Caltech dataset; subfigures (i)–(l) show the results on the Hdigit dataset; subfigures (m)–(p) correspond to the 100leaves dataset; subfigures (q)–(t) present the results on the MSRC-v1 dataset; and subfigures (u)–(x) show the results on the Handwritten dataset.

From the visualizations, it can be observed that the learned representations progressively form more compact and distinguishable cluster structures as training proceeds. Specifically, samples from the same class gradually become more concentrated, while the boundaries between different classes become increasing



(a) MNIST_USPS: λ_2, λ_3 (b) MNIST_USPS: λ_2, λ_4 (c) MNIST_USPS: λ_3, λ_4

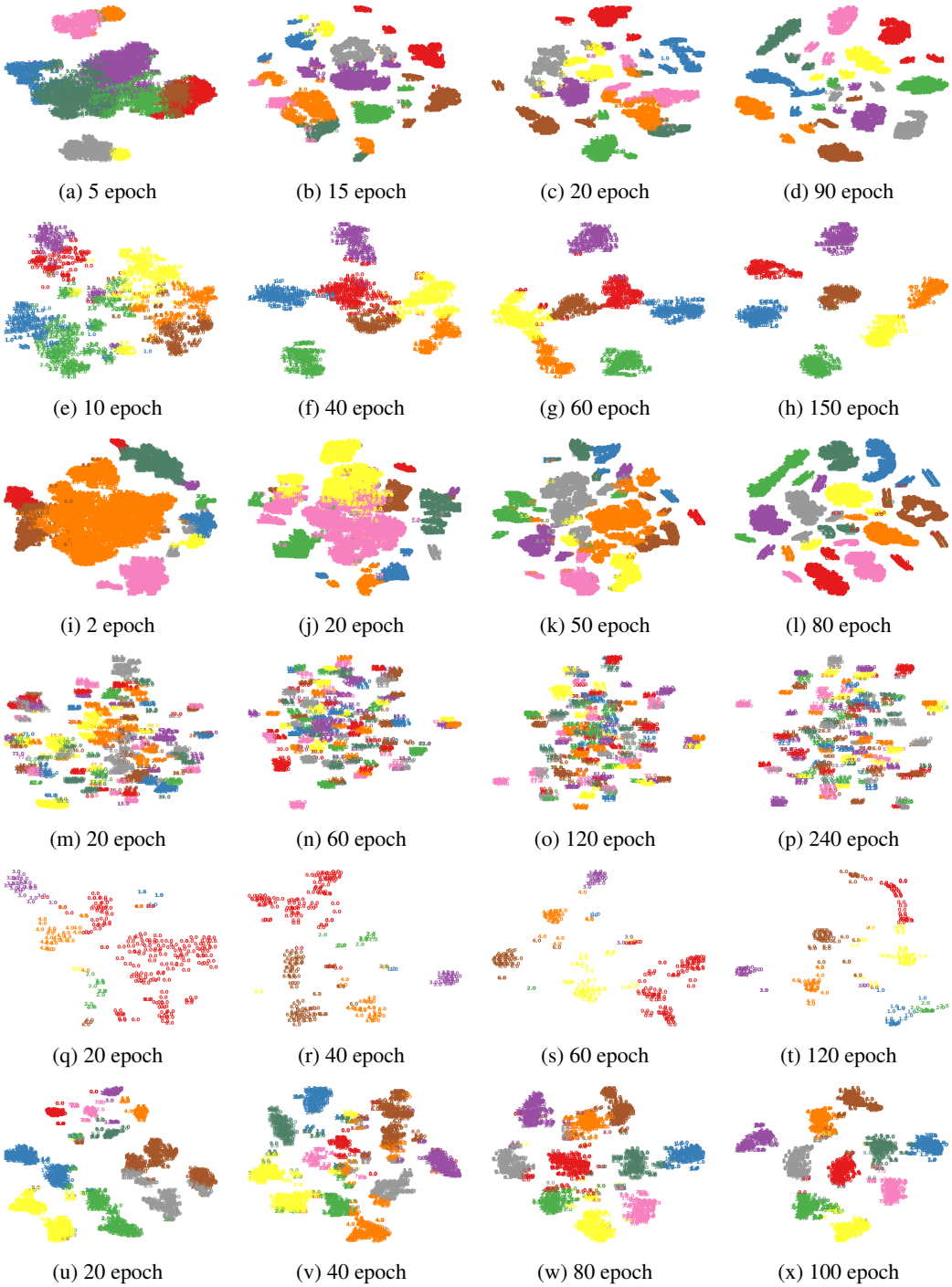
(d) Caltech-5V: λ_2, λ_3 (e) Caltech-5V: λ_2, λ_4 (f) Caltech-5V: λ_3, λ_4

(g) Hdigit: λ_2, λ_3 (h) Hdigit: λ_2, λ_4 (i) Hdigit: λ_3, λ_4

(j) 100leaves: λ_2, λ_3 (k) 100leaves: λ_2, λ_4 (l) 100leaves: λ_3, λ_4

(m) MSRC-v1: λ_2, λ_3 (n) MSRC-v1: λ_2, λ_4 (o) MSRC-v1: λ_3, λ_4

Figure 3: Parameter sensitivity analysis on other datasets with a missing rate of 0.5 and $\lambda_1 = 0.1$.



(a) 5 epoch (b) 15 epoch (c) 20 epoch (d) 90 epoch

(e) 10 epoch (f) 40 epoch (g) 60 epoch (h) 150 epoch

(i) 2 epoch (j) 20 epoch (k) 50 epoch (l) 80 epoch

(m) 20 epoch (n) 60 epoch (o) 120 epoch (p) 240 epoch

(q) 20 epoch (r) 40 epoch (s) 60 epoch (t) 120 epoch

(u) 20 epoch (v) 40 epoch (w) 80 epoch (x) 100 epoch

Figure 4: t-SNE visualizations of clustering results on all datasets with increasing training iterations. Each row corresponds to one dataset.