PROTOTYPE-DRIVEN RELIABILITY-AWARE SEMANTIC CONSENSUS LEARNING FOR INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete multi-view clustering (IMVC) faces challenges from missing views, heterogeneous view distributions, and unequal view quality, making it difficult to learn reliable and consistent semantic consensus representations. To address these issues, we propose a Prototype-driven Reliability-aware Semantic Consensus Learning (PRSCL) framework, which centers on consensus semantic prototypes to align semantic representations across different views and ultimately achieves the fusion of reliable semantic representations. Specifically, PRSCL first employs a reliability-aware gating mechanism to adaptively evaluate the quality of available views and perform instance-level weighted fusion, thereby producing reliable consensus semantic representations and semantic prototypes. PRSCL then employs entropy-regularized optimal transport to derive reliable soft assignments between samples and consensus semantic prototypes, while enforcing cross-view consistency at the clustering level through bidirectional mutual supervision. By jointly modeling feature consistency, view reliability, and prototype-driven semantic structure, PRSCL learns robust consensus semantic representations without explicitly imputing missing views. Extensive experiments on benchmark datasets demonstrate its effectiveness under various missing-view conditions.

1 INTRODUCTION

Data collected in real-world applications often originates from diverse observational dimensions, heterogeneous sensors, or different acquisition mechanisms, and thus naturally exhibits multi-source and heterogeneous characteristics, commonly referred to as multi-view data (Chao et al., 2021; Liu et al., 2026). Multi-view clustering (MVC) is a representative unsupervised learning paradigm that exploits the complementary and consistent information across multi-views to partition samples into meaningful clusters without relying on label information (Chen et al., 2023; Wang et al., 2026b). However, complete multi-view data are difficult to obtain due to sensor failures, privacy restrictions, and heterogeneous data collection processes, resulting in the prevalent problem of incomplete multi-view clustering (IMVC), where some samples are observed only in a subset of views. Such missing views not only reduce the available information for individual samples but also make it challenging to learn reliable and semantically consistent representations across views. Therefore, effectively exploiting available observations while mitigating the adverse effects of missing views has become a fundamental challenge in IMVC (Wen et al., 2023; Wang et al., 2025).

Benefiting from the strong representation capability of deep neural networks, deep incomplete multi-view clustering (DIMVC) has attracted increasing attention. Existing methods can be broadly categorized into three paradigms, as illustrated in Figure 1. Filling-driven methods explicitly recover missing views through cross-view neighbor imputation or neural generation (Wong et al., 2023; Wang et al., 2023), but their performance is highly dependent on completion quality, and inaccurate recovery may introduce errors that propagate to subsequent clustering. Probabilistic assignment-driven methods bypass explicit completion by independently learning view-specific category distributions and fusing them for final prediction (Wu et al., 2025; Wang et al., 2026a); however, their predictions can become unreliable under severe incompleteness, while the lack of cross-view semantic regularization may lead to inconsistent clustering decisions. Prototype-driven methods (Yu et al., 2024; Zhu et al., 2026) exploit sample-prototype relationships to preserve view-specific se-

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.

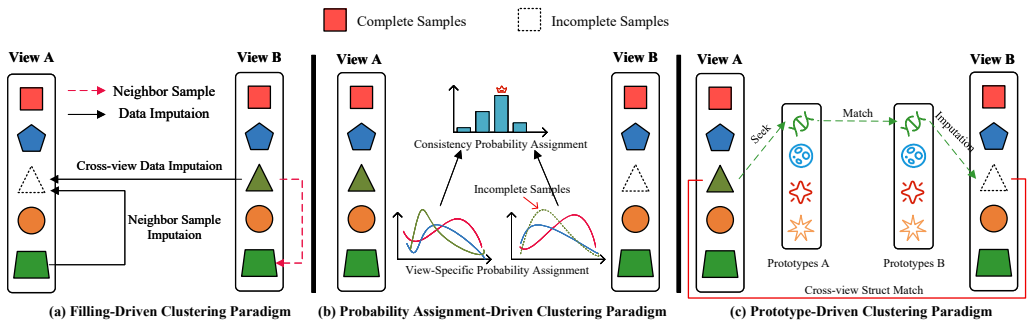

Figure 1: An overview of three existing deep incomplete multi-view clustering paradigms: (a) filling-driven methods recover missing views yet risk error propagation; (b) probabilistic assignment-driven methods avoid explicit completion but yield unstable clustering results under severe view incompleteness; and (c) prototype-driven methods model sample-prototype relations but fail to eliminate semantic discrepancies across view-specific prototypes.

mantic structures, and are regarded as the most promising paradigm among the three; however, they still struggle with the semantic discrepancy among view-specific prototypes under severe missingness, and existing prototype-based consensus learning further leaves two important issues insufficiently addressed. First, existing prototype-based methods construct consensus prototypes to establish a shared semantic space without explicit view completion, yet they generally lack instance-level view reliability modeling, and their consensus learning is often built upon graph modularity-based clustering (e.g., FreeCLS(Dai et al., 2025)). Second, consensus learning mainly relies on graph modularity-based clustering. Moreover, the complementary prototype-based assignment structures across views are not fully exploited for cross-view semantic supervision. These limitations motivate us to jointly consider view reliability and prototype-driven cross-view semantic consistency in consensus representation learning, thereby moving beyond existing paradigms that either rely on potentially unreliable missing-data recovery or lack effective mechanisms for jointly exploiting view reliability and cross-view semantic structures.

To address these challenges, we propose a novel Prototype-driven Reliability-aware Semantic Consensus Learning (PRSCL) framework for incomplete multi-view clustering. As illustrated in Figure 2, PRSCL is built around two key ideas, i.e., prototype-driven cross-view semantic representation alignment and reliability-aware semantic fusion. Specifically, PRSCL first maps view-specific representations into a common semantic space and constructs shared consensus prototypes, which serve as a unified semantic reference for aligning the semantic representations across views. To obtain reliable consensus semantics, a reliability-aware gating mechanism then adaptively estimates the quality of each available view and performs instance-level weighted fusion, assigning greater contributions to reliable views while discarding missing ones. Based on consensus prototypes, entropy-regularized optimal transport produces reliable soft sample-prototype assignments, which are further refined via bidirectional mutual supervision to enhance clustering-level consistency across views. By jointly aligning cross-view semantic prototypes and adaptively fusing reliable semantic representations, PRSCL learns robust and discriminative consensus representations without explicitly imputing missing views. The main contributions of this paper are summarized as follows:

- We develop a prototype-driven cross-view semantic alignment mechanism that constructs shared consensus semantic prototypes as a unified semantic reference, enabling consistent semantic representation learning across different views.
- To adaptively weights available views for reliable consensus construction and exploits prototype-based soft assignments to enhance cross-view clustering consistency, we design a reliability-aware semantic fusion and bidirectional mutual supervision strategy.
- We analyze the performance of PRSCL with extensive experiments, whose results demonstrate that it consistently and clearly outperforms state-of-the-art methods across four benchmark datasets under varying missing rates.

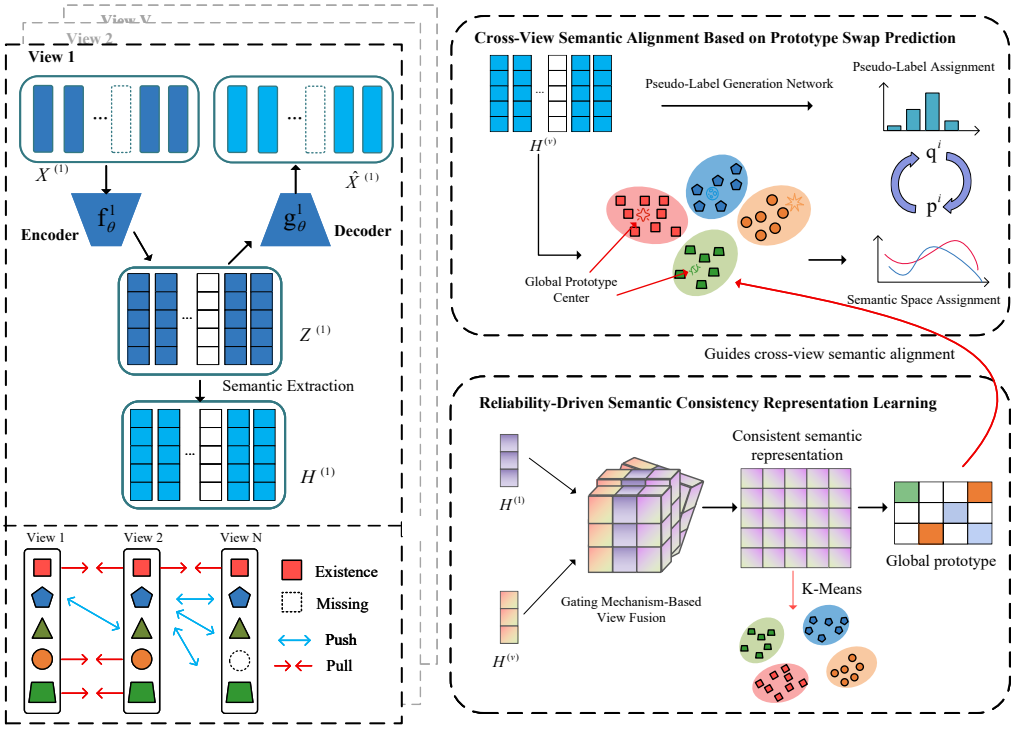

Figure 2: The architecture of our proposed PRSCL framework consists of three modules: (a) a semantic representation learning module that encodes observations from each view into view-specific semantic representations; (b) a reliability-aware semantic consensus learning module that estimates the reliability of each view and adaptively aggregates reliable semantic representations to learn a robust consensus representation, while suppressing noisy and incomplete observations; and (c) a cross-view semantic alignment module that exchanges semantic prototypes across views to predict missing-view semantics and aligns semantic representations of different views in a shared space.

2 METHODOLOGY

2.1 NOTATIONS

We consider a multi-view dataset $\mathcal{X} = \{\mathbf{X}^v \in \mathcal{R}^{N\times D_v}\}_{v=1}^V$ comprising N instances, where each instance is characterized by V views with D_v dimensions. To handle incomplete data, we define a binary indicator matrix $\mathbf{M} \in \{0,1\}^{N\times V}$. Specifically, the entry $\mathbf{M}_{iv}$ equals 1 if the observation for the i-th sample in the v-th view is available, and 0 otherwise. It is assumed that no sample is missing in all views simultaneously, i.e., $\forall i \in \{1,\ldots,N\}, \sum_{v=1}^V \mathbf{M}_{iv} \geq 1$. Our objective is to partition these N samples, which may suffer from missing views, into K distinct clusters.

2.2 INTRA-VIEW REPRESENTATION RECONSTRUCTION AND ENHANCEMENT

To mitigate the instability of clustering caused by similarity measurements on high-dimensional manifolds, we leverage autoencoders to transform the observed views $\mathbf{X}^v$ into view-specific representations $\mathbf{Z}^v$. To accommodate the heterogeneity and distribution shifts across different views, we employ a dedicated encoder-decoder pair $\{\mathcal{E}_v, \mathcal{D}_v\}_{v=1}^V$ for each view. These networks are trained to minimize the following reconstruction objective:

$$\mathcal{L}_{rec} = \sum_{v=1}^{V} \|M \odot (\mathbf{X}^v - \mathcal{D}_v(\mathbf{Z}^v))\|_F^2 = \sum_{v=1}^{V} \sum_{i=1}^{N_v} \|M_{i,v} \cdot (\mathbf{x}_i^v - \mathcal{D}_v(\mathcal{E}_v(\mathbf{x}_i^v)))\|_2^2, \quad (1)$$

where $\mathcal{E}_v(\mathbf{X}^v;\theta_v) : \mathbf{X}^v \in \mathcal{R}^{N_v\times D_v} \to \mathbf{Z}^v \in \mathcal{R}^{N_v\times d}$ and $\mathcal{D}_v(\mathbf{Z}^v;\phi_v) : \tilde{\mathbf{Z}}^v \in \mathcal{R}^{N_v\times d} \to \hat{\mathbf{X}}^v \in \mathcal{R}^{N_v\times D_v}$, $\odot$ represents Hadamard product. To alleviate the inherent heterogeneity of multi-view

data, representations of each view require further enhancement to support downstream tasks and improve relation alignment among view-specific embeddings. Therefore, recent contrastive multi-view clustering methods (Zhou et al., 2026) pull positive pairs (different views of the same instance) together and push negative pairs (cross-view samples from distinct instances) apart in the embedding space. We adopt cross-view contrastive learning (CCL) to measure cross-view discrepancy:

$$l_{ccl}^{vu} = -\frac{1}{N_v} \sum\nolimits_{i=1}^{N_v} \log \frac{\exp(S(z_i^v, z_i^u)\tau)}{\sum_{j=1}^{N_v} \exp(S(z_i^v, z_j^u)/\tau)}, \tag{2}$$

where $S(a, b)$ denotes the similarity between two samples (e.g., cosine similarity), and the temperature hyper-parameter τ is set to 1. The overall loss over all view pairs is formulated as:

$$\mathcal{L}_{ccl} = \sum\nolimits_{v=1}^{V} \sum\nolimits_{\substack{u=1 \\ u \neq v}}^{V} l_{ccl}^{vu}. \tag{3}$$

Although CCL reduces the feature-level heterogeneity across views, the resulting representations $\mathbf{Z}^{(v)}$ are not yet explicitly grounded in a common semantic space. We therefore introduce a semantic mapping to transform each view-specific representation into a unified semantic space:

$$\mathbf{H}^{(v)} = Sem^{(v)}(\mathbf{Z}^{(v)}), \tag{4}$$

where $Sem^{(v)}(\cdot)$ denotes the semantic extractor for the v-th view, implemented as a lightweight projection layer. In this shared space, cross-view samples can be consistently characterized by their relationships to a common set of semantic prototypes, providing a unified semantic reference for subsequent prototype-driven alignment.

2.3 Reliability-Driven Semantic Consistency Representation Learning

In multi-view data, views exhibit diverse information quality and some views may be missing. Direct equal-weight fusion over all views tends to introduce noise and degrade shared semantic information. To address this issue, this paper proposes a reliability-driven semantic consistency representation learning module. A learnable gating mechanism is adopted to adaptively estimate the reliability of each view, and cluster prototypes are further utilized to regularize the semantic structure of fused representations, so as to obtain discriminative consensus semantic representations.

For sample i, the reliability score under view v is first computed by the quality gating network:

$$s_i^{(v)} = f_{\text{gate}}\left(\mathbf{h}_i^{(v)}\right), \tag{5}$$

where $f_{\text{gate}}(\cdot)$ is a multi-layer perceptron that learns view quality information from view-specific semantic representations. The reliability scores are converted into view weights via Softmax:

$$\alpha_i^{(v)} = \left(M_{i,v} \exp\left(s_i^{(v)}/\tau\right)\right) \Bigg/ \left(\sum_{u=1}^{V} M_{i,u} \exp\left(s_i^{(u)}/\tau\right)\right). \tag{6}$$

When view v is missing which forces its fusion weight to zero and prevents missing information from participating in the construction of consensus representations. Adaptive weighted fusion is performed over view-level representations using the learned reliability weights:

$$\boldsymbol{g}_i = \sum\nolimits_{v=1}^{V} \alpha_i^{(v)} \mathbf{h}_i^{(v)}, \tag{7}$$

where $\boldsymbol{g}_i$ is the consensus semantic representation of sample i. This mechanism dynamically adjusts view contributions according to semantic quality, so that high-reliability views dominate in the final consensus representation.

The resulting consensus representations provide a reliable semantic space for capturing the shared cluster structure across views. We therefore extract shared semantic prototypes from the consensus space and use them as a unified semantic reference for subsequent cross-view semantic alignment. To continuously adapt the prototypes to the evolving consensus representation space during training, K-Means clustering is performed over the current consensus representations $\boldsymbol{G}^{(t)}$, and the resulting cluster centers are assigned as the shared semantic prototypes for the next epoch:

$$\boldsymbol{C}^{(t+1)} = \text{KMeans}\left(\boldsymbol{G}^{(t)}\right), \tag{8}$$

To further improve the clustering discriminability of consensus semantic representations, the shared semantic prototypes $\boldsymbol{C} = \{\boldsymbol{c}_k\}_{k=1}^K$ are employed to impose semantic constraints on fused representations. Specifically, the clustering prediction distribution is derived from the similarity between consensus representations and shared prototypes:

$$p_{i,k}^c = \frac{\exp(\boldsymbol{g}_i^\top \boldsymbol{c}_k/\tau_c)}{\sum_{k'=1}^K \exp(\boldsymbol{g}_i^\top \boldsymbol{c}_{k'}/\tau_c)}, \tag{9}$$

where τ_c denotes the temperature coefficient for prototype prediction of consensus representations. Since the shared prototypes have built a unified clustering semantic space in the preceding cross-view semantic alignment stage, this module does not construct extra prototypes and directly constrains fused representations within this pre-established shared semantic space. Soft pseudo-labels $\boldsymbol{q}_i$ with clustering balance constraints are then obtained, and cross-entropy between predicted distribution and soft pseudo-labels is adopted to build the consensus semantic loss:

$$\mathcal{L}_{\text{sem}} = -\frac{1}{N}\sum\nolimits_{i=1}^{N}\sum\nolimits_{k=1}^{K} q_{i,k}\log p_{i,k}. \tag{10}$$

This loss enforces the consensus representations learned via reliability-weighted fusion to preserve explicit clustering semantic structure, avoiding the gating mechanism merely capturing low-level feature similarities while ignoring discriminative information required by the downstream clustering task. In addition, an entropy regularization term is introduced to control the concentration of view attention weights:

$$\mathcal{L}_{\text{ent}} = -\frac{1}{N}\sum\nolimits_{i=1}^{N}\sum\nolimits_{v=1}^{V} \alpha_i^{(v)}\log(\alpha_i^{(v)}+\epsilon), \tag{11}$$

where ϵ is a small smoothing constant to prevent numerical issues caused by $\log 0$. The overall objective of reliability-driven consensus representation learning is formulated as:

$$\mathcal{L}_{\text{fusion}} = \mathcal{L}_{\text{sem}} + \mathcal{L}_{\text{ent}}, \tag{12}$$

2.4 Cross-View Semantic Alignment Based on Prototype Swap Prediction

Definition 1 (Cross-View Semantic Consistency)**.** *If the semantic representations of the same sample from different views yield consistent or approximately consistent cluster assignments when measured by shared semantic prototypes, cross-view semantic consistency is achieved.*

Due to distinct data distributions and feature representations across different views, especially under partial view missing conditions, representations of the same sample across views tend to exhibit significant semantic drift. This further results in inconsistent cross-view feature distributions and heterogeneous clustering assignments. According to Definition 1, the same sample from different views ought to possess consistent cluster membership within the shared semantic space. To this end, cross-view semantic information are propagated and aligned effectively, by exploiting bidirectional swap-based supervision among different views. To further break the limitation of single-view representations and enforce deep cross-view semantic consistency constraints, we construct a prototype-swap prediction constraint for bidirectional cross-view supervision.

Cross-View Semantic Measurement: The shared semantic prototypes $\boldsymbol{C}$ serves as the shared set of cluster centers for all views, establishing a unified benchmark for cross-view semantic interaction and clustering measurement. For an arbitrary view v, the similarity matrix between samples and global cluster prototypes is computed as $S^{(v)} = H^{(v)\top}C$, which quantifies the correlation between samples and clustering bases. The initial cluster assignment probability of each sample in the single-view semantic space is formulated as:

$$p_{i,k}^v = \exp\left(\mathbf{h}_i^{v\top}\mathbf{c}_k/\tau\right)\Big/\sum\nolimits_{k'}^{K}\exp\left(\mathbf{h}_i^{v\top}\mathbf{c}_{k'}/\tau\right), \tag{13}$$

where τ is the temperature coefficient, and $\boldsymbol{p}_i^{(v)} = \{p_{i,k}^{(v)}\}_{k=1}^K$ denotes the initial soft clustering assignment of sample $\boldsymbol{x}_i^{(v)}$ in view v obtained based on shared semantic prototypes.

Consistent Semantic Measurement: Pure sample-wise softmax assignment may cause over-utilization of partial cluster prototypes and further lead to clustering collapse and other trivial solutions. To improve the utilization balance of shared prototypes and obtain more reliable soft

pseudo-labels, we redistribute the similarity matrix from the perspective of entropy-regularized optimal transport:

$$\max_{Q^{(v)}\in\mathcal{C}} \left\langle \boldsymbol{S}^{(v)}, \boldsymbol{Q}^{(v)} \right\rangle + \tau\mathcal{H}\left(Q^{(v)}\right), \quad \mathcal{H}(Q^{(v)}) = -\sum_i \sum_k q_{i,k}^{(v)} \log q_{i,k}^{(v)}, \tag{14}$$

where $\mathcal{C}$ denotes the feasible region satisfying sample normalization and cluster equalization constraints. The entropy regularization term guarantees sufficient utilization of cluster prototypes while avoiding over-sharp pseudo-label assignments, yielding more stable semantic supervision signals.

As the same shared sample should produce consistent clustering assignments across different views, pseudo-labels $q_i^{(m)}$ and $q_i^{(n)}$ are treated as mutual supervision signals to encourage their clustering assignments $p_i^{(m)}$ and $p_i^{(n)}$ to be as consistent as possible:

$$l_{csa}^{m,n} = D_{KL}(Q^{(m)} \| P^{(n)}) + D_{KL}(P^{(m)} \| Q^{(n)}), \tag{15}$$

where $D_{KL}(\cdot)$ denotes the Kullback-Leibler divergence (Yu et al., 2025). We extend Eq. (15) to multi-view scenarios to facilitate broader view collaboration and enhance semantic learning:

$$\mathcal{L}_{csa} = \sum\nolimits_{m=1}^{V} \sum\nolimits_{\substack{n=1 \\ n\neq m}}^{V} l_{csa}^{m,n}. \tag{16}$$

2.5 Clustering Driven by Consensus Semantics Representation

As shown in Algorithm 1, our method learns high-reliability consensus semantic representations $\mathbf{G}$ in an end-to-end manner via the following composite loss:

$$\mathcal{L} = \mathcal{L}_{rec} + \lambda\mathcal{L}_{ccl} + \beta\mathcal{L}_{csa} + \gamma\mathcal{L}_{fusion}. \tag{17}$$

Finally, K-means is performed on $\mathbf{G}$ to obtain the final cluster assignments. The implementation of PRSCL consists of two stages, namely, warm-up training of the encoder-decoder module and fine-tuning for semantic representation alignment and clustering.

Algorithm 1 Learning Procedure of PRSCL

Input: Incomplete multi-view dataset $\{\mathbf{X}^v\}_{v=1}^V$, number of clusters K, warm-up epochs e, maximum fine-tuning iterations E, hyper-parameters $\lambda, \beta, \gamma, \tau_c$.
Output: Cluster prediction result $\hat{\mathbf{Y}}$.

while epochs $\leq e$ **do**
  Pre-training for each view's encoder-decoder network via Eq. (1).
end while
while epochs $\leq E$ **do**
  Each view encoder obtains single-view semantic representation $\mathbf{H}^v$.
  Construct shared cluster prototypes $(\mathbf{C})$ according to Section 2.3.
  Compute reliability weights and consensus semantic representation $(\mathbf{G})$ via Eq. (7).
  Calculate reconstruction loss $\mathcal{L}_{rec}$, cross-view contrastive clustering loss $\mathcal{L}_{ccl}$, cross-view semantic alignment loss $\mathcal{L}_{csa}$ and fusion constraint loss $\mathcal{L}_{fusion}$.
  Back-propagate and update all network parameters according to total loss Eq. (17).
  Periodically update shared cluster prototypes $(\mathbf{C})$ via Eq. (8).
end while
Run K-means on consensus semantic representation $(\mathbf{G})$ to obtain final cluster prediction $\hat{\mathbf{Y}}$.
return The cluster prediction result $\hat{\mathbf{Y}}$.

3 Experiments

3.1 Experimental Settings

Dataset. Following prior work (Dai et al., 2025; Li et al., 2026), we evaluate PRSCL on four benchmark datasets, including ALOI-100, BDGP, Scene-15 and MNIST. Among these benchmark datasets, the maximum number of instances is equal to 10800, the maximum number of features is

Table 1: Performance comparison on four multi-view benchmarks. The best and second-best results are highlighted in red and blue.

Missing rate		$r = 0.0$			$r = 0.1$			$r = 0.3$			$r = 0.5$		
Dataset	**Method**	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)
ALOI-100	DCP	28.67	56.95	21.06	28.33	54.98	19.59	25.47	51.79	16.73	22.75	48.83	13.38
	CPSPAN	61.59	80.15	48.76	66.26	82.42	54.39	66.69	82.94	54.74	65.41	82.23	54.50
	ProImp	82.21	92.95	79.29	78.83	91.13	75.44	69.93	86.18	64.98	57.92	79.76	51.73
	RPCIC	58.04	78.22	45.17	72.22	86.75	62.40	70.86	86.68	61.05	57.15	80.37	48.26
	PMIMC	57.13	77.75	42.82	72.14	87.04	63.12	70.10	86.56	60.89	59.19	80.29	46.85
	FreeCLS	87.56	95.06	85.34	86.80	94.63	84.69	82.89	92.74	79.91	81.07	91.46	78.04
	HSACC	69.23	84.79	61.50	70.74	85.06	62.43	66.49	83.30	58.89	64.04	80.47	54.05
	DCG	49.07	76.20	36.21	45.64	72.92	24.80	37.66	67.06	26.63	35.81	66.79	23.09
	CAMERA	38.56	63.06	26.98	34.46	58.92	23.04	29.11	54.36	17.69	23.67	51.11	12.15
	GLGC	88.83	93.68	84.73	84.73	93.61	84.60	87.54	92.66	82.93	85.99	91.56	80.94
	OURS	**95.63**	**96.56**	**92.44**	**95.11**	**96.10**	**91.66**	**93.07**	**94.53**	**88.43**	87.36	90.65	80.76
BDGP	DCP	48.39	39.43	12.80	49.10	40.82	14.37	46.78	38.10	11.75	47.22	38.68	12.69
	CPSPAN	52.12	37.98	30.54	71.12	49.44	44.22	53.12	33.36	28.80	55.56	32.61	25.71
	ProImp	59.04	50.71	39.33	68.22	59.06	45.69	50.62	39.72	26.20	56.82	44.26	34.62
	RPCIC	86.44	69.74	70.01	72.72	50.58	48.44	61.92	42.07	37.91	58.00	39.11	34.83
	PMIMC	84.84	68.19	67.56	55.36	39.59	34.32	58.12	37.29	34.15	56.76	35.79	31.91
	FreeCLS	97.92	93.53	94.92	96.29	89.36	90.96	91.16	76.45	79.22	87.28	67.55	70.85
	HSACC	49.24	41.75	16.07	52.68	45.68	21.04	52.54	45.88	20.95	48.34	40.27	15.05
	DCG	76.72	71.76	62.21	70.84	67.98	59.25	85.48	76.43	79.06	72.76	61.22	53.68
	CAMERA	96.84	92.41	94.72	96.12	89.76	92.98	94.60	84.65	86.36	92.64	73.65	80.15
	GLGC	33.81	12.33	5.24	33.76	13.88	7.16	29.58	7.20	3.21	30.22	6.88	2.68
	OURS	**97.96**	**94.36**	**95.00**	**97.16**	**91.50**	**93.07**	**94.76**	**85.50**	**87.47**	**93.28**	**81.92**	**84.05**
MNIST	DCP	74.11	65.37	50.36	76.78	68.07	59.07	76.57	67.40	58.36	76.10	67.23	57.69
	CPSPAN	61.15	55.23	45.04	67.58	56.20	47.89	67.30	56.78	47.49	63.37	53.84	43.13
	ProImp	49.45	46.44	32.20	51.67	46.14	32.36	51.34	46.86	32.75	50.80	45.18	31.23
	RPCIC	69.06	62.37	54.61	69.78	64.36	55.40	80.88	67.29	63.36	81.42	68.01	64.32
	PMIMC	77.18	64.26	58.10	70.43	63.29	55.22	80.99	67.58	63.61	74.68	64.29	58.47
	FreeCLS	85.60	74.25	71.87	85.25	73.51	71.18	85.74	74.12	72.05	84.82	72.64	70.42
	HSACC	51.73	50.70	22.03	51.38	52.64	22.31	58.87	56.12	30.28	55.37	53.17	27.01
	DCG	30.26	27.27	16.43	58.23	53.00	40.96	60.54	52.31	41.62	48.56	47.12	34.98
	CAMERA	84.14	71.97	69.16	83.86	71.81	68.71	84.23	71.82	69.37	83.51	70.69	68.07
	GLGC	70.56	62.03	37.33	67.85	60.12	36.11	68.42	60.35	35.63	66.95	58.90	34.32
	OURS	**86.10**	**74.87**	**72.60**	**85.31**	**73.88**	**71.36**	**85.97**	**74.71**	**72.57**	**85.69**	**73.96**	**71.70**
Scene-15	DCP	41.66	44.88	25.81	38.54	42.39	23.33	40.49	43.01	24.14	39.50	42.35	23.51
	CPSPAN	37.95	39.03	22.42	35.67	39.65	21.73	42.94	41.52	25.29	41.07	39.46	24.32
	ProImp	43.61	45.02	26.84	43.55	45.08	27.57	41.01	43.45	25.23	42.52	43.43	26.22
	RPCIC	29.48	28.19	14.70	36.25	34.56	20.53	36.54	36.36	19.83	29.77	34.49	16.13
	PMIMC	26.64	25.69	12.21	34.76	34.00	19.57	33.62	34.19	17.37	27.09	31.27	13.75
	FreeCLS	37.70	41.76	22.79	33.98	34.71	18.88	35.03	34.90	19.31	34.85	34.31	19.73
	HSACC	32.16	31.47	15.74	32.52	31.89	16.32	31.46	30.03	15.03	29.40	30.34	14.14
	DCG	39.67	43.34	23.61	40.27	43.42	24.28	40.60	43.30	24.07	41.03	42.14	25.05
	CAMERA	46.51	46.61	28.86	45.77	46.57	28.46	44.40	45.01	27.14	44.13	43.92	26.29
	GLGC	36.50	38.58	21.32	35.59	37.16	20.18	34.90	35.75	19.15	34.22	35.34	19.01
	OURS	**47.87**	**47.26**	**30.59**	44.30	45.11	27.73	44.28	43.10	26.71	41.18	40.01	24.38

equal to 4096, the number of views varies from 2 to 4, and the number of clusters varies from 5 to 100.

Comparison methods. We compare PRSCL with two categories of SOTA baselines: ① Prototype-driven IMVC, including CPSPAN (Jin et al., 2023), ProImp (Li et al., 2023), RPCIC(Yuan et al., 2024), PMIMC(Yuan et al., 2025), CAMERA(Li et al., 2026). ② Other MVC methods, including DCP(Lin et al., 2023), FreeCLS(Dai et al., 2025), HSACC(Ding et al., 2025), DCG(Zhang et al., 2025) and GLGC(He et al., 2026).

3.2 Performance Comparison

As shown in Table 1, we compare PRSCL with existing methods on four datasets with different scales and clustering complexities, using ACC, NMI, and ARI under four missing rates $r = 0.0, 0.1, 0.3$, and 0.5. The implementation code is provided in the supplementary materials.

From the perspective of effectiveness, PRSCL demonstrates clear performance advantages over competing methods across most datasets. On ALOI-100, PRSCL achieves 95.63%, 96.56%, and 92.44% , respectively, under the complete setting $r = 0.0$, substantially outperforming competing methods. When $r = 0.5$, it maintains 87.36% ACC, showing its ability to preserve reliable clustering structures under substantial missingness. Similarly, on BDGP, PRSCL consistently achieves the best ACC, NMI, and ARI across all missing rates. These results demonstrate its ability to exploit complementary information across views under different data scales and distributions. Additionally, it is worth noting that all methods attain relatively low absolute performance on Scene-15, and PRSCL fails to consistently achieve the best results as the missing rate increases. This stems from its complex semantics and ambiguous boundaries, which hinder learning discriminative clusters from incomplete multi-view data. While PRSCL's reliability-aware gating suppresses uninformative views,

it may discard complementary information when few views remain. Meanwhile, inherent semantic ambiguity yields weakly discriminative shared prototypes and unreliable prototype assignments, limiting cross-view mutual supervision. Finally, for a comprehensive overall evaluation, we provide the Critical Difference (CD) analysis in the appendix, which aggregates average ranks across datasets and missing rates for supplementary statistical comparison.

From the perspective of robustness to missing views, PRSCL exhibits relatively stable performance degradation as the missing rate increases. On ALOI-100, its ACC, NMI, and ARI decrease from 95.63%, 96.56%, and 92.44% at $r = 0.0$ to 87.36%, 90.65%, and 80.76% at $r = 0.5$, respectively. In contrast, several existing methods show larger performance drops. For example, ProImp's ACC decreases from 82.21% to 57.92%. This indicates that methods relying heavily on cross-view consistency or explicit imputation are more vulnerable to missing information, whereas PRSCL shows stronger adaptability to high missing rates. Overall, PRSCL achieves strong clustering effectiveness under low missing rates while maintaining relatively stable performance as missingness increases. These results indicate that PRSCL can exploit consensus semantic information across views while reducing the impact of incomplete observations on the shared clustering structure, resulting in strong generalization and robustness across datasets with different scales and clustering complexities.

3.3 Method Analysis

Table 2: Ablation study on the BDGP dataset under different missing rates.

Loss Components			$r = 0.0$			$r = 0.1$			$r = 0.3$			$r = 0.5$		
L_{ccl}	L_{csa}	L_{fusion}	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)
✓			33.52	9.03	4.89	39.44	17.77	10.34	34.08	13.68	7.35	35.21	9.38	6.52
	✓		94.04	90.87	91.21	93.64	90.32	91.23	92.12	84.74	86.59	91.61	**83.15**	**85.36**
		✓	92.52	84.17	83.03	94.24	84.60	86.32	64.28	52.54	46.06	62.04	43.46	36.83
	✓	✓	96.20	91.11	90.97	95.92	90.22	92.87	94.12	84.63	87.27	85.84	68.96	68.71
✓		✓	96.16	91.06	90.88	95.04	90.54	92.16	94.04	84.36	87.08	86.41	70.43	69.93
✓	✓		96.84	93.16	93.71	96.28	90.96	92.37	91.16	83.35	85.68	89.72	79.37	82.29
✓	✓	✓	**97.96**	**94.36**	**95.00**	**97.16**	**91.50**	**93.07**	**94.76**	**85.50**	**87.47**	**93.28**	81.92	84.05

(1) Ablation Studies. To investigate the importance of each component, we conducted ablation studies on PRSCL using BDGP under different missing rates. Table 2 verifies that $\mathcal{L}_{ccl}$, $\mathcal{L}_{csa}$, and $\mathcal{L}_{fusion}$ each make an independent contribution to the proposed framework while further improving performance through their synergistic interaction. Specifically, $\mathcal{L}_{csa}$ plays an important role in establishing cross-view semantic consistency, $\mathcal{L}_{ccl}$ enhances clustering discriminability, and $\mathcal{L}_{fusion}$ facilitates the effective interaction of complementary information across views. Their joint optimization enables the model to simultaneously capture semantic consistency, clustering discriminability, and cross-view information fusion, resulting in more accurate and robust clustering performance.

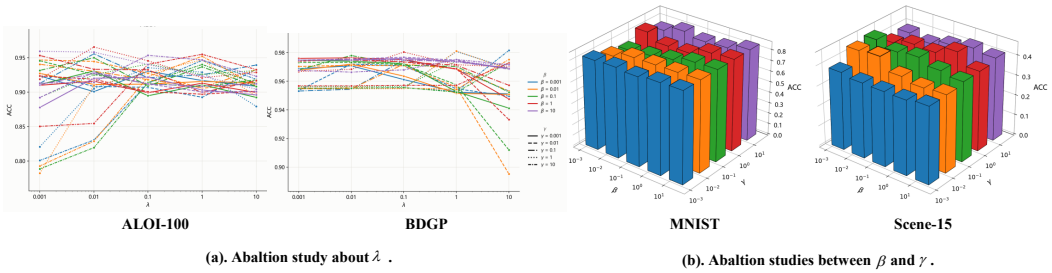

Figure 3: Parameters analysis of PRSCL with the $r = 0.3$. (a) Parameters sensitivity analysis of λ with fixed β and γ; (b) Parameter sensitivity analysis of β and γ conducted under the optimal λ.

(2) Parametes Analysis. To evaluate the robustness of PRSCL to different parameter configurations, we investigate the effects of λ, β, and γ. We conducted a parameter analysis by setting both parameters in the range from 10^{-3} to 10. It can be observed that PRSCL generally maintains stable clustering performance over a relatively wide parameter range. As shown in Figure 3, excessively high or low parameter values are unfavorable for clustering. Based on our parameter experiments,

we recommend setting the parameters range between 0.01 and 1. The detailed parameter settings for our datasets are provided in the Appendix.

(3) Visualization Analysis. As shown in Figure 4, we utilize t-SNE to conduct visual analysis on the consensus semantic representations at different training epochs on the MNIST dataset with $r = 0.1$. At the early training stage, samples are mixed together with low inter-class discrimination. After multiple iterative optimizations, samples are effectively separated according to their categories, and distinct boundaries emerge between different classes.

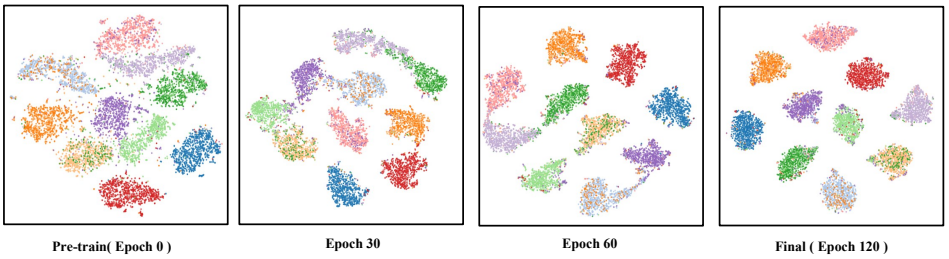

Figure 4: Visualization for MNIST with r = 0.1 (K = 10).

(4) Discussion of Reliability-Aware Semantic Consensus Learning. we construct the cosine similarity matrices of the final semantic representations from individual views and the consensus semantic representation on the ALOI-100 dataset. As shown in Figure 5, the semantic representations of individual views contain a considerable amount of noisy or misleading information, resulting in weak and indistinct diagonal structures. In contrast, the consensus representation exhibits a much clearer block-diagonal structure after reliability-aware fusion. This indicates that the proposed mechanism can effectively identify and suppress unreliable information from individual views while retaining complementary and discriminative semantic information, thereby producing a more coherent and cluster-discriminative consensus representation.

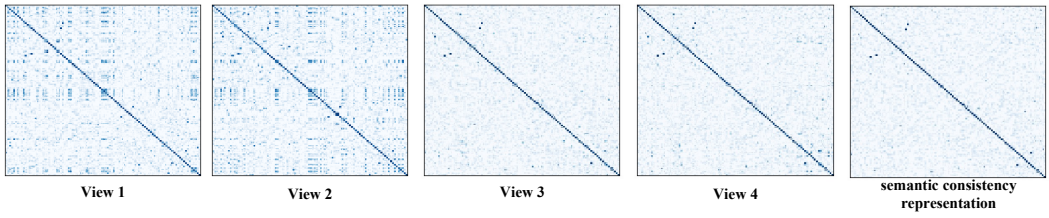

Figure 5: Similarity matrices of different views and semantic consistency representation on ALOI-100 with r = 0.1.

4 Conclusion

In this paper, we proposed a novel Prototype-driven Reliability-aware Semantic Consensus Learning (PRSCL) framework for incomplete multi-view clustering. PRSCL directly learns robust consensus semantics from available observations without explicitly imputing missing views. By constructing shared consensus prototypes, PRSCL establishes a unified semantic reference for cross-view semantic representation alignment, while a reliability-aware gating mechanism adaptively weights available views to obtain reliable consensus representations. Entropy-regularized optimal transport and bidirectional mutual supervision further exploit prototype-driven assignment structures to enhance clustering-level consistency across views. Extensive experiments on multiple benchmark datasets demonstrate the effectiveness and robustness of PRSCL under various missing-view conditions. In future work, we will further explore reliable semantic consensus representations, investigate more robust reliability modeling and prototype learning under extremely high missing rates, and extend PRSCL to large-scale, dynamically evolving multi-view data.

AI USE STATEMENT

In this work, generative AI tools are only used for grammatical correction, sentence polishing, and academic expression standardization. We did not use generative AI tools for algorithm design, model implementation, experimental analysis, and result interpretation. Other mandatory disclosure items are not applicable to this work.

Additionally, generative AI tools are adopted to improve the readability and fluency of the manuscript. All AI-assisted contents have been fully manually reviewed and verified. Specifically, all AI-polished sentences were checked sentence by sentence to avoid plagiarism, logical errors and inconsistent descriptions. The core technical contributions, experimental designs and conclusions are independently completed and validated by the authors.

REFERENCES

Guoqing Chao, Shiliang Sun, and Jinbo Bi. A survey on multiview clustering. *IEEE Transactions on Artificial Intelligence*, 2(2):146–168, 2021.

Jie Chen, Hua Mao, Wai Lok Woo, and Xi Peng. Deep multiview clustering by contrasting cluster assignments. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 16752–16761, 8 2023.

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 *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5071–5081, 2025.

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

Hongqing He, Jie Xu, Wenyuan Yang, Yonghua Zhu, Guoqiu Wen, and Xiaofeng Zhu. Global-graph guided and local-graph weighted contrastive learning for unified clustering on incomplete and noise multi-view data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24813–24822, June 2026.

Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11600–11609, 2023.

Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. In *Proceedings of the 32th International Joint Conference on Artificial Intelligence*, 2023.

Haobin Li, Yijie Lin, Peng Hu, Mouxing Yang, and Xi Peng. Community-aware multi-view representation learning with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(3):3517–3529, 2026.

Yueyao Li and Bin Wu. Reliability-guided adversarial semantic completion for incomplete multi-view clustering. *Applied Soft Computing*, 204:116324, 2027.

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, 2023.

Yue Liu, Jun Xia, Benyu Wu, Sihang Zhou, Xihong Yang, Ke Liang, Chenchen Fan, Yan Zhuang, Guoxian Yu, Stan Z. Li, Xinwang Liu, and Kunlun He. A survey of deep graph clustering: Taxonomy, challenge, application, and open resource. *IEEE Transactions on Knowledge and Data Engineering*, 38(6):3365–3384, 2026.

Huibing Wang, Yawei Chen, Mingze Yao, Wenzhe Liu, Jinjia Peng, and Xianping Fu. Tensor completion framework by graph refinement for incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 2025.

Jiatai Wang, Zhiwei Xu, Xuewen Yang, Dongjin Guo, and Limin Liu. Self-supervised image clustering from multiple incomplete views via constrastive complementary generation. *IET Computer Vision*, 17(2):189–202, 2023.

Qianqian Wang, Xurui Liao, Wei Feng, and Quanxue Gao. Debiased and denoised representation learning for incomplete multi-view clustering. In *International Conference on Learning Representations*, volume 2026, pp. 73338–73350, 2026a.

Shuaijun Wang, Hui Huang, Nan Zhang, and Shiliang Sun. Federated incomplete multi-view clustering with cross-view relationship imputation. *IEEE Transactions on Knowledge and Data Engineering*, 38(7):4344–4356, 2026b.

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, 2023.

Wai Keung Wong, Chengliang Liu, Shijie Deng, Lunke Fei, Lusi Li, Yuwu Lu, and Jie Wen. Neighbor group structure preserving based consensus graph learning for incomplete multi-view clustering. *Information Fusion*, 100:101917, 2023.

Benyu Wu, Wei Du, Jun Wang, and Guoxian Yu. Imputation-free incomplete multi-view clustering via knowledge distillation. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 6570–6578, 8 2025.

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.

Xuejiao Yu, Yi Jiang, Guoqing Chao, and Dianhui Chu. Deep contrastive multi-view subspace clustering with representation and cluster interactive learning. *IEEE Transactions on Knowledge and Data Engineering*, 37(1):188–199, 2025.

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*, MM '24, 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.

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.

Haochen Zhou, Guofeng Ding, Mouxing Yang, Peng Hu, Yijie Lin, and Xi Peng. Uncover underlying correspondence for robust multi-view clustering. In *The Fourteenth International Conference on Learning Representations*, 2026.

Zhengzhong Zhu, Pei Zhou, Lanxi Bai, Jia Nie, Li Cheng, Shiquan Min, and Jiangping Zhu. Prototype and sample level semantic alignment for incomplete multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5818–5827, 2026.

APPENDIX

A Related Work

A.1 Prototype-driven Incomplete Multi-view Clustering

Traditional IMVC methods build anchor graphs for scalability, yet anchors sampled or clustered in the original feature space are unreliable under heavy view incompleteness due to heterogeneous distributions and limited observed information. With the development of deep representation learning, recent studies have increasingly explored semantic representations and prototype-based strategies to obtain more representative and structurally meaningful information from incomplete multi-view data. These methods establish a prototype-driven paradigm by introducing representative cluster-level semantic structures to alleviate the adverse effects of missing views.

For example, ProImp (Li et al., 2023) employs a dual-stream encoder to jointly learn view representations and prototypes, thereby improving the quality of clustering-oriented semantic information. CPSPAN (Jin et al., 2023) introduces cross-view local sample alignment and shifted prototype alignment to alleviate the misalignment between incomplete views. PMIMC (Yuan et al., 2025) learns view-specific semantic prototypes and identifies cross-view neighbors through structural matching, together with a robust prototype contrastive loss to reduce prototype misalignment. RPCIC (Yuan et al., 2024) further investigates correlations among cross-view prototypes and incorporates prototype fusion into missing-view imputation, attempting to exploit prototype-level relationships for more accurate view recovery. More recently, GAN-IMVC (Li & Wu, 2027) adopts a reliability-guided adversarial semantic completion framework, leveraging adversarial missing-view completion and prototype-based soft assignments together with graph enhancement to unify local and global prototype constraints for robust clustering under various missing rates.

Despite the progress achieved by these prototype-driven methods, most existing approaches still suffer from limitations. Their prototypes are predominantly learned within individual views or are mainly utilized to facilitate view completion, local structural matching, or prototype-level contrastive learning. Such strategies may remain vulnerable when the available observations are severely incomplete or when the semantic structures of different views exhibit substantial discrepancies. More importantly, existing methods pay limited attention to the consensus semantic space that captures the common cluster semantics shared across available views. Treating view-specific prototypes as independent semantic references makes it difficult to establish a unified semantic correspondence across views and may introduce view-dependent biases into the final clustering representation.

In contrast, our method places the consensus semantic space at the center of prototype-driven learning. Rather than explicitly reconstructing missing views, PRSCL constructs shared consensus semantic prototypes to provide a unified semantic reference for aligning the semantic representations of different views. The discrepancies between view-specific semantic representations and the consensus semantics are further exploited to characterize view-specific information, while a reliability-aware gating mechanism adaptively determines the contribution of each available view during consensus construction. In this way, prototypes are not merely used for missing-view recovery or cluster assignment, but serve as a semantic bridge connecting cross-view alignment and reliable consensus fusion. This design enables PRSCL to jointly preserve cross-view semantic consistency and view-specific complementary information, while reducing the influence of unreliable or missing observations.

B Experimental Details

B.1 Dataset and Comparison Methods

As shown in Table 3, we evaluate PRSCL on four benchmark datasets, including ALOI-100, BDGP, Scene-15 and MNIST. Among these benchmark datasets, the maximum number of instances is equal to 10800, the maximum number of features is equal to 1750, the number of views varies from 2 to 4, and the number of clusters varies from 5 to 100. The details are summarized as follows.

Table 3: Details of Benchmark Datasets

Datasets	#Clusters	#Instances	View Dimensions
ALOI-100	100	10800	77, 13, 64, 125
BDGP	5	2500	1750, 79
MNIST	10	10000	30, 9, 30
Scene-15	15	4485	20, 59, 40

- **ALOI-100** consists of a large number of images of small objects captured under varying illumination conditions and rotation angles. It contains 10800 samples from 100 categories with 4 views, whose feature dimensions are 77, 13, 64 and 125 respectively.
- **BDGP** is a gene-expression dataset derived from the Drosophila genome project. It comprises 2 views with feature dimensions of 1750 and 79 respectively, containing ground-truth labels for clustering evaluation.
- **MNIST** contains 10,000 instances belonging to 10 classes, which are composed of three views and represent the handwritten digits from 0 to 9.
- **Scene-15** contains 4,485 images from 15 categories with two image features as views.

We compare PRSCL with two categories of SOTA baselines: ① Prototype-driven IMVC, including CPSPAN, ProImp, RPCIC, PMIMC, CAMERA. ② Other MVC methods, including DCP, FreeCLS, HSACC, DCG and GLGC. In the experimental validation, all parameter settings follow the optimal values reported in the original paper. For parameters not specified in the literature, the default values are adopted.

- **DCP** provides an information theoretical framework under which the consistency learning and data recovery are treated as a whole.
- **CPSPAN** employs cross-view partial sample alignment and shifted prototype alignment to address data mismatchment issues.
- **ProImp** adopts dual attention layers and contrastive learning loss to optimize view-aware prototypes, thereby enabling robust data reconstruction against incomplete view information.
- **RPCIC** mitigates prototype noisy correspondence for incomplete multi-view clustering by cross-view contrastive learning, robust prototype contrastive loss and prototype-fusion-driven missing-data completion.
- **PMIMC** addresses prototype misalignment and imputation instability in IMVC through relational consistency learning and robust prototype contrastive learning.
- **FreeCLS** learns consensus prototypes to build a shared semantic space and leverages modularity-driven heuristic graph clustering to enhance cluster semantics, avoiding explicit data imputation and alignment and obtaining more robust cluster assignments.
- **HSACC** ealizes implicit missing-view recovery via cooperative optimization within a two-tier semantic space consisting of low-level mutual-information consistency and high-level adaptive weighted fusion, and outperforms existing incomplete multi-view clustering methods on five benchmarks.
- **DCG** unifies incomplete multi-view clustering and view generation within a diffusion-based contrastive framework, where view recovery, cross-view alignment, and discriminative clustering are jointly optimized.
- **CAMERA** integrates instance restoration, view alignment, and diversity preservation into a unified framework, jointly optimizing them through community commonality and community versatility.
- **GLGC** constructs global and local affinity graphs to respectively generate supplementary cross-view sample pairs and assign adaptive pairwise weights, and integrates them into a unified contrastive learning framework to achieve superior performance on incomplete and noisy multi-view clustering.

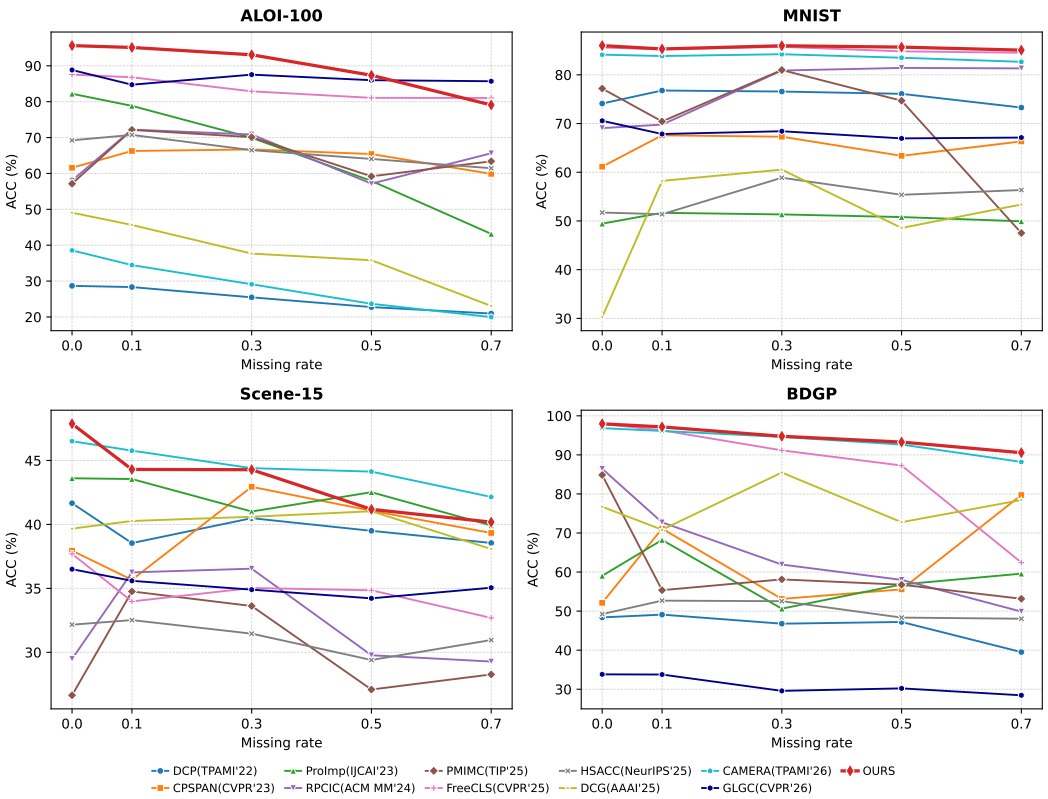

Figure 6: The clustering performance under varying missing rates on four datasets.

B.2 Implementation Details

Models are trained on an NVIDIA 4060 GPU using Adam optimizer. Our model consists of three modules: inter-view representation reconstruct and enhancement, cross-view semantic alignment based on prototype swap prediction, reliability-driven semantic consistency representation learning, as well as four components: encoder, decoder, semantic extraction and the quality gating network. The implementation details are shown in Figure 4, the recon_fea_dim in our experience is 128.

Table 4: Detailed network architecture of the proposed model.

Components	Layer	Dimension
Encoder	4-layer MLPs	Input_dim $\rightarrow$ 500 $\rightarrow$ 500 $\rightarrow$ 2000 $\rightarrow$ Recon_fea_dim
Decoder	4-layer MLPs	Recon_fea_dim $\rightarrow$ 2000 $\rightarrow$ 500 $\rightarrow$ 500 $\rightarrow$ Input_dim
Semantic extraction	MLPs	Recon_fea_dim $\rightarrow$ Recon_fea_dim
Quality gating network	3-layer MLPs	Recon_fea_dim $\rightarrow$ Recon_fea_dim $\rightarrow$ Recon_fea_dim/2 $\rightarrow$ 1

B.3 Evaluation Protocols

We adopt three widely used clustering metrics: accuracy (ACC), normalized mutual information (NMI), and adjusted Rand index (ARI). Higher values indicate better performance for all metrics.

C Additional Experimental Results

C.1 Performance Comparison

To evaluate the robustness of PRSCL to incomplete observations, we compare its clustering performance with representative state-of-the-art methods under different missing rates, ranging from 0 to 0.7. The results on ALOI-100, MNIST, Scene-15, and BDGP are reported in Figure 6. The performance of most methods generally deteriorates as the missing rate increases, demonstrating the inherent difficulty of clustering with increasingly incomplete observations. In contrast, PRSCL maintains relatively stable and competitive performance across different datasets. On ALOI-100, PRSCL achieves an ACC of approximately 96% under the complete setting and remains around 79% even when the missing rate reaches 0.7. Similarly, on BDGP, its ACC decreases gradually from approximately 98% to 91%, indicating strong robustness to severe view incompleteness. On MNIST, PRSCL exhibits particularly stable performance, with ACC remaining above 85% across all missing rates. The Scene-15 dataset presents a more challenging case, where PRSCL decreases from approximately 48% to 40% as the missing rate increases. Although several competing methods achieve slightly higher performance under severe missingness, PRSCL still maintains a relatively competitive performance and exhibits a gradual degradation trend. This suggests that the effectiveness of consensus semantic learning is influenced by the intrinsic difficulty and heterogeneity of the dataset. Overall, these results demonstrate that PRSCL is relatively robust to increasing missing rates. Rather than relying on explicit reconstruction of missing views, PRSCL directly exploits the available observations and adaptively integrates reliable semantic information, which helps alleviate the adverse impact of missing views on consensus representation learning.

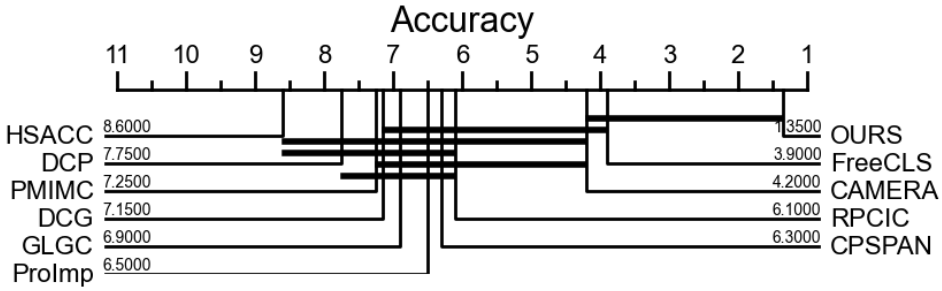

Figure 7: Comparison of PRSCL against contrast methods with the Wilcoxon rank test and Holm's alpha correction on all datasets under all missing rates.

As shown in Figure 7, the average rank of our proposed method is 1.35 in terms of accuracy, which is significantly lower than the ranks for FreeCLS, CAMERA and other contrast baselines. Our method excavates shared semantic prototypes across incomplete views and exploits complementary information from missing-aware representation learning. Compared with FreeCLS, our method outperforms it in terms of mean accuracy, with a clear performance gain across various missing rates. This finding serves to demonstrate the effectiveness of our missing-robust prototype consensus learning paradigm. Furthermore, our method achieves consistent improvements over other competitive approaches including CAMERA, RPCIC and CPSPAN. In comparison with our method, these competitors fail to sufficiently mitigate the adverse impact introduced by random view corruption. By contrast, our method explicitly models the uncertainty brought by missing samples and aligns cross-view semantics via prototype-driven contrastive constraints. Subsequently, our method separates consistent and complementary view information during joint representation learning, thereby enabling the model to maintain stable clustering performance even under high missing ratios and achieving state-of-the-art results.

C.2 Parametes Analysis

We conducted parameter sensitivity experiments on four datasets to verify the extent to which each parameter affects the experimental results. Three parameters are examined in this study. Specif-

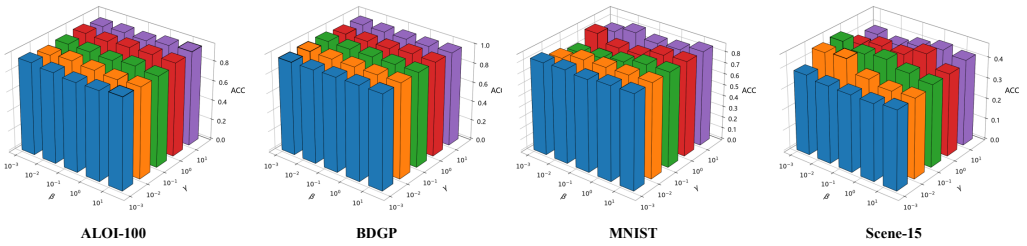

Figure 8: The clustering performance under varying missing rates on four datasets.

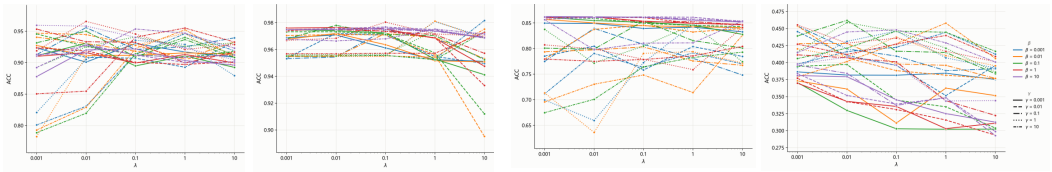

Figure 9: The clustering performance under varying missing rates on four datasets.

ically, by fixing λ, we compare the effects of β and γ on accuracy (ACC), as shown in Figure 8. Furthermore, we group β and γ into fixed combinations to investigate how the value of λ influences the experimental performance in the Figure 9.

PRSCL exhibits relatively stable performance across a wide range of parameter configurations on ALOI-100, BDGP, and MNIST. In particular, the ACC curves remain within a relatively narrow range for most combinations of β and γ, indicating that the model is not highly sensitive to their exact values. The joint analysis of β and γ also shows that satisfactory performance can be obtained across a broad parameter region rather than being restricted to a specific combination. This demonstrates the robustness of PRSCL to the relative weighting of different semantic learning objectives. The effect of λ is relatively more evident on the Scene-15 dataset, where larger fluctuations can be observed under different β and γ configurations. This may be attributed to the more complex and heterogeneous characteristics of Scene-15, which make the balance among different learning objectives more important. Nevertheless, the performance on the other datasets remains comparatively stable as λ varies, suggesting that PRSCL can maintain reliable clustering performance without requiring precise parameter tuning. Overall, the sensitivity results demonstrate that PRSCL is robust to a broad range of hyperparameter setting.