MGRL-IMVC: MULTI-GRANULARITY RELATIONAL LEARNING FOR INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete multi-view clustering (IMVC) aims to uncover latent cluster structures from partially observed views. However, missing views weaken instance-level cross-view supervision, noisy observations reduce the reliability of available pairs, and cluster-level semantic discrepancies across views hinder consistent representation learning. This paper proposes MGRL-IMVC, a multi-granularity relational learning framework that avoids imputing missing features and jointly models sample relations, neighborhood structures, and prototype correspondences. Specifically, unified-space relational learning aggregates all observable representations and mines latent positive and negative relations, enabling samples without direct cross-view pairs to participate in contrastive learning. Cross-view structural learning constructs affinity-based neighborhood weights to provide structural evidence beyond direct similarity for same-instance pairs and to modulate contrastive constraints. Cross-view soft prototype alignment estimates soft correspondences between view-specific prototypes via the co-occurrence of soft assignments on commonly observed samples, and enforces bidirectional prototype matching for cluster-level semantic alignment. These objectives are jointly optimized with view reconstruction, allowing limited observations to complement each other across multiple relational granularities. Experiments on five benchmark datasets under various missing rates demonstrate that MGRL-IMVC achieves superior clustering accuracy and adaptability compared with seven representative deep IMVC methods, and ablation studies verify the complementary contributions of the three granularities.

1 INTRODUCTION

Incomplete multi-view clustering (IMVC) aims to discover latent representations from data in which only a subset of views is available for each sample (Wen et al., 2023b). However, the absence of views weakens the correspondence information through which different views complement one another, making it difficult to recover a consistent clustering structure. Existing approaches mainly follow two directions. One direction reconstructs missing features, relations, or predictions from the available views (Xu et al., 2019; Lin et al., 2023); the other learns shared clustering information directly from observed data to avoid the errors introduced by uncertain recovery (Xu et al., 2022a; 2024a). Despite their different strategies, both directions ultimately depend on whether the limited observations can provide sufficient and reliable relational evidence. From this perspective, the central problem of IMVC is not merely how to recover missing content, but how to organize the remaining observations into relations that faithfully reveal the underlying cluster structure.

The first challenge is the scarcity of sample-level supervision. Contrastive IMVC methods generally regard different observed views of the same instance as a positive pair. When one view is missing, this pair can no longer be constructed, even though the remaining representation may still contain useful semantic information. (Dai et al., 2025)Moreover, instance identity describes only a small subset of the relations needed for clustering. Different instances from the same latent category may be incorrectly separated as negatives, thereby distorting the representation space (Chuang et al., 2020). At present, graph-based methods alleviate this limitation by discovering additional relationships among observed samples (Wang et al., 2023; Wen et al., 2023a; He et al., 2026). These studies

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.

indicate that effective IMVC should exploit relations beyond directly paired instances so that all available observations can contribute to representation learning.

Expanding the range of sample relations, however, introduces a second challenge: the discovered relations are not necessarily reliable. View-specific characteristics and noisy observations can cause representations of the same instance to deviate across views (Yang et al., 2021; Sun et al., 2024). Conversely, a high pairwise similarity does not necessarily imply semantic consistency. Direct similarity alone therefore provides insufficient evidence for deciding how strongly a pair should be constrained. Local neighborhoods offer a more informative structural context because semantically consistent samples tend to exhibit similar relational patterns with their surroundings (Wen et al., 2023a; Lu et al., 2024). Recent global-local learning further demonstrates that global relation discovery and local structural weighting can jointly alleviate scarce and unreliable cross-view pairs (He et al., 2026). This observation suggests that relation discovery and relation assessment should form a coordinated learning process.

Even reliable sample and neighborhood relations do not fully determine cross-view clustering semantics. Missing observations change the sample distributions from which individual views form their clusters, while view heterogeneity can cause the resulting cluster structures to shift (Jin et al., 2023). Consequently, a cluster identified in one view does not have a naturally known counterpart in another view. Partial sample alignment, semantic invariance, and prototype matching have been introduced to reduce such discrepancies (Zeng et al., 2024; Wang et al., 2025). Nevertheless, prototype correspondences remain uncertain when only a limited number of samples are observed in both views. More importantly, cluster-level alignment is often treated separately from sample relation learning and neighborhood structure modeling. The information learned at these different granularities is therefore not fully connected.

The above analysis raises three coupled issues at the sample, neighborhood, and prototype levels: which observed representations are related, how reliable their cross-view relations are, and which view-specific prototypes express corresponding cluster semantics. We therefore propose MGRL-IMVC, an imputation-free framework with three granularities. Unified-space relational learning mines potential positive/negative pairs from all observed representations; cross-view structural learning uses affinity-based neighborhoods to modulate same-instance positive pairs; and cross-view soft prototype alignment estimates prototype correspondence from soft-assignment co-occurrence and enforces bidirectional matching. Jointly optimized with view reconstruction, these objectives connect global sample relations, local structural evidence, and cross-view cluster semantics.

Our main contributions are summarized as follows:

1. We formulate IMVC from a multi-granularity relational perspective and develop a unified framework that connects global sample relations, cross-view neighborhood structure, and semantic correspondences between cluster prototypes.
2. We develop unified-space relational learning and cross-view structural learning. The former mines potential positive and negative pairs from all observed representations to expand supervision beyond instance identity, while the latter uses cross-view neighborhood evidence to regulate same-instance positive pairs and reduce the influence of structurally unreliable relations.
3. We introduce a soft prototype alignment mechanism guided by commonly observed samples. It estimates cross-view prototype correspondence weights from soft-assignment co-occurrence and applies bidirectional prototype matching, extending relational learning from the sample level to cluster-level semantics without imputing missing features.

2 Related Work

2.1 Deep Incomplete Multi-View Clustering

Early IMVC methods primarily focus on learning shared latent representations or recovering similarity structure (Li et al., 2014; Hu & Chen, 2018; Wang et al., 2019), while late-fusion methods align view-specific clustering assignments (Liu et al., 2019; 2021). However, their reliance on predefined features and shallow mappings limits semantic representation learning.

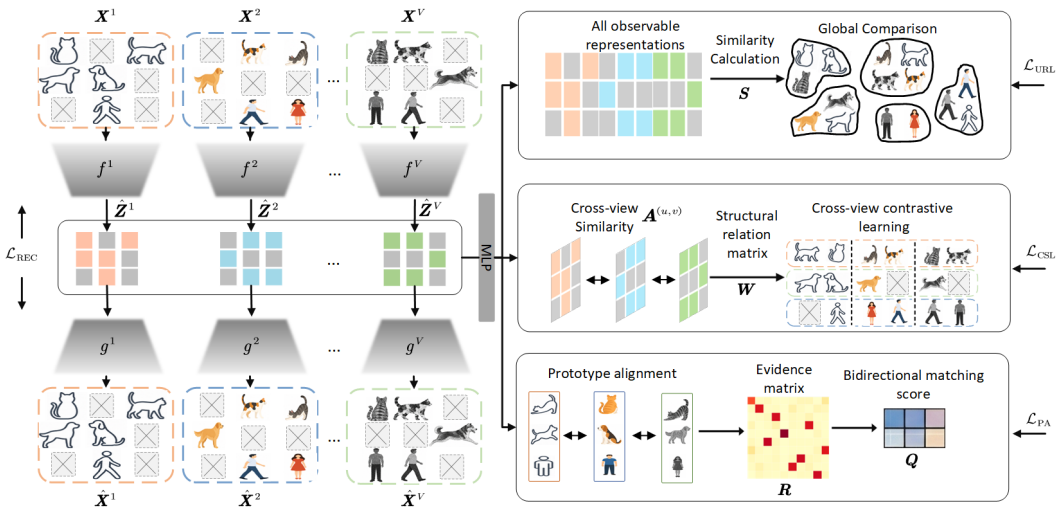


Figure 1: Overview of the proposed MGRL-IMVC framework.

Deep methods jointly optimize representation learning and clustering. Specifically, AIMC recovers missing views through adversarial learning (Xu et al., 2019); CDIMC-net combines graph embedding with self-paced learning (Wen et al., 2020); and COMPLETER couples consistency learning with missing-view prediction (Lin et al., 2021). To further address recovery errors and semantic discrepancies, DSIMVC selects reliable imputed samples (Tang & Liu, 2022a), CPSPAN aligns partial samples and prototypes (Jin et al., 2023), and DRUMVC constructs additional correspondences using robust contrastive objectives (Zhao et al., 2025). Unlike these methods, our framework retains view-specific representations and aligns prototypes through soft-assignment co-occurrence on commonly observed samples, without feature imputation.

2.2 Multi-View Contrastive Learning

SimCLR learns representations through instance discrimination (Chen et al., 2020). Contrastive Clustering further combines instance-level and cluster-level objectives (Li et al., 2021b), while PCL and SwAV incorporate prototypes or cluster assignments (Li et al., 2021a; Caron et al., 2020). In multi-view clustering, MFLVC learns consistency at multiple feature and semantic levels (Xu et al., 2022b), and MCGC introduces graph structure through consensus graph learning (Pan & Kang, 2021). Graph contrastive partial multi-view clustering further exploits graph relations under incomplete observations (Wang et al., 2023).

In addition, reliable supervision is also essential: low-quality and noisy views can undermine consistency learning (Tang & Liu, 2022b; Xu et al., 2024b). To mitigate this issue, EASEMVC addresses it through view-pair selection and graph-based sample weighting (Xiao et al., 2025). Particularly relevant, He et al. (2026) expand sample pairs using a global affinity graph and weight cross-view pairs with local neighborhood information. However, their modeling primarily concerns samples and neighborhoods. In contrast, MGRL-IMVC integrates these relational granularities with soft prototype alignment to jointly model cross-view cluster semantics.

3 Method

3.1 Notation

Given an incomplete multi-view dataset with V views, let $\boldsymbol{X}^v \in \mathbb{R}^{N \times D_v}$ denote the feature matrix of view v, where N is the number of samples and D_v is the feature dimension of that view. Its i-th row is denoted by $\boldsymbol{x}_i^v \in \mathbb{R}^{D_v}$.

View availability is represented by an observation matrix $\boldsymbol{M} \in \{0,1\}^{N\times V}$. Its entry $m_i^v = 1$ indicates that sample i is observed in view v, whereas $m_i^v = 0$ indicates that the corresponding view is missing. All subsequent operations involving $\boldsymbol{x}_i^v$ are restricted to observations satisfying $m_i^v = 1$.

We use i and j to index samples, v and u to index views, and K to denote the number of clusters.For a view pair (v, u), the indices of samples that are observable in both views are defined as

$$\mathcal{I}^{(v,u)} = \{i \,|\, 1 \le i \le N,\ m_i^v m_i^u = 1\} \tag{1}$$

Furthermore, the valid view pairs that contain at least one commonly observed sample are defined as

$$\mathcal{E} = \left\{(v,u) \,\middle|\, 1 \le v < u \le V,\ \left|\mathcal{I}^{(v,u)}\right| > 0\right\} \tag{2}$$

3.2 Unified-Space Relational Learning

Missing views prevent many samples from forming direct cross-view pairs, leaving only limited instance-level supervision. To allow all available observations to participate in relation learning, we first map them into a unified representation space.

For each observed input $\boldsymbol{x}_i^v$ satisfying $m_i^v = 1$, a view-specific encoder f_v extracts its latent features, and a view-specific projection head ϕ_v maps them into a common d_h-dimensional contrastive space. The projected representation is then normalized as

$$\boldsymbol{h}_i^v = \frac{\phi_v(f_v(\boldsymbol{x}_i^v))}{\|\phi_v(f_v(\boldsymbol{x}_i^v))\|_2} \tag{3}$$

All observed representations are subsequently concatenated by rows to form $\boldsymbol{H} \in \mathbb{R}^{N_c\times d_h}$, where $N_c = \sum_{v=1}^{V}\sum_{i=1}^{N} m_i^v$ is the total number of observed sample representations.We use $a, b \in \{1, \ldots, N_c\}$ to index rows of $\boldsymbol{H}$. Each row represents an observed sample in a particular view; therefore, different rows may correspond to different views of the same sample or to different samples. We construct $\boldsymbol{S} \in \mathbb{R}^{N_c\times N_c}$ using cosine similarity, with $S_{ab} = \langle \boldsymbol{h}_a, \boldsymbol{h}_b\rangle/(\|\boldsymbol{h}_a\|_2\|\boldsymbol{h}_b\|_2)$.

For each anchor a, potential positive and negative relations are selected from its similarity scores:

$$\mathcal{P}(a) = \mathrm{Top}_{\rho_{\mathrm{pos}}}(\boldsymbol{S}_{a,:}) \qquad \mathcal{N}(a) = \mathrm{Bottom}_{\rho_{\mathrm{neg}}}(\boldsymbol{S}_{a,:}) \tag{4}$$

Here, ρ_{pos} and ρ_{neg} denote the positive and negative selection ratios, respectively. The anchor itself is excluded before selection, and the two sets are constrained to be nonempty and disjoint. These sets represent relations inferred from the current representation space rather than ground-truth class correspondences.

The unified-space relational objective is

$$\mathcal{L}_{\mathrm{URL}} = -\frac{1}{N_c}\sum_{a=1}^{N_c}\frac{1}{|\mathcal{P}(a)|}\sum_{p\in\mathcal{P}(a)}\log\frac{\exp(S_{ap}/\tau)}{\exp(S_{ap}/\tau) + \sum_{n\in\mathcal{N}(a)}\exp(S_{an}/\tau)} \tag{5}$$

where $\tau > 0$ is the contrastive temperature. For each potential positive relation, the objective contrasts its similarity against the selected negative relations. In this way, observations without direct cross-view counterparts can still contribute to learning through their relations with other observed representations.

3.3 Cross-View Structural Learning

Although unified-space relational learning expands the available supervision, relations inferred from individual representations may still be unreliable. In particular, view-specific characteristics and noisy observations can cause the representations of the same instance to deviate across views. We therefore use cross-view neighborhood structure to assess the structural reliability of instance-level relations.

For two views v and u, we construct a cross-view affinity matrix $\boldsymbol{A}^{(v,u)} \in \mathbb{R}^{N\times N}$. For an observed pair satisfying $m_i^v m_j^u = 1$, its affinity is defined by the Gaussian kernel $A_{ij}^{(v,u)} = \exp(-\|\boldsymbol{h}_i^v -$

$\boldsymbol{h}_j^u\|_2^2/\sigma)$, where $\sigma > 0$ controls the distance decay. The affinity is set to zero whenever either observation is missing.

Before computing structural relations, each nonzero row of $\boldsymbol{A}^{(v,u)}$ is normalized to have unit sum, while an all-zero row remains unchanged. For simplicity, the row-normalized affinity matrix is still denoted by $\boldsymbol{A}^{(v,u)}$ in the following. The structural relation matrix is then constructed as

$$\boldsymbol{W}^{(v,u)} = \boldsymbol{A}^{(v,u)} \left(\boldsymbol{A}^{(v,u)}\right)^{\mathrm{T}} \tag{6}$$

The entry $W_{ij}^{(v,u)}$ measures the agreement between the cross-view affinity profiles of samples i and j. In particular, $W_{ii}^{(v,u)} = \|\boldsymbol{A}_{i,:}^{(v,u)}\|_2^2$ measures the concentration of the neighborhood affinity profile of sample i. A larger diagonal value indicates a more concentrated and structurally confident cross-view neighborhood.

For each valid view pair $(v,u) \in \mathcal{E}$ and each commonly observed sample $i \in \mathcal{I}^{(v,u)}$, $\boldsymbol{h}_i^v$ and $\boldsymbol{h}_i^u$ form a same-instance positive pair. We denote by $\mathcal{N}_i^{(v,u)}$ the indices of the lowest ρ_{neg} proportion of cosine similarities between $\boldsymbol{h}_i^v$ and the observed representations in view u, excluding sample i. Let $S_{ij}^{(v,u)}$ denote the cosine similarity between $\boldsymbol{h}_i^v$ and $\boldsymbol{h}_j^u$, computed using the same cosine measure as in unified-space relational learning.

For a view pair containing at least one commonly observed sample, the cross-view structural objective is

$$\mathcal{L}_{\mathrm{CSL}}^{(v,u)} = -\frac{1}{|\mathcal{I}^{(v,u)}|} \sum_{i \in \mathcal{I}^{(v,u)}} \log \frac{\exp\left(W_{ii}^{(v,u)} S_{ii}^{(v,u)}/\tau\right)}{\exp\left(W_{ii}^{(v,u)} S_{ii}^{(v,u)}/\tau\right) + \sum_{j \in \mathcal{N}_i^{(v,u)}} \exp\left(S_{ij}^{(v,u)}/\tau\right)} \qquad (v,u) \in \mathcal{E} \tag{7}$$

The diagonal structural weight modulates the positive logit according to the concentration of the anchor's cross-view neighborhood. Thus, same-instance pairs supported by clearer neighborhood structure receive different contrastive strengths from structurally ambiguous pairs.

Finally, the structural objectives are accumulated over view pairs that contain common observations:

$$\mathcal{L}_{\mathrm{CSL}} = \sum_{(v,u) \in \mathcal{E}} \mathcal{L}_{\mathrm{CSL}}^{(v,u)} \tag{8}$$

View pairs without commonly observed samples make no contribution to this objective.

3.4 Cross-View Soft Prototype Alignment

Sample and neighborhood relations improve representation learning, but they do not explicitly establish semantic correspondences between clusters learned from different views. Because missing observations change the sample distribution of each view, independently learned clusters may represent different semantic partitions. We therefore extend relational learning from samples to cluster prototypes.

For each view v, K-means is applied to its observed contrastive representations to obtain K clusters. The mean representation of cluster k is used as its prototype $\boldsymbol{c}_k^v \in \mathbb{R}^{d_h}$, and all prototypes are collected by rows in $\boldsymbol{C}^v \in \mathbb{R}^{K \times d_h}$.

For each observed representation $\boldsymbol{h}_i^v$, its soft assignment over the K prototypes is defined as

$$\boldsymbol{p}_i^v = \mathrm{Softmax}\left(\frac{\boldsymbol{C}^v \boldsymbol{h}_i^v}{\theta}\right) \qquad m_i^v = 1 \tag{9}$$

where $\boldsymbol{p}_i^v \in \mathbb{R}^K$ and $\theta > 0$ is the assignment temperature. Its k-th entry p_{ik}^v represents the soft association between sample i and prototype k in view v.

For each valid view pair $(v,u) \in \mathcal{E}$, commonly observed samples provide semantic evidence for matching view-specific prototypes. We accumulate their joint soft assignments to construct the prototype correspondence evidence matrix

$$\boldsymbol{R}^{(v,u)} = \sum_{i \in \mathcal{I}^{(v,u)}} \boldsymbol{p}_i^v \left(\boldsymbol{p}_i^u\right)^{\mathrm{T}} \in \mathbb{R}^{K \times K} \tag{10}$$

The entry $R_{kl}^{(v,u)}$ accumulates the support that commonly observed samples assign simultaneously to prototype k in view v and prototype l in view u. A larger value therefore indicates stronger evidence that the two prototypes express corresponding cluster semantics. The matrix $\boldsymbol{R}^{(v,u)}$ contains accumulated correspondence evidence rather than normalized probabilities.

We next evaluate the candidate correspondences according to the current prototype similarities. Let $\boldsymbol{B}^{(v,u)} \in \mathbb{R}^{K\times K}$ denote the prototype similarity matrix, whose entry is $B_{kl}^{(v,u)} = \langle \boldsymbol{c}_k^v, \boldsymbol{c}_l^u \rangle / (\|\boldsymbol{c}_k^v\|_2 \|\boldsymbol{c}_l^u\|_2)$. Row-wise and column-wise normalization produces the bidirectional matching score

$$\boldsymbol{Q}^{(v,u)} = \frac{1}{2}\left[\mathrm{Softmax}_{\mathrm{row}}\left(\boldsymbol{B}^{(v,u)}\right) + \mathrm{Softmax}_{\mathrm{col}}\left(\boldsymbol{B}^{(v,u)}\right)\right] \tag{11}$$

The entry $Q_{kl}^{(v,u)}$ combines the preference of prototype k for prototype l with the reverse preference of prototype l for prototype k. A larger value indicates a stronger relative match. The bidirectional score is not required to be symmetric or simultaneously row- and column-stochastic.

The prototype alignment objective for a valid view pair is

$$\mathcal{L}_{\mathrm{PA}}^{(v,u)} = -\frac{1}{K}\sum_{k=1}^{K}\sum_{l=1}^{K} R_{kl}^{(v,u)} \log Q_{kl}^{(v,u)} \qquad (v,u) \in \mathcal{E} \tag{12}$$

Prototype pairs supported by stronger soft-assignment co-occurrence therefore contribute more to semantic alignment.

When $\mathcal{E}$ is nonempty, the prototype alignment objective is averaged over all valid view pairs:

$$\mathcal{L}_{\mathrm{PA}} = \frac{1}{|\mathcal{E}|}\sum_{(v,u)\in\mathcal{E}} \mathcal{L}_{\mathrm{PA}}^{(v,u)} \tag{13}$$

3.5 Overall Objective

The proposed relational objectives are optimized together with a standard view-wise reconstruction objective:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{REC}} + \alpha\,\mathcal{L}_{\mathrm{URL}} + \beta\,\mathcal{L}_{\mathrm{CSL}} + \lambda\,\mathcal{L}_{\mathrm{PA}}, \tag{14}$$

where $\mathcal{L}_{\mathrm{REC}}$ denotes the standard reconstruction loss of the view-specific autoencoders, evaluated only on observed inputs. The nonnegative coefficients α, β, and λ control the contributions of unified-space relational learning, cross-view structural learning, and cross-view soft prototype alignment, respectively.

4 Experiments

4.1 Experimental Setup

We evaluate MGRL-IMVC on DHA (Lin et al., 2012), LandUse-21 (Yang & Newsam, 2010), Hand-Written (Dua & Graff, 2019), NoisyMNIST (LeCun et al., 1998; Wang et al., 2015), and ALOI (Geusebroek et al., 2005). We report mean ACC and NMI over five independent runs with randomized initializations and missing-view masks; ARI is used only to analyze performance during training.

The seven baselines are COMPLETER (Lin et al., 2021), DSIMVC (Tang & Liu, 2022a), CPSPAN (Jin et al., 2023), RPCIC (Yuan et al., 2024), DCG (Zhang et al., 2025), FreeCSL (Dai et al., 2025), and GLGC (He et al., 2026). MGRL-IMVC is implemented in Python 3.10 and PyTorch 2.1.2, running on Ubuntu 22.04 with CUDA 12.2 and two NVIDIA Tesla T4 GPUs (16 GB each).

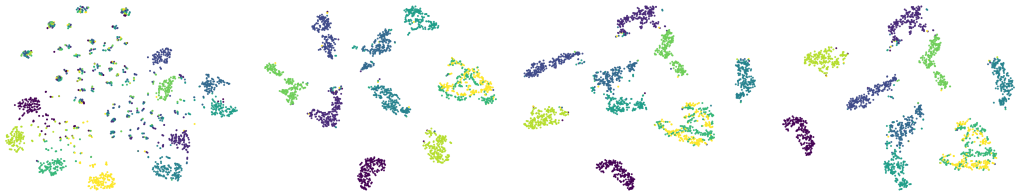
For each missing rate $p_m \in \{0.1, 0.3, 0.5, 0.7, 1.0\}$, we randomly select $p_m N$ samples and remove between 1 and $V-1$ views per selected sample, retaining at least one observed view. Thus, $p_m = 1.0$ means that every sample is incomplete.

Table 1: Clustering performance (%) under different missing rates. Each entry reports ACC/NMI. The best and second-best results are highlighted in bold and underlined, respectively.

Dataset	Method (Year)	Missing rate p_m (ACC/NMI)				
		0.1	0.3	0.5	0.7	1.0
DHA	COMPLETER (2021)	60.90/56.33	51.60/49.34	44.10/54.35	44.00/37.07	31.65/26.35
	DSIMVC (2022)	48.97/63.25	44.95/62.75	39.39/59.15	39.46/56.25	35.99/43.62
	CPSPAN (2023)	63.78/76.23	57.91/73.31	52.13/70.50	45.82/63.75	34.11/48.81
	RPCIC (2024)	56.73/68.28	48.65/58.92	35.21/44.98	21.82/31.70	16.61/23.42
	DCG (2025)	72.83/78.79	66.44/78.65	63.02/76.58	57.47/68.75	23.80/35.54
	FreeCSL (2025)	78.46/84.10	75.00/80.87	68.66/77.00	56.26/62.77	31.68/44.26
	GLGC (2026)	84.02/84.84	79.92/79.70	76.33/79.76	65.34/70.00	41.73/58.08
	MGRL-IMVC	**86.96/86.90**	**83.99/84.04**	**79.57/80.91**	**72.05/75.29**	**43.00/58.12**
LandUse-21	COMPLETER (2021)	15.72/28.64	15.74/23.06	12.89/19.76	10.63/19.70	8.05/14.28
	DSIMVC (2022)	20.35/19.70	18.85/21.35	19.72/21.37	18.60/15.73	7.46/11.38
	CPSPAN (2023)	24.71/**32.47**	21.42/29.53	20.14/**30.29**	21.51/26.42	15.91/18.04
	RPCIC (2024)	22.49/28.94	16.87/21.41	16.99/21.19	18.67/21.25	15.66/19.02
	DCG (2025)	27.44/29.52	24.43/30.50	23.23/28.22	24.05/25.20	11.81/11.84
	FreeCSL (2025)	26.61/28.84	26.52/27.57	23.93/27.38	25.17/25.85	14.44/14.05
	GLGC (2026)	27.39/29.24	26.60/29.75	25.88/29.45	24.59/27.11	**19.95**/19.86
	MGRL-IMVC	**28.62**/32.36	**27.67/31.53**	**26.83**/30.17	**25.77/28.32**	18.73/**21.52**
HandWritten	COMPLETER (2021)	84.71/85.21	61.08/67.97	48.19/55.44	43.02/41.79	26.30/32.81
	DSIMVC (2022)	79.74/74.95	78.29/73.16	74.30/68.72	65.61/62.04	61.33/66.70
	CPSPAN (2023)	88.43/88.36	85.88/84.19	79.40/75.24	77.49/71.68	72.61/68.96
	RPCIC (2024)	91.13/**95.22**	84.28/**93.34**	82.56/**90.53**	76.44/**85.00**	68.06/67.68
	DCG (2025)	85.32/83.14	81.26/78.68	80.01/78.02	71.79/65.72	62.78/65.64
	FreeCSL (2025)	86.28/86.91	84.04/81.89	76.73/73.47	76.24/69.13	69.71/66.98
	GLGC (2026)	90.72/88.42	84.04/81.66	82.38/76.98	**80.35**/75.45	75.62/69.39
	MGRL-IMVC	**94.99**/94.03	**87.23**/83.82	**84.40**/80.41	79.34/77.94	**77.75/72.58**
NoisyMNIST	COMPLETER (2021)	76.55/87.95	78.16/86.11	64.79/65.77	41.15/40.60	39.92/37.82
	DSIMVC (2022)	85.72/74.62	83.39/73.44	82.11/73.07	**80.45**/71.22	**78.22**/69.88
	CPSPAN (2023)	85.86/88.59	77.11/84.36	74.20/79.25	74.55/81.33	71.80/64.65
	RPCIC (2024)	89.77/85.43	84.09/86.06	74.43/83.66	61.91/64.65	55.42/60.44
	DCG (2025)	93.89/87.23	91.75/**93.94**	**89.06/89.81**	78.25/**82.50**	63.29/74.86
	FreeCSL (2025)	90.78/90.36	89.75/93.71	75.64/87.24	73.91/80.91	70.24/66.83
	GLGC (2026)	94.72/91.44	90.85/89.10	87.09/87.29	68.57/75.61	53.58/66.98
	MGRL-IMVC	**95.61/98.30**	**93.11**/93.29	88.85/87.16	76.00/80.04	72.25/**77.20**
ALOI	COMPLETER (2021)	20.91/40.20	17.49/38.96	14.43/32.50	13.08/24.82	9.86/15.37
	DSIMVC (2022)	52.18/67.57	49.05/65.06	48.22/64.58	43.73/59.28	30.50/38.42
	CPSPAN (2023)	67.55/82.62	68.83/82.51	67.15/81.47	67.18/80.38	66.49/80.98
	RPCIC (2024)	77.57/88.45	66.90/86.59	56.11/78.62	59.48/82.46	45.13/67.70
	DCG (2025)	60.44/84.20	56.90/83.98	52.90/80.12	42.40/74.44	28.58/23.06
	FreeCSL (2025)	86.27/93.28	85.63/89.16	82.37/87.68	76.51/83.89	46.26/66.71
	GLGC (2026)	88.84/**94.94**	89.57/92.69	**89.18/92.31**	**83.80/90.31**	**81.14/87.36**
	MGRL-IMVC	**91.95**/94.85	**91.25/93.86**	88.20/89.78	80.30/87.03	78.85/79.05

4.2 Clustering Performance Comparison

Table 1 compares the methods across five datasets, five missing rates, and two evaluation metrics. MGRL-IMVC achieves an overall mean score of 71.84%, exceeding the strongest baseline, GLGC, by 2.24 percentage points. It ranks first in 29 of the 50 dataset–rate–metric settings and obtains the highest dataset-level average in seven of the ten dataset–metric combinations. The improvement is particularly consistent on DHA, where MGRL-IMVC achieves the best ACC and NMI at every missing rate. Under severe incompleteness, it is not always optimal on HandWritten and ALOI, possibly because fewer commonly observed samples provide weaker co-occurrence evidence for estimating cross-view prototype correspondences. Nevertheless, the strong overall results demonstrate the effectiveness of jointly modeling sample relations, neighborhood structure, and prototype-level semantics.



(a) Raw (b) CSL (c) URL+CSL (d) Full model

Figure 2: t-SNE visualizations of different ablation variants on HandWritten at $p_m = 0.7$. The full model combines URL, CSL, and PA.

Table 2: Ablation results in terms of ACC (%) under different missing rates.

Dataset	Included modules			Missing rate p_m				
	URL	CSL	PA	0.1	0.3	0.5	0.7	1.0
DHA	✓			71.47	67.03	66.88	63.15	40.62
	✓	✓		84.47	79.14	76.39	65.34	41.73
	✓	✓	✓	86.96	83.99	79.57	72.05	43.00
LandUse-21	✓			22.64	20.19	19.86	15.07	11.30
	✓	✓		27.35	26.56	25.44	23.19	19.18
	✓	✓	✓	28.62	27.67	26.83	25.77	18.73
HandWritten	✓			85.11	80.73	78.90	75.82	74.49
	✓	✓		90.70	83.98	82.38	78.35	75.47
	✓	✓	✓	94.99	87.23	84.40	79.34	77.75
NoisyMNIST	✓			81.70	76.64	73.75	66.08	60.92
	✓	✓		94.73	90.96	87.63	68.57	56.36
	✓	✓	✓	95.61	93.11	88.85	76.00	72.25
ALOI	✓			85.16	83.72	80.44	76.95	70.99
	✓	✓		88.80	89.42	88.18	80.73	79.12
	✓	✓	✓	91.95	91.25	88.20	80.30	78.85

4.3 Ablation Study and Visualization

We evaluate the three relational modules through progressively combined variants on all five datasets. All variants use identical encoders, decoders, and reconstruction objectives, so their differences arise only from URL, CSL, and PA. As reported in Table 2, adding CSL to URL improves ACC in 24 of the 25 settings, with an average gain of 6.18 percentage points. This improvement demonstrates that cross-view neighborhood structure provides useful evidence for correcting relations inferred solely from global similarity. Incorporating PA further improves 22 settings by 2.76 percentage points on average, confirming the complementary contribution of prototype-level semantic correspondence. Consequently, the full model outperforms URL alone in all 25 settings, with an average improvement of 8.95 percentage points.

The slight decreases introduced by PA under several high-missing-rate settings on LandUse-21 and ALOI indicate that prototype alignment depends on the availability of commonly observed samples. Nevertheless, the consistent overall improvements verify that sample relations, neighborhood structure, and prototype correspondence provide complementary supervision. The t-SNE results in Figure 2 further support this observation: progressively incorporating the relational modules produces more compact and better-separated clusters on HandWritten at $p_m = 0.7$.

4.4 Cross-View Prototype Alignment Analysis

To investigate whether the estimated prototype correspondences can resolve cross-view semantic permutations, we visualize the correspondence evidence matrices on HandWritten at $p_m = 0.5$. HandWritten contains six views and therefore yields 15 view pairs. For a concise and representative comparison, we select V_1 as the reference view and present its correspondence matrices with each

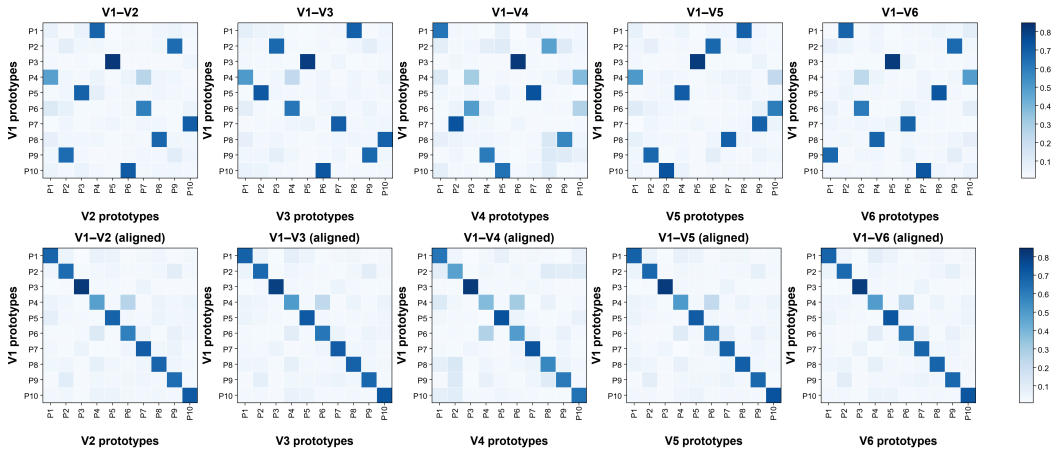


Figure 3: Prototype correspondence matrices before alignment (top) and after alignment (bottom) between V_1 and V_2–V_6 on HandWritten at $p_m = 0.5$.

of the remaining five views, i.e., V_1–V_2 through V_1–V_6. Figure 3 compares these matrices before and after correspondence-based alignment.

Before alignment, the dominant responses are frequently distributed across off-diagonal positions because independently learned view-specific clusters may use different prototype orders. After reordering the prototypes according to the estimated correspondences, most matrices exhibit a clearer diagonally dominant structure. Since the reordering changes only the prototype indices and leaves the correspondence weights unchanged, this concentration indicates that soft-assignment co-occurrence provides informative evidence for identifying cross-view prototype correspondences.

Several view pairs still contain noticeable off-diagonal responses, reflecting residual semantic ambiguity caused by view heterogeneity and incomplete observations. These uncertain correspondences support the use of soft alignment weights, which preserve multiple plausible cross-view relations instead of imposing a potentially unreliable hard one-to-one match.

5 Conclusion

We proposed MGRL-IMVC, an imputation-free framework for incomplete multi-view clustering that connects relational learning at three granularities. It expands contrastive supervision through global sample relations, evaluates cross-view pairs using neighborhood structure, and aligns cluster semantics through soft prototype correspondences. Experiments on five benchmark datasets under different missing rates demonstrate the effectiveness and robustness of the proposed framework.

AI USE STATEMENT

We have reviewed all AI-assisted work. Specifically, we manually checked the references compiled with the help of large language models to ensure that all citations actually exist and match the content discussed. We take responsibility for the final content of this work, including any text, statements, or results generated with the aid of generative AI. For more details, please refer to the ICLR 2027 Author AI Policy.

REFERENCES

Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. In *Advances in Neural Information Processing Systems*, pp. 9912–9924, 2020.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the International Conference on Machine Learning*, pp. 1597–1607, 2020.

Ching-Yao Chuang, Joshua W. Robinson, Yen-Chen Lin, Antonio Torralba, and Stefanie Jegelka. Debiased Contrastive Learning. In *Advances in Neural Information Processing Systems*, pp. 8765–8775, 2020.

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

Dheeru Dua and Casey Graff. UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences, 2019.

Jan-Mark Geusebroek, Gertjan J. Burghouts, and Arnold W. M. Smeulders. The Amsterdam Library of Object Images. *International Journal of Computer Vision*, pp. 103–112, 2005.

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*, pp. 24813–24822, 2026.

Menglei Hu and Songcan Chen. Doubly Aligned Incomplete Multi-view Clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 2262–2268, 2018.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11600–11609, 2023.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, pp. 2278–2324, 1998.

Junnan Li, Pan Zhou, Caiming Xiong, and Steven C. H. Hoi. Prototypical Contrastive Learning of Unsupervised Representations. In *Proceedings of the International Conference on Learning Representations*, 2021a.

Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. Partial Multi-View Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 1968–1974, 2014.

Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 8547–8555, 2021b.

Yan-Ching Lin, Min-Chun Hu, Wen-Huang Cheng, Yung-Huan Hsieh, and Hong-Ming Chen. Human Action Recognition and Retrieval Using Sole Depth Information. In *Proceedings of the 20th ACM International Conference on Multimedia*, pp. 1053–1056, 2012.

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*, pp. 4447–4461, 2023.

Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, Chang Tang, Jianping Yin, Dinggang Shen, Huaimin Wang, and Wen Gao. Late Fusion Incomplete Multi-View Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2410–2423, 2019.

Xinwang Liu, Miaomiao Li, Chang Tang, Jingyuan Xia, Jian Xiong, Li Liu, Marius Kloft, and En Zhu. Efficient and Effective Regularized Incomplete Multi-view Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2634–2646, 2021.

Yiding Lu, Yijie Lin, Mouxing Yang, Dezhong Peng, Peng Hu, and Xi Peng. Decoupled Contrastive Multi-View Clustering with High-Order Random Walks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 14193–14201, 2024.

Erlin Pan and Zhao Kang. Multi-view Contrastive Graph Clustering. In *Advances in Neural Information Processing Systems*, pp. 2148–2159, 2021.

Yuan Sun, Yang Qin, Yongxiang Li, Dezhong Peng, Xi Peng, and Peng Hu. Robust Multi-View Clustering With Noisy Correspondence. *IEEE Transactions on Knowledge and Data Engineering*, pp. 9150–9162, 2024.

Huayi Tang and Yong Liu. Deep Safe Incomplete Multi-view Clustering: Theorem and Algorithm. In *Proceedings of the International Conference on Machine Learning*, pp. 21090–21110, 2022a.

Huayi Tang and Yong Liu. Deep Safe Multi-View Clustering: Reducing the Risk of Clustering Performance Degradation Caused by View Increase. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 202–211, 2022b.

Hao Wang, Linlin Zong, Bing Liu, Yan Yang, and Wei Zhou. Spectral Perturbation Meets Incomplete Multi-view Data. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 3677–3683, 2019.

Weiran Wang, Raman Arora, Karen Livescu, and Jeff Bilmes. On Deep Multi-View Representation Learning. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 1083–1092, 2015.

Yiming Wang, Dongxia Chang, Zhiqiang Fu, Jie Wen, and Yao Zhao. Graph Contrastive Partial Multi-View Clustering. *IEEE Transactions on Multimedia*, pp. 6551–6562, 2023.

Yiming Wang, Qun Li, Dongxia Chang, Jie Wen, Hua Dai, Fu Xiao, and Yao Zhao. Learning from Disjoint Views: A Contrastive Prototype Matching Network for Fully Incomplete Multi-View Clustering. In *Advances in Neural Information Processing Systems*, pp. 18265–18284, 2025.

Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei, and Guo-Sen Xie. CDIMC-net: Cognitive Deep Incomplete Multi-view Clustering Network. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 3230–3236, 2020.

Jie Wen, Chengliang Liu, Gehui Xu, Zhihao Wu, Chao Huang, Lunke Fei, and Yong Xu. Highly Confident Local Structure Based Consensus Graph Learning for Incomplete Multi-View Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15712–15721, 2023a.

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*, pp. 1136–1149, 2023b.

Baili Xiao, Zhibin Dong, Ke Liang, Suyuan Liu, Siwei Wang, Tianrui Liu, Xingchen Hu, En Zhu, and Xinwang Liu. EASEMVC: Efficient Dual Selection Mechanism for Deep Multi-View Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20716–20726, 2025.

Cai Xu, Ziyu Guan, Wei Zhao, Hongchang Wu, Yunfei Niu, and Beilei Ling. Adversarial Incomplete Multi-view Clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 3933–3939, 2019.

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*, pp. 16147–16155, 2024a.

Jie Xu, Chao Li, Yazhou Ren, Liang Peng, Yujie Mo, Xiaoshuang Shi, and Xiaofeng Zhu. Deep Incomplete Multi-View Clustering via Mining Cluster Complementarity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 8761–8769, 2022a.

Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-Level Feature Learning for Contrastive Multi-View Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16051–16060, 2022b.

Jie Xu, Yazhou Ren, Xiaolong Wang, Lei Feng, Zheng Zhang, Gang Niu, and Xiaofeng Zhu. Investigating and Mitigating the Side Effects of Noisy Views for Self-Supervised Clustering Algorithms in Practical Multi-View Scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22957–22966, 2024b.

Mouxing Yang, Yunfan Li, Zhenyu Huang, Zitao Liu, Peng Hu, and Xi Peng. Partially View-Aligned Representation Learning With Noise-Robust Contrastive Loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1134–1143, 2021.

Yi Yang and Shawn Newsam. Bag-of-Visual-Words and Spatial Extensions for Land-Use Classification. In *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 270–279, 2010.

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.

Pengxin Zeng, Mouxing Yang, Yiding Lu, Changqing Zhang, Peng Hu, and Xi Peng. Semantic Invariant Multi-View Clustering With Fully Incomplete Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2139–2150, 2024.

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*, pp. 22650–22658, 2025.

Liang Zhao, Ziyue Wang, Chuanye He, Qingchen Zhang, and Bo Xu. Dual Robust Unbiased Multi-View Clustering for Incomplete and Unpaired Information. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 7092–7100, 2025.