TOWARDS DISCRIMINATIVE INCOMPLETE MULTI-VIEW CLUSTERING VIA PROTOTYPE DIFFUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete multi-view clustering (IMVC) aims to learn unified representations from multi-source data with partially unavailable views for unsupervised clustering. Existing methods often rely heavily on limited paired observations for instance-level cross-view alignment, making it difficult to preserve globally consistent semantic structures under severe view incompleteness. To address this issue, we propose a Discriminative Prototype Diffusion framework, termed DPD. DPD introduces learnable shared prototypes as global semantic anchors and aligns view-specific latent distributions with the prototype distribution through prototype-guided maximum mean discrepancy regularization. In addition to paired-sample contrastive alignment, we further introduce a lightweight latent diffusion objective to regularize view-specific representations in the prototype-aligned latent space. Since the diffusion objective and prototype alignment operate on the same latent representations, they jointly encourage a semantically structured and discriminative representation space without requiring prototypes to be explicitly injected into the denoising network. Finally, a mask-aware prototypical spherical fusion module enhances each available view with shared prototype information and aggregates the observed representations into a normalized unified embedding for clustering. Experiments on five incomplete multi-view benchmark datasets under different missing rates demonstrate that DPD achieves competitive or superior clustering performance compared with representative methods.

1 INTRODUCTION

Multi-view data describe the same objects through heterogeneous sources and provide complementary information for representation learning and clustering. In practical applications, however, some views can be unavailable because of acquisition failures, privacy constraints, or heterogeneous data-collection processes, giving rise to incomplete multi-view clustering (IMVC) (Wang et al., 2024; Lei et al., 2024; Yuan et al., 2025; Wen et al., 2022). The central challenge of IMVC is to learn a unified representation from partially observed multi-view data while preserving cross-view consistency and clustering-discriminative structures.

A major line of IMVC research exploits the correspondence between available views to learn consistent representations. For example, Completer (Lin et al., 2021) performs cross-view contrastive prediction, while DCP (Lin et al., 2022) further introduces dual contrastive prediction for incomplete multi-view representation learning. Although instance-level correspondence is effective when paired observations are sufficiently available(Dong et al., 2025; Yin et al., 2025; Zhuang et al., 2025), its supervision becomes more limited as view incompleteness increases.

To capture higher-level semantic structures, recent studies introduce prototype-level information into IMVC. CPSPAN (Jin et al., 2023) jointly considers cross-view partial-sample and prototype alignment, RPCIC (Yuan et al., 2024) exploits robust prototype completion, and FreeCSL (Dai et al., 2025) learns consensus prototypes to construct a semantic space shared across available views. These results suggest that prototype-level constraints provide a useful complement to purely instance-level correspondence. However, prototype-based alignment primarily structures the global semantic organization of the representation space, leaving the local distributional regularization of latent representations less explored.

This observation motivates the qualitative comparison in Figure 1: diffusion-based regularization

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.

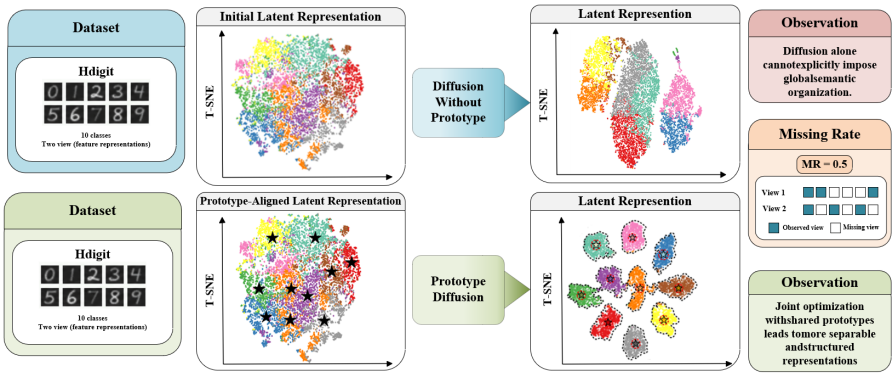


Figure 1: Qualitative motivation for coupling prototype alignment with latent diffusion regularization. Diffusion regularization without global prototype alignment may yield entangled latent structures, whereas joint optimization with shared prototype alignment leads to more separable representations. The visualization is obtained on Hdigit at MR=0.5.

alone does not explicitly impose a global semantic organization, whereas coupling it with shared prototype alignment yields a more structured latent representation.

Another family of IMVC methods focuses on reconstructing or inferring missing information(Wen et al., 2020; Liu et al., 2024b; Yao et al., 2024). LaSA (Liu et al., 2024a) combines latent structure modeling with view recovery, while DCG (Zhang et al., 2025) introduces diffusion contrastive generation for missing-view recovery. More recently, HSACC (Ding et al., 2026) combines hierarchical semantic alignment with cooperative completion and adaptive multi-view fusion. These methods demonstrate the effectiveness of completion-oriented learning for incomplete observations(Zheng et al., 2026; Jin et al., 2025; Wang et al., 2026).

Different from using diffusion primarily as a cross-view completion mechanism, we investigate whether diffusion can instead act as a regularizer of a semantically structured latent space. Specifically, we couple a denoising-based latent diffusion objective with prototype-guided distribution alignment, so that both objectives jointly optimize the same view-specific representations.

To address these challenges, we propose a novel discriminative incomplete multi-view clustering framework termed Discriminative Prototype Diffusion (DPD). The core idea is to jointly exploit global prototype-based semantic alignment and latent diffusion regularization within a common representation space. Specifically, learnable shared prototypes serve as global semantic anchors to regularize view-specific latent distributions, alleviating semantic inconsistency caused by limited paired observations. Meanwhile, paired-sample contrastive learning preserves instance-level correspondence between available cross-view pairs. On top of the prototype-aligned latent space, we introduce lightweight view-specific diffusion objectives that model and regularize the latent distributions through noise prediction. Importantly, prototypes are not explicitly provided as inputs to the denoising networks; instead, prototype information influences diffusion optimization implicitly through the shared prototype-aligned latent representations. Finally, a mask-aware prototypical spherical fusion module enriches each available representation through prototype cross-attention and aggregates the observed views into a normalized unified embedding for clustering. The main contributions of this paper are summarized as follows:

- A discriminative incomplete multi-view clustering framework, termed DPD, is developed to jointly exploit instance-level correspondence, global prototype-guided semantic alignment, and latent distribution regularization within a unified representation space.
- Learnable shared prototypes are employed as global semantic anchors through prototype-guided MMD alignment, complementing paired-sample contrastive learning with distribution-level semantic regularization under incomplete observations.
- Prototype-aligned latent diffusion regularization and mask-aware prototypical spherical fusion are further designed to improve local latent regularity and construct robust unified representations from the available views.

2 METHOD

In this section, we introduce the proposed method consisting of four modules. Figure 2 illustrates the overall architecture of the method.

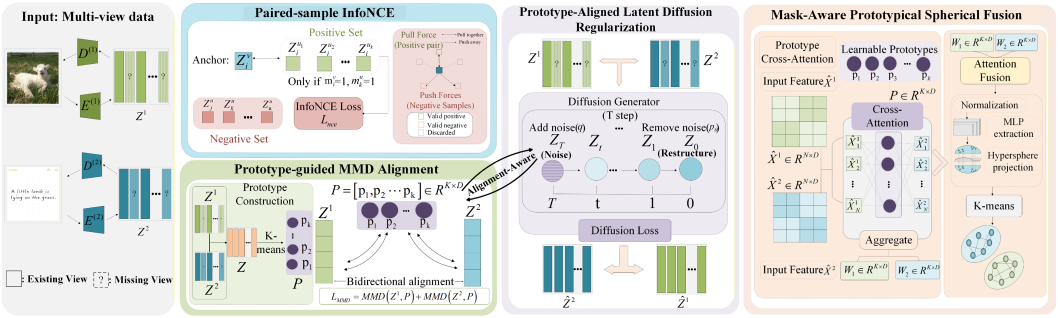


Figure 2: Schematic overview of the proposed DPD framework. View-specific autoencoders first map available observations into a common latent dimensionality. Paired-sample InfoNCE preserves instance-level correspondence between available cross-view pairs, while Prototype-guided MMD Alignment regularizes the view-specific latent distributions using shared learnable prototypes as global semantic anchors. Lightweight view-specific diffusion objectives further regularize representations in the prototype-aligned latent space. Finally, prototype cross-attention enhances the available latent representations, which are aggregated according to view availability and projected onto a hypersphere for K-means clustering.

2.1 PROBLEM FORMULATION

Let $\mathcal{X} = \{X^a, X^b\}$ denote an incomplete two-view dataset with N samples, where $X^v = \{x_i^v\}_{i=1}^N$ represents the v-th view. The observation status is described by a binary matrix $M \in \{0,1\}^{N\times 2}$, where $m_i^v = 1$ indicates that the v-th view of sample i is available and $m_i^v = 0$ otherwise.

For each sample, we define the set of available views as

$$\Omega_i = \{v \in \{a, b\} \mid m_i^v = 1\}. \tag{1}$$

We further denote the paired subset containing samples whose two views are both observed as

$$\mathcal{D}_c = \{i \mid m_i^a = 1,\ m_i^b = 1\}. \tag{2}$$

Following our implementation, model optimization is performed on the available paired subset $\mathcal{D}_c$, while incomplete samples are handled at representation construction time through mask-aware aggregation of their observed views. Our goal is to learn a unified representation h_i for each incomplete sample such that the resulting latent space preserves cross-view semantic consistency and discriminative clustering structures.

2.2 VIEW-SPECIFIC LATENT ENCODING

We employ two view-specific autoencoders to map heterogeneous observations into latent representations of the same dimensionality. For view $v \in \{a, b\}$,

$$z_i^v = E^v(x_i^v), \tag{3}$$
$$\hat{x}_i^v = D^v(z_i^v), \tag{4}$$

where $E^v(\cdot)$ and $D^v(\cdot)$ denote the encoder and decoder, respectively.

During optimization, reconstruction is performed on the paired training subset $\mathcal{D}_c$:

$$\mathcal{L}_{\text{rec}} = \sum_{v\in\{a,b\}} \mathbb{E}_{i\in\mathcal{D}_c} \left[\|x_i^v - \hat{x}_i^v\|_2^2 \right]. \tag{5}$$

The reconstruction objective encourages each encoder to preserve view-specific information before cross-view semantic regularization.

2.3 PAIRED-SAMPLE INFONCE

To preserve instance-level correspondence between the two views, we apply contrastive alignment to paired observations in $\mathcal{D}_c$. The latent representations are first ℓ_2-normalized:

$$\bar{z}_i^v = \frac{z_i^v}{\|z_i^v\|_2}. \tag{6}$$

For a paired mini-batch $\mathcal{B} \subseteq \mathcal{D}_c$, the cross-view similarity is defined as

$$s(z_i^a, z_j^b) = (\bar{z}_i^a)^\top \bar{z}_j^b. \tag{7}$$

The contrastive objective is

$$\mathcal{L}_{\text{nce}} = -\frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \log \frac{\exp(s(z_i^a, z_i^b)/\tau)}{\exp(s(z_i^a, z_i^b)/\tau) + \sum_{\substack{j \in \mathcal{B} \\ j \neq i}} \exp(s(z_i^a, z_j^b)/\tau)}, \tag{8}$$

where τ is the temperature parameter.

This objective preserves sample-level correspondence whenever paired observations are available. However, under a high missing rate, the number of such paired observations can become limited, making instance-level alignment alone insufficient to characterize the global semantic geometry. We therefore introduce shared prototypes as distribution-level semantic anchors.

2.4 PROTOTYPE-GUIDED MMD ALIGNMENT

Instance-level alignment alone does not explicitly impose a global semantic structure on the latent space. We therefore introduce K learnable shared prototypes

$$P = \{p_k\}_{k=1}^K \in \mathbb{R}^{K \times d}, \tag{9}$$

where d is the latent dimensionality. The prototypes are optimized jointly with the representation networks and serve as shared semantic anchors for both views.

Given two sets of representations $A = \{a_i\}_{i=1}^n$ and $B = \{b_j\}_{j=1}^m$, we measure their distribution discrepancy using maximum mean discrepancy:

$$\text{MMD}^2(A, B) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n k(a_i, a_j) + \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m k(b_i, b_j) - \frac{2}{nm} \sum_{i=1}^n \sum_{j=1}^m k(a_i, b_j), \tag{10}$$

where $k(\cdot, \cdot)$ denotes a multi-scale Gaussian kernel:

$$k(a, b) = \sum_{q=1}^Q \exp\left(-\frac{\|a - b\|_2^2}{\sigma_q}\right). \tag{11}$$

Instead of directly minimizing only the discrepancy between the two view distributions, we align both of them toward the same prototype distribution:

$$\mathcal{L}_{\text{pga}} = \frac{1}{2}\left[\text{MMD}^2(Z^a, P) + \text{MMD}^2(Z^b, P)\right]. \tag{12}$$

The shared prototype distribution provides a common reference for both views and therefore complements the local correspondence captured by $\mathcal{L}_{\text{nce}}$. Since the prototypes and view-specific encoders are jointly optimized, this objective shapes the global geometry of the latent representation space rather than requiring explicit prototype assignments for individual samples.

2.5 PROTOTYPE-ALIGNED LATENT DIFFUSION REGULARIZATION

Although prototype-guided MMD alignment imposes a shared distribution-level semantic structure, the latent representations can still exhibit local distributional variations. We therefore introduce a lightweight diffusion objective as an additional regularizer of the view-specific latent representations.

For a latent representation z_0^v from view $v \in \{a, b\}$, the forward diffusion process generates a noisy latent variable

$$z_t^v = \sqrt{\bar{\alpha}_t} z_0^v + \sqrt{1 - \bar{\alpha}_t}\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I), \tag{13}$$

where $t \in \{1, \ldots, T\}$, $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$.

For each view, a lightweight denoising network G_{θ_v} takes the noisy latent representation together with the timestep embedding and predicts the injected noise:

$$\hat{\epsilon} = G_{\theta_v}(z_t^v, t). \tag{14}$$

The corresponding diffusion regularization objective is

$$\mathcal{L}_{\text{diff}}^v = \mathbb{E}_{z_0^v, t, \epsilon}\left[\|\epsilon - G_{\theta_v}(z_t^v, t)\|_2^2\right]. \tag{15}$$

For the two-view setting, we optimize

$$\mathcal{L}_{\text{diff}} = \mathcal{L}_{\text{diff}}^a + \mathcal{L}_{\text{diff}}^b. \tag{16}$$

We emphasize that the shared prototypes are not explicitly provided as conditioning inputs to the denoising networks. Instead, prototype alignment and diffusion regularization are coupled through the same latent representations: the encoders are jointly optimized by $\mathcal{L}_{\text{pga}}$ and $\mathcal{L}_{\text{diff}}$. Consequently, the diffusion objective operates on and regularizes a latent space whose global geometry is simultaneously shaped by the shared prototypes. We refer to this mechanism as *prototype-aligned latent diffusion regularization*.

2.6 MASK-AWARE PROTOTYPICAL SPHERICAL FUSION

At clustering time, the representation of each available view is further enhanced by interaction with the shared prototypes. For $v \in \Omega_i$, we compute

$$A_i^v = \text{softmax}\left(z_i^v P^\top\right), \tag{17}$$

and obtain the prototype context

$$\tilde{z}_i^v = A_i^v P. \tag{18}$$

The prototype-enhanced representation is defined as

$$r_i^v = z_i^v + \tilde{z}_i^v. \tag{19}$$

We use the observed-view indicator directly as a binary aggregation mask:

$$c_i^v = m_i^v. \tag{20}$$

The unified representation is constructed as

$$h_i = \sum_{v \in \{a,b\}} m_i^v r_i^v. \tag{21}$$

Therefore, only observed views contribute directly to the final clustering representation. Missing-view entries are excluded from the aggregation rather than being assigned a learned confidence score. This conservative fusion rule prevents unavailable views from introducing unreliable information into the clustering embedding. We project the aggregated representation onto the unit hypersphere:

$$h_i \leftarrow \frac{h_i}{\|h_i\|_2}. \tag{22}$$

The resulting representation matrix $H = \{h_i\}_{i=1}^N$ is finally clustered using K-means.

2.7 Overall Objective and Optimization

The complete objective is

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{rec}} + \eta(e)\lambda_2 \mathcal{L}_{\text{diff}} + \lambda_3 \mathcal{L}_{\text{nce}} + \lambda_4 \mathcal{L}_{\text{pga}}, \tag{23}$$

where λ_1, λ_2, λ_3, and λ_4 control the relative contributions of the four objectives. To avoid imposing the diffusion objective on an insufficiently structured latent space during early optimization, we activate diffusion regularization only after a predefined warm-up stage:

$$\eta(e) = \begin{cases} 0, & e < e_0, \\ 1, & e \geq e_0, \end{cases} \tag{24}$$

where e denotes the current training epoch and e_0 is the diffusion activation epoch. The encoders, decoders, shared prototypes, and view-specific denoising networks are optimized jointly. Prototype-guided alignment and latent diffusion regularization therefore influence the same view-specific representations through a shared optimization process, while the final clustering representation is constructed solely from the available views.

3 Experiments

3.1 Experimental Setting

We validate the performance of the proposed DPD algorithm on five common multi-view datasets and compare it with ten state-of-the-art multi-view clustering methods.

MR	Method	Scene-15			LandUse-21			Hdigit			CUB			NoisyMNIST		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
0.3	COMP	39.24	42.57	**37.85**	26.85	31.10	12.87	93.84	85.66	86.86	60.73	68.88	52.78	84.76	8.89	77.00
	DSIMVC	31.38	32.54	16.29	21.43	25.25	8.04	94.30	91.38	91.30	49.22	49.93	32.80	63.70	60.70	51.40
	DCP	40.11	42.02	23.32	25.48	30.18	11.08	94.31	93.33	89.43	40.60	52.37	31.41	85.53	84.57	81.35
	CPSPAN	38.39	42.66	24.93	18.08	25.75	6.90	95.75	90.02	90.84	61.30	64.21	48.93	49.94	50.99	36.95
	ICMVC	36.20	34.21	19.69	26.12	29.83	12.28	16.34	8.67	3.64	45.13	41.32	25.11	**96.89**	**92.09**	**93.29**
	RPCIC	38.04	39.67	20.80	21.29	26.52	8.21	91.06	82.49	81.23	67.50	67.60	54.28	53.19	52.18	39.21
	MCMVC	38.20	41.88	21.52	23.67	29.38	11.20	85.23	81.37	78.19	55.67	58.98	35.83	69.78	72.02	52.44
	MASK-IMVC	27.72	32.09	16.36	18.84	24.43	7.27	59.39	53.06	39.95	73.91	71.60	57.89	31.58	28.10	17.80
	DCG	37.59	42.74	24.61	27.33	32.09	14.47	95.47	93.41	94.46	72.67	66.98	53.82	93.28	84.46	85.84
	DGIMVCM	41.00	43.68	26.51	28.56	**34.18**	13.89	81.04	81.66	75.78	49.50	55.40	39.11	59.18	56.79	46.69
	DPD(Ours)	**41.98**	**43.78**	26.83	**30.00**	32.53	**15.34**	**98.24**	**95.60**	**96.17**	**75.00**	**72.71**	**61.64**	93.61	85.62	86.48
0.5	COMP	38.73	41.76	**36.91**	25.24	31.69	13.53	91.36	82.65	81.85	51.90	61.84	45.11	78.68	74.28	67.37
	DSIMVC	27.24	28.68	13.38	21.53	25.29	8.34	94.40	89.37	90.40	49.89	49.61	32.35	83.93	76.72	73.95
	DCP	38.14	40.97	22.07	21.52	26.32	11.11	94.10	93.30	87.40	38.87	50.15	31.18	85.37	77.97	74.63
	CPSPAN	41.12	41.30	22.85	20.43	26.99	8.05	96.13	90.43	91.60	60.07	64.18	46.42	46.88	49.03	33.58
	ICMVC	37.34	38.59	21.06	24.38	27.75	11.12	16.44	8.67	3.64	41.70	37.78	21.66	85.37	**81.42**	78.31
	RPCIC	27.33	32.74	14.66	19.24	22.49	5.83	83.58	73.18	66.95	59.43	60.98	44.52	51.63	47.77	35.75
	MCMVC	37.51	40.03	18.18	21.01	26.82	10.07	87.00	78.19	77.34	52.56	60.04	42.50	73.24	72.48	59.51
	MASK-IMVC	25.54	30.10	13.86	16.26	19.15	4.95	55.59	51.43	36.65	68.46	68.39	52.74	22.08	18.95	9.96
	DCG	39.29	42.97	24.89	25.76	29.07	13.17	93.85	89.23	91.02	63.67	63.47	49.37	90.36	79.41	80.13
	DGIMVCM	40.54	41.16	23.16	27.08	31.56	11.98	73.91	70.18	62.73	42.83	48.85	30.04	51.07	49.90	36.99
	DPD(Ours)	**46.53**	**43.91**	28.34	**29.66**	**32.16**	**15.40**	**97.71**	**93.68**	**94.97**	**74.83**	**70.39**	**59.59**	**90.88**	80.89	**81.32**
0.7	COMP	36.10	36.50	**33.23**	24.25	29.11	10.43	87.45	73.86	74.29	19.43	17.37	0.73	68.72	65.59	55.88
	DSIMVC	28.42	29.09	13.85	20.87	23.69	7.76	91.05	85.67	84.67	36.78	40.30	23.72	58.10	55.50	44.30
	DCP	38.54	40.53	22.43	22.18	27.00	10.13	75.32	86.98	80.95	38.20	47.97	29.18	68.91	**69.99**	53.59
	CPSPAN	37.48	40.07	21.90	23.52	26.33	10.31	91.33	**88.48**	82.95	58.60	62.16	45.23	53.46	50.14	37.58
	ICMVC	30.49	33.24	16.27	22.97	25.15	9.65	16.53	8.94	3.67	39.30	37.08	21.06	64.78	65.73	54.73
	RPCIC	27.38	32.03	15.47	17.71	18.73	4.45	49.63	42.52	31.86	58.53	61.14	45.14	47.24	43.19	30.47
	MCMVC	36.42	39.85	21.97	23.11	27.99	10.86	81.97	69.46	67.55	38.11	38.25	11.61	70.20	62.18	48.07
	MASK-IMVC	33.25	28.37	11.43	14.19	14.98	3.10	57.53	51.82	38.46	34.29	39.53	16.69	21.36	16.77	6.83
	DCG	39.29	40.53	23.46	25.14	7.23	11.85	87.47	80.31	81.92	61.67	62.49	46.64	76.74	65.88	60.98
	DGIMVCM	23.62	35.09	17.35	23.75	28.21	9.77	47.21	52.24	34.36	41.33	48.05	27.96	39.23	41.34	23.95
	DPD(Ours)	**39.93**	**41.68**	25.61	**27.95**	**29.37**	**12.27**	**94.08**	86.05	**87.30**	**62.21**	**62.77**	**48.19**	**80.16**	69.53	**65.52**

Table 1: Clustering performance comparison on five benchmark datasets under different missing rates (MR). The best results are in **bold**, and the second-best results are underlined.

3.1.1 DATASETS AND EVALUATION METRICS

To comprehensively evaluate the proposed DPD, experiments are conducted on five representative multi-view clustering benchmarks: CUB, LandUse-21, Scene-15, Hdigit, and NoisyMNIST. These datasets cover diverse application scenarios, including fine-grained image recognition, remote sensing imagery, scene understanding, and handwritten digit analysis. Following the standard setting for incomplete multi-view clustering, we randomly discard partial views to construct incomplete datasets with missing rates of 0.3, 0.5, and 0.7. All competing methods are evaluated under identical experimental settings for a fair comparison. Clustering performance is assessed using three widely adopted metrics: ACC, NMI, and ARI, where higher values indicate better performance.

3.1.2 COMPARED METHODS

We compare the proposed DPD with ten representative incomplete multi-view clustering methods, including COMP(Lin et al., 2021), DSIMVC (Tang & Liu, 2022) , DCP (Lin et al., 2022), CPSPAN (Jin et al., 2023), ICMVC (Chao et al., 2024), RPCIC (Yuan et al., 2024), MCMVC (Geng et al., 2024), MASK-IMVC (Li et al., 2025), DCG (Zhang et al., 2025), and DGIMVCM (Zhang et al., 2026). These methods cover several mainstream IMVC paradigms, such as incomplete-view completion, contrastive representation learning, prototype-based alignment, masked modeling, and generative recovery. Among them, COMP and DSIMVC are representative incomplete multi-view representation learning methods; DCP, ICMVC, RPCIC and MCMVC mainly focus on cross-view consistency and contrastive learning; CPSPAN introduces prototype and structure-aware alignment; MASK-IMVC explores masked view modeling; DCG and DGIMVCM are recent generative or graph-enhanced IMVC methods. By comparing DPD with these representative baselines, we evaluate its effectiveness in semantic alignment, robustness to missing views, and discriminative representation learning. For a fair comparison, all competing methods are reproduced based on their original papers or official implementations, with hyperparameters tuned within the recommended ranges. Unless otherwise specified, the proposed DPD uses the same experimental settings across all datasets and missing rates.

3.1.3 IMPLEMENTATION DETAILS

All experiments are implemented using PyTorch 2.3.1 on an NVIDIA RTX 3050Ti GPU. For each dataset, all methods are evaluated under three missing rates (0.3, 0.5, and 0.7). The same data partition and missing-view generation protocol are adopted for all methods to ensure fair comparison. Each experiment is repeated five times using independently generated missing-view masks, while the model initialization and training random seed are fixed within each dataset. The average results over the five runs are reported.

3.2 EXPERIMENT RESULTS AND ANALYSIS

3.2.1 PERFORMANCE COMPARISON

Table 1 reports the clustering performance of DPD and ten representative incomplete multi-view clustering methods on five benchmark datasets under missing rates of 0.3, 0.5, and 0.7. Overall, DPD achieves the best result in 35 out of the 45 dataset–metric combinations, demonstrating consistently strong performance across different data distributions and degrees of view incompleteness.

At MR=0.3, DPD achieves the best results on all three metrics on Scene-15, Hdigit, and CUB, and obtains the highest ACC and ARI on LandUse-21. Its advantage becomes particularly evident at the moderate missing rate of MR=0.5, where DPD achieves the best ACC, NMI, and ARI on all five datasets. For example, DPD obtains 46.53% ACC on Scene-15, compared with 41.12% for the strongest competing result, corresponding to an improvement of 5.41 percentage points. On Hdigit, DPD reaches 97.71% ACC, exceeding the best baseline result of 96.13% by 1.58 percentage points.

DPD also remains competitive under the more challenging setting of MR=0.7. It achieves the best result in 12 out of the 15 dataset–metric combinations, including all three metrics on Scene-15, LandUse-21, and CUB. On Hdigit and NoisyMNIST, DPD obtains the highest ACC and ARI, while its NMI remains close to the strongest competing methods. These results indicate that DPD maintains effective clustering representations as the amount of available cross-view information de-

creases. We further investigate the contribution of individual components through the ablation study below.

Reconstruct	Diffusion	InfoNCE	Proto-MMD	ACC	NMI	ARI
✓	✓	✓		32.95	20.39	12.18
✓		✓	✓	95.78	90.39	90.88
✓	✓		✓	91.82	88.07	84.84
	✓	✓	✓	88.32	91.79	86.11
✓	✓	✓	✓	**97.71**	**93.68**	**94.97**

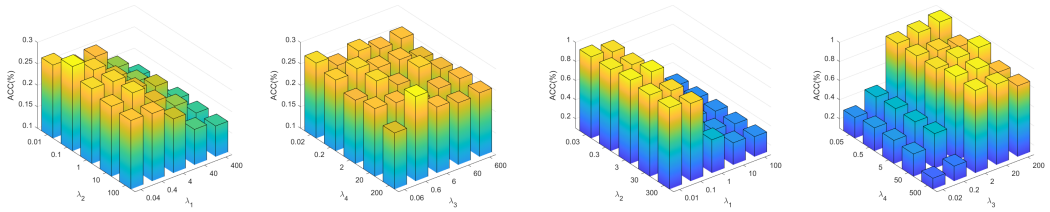
Table 2: Ablation study on Hdigit at MR=0.5.

3.2.2 Ablation Study

We conduct a leave-one-component-out ablation study on Hdigit at MR=0.5 to evaluate the contribution of each training objective. As shown in Table 2, the complete model achieves 97.71 ACC, 93.68 NMI, and 94.97 ARI.

Removing prototype-guided MMD alignment causes the largest degradation, reducing ACC to 32.95, which highlights the importance of shared prototypes in preserving global semantic structure. Removing reconstruction or paired-sample InfoNCE reduces ACC to 88.32 and 91.82, respectively, confirming the complementary roles of view-specific information preservation and instance-level correspondence. Without latent diffusion regularization, ACC decreases to 95.78, indicating that the denoising objective provides an additional refinement of the prototype-aligned latent space. Overall, these results verify the complementary contributions of the four objectives.

3.3 Model Analysis



(a) Parameter sensitivity analysis on the LandUse-21. (b) Parameter sensitivity analysis on the Hdigit.

Figure 3: Parameter sensitivity analysis of DPD with respect to different hyperparameter settings. The left two subplots correspond to the LandUse-21 dataset, and the right two subplots correspond to the Hdigit dataset.

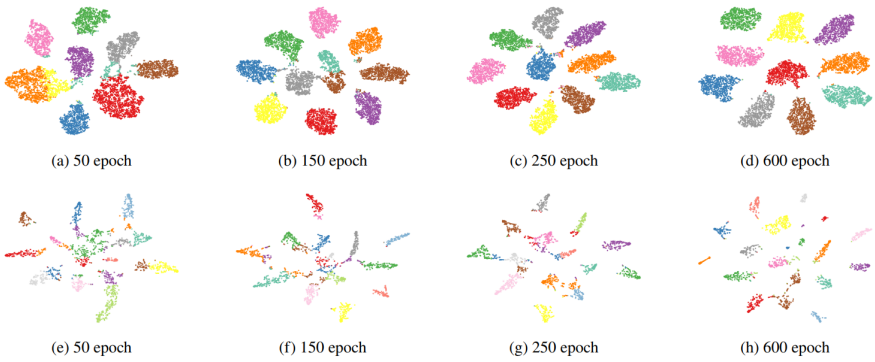
3.3.1 Parameter Analysis

We analyze the sensitivity of the four trade-off hyperparameters in Eq. (23), where λ_1, λ_2, λ_3, and λ_4 correspond to the reconstruction loss $\mathcal{L}_{\text{rec}}$, latent diffusion regularization loss $\mathcal{L}_{\text{diff}}$, paired-sample InfoNCE loss $\mathcal{L}_{\text{nce}}$, and prototype-guided MMD alignment loss $\mathcal{L}_{\text{pga}}$, respectively. Joint parameter sensitivity experiments are conducted on the LandUse-21 and Hdigit datasets, and the corresponding ACC results are shown in Figure 3. For visualization, the first and third subplots jointly vary λ_1 and λ_2, whereas the second and fourth subplots vary λ_3 and λ_4.

Overall, both datasets exhibit relatively stable performance around the selected hyperparameter configurations. Performance degradation can be observed when individual loss weights deviate substantially from suitable ranges, indicating that an appropriate balance among reconstruction, diffusion regularization, instance-level contrastive alignment, and prototype-level distribution alignment is important for optimization.

3.3.2 VISUALIZATION

Figure 4 presents t-SNE visualizations of the learned representations on Hdigit and LandUse-21 at different training epochs. As optimization proceeds, the projected representations exhibit increasingly separated cluster structures, particularly on Hdigit. These qualitative observations are consistent with the improvement of the quantitative clustering metrics, although the two-dimensional visualization itself is not used as quantitative evidence.



(a) 50 epoch (b) 150 epoch (c) 250 epoch (d) 600 epoch

(e) 50 epoch (f) 150 epoch (g) 250 epoch (h) 600 epoch

Figure 4: t-SNE visualizations of clustering results on the Hdigit and LandUse-21 datasets with increasing training epochs. (a)-(d) show the results on the Hdigit dataset. (e)-(h) show the results on the LandUse-21 dataset.

3.3.3 CONVERGENCE ANALYSIS

Figure 5 shows the training loss together with ACC, NMI, and ARI on LandUse-21 and Scene-15 at MR=0.5. During the early stage of training, the clustering metrics improve rapidly while the overall objective decreases substantially. The curves subsequently exhibit smaller variations and gradually stabilize as training proceeds. Although local fluctuations are observable, particularly on LandUse-21, the overall optimization remains stable and the clustering metrics reach a relatively steady regime at later epochs.



(a) LandUse-21 Dataset (b) Scene-15 Dataset

Figure 5: Clustering metrics (ACC, NMI, ARI) and loss curves over training epochs on the LandUse-21 and Scene-15 datasets.

4 CONCLUSION

In this paper, we proposed Discriminative Prototype Diffusion (DPD) for incomplete multi-view clustering. DPD combines paired-sample contrastive alignment with prototype-guided distribution alignment, where learnable shared prototypes provide global semantic anchors for view-specific representations. A lightweight latent diffusion objective further regularizes the prototype-aligned latent space, while mask-aware prototypical spherical fusion integrates the available views into a unified representation for clustering. Experiments on five benchmark datasets under different missing rates demonstrate the effectiveness and robustness of DPD across diverse incomplete multi-view settings. Future work will explore extending the framework to more general multi-view scenarios.

5 Ethics Statement

This work does not involve human subjects, animal experiments, or the collection of personally identifiable information. All datasets used in this study are publicly available benchmark datasets for incomplete multi-view clustering research. We follow the original usage guidelines of these datasets and use them solely for academic evaluation of machine learning algorithms. Since our study focuses on representation learning and clustering methodology, it does not involve decision-making applications that may directly affect individuals or social groups. We have made efforts to ensure that the proposed method is developed and evaluated in a responsible and transparent manner.

6 AI Usage Statement

Large Language Models (LLMs) were used solely to assist with the writing and polishing of this manuscript. Specifically, LLMs were utilized for improving language clarity, refining sentence structure, and assisting with the organization of textual descriptions. In addition, LLMs were used to aid retrieval and discovery of related literature during manuscript preparation.

LLMs were not involved in the development of the research ideas, formulation of the proposed method, design of experiments, implementation of algorithms, analysis of experimental results, or generation of scientific conclusions. All technical content, experimental findings, and claims presented in this paper were developed and verified by the authors, who take full responsibility for the accuracy and integrity of the manuscript.

References

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

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

Xiaojian Ding, Lin Zhao, Xian Li, and Xiaoying Zhu. Incomplete multi-view clustering via hierarchical semantic alignment and cooperative completion. *Advances in Neural Information Processing Systems*, 38:52175–52195, 2026.

Zhibin Dong, Dayu Hu, Jiaqi Jin, Siwei Wang, Xinwang Liu, and En Zhu. Selective cross-view topology for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 2025.

Chuanxing Geng, Aiyang Han, and Songcan Chen. Explicit view-labels matter: A multifacet complementarity study of multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4):2520–2532, 2024.

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.

Jiaqi Jin, Siwei Wang, Zhibin Dong, Xihong Yang, Xinwang Liu, En Zhu, and Kunlun He. Deep incomplete multi-view clustering with distribution dual-consistency recovery guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1016–1026, 2025.

Yu Lei, Zuoyuan Niu, Qianqian Wang, Quanxue Gao, and Ming Yang. Anchor graph-based multi-view spectral clustering. *Neurocomputing*, 583:127579, 2024.

Zhenglai Li, Yuqi Shi, Xiao He, and Chang Tang. Mask-informed deep contrastive incomplete multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(12): 11822–11833, 2025. doi: 10.1109/TCSVT.2025.3585600.

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

Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4447–4461, 2022.

Cheng Liu, Rui Li, Hangjun Che, Man-Fai Leung, Si Wu, Zhiwen Yu, and Hau-San Wong. Latent structure-aware view recovery for incomplete multi-view clustering. *IEEE transactions on knowledge and data engineering*, 36(12):8655–8669, 2024a.

Suyuan Liu, Siwei Wang, Ke Liang, Junpu Zhang, Zhibin Dong, Tianrui Liu, En Zhu, Kunlun He, and Xinwang Liu. Alleviate anchor-shift: Explore blind spots with cross-view reconstruction for incomplete multi-view clustering. *Advances in Neural Information Processing Systems*, 37: 87509–87531, 2024b.

Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International conference on machine learning*, pp. 21090–21110. PMLR, 2022.

Huibing Wang, Mingze Yao, Yawei Chen, Yunqiu Xu, Haipeng Liu, Wei Jia, Xianping Fu, and Yang Wang. Manifold-based incomplete multi-view clustering via bi-consistency guidance. *IEEE Transactions on Multimedia*, 26:10001–10014, 2024.

Huibing Wang, Luyan Cui, Mingze Yao, and Xianping Fu. Tensorized fine-grained incomplete multiview clustering via intrinsic structure recovery. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2026.

Jie Wen, Ke Yan, Zheng Zhang, Yong Xu, and Bob Zhang. Adaptive graph completion based incomplete multi-view clustering. *IEEE Transactions on Multimedia*, PP(99):1–1, 2020.

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

Mingze Yao, Huibing Wang, Yawei Chen, and Xianping Fu. Between/within view information completing for tensorial incomplete multi-view clustering. *IEEE transactions on multimedia*, 27: 1538–1550, 2024.

Jun Yin, Pei Wang, Shiliang Sun, and Zhonglong Zheng. Incomplete multi-view clustering via multi-level contrastive learning. *IEEE Transactions on Knowledge and Data Engineering*, 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*, 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.

Zhenghao Zhang, Jun Xie, Xingchen Chen, Tao Yu, Hongzhu Yi, Kaixin Xu, Yuanxiang Wang, Tianyu Zong, Xinming Wang, Jiahuan Chen, et al. Dynamic deep graph learning for incomplete multi-view clustering with masked graph reconstruction loss. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 28600–28608, 2026.

Rui Zheng, Zhuojie Huang, Shuping Zhao, and Jigang Wu. Incomplete multi-view clustering via reconstructing views and structure unification. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6161–6165. IEEE, 2026.

Deyin Zhuang, Jian Dai, Xingfeng Li, Xi Wu, Yuan Sun, and Zhenwen Ren. Robust graph contrastive learning for incomplete multi-view clustering. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 7282–7290, 2025.