GRADE: EVIDENCE-ALIGNED ADAPTIVE GRANULARITY FOR STRUCTURE-AWARE RESIDUAL DIFFUSION IN INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete multi-view clustering aims to discover a consistent clustering structure from partially observed multi-view data. However, recovering missing representations still faces two key challenges: (i) relying solely on shared semantics from the observed views can preserve cross-view consistency but cannot fully characterize the local structure specific to the target view; and (ii) view heterogeneity makes it difficult to establish reliable semantic correspondences between local structures across views, hindering the acquisition of query-relevant local priors from the target view. To address these challenges, we propose GRADE. First, we construct local regions for each view at adaptive granularities and use jointly observed members as evidence to establish soft correspondences across views. This extends sparse pairwise evidence to adaptive local regions and enables the retrieval of query-relevant local priors from the target view. Second, we map the shared semantics of the observed views into a target-view base representation. Rather than directly generating the entire missing representation, GRADE uses the retrieved local priors to guide residual diffusion, generating only the target-view-specific local information not captured by the base representation. Experiments on eight multi-view benchmark datasets across multiple missing rates demonstrate the superior overall clustering performance of GRADE.

1 INTRODUCTION

Incomplete multi-view clustering (IMVC) aims to discover a reliable clustering structure from partially observed multi-view data (Wen et al., 2022). In real-world applications, sensor failures (Jin et al., 2023), data acquisition costs (Wu et al., 2024), transmission disruptions (Chao et al., 2024), and privacy restrictions (Zhao et al., 2025) may render one or more views of some instances unavailable. Such view incompleteness not only leads to information loss but also disrupts cross-view instance correspondences, hindering the exploitation of cross-view consistency and complementarity.

To address the aforementioned problems caused by view incompleteness, existing incomplete multi-view clustering methods can generally be divided into imputation-free and imputation-based approaches. Imputation-free methods bypass missing-view recovery and learn clustering structures directly from observed data, for example, by learning shared representations (Chen et al., 2026) or mining cluster-level complementary information (Xu et al., 2022) to enhance clustering consistency. In contrast, imputation-based methods recover missing representations to construct complete multi-view representations. Representative strategies include cross-view prediction (Lin et al., 2021), shared latent representation projection (Liu et al., 2025), sample–prototype relation transfer (Yuan et al., 2025), and hierarchical semantic alignment (Ding et al., 2025). Despite their effectiveness, existing imputation-based methods still rely primarily on cross-view shared semantics to infer missing representations, leaving target-view local structure underexploited. As illustrated in Fig. 1, without such structural guidance, recovered representations may deviate from query-relevant local regions and fail to preserve their local distributions, making the joint preservation of cross-view semantic consistency and target-view local structure a key challenge in IMVC.

Addressing this challenge requires establishing reliable local correspondences between heterogeneous observed views and the target view. Existing approaches to cross-view relation learning generally

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.

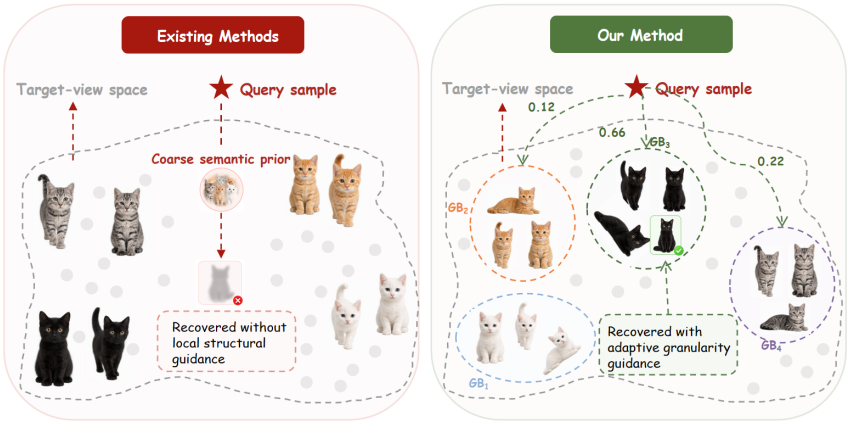


Figure 1: Conceptual comparison of missing-representation recovery. Existing methods (left) mainly rely on cross-view shared semantics, whereas GRADE (right) leverages target-view local priors at adaptive granularity to better preserve local structure.

operate at either the instance or cluster level. At the instance level, generated representations can be aligned with real ones using samples jointly observed in two views (Zhang et al., 2025). However, such alignment evidence becomes increasingly sparse as the missing rate rises. At the cluster level, methods learn cross-view prototype correspondences (Yuan et al., 2024) or consensus prototypes (Dai et al., 2025) to reduce their reliance on instance-level matches. Nevertheless, their fixed granularity at the cluster or prototype level cannot accommodate structural variations across views and data regions, and may compress diverse local structures within a cluster into uniform relations. Therefore, an adaptive granularity alignment mechanism is needed to extend limited pairwise evidence to local regions and retrieve query-relevant local structural priors from the target view for missing-representation recovery.

To this end, we propose GRADE, whose framework is illustrated in Figure 2. It comprises two core modules: evidence-aligned adaptive granularity learning and structure-aware residual diffusion based on denoising diffusion models (Ho et al., 2020). The former constructs local regions at adaptive granularity in each view and establishes soft cross-view correspondences based on the membership overlap of jointly observed samples. In this way, it extends sparse pairwise evidence to local regions and retrieves query-relevant local priors from the target view. The latter first maps the shared query to a base representation carrying semantics shared across views, and then uses the retrieved local priors to condition residual diffusion, which generates only the target-view local structural information not captured by the base representation. Through this decomposition, GRADE recovers missing representations that preserve both cross-view semantic consistency and target-view local structure.

The main contributions of this work are summarized as follows:

- We develop adaptive granularity learning through evidence alignment to turn sparse pairwise evidence into correspondences between local regions and retrieve local priors for each query from the target view, addressing limited instance coverage and fixed cluster granularity.
- We introduce structure-aware residual diffusion, which decomposes missing-representation recovery into a cross-view shared semantic base and a target-view local structural residual. This design avoids directly generating the entire missing representation while preserving both cross-view semantic consistency and target-view local structure.
- Extensive experiments on eight multi-view benchmark datasets under five missing ratios demonstrate the superior overall performance of GRADE against eight representative IMVC methods, while ablation and parameter analyses further validate the proposed design.

2 Related Works

Adaptive Granularity adjusts partition scales according to the distributions and structures of different data regions, enabling local structures to be characterized more flexibly than with a fixed granularity.

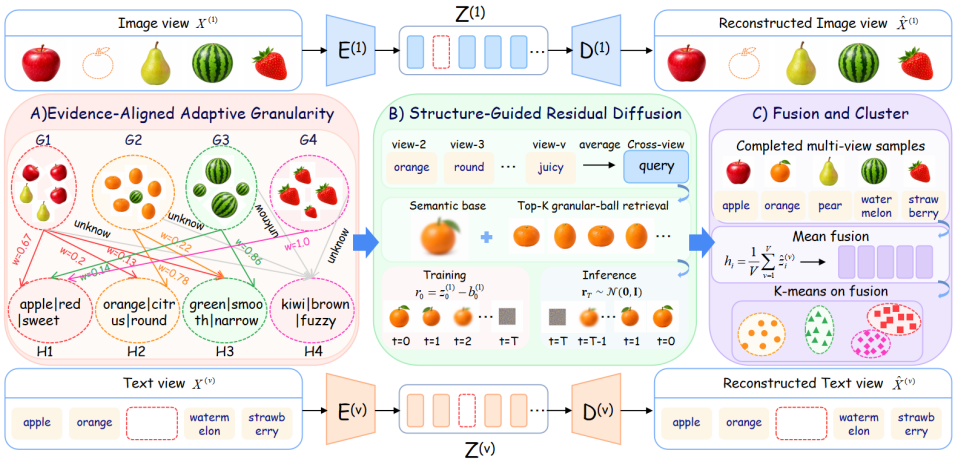


Figure 2: Overall framework of GRADE

Granular-ball computing (Xia et al., 2019) provides an effective means of achieving adaptive granularity by partitioning a complex sample space into compact local regions at multiple granularities. It has been applied to adaptive clustering (Xia et al., 2020), multi-granularity neighborhood relation modeling (Xie et al., 2024), and graph clustering (Xie et al., 2025). More recently, granular-ball computing has been extended to multiview learning. MGBCC (Su et al., 2025) constructs binary associations from geometric overlaps within each view and thresholded member intersections across views, and uses these associations for contrastive learning at multiple granularities. GFSAF (Liu et al., 2026) performs granular-ball-based fuzzy contrastive learning using paired samples across views and fuses mutual and complementary representations. However, the use of adaptive granularity in incomplete multi-view clustering remains underexplored. GRADE adaptively constructs local regions according to the data distribution of each view and uses jointly observed samples as evidence to establish soft cross-view correspondences between regions, thereby extending sparse evidence from the sample level to the region level and obtaining local structural priors from the target view.

Diffusion Models for Missing Representation Recovery progressively add Gaussian noise to data through a forward process and generate samples by learning a reverse denoising process. DCG (Zhang et al., 2025) introduces diffusion models into missing-representation recovery for incomplete multi-view clustering by modeling within-view distributions and incorporating cross-view contrastive learning. Subsequent conditional diffusion methods further exploit neighborhood structures from observed views as recovery guidance: IMVC-NCD (Guo et al., 2026) aggregates nearest neighbor representations into a unified neighborhood condition, whereas SACDG (Zhang et al., 2026) fuses adjacency structures derived from adaptive neighborhood graphs through cross-attention. However, these conditions are constructed primarily by aggregating or fusing local structures from observed views, without sufficiently accounting for local distribution discrepancies across heterogeneous views. Consequently, they may fail to locate the target-view local region corresponding to a query. In contrast, GRADE retrieves a query-relevant local prior from adaptive granularity regions in the target view and uses it to guide residual diffusion, thereby supplementing the shared semantic base representation with the target-view local structure that it does not capture.

3 Preliminaries

This section first formulates the incomplete multi-view clustering problem and then introduces granular-ball computing (GBC) for obtaining adaptive granularity structures. The main notation is summarized in Appendix A.

Problem Formulation. Given a V-view dataset $\mathcal{X} = \{\mathbf{X}^{(v)} \in \mathbb{R}^{N \times D_v}\}_{v=1}^{V}$ with N instances, let $\mathbf{M} \in \{0,1\}^{N \times V}$ be the observation mask, where $M_{iv} = 1$ if instance i is observed in view v, and 0 otherwise. The observed subset of view v is $\widetilde{\mathbf{X}}^{(v)} = \{\mathbf{x}_i^{(v)} \mid M_{iv} = 1\}$, where $\mathbf{x}_i^{(v)} \in \mathbb{R}^{D_v}$ denotes the representation of the i-th instance in view v. The goal is to partition all instances into K clusters.

Formally, in the incomplete multi-view setting considered in this work, GBC partitions the observed subset $\widetilde{\mathbf{X}}^{(v)}$ of view v into a set of view-specific granular balls $GB^{(v)} = \{GB_g^{(v)}\}_{g=1}^{G_v}$, where G_v denotes the number of granular balls in view v and each granular-ball is defined in Definition 1.

Definition 1 (**Granular-Ball**). *For view v, the granular-ball $GB_g^{(v)} = \{\widetilde{\mathbf{x}}_i^{(v)} \mid i \in I_g^{(v)}\}$ consists of the observed samples indexed by $I_g^{(v)}$. Its center and radius are defined as:*

$$\mathbf{c}_g^{(v)} = \frac{1}{|I_g^{(v)}|} \sum_{i \in I_g^{(v)}} \widetilde{\mathbf{x}}_i^{(v)}, \qquad r_g^{(v)} = \max_{i \in I_g^{(v)}} \left\| \widetilde{\mathbf{x}}_i^{(v)} - \mathbf{c}_g^{(v)} \right\|_2 . \tag{1}$$

GBC recursively divides each granular ball into two candidate children, using the sample farthest from its center and the sample farthest from that point as anchors. The candidate split is evaluated using the distribution measure in Definition 2.

Definition 2 (**Distribution Measure**). *For a granular-ball $GB_g^{(v)}$, its distribution measure is defined as the average distance between its samples and center:*

$$\mathrm{DM}\Big(GB_g^{(v)}\Big) = \frac{1}{|I_g^{(v)}|} \sum_{i \in I_g^{(v)}} \left\| \widetilde{\mathbf{x}}_i^{(v)} - \mathbf{c}_g^{(v)} \right\|_2 . \tag{2}$$

A smaller DM indicates a more compact distribution within the granular-ball. Suppose that $GB_g^{(v)}$ is divided into two candidate child balls $GB_{g,1}^{(v)}$ and $GB_{g,2}^{(v)}$. The weighted distribution measure after splitting is defined as:

$$\mathrm{DM_w}\Big(GB_g^{(v)}\Big) = \frac{|I_{g,1}^{(v)}|}{|I_g^{(v)}|} \mathrm{DM}\Big(GB_{g,1}^{(v)}\Big) + \frac{|I_{g,2}^{(v)}|}{|I_g^{(v)}|} \mathrm{DM}\Big(GB_{g,2}^{(v)}\Big) . \tag{3}$$

The candidate split is accepted if $\mathrm{DM_w}(GB_g^{(v)}) < \mathrm{DM}(GB_g^{(v)})$; otherwise, the original granular-ball is retained. Appendix B analyzes the compactness induced by this criterion, and Appendix C provides the complete algorithm. The recursion terminates when no valid split remains, yielding an adaptive granularity structure composed of compact regions that adapt to local data distributions.

4 METHODOLOGY

In this section, we present GRADE, an IMVC method that combines evidence-aligned adaptive granularity with structure-guided residual diffusion. As illustrated in Figure 2, GRADE comprises view-specific representation learning, adaptive local correspondence modeling, and target-view structure-guided residual diffusion for missing representation recovery and clustering.

4.1 WITHIN-VIEW AUTOENCODING

To account for the heterogeneous feature spaces and data distributions across views, we employ an independent encoder–decoder pair $\{\mathcal{E}_v, \mathcal{D}_v\}_{v=1}^{V}$ for each view to learn compact and clustering-friendly representations. For the v-th view, the encoder $\mathcal{E}_v$ maps the i-th observed instance $\widetilde{\mathbf{x}}_i^{(v)}$ to a d-dimensional latent representation $\widetilde{\mathbf{z}}_i^{(v)} = \mathcal{E}_v(\widetilde{\mathbf{x}}_i^{(v)})$, while the decoder $\mathcal{D}_v$ reconstructs the input features from this representation. With $N_v = \sum_{i=1}^{N} M_{iv}$ denoting the number of observed instances in the v-th view, the view-specific autoencoders are optimized by minimizing the following reconstruction loss:

$$\mathcal{L}_{\mathrm{rec}} = \sum_{v=1}^{V} \sum_{i=1}^{N_v} \left\| \widetilde{\mathbf{x}}_i^{(v)} - \mathcal{D}_v\Big(\mathcal{E}_v\Big(\widetilde{\mathbf{x}}_i^{(v)}\Big)\Big) \right\|_2^2 . \tag{4}$$

The latent representations learned by the view specific encoders are further aligned across views by the evidence-aligned adaptive granularity module in Section 4.2. Accordingly, all subsequent cross-view aggregation and similarity computations are performed in the aligned latent space.

4.2 EVIDENCE-ALIGNED ADAPTIVE GRANULARITY

Based on the observed latent representations $\widetilde{\mathbf{Z}}^{(v)}$ obtained in the preceding section, we construct a granular-ball set $GB^{(v)}$ for each view according to Definitions 1 and 2. Since granular balls are independently formed in each view without explicit correspondences, we use co-observed members as evidence to establish soft cross-view correspondences. For an arbitrary pair of views u and v, let $I_g^{(u)}$ and $I_h^{(v)}$ denote the member index sets of $GB_g^{(u)}$ and $GB_h^{(v)}$, respectively. Since each instance shares the same global index across views, member intersections provide direct correspondence evidence:

$$O_{gh}^{(u,v)} = \left| I_g^{(u)} \cap I_h^{(v)} \right|, \qquad n_g^{u\to v} = \sum_h O_{gh}^{(u,v)}, \qquad n_h^{v\to u} = \sum_g O_{gh}^{(u,v)}. \tag{5}$$

Here, $O_{gh}^{(u,v)}$ denotes the number of members shared by the two granular balls, while $n_g^{u\to v}$ and $n_h^{v\to u}$ count the members of each ball with corresponding observations in the other view, respectively. To avoid misinterpreting the absence of evidence as semantic inconsistency, we define an evidence-validity mask:

$$A_{gh}^{(u,v)} = \mathbb{I}\big(n_g^{u\to v} > 0 \wedge n_h^{v\to u} > 0\big). \tag{6}$$

A pair with $O_{gh}^{(u,v)} > 0$ is regarded as a soft positive correspondence. A pair with $O_{gh}^{(u,v)} = 0$ and $A_{gh}^{(u,v)} = 1$ is treated as a valid negative correspondence, whereas a pair with $A_{gh}^{(u,v)} = 0$ is regarded as unknown and excluded from alignment. We further normalize the intersection evidence in both directions to obtain evidence-driven soft correspondence targets:

$$P_{gh}^{u\to v} = \frac{O_{gh}^{(u,v)}}{n_g^{u\to v}}, \qquad P_{hg}^{v\to u} = \frac{O_{gh}^{(u,v)}}{n_h^{v\to u}}. \tag{7}$$

The resulting soft targets permit one-to-many correspondences between local regions and aggregate sparse instance-level evidence into region-level supervision. We employ a bidirectional masked soft contrastive objective to match the similarity-induced prediction with the evidence-derived targets P. Let $\widehat{\mathbf{c}}_g^{(u)}$ and $\widehat{\mathbf{c}}_h^{(v)}$ denote the normalized latent representations obtained by averaging the member representations of the corresponding local regions. Their similarity and the predicted correspondence distribution are defined as

$$S_{gh}^{(u,v)} = \widehat{\mathbf{c}}_g^{(u)\top} \widehat{\mathbf{c}}_h^{(v)}, \qquad Q_{gh}^{u\to v} = \frac{\exp\left(S_{gh}^{(u,v)}/\tau\right)}{\sum_{l:A_{gl}^{(u,v)}=1} \exp\left(S_{gl}^{(u,v)}/\tau\right)}, \tag{8}$$

where τ is a temperature parameter. Valid negative correspondences are retained in the softmax denominator, whereas unknown relations are masked out. To reduce the contribution of soft targets supported by insufficient co-observation evidence, we define the evidence coverage

$$\rho_g^{u\to v} = \frac{n_g^{u\to v}}{|I_g^{(u)}|}, \qquad \rho_h^{v\to u} = \frac{n_h^{v\to u}}{|I_h^{(v)}|}. \tag{9}$$

Based on this coverage, the forward alignment loss is defined as

$$\mathcal{L}_{u\to v} = -\frac{\sum_{g:n_g^{u\to v}>0} \rho_g^{u\to v} \sum_{h:A_{gh}^{(u,v)}=1} P_{gh}^{u\to v} \log Q_{gh}^{u\to v}}{\sum_{g:n_g^{u\to v}>0} \rho_g^{u\to v}}. \tag{10}$$

The reverse loss $\mathcal{L}_{v\to u}$ is computed analogously. Let $\omega_{uv} = \sum_{g,h} O_{gh}^{(u,v)}$ denote the total intersection evidence between the two views. The final adaptive granularity alignment loss is

$$\mathcal{L}_{\mathrm{GB}} = \frac{\sum_{u<v} \omega_{uv} \left(\mathcal{L}_{u\to v} + \mathcal{L}_{v\to u}\right)/2}{\sum_{u<v} \omega_{uv}}. \tag{11}$$

This objective directly aligns the local structures of different views in the latent space, making their representations comparable for subsequent cross-view aggregation and target view region retrieval.

4.3 STRUCTURE-AWARE RESIDUAL DIFFUSION

Building on the local structures obtained through evidence-aligned adaptive granularity modeling, we develop a structure-aware residual diffusion model to recover missing representations while preserving target-view-specific local structure. The completion process consists of training and inference stages, as described below.

Training Stage. During training, an observed view v is treated as the target view, while the remaining observed views $\Omega_i^{-v} = \{u \neq v \mid M_{iu} = 1\}$ are used to simulate missing-view recovery. Only instances satisfying $M_{iv} = 1$ and $|\Omega_i^{-v}| \geq 1$ are used. Since the view representations are aligned in the latent space, the cross-view query is defined as

$$\mathbf{q}_{i\to v} = \frac{1}{|\Omega_i^{-v}|} \sum_{u\in\Omega_i^{-v}} \widetilde{\mathbf{z}}_i^{(u)}. \tag{12}$$

Based on the cross-view query, the target-view mapper B_v predicts a semantic base representation, while the unexplained target-view information is defined as a residual:

$$\mathbf{b}_i^{(v)} = B_v(\mathbf{q}_{i\to v}), \qquad \mathbf{r}_{0,i}^{(v)} = \widetilde{\mathbf{z}}_i^{(v)} - \mathbf{b}_i^{(v)}. \tag{13}$$

Consequently, the diffusion model only recovers the target-view local structural information not captured by the base representation, rather than generating the entire target representation. To incorporate target-view local structure, we retrieve the K_{gb} most relevant local region prototypes according to their cosine similarities with the cross-view query in the aligned latent space:

$$e_{ig}^{(v)} = \mathrm{sim}\big(\mathbf{q}_{i\to v}, \mathbf{c}_g^{(v)}\big), \quad \mathcal{G}_{i\to v} = \mathrm{TopK}_{K_{gb}}\big(\{e_{ig}^{(v)}\}_{g=1}^{G_v}\big), \quad \mathbf{C}_{i\to v} = \{\mathbf{c}_g^{(v)} \mid g \in \mathcal{G}_{i\to v}\}. \tag{14}$$

Here, $G_v = |GB^{(v)}|$ is the number of local regions in view v, sim denotes cosine similarity, and $\mathbf{C}_{i\to v}$ is the query-relevant target-view local structural condition. Let T denote the total number of diffusion steps and $\{\beta_t\}_{t=1}^T$ the noise variance schedule, with $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$. For a randomly sampled timestep t and Gaussian noise $\boldsymbol{\epsilon}$, the target residual is diffused as

$$\mathbf{r}_{t,i}^{(v)} = \sqrt{\bar{\alpha}_t}\, \mathbf{r}_{0,i}^{(v)} + \sqrt{1-\bar{\alpha}_t}\, \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{15}$$

We inject the local structural condition into the denoising process through multi-head cross-attention (MHA), using the sum of the noisy residual and timestep embedding as the query and the retrieved local region prototypes as the keys and values:

$$\mathbf{a}_{t,i\to v} = \mathrm{MHA}\left(\mathbf{r}_{t,i}^{(v)} + \mathbf{e}_t,\, \mathbf{C}_{i\to v},\, \mathbf{C}_{i\to v}\right), \tag{16}$$

Here, $\mathbf{e}_t$ is the timestep embedding and $\mathbf{a}_{t,i\to v}$ is the resulting target-view local structural condition. Conditioned on $\mathbf{a}_{t,i\to v}$, the target-view denoiser ϵ_{θ_v} predicts the noise and is optimized by

$$\mathcal{L}_{\mathrm{diff}} = \mathbb{E}_{i,v,t,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \epsilon_{\theta_v}\left(\mathbf{r}_{t,i}^{(v)}, t, \mathbf{a}_{t,i\to v}\right)\right\|_2^2\right]. \tag{17}$$

Inference Stage. For an instance with a missing target view v, we compute $\mathbf{q}_{i\to v}$, $\mathbf{b}_i^{(v)}$, and $\mathbf{C}_{i\to v}$ from all observed views following the training procedure. Since the semantic base already provides an initial estimate of the target-view representation, reverse diffusion is applied only to the residual, whose terminal state is initialized as

$$\mathbf{r}_{T,i}^{(v)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{18}$$

We then perform deterministic DDIM (Song et al., 2021) sampling over $t = T, \ldots, 1$ with $\eta = 0$. At each reverse timestep t, the local structural representation $\mathbf{a}_{t,i\to v}$ is recomputed through cross-attention using the current residual state and the fixed target-view local structural condition. The target-view denoiser predicts the noise in the current residual state as

$$\widehat{\boldsymbol{\epsilon}}_{t,i}^{(v)} = \epsilon_{\theta_v}\left(\mathbf{r}_{t,i}^{(v)}, t, \mathbf{a}_{t,i\to v}\right). \tag{19}$$

Based on the predicted noise, the clean target residual is first estimated by

$$\widehat{\mathbf{r}}_{0,i}^{(v,t)} = \frac{\mathbf{r}_{t,i}^{(v)} - \sqrt{1-\bar{\alpha}_t}\, \widehat{\boldsymbol{\epsilon}}_{t,i}^{(v)}}{\sqrt{\bar{\alpha}_t}}. \tag{20}$$

The residual state at the preceding timestep is then obtained through the deterministic DDIM update:

$$\mathbf{r}_{t-1,i}^{(v)} = \sqrt{\bar{\alpha}_{t-1}}\,\widehat{\mathbf{r}}_{0,i}^{(v,t)} + \sqrt{1-\bar{\alpha}_{t-1}}\,\widehat{\boldsymbol{\epsilon}}_{t,i}^{(v)}, \tag{21}$$

where $\bar{\alpha}_0 = 1$. After iterative denoising, the recovered residual is combined with the semantic base to obtain the completed target-view representation:

$$\widehat{\mathbf{z}}_i^{(v)} = \mathbf{b}_i^{(v)} + \widehat{\mathbf{r}}_{0,i}^{(v)}. \tag{22}$$

For observed views, we retain the original representations, i.e., $\widehat{\mathbf{z}}_i^{(v)} = \widetilde{\mathbf{z}}_i^{(v)}$ for $M_{iv} = 1$.

4.4 Overall Objective and Clustering

Combining the three loss components above, we define the overall training objective as

$$\mathcal{L}_{\text{overall}} = \mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{GB}} + \mathcal{L}_{\text{diff}}. \tag{23}$$

After inference, the completed representations are averaged across views in the aligned latent space, and K-means is then applied to the resulting unified representations to obtain the final clustering assignments. The complete procedure is summarized in Appendix D.

5 Experiments

To evaluate the effectiveness of the proposed method, we conduct extensive experiments on eight widely used multi-view benchmark datasets to answer the following three questions: **Q1:** Does the proposed method outperform existing state-of-the-art incomplete multi-view clustering methods? **Q2:** Does each component contribute to the overall performance? **Q3:** How do the key hyperparameters affect the clustering performance?

5.1 Experimental Setup

Datasets. We evaluate GRADE on eight widely used multi-view benchmark datasets: Synthetic3D (Kumar et al., 2011), EYaleB10 (Liu et al., 2010), COIL20MV (Nene, 1996), 100Leaves (Zheng et al., 2021), Mfeat (Nie et al., 2018), LandUse-21 (Yang & Newsam, 2010), Scene-15 (Fei-Fei & Perona, 2005), and Animal (Lampert et al., 2013). Detailed dataset statistics are provided in Appendix E.

Incomplete Multi-View Data Construction. We randomly select a proportion r of samples as incomplete and retain one randomly chosen view for each incomplete sample while treating all remaining views as missing. We set $r \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, where r denotes the proportion of incomplete samples.

Baselines. We compare GRADE with eight representative IMVC methods: DCP (Lin et al., 2022), PMIMC (Yuan et al., 2025), FreeCSL (Dai et al., 2025), GUITAR (Zhong et al., 2025), HSACC (Ding et al., 2025), IMDC-DIA (Liu et al., 2025), IMVC-NCD (Guo et al., 2026), and OPTION (Zhou & Gu, 2026). All methods are evaluated under identical incomplete-data settings using the hyperparameters recommended in their papers or released code; further details are provided in Appendix F.

Evaluation Metrics. We evaluate clustering performance using clustering accuracy (ACC) (Wu & Schölkopf, 2006), normalized mutual information (NMI) (Strehl & Ghosh, 2002), and adjusted Rand index (ARI) (Qian et al., 2012). Implementation details are provided in Appendix G.

5.2 Performance Comparison (Q1)

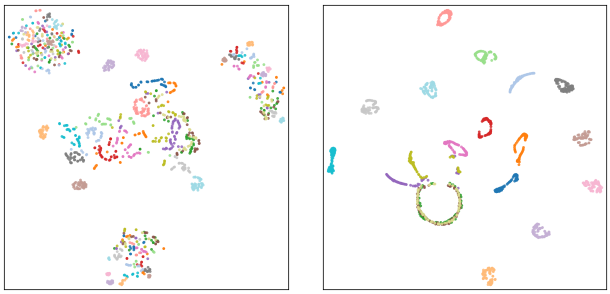
The clustering results of our method and eight competing methods under different missing ratios are reported in Table 1 and Appendix H. Overall, our method outperforms the competing methods on most datasets. Since GUITAR incurs prohibitively high computational costs on Animal and cannot be completed within a reasonable amount of time, its corresponding results are reported as N/A. Specifically, at a missing ratio of 0.3, our method improves ACC and ARI on 100Leaves by 21.08 and 21.70 percentage points, respectively, over the best competing results. At a missing ratio of 0.1, it improves NMI on EYaleB10 by 16.55 percentage points. For a more comprehensive comparison,

Table 1: Clustering performance (%, mean with standard deviation in parentheses) under different missing ratios. The best and second-best results are boldfaced and underlined, respectively.

MR Ratio	Method	Synthetic3D			EYaleB10			COIL20MV			100Leaves		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
10%	DCP	95.86 (0.99)	83.44 (3.16)	88.09 (2.73)	28.71 (1.69)	24.66 (1.20)	7.97 (0.58)	58.58 (2.88)	74.66 (1.15)	48.30 (2.96)	50.21 (1.99)	79.57 (0.52)	39.08 (1.05)
	PMIMC	93.57 (0.84)	75.82 (2.59)	81.67 (2.29)	20.34 (1.04)	10.66 (1.51)	3.41 (0.58)	57.99 (2.95)	71.04 (0.83)	51.70 (2.63)	76.95 (1.54)	88.02 (0.53)	66.84 (1.34)
	FreeCSL	87.37 (3.92)	64.46 (7.08)	67.81 (8.42)	28.69 (1.28)	26.96 (1.14)	12.73 (1.13)	75.67 (2.06)	83.37 (1.48)	71.74 (2.60)	84.26 (0.68)	92.02 (0.26)	76.89 (0.47)
	GUITAR	46.36 (17.64)	8.50 (11.27)	9.55 (13.16)	31.34 (1.80)	23.72 (2.01)	12.17 (1.29)	76.01 (1.33)	83.99 (1.00)	70.91 (1.49)	71.21 (1.50)	86.17 (0.67)	62.93 (1.79)
	HSACC	96.16 (0.55)	84.76 (1.86)	88.92 (1.53)	25.56 (4.40)	19.37 (5.59)	7.00 (2.18)	69.47 (2.77)	77.87 (2.14)	57.43 (4.34)	60.74 (2.82)	81.81 (1.09)	46.90 (3.49)
	IMDC-DIA	95.93 (0.73)	83.72 (2.41)	88.28 (2.01)	19.78 (1.11)	8.74 (1.66)	2.76 (0.80)	71.46 (1.34)	76.57 (1.07)	62.16 (1.79)	65.20 (1.47)	82.88 (0.45)	54.30 (1.26)
	IMVC-NCD	94.33 (0.61)	79.90 (1.39)	84.04 (1.58)	29.68 (1.19)	27.24 (2.57)	13.78 (2.18)	75.18 (2.42)	82.12 (1.30)	69.32 (2.44)	83.40 (1.15)	91.20 (0.47)	74.61 (1.38)
	OPTION	95.03 (0.61)	80.06 (1.47)	85.75 (1.62)	29.37 (1.91)	22.79 (2.05)	10.06 (0.74)	72.58 (1.08)	80.31 (1.27)	63.95 (1.90)	66.98 (1.90)	82.45 (1.45)	52.95 (2.92)
	Ours	**96.86** (0.70)	**86.52** (2.17)	**90.84** (1.97)	**42.81** (2.76)	**43.79** (4.34)	**28.19** (5.35)	**81.26** (1.94)	**87.41** (1.08)	**78.13** (2.54)	**92.95** (0.65)	**95.98** (0.30)	**88.31** (0.89)
30%	DCP	93.33 (0.28)	75.78 (0.51)	81.15 (0.68)	30.18 (1.78)	28.65 (2.23)	7.70 (1.58)	59.65 (4.49)	75.49 (1.43)	46.93 (2.96)	46.35 (1.12)	73.17 (0.36)	29.36 (0.68)
	PMIMC	84.07 (0.80)	52.60 (2.44)	57.96 (1.84)	17.66 (2.05)	6.75 (2.98)	1.80 (1.49)	46.42 (1.87)	52.98 (1.50)	34.88 (1.56)	60.65 (0.41)	76.47 (0.66)	43.46 (0.92)
	FreeCSL	84.00 (3.42)	54.32 (7.20)	58.89 (7.32)	28.69 (2.44)	26.31 (2.39)	12.17 (2.07)	75.71 (2.13)	83.50 (1.55)	71.77 (2.67)	66.01 (1.00)	81.43 (0.91)	53.00 (1.52)
	GUITAR	38.83 (11.92)	3.15 (6.31)	3.28 (7.33)	30.91 (3.63)	22.61 (2.47)	11.87 (1.73)	74.33 (3.33)	82.23 (1.62)	68.67 (3.75)	65.78 (1.59)	81.49 (0.69)	53.94 (1.42)
	HSACC	92.74 (0.43)	75.00 (1.02)	79.78 (1.12)	24.94 (3.26)	18.35 (4.60)	8.08 (2.71)	65.17 (2.61)	75.75 (2.87)	52.03 (6.25)	47.88 (2.21)	70.88 (2.54)	26.91 (5.95)
	IMDC-DIA	91.73 (0.67)	72.40 (1.89)	77.21 (1.75)	18.22 (0.69)	5.70 (1.23)	1.30 (0.46)	61.11 (3.05)	68.58 (1.27)	49.82 (2.48)	48.90 (1.08)	71.62 (0.39)	31.09 (0.94)
	IMVC-NCD	90.00 (0.76)	68.78 (1.35)	73.10 (1.73)	27.09 (1.31)	23.12 (3.26)	11.07 (1.59)	72.61 (2.60)	81.57 (1.10)	67.38 (2.56)	67.23 (1.21)	81.42 (0.52)	51.47 (1.75)
	OPTION	91.86 (0.60)	73.00 (0.87)	77.66 (1.40)	26.06 (0.68)	17.74 (1.08)	6.06 (0.82)	70.09 (2.70)	78.66 (1.83)	61.42 (2.65)	54.57 (1.99)	72.63 (1.50)	36.16 (2.82)
	Ours	**94.43** (0.82)	**78.80** (2.44)	**84.11** (2.23)	**42.96** (2.99)	**43.84** (2.57)	**28.32** (3.45)	**80.94** (1.81)	**87.74** (0.93)	**77.94** (1.87)	**88.31** (0.89)	**90.72** (0.20)	**75.64** (0.78)
50%	DCP	67.59 (1.84)	51.99 (2.02)	46.06 (1.71)	31.71 (1.70)	30.65 (1.88)	5.05 (0.86)	61.90 (2.10)	76.17 (1.19)	51.33 (4.60)	40.55 (3.38)	66.76 (1.07)	19.45 (2.11)
	PMIMC	71.77 (1.41)	30.86 (2.28)	33.56 (2.46)	18.37 (0.47)	6.21 (0.54)	1.77 (0.21)	38.57 (1.77)	41.54 (1.47)	24.29 (1.54)	42.31 (1.79)	64.86 (0.52)	23.26 (0.89)
	FreeCSL	80.53 (4.04)	49.04 (6.46)	52.30 (7.67)	28.59 (1.38)	26.26 (2.43)	12.07 (1.43)	75.40 (1.56)	82.60 (1.46)	70.63 (2.57)	50.45 (1.52)	72.93 (0.62)	35.27 (1.32)
	GUITAR	40.63 (10.64)	4.17 (5.48)	4.27 (6.18)	29.63 (0.95)	21.28 (1.51)	11.50 (0.98)	71.47 (3.86)	81.27 (2.14)	66.38 (3.83)	58.56 (0.82)	75.59 (0.71)	43.19 (1.35)
	HSACC	87.44 (1.01)	63.40 (2.01)	66.30 (2.36)	24.50 (2.50)	19.45 (3.88)	7.42 (1.10)	68.86 (3.12)	78.46 (2.07)	60.02 (1.97)	37.08 (2.11)	63.55 (3.23)	15.92 (3.24)
	IMDC-DIA	87.97 (0.62)	63.70 (1.09)	68.00 (1.38)	18.16 (0.84)	5.15 (1.48)	1.10 (0.60)	56.28 (1.48)	63.73 (1.02)	42.96 (1.70)	36.68 (0.63)	63.92 (0.37)	18.25 (0.47)
	IMVC-NCD	85.66 (1.03)	59.73 (2.75)	62.83 (2.34)	27.68 (1.91)	22.84 (1.53)	10.67 (0.70)	73.98 (1.53)	81.30 (1.40)	67.44 (2.30)	52.62 (1.26)	73.41 (0.46)	34.01 (1.42)
	OPTION	90.76 (0.84)	70.33 (1.54)	74.96 (1.94)	23.37 (1.29)	15.48 (1.59)	5.31 (1.32)	67.66 (3.68)	77.05 (1.77)	59.08 (2.00)	41.41 (2.27)	62.96 (2.00)	20.72 (2.92)
	Ours	**91.46** (0.92)	**71.16** (2.15)	**76.41** (2.31)	**41.75** (3.84)	**39.79** (4.40)	**25.34** (4.74)	**81.83** (1.28)	**87.48** (0.84)	**78.08** (2.02)	**76.58** (0.52)	**85.20** (0.25)	**62.72** (0.58)
70%	DCP	45.03 (2.61)	8.89 (1.64)	5.87 (1.45)	17.37 (1.19)	9.47 (2.64)	0.73 (0.38)	53.29 (4.45)	69.59 (3.70)	30.14 (7.66)	33.30 (1.44)	59.17 (1.46)	8.87 (1.52)
	PMIMC	63.03 (1.74)	19.66 (2.55)	20.39 (2.27)	16.60 (0.91)	4.15 (0.58)	0.62 (0.27)	29.15 (1.51)	31.10 (0.97)	14.03 (0.68)	29.19 (0.66)	57.08 (0.48)	12.34 (0.64)
	FreeCSL	76.00 (4.45)	42.24 (7.23)	43.73 (7.29)	27.44 (1.19)	23.89 (2.35)	11.07 (0.98)	69.44 (2.07)	77.81 (1.05)	61.22 (2.72)	39.06 (1.05)	65.08 (0.63)	22.26 (1.11)
	GUITAR	40.96 (10.35)	2.31 (4.44)	2.25 (5.04)	26.59 (1.00)	17.97 (1.23)	8.26 (0.91)	71.58 (2.24)	81.35 (1.44)	66.44 (2.36)	48.85 (1.47)	68.79 (0.61)	30.74 (1.09)
	HSACC	74.66 (4.69)	47.66 (4.70)	38.14 (9.70)	26.25 (2.74)	20.65 (3.82)	7.77 (1.48)	64.47 (3.82)	75.00 (2.71)	50.90 (6.33)	28.14 (1.11)	58.51 (0.96)	11.10 (0.99)
	IMDC-DIA	83.13 (1.01)	54.34 (2.04)	57.25 (2.07)	17.84 (0.30)	5.10 (1.03)	0.97 (0.36)	47.39 (3.06)	58.67 (2.17)	35.38 (4.12)	28.76 (1.30)	59.15 (1.10)	12.17 (1.14)
	IMVC-NCD	78.96 (6.16)	50.53 (6.06)	50.72 (8.33)	26.15 (2.36)	20.49 (2.37)	9.46 (1.47)	73.24 (2.05)	80.85 (0.78)	67.18 (1.79)	41.79 (1.49)	68.01 (0.45)	23.69 (0.90)
	OPTION	86.10 (2.91)	61.18 (3.31)	64.79 (5.26)	24.87 (1.95)	17.59 (3.07)	7.67 (1.62)	65.94 (2.30)	74.77 (1.20)	56.63 (2.19)	28.71 (1.66)	54.57 (1.35)	9.23 (2.54)
	Ours	**87.95** (0.69)	**63.04** (1.50)	**67.97** (1.48)	**36.87** (1.40)	**32.56** (0.90)	**17.71** (0.92)	**80.80** (1.45)	**86.90** (0.70)	**77.14** (1.12)	**63.43** (1.27)	**76.94** (0.52)	**45.66** (1.19)
90%	DCP	39.50 (3.82)	2.08 (1.93)	1.94 (2.11)	16.68 (1.38)	9.27 (2.31)	0.42 (0.16)	9.30 (0.30)	4.89 (0.36)	−0.07 (0.06)	10.76 (4.81)	36.21 (2.65)	0.47 (0.10)
	PMIMC	52.57 (4.58)	11.74 (2.98)	11.45 (3.26)	16.72 (1.00)	4.06 (0.77)	0.57 (0.36)	20.24 (1.21)	24.23 (1.29)	7.26 (0.69)	23.09 (0.73)	53.28 (0.71)	7.83 (0.63)
	FreeCSL	76.13 (5.89)	42.61 (8.34)	44.22 (9.90)	27.84 (1.35)	20.70 (0.80)	9.49 (0.64)	46.60 (3.54)	56.50 (3.20)	32.94 (3.30)	29.00 (0.70)	58.25 (0.37)	13.09 (0.36)
	GUITAR	40.66 (6.56)	2.43 (1.95)	2.07 (1.91)	26.66 (2.78)	15.72 (2.84)	7.47 (2.38)	70.49 (2.39)	78.50 (0.99)	62.38 (2.29)	36.29 (1.19)	61.42 (0.43)	19.13 (0.72)
	HSACC	35.78 (0.86)	0.30 (0.13)	−0.04 (0.15)	16.72 (0.86)	5.93 (1.58)	0.48 (0.20)	34.71 (5.08)	49.52 (4.49)	8.45 (2.31)	23.41 (0.60)	52.66 (1.05)	5.37 (0.76)
	IMDC-DIA	79.53 (1.33)	48.88 (1.82)	50.09 (2.47)	18.47 (0.68)	5.99 (1.11)	1.20 (0.60)	39.69 (1.79)	53.43 (1.69)	26.33 (2.16)	25.19 (0.72)	56.91 (0.79)	10.52 (0.82)
	IMVC-NCD	62.16 (3.15)	37.31 (4.29)	32.20 (3.52)	28.12 (2.20)	20.53 (2.98)	9.36 (1.99)	71.94 (1.17)	79.13 (1.20)	64.37 (1.84)	32.82 (1.15)	63.19 (0.80)	15.94 (0.63)
	OPTION	**86.40** (0.99)	**59.87** (1.77)	**64.59** (2.06)	20.00 (1.64)	8.57 (2.28)	2.75 (0.94)	67.72 (1.07)	76.88 (0.64)	58.76 (0.79)	26.60 (1.26)	57.98 (1.16)	10.93 (0.84)
	Ours	83.03 (2.52)	54.24 (4.28)	57.77 (5.00)	**30.37** (1.14)	**21.38** (1.37)	**10.56** (0.84)	**76.25** (2.31)	**82.01** (2.04)	**69.53** (3.04)	**40.95** (1.25)	**65.16** (0.72)	**23.34** (1.13)

we further average the performance of each method over the five missing ratios. On Scene-15, our method surpasses the strongest baseline in terms of averaged ACC, NMI, and ARI by 7.90, 0.57, and 4.23 percentage points, respectively. The other datasets exhibit the same trend, with our method achieving higher averaged results than the competing methods on all three evaluation metrics. A detailed comparison of computational efficiency is provided in Appendix I.

To qualitatively assess representation evolution, Figure 3 shows that samples from different classes are initially intermingled, resulting in ambiguous class boundaries. After training, the representations form more compact clusters with substantially reduced inter-class overlap, revealing clearer local structures and greater class separability. Results on additional datasets are provided in Appendix H. This comparison provides qualitative evidence that GRADE captures local structural information and learns more discriminative representations from incomplete multi-view data.



(a) COIL20MV: Initial (b) COIL20MV: Final

Figure 3: t-SNE on COIL20MV at a missing ratio of 0.5. Colors denote ground-truth labels.

5.3 Ablation Study (Q2)

Table 2 presents the ablations of the training objectives and diffusion formulation. Adding $\mathcal{L}_{\mathrm{GB}}$ or $\mathcal{L}_{\mathrm{diff}}$ to the base objective $\mathcal{L}_{\mathrm{rec}}$ increases the average ACC from 25.30% to 35.82% and 28.01%, respectively, with $\mathcal{L}_{\mathrm{GB}}$ yielding improvements on all eight datasets. Combining both components under the residual formulation further raises the average ACC to 48.95% and outperforms all single-component variants, confirming their complementarity. Compared with full representation generation, residual generation performs better on seven of the eight datasets and improves the average ACC

from 42.51% to 48.95%. This demonstrates that recovering only the information not captured by the semantic base is generally more effective than directly generating the entire target representation.

Table 2: ACC results for ablations of the training objectives and diffusion formulation at a missing ratio of 0.9 with random seed 42. Corresponding NMI and ARI results are provided in Appendix J.

$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{GB}}$	$\mathcal{L}_{\text{diff}}$	Diffusion	Synthetic3D	EYaleB10	COIL20MV	100Leaves	Mfeat	LandUse-21	Scene-15	Animal
Training objectives											
✓	×	×	–	60.83	16.56	29.16	27.56	18.75	15.19	22.14	12.20
✓	✓	×	–	81.66	28.12	52.91	29.75	32.70	20.61	26.64	14.14
✓	×	✓	Residual	66.83	19.06	28.05	26.25	32.95	15.90	20.20	14.83
Diffusion formulation											
✓	✓	✓	Full	**84.50**	26.09	66.67	30.25	59.90	19.90	37.59	15.17
✓	✓	✓	Residual	82.17	**30.00**	**77.15**	**40.56**	**80.10**	**24.43**	**41.36**	**15.79**

Figure 4 compares the cosine similarity matrices under different component configurations. With only $\mathcal{L}_{\text{rec}}$, intra-class and inter-class similarities remain insufficiently distinguished. Adding $\mathcal{L}_{\text{GB}}$ suppresses inter-class similarities but leaves limited intra-class compactness, whereas $\mathcal{L}_{\text{diff}}$ alone increases both without establishing clear class boundaries. The complete model produces the clearest block-diagonal structure, confirming the complementary roles of adaptive granularity alignment and residual diffusion in improving intra-class compactness and inter-class separability.

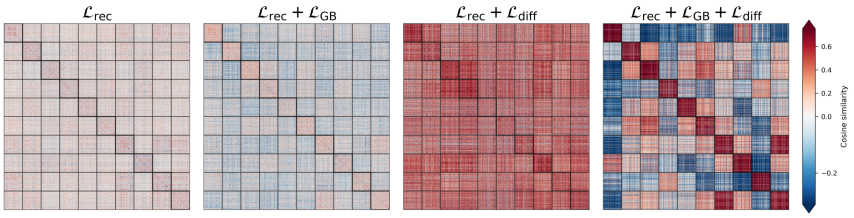


Figure 4: Cosine similarity matrices for different component configurations on Mfeat at a missing ratio of 0.9 (seed 42). Results on other datasets are provided in Appendix J.

5.4 Parameter analysis (Q3)

Figure 7 in Appendix K illustrates the effects of the temperature coefficient τ and the number of retrieved granular-ball prototypes K_{gb} on clustering performance. For τ, Synthetic3D, Mfeat, and COIL20MV exhibit performance improvements as τ increases from 0.1 to 0.3 and remain relatively stable at larger values, while Scene-15, LandUse-21, and Animal show only minor fluctuations. In contrast, the performance on 100Leaves and EYaleB10 decreases with larger values of τ, suggesting that an overly smoothed relation distribution may weaken the discrimination of local semantics on these datasets. As K_{gb} increases from 1 to 9, only minor performance variations are observed across the eight datasets, indicating that different choices of K_{gb} have a limited influence on the results. Overall, an appropriate τ facilitates favorable clustering performance, while the model maintains relatively stable performance across different values of K_{gb}.

6 Conclusion

This paper addresses missing representation recovery in incomplete multi-view clustering and proposes GRADE to jointly preserve cross-view semantic consistency and target-view-specific local structure. GRADE establishes correspondences between adaptive local regions using co-observation evidence and retrieves query-relevant local priors from the target view as structural conditions, guiding residual diffusion to complement the semantic base representation with target-view-specific information. Extensive experiments on eight multi-view benchmark datasets under various missing rates demonstrate the superior overall clustering performance of GRADE, while ablation studies further validate the effectiveness of its individual components. Future work will explore more efficient residual diffusion strategies and extend the proposed method to more complex missing scenarios.

AI USE STATEMENT

We used generative AI tools for translation, language polishing, consistency checks, and limited assistance with code review and debugging. These tools were not used to formulate the core research idea, determine the proposed methodology, generate or fabricate data, or independently draw scientific conclusions. All AI-assisted text and technical suggestions were carefully reviewed and verified by the authors. The authors take full responsibility for the methods, experiments, claims, and final content presented in this work.

ETHICS STATEMENT

This work uses existing publicly available benchmark datasets and involves no recruitment of human participants or collection of new personal or sensitive data. We comply with the applicable dataset licenses and cite the original sources. As inaccurate representation recovery may produce misleading clusters or amplify existing data biases, GRADE should not be used in high-stakes applications without domain-specific validation and appropriate human oversight.

REPRODUCIBILITY STATEMENT

To support reproducibility, we provide an anonymous link to the source code in Appendix L. The complete training and inference procedure of GRADE, together with the construction of adaptive granularity, is summarized in pseudocode in Appendix D. Dataset statistics and sources are provided in Appendix E, while the implementations, code sources, and hyperparameter settings of the baseline methods are documented in Appendix F. In addition, the relevant theoretical analysis and complete proof are provided in Appendix B.

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.

Wenlan Chen, Lu Gao, Daoyuan Wang, Fei Guo, and Cheng Liang. Geometry-aware variational information maximization for deep incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 20289–20297, 2026.

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. IEEE, 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, 2025.

Li Fei-Fei and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 2, pp. 524–531. IEEE, 2005.

Qian Guo, Gaohui Zuo, Bingbing Jiang, Guangrui Fan, Zhihua Cui, Xinyan Liang, and Jianjian Ding. Incomplete multi-view clustering via neighborhood-conditioned diffusion. In *Forty-third International Conference on Machine Learning*, 2026.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

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. IEEE, 2023.

Abhishek Kumar, Piyush Rai, and Hal Daume. Co-regularized multi-view spectral clustering. *Advances in neural information processing systems*, 24, 2011.

Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE transactions on pattern analysis and machine intelligence*, 36(3):453–465, 2013.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11169–11178. IEEE, 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.

Guangcan Liu, Zhouchen Lin, and Yong Yu. Robust subspace segmentation by low-rank representation. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 663–670, 2010.

Jiyuan Liu, Xinwang Liu, Xinhang Wan, Ke Liang, Weixuan Liang, Huijun Wu, Kehua Guo, et al. Incomplete multi-view deep clustering with data imputation and alignment. *Advances in Neural Information Processing Systems*, 38:89119–89142, 2025.

Shuaiyu Liu, Song Wu, Jie Xu, Yazhou Ren, Yang Yang, Xiaorong Pu, and Guoying Wang. Views attention fusion of granular-ball fuzzy representations split for improved multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 23810–23818, 2026.

S Nene. Columbia object image library. *COIL-100. Technical Report*, 6, 1996.

Feiping Nie, Lai Tian, and Xuelong Li. Multiview clustering via adaptively weighted procrustes. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2022–2030, 2018.

Pengjiang Qian, Fu-Lai Chung, Shitong Wang, and Zhaohong Deng. Fast graph-based relaxed clustering for large data sets using minimal enclosing ball. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(3):672–687, 2012.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research*, 3(Dec):583–617, 2002.

Peng Su, Shudong Huang, Weihong Ma, Deng Xiong, and Jiancheng Lv. Multi-view granular-ball contrastive clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20637–20645, 2025.

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.

Mingrui Wu and Bernhard Schölkopf. A local learning approach for clustering. *Advances in neural information processing systems*, 19, 2006.

Renjie Wu, Hu Wang, Hsiang-Ting Chen, and Gustavo Carneiro. Deep multimodal learning with missing modality: A survey. *arXiv preprint arXiv:2409.07825*, 2024.

Shuyin Xia, Yunsheng Liu, Xin Ding, Guoyin Wang, Hong Yu, and Yuoguo Luo. Granular ball computing classifiers for efficient, scalable and robust learning. *Information Sciences*, 483:136–152, 2019.

Shuyin Xia, Daowan Peng, Deyu Meng, Changqing Zhang, Guoyin Wang, Elisabeth Giem, Wei Wei, and Zizhong Chen. Ball k k-means: Fast adaptive clustering with no bounds. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):87–99, 2020.

Jiang Xie, Xuexin Xiang, Shuyin Xia, Lian Jiang, Guoyin Wang, and Xinbo Gao. Mgnr: A multi-granularity neighbor relationship and its application in knn classification and clustering methods. *IEEE transactions on pattern analysis and machine intelligence*, 46(12):7956–7972, 2024.

Jiang Xie, Yuxin Cheng, Shuyin Xia, Chunfeng Hua, Guoyin Wang, and Xinbo Gao. Aw-gbgae: an adaptive weighted graph autoencoder based on granular-balls for general data clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Jie Xu, Chao Li, Yazhou Ren, Liang Peng, Yujie Mo, Xiaoshuang Shi, and Xiaofeng Zhu. Deep incomplete multi-view clustering via mining cluster complementarity. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 8761–8769, 2022.

Yi Yang and Shawn Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th 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.

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.

Yuanyang Zhang, Yijie Lin, Xinhang Wan, Jie Xu, Li Yao, Weiqing Yan, and Chang Tang. Structure-aware conditional diffusion generation for incomplete multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2026.

Bowen Zhao, Qianqian Wang, Zhengming Ding, and Quanxue Gao. Attribute-missing multi-view graph clustering. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25832–25841. IEEE, 2025.

Qinghai Zheng, Jihua Zhu, and Zhongyu Li. Collaborative unsupervised multi-view representation learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(7):4202–4210, 2021.

Zhenhao Zhong, Zhibin Gu, Pengpeng Yang, Ruiqiang Guo, et al. Gaussian regression-driven tensorized incomplete multi-view clustering with dual manifold regularization. *Advances in Neural Information Processing Systems*, 38:72102–72131, 2025.

Siyuan Zhou and Zhibin Gu. Option: Optimal transport–guided flow matching for incomplete and unaligned multi-view clustering. In *Forty-third International Conference on Machine Learning*, 2026.

A Notation Summary

The main notation used in GRADE is summarized in Table 3.

Table 3: Summary of the main notation used in GRADE.

Symbol	Description	Symbol	Description
Basic notation			
N	Number of instances	V	Number of views
K	Number of clusters	d	Latent representation dimension
$\mathbf{M}$	View-availability matrix	$\widetilde{\mathbf{z}}_i^{(v)}$	Observed latent representation
Evidence-aligned adaptive granularity			
$GB^{(v)}$	Granular-ball set in view v	$GB_g^{(v)}$	The g-th granular ball
$I_g^{(v)}$	Member-index set of $GB_g^{(v)}$	$\mathbf{c}_g^{(v)}$	Granular-ball center
$\widehat{\mathbf{c}}_g^{(v)}$	Normalized local prototype	$O_{gh}^{(u,v)}$	Intersection evidence
$A_{gh}^{(u,v)}$	Evidence-validity mask	$P_{gh}^{u\to v}$	Evidence-derived soft target
$Q_{gh}^{u\to v}$	Predicted correspondence distribution	$\mathcal{L}_{\mathrm{GB}}$	Adaptive-granularity alignment loss
Structure-aware residual diffusion			
$\mathbf{q}_{i\to v}$	Cross-view query	$\mathbf{b}_i^{(v)}$	Semantic base representation
$\mathbf{r}_{0,i}^{(v)}$	Clean target residual	$\mathbf{r}_{t,i}^{(v)}$	Noisy residual at timestep t
K_{gb}	Number of retrieved prototypes	$\mathbf{C}_{i\to v}$	Retrieved local structural condition
$\mathbf{a}_{t,i\to v}$	Cross-attention condition	ϵ_{θ_v}	Target-view denoiser
$\widehat{\mathbf{r}}_{0,i}^{(v,t)}$	Clean-residual estimate at timestep t	$\widehat{\mathbf{r}}_{0,i}^{(v)}$	Generated residual
$\widehat{\mathbf{z}}_i^{(v)}$	Completed target-view representation	$\mathbf{h}_i$	Final fused representation
Learning objectives			
$\mathcal{L}_{\mathrm{rec}}$	Reconstruction loss	$\mathcal{L}_{\mathrm{diff}}$	Residual diffusion loss
$\mathcal{L}_{\mathrm{overall}}$	Overall training objective		

B Compactness Analysis of DM Based Splitting

To analyze the spatial compactness induced by the splitting process, we introduce the sum of squared errors (SSE). For a granular-ball $GB_g^{(v)}$, its SSE is defined as

$$\mathrm{SSE}\Big(GB_g^{(v)}\Big) = \sum_{i\in I_g^{(v)}} \Big\|\widetilde{\mathbf{x}}_i^{(v)} - \mathbf{c}_g^{(v)}\Big\|_2^2. \tag{24}$$

Unlike DM, which determines whether a candidate split is accepted, SSE measures the overall squared dispersion of the samples around their granular-ball center.

Suppose that a parent granular-ball $GB_g^{(v)}$ is divided into two nonempty child balls $GB_{g,1}^{(v)}$ and $GB_{g,2}^{(v)}$, whose member index sets satisfy

$$I_g^{(v)} = I_{g,1}^{(v)} \cup I_{g,2}^{(v)}, \qquad I_{g,1}^{(v)} \cap I_{g,2}^{(v)} = \varnothing. \tag{25}$$

If the candidate split satisfies

$$\mathrm{DM_w}\Big(GB_g^{(v)}\Big) < \mathrm{DM}\Big(GB_g^{(v)}\Big), \tag{26}$$

we prove that

$$\mathrm{SSE}\Big(GB_g^{(v)}\Big) > \mathrm{SSE}\Big(GB_{g,1}^{(v)}\Big) + \mathrm{SSE}\Big(GB_{g,2}^{(v)}\Big). \tag{27}$$

Proof. For clarity, we omit the view superscript (v). Let $\mathbf{c}_g$ denote the center of the parent ball and $\mathbf{c}_{g,t}$ denote the center of child ball t, where $t \in \{1,2\}$. Since $\mathbf{c}_{g,t}$ is the mean center of its members, we have

$$\sum_{i\in I_{g,t}} (\widetilde{\mathbf{x}}_i - \mathbf{c}_{g,t}) = \mathbf{0}. \tag{28}$$

Therefore,

$$\begin{aligned}
&\sum_{i\in I_{g,t}} \|\widetilde{\mathbf{x}}_i - \mathbf{c}_g\|_2^2 \\
&= \sum_{i\in I_{g,t}} \|(\widetilde{\mathbf{x}}_i - \mathbf{c}_{g,t}) + (\mathbf{c}_{g,t} - \mathbf{c}_g)\|_2^2 \\
&= \mathrm{SSE}(GB_{g,t}) + |I_{g,t}|\,\|\mathbf{c}_{g,t} - \mathbf{c}_g\|_2^2 .
\end{aligned} \tag{29}$$

Summing over the two child balls gives

$$\begin{aligned}
\mathrm{SSE}(GB_g) = \mathrm{SSE}(GB_{g,1}) + \mathrm{SSE}(GB_{g,2}) \\
+ \sum_{t=1}^{2} |I_{g,t}|\,\|\mathbf{c}_{g,t} - \mathbf{c}_g\|_2^2 .
\end{aligned} \tag{30}$$

Since the last term is nonnegative, splitting cannot increase the total within-ball SSE.

It remains to exclude equality. If equality held, the nonemptiness of both child balls would imply

$$\mathbf{c}_{g,1} = \mathbf{c}_{g,2} = \mathbf{c}_g. \tag{31}$$

Consequently,

$$\begin{aligned}
\mathrm{DM_w}(GB_g) &= \frac{1}{|I_g|}\sum_{t=1}^{2}\sum_{i\in I_{g,t}} \|\widetilde{\mathbf{x}}_i - \mathbf{c}_{g,t}\|_2 \\
&= \frac{1}{|I_g|}\sum_{i\in I_g} \|\widetilde{\mathbf{x}}_i - \mathbf{c}_g\|_2 \\
&= \mathrm{DM}(GB_g)\,,
\end{aligned} \tag{32}$$

which contradicts the strict acceptance condition in Eq. 26. Hence, equality is impossible, and Eq. 27 holds. Since all other granular-balls remain unchanged, replacing the parent ball with its two child balls also strictly reduces the total SSE of the entire partition. Therefore, every split accepted by the DM criterion produces a more compact adaptive granularity structure from the SSE perspective.

C Adaptive Granularity Construction via Recursive Splitting

This section details the recursive splitting procedure used to construct adaptive local regions in each view. As summarized in Algorithm 1, the procedure recursively evaluates candidate splits using the DM criterion until no further split is accepted.

D Overall Procedure of GRADE

This section summarizes the complete training and inference procedure of GRADE. As detailed in Algorithm 2, the framework integrates representation learning, evidence-aligned adaptive granularity, and structure guided residual diffusion for missing representation recovery and final clustering.

E Dataset Details

Table 4 summarizes the statistics of the eight multi-view benchmark datasets used in our experiments.

F Details of Compared Methods

This section provides brief descriptions, code repositories, and principal hyperparameter settings for the eight compared methods used in our experiments.

Algorithm 1 Adaptive Granularity Construction via Recursive Splitting

Require: Observed sample set $\widetilde{\mathbf{X}}^{(v)}$ of view v
Ensure: Granular ball set $GB^{(v)} = \{GB_g^{(v)}\}_{g=1}^{G_v}$

Initialize $GB^{(v)} \leftarrow \{GB_1^{(v)}\}$, where $I_1^{(v)} = \{i \mid M_{iv} = 1\}$
repeat
  split $\leftarrow$ False
  for each $GB_g^{(v)} \in GB^{(v)}$ **do**
    Compute the center $\mathbf{c}_g^{(v)}$ using Eq. (1)
    $a_1 \leftarrow \arg\max_{i \in I_g^{(v)}} \left\| \widetilde{\mathbf{x}}_i^{(v)} - \mathbf{c}_g^{(v)} \right\|_2$
    $a_2 \leftarrow \arg\max_{i \in I_g^{(v)}} \left\| \widetilde{\mathbf{x}}_i^{(v)} - \widetilde{\mathbf{x}}_{a_1}^{(v)} \right\|_2$
    $I_{g,1}^{(v)} \leftarrow \varnothing$, $I_{g,2}^{(v)} \leftarrow \varnothing$
    for each $i \in I_g^{(v)}$ **do**
      if $\left\| \widetilde{\mathbf{x}}_i^{(v)} - \widetilde{\mathbf{x}}_{a_1}^{(v)} \right\|_2 \leq \left\| \widetilde{\mathbf{x}}_i^{(v)} - \widetilde{\mathbf{x}}_{a_2}^{(v)} \right\|_2$ **then**
        $I_{g,1}^{(v)} \leftarrow I_{g,1}^{(v)} \cup \{i\}$
      else
        $I_{g,2}^{(v)} \leftarrow I_{g,2}^{(v)} \cup \{i\}$
      end if
    end for
    Construct $GB_{g,1}^{(v)}$ and $GB_{g,2}^{(v)}$ from $I_{g,1}^{(v)}$ and $I_{g,2}^{(v)}$
    Compute $\mathrm{DM_w}\left(GB_g^{(v)}\right)$ using Eqs. (2) and (3)
    if $I_{g,1}^{(v)} \neq \varnothing$ and $I_{g,2}^{(v)} \neq \varnothing$ and $\mathrm{DM_w}\left(GB_g^{(v)}\right) < \mathrm{DM}\left(GB_g^{(v)}\right)$ **then**
      $GB^{(v)} \leftarrow \left(GB^{(v)} \setminus \{GB_g^{(v)}\}\right) \cup \{GB_{g,1}^{(v)}, GB_{g,2}^{(v)}\}$
      split $\leftarrow$ True
    end if
  end for
until split $=$ False
return $GB^{(v)}$

Table 4: Statistics of the eight multi-view benchmark datasets.

Dataset	Source	Samples	Views	Classes	View Dimensions
Synthetic3D	(Kumar et al., 2011)	600	3	3	3 / 3 / 3
EYaleB10	(Liu et al., 2010)	640	3	10	1024 / 1239 / 256
COIL20MV	(Nene, 1996)	1440	3	20	1024 / 3304 / 6750
100Leaves	(Zheng et al., 2021)	1600	3	100	64 / 64 / 64
Mfeat	(Nie et al., 2018)	2000	6	10	216 / 76 / 64 / 6 / 240 / 47
LandUse-21	(Yang & Newsam, 2010)	2100	3	21	20 / 59 / 40
Scene-15	(Fei-Fei & Perona, 2005)	4485	3	15	20 / 59 / 40
Animal	(Lampert et al., 2013)	11673	4	20	2689 / 2000 / 2001 / 2000

Algorithm 2 Overall Procedure of GRADE

Require: Multi-view data $\mathbf{X} = \{\mathbf{X}^{(v)}\}_{v=1}^{V}$, observation mask $\mathbf{M}$, and number of clusters K
Ensure: Cluster assignments $\mathbf{y}$

Initialize all learnable parameters
Training
repeat
  Encode and reconstruct the observed samples of each view
  for $v = 1, \ldots, V$ **do**
    Construct $GB^{(v)}$ from $\widetilde{\mathbf{Z}}^{(v)}$ using Algorithm 1
  end for
  Compute $\mathcal{L}_{\mathrm{rec}}$ using Eq. 4
  Establish evidence-aligned correspondences across views and compute $\mathcal{L}_{\mathrm{GB}}$ using Eqs. 5–11
  for each valid training pair (i, v) **do**
    Construct $\mathbf{q}_{i\to v}$, $\mathbf{b}_i^{(v)}$, $\mathbf{r}_{0,i}^{(v)}$, and $\mathbf{C}_{i\to v}$
    Perform residual diffusion and compute $\mathcal{L}_{\mathrm{diff}}$
  end for
  Update the model by minimizing $\mathcal{L}_{\mathrm{overall}}$ in Eq. 23
until convergence
Inference
for each missing pair (i, v) **do**
  Construct $\mathbf{q}_{i\to v}$, $\mathbf{b}_i^{(v)}$, and $\mathbf{C}_{i\to v}$
  Sample $\mathbf{r}_{T,i}^{(v)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Recover $\widehat{\mathbf{r}}_{0,i}^{(v)}$ through DDIM sampling
  Complete the missing representation as $\widehat{\mathbf{z}}_i^{(v)} \leftarrow \mathbf{b}_i^{(v)} + \widehat{\mathbf{r}}_{0,i}^{(v)}$
end for
Retain the original representations of the observed views
Fuse the completed representations
Apply K-means to obtain $\mathbf{y}$
return $\mathbf{y}$

F.1 Brief Descriptions

The main mechanisms of the eight compared methods are briefly summarized below.

DCP (Lin et al., 2022) learns consistent representations by maximizing mutual information across views through contrastive learning and employs dual prediction networks to recover missing representations.

PMIMC (Yuan et al., 2025) reduces discrepancies across views through relational consistency, aligns prototypes across views using robust prototype contrastive learning, and transfers relationships between samples and prototypes to recover missing data.

FreeCSL (Dai et al., 2025) avoids both imputation and explicit alignment. It uses consensus prototype contrastive learning to map available samples into a shared semantic space and modularity based graph clustering to enhance the cluster structure of each view.

GUITAR (Zhong et al., 2025) estimates missing data through anchor decomposition and tensorized representation learning, models complex noise using Gaussian regression, and preserves local manifolds within and across views through dual manifold regularization.

HSACC (Ding et al., 2025) aligns low level features by maximizing mutual information, learns a unified global representation through adaptive fusion weighted by distribution affinity, and projects the aligned latent representations into the semantic spaces of individual views to infer missing representations.

IMDC-DIA (Liu et al., 2025) learns a shared latent representation for each sample and projects it into individual feature spaces through networks specific to each view for observed data reconstruction and missing data recovery. It further uses linear alignment to preserve pairwise similarities.

IMVC-NCD (Guo et al., 2026) constructs a unified neighborhood condition by aggregating neighboring latent representations from the observed views and encoding the missing pattern. This condition then guides a conditional diffusion model to recover missing latent representations.

OPTION (Zhou & Gu, 2026) uses conditional flow matching to learn deterministic transport trajectories for missing data recovery and introduces a structural regularizer inspired by the Gromov–Wasserstein distance to align the internal geometries of different views without requiring hard correspondences.

F.2 Code Repositories

Table 5 lists the code repositories used to reproduce the compared methods.

Table 5: Code repositories of the compared methods.

Method	Code Repository
DCP	`https://github.com/XLearning-SCU/2022-TPAMI-DCP`
PMIMC	`https://github.com/hl-yuan/PMIMC`
FreeCSL	`https://github.com/zoyadai/2025_CVPR_FreeCSL`
GUITAR	`https://github.com/RockfireTip/GUITAR`
HSACC	`https://github.com/XiaojianDing/2025-NeurIPS-HSACC`
IMDC-DIA	`https://github.com/liujiyuan13/IMDC-DIA-code_release`
IMVC-NCD	`https://github.com/zgh1115/IMVC-NCD`
OPTION	`https://github.com/TimoZhou1024/OPTION`

F.3 Hyperparameter Settings of Compared Methods

For a fair comparison, we use the official implementations and follow the recommended settings of the compared methods whenever available. The principal hyperparameter settings are reported in Table 6.

Table 6: Hyperparameter settings of the compared methods.

Method	Hyperparameter 1	Hyperparameter 2	Hyperparameter 3	Hyperparameter 4
DCP	λ_1: 0.1	λ_2: 0.1	–	–
PMIMC	α: 0.0001	β: 0.01	t: 4	
FreeCSL	ζ: 3	λ: 0.1	–	–
GUITAR	λ_1: 0.001	λ_2: 0.1	λ_3: 1	–
HSACC	λ_1: 0.1	λ_2: 10	λ_3: 5	λ_4: 0.1
IMDC-DIA	β: 1.0	–	–	–
IMVC-NCD	k: 5	w: 1.0	p: 0.1	–
OPTION	λ_{gw}: 0.1	$\lambda_{cluster}$: 0.5	λ_{recon}: 1.0	$ODESteps$: 10

G Implementation Details

All experiments are conducted on a workstation equipped with an NVIDIA RTX 4090 GPU with 24 GB of GPU memory, using Python 3.10 and PyTorch 1.13. Each encoder is structured as Input–512–256–Output (128 dimensions), with a mirrored decoder. The model is trained using the Adam optimizer with a learning rate of 10^{-3} and a batch size of 512. To reduce the influence of randomness, we conduct independent experiments using five random seeds, i.e., $\{0, 1, 42, 123, 1000\}$, and report the mean and standard deviation.

G.1 Dataset-Specific Hyperparameter Settings

The shared implementation settings are described above. The dataset-specific values of the temperature coefficient τ and the number of retrieved granular-ball prototypes K_{gb} used by GRADE are summarized in Table 7.

Table 7: Dataset-specific hyperparameter settings of GRADE.

Dataset	τ	K_{gb}
Synthetic3D	1.0	5
EYaleB10	0.2	5
COIL20MV	1.0	5
100Leaves	0.2	5
Mfeat	0.5	5
LandUse-21	0.2	5
Scene-15	0.2	5
Animal	0.2	5

H Additional Performance Comparisons

This section provides additional quantitative and qualitative comparisons that are omitted from the main text due to space limitations.

H.1 Quantitative Comparison

Table 8 reports the clustering results of GRADE and eight competing methods on the remaining datasets under different missing ratios. These results provide a more comprehensive evaluation of the performance and stability of GRADE under varying degrees of view incompleteness.

H.2 Visualization Comparison

To further illustrate the clustering quality of GRADE, Figure 5 presents the initial representations and the final representations learned by GRADE on the remaining datasets. Compared with the initial

Table 8: Clustering performance (%) on Mfeat, LandUse-21, Scene-15, and Animal under different missing ratios. Results are reported as mean with standard deviation in parentheses. The best and second-best results are boldfaced and underlined, respectively.

MR Ratio	Method	Mfeat			LandUse-21			Scene-15			Animal		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
10%	DCP	78.29 (0.75)	78.36 (1.37)	70.42 (1.61)	27.16 (1.24)	30.44 (1.03)	13.38 (1.05)	40.93 (1.46)	**43.27** (0.79)	24.54 (1.14)	14.31 (0.57)	12.27 (0.50)	4.40 (0.38)
	PMIMC	72.41 (5.28)	66.26 (2.34)	59.40 (4.07)	24.52 (0.63)	29.96 (0.51)	10.67 (0.34)	29.46 (1.35)	28.87 (0.29)	14.28 (0.45)	14.91 (0.15)	13.92 (0.22)	4.42 (0.08)
	FreeCSL	79.64 (3.35)	76.44 (2.93)	69.07 (3.69)	22.60 (0.20)	25.47 (1.08)	8.92 (0.49)	39.07 (1.42)	36.16 (1.29)	22.60 (0.99)	18.06 (0.80)	18.23 (0.81)	7.41 (0.51)
	GUITAR	79.22 (4.49)	68.87 (3.00)	62.93 (4.42)	**28.10** (0.90)	30.42 (0.85)	**13.52** (0.61)	32.58 (1.28)	26.67 (1.22)	15.05 (0.78)		N/A	
	HSACC	74.94 (5.25)	73.59 (1.28)	53.90 (6.71)	24.60 (1.29)	32.32 (1.78)	7.37 (1.30)	40.21 (1.76)	42.50 (0.80)	23.39 (1.18)	18.57 (0.76)	16.83 (0.45)	6.69 (0.90)
	IMDC-DIA	86.54 (0.65)	76.63 (0.90)	70.71 (1.09)	20.37 (0.63)	24.44 (1.31)	7.33 (0.59)	39.78 (0.90)	38.19 (1.14)	21.71 (0.59)	16.74 (0.41)	13.02 (0.26)	5.17 (0.33)
	IMVC-NCD	**92.71** (0.27)	**84.59** (0.36)	**84.29** (0.64)	27.12 (0.67)	31.70 (0.68)	12.52 (0.49)	44.49 (1.20)	42.50 (0.48)	26.58 (0.94)	16.77 (0.25)	13.22 (0.26)	6.26 (0.55)
	OPTION	78.62 (2.59)	74.27 (3.50)	67.10 (3.08)	25.26 (0.59)	27.79 (0.91)	10.81 (0.41)	39.05 (2.30)	39.48 (2.68)	24.43 (2.72)	20.78 (0.31)	17.57 (0.42)	8.90 (0.59)
	Ours	84.69 (0.23)	84.04 (0.69)	77.61 (1.00)	27.47 (0.44)	**32.48** (0.16)	13.37 (0.18)	**46.98** (0.84)	42.69 (0.72)	**28.71** (0.62)	**23.64** (0.58)	**19.89** (0.28)	**9.31** (0.27)
30%	DCP	72.82 (3.40)	73.26 (1.56)	63.32 (2.04)	26.51 (0.78)	29.92 (0.84)	13.03 (0.74)	39.50 (0.97)	**41.26** (0.95)	23.46 (0.72)	13.92 (0.64)	11.57 (0.40)	4.43 (0.77)
	PMIMC	64.25 (3.53)	46.63 (1.73)	41.65 (2.90)	21.53 (1.05)	22.77 (0.85)	7.69 (0.32)	30.81 (0.88)	26.60 (0.60)	14.46 (0.63)	15.31 (0.16)	14.86 (0.22)	4.99 (0.20)
	FreeCSL	82.85 (0.70)	79.76 (0.84)	73.41 (0.96)	22.17 (0.53)	24.61 (1.16)	8.47 (0.66)	36.53 (0.79)	34.11 (1.23)	20.10 (0.85)	17.80 (0.30)	16.81 (0.39)	6.72 (0.25)
	GUITAR	75.09 (3.43)	64.05 (1.37)	57.13 (2.66)	27.14 (0.98)	27.46 (0.65)	11.93 (0.43)	32.56 (1.28)	26.42 (0.46)	15.16 (0.42)		N/A	
	HSACC	64.37 (2.10)	62.38 (0.65)	34.12 (1.96)	23.23 (0.36)	30.77 (1.10)	7.07 (0.93)	38.14 (1.09)	39.06 (2.29)	21.38 (0.65)	17.88 (1.10)	15.88 (0.63)	6.22 (1.06)
	IMDC-DIA	67.56 (1.93)	59.85 (0.39)	42.45 (0.94)	18.36 (0.70)	21.75 (1.56)	5.89 (0.69)	35.49 (0.81)	32.39 (0.51)	16.95 (0.55)	15.05 (0.27)	11.21 (0.14)	3.86 (0.23)
	IMVC-NCD	81.07 (4.60)	71.67 (2.51)	64.99 (4.13)	23.44 (0.54)	27.14 (0.25)	9.47 (0.45)	37.47 (3.40)	36.62 (1.69)	20.82 (1.84)	14.67 (0.41)	12.02 (0.28)	4.02 (0.07)
	OPTION	61.27 (1.82)	65.30 (4.09)	48.50 (6.12)	23.85 (0.56)	25.12 (1.08)	9.68 (0.61)	34.42 (2.69)	34.73 (2.97)	20.98 (2.64)	18.08 (0.77)	15.02 (0.73)	6.31 (0.39)
	Ours	**83.77** (0.47)	**81.51** (1.27)	**75.44** (1.09)	**27.59** (0.44)	**31.90** (0.54)	**13.14** (0.27)	**46.05** (0.80)	40.77 (0.62)	**27.66** (0.71)	**22.34** (0.54)	**18.26** (0.37)	**8.34** (0.18)
50%	DCP	61.39 (3.63)	61.71 (4.45)	38.02 (6.48)	25.89 (0.75)	28.50 (0.65)	11.37 (0.17)	39.29 (1.26)	38.95 (0.64)	22.38 (0.80)	13.59 (0.89)	10.88 (0.41)	3.44 (0.55)
	PMIMC	49.40 (2.89)	30.01 (1.95)	24.35 (2.55)	17.76 (0.94)	16.48 (0.92)	4.73 (0.47)	26.05 (1.41)	20.25 (1.01)	10.52 (1.01)	15.78 (0.37)	14.73 (0.18)	5.43 (0.08)
	FreeCSL	80.59 (1.42)	75.34 (1.45)	68.92 (2.11)	22.32 (0.51)	24.36 (0.80)	8.55 (0.42)	35.38 (0.90)	33.15 (1.02)	19.52 (1.04)	16.52 (0.61)	14.62 (0.77)	5.80 (0.32)
	GUITAR	72.24 (5.14)	59.90 (3.74)	52.66 (5.27)	25.30 (1.23)	25.56 (1.24)	10.81 (0.78)	31.20 (0.97)	25.65 (0.66)	14.52 (0.41)		N/A	
	HSACC	40.34 (3.47)	42.34 (1.56)	7.50 (1.44)	23.74 (1.27)	29.07 (0.72)	7.12 (1.13)	36.32 (0.94)	36.10 (0.86)	18.40 (1.19)	17.39 (0.77)	14.39 (0.21)	5.97 (0.87)
	IMDC-DIA	51.21 (2.57)	45.91 (0.87)	21.81 (1.33)	17.18 (0.22)	19.78 (0.69)	4.85 (0.26)	30.68 (0.70)	28.01 (0.69)	13.20 (0.71)	13.74 (0.29)	9.81 (0.29)	3.06 (0.14)
	IMVC-NCD	76.25 (3.87)	65.10 (1.95)	56.67 (2.30)	18.89 (0.74)	22.02 (1.10)	6.02 (0.44)	28.18 (0.59)	30.03 (0.76)	15.01 (0.69)	12.84 (0.48)	9.42 (0.24)	2.81 (0.15)
	OPTION	65.14 (6.64)	69.80 (3.80)	56.19 (7.13)	21.47 (0.60)	22.13 (1.54)	7.93 (0.56)	35.40 (1.23)	34.78 (1.70)	20.81 (1.47)	17.41 (0.52)	14.03 (0.35)	6.63 (0.52)
	Ours	**83.63** (0.69)	**79.27** (0.64)	**74.21** (1.00)	**27.03** (0.23)	**31.58** (0.49)	**13.18** (0.59)	**45.15** (1.00)	**39.14** (0.51)	**27.04** (0.78)	**20.84** (0.51)	**16.36** (0.29)	**7.35** (0.26)
70%	DCP	43.94 (3.09)	44.97 (2.90)	15.58 (1.62)	23.50 (0.66)	26.95 (1.15)	7.44 (0.54)	37.38 (0.92)	36.03 (0.78)	20.15 (0.92)	12.43 (0.18)	10.38 (0.31)	2.85 (0.23)
	PMIMC	35.79 (1.10)	16.56 (0.42)	11.97 (0.20)	14.78 (0.78)	12.68 (0.51)	2.99 (0.33)	21.55 (0.95)	14.26 (0.45)	6.56 (0.26)	16.06 (0.13)	**14.68** (0.16)	5.58 (0.18)
	FreeCSL	78.41 (0.91)	71.11 (1.48)	64.81 (1.41)	21.48 (0.42)	23.56 (0.33)	8.20 (0.33)	34.59 (1.03)	31.60 (0.92)	18.00 (0.75)	15.82 (0.30)	13.06 (0.37)	5.05 (0.13)
	GUITAR	70.38 (4.97)	56.67 (3.57)	49.54 (5.14)	23.08 (0.71)	22.85 (0.65)	9.27 (0.66)	31.46 (1.18)	24.41 (1.08)	13.90 (0.69)		N/A	
	HSACC	28.76 (0.87)	21.27 (0.95)	4.81 (0.68)	22.01 (0.85)	28.46 (0.54)	6.99 (0.76)	36.69 (1.99)	35.45 (0.72)	17.63 (1.77)	16.16 (0.60)	12.98 (0.53)	4.80 (0.33)
	IMDC-DIA	42.03 (1.22)	37.20 (1.09)	14.85 (1.06)	15.94 (0.51)	17.25 (0.85)	3.90 (0.25)	25.89 (0.64)	25.11 (0.92)	10.33 (0.50)	12.87 (0.08)	8.59 (0.18)	2.67 (0.14)
	IMVC-NCD	61.80 (2.35)	55.52 (1.94)	42.71 (3.08)	17.10 (1.12)	19.81 (1.04)	4.75 (0.63)	23.77 (0.87)	26.64 (0.53)	10.94 (0.52)	11.48 (0.51)	7.78 (0.22)	2.10 (0.13)
	OPTION	76.01 (6.11)	73.63 (3.38)	65.65 (5.70)	21.35 (0.91)	23.11 (1.22)	8.32 (0.96)	38.12 (2.34)	36.32 (1.11)	22.53 (1.26)	18.62 (0.57)	14.29 (0.51)	**6.36** (0.72)
	Ours	**82.91** (0.62)	**77.15** (0.69)	**72.31** (1.26)	**27.20** (0.56)	**30.47** (0.52)	**13.07** (0.53)	**43.96** (0.51)	**37.04** (0.72)	**25.12** (0.53)	**19.34** (0.56)	14.39 (0.28)	6.24 (0.14)
90%	DCP	14.98 (0.60)	5.82 (1.67)	0.52 (0.21)	12.08 (0.54)	12.11 (0.75)	1.80 (0.38)	24.74 (1.91)	29.26 (1.10)	7.89 (1.04)	12.73 (0.37)	9.60 (0.49)	2.16 (0.44)
	PMIMC	26.33 (0.83)	8.21 (0.69)	4.62 (0.55)	12.25 (0.56)	10.01 (0.46)	1.89 (0.21)	16.60 (0.99)	9.00 (0.64)	3.57 (0.38)	16.53 (0.41)	**14.26** (0.36)	**5.56** (0.27)
	FreeCSL	71.58 (1.57)	60.79 (2.08)	53.69 (2.35)	21.52 (0.31)	21.44 (0.71)	7.83 (0.28)	32.67 (1.35)	28.34 (1.37)	16.05 (1.32)	15.23 (0.40)	11.97 (0.45)	4.52 (0.40)
	GUITAR	64.49 (4.87)	49.61 (2.25)	42.14 (2.66)	20.29 (0.37)	17.71 (0.29)	6.42 (0.19)	27.00 (0.97)	20.41 (0.37)	10.37 (0.53)		N/A	
	HSACC	17.31 (0.19)	8.05 (0.37)	1.46 (0.12)	12.81 (0.49)	12.69 (0.56)	2.10 (0.29)	24.16 (2.28)	24.46 (3.13)	6.04 (2.45)	15.28 (0.76)	10.84 (0.50)	4.02 (0.73)
	IMDC-DIA	34.75 (2.02)	27.81 (1.44)	13.07 (2.27)	15.03 (0.41)	16.11 (0.67)	3.38 (0.26)	23.01 (0.37)	21.95 (0.75)	8.39 (0.41)	12.09 (0.22)	7.83 (0.25)	2.26 (0.06)
	IMVC-NCD	57.10 (6.12)	52.31 (3.04)	38.46 (4.30)	16.47 (0.92)	19.07 (0.86)	4.31 (0.35)	23.89 (0.33)	26.46 (0.67)	10.51 (0.19)	11.14 (0.17)	7.02 (0.35)	1.90 (0.12)
	OPTION	**81.94** (4.92)	**74.76** (4.59)	**68.46** (6.33)	18.95 (0.41)	21.15 (0.90)	6.51 (0.50)	36.20 (1.82)	**33.05** (1.03)	19.10 (1.30)	16.30 (0.68)	11.84 (0.72)	4.80 (0.49)
	Ours	78.83 (2.01)	69.22 (1.48)	63.60 (1.58)	**25.26** (0.61)	**26.03** (0.48)	**10.50** (0.44)	**40.54** (1.40)	31.99 (1.15)	**20.46** (1.46)	**16.82** (0.74)	12.31 (0.64)	5.22 (0.49)

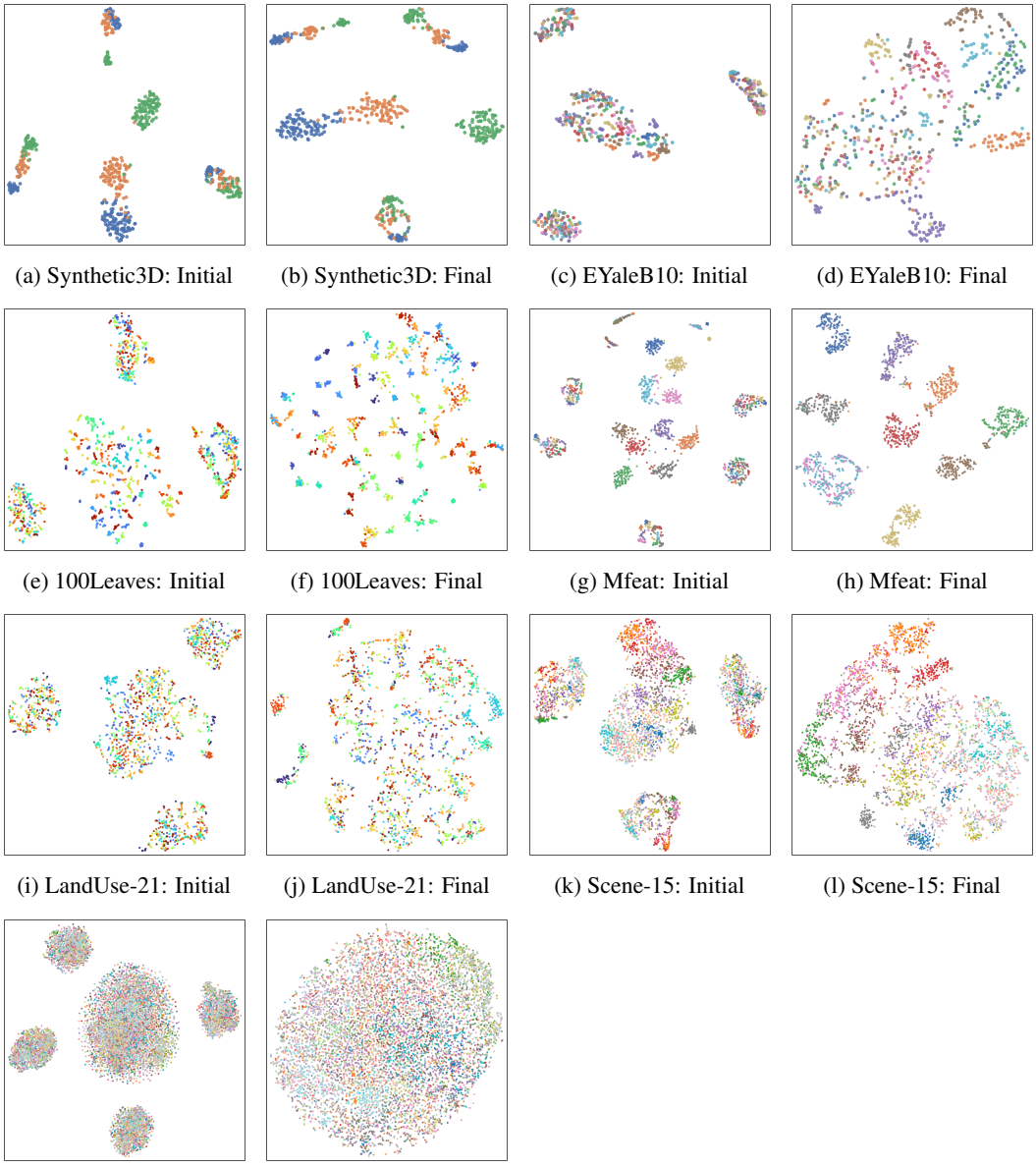
representations, the representations learned by GRADE exhibit more compact intracluster distributions and clearer intercluster separation, further demonstrating its ability to recover discriminative clustering structures from incomplete multi-view data.

I COMPUTATIONAL EFFICIENCY ANALYSIS

As shown in Table 9, lightweight methods such as DCP and IMDC-DIA generally require less running time. Compared with these lightweight approaches, GRADE introduces additional computational cost due to granular-ball structure modeling and diffusion-based missing-representation recovery. Nevertheless, GRADE remains computationally competitive with recent generative imputation methods. Specifically, compared with the diffusion-based IMVC-NCD and the flow-matching-based OPTION, GRADE achieves lower running time on six of the eight datasets. For example, on Animal, GRADE requires 984.87 s, versus 1363.82 s for IMVC-NCD and 1583.86 s for OPTION; similar trends are observed on Mfeat and Scene-15. Overall, these results demonstrate that GRADE maintains acceptable computational cost while improving clustering performance, achieving a favorable trade-off between clustering effectiveness and computational efficiency.

Table 9: Running time (seconds) of different methods on eight benchmark datasets. All methods are executed on the same hardware and timed until convergence under their respective training protocols.

Dataset	DCP	PMIMC	FreeCSL	GUITAR	HSACC	IMDC-DIA	IMVC-NCD	OPTION	Ours
Synthetic3D	12.06	39.60	39.55	13.33	44.88	13.73	59.83	59.77	28.95
EYaleB10	12.71	45.01	47.63	17.68	48.76	15.34	64.51	68.19	49.92
COIL20MV	26.94	108.28	91.89	84.37	61.97	31.46	132.23	132.63	266.30
100Leaves	48.17	118.47	84.02	116.66	79.79	28.90	180.03	146.97	302.18
Mfeat	55.57	319.43	134.39	396.94	24.42	53.21	381.18	327.38	238.98
LandUse-21	37.67	143.56	90.39	203.92	54.37	36.01	199.83	172.97	151.74
Scene-15	44.64	309.72	160.72	1474.36	101.13	66.78	391.08	366.03	267.35
Animal	302.74	1138.11	488.92	N/A	334.96	222.24	1363.82	1583.86	984.87



(a) Synthetic3D: Initial (b) Synthetic3D: Final (c) EYaleB10: Initial (d) EYaleB10: Final

(e) 100Leaves: Initial (f) 100Leaves: Final (g) Mfeat: Initial (h) Mfeat: Final

(i) LandUse-21: Initial (j) LandUse-21: Final (k) Scene-15: Initial (l) Scene-15: Final

(m) Animal: Initial (n) Animal: Final

Figure 5: t-SNE visualizations of the initial and final representations on Synthetic3D, EYaleB10, 100Leaves, Mfeat, LandUse-21, Scene-15, and Animal at a missing ratio of 0.5. Colors denote ground-truth classes.

J Additional Ablation Study

This section further evaluates the effectiveness of the key components in GRADE through additional quantitative and qualitative ablation results. We first report the NMI and ARI results of different component combinations, complementing the ACC results in the main text. We then visualize the cosine similarity matrices learned under different component configurations on additional datasets.

J.1 Ablation Results on NMI and ARI

Table 10 reports the NMI and ARI results of different component combinations under a missing ratio of 0.9. Regarding the training objectives, the complete model consistently achieves the best performance across all datasets on both metrics. In particular, evidence-aligned adaptive granularity generally improves the clustering results by providing target-view local structural information, while residual diffusion further complements the semantic base representation. Their combination produces substantial improvements, demonstrating that the two components contribute complementary information to missing-representation recovery. Regarding the diffusion formulation, residual generation outperforms full-representation generation on seven of the eight datasets for both NMI and ARI, with Synthetic3D being the only exception. This comparison demonstrates that recovering only the information not captured by the semantic base is generally more effective than directly generating the entire target representation.

Table 10: NMI and ARI results under a missing ratio of 0.9 with random seed 42.

(a) NMI

$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{GB}}$	$\mathcal{L}_{\text{diff}}$	Diffusion	Synthetic3D	EYaleB10	COIL20MV	100Leaves	Mfeat	LandUse-21	Scene-15	Animal
Training objectives											
✓	×	×	–	31.55	4.27	46.51	59.88	11.27	17.60	22.72	7.45
✓	✓	×	–	49.23	16.11	65.95	61.62	24.33	20.41	23.80	10.13
✓	×	✓	Residual	40.82	9.32	44.39	58.92	19.75	15.10	18.82	10.29
Diffusion formulation											
✓	✓	✓	Full	**55.51**	16.03	75.05	61.33	48.91	20.50	30.12	10.66
✓	✓	✓	Residual	53.81	**21.81**	**85.16**	**65.05**	**71.01**	**25.61**	**32.97**	**11.68**

(b) ARI

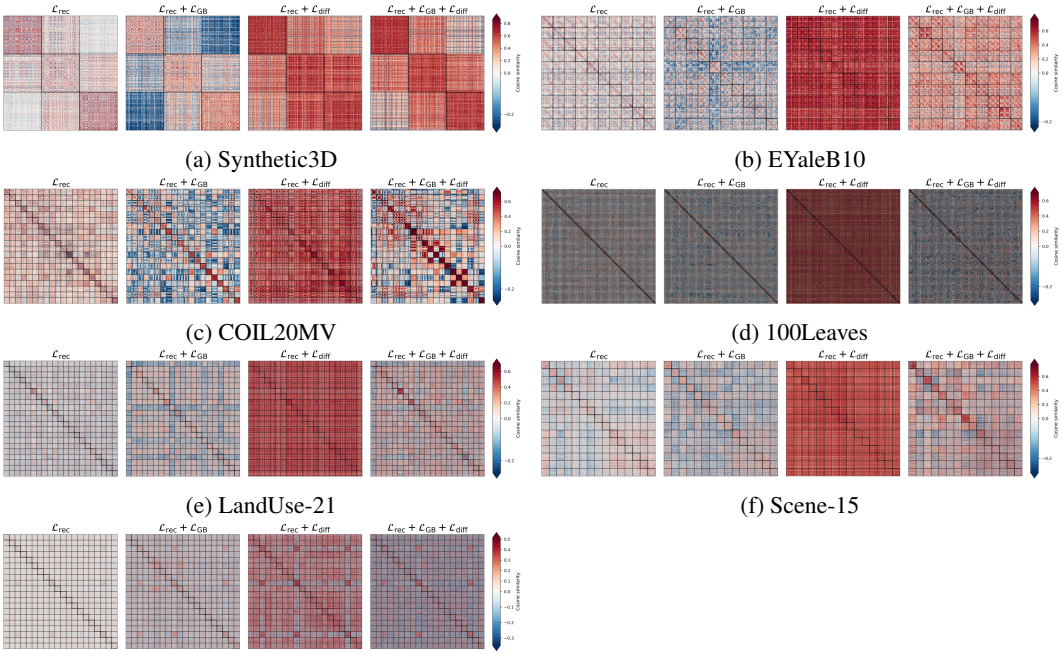
$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{GB}}$	$\mathcal{L}_{\text{diff}}$	Diffusion	Synthetic3D	EYaleB10	COIL20MV	100Leaves	Mfeat	LandUse-21	Scene-15	Animal
Training objectives											
✓	×	×	–	18.63	0.07	17.09	12.63	2.77	3.28	7.36	1.86
✓	✓	×	–	54.17	8.10	42.25	14.91	12.69	6.62	10.47	3.29
✓	×	✓	Residual	37.47	2.56	15.53	11.59	11.10	3.95	7.34	3.64
Diffusion formulation											
✓	✓	✓	Full	**60.73**	7.83	57.52	14.84	41.07	7.45	17.25	4.10
✓	✓	✓	Residual	55.69	**10.80**	**73.52**	**22.68**	**65.30**	**10.25**	**20.61**	**4.55**

J.2 Similarity Matrix Visualization

Figure 6 presents the cosine similarity matrices learned under different component configurations on seven additional datasets at a missing ratio of 0.9. Compared with the ablated variants, the complete model generally produces clearer block diagonal structures, indicating higher intraclass similarity and better interclass separation. These results further demonstrate the complementary roles of evidence-aligned adaptive granularity and residual diffusion in learning discriminative representations.

K Additional Parameter analysis

To complement the parameter analysis using ACC in the main text, this section further evaluates the effects of the temperature coefficient τ and the number of retrieved local region prototypes K_{gb} using NMI and ARI. All experimental settings are identical to those used in the main text.



(a) Synthetic3D (b) EYaleB10

(c) COIL20MV (d) 100Leaves

(e) LandUse-21 (f) Scene-15

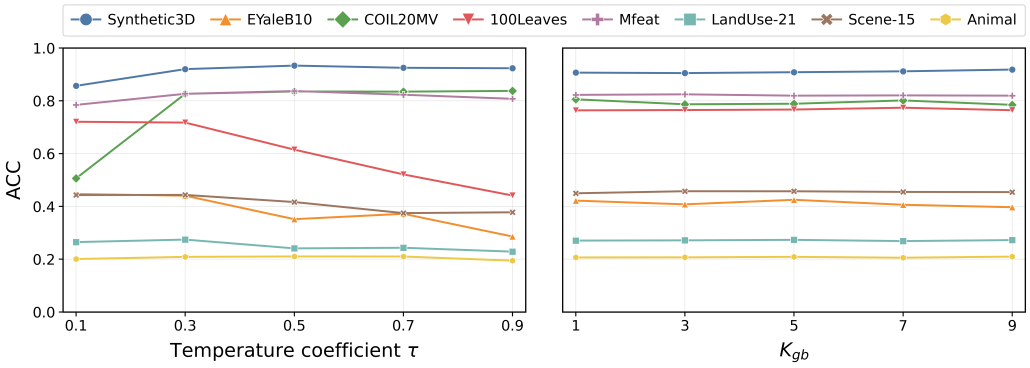
(g) Animal

Figure 6: Cosine similarity matrices under different component configurations at a missing ratio of 0.9 (seed 42). Samples are ordered by ground-truth class.

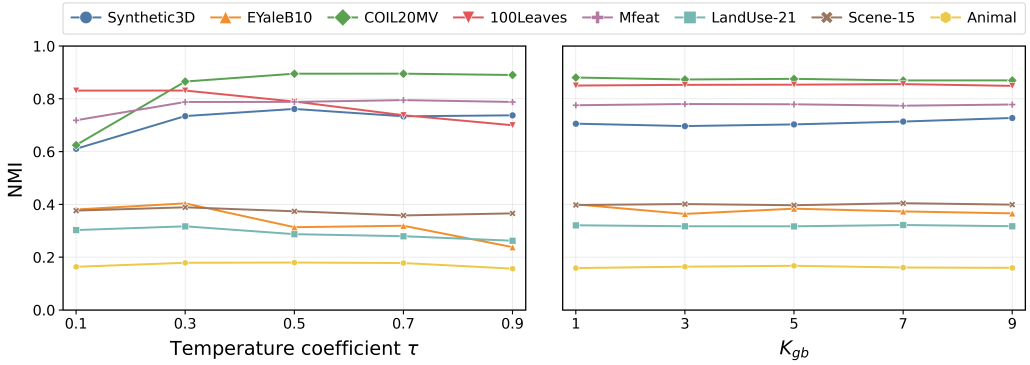
As shown in Figure 7, increasing τ from 0.1 improves both metrics on Synthetic3D, COIL20MV, and Mfeat, after which their performance remains stable or varies only slightly. In contrast, larger values of τ lead to performance degradation on 100Leaves and EYaleB10, while the remaining datasets exhibit comparatively moderate variations. Across different values of K_{gb}, only minor fluctuations are observed on all eight datasets. Together with the ACC results in the main text, these results indicate that an appropriate τ facilitates effective local correspondence learning, while GRADE remains relatively stable with respect to K_{gb}.

L Code Availability

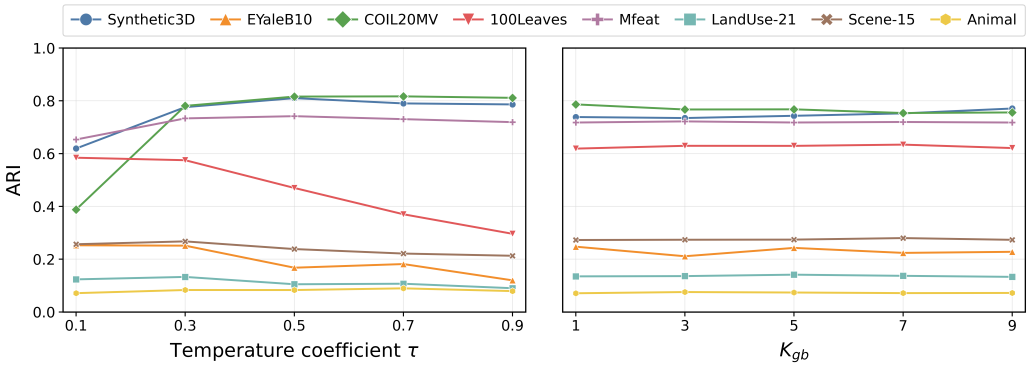
The anonymous source code for reproducing the experiments is available at: `https://anonymous.4open.science/r/GRADE-51A2`.



(a) ACC



(b) NMI



(c) ARI

Figure 7: Parameter sensitivity to the temperature coefficient τ and the number of retrieved local region prototypes K_{gb} on eight datasets. When varying τ, K_{gb} is fixed to 5; when varying K_{gb}, τ is fixed to 0.2. All results are obtained at a missing ratio of 0.9 with random seed 42.