CROSS-VIEW GRANULAR-BALL STRUCTURE TRANSFER FOR INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete multi-view clustering (IMVC) groups unlabeled samples when one or more views are unavailable. Existing graph-based methods typically recover missing structure through instance-level pairwise neighborhoods, which are vulnerable to view-specific geometry and local noise. Granular balls instead aggregate locally coherent samples into mesoscopic units that preserve neighborhood organization while reducing sensitivity to isolated perturbations. Building on this insight, we propose Cross-View Granular-Ball Structure Transfer (CGBST), a region-mediated framework that transfers granular structure rather than missing features or individual relations. CGBST partitions observed embeddings into adaptive regions and constructs their local topology. Co-observed samples establish soft cross-view region correspondences, allowing source-view affiliations to infer target-region distributions for missing samples. The inferred affiliations and target topology are projected back to the sample level to recover missing relations. Experiments on four benchmarks and three missing rates show that CGBST achieves state-of-the-art performance in 42 of 48 evaluation settings, validating region-level structure transfer for incomplete multi-view clustering.

1 INTRODUCTION

Multi-view clustering groups unlabeled samples by combining heterogeneous descriptions, such as visual descriptors, text, and sensor streams. Its central promise is that complementary views can reveal a cluster structure that no individual view captures reliably (Zhou et al., 2024; Wu et al., 2024; Ren et al., 2026). Real collections rarely satisfy the assumption that every sample is observed in every view. Sensor failures, privacy restrictions, and acquisition cost instead produce incomplete multi-view data, making both representation fusion and neighborhood estimation unreliable.

Deep incomplete multi-view clustering (IMVC) has addressed this problem through feature imputation, cross-view prediction, prototype alignment, and imputation-free fusion. COMPLETER predicts missing views with cross-view contrastive objectives (Lin et al., 2021); CPSPAN aligns partially observed samples and shifted prototypes (Jin et al., 2023); latent-graph imputation reconstructs missing features from cluster assignments and global centroids (Pu et al., 2024); and Mask-IMvC suppresses unavailable inputs while injecting observed-view graphs into contrastive learning (Li et al., 2025). Recent methods further transfer global graph information (Chao et al., 2025), learn unified similarity graphs for consistency-guided recovery (Jiang et al., 2025), or reconstruct missing distributions (Jin et al., 2025). Despite these different designs, they commonly transfer cross-view information at a relatively fine granularity, through individual instances or features, isolated prototypes, or pairwise relations in dense graphs. Such fine-grained transfer is sensitive to noise because local perturbations can directly corrupt the units or relations being propagated. Moreover, when cross-view overlap is sparse, limited co-observed evidence makes the correspondence between these units increasingly unreliable.

The central unresolved problem is therefore how to recover reliable neighborhood structure for clustering when local observations are noisy and co-observed samples are sparse. Fine-grained units then provide insufficient structural evidence, motivating an intermediate representation that aggregates local evidence while preserving neighborhood organization. Granular-ball regions offer this mesoscopic scale: coherent regional aggregation suppresses isolated perturbations, while sparse

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.

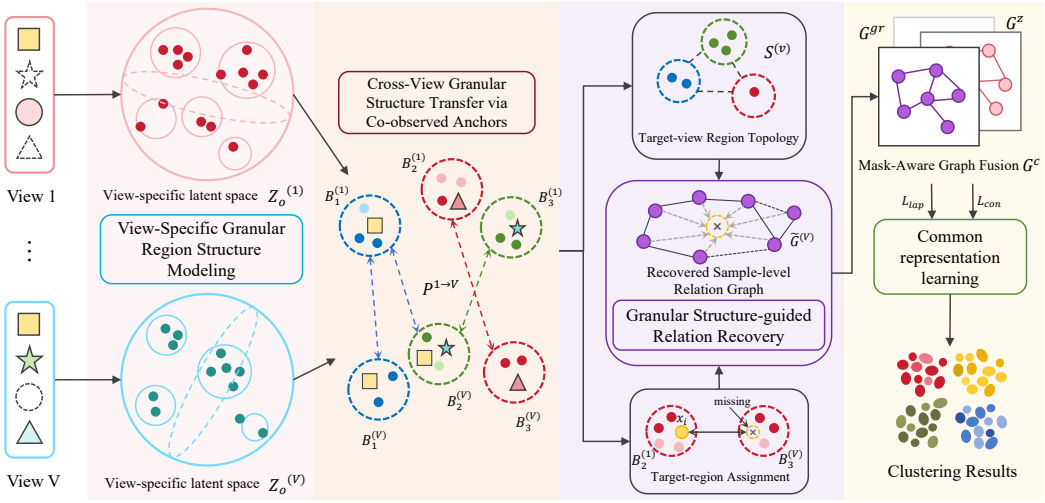

Figure 1: Overview of CGBST. Co-observed samples align view-specific granular-ball structures to recover missing relations for common representation learning.

inter-region topology retains neighborhood relations. However, incomplete views induce region systems with unequal cardinalities that can be related only through co-observed samples.

To address these challenges, we propose Cross-View Granular-Ball Structure Transfer (CGBST), which transfers mesoscopic structure without synthesizing target-view features. Each encoder maps observed samples into a latent space, where a quality-driven divisive procedure forms granular-ball regions. Co-observed samples define row-normalized co-occurrence maps between source and target regions, accommodating unequal region sets and many-to-many associations. For a target-missing sample, donor-view affiliations are transported through these maps and aggregated into a target-region distribution. Same-region affinity and target-region topology then recover its sample relations.

Our contributions are summarized as follows:

- Unlike IMVC methods that transfer instances or pairwise relations, we propose CGBST, a region-mediated framework that aggregates locally coherent samples into adaptive regions with sparse topology, reducing reliance on fragile instance-level correspondences.
- We develop co-observed-anchor-driven structure transfer and relation recovery. Row-normalized co-occurrence maps handle unequal region sets and many-to-many correspondences. Transferred soft affiliations and target topology recover missing relations without feature synthesis.
- CGBST uses recovered relations for mask-aware contrastive learning and Laplacian smoothing. Results on four benchmarks and multiple missing rates show competitive performance.

2 Related Work

2.1 Graph-based incomplete multi-view clustering

Graph-based IMVC avoids explicit feature synthesis by completing adaptive affinities (Wen et al., 2021), transferring cross-view relations (Wang et al., 2023), constructing bipartite graphs (Wang et al., 2022), or coupling graph completion with clustering (Yang et al., 2024). Deep variants integrate representation learning with fusion graphs (Wen et al., 2020), mask-aware contrast (Li et al., 2025), global propagation (Chao et al., 2025), robust graph contrast (Zhuang et al., 2025), and masked-graph reconstruction (Zhang et al., 2026). However, their fine-grained sample or anchor connections can propagate local errors when neighborhoods are noisy or cross-view overlap is sparse.

2.2 GRANULAR-BALL LEARNING

Granular-ball learning replaces points with adaptive regions. Recent methods improve region construction and clustering through coverage–specificity partitioning (Jia et al., 2025), task-adaptive multi-granularity neighborhoods (Xie et al., 2024b), local feature weighting (Xie et al., 2024a), and inter-ball relation clustering (Xia et al., 2025). In complete multi-view clustering, MGBCC associates within- and cross-view balls (Su et al., 2025), GFSAF separates mutual and complementary representations through fuzzy ball splitting (Liu et al., 2026), and HOARD performs ball-level contrastive alignment within hierarchical distillation (Zhou et al., 2026). These methods establish granular balls as mesoscopic units but do not infer target-view regions or recover missing sample relations under incomplete observations.

3 METHODOLOGY

3.1 PRELIMINARY AND PROBLEM SETUP

Let $\{\boldsymbol{X}^{(v)}\}_{v=1}^{V}$ be a multi-view dataset with N samples, where $\boldsymbol{X}^{(v)} \in \mathbb{R}^{N\times d_v}$. The observation mask $\boldsymbol{M} \in \{0,1\}^{N\times V}$ has $M_{iv} = 1$ if sample i is present in view v. We denote the observed index set by $\mathcal{O}_v = \{i : M_{iv} = 1\}$ and the pairwise co-observed set by $\mathcal{O}_{uv} = \mathcal{O}_u \cap \mathcal{O}_v$.

Let $\boldsymbol{T}^{(v)} \in \{0,1\}^{N\times|\mathcal{O}_v|}$ select the observed rows, giving $\boldsymbol{X}_o^{(v)} = \boldsymbol{T}^{(v)\top}\boldsymbol{X}^{(v)}$. Following Mask-IMvC (Li et al., 2025), we construct an adaptive k_{sam}-nearest-neighbor graph only on these rows. With $d_{ij}^{(v)} = \|\boldsymbol{x}_i^{(v)} - \boldsymbol{x}_j^{(v)}\|_2^2$ and k_{sam}-th neighbor distance $d_{i(k_{\mathrm{sam}})}^{(v)}$, its directed weight is

$$G_{o,ij}^{(v)} = \begin{cases} \dfrac{d_{i(k_{\mathrm{sam}})}^{(v)} - d_{ij}^{(v)}}{k_{\mathrm{sam}} d_{i(k_{\mathrm{sam}})}^{(v)} - \sum_{j'\in\mathcal{N}_{k_{\mathrm{sam}}}^{(v)}(i)} d_{ij'}^{(v)} + \epsilon}, & j \in \mathcal{N}_{k_{\mathrm{sam}}}^{(v)}(i), \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

followed by symmetrization and a unit diagonal. The view-specific encoder and the corresponding N-row latent matrix are

$$\boldsymbol{Z}_o^{(v)} = f_v(\boldsymbol{X}_o^{(v)}, \boldsymbol{G}_o^{(v)}), \qquad \bar{\boldsymbol{Z}}^{(v)} = \boldsymbol{T}^{(v)}\boldsymbol{Z}_o^{(v)}. \tag{2}$$

For later relation operations in the common N-sample space, the observed graph is lifted as

$$\bar{\boldsymbol{G}}_0^{(v)} = \boldsymbol{T}^{(v)}\boldsymbol{G}_o^{(v)}\boldsymbol{T}^{(v)\top}, \qquad \bar{G}_{0,ij}^{(v)} = 0 \quad \text{if } M_{iv}M_{jv} = 0. \tag{3}$$

Missing rows are zero-filled only after encoding and never participate in constructing or propagating $\boldsymbol{G}_o^{(v)}$.

Figure 1 summarizes the pipeline. Observed embeddings form view-specific granular-ball regions and topology graphs. Co-observed samples establish cross-view region correspondences, which infer target-region assignments for missing samples; the target topology then recovers sample relations for common representation learning. Thus, transfer is performed at the region level while clustering remains at the sample level.

3.2 VIEW-SPECIFIC GRANULAR REGION STRUCTURE MODELING

The first stage of Figure 1 partitions the detached observed embeddings $\boldsymbol{Z}_o^{(v)}$ into adaptive regions $\mathcal{B}^{(v)} = \{B_k^{(v)}\}_{k=1}^{K_v}$. Region B has center and mean radius

$$\boldsymbol{\mu}_B = \frac{1}{|B|}\sum_{i\in B}\boldsymbol{z}_i^{(v)}, \qquad r(B) = \frac{1}{|B|}\sum_{i\in B}\|\boldsymbol{z}_i^{(v)} - \boldsymbol{\mu}_B\|_2. \tag{4}$$

Its quality balances coverage and compactness:

$$q(B) = \frac{|B|}{|B| + m_{\min}} \exp\left(-\frac{r(B)}{r(B_{\mathrm{root}}) + \epsilon}\right). \tag{5}$$

This favors sufficiently large, compact regions. A candidate is bisected using two approximate farthest pivots:

$$p_1(B) = \arg\max_{i \in B} \|\boldsymbol{z}_i^{(v)} - \boldsymbol{\mu}_B\|_2, \tag{6}$$

$$p_2(B) = \arg\max_{i \in B} \|\boldsymbol{z}_i^{(v)} - \boldsymbol{z}_{p_1(B)}^{(v)}\|_2, \tag{7}$$

$$B_l = \left\{ i \in B : \|\boldsymbol{z}_i^{(v)} - \boldsymbol{z}_{p_1(B)}^{(v)}\|_2^2 \le \|\boldsymbol{z}_i^{(v)} - \boldsymbol{z}_{p_2(B)}^{(v)}\|_2^2 \right\}, \qquad B_r = B \setminus B_l. \tag{8}$$

For the resulting children, the split gain is

$$\Delta q = \frac{|B_l|}{|B|} q(B_l) + \frac{|B_r|}{|B|} q(B_r) - q(B). \tag{9}$$

At iteration t, the algorithm selects

$$B_t^* = \arg\max_{B \in \mathcal{B}_t^{(v)}} \Delta q(B) \quad \text{s.t.} \quad |B_l|, |B_r| \ge m_{\min}, \quad \operatorname{depth}(B) < d_{\max}, \tag{10}$$

and accepts the split only when $\Delta q(B_t^*) > \delta$ and $|\mathcal{B}_t^{(v)}| < K_{\max}$. Thus K_v adapts to the observed geometry instead of being fixed to the number of semantic clusters.

Cross-view support uses the hard membership produced by the divisive construction:

$$\bar{A}_{ik}^{(v)} = \begin{cases} \mathbb{I}[i \in B_k^{(v)}], & i \in \mathcal{O}_v, \\ 0, & i \notin \mathcal{O}_v. \end{cases} \tag{11}$$

Relation recovery uses a soft affiliation $\boldsymbol{A}^{(v)} \in \mathbb{R}^{N \times K_v}$. With $D_{ik}^{(v)} = \|\boldsymbol{z}_i^{(v)} - \boldsymbol{\mu}_k^{(v)}\|_2$ and $\mathcal{N}_{k_{\text{aff}}}^{(v)}(i)$ the k_{aff} nearest region centers, the bandwidth is their global mean distance,

$$\sigma_v = \frac{1}{k_{\text{aff}} |\mathcal{O}_v|} \sum_{i \in \mathcal{O}_v} \sum_{k \in \mathcal{N}_{k_{\text{aff}}}^{(v)}(i)} D_{ik}^{(v)}. \tag{12}$$

The sparse Gaussian affiliation is

$$w_{ik}^{(v)} = \mathbb{I}[k \in \mathcal{N}_{k_{\text{aff}}}^{(v)}(i)] \exp\left(-\frac{(D_{ik}^{(v)})^2}{2\sigma_v^2 + \epsilon} \right), \tag{13}$$

$$A_{ik}^{(v)} = \begin{cases} \dfrac{w_{ik}^{(v)}}{\sum_{a=1}^{K_v} w_{ia}^{(v)} + \epsilon}, & i \in \mathcal{O}_v, \\ 0, & i \notin \mathcal{O}_v. \end{cases} \tag{14}$$

To preserve region adjacency, we connect the k_{topo} nearest centers $\mathcal{N}_{k_{\text{topo}}}^{(v)}(a)$ in an undirected binary graph and row-normalize it:

$$J_{ab}^{(v)} = \mathbb{I}[b \in \mathcal{N}_{k_{\text{topo}}}^{(v)}(a) \ \vee \ a \in \mathcal{N}_{k_{\text{topo}}}^{(v)}(b)], \qquad S_{ab}^{(v)} = \frac{J_{ab}^{(v)}}{\sum_{c=1}^{K_v} J_{ac}^{(v)} + \epsilon}, \qquad J_{aa}^{(v)} = 0. \tag{15}$$

Thus, $(\boldsymbol{A}^{(v)}, \boldsymbol{S}^{(v)})$ retains membership and mesoscopic topology; hard affiliation provides anchor counts, while soft affiliation preserves assignment uncertainty during transfer and recovery.

The two affiliations play deliberately different roles. Divisive membership prevents diffuse weights from blurring the co-observed anchor counts, whereas the Gaussian affiliation lets samples near region boundaries retain multiple plausible memberships. The region graph supplies a complementary neighborhood prior between these memberships. Consequently, cross-view mappings are estimated from discrete and interpretable co-occurrences, while the subsequent transfer remains uncertainty-aware. This separation also allows every view to construct its regions independently, without requiring equal region cardinalities or a one-to-one match across views.

3.3 CROSS-VIEW GRANULAR STRUCTURE TRANSFER VIA CO-OBSERVED ANCHORS

We transfer region affiliations rather than features or instance edges. Co-observed samples in source view u and target view v define

$$H_{ab}^{u\to v} = \sum_{i=1}^{N} M_{iu} M_{iv} \bar{A}_{ia}^{(u)} \bar{A}_{ib}^{(v)}, \qquad P_{ab}^{u\to v} = \frac{H_{ab}^{u\to v}}{\sum_{c=1}^{K_v} H_{ac}^{u\to v} + \epsilon}. \tag{16}$$

Equivalently, $\boldsymbol{H}^{u\to v} = (\bar{\boldsymbol{A}}_{\mathcal{O}_{uv}}^{(u)})^\top \bar{\boldsymbol{A}}_{\mathcal{O}_{uv}}^{(v)}$ counts shared anchors. Row normalization yields a soft $K_u \times K_v$ mapping that supports unequal region sets and one-to-many associations; unsupported source regions give zero rows. Writing $n_a^{u\to v} = \sum_{b=1}^{K_v} H_{ab}^{u\to v}$,

$$\sum_{b=1}^{K_v} P_{ab}^{u\to v} = \frac{n_a^{u\to v}}{n_a^{u\to v} + \epsilon} \approx 1, \qquad n_a^{u\to v} > 0. \tag{17}$$

Hence, each source region distributes mass over compatible target regions without one-to-one matching. We use this raw mapping without confidence scoring, thresholding, or reweighting.

For sample i observed in u but missing in v, the donor estimate is

$$\hat{\boldsymbol{a}}_i^{u\to v} = M_{iu}(1 - M_{iv})\, \boldsymbol{a}_i^{(u)} \boldsymbol{P}^{u\to v}. \tag{18}$$

Here $\boldsymbol{a}_i^{(u)}$ is the soft source-region affiliation. All eligible samples can be transferred jointly as

$$\widehat{\boldsymbol{A}}^{u\to v} = \boldsymbol{D}_{\text{elig}}^{u\to v} \boldsymbol{A}^{(u)} \boldsymbol{P}^{u\to v}, \qquad \boldsymbol{D}_{\text{elig}}^{u\to v} = \operatorname{diag}(\boldsymbol{M}_{:u} \odot (\boldsymbol{1} - \boldsymbol{M}_{:v})). \tag{19}$$

The eligibility mask $\boldsymbol{D}_{\text{elig}}^{u\to v}$ excludes unobserved source rows and observed target rows. Valid donors observe sample i and yield nonzero mapped mass:

$$\mathcal{D}_i^{(v)} = \{u \neq v : M_{iu} = 1,\ M_{iv} = 0,\ \|\hat{\boldsymbol{a}}_i^{u\to v}\|_1 > 0\}. \tag{20}$$

The completed target-region affiliation is

$$\tilde{\boldsymbol{a}}_i^{(v)} = \begin{cases} \boldsymbol{a}_i^{(v)}, & M_{iv} = 1, \\ \dfrac{\sum_{u\in\mathcal{D}_i^{(v)}} \hat{\boldsymbol{a}}_i^{u\to v}}{\left\|\sum_{u\in\mathcal{D}_i^{(v)}} \hat{\boldsymbol{a}}_i^{u\to v}\right\|_1 + \epsilon}, & M_{iv} = 0,\ |\mathcal{D}_i^{(v)}| > 0, \\ \boldsymbol{0}, & M_{iv} = 0,\ |\mathcal{D}_i^{(v)}| = 0. \end{cases} \tag{21}$$

The donor-average factor $1/|\mathcal{D}_i^{(v)}|$ cancels under normalization. Observed affiliations remain unchanged, while missing samples aggregate all valid donors without reconstructing $\boldsymbol{x}_i^{(v)}$.

3.4 GRANULAR STRUCTURE-GUIDED RELATION RECOVERY

After target-region assignment, regional evidence is projected to samples. The same-region term measures membership overlap, while the topology term connects adjacent regions:

$$\boldsymbol{Q}_{\text{same}}^{(v)} = \tilde{\boldsymbol{A}}^{(v)} \tilde{\boldsymbol{A}}^{(v)\top}, \qquad \boldsymbol{Q}_{\text{topo}}^{(v)} = \tilde{\boldsymbol{A}}^{(v)} \boldsymbol{S}^{(v)} \tilde{\boldsymbol{A}}^{(v)\top}. \tag{22}$$

Because the row-normalized region graph may be asymmetric, we symmetrize the combined relation:

$$\boldsymbol{Q}^{(v)} = \frac{\boldsymbol{Q}_{\text{same}}^{(v)} + \beta \boldsymbol{Q}_{\text{topo}}^{(v)}}{1 + \beta}, \qquad \boldsymbol{R}^{(v)} = \frac{\boldsymbol{Q}^{(v)} + \boldsymbol{Q}^{(v)\top}}{2}, \qquad R_{ii}^{(v)} = 1. \tag{23}$$

For $i \neq j$, its element-wise form is

$$R_{ij}^{(v)} = \frac{\langle \tilde{\boldsymbol{a}}_i^{(v)}, \tilde{\boldsymbol{a}}_j^{(v)} \rangle + \frac{\beta}{2} \left(\tilde{\boldsymbol{a}}_i^{(v)} \boldsymbol{S}^{(v)} \tilde{\boldsymbol{a}}_j^{(v)\top} + \tilde{\boldsymbol{a}}_j^{(v)} \boldsymbol{S}^{(v)} \tilde{\boldsymbol{a}}_i^{(v)\top} \right)}{1 + \beta}. \tag{24}$$

A pair is therefore supported by shared-region membership or region adjacency, with the bidirectional term removing row-normalization bias.

We sparsify $\boldsymbol{R}^{(v)}$ according to observation status. Let $\mathrm{TopK}_{k_{\mathrm{rel}}}(\boldsymbol{\rho};\mathcal{C})$ select the k_{rel} largest positive entries of $\boldsymbol{\rho}$ within $\mathcal{C}$. For a target-missing row,

$$G^{(v)}_{\mathrm{mis},ij} = R^{(v)}_{ij}\mathbb{I}\Big[M_{iv}=0,\ j\in \mathrm{TopK}_{k_{\mathrm{rel}}}(R^{(v)}_{i:};\{1,\dots,N\}\setminus\{i\})\Big]. \tag{25}$$

After symmetrization, observed samples may receive edges selected by missing samples. For target-observed rows, granular structure only reinforces existing observed-view edges:

$$G^{(v)}_{\mathrm{obs},ij} = R^{(v)}_{ij}\mathbb{I}[\bar{G}^{(v)}_{0,ij}>0]\mathbb{I}\Big[M_{iv}M_{jv}=1,\ j\in \mathrm{TopK}_{k_{\mathrm{rel}}}(R^{(v)}_{i:};\mathcal{O}_v\setminus\{i\})\Big]. \tag{26}$$

Writing $\mathrm{SymMax}(\boldsymbol{B}) = \max(\boldsymbol{B},\boldsymbol{B}^\top)$ element-wise, the recovered graph for view v is

$$\tilde{\boldsymbol{G}}^{(v)} = \mathrm{SymMax}(\boldsymbol{G}^{(v)}_{\mathrm{mis}}) \vee \mathrm{SymMax}(\boldsymbol{G}^{(v)}_{\mathrm{obs}}), \qquad \tilde{G}^{(v)}_{ii}=1, \tag{27}$$

where $\vee$ is the element-wise maximum. Consequently,

$$\mathbb{I}[\tilde{G}^{(v)}_{ij}>0] \le \mathbb{I}[\bar{G}^{(v)}_{0,ij}>0] \quad \text{whenever } M_{iv}M_{jv}=1, \tag{28}$$

so no new observed–observed edge is introduced; completion is confined to missingness-affected relations. With $E^{(v)}_{ij} = \mathbb{I}[\tilde{G}^{(v)}_{ij}>0]$, recovered graphs are averaged only over existing edges:

$$G^{\mathrm{gr}}_{ij} = \frac{\sum_{v=1}^{V} E^{(v)}_{ij}\tilde{G}^{(v)}_{ij}}{\sum_{v=1}^{V} E^{(v)}_{ij}+\epsilon}, \qquad E^{\mathrm{gr}}_{ij} = \mathbb{I}[G^{\mathrm{gr}}_{ij}>0]. \tag{29}$$

3.5 Common representation learning

Recovered relations are injected into the clustering backbone. Available view-specific embeddings are aggregated into a common space:

$$\boldsymbol{h}_i = \frac{\sum_{v=1}^{V} M_{iv}\bar{\boldsymbol{z}}^{(v)}_i}{\sum_{v=1}^{V} M_{iv}}, \qquad \boldsymbol{z}_i = \boldsymbol{W}_f\boldsymbol{h}_i+\boldsymbol{b}_f, \qquad \boldsymbol{y}_i = \boldsymbol{W}_c\boldsymbol{z}_i+\boldsymbol{b}_c. \tag{30}$$

Let $\boldsymbol{Z}=[\boldsymbol{z}_1,\dots,\boldsymbol{z}_N]^\top$ and $\boldsymbol{Y}=[\boldsymbol{y}_1,\dots,\boldsymbol{y}_N]^\top$. With $\boldsymbol{m}^{(v)}=\boldsymbol{M}_{:v}$, $\boldsymbol{C}^{(v)}=\boldsymbol{m}^{(v)}\boldsymbol{m}^{(v)\top}$ marks jointly observed pairs, and $\boldsymbol{G}^z$ is the adaptive k_{sam}-nearest-neighbor graph of $\boldsymbol{Z}$. Their normalized fusion is

$$\boldsymbol{G}^c = \frac{\boldsymbol{G}^z+\sum_{v=1}^{V}\boldsymbol{C}^{(v)}\odot\bar{\boldsymbol{G}}^{(v)}_0+\lambda_g\boldsymbol{E}^{\mathrm{gr}}\odot\boldsymbol{G}^{\mathrm{gr}}}{\boldsymbol{1}\boldsymbol{1}^\top+\sum_{v=1}^{V}\boldsymbol{C}^{(v)}+\lambda_g\boldsymbol{E}^{\mathrm{gr}}}, \tag{31}$$

where $\boldsymbol{E}^{\mathrm{gr}}$ masks absent recovered edges. For ℓ_2-normalized outputs $\bar{\boldsymbol{y}}_i=\boldsymbol{y}_i/\|\boldsymbol{y}_i\|_2$ and $\kappa_{ij}=\exp(\bar{\boldsymbol{y}}_i^\top\bar{\boldsymbol{y}}_j/\tau)$, the fused graph defines the graph-guided contrastive objective:

$$\mathcal{L}_{\mathrm{con}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\sum_{j\neq i}G^c_{ij}\kappa_{ij}}{\sum_{j\neq i}(1-G^c_{ij})\kappa_{ij}}, \qquad \tau=1. \tag{32}$$

The recovered graph also regularizes the common embedding. After removing the diagonal, let $\boldsymbol{W}_{\mathrm{lap}}=(\boldsymbol{G}^{\mathrm{gr}}+\boldsymbol{G}^{\mathrm{gr}\top})/2$, $\boldsymbol{D}_{\mathrm{lap}}=\mathrm{diag}(\boldsymbol{W}_{\mathrm{lap}}\boldsymbol{1})$, and $\boldsymbol{L}=\boldsymbol{D}_{\mathrm{lap}}-\boldsymbol{W}_{\mathrm{lap}}$. Then

$$\mathcal{L}_{\mathrm{lap}} = \frac{\mathrm{Tr}(\boldsymbol{Z}^\top\boldsymbol{L}\boldsymbol{Z})}{d_z\sum_{ij}W_{\mathrm{lap},ij}+\epsilon} = \frac{\sum_{i,j}W_{\mathrm{lap},ij}\|\boldsymbol{z}_i-\boldsymbol{z}_j\|_2^2}{2\left(d_z\sum_{ij}W_{\mathrm{lap},ij}+\epsilon\right)}. \tag{33}$$

Thus, recovered structure guides both contrastive relations and local smoothness. Decoder ψ_v reconstructs only observed samples:

$$\mathcal{L}_{\mathrm{rec}} = \sum_{v=1}^{V}\frac{1}{|\mathcal{O}_v|d_v}\sum_{i\in\mathcal{O}_v}\left\|\boldsymbol{x}^{(v)}_i-\psi_v(\boldsymbol{z}_i)\right\|_2^2. \tag{34}$$

The overall objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{rec}}+10^{\alpha}\mathcal{L}_{\mathrm{con}}+\lambda_{\mathrm{lap}}\mathcal{L}_{\mathrm{lap}}. \tag{35}$$

After optimization, the common embeddings are partitioned into C clusters by

$$(\hat{\boldsymbol{c}},\hat{\boldsymbol{U}}) = \arg\min_{\substack{c_i\in\{1,\dots,C\}\\ \boldsymbol{U}=[\boldsymbol{u}_1,\dots,\boldsymbol{u}_C]}}\sum_{i=1}^{N}\|\boldsymbol{z}_i-\boldsymbol{u}_{c_i}\|_2^2, \tag{36}$$

where $\hat{c}_i$ is the final assignment. After warm-up, granular structures are periodically rebuilt from detached features, avoiding discrete reconstruction at every gradient step.

Complexity. Let $K = \max_v K_v$ and $N_{uv} = |\mathcal{O}_{uv}|$. Affiliation, cross-view support, and dense relation recovery cost $O(VNKd_z)$, $O(\sum_{u\neq v} N_{uv}K_uK_v)$, and $O(\sum_v N^2K_v)$, with $O(N^2+\sum_v NK_v+\sum_v K_v^2)$ storage. The overall order remains quadratic in the sample graph, while the candidate-edge implementation avoids materializing all recovered pairs.

4 Experiments and Analysis

4.1 Experimental setup

We evaluate on four multi-view benchmarks: MSRCV1 (210 samples, five views, seven classes), CUB (600 samples, two views, ten classes), OutdoorScene (2,688 samples, four views, eight classes), and NUS-WIDE (3,000 samples, two views, ten classes). CUB and NUS-WIDE originate from established image benchmarks (Wah et al., 2011; Chua et al., 2009).

Incomplete multi-view data construction. For a benchmark with V views, the mask generator sets the target entry-wise observation rate to $1-\eta/V$, where $\eta \in \{0.1, 0.3, 0.5\}$. It first preserves one uniformly sampled view for every instance and then samples the remaining entries until the realized observation rate differs from the target by less than 0.005. Thus, every instance retains at least one observed view and the mask contains approximately $N\eta$ missing view entries in total. We generate five masks per rate using seeds 1–5, average the corresponding results, and min–max normalize each view.

Unless specified otherwise, we use $m_{\min} = 8$, $K_{\max} = d_{\max} = 20$, and $\delta = 10^{-4}$; soft affiliations, region graphs, and recovered relations use 2, 3, and 10 neighbors, respectively, with $\beta = 1$. We set $\lambda_g = 0.2$ and $\lambda_{\text{lap}} = 0.1$ and use raw anchor-derived correspondences without reliability filtering or semantic calibration.

We compare with DCCA (Andrew et al., 2013), DSIMVC++ (Tang & Liu, 2022; Yan et al., 2023), CPSPAN (Jin et al., 2023), GIGA (Yang et al., 2024), SCM (Luo et al., 2024), DCG (Zhang et al., 2025), Mask-IMvC (Li et al., 2025), GFSAF (Liu et al., 2026), SMHGC (Chen et al., 2026), and DIMVC-HIA (Du et al., 2026). The model is trained for 1,500 epochs with Adam (Kingma & Ba, 2015) at a learning rate of 10^{-3}; granular structures are activated after 100 epochs and updated every 20 epochs. We report ACC, NMI, ARI, and F-score after k-means with ten fixed initializations. For $\eta = (0.1, 0.3, 0.5)$, α is $(0, 0, -1)$, $(-1, -1, -1)$, $(0, -1, -1)$, and $(2, -3, 3)$ for MSRCV1, CUB, OutdoorScene, and NUS-WIDE, respectively.

4.2 Performance Comparison

Table 1 shows that CGBST ranks first in 42 of the 48 settings. Specifically, CGBST achieves the best result in all 12 MSRCV1 settings, 11 of 12 CUB settings, 10 of 12 NUS-WIDE settings, and 9 of 12 OutdoorScene settings. This consistency follows from region-level recovery: co-observed anchors pool noisy pairwise matches into region correspondences, while target-view topology supplies relations beyond exact region co-membership. The row-normalized correspondence accommodates unequal granular partitions by expressing each source region as a distribution over target regions. Aggregating these distributions across the views observed for a sample suppresses donor-specific errors before the target topology projects regional evidence back to sample relations. Observation-aware sparsification retains valid observed relations, and the recovered edges couple contrastive separation with Laplacian smoothness. Multi-donor aggregation is most effective when several views remain available.

4.3 Ablation Study

Table 2 separates granular structure transfer from recovered-graph learning. Instance-level transfer reduces MSRCV1 ARI by 5.03 points but changes most CUB metrics by less than one point, indicating that region aggregation is most valuable when view geometries conflict. Removing missing-view recovery drops OutdoorScene ACC by 4.32 points; removing observed-view reinforcement lowers ARI by 4.89 points on MSRCV1 and 4.13 points on NUS-WIDE. Together with the losses without inter-region topology, these results show that target-missing connections, observed target edges, and adjacent-region relations provide complementary evidence.

Table 1: Clustering performance comparison (%) under different missing rates. The best result is highlighted in red bold, and the second-best result is marked with gold bold.

η	Datasets	CUB				MSRCV1				OutdoorScene				NUS-WIDE			
	Methods	ACC	NMI	ARI	F-score	ACC	NMI	ARI	F-score	ACC	NMI	ARI	F-score	ACC	NMI	ARI	F-score
0.1	DCCA (2013)	33.01	28.50	11.02	29.80	46.69	38.94	21.14	44.58	50.01	49.53	34.41	47.58	34.51	24.30	10.78	31.46
	DSIMVC++ (2023)	59.70	57.25	41.07	47.52	44.70	37.65	24.85	36.85	54.89	49.23	35.27	43.63	41.32	35.12	24.15	31.58
	CPSPAN (2023)	63.37	62.80	48.18	60.86	73.43	66.40	56.24	70.87	57.96	49.55	37.15	57.83	39.47	28.74	19.00	39.25
	GIGA (2024)	73.77	**77.54**	63.42	71.96	74.48	76.69	62.52	76.14	46.57	54.58	35.13	52.72	30.61	30.54	7.18	41.01
	SCM (2024)	**79.37**	73.75	64.10	69.01	76.95	67.85	60.35	67.63	59.73	48.46	37.72	47.26	56.43	41.29	34.69	41.80
	DCG (2025)	74.33	72.62	60.56	72.30	57.62	48.57	35.42	57.32	67.18	55.45	47.23	**65.88**	55.45	**52.52**	38.52	45.46
	Mask-IMvC (2025)	78.95	77.31	**65.62**	**72.37**	**89.69**	**80.01**	**77.20**	**81.63**	**74.20**	**61.62**	**51.96**	61.04	**60.67**	46.15	**39.64**	**46.74**
	GFSAF (2026)	59.33	62.49	49.16	56.40	50.00	32.97	23.24	46.13	65.96	54.43	47.45	**64.63**	49.83	47.02	34.00	43.13
	SMHGC (2026)	58.33	61.83	50.32	53.39	68.57	59.83	49.31	64.57	53.50	47.64	36.44	50.78	50.13	33.31	36.29	32.76
	DIMVC-HIA (2026)	69.23	65.79	51.48	56.53	67.90	60.26	48.59	55.77	70.50	56.37	47.62	54.25	42.63	50.12	29.69	40.10
	CGBST (ours)	**86.62**	**82.20**	**74.50**	**78.44**	**94.68**	**88.11**	**86.30**	**89.51**	**75.18**	**61.70**	**52.79**	61.26	**67.95**	**61.97**	**44.03**	**62.95**
0.3	DCCA (2013)	19.31	12.90	1.83	24.31	22.10	12.65	1.51	31.81	49.99	46.92	32.27	45.53	20.31	7.42	1.79	20.15
	DSIMVC++ (2023)	61.77	58.04	42.58	48.88	52.10	43.77	32.54	42.97	53.92	48.83	34.42	42.78	45.41	37.04	26.84	34.22
	CPSPAN (2023)	60.37	59.56	43.70	59.62	71.24	65.47	54.22	68.34	56.05	48.13	35.32	56.02	39.67	28.63	19.44	39.25
	GIGA (2024)	71.47	**74.48**	**60.16**	69.15	70.76	**73.16**	56.96	72.55	43.65	49.82	28.52	51.73	28.57	26.90	6.15	38.33
	SCM (2024)	68.63	63.79	50.93	58.65	80.10	66.44	60.54	67.77	59.51	47.66	36.72	46.29	47.21	33.45	24.57	**53.75**
	DCG (2025)	73.87	70.57	58.95	**72.73**	55.43	47.04	34.21	54.77	**69.97**	56.10	**47.99**	**70.06**	51.21	**45.39**	32.24	43.15
	Mask-IMvC (2025)	**73.91**	71.60	57.89	66.94	**82.62**	72.31	**65.80**	**73.09**	67.58	**56.93**	46.71	57.22	**53.33**	40.48	28.82	40.21
	GFSAF (2026)	45.00	48.95	30.33	45.12	40.95	25.80	17.82	36.04	38.50	34.87	28.53	33.35	46.03	39.78	28.62	36.47
	SMHGC (2026)	44.33	47.36	30.73	43.17	71.90	62.12	54.02	69.56	47.73	42.85	31.63	43.56	41.80	28.94	26.15	27.65
	DIMVC-HIA (2026)	63.67	59.02	44.42	50.09	67.24	61.54	49.94	56.92	56.62	50.44	38.50	46.29	53.10	39.13	**32.72**	45.14
	CGBST (ours)	**77.37**	**76.64**	**65.51**	**73.01**	**93.40**	**85.93**	**84.40**	**87.32**	**74.13**	**60.79**	**52.40**	**60.50**	**55.04**	**53.08**	**34.53**	**52.29**
0.5	DCCA (2013)	14.01	5.30	0.29	20.23	19.08	7.44	0.71	28.58	47.49	42.76	29.95	42.57	14.24	2.17	0.18	18.11
	DSIMVC++ (2023)	60.30	56.93	40.83	47.29	47.33	38.09	26.60	38.28	52.37	47.10	32.94	41.45	42.70	34.24	24.45	31.68
	CPSPAN (2023)	61.27	61.52	45.93	59.29	70.09	62.06	50.84	67.23	56.98	47.45	35.47	56.83	42.64	30.34	21.81	42.53
	GIGA (2024)	69.30	**70.35**	54.35	64.85	72.19	**70.13**	54.00	**70.11**	48.57	**54.97**	34.06	54.36	33.51	30.91	7.23	41.27
	SCM (2024)	61.83	59.83	45.41	52.21	73.71	58.33	50.30	59.84	59.19	48.32	38.49	47.36	41.76	28.57	15.88	29.80
	DCG (2025)	**72.20**	67.61	**55.26**	**70.86**	54.95	43.29	32.16	53.59	63.45	52.24	42.30	**62.99**	**48.33**	38.17	**26.83**	41.75
	Mask-IMvC (2025)	68.46	68.39	52.74	63.05	**80.25**	69.42	**60.73**	70.08	60.84	49.47	38.51	50.19	45.51	34.12	19.13	34.96
	GFSAF (2026)	35.33	41.37	20.75	35.89	29.05	12.47	5.80	27.56	33.26	25.83	17.34	32.20	35.37	32.81	17.29	29.37
	SMHGC (2026)	32.50	33.75	16.74	31.27	55.24	49.81	38.66	50.84	34.49	25.83	16.55	29.66	31.83	21.65	13.27	22.53
	DIMVC-HIA (2026)	57.77	52.51	37.32	43.75	64.29	57.59	45.74	53.31	**67.82**	54.23	**45.30**	52.21	46.10	**38.48**	**30.64**	**43.65**
	CGBST (ours)	**72.88**	**71.37**	**60.13**	**67.24**	**91.60**	**82.81**	**80.92**	**84.60**	**73.52**	**59.35**	**50.88**	**59.56**	**48.62**	**45.44**	24.78	**47.17**

Table 2: Ablation study on four datasets under missing rate $\eta = 0.3$. All entries are mean percentages over five runs.

Variant	CUB				MSRCV1				OutdoorScene				NUS-WIDE			
	ACC	NMI	ARI	F-score	ACC	NMI	ARI	F-score	ACC	NMI	ARI	F-score	ACC	NMI	ARI	F-score
Full CGBST	77.37	76.64	65.51	73.01	93.40	85.93	84.40	87.32	74.13	60.79	52.40	60.50	55.04	53.08	34.53	52.29
(a) Effectiveness of granular-ball structure transfer																
Instance-level transfer	77.77	76.91	66.04	72.19	90.71	82.89	79.37	83.84	73.63	60.25	52.06	59.87	52.97	53.58	33.50	51.85
w/o missing-view relation recovery	77.76	76.51	65.51	71.65	91.78	84.32	81.47	85.33	69.81	58.53	50.21	59.02	53.49	53.17	33.95	51.94
w/o observed-view reinforcement	76.38	75.15	63.41	69.83	90.74	83.40	79.51	84.02	74.03	60.61	52.39	60.34	53.55	50.53	30.40	50.48
w/o inter-region topology	75.91	75.55	63.51	70.23	91.90	84.64	81.71	85.54	73.20	60.01	51.51	59.52	54.05	53.26	32.80	52.28
(b) Effectiveness of recovered-graph learning																
w/o graph fusion	77.90	76.19	64.83	71.31	88.40	82.39	77.89	82.61	72.99	59.97	51.34	59.39	55.58	53.17	33.15	52.87
w/o Laplacian regularization	71.33	71.59	57.45	65.76	90.09	82.56	78.52	83.30	64.86	54.52	45.52	54.86	54.22	45.14	38.85	48.04

Removing Laplacian regularization degrades every metric on CUB, MSRCV1, and OutdoorScene, including a 9.27-point ACC loss on OutdoorScene; graph fusion has smaller and less consistent effects outside MSRCV1. Fusion reweights training pairs, whereas the Laplacian directly smooths embeddings along recovered edges. NUS-WIDE ARI instead improves without this term, showing that direct smoothing is beneficial overall but not uniformly optimal in the single-donor two-view setting.

4.4 Parameter Sensitivity Analysis

Figure 2 varies $\alpha \in \{-3, -2, \ldots, 3\}$ on MSRCV1, corresponding to a contrastive-loss weight of 10^{α}. Very small weights consistently underperform the best setting, while further increasing the weight does not yield monotonic gains. Under severe missingness, the preferred value shifts toward a smaller weight, indicating that graph-guided contrast should be moderated when fewer observed relations are available.

4.5 Visualization Analysis

Figure 3 shows clear class-aligned blocks on CUB and MSRCV1 and weaker block structure on OutdoorScene and NUS-WIDE. Because class labels only reorder the matrices for display, the

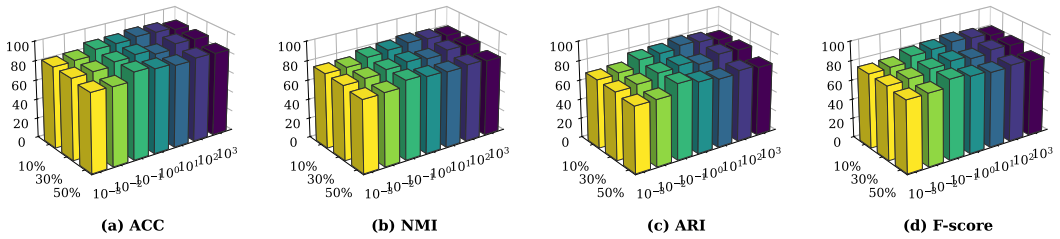

Figure 2: Sensitivity to the contrastive-loss exponent α on MSRCV1 under missing rates $\eta \in \{0.1, 0.3, 0.5\}$. Bars report the five-run mean performance (%).

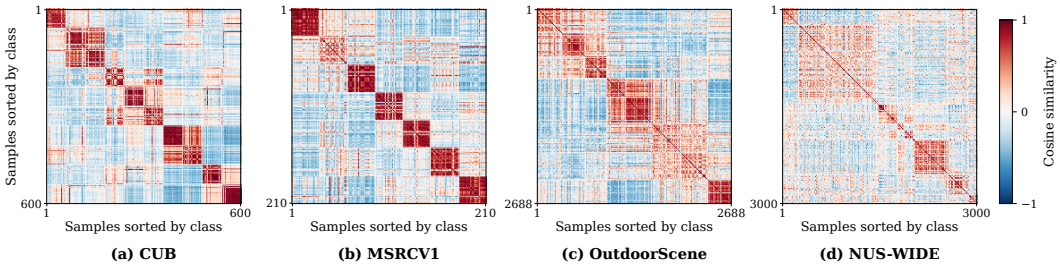

Figure 3: Class-sorted cosine-similarity matrices of the mean-centered final common embeddings under $\eta = 0.3$. Labels are used only for display ordering.

patterns arise from the unsupervised common representations rather than supervised optimization. The remaining off-diagonal responses expose ambiguity on the two more challenging datasets.

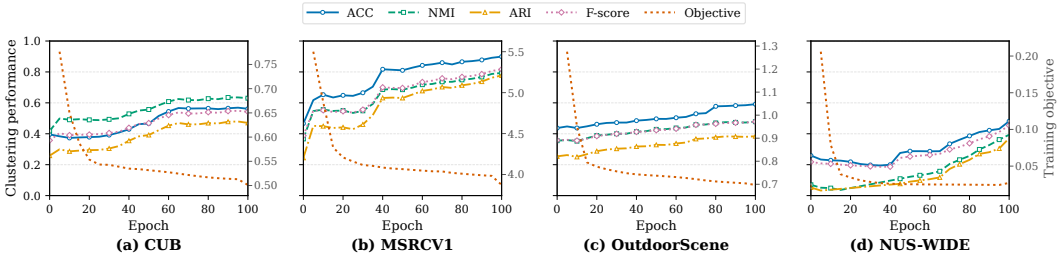

Figure 4: Training loss and clustering metrics over epochs.

Figure 4 shows rapid objective stabilization alongside improving clustering metrics. Across datasets, the metric gains persist as the objective approaches a plateau, so the final quality is not an isolated late-epoch fluctuation. The absence of sustained metric oscillation indicates that the coupled learning objectives refine the embedding without sacrificing training stability.

5 Conclusion

We presented CGBST, a region-mediated IMVC framework that replaces fragile instance-level transfer with mesoscopic granular-ball structure. It relates adaptive view-specific regions through co-observed anchors and projects target-region affiliations to sample relations without synthesizing missing features. The recovered relations provide structural guidance for common representation learning; results on four benchmarks and three missing rates validate the transfer and recovery mechanisms. The current formulation is limited by the quadratic sample-level relation graph and the reduced correspondence evidence under extremely sparse cross-view overlap. Future work will investigate sparse relation recovery and uncertainty-aware region transfer for larger and more severely incomplete datasets.

AI USE STATEMENT

In this work, generative AI tools were used to aid and polish the writing and to support the retrieval and discovery of relevant literature. The authors reviewed all AI-assisted work and take responsibility for the final content.

REPRODUCIBILITY STATEMENT

The complete formulation of CGBST, including granular-region construction, cross-view structure transfer, relation recovery, the learning objective, and complexity, is provided in Section 3. Section 4.1 specifies the data processing, incomplete-view protocol, evaluation procedure, and training configuration. Appendix A documents the dataset dimensions, architecture, optimization settings, and hyperparameters, while Appendix B gives step-by-step algorithm flows; the additional sensitivity and representation analyses are provided in Appendices D and E.

REFERENCES

Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *International Conference on Machine Learning*, volume 28, pp. 1247–1255, 2013. URL https://proceedings.mlr.press/v28/andrew13.html.

Guoqing Chao, Kaixin Xu, Xijiong Xie, and Yongyong Chen. Global graph propagation with hierarchical information transfer for incomplete contrastive multi-view clustering. In *AAAI Conference on Artificial Intelligence*, volume 39, pp. 15713–15721, 2025. doi: 10.1609/aaai.v39i15.33725.

Jianpeng Chen, Yawen Ling, Yazhou Ren, Shufei Zhang, and Lifang He. SMHGC: Homophily-agnostic multi-view heterophilous graph clustering. *Pattern Recognition*, 179:113668, 2026. doi: 10.1016/j.patcog.2026.113668.

Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng. NUS-WIDE: A real-world web image database from national university of singapore. In *ACM International Conference on Image and Video Retrieval*, pp. 1–9, 2009. doi: 10.1145/1646396.1646452.

Yiming Du, Ziyu Wang, Jian Li, Rui Ning, and Lusi Li. Deep incomplete multi-view clustering via hierarchical imputation and alignment. In *AAAI Conference on Artificial Intelligence*, volume 40, pp. 20941–20949, 2026. doi: 10.1609/aaai.v40i25.39235.

Zihang Jia, Zhen Zhang, and Witold Pedrycz. Generation of granular-balls for clustering based on the principle of justifiable granularity. *IEEE Transactions on Cybernetics*, 55(4):1687–1700, 2025. doi: 10.1109/TCYB.2025.3534195.

Bingbing Jiang, Chenglong Zhang, Xinyan Liang, Peng Zhou, Jie Yang, Xingyu Wu, Junyi Guan, Weiping Ding, and Weiguo Sheng. Collaborative similarity fusion and consistency recovery for incomplete multi-view clustering. In *AAAI Conference on Artificial Intelligence*, volume 39, pp. 17617–17625, 2025. doi: 10.1609/aaai.v39i17.33937.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11600–11609, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Jin_Deep_Incomplete_Multi-View_Clustering_With_Cross-View_Partial_Sample_and_Prototype_CVPR_2023_paper.html.

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 *IEEE/CVF International Conference on Computer Vision*, pp. 1016–1026, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Jin_Deep_Incomplete_Multi-view_Clustering_with_Distribution_Dual-Consistency_Recovery_Guidance_ICCV_2025_paper.html.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. URL https://arxiv.org/abs/1412.6980.

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, Yiding Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. COMPLETER: Incomplete multi-view clustering via contrastive prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11174–11183, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Lin_COMPLETER_Incomplete_Multi-View_Clustering_via_Contrastive_Prediction_CVPR_2021_paper.html.

Shuaiyu Liu, Song Wu, Jie Xu, Yazhou Ren, Yang Yang, Xiaorong Pu, and Guoyin Wang. Views attention fusion of granular-ball fuzzy representations split for improved multi-view clustering. In *AAAI Conference on Artificial Intelligence*, volume 40, pp. 23810–23818, 2026. doi: 10.1609/aaai.v40i28.39556.

Caixuan Luo, Jie Xu, Yazhou Ren, Junbo Ma, and Xiaofeng Zhu. Simple contrastive multi-view clustering with data-level fusion. In *International Joint Conference on Artificial Intelligence*, pp. 4697–4705, 2024.

Jingyu Pu, Chenhang Cui, Xinyue Chen, Yazhou Ren, Xiaorong Pu, Zhifeng Hao, Philip S. Yu, and Lifang He. Adaptive feature imputation with latent graph for deep incomplete multi-view clustering. In *AAAI Conference on Artificial Intelligence*, volume 38, pp. 14633–14641, 2024. doi: 10.1609/aaai.v38i13.29378.

Yazhou Ren, Zichen Wen, Junlong Ke, Chenhang Cui, Yonghao Huang, Xinyue Chen, Philip S. Yu, and Lifang He. Enhancing multi-view clustering: A sufficient information-theoretic approach for consistency acquisition and redundancy elimination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48 (7):7420–7435, 2026. doi: 10.1109/TPAMI.2026.3665701.

Peng Su, Shudong Huang, Weihong Ma, Deng Xiong, and Jiancheng Lv. Multi-view granular-ball contrastive clustering. In *AAAI Conference on Artificial Intelligence*, volume 39, pp. 20637–20645, 2025. doi: 10.1609/aaai.v39i19.34274.

Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International Conference on Machine Learning*, volume 162, pp. 21090–21110, 2022. URL https://proceedings.mlr.press/v162/tang22c.html.

Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The Caltech-UCSD birds-200-2011 dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.

Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu, Jiyuan Liu, Sihang Zhou, and En Zhu. Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9776–9785, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Wang_Highly-Efficient_Incomplete_Large-Scale_Multi-View_Clustering_With_Consensus_Bipartite_Graph_CVPR_2022_paper.html.

Yiming Wang, Dongxia Chang, Zhiqiang Fu, Jie Wen, and Yao Zhao. Incomplete multiview clustering via cross-view relation transfer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):367–378, 2023. doi: 10.1109/TCSVT.2022.3201822.

Jie Wen, Zheng Zhang, Zhao Zhang, Zhihao Wu, Lunke Fei, Yong Xu, and Bob Zhang. DIMC-Net: Deep incomplete multi-view clustering network. In *ACM International Conference on Multimedia*, pp. 3753–3761, 2020. doi: 10.1145/3394171.3413726.

Jie Wen, Ke Yan, Zheng Zhang, Yong Xu, Junqian Wang, Lunke Fei, and Bob Zhang. Adaptive graph completion based incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 23:2493–2504, 2021. doi: 10.1109/TMM.2020.3013408.

Song Wu, Yan Zheng, Yazhou Ren, Jing He, Xiaorong Pu, Shudong Huang, Zhifeng Hao, and Lifang He. Self-weighted contrastive fusion for deep multi-view clustering. *IEEE Transactions on Multimedia*, 26: 9150–9162, 2024. doi: 10.1109/TMM.2024.3387298.

Shuyin Xia, Bolun Shi, Yifan Wang, Jiang Xie, Guoyin Wang, and Xinbo Gao. GBCT: Efficient and adaptive clustering via granular-ball computing for complex data. *IEEE Transactions on Neural Networks and Learning Systems*, 36(7):12159–12172, 2025. doi: 10.1109/TNNLS.2024.3497174.

Jiang Xie, Chunfeng Hua, Shuyin Xia, Yuxin Cheng, Guoyin Wang, and Xinbo Gao. W-GBC: An adaptive weighted clustering method based on granular-ball structure. In *IEEE 40th International Conference on Data Engineering*, pp. 914–925, 2024a. doi: 10.1109/ICDE60146.2024.00075.

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, 2024b. doi: 10.1109/TPAMI.2024.3400281.

Weiqing Yan, Yuanyang Zhang, Chenlei Lv, Chang Tang, Guanghui Yue, Liang Liao, and Weisi Lin. GCFAgg: Global and cross-view feature aggregation for multi-view clustering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19863–19872, 2023.

Zequn Yang, Han Zhang, Yake Wei, Zheng Wang, Feiping Nie, and Di Hu. Geometric-inspired graph-based incomplete multi-view clustering. *Pattern Recognition*, 147:110082, 2024. doi: 10.1016/j.patcog.2023.110082.

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 *AAAI Conference on Artificial Intelligence*, volume 39, pp. 22650–22658, 2025. doi: 10.1609/aaai.v39i21.34424.

Zhenghao Zhang, Jun Xie, Xingchen Chen, Tao Yu, Hongzhu Yi, Kaixin Xu, Yuanxiang Wang, Tianyu Zong, Xinming Wang, Jiahuan Chen, Guoqing Chao, Feng Chen, Zhepeng Wang, and Jungang Xu. Dynamic deep graph learning for incomplete multi-view clustering with masked graph reconstruction loss. In *AAAI Conference on Artificial Intelligence*, volume 40, pp. 28600–28608, 2026. doi: 10.1609/aaai.v40i34.40091.

Lihua Zhou, Guowang Du, Kevin Lü, Lizheng Wang, and Jingwei Du. A survey and an empirical evaluation of multi-view clustering approaches. *ACM Computing Surveys*, 56(7):1–38, 2024. doi: 10.1145/3649447.

Taichun Zhou, Siwei Wang, Zhibin Dong, Jiaqi Jin, Ke Liang, Baili Xiao, Miaomiao Li, Xinwang Liu, and En Zhu. Hierarchical cross-view alignment for multi-view clustering via decoupled information distillation. In *AAAI Conference on Artificial Intelligence*, volume 40, pp. 29000–29008, 2026. doi: 10.1609/aaai.v40i34.40136.

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

2025. doi: 10.24963/ijcai.2025/810.

APPENDIX

A IMPLEMENTATION AND REPRODUCIBILITY DETAILS

A.1 DATASETS AND PREPROCESSING

Table 3 summarizes the datasets used in the experiments. Each view is independently transformed to $[0, 1]$ by feature-wise min–max normalization. Missing entries are excluded before constructing the view-specific graph and are never passed through the corresponding encoder.

Table 3: Dataset statistics. The view dimensions follow their order in the released data files.

Dataset	Samples	Views	Classes	View dimensions
MSRCV1	210	5	7	24, 576, 512, 256, 254
CUB	600	2	10	1,024, 300
OutdoorScene	2,688	4	8	512, 432, 256, 48
NUS-WIDE	3,000	2	10	4,096, 300

A.2 INCOMPLETE-VIEW PROTOCOL

For a dataset with N samples and V views, the implementation interprets η through the target observation fraction

$$\rho_{\text{obs}} = 1 - \frac{\eta}{V}. \tag{37}$$

The generator first assigns every sample one uniformly selected preserved view. It then samples the other mask entries and adjusts their sampling probability until the realized observation fraction is within 0.005 of ρ_{obs}. Consequently, each row of $\boldsymbol{M}$ has at least one nonzero entry, while the expected total number of missing entries is approximately $NV(1 - \rho_{\text{obs}}) = N\eta$. Mask seeds 1–5 define the five repeated trials at each value of η. This protocol differs from independently deleting an η fraction within every view; η controls the total number of missing view entries relative to the number of samples.

A.3 NETWORK AND OPTIMIZATION

Each view-specific encoder contains three graph-convolution layers with dimensions $d_v \to 192 \to 128 \to 64$ and ReLU activations. The available 64-dimensional view embeddings are averaged according to the observation mask and linearly mapped to a 96-dimensional common representation. A linear clustering head maps this representation to C outputs. Each decoder follows $96 \to 128 \to 192 \to d_v$ and reconstructs only observed samples.

All models are optimized with Adam for 1,500 epochs using a learning rate of 10^{-3}. The implementation fixes the model-level random seed to 42. Granular structures are activated at epoch 100 and reconstructed from detached view-specific embeddings every 20 epochs. Final clustering uses k-means with ten fixed initializations. The reported score for each setting is averaged over the five missing masks.

A.4 HYPERPARAMETER CONFIGURATION

Table 4 lists the shared settings. Reliability filtering, confidence reweighting, and semantic calibration are disabled, so the experiments use the raw row-normalized co-observation mapping in Equation 16.

Table 4: Shared hyperparameters used by CGBST.

Component	Parameter	Value
Granular construction	$m_{\min}$ / $K_{\max}$ / $d_{\max}$	8 / 20 / 20
Granular construction	minimum split gain δ	10^{-4}
Sample graphs	nearest neighbors k_{sam}	15
Soft affiliation	nearest regions k_{aff}	2
Region topology	nearest regions k_{topo}	3
Relation recovery	nearest relations k_{rel} / topology weight β	10 / 1
Graph fusion	λ_g	0.2
Laplacian regularization	λ_{lap}	0.1
Contrastive learning	temperature τ	1
Structure schedule	warm-up / update interval	100 / 20 epochs

The contrastive exponent α is selected separately for each dataset and missing rate. For $\eta = (0.1, 0.3, 0.5)$, the respective settings are $(0, 0, -1)$ for MSRCV1, $(-1, -1, -1)$ for CUB, $(0, -1, -1)$ for OutdoorScene, and $(2, -3, 3)$ for NUS-WIDE.

B Detailed Algorithms

Algorithm 1 gives the complete training and clustering procedure. Algorithms 2 and 3 then detail its two structural update stages.

Algorithm 1: Overall training and clustering procedure of CGBST

Input: Incomplete views $\{\boldsymbol{X}^{(v)}\}_{v=1}^{V}$; observation mask $\boldsymbol{M}$; cluster number C; training epochs T; warm-up epoch t_{w}; structure-update interval T_{u}; model and structural hyperparameters.
Output: Common embedding $\boldsymbol{Z}$ and cluster assignments $\hat{\boldsymbol{c}}$.

Determine $\{\mathcal{O}_v, \boldsymbol{T}^{(v)}\}_{v=1}^{V}$ from $\boldsymbol{M}$, normalize each view, and build $\{\boldsymbol{G}_o^{(v)}, \bar{\boldsymbol{G}}_0^{(v)}\}_{v=1}^{V}$.
Initialize the encoders $\{f_v\}$, common projection, clustering head, decoders $\{\psi_v\}$, and Adam optimizer.
Initialize the recovered graph $\boldsymbol{G}^{\mathrm{gr}}$ and its edge mask $\boldsymbol{E}^{\mathrm{gr}}$ as unavailable.
for epoch $t = 1$ to T **do**
  Encode observed samples to obtain $\{\boldsymbol{Z}_o^{(v)}, \bar{\boldsymbol{Z}}^{(v)}\}_{v=1}^{V}$ and form $\boldsymbol{Z}$ and $\boldsymbol{Y}$ by mask-aware fusion.
  if $t \geq t_{\mathrm{w}}$ and either $\boldsymbol{G}^{\mathrm{gr}}$ is unavailable or $(t - t_{\mathrm{w}}) \bmod T_{\mathrm{u}} = 0$ **then**
    For each view, construct $\mathcal{B}^{(v)}$, $\bar{\boldsymbol{A}}^{(v)}$, $\boldsymbol{A}^{(v)}$, and $\boldsymbol{S}^{(v)}$ from detached embeddings using Algorithm 2.
    Update $\boldsymbol{G}^{\mathrm{gr}}$ and $\boldsymbol{E}^{\mathrm{gr}}$ through cross-view transfer and relation recovery using Algorithm 3.
  end if
  Construct $\boldsymbol{G}^{z}$ and fuse it with the observed-view graphs and, when available, $\boldsymbol{G}^{\mathrm{gr}}$ to obtain $\boldsymbol{G}^{c}$.
  Compute $\mathcal{L}_{\mathrm{rec}}$ and $\mathcal{L}_{\mathrm{con}}$ using Equations 34 and 32.
  Compute $\mathcal{L}_{\mathrm{lap}}$ using Equation 33 if $\boldsymbol{G}^{\mathrm{gr}}$ is available; otherwise set $\mathcal{L}_{\mathrm{lap}} \leftarrow 0$.
  Update all trainable parameters by minimizing the objective in Equation 35.
end for
Apply k-means to the final $\boldsymbol{Z}$ according to Equation 36.
return $\boldsymbol{Z}, \hat{\boldsymbol{c}}$.

Algorithm 2 summarizes the adaptive construction performed independently in each view. Detaching the embeddings makes the discrete partition a periodically updated structural target rather than a differentiable operation inside every optimization step.

Algorithm 2: Adaptive view-specific granular-region construction

Input: Detached observed embedding $\boldsymbol{Z}_o^{(v)}$; $m_{\min}$, $K_{\max}$, $d_{\max}$, δ; neighborhood sizes k_{aff} and k_{topo}.
Output: Regions $\mathcal{B}^{(v)}$; affiliations $\bar{\boldsymbol{A}}^{(v)}$, $\boldsymbol{A}^{(v)}$; region topology $\boldsymbol{S}^{(v)}$.

Initialize $\mathcal{B}^{(v)}$ with one root region containing all observed samples.
Set the specificity scale to the mean radius of the root region.
while $|\mathcal{B}^{(v)}| < K_{\max}$ **do**
  $\mathcal{C} \leftarrow \varnothing$ as the valid split-candidate set.
  for each $B \in \mathcal{B}^{(v)}$ **do**
    if $\text{depth}(B) < d_{\max}$ and $|B| \geq 2m_{\min}$ **then**
      $(B_l, B_r) \leftarrow \text{FarthestBisect}(B)$ using Equation 8.
      if $|B_l|, |B_r| \geq m_{\min}$ and $\Delta q(B) > \delta$ **then**
        Add $(B, B_l, B_r, \Delta q(B))$ to $\mathcal{C}$.
      end if
    end if
  end for
  if $\mathcal{C} = \varnothing$ **then break**.
  Replace the candidate parent having the largest gain with its children (B_l, B_r).
end while
Compute $\bar{\boldsymbol{A}}^{(v)}$ from the final divisive-region membership.
Compute $\boldsymbol{A}^{(v)}$ using the k_{aff}-nearest regions in Equation 14.
Construct the k_{topo}-nearest-neighbor region topology $\boldsymbol{S}^{(v)}$.
return $\mathcal{B}^{(v)}, \bar{\boldsymbol{A}}^{(v)}, \boldsymbol{A}^{(v)}, \boldsymbol{S}^{(v)}$.

Algorithm 3 presents one structure-transfer and relation-recovery update. It assigns a missing sample to target-view regions and reconstructs its relations; it neither moves the sample between view-specific regions nor synthesizes a missing feature vector.

Algorithm 3: Cross-view granular structure transfer and relation recovery

Input: Observation mask $\boldsymbol{M}$; $\{\bar{\boldsymbol{A}}^{(v)}, \boldsymbol{A}^{(v)}, \boldsymbol{S}^{(v)}, \bar{\boldsymbol{G}}_0^{(v)}\}_{v=1}^{V}$; topology weight β; relation neighborhood size k_{rel}.
Output: Aggregated recovered graph $\boldsymbol{G}^{\text{gr}}$ and its edge mask $\boldsymbol{E}^{\text{gr}}$.

for target view $v = 1$ to V **do**
  Initialize $\tilde{\boldsymbol{A}}^{(v)} \leftarrow \boldsymbol{A}^{(v)}$.
  for each source view $u \neq v$ **do**
    $\boldsymbol{H}^{u \to v} \leftarrow (\bar{\boldsymbol{A}}_{\mathcal{O}_{uv}}^{(u)})^{\top} \bar{\boldsymbol{A}}_{\mathcal{O}_{uv}}^{(v)}$.
    $\boldsymbol{P}^{u \to v} \leftarrow \text{RowNormalize}(\boldsymbol{H}^{u \to v})$.
    $\widehat{\boldsymbol{A}}^{u \to v} \leftarrow \boldsymbol{D}_{\text{elig}}^{u \to v} \boldsymbol{A}^{(u)} \boldsymbol{P}^{u \to v}$.
  end for
  for each sample i with $M_{iv} = 0$ **do**
    Determine valid donors $\mathcal{D}_i^{(v)}$ using Equation 20.
    if $|\mathcal{D}_i^{(v)}| > 0$ **then**
      $\tilde{\boldsymbol{a}}_i^{(v)} \leftarrow \text{RowNormalize}(\sum_{u \in \mathcal{D}_i^{(v)}} \hat{\boldsymbol{a}}_i^{u \to v})$.
    else $\tilde{\boldsymbol{a}}_i^{(v)} \leftarrow \boldsymbol{0}$.
    end if
  end for
  Compute $\boldsymbol{R}^{(v)}$ using the same-region and topology terms in Equation 23.
  Build $\boldsymbol{G}_{\text{mis}}^{(v)}$ from the top-k_{rel} positive relations of target-missing rows.
  Build $\boldsymbol{G}_{\text{obs}}^{(v)}$ by reinforcing existing observed-view edges only.
  $\tilde{\boldsymbol{G}}^{(v)} \leftarrow \text{SymMax}(\boldsymbol{G}_{\text{mis}}^{(v)}) \vee \text{SymMax}(\boldsymbol{G}_{\text{obs}}^{(v)})$.
end for
Aggregate $\{\tilde{\boldsymbol{G}}^{(v)}\}_{v=1}^{V}$ into $\boldsymbol{G}^{\text{gr}}$ using Equation 29 and set $\boldsymbol{E}^{\text{gr}} = \mathbb{I}[\boldsymbol{G}^{\text{gr}} > 0]$.
return $\boldsymbol{G}^{\text{gr}}, \boldsymbol{E}^{\text{gr}}$.

C Computational Complexity and Scalability

Let $K = \max_v K_v$, d_z be the latent dimension, and $N_{uv} = |\mathcal{O}_{uv}|$. Sparse sample-to-region affiliation costs $O(VNKd_z)$, cross-view support costs $O(\sum_{u \neq v} N_{uv} K_u K_v)$ in the dense upper bound, and dense relation recovery costs $O(\sum_v N^2 K_v)$. The corresponding storage is $O(N^2 + \sum_v NK_v + \sum_v K_v^2)$. Since $K_{\max} = 20$, the main bottleneck is the $N \times N$ sample graph rather than the region structures. The candidate-edge implementation avoids materializing all recovered pairs and provides a sparse route to larger datasets.

D Additional Parameter Sensitivity Analysis

Figures 5–7 complement Figure 2 with the remaining datasets; each bar reports the mean over five masks (%). The preferred exponent depends on both the dataset and the missing rate, and increasing the contrastive weight does not produce a consistent monotonic improvement. This behavior motivates selecting α on the same fixed search grid for every benchmark rather than imposing one global value.

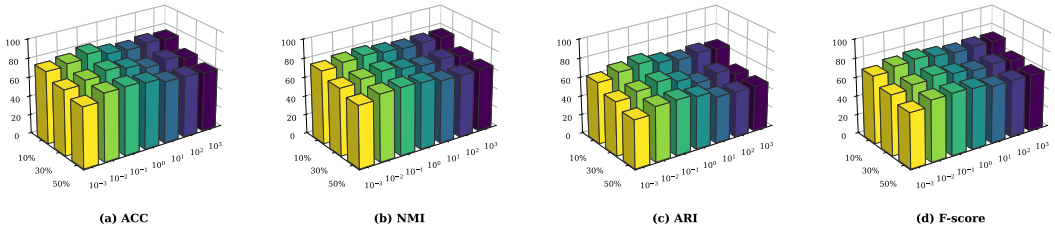

Figure 5: Sensitivity to α on CUB under $\eta \in \{0.1, 0.3, 0.5\}$.

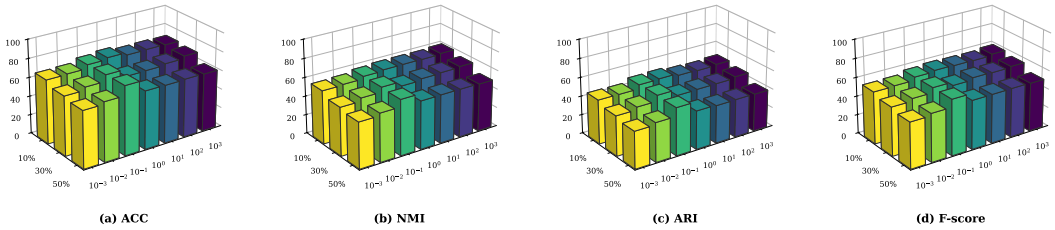

Figure 6: Sensitivity to α on OutdoorScene under $\eta \in \{0.1, 0.3, 0.5\}$.

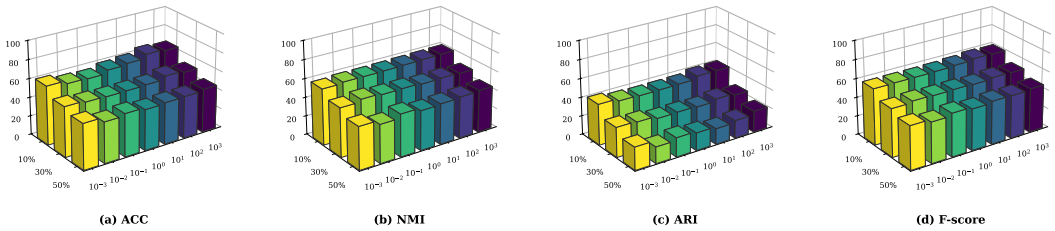

Figure 7: Sensitivity to α on NUS-WIDE under $\eta \in \{0.1, 0.3, 0.5\}$.

E Additional Representation Visualizations

Figures 8 and 9 visualize the common representations before training, at epoch 300, and after 1,500 epochs under $\eta = 0.3$. Class labels are used only for coloring. The representations become more

class-structured during optimization, while OutdoorScene and NUS-WIDE retain visible overlap among several classes. The latter cases are consistent with their more challenging clustering results but still show improved separation over training.

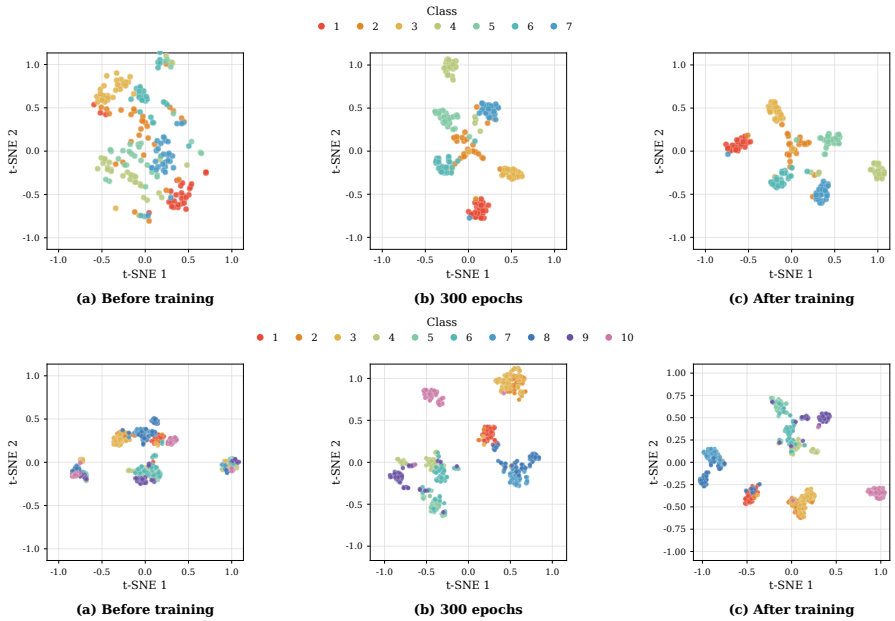

Figure 8: t-SNE visualizations on MSRCV1 and CUB at three training stages under $\eta = 0.3$. All samples are shown.

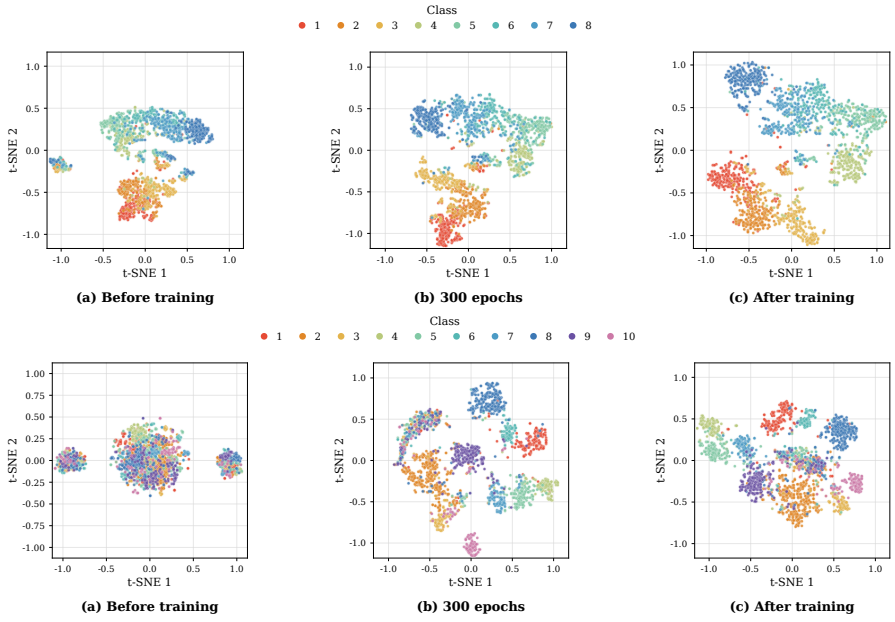

Figure 9: t-SNE visualizations on OutdoorScene and NUS-WIDE at three training stages under $\eta = 0.3$. Each dataset uses a fixed subset of 1,800 samples for readability.