INCOMPLETE MULTI-VIEW CLUSTERING VIA REPRESENTATION RECOVERY AND DIRECT SOFT-ASSIGNMENT ESTIMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete multi-view clustering (IMVC) aims to discover latent cluster structure from multi-view data with partially observed views. Existing imputation-based methods commonly recover missing representations from observed views and then obtain cluster assignments from the completed representations. For a sample missing a target view, however, representation recovery and soft cluster-assignment estimation generally do not share the same reference samples and relation weights, so the latent representation and cluster-membership information already available in the target view are not used under a common sample relation. We propose ReSA, an IMVC method based on representation recovery and direct soft-assignment estimation. ReSA first learns cross-view relation representations for reference retrieval and relation-weight computation. Cross-view contrastive learning treats the relation representations of the same jointly observed instance as a positive pair and those of different instances as negative pairs; neighborhood distribution consistency further preserves local cross-view relations. For a sample with a missing target view, ReSA selects the most strongly related references from samples observed in both the source and target views and computes their relation weights. The same references and weights are then reused to aggregate target-view latent representations for missing-representation recovery and target-view soft assignments for direct assignment estimation. Completed multi-view representations provide global clustering information, whereas observed or estimated view-specific assignments preserve view-specific cluster membership; both are integrated by global–view consensus clustering. Experiments on four IMVC benchmarks and multiple missing rates show that ReSA achieves leading clustering performance. Comparisons with the trained model held fixed show that direct soft-assignment estimation is closer to the reference assignment computed from the complete target view under identical references and weights, and missing-view reconstruction confirms that the recovered representations retain decodable target-view content.

1 INTRODUCTION

Multi-view data are often incomplete. In assisted diagnosis, MRI, PET, or cerebrospinal-fluid measurements may be unavailable because of cost, patient preference, or data quality; similar patterns arise in multimedia analysis, sensor systems, and recommender systems. Incomplete multi-view clustering (IMVC) seeks to discover clusters from partial observations by exploiting cross-view consistency and complementarity (Wen et al., 2023b; Tang et al., 2024). The challenge is to compensate for missing views while preserving structure useful for clustering.

Deep IMVC methods address missing information through imputation-based or imputation-free learning. The former estimate missing views or representations using contrastive prediction, semantic neighbors, prototype relations, or diffusion generation (Lin et al., 2021; Tang and Liu, 2022; Li et al., 2023; Wen et al., 2024). The latter learn clustering structure directly from observations, as in DVIMC and FreeCSL (Xu et al., 2024; Dai et al., 2025). They avoid recovery errors, but increasing missingness reduces the available cross-view information and can increase view imbalance and assignment uncertainty.

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.

For imputation-based methods, two issues motivate our work. First, representation recovery and target-view assignment estimation usually follow serial procedures or separate mechanisms. MICA learns assignments from completed features (Wang et al., 2025b); DIMVC-HIA first estimates assignments, then reconstructs features from target-view within-cluster statistics (Du et al., 2026); and JVILG uses separate anchor-based imputation and label-generation modules (Wang et al., 2025a). A representation-first procedure infers the missing assignment from a recovered representation without directly aggregating the references' target-view memberships. This motivates a common sample-level relation that provides the same references and weights for estimating both quantities.

Second, this relation must support transfer across heterogeneous views: source-view similarities need not reflect target-view local relations. Cross-view alignment and relation transfer provide useful semantic links (Jin et al., 2023; Chao et al., 2024), but instance-level contrastive objectives do not explicitly match the neighborhood distributions used for reference retrieval. Aligning the same instance across views alone does not characterize its relative relations to neighboring references.

We propose ReSA, which learns relation representations on top of view-specific latent representations. Cross-view contrastive learning (CCL) aligns jointly observed instances, while neighborhood distribution consistency (NDC) matches their local relation distributions. For a missing target view, an observed source view selects references observed in both views and determines their weights. The same references and weights aggregate target-view latent representations and soft assignments separately. Global–view consensus then combines assignments from completed multi-view representations with observed or estimated view-specific assignments. Our contributions are:

- A shared sample-level cross-view relation for missing-representation recovery and direct target-view soft-assignment estimation, with both outputs used for clustering.
- CCL and NDC for reference retrieval and weighting, followed by global–view consensus integrating global structure and view-specific membership.
- Experiments across benchmarks and missing rates, with component ablations, controlled assignment comparisons, and reconstruction analysis of retained target-view content.

2 Related Work

Incomplete multi-view clustering. Early methods use matrix factorization, kernels, or graphs (Wen et al., 2023b; Tang et al., 2024). MIC learns a common representation with weighted nonnegative factorization (Shao et al., 2015); DAIMC adds basis alignment to weighted semi-nonnegative factorization (Hu and Chen, 2018); and OPIMC supports one-pass clustering (Hu and Chen, 2019). Their shallow formulations have limited ability to capture nonlinear semantics. Imputation-free methods include consensus graph learning from high-confidence local structures in HCLS-CGL (Wen et al., 2023a), shared clustering structures in DVIMC (Xu et al., 2024), and consensus semantics and within-view clusters in FreeCSL (Dai et al., 2025).

Imputation-based methods use bidirectional contrastive prediction in COMPLETER (Lin et al., 2021), reliable semantic-neighbor recovery in DSIMVC (Tang and Liu, 2022), sample–prototype transfer in ProImp (Li et al., 2023), and diffusion generation conditioned on observed views in DMVG (Wen et al., 2024). DCG adds contrastive instance- and class-level interactions (Zhang et al., 2025). MICA uses an adaptive cross-view graph and feature-, data-, and reconstruction-level imputation (Wang et al., 2025b). ARSL-IMVC couples collaborative recovery and tensor subspace learning through a shared latent representation (Wang et al., 2026). DIMVC-HIA refines recovered features and assignments with energy-based semantic and cluster-assignment alignment (Du et al., 2026), while JVILG jointly models imputation and soft labels with separate modules (Wang et al., 2025a).

Cross-view relations. CPSPAN aligns jointly observed instances and prototypes (Jin et al., 2023); PMIMC combines relation consistency with prototype matching (Yuan et al., 2025); and ICMVC uses consistency relation transfer, graph convolution, and contrastive learning (Chao et al., 2024). ReSA focuses on sharing one reference set and its weights between representation recovery and assignment estimation, then integrating both outputs through global–view consensus.

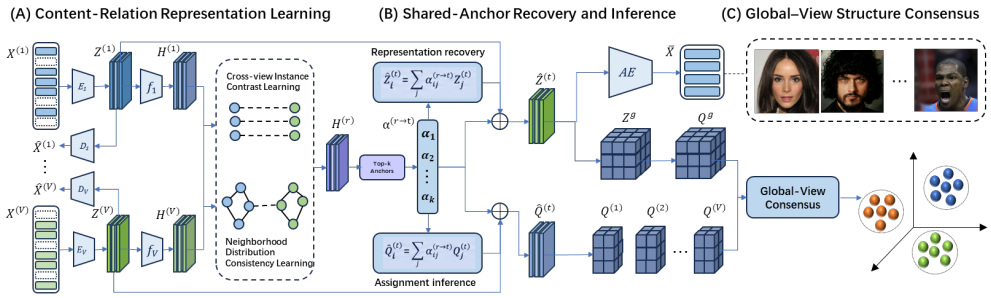


Figure 1: Overview of ReSA. Relation representations $H^{(v)}$ are learned from view-specific latent representations $Z^{(v)}$ with cross-view contrastive learning (CCL) and neighborhood distribution consistency (NDC). For a missing target view, the same jointly observed reference samples and relation weights are used for representation recovery and direct soft-assignment estimation. Completed representations provide global clustering information, whereas observed or estimated assignments preserve view-specific cluster membership for global–view consensus. Missing-view reconstruction is used only for auxiliary evaluation.

3 METHOD

3.1 NOTATION AND OVERALL FRAMEWORK

Let $\mathcal{X} = \{X^{(v)}\}_{v=1}^{V}$ contain N samples and V views, with $X^{(v)} \in \mathbb{R}^{N \times d_v}$ and input dimension d_v. The observation matrix $M \in \{0,1\}^{N \times V}$ has $M_{iv} = 1$ for an observed entry and $M_{iv} = 0$ otherwise. Define $\mathcal{O}_v = \{i : M_{iv} = 1\}$ and assume $\sum_v M_{iv} \geq 1$ for every sample. The goal is to partition all N samples into K clusters without class supervision. An observed view used to estimate missing information is the *source view* r; the missing view is the *target view* t.

Figure 1 shows the three components of ReSA: cross-view relation learning, relation-guided estimation, and global–view consensus. The first learns relations for retrieval and weighting, the second uses shared references and weights to estimate target representations and assignments, and the third integrates their clustering information.

3.2 CROSS-VIEW RELATION LEARNING

View-specific latent representations. Each view has an independent encoder E_v and decoder D_v. For $i \in \mathcal{O}_v$, the latent representation is $z_i^{(v)} = E_v(x_i^{(v)})$ and its reconstruction is $\widehat{x}_i^{(v)} = D_v(z_i^{(v)})$. Training uses only observed samples:

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

The resulting $Z^{(v)}$ provides view content for representation recovery and input to relation learning.

Relation representations. A view-specific projection head f_v produces normalized relation representations for cross-view reference retrieval and weight computation:

$$h_i^{(v)} = \frac{f_v\left(z_i^{(v)}\right)}{\left\| f_v\left(z_i^{(v)}\right) \right\|_2}, \qquad H^{(v)} = \left\{ h_i^{(v)} \right\}_{i \in \mathcal{O}_v}. \tag{2}$$

Thus $H^{(v)}$ determines sample relations, whereas $Z^{(v)}$ supplies target-view content. The jointly observed set $\Omega_{r,t} = \{i : M_{ir} = M_{it} = 1\}$ provides genuine observations in both views for relation learning.

Cross-view contrastive learning (CCL). CCL reduces view heterogeneity by treating $(h_i^{(r)}, h_i^{(t)})$ as a positive pair and $(h_i^{(r)}, h_j^{(t)})$, $j \neq i$, as negative pairs for $i, j \in \Omega_{r,t}$. Normalization gives

$\mathrm{sim}(a,b) = a^\top b$. The one-way loss is

$$\mathcal{L}_{\mathrm{CCL}}^{r\to t} = -\frac{1}{|\Omega_{r,t}|}\sum_{i\in\Omega_{r,t}} \log \frac{\exp\left(\mathrm{sim}\left(h_i^{(r)}, h_i^{(t)}\right)/\tau_c\right)}{\sum_{j\in\Omega_{r,t}} \exp\left(\mathrm{sim}\left(h_i^{(r)}, h_j^{(t)}\right)/\tau_c\right)}, \tag{3}$$

where $\tau_c = 0.07$. For valid view pairs $\mathcal{P} = \{(r,t) : r < t,\ |\Omega_{r,t}| \geq 2\}$, the symmetric loss is $\mathcal{L}_{\mathrm{CCL}} = (2|\mathcal{P}|)^{-1}\sum_{(r,t)\in\mathcal{P}}(\mathcal{L}_{\mathrm{CCL}}^{r\to t} + \mathcal{L}_{\mathrm{CCL}}^{t\to r})$.

Neighborhood distribution consistency (NDC). Reference selection depends on relations to neighboring candidates as well as alignment of the same instance. NDC therefore matches local relation distributions. For $(r,t)\in\mathcal{P}$, $i,j\in\Omega_{r,t}$, and $v\in\{r,t\}$, define

$$\kappa\left(h_i^{(v)}, h_j^{(v)}\right) = \exp\left[-\frac{2 - 2\left\langle h_i^{(v)}, h_j^{(v)}\right\rangle}{\tau_r}\right], \tag{4}$$

where $\tau_r = 0.07$ and normalization gives $2 - 2\langle h_i^{(v)}, h_j^{(v)}\rangle = \|h_i^{(v)} - h_j^{(v)}\|_2^2$. For each query, normalize over the same jointly observed support:

$$P_{ij}^{(v)} = \frac{\mathbb{I}[i\neq j]\,\kappa\left(h_i^{(v)}, h_j^{(v)}\right)}{\sum_{\ell\in\Omega_{r,t},\,\ell\neq i}\kappa\left(h_i^{(v)}, h_\ell^{(v)}\right)}, \qquad j\in\Omega_{r,t}. \tag{5}$$

The squared Hellinger distance between these distributions is $D_{\mathrm{Hel}}^2(P_i^{(r)}, P_i^{(t)}) = \frac{1}{2}\sum_{j\in\Omega_{r,t}}(\sqrt{P_{ij}^{(r)}} - \sqrt{P_{ij}^{(t)}})^2$. The NDC loss is

$$\mathcal{L}_{\mathrm{NDC}} = \frac{1}{|\mathcal{P}|}\sum_{(r,t)\in\mathcal{P}}\frac{1}{|\Omega_{r,t}|}\sum_{i\in\Omega_{r,t}} D_{\mathrm{Hel}}^2\left(P_i^{(r)}, P_i^{(t)}\right). \tag{6}$$

Together, CCL and NDC constrain cross-view relation representations and their local neighborhoods. The full objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{sem}}\left(\mathcal{L}_{\mathrm{CCL}} + \mathcal{L}_{\mathrm{NDC}}\right), \tag{7}$$

with $\lambda_{\mathrm{sem}} = 0.05$ balancing relation learning against reconstruction.

3.3 Relation-Guided Representation Recovery and Soft-Assignment Estimation

Shared references and weights. For sample i missing target view t, source view r retrieves candidates from $\Omega_{r,t}$. Each candidate has both an observed source relation and target-view content. Set $s_{ij}^{r\to t} = \kappa(h_i^{(r)}, h_j^{(r)})$ using Eq. (4) and select $\mathcal{N}_{i,k}^{r\to t} = \mathrm{TopK}_{j\in\Omega_{r,t}}(s_{ij}^{r\to t})$. Their normalized weights are

$$\alpha_{ij}^{r\to t} = \frac{s_{ij}^{r\to t}}{\sum_{\ell\in\mathcal{N}_{i,k}^{r\to t}} s_{i\ell}^{r\to t}}, \qquad j\in\mathcal{N}_{i,k}^{r\to t}, \tag{8}$$

with $\alpha_{ij}^{r\to t} \geq 0$ and $\sum_{j\in\mathcal{N}_{i,k}^{r\to t}}\alpha_{ij}^{r\to t} = 1$. The references and weights define the sample-level cross-view relation.

Target-view soft cluster assignments. Run K-means only on observed target representations $\{z_j^{(t)} : j\in\mathcal{O}_t\}$ to obtain centroids $\{\mu_c^{(t)}\}_{c=1}^K$. Define

$$[C_t(z)]_c = \frac{\exp\left(-\left\|z - \mu_c^{(t)}\right\|_2\right)}{\sum_{\ell=1}^{K}\exp\left(-\left\|z - \mu_\ell^{(t)}\right\|_2\right)}, \qquad c = 1,\ldots,K. \tag{9}$$

Each observed reference has $q_j^{(t)} = C_t(z_j^{(t)}) \in \mathbb{R}^K$, computed from its observed target representation.

Representation recovery and direct assignment estimation. The same references and weights aggregate target-view latent representations and soft assignments separately:

$$\widehat{z}_i^{(t|r)} = \sum_{j \in \mathcal{N}_{i,k}^{r \to t}} \alpha_{ij}^{r \to t} z_j^{(t)}. \tag{10}$$

$$\widehat{q}_i^{(t|r)} = \sum_{j \in \mathcal{N}_{i,k}^{r \to t}} \alpha_{ij}^{r \to t} q_j^{(t)}. \tag{11}$$

Here $H^{(r)}$ determines retrieval and weighting, while the aggregated quantities are $z_j^{(t)}$ and $q_j^{(t)}$. The target assignment need not be inferred indirectly from the recovered representation.

Multiple sources and completed outputs. For multiple observed sources, define $\mathcal{R}_i^t = \{r \neq t : M_{ir} = 1,\ |\Omega_{r,t}| \geq k\}$. Average over valid sources: $\widehat{z}_i^{(t)} = |\mathcal{R}_i^t|^{-1} \sum_{r \in \mathcal{R}_i^t} \widehat{z}_i^{(t|r)}$ and $\widehat{q}_i^{(t)} = |\mathcal{R}_i^t|^{-1} \sum_{r \in \mathcal{R}_i^t} \widehat{q}_i^{(t|r)}$. The completed entries $(\widetilde{z}_i^{(t)}, \widetilde{q}_i^{(t)})$ equal $(z_i^{(t)}, q_i^{(t)})$ when $M_{it} = 1$ and $(\widehat{z}_i^{(t)}, \widehat{q}_i^{(t)})$ otherwise. Stacking rows yields $\widetilde{Z}^{(t)} = [\widetilde{z}_1^{(t)}, \ldots, \widetilde{z}_N^{(t)}]^\top$ and $\widetilde{Q}^{(t)} = [\widetilde{q}_1^{(t)}, \ldots, \widetilde{q}_N^{(t)}]^\top \in \mathbb{R}^{N \times K}$.

Auxiliary reconstruction. The target decoder produces $\widehat{x}_i^{(t)} = D_t(\widehat{z}_i^{(t)})$ only to evaluate decodable target-view content. Reconstruction does not participate in assignment estimation or consensus clustering.

3.4 Global–View Consensus Clustering

Concatenate completed representations as $z_i^g = [\widetilde{z}_i^{(1)} \| \cdots \| \widetilde{z}_i^{(V)}]$. Run K-means on these joint representations and compute global soft assignments $Q^g \in \mathbb{R}^{N \times K}$ from distances to global centers. The consensus input $\mathcal{Q} = \{Q^g, \widetilde{Q}^{(1)}, \ldots, \widetilde{Q}^{(V)}\}$ combines global clustering structure with view-specific membership.

To resolve permutation ambiguity, align cluster indices with the Hungarian algorithm (Kuhn, 1955). Retain and renormalize each soft assignment's three largest components, then mix soft and hard edges with fixed coefficient $\gamma = 0.5$. Construct the sample–assignment matrix and obtain final clusters using a T-cut/FastMICE-style solver (Huang et al., 2023).

4 Experiments

We first evaluate the complete ReSA clustering framework on four IMVC benchmarks, and then analyze its relation representation, direct soft-assignment estimator, global–view consensus, and missing-view reconstruction branch.

4.1 Experimental Setup

Datasets. We evaluate ALOI-100, HandWritten, Hdigit, and YouTubeFace10, which have different scales and view structures. ALOI-100 contains 10,800 samples from 100 object classes, represented by HSB, RGB, Color Similarity, and Haralick color or texture features. HandWritten contains 2,000 handwritten-digit samples from ten classes and six heterogeneous visual descriptors. Hdigit contains 10,000 digit samples from ten classes with MNIST-style and USPS-style views. YouTubeFace10 contains 38,654 face samples from ten identities and four feature views. Detailed statistics are reported in Table 1.

Baselines. We compare with representative IMVC methods. COMPLETER uses contrastive representation learning and cross-view prediction (Lin et al., 2021); DSIMVC recovers missing views from semantic neighbors (Tang and Liu, 2022); CPSPAN learns clustering structure with instance- and prototype-level cross-view alignment (Jin et al., 2023); DVIMC aggregates shared clustering information from available views without explicit imputation (Xu et al., 2024); FreeCSL learns cross-view consensus semantics and within-view cluster structure (Dai et al., 2025); PMIMC uses relation consistency and prototype matching for missing-information recovery (Yuan et al., 2025); and

Table 1: Statistics of the four IMVC benchmarks.

Dataset	Samples	Classes	Views	View dimensions
ALOI-100	10,800	100	4	77/13/64/125
HandWritten	2,000	10	6	240/76/216/47/64/6
Hdigit	10,000	10	2	784/256
YouTubeFace10	38,654	10	4	944/576/512/640

DIMVC-HIA jointly uses hierarchical soft-assignment estimation and feature recovery (Du et al., 2026). We additionally compare with MVP (Gao and Pu, 2025) and GAVIM (Chen et al., 2026). On the largest dataset, YouTubeFace10, we also include DIMVC (Xu et al., 2022), DCP (Lin et al., 2023), and GLGC (He et al., 2026). Methods requiring a rerun are evaluated from their public implementations while keeping their model structures and optimization procedures unchanged; input features, observation masks, and evaluation protocols are unified.

Incomplete-data construction. We use the sample-level missing rate, the fraction of samples missing at least one view:

$$\rho = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I} \left[\sum_{v=1}^{V} M_{iv} < V \right], \tag{12}$$

where ρ denotes the missing rate. The main experiments use $\rho \in \{0.3, 0.5, 0.7\}$, and every sample retains at least one observed view. For the same dataset and missing rate, all rerun baselines use exactly the same observation mask, input features, sample order, and evaluation code. The three training seeds $\{0, 1, 2\}$ share one observation mask for each dataset–missing rate setting, so that changes in the missing pattern do not confound the comparison.

Implementation details. Each view-specific encoder uses a fully connected d_v–512–512–1024–256 architecture with a symmetric decoder. The relation projection head f_v has architecture 256–256–256. We use ReLU activations, Adam, and batch size 256. The autoencoders are trained for 200 epochs with learning rate 5×10^{-4}, followed by 100 epochs of cross-view relation learning with learning rate 10^{-4}. Experiments are run on an NVIDIA GeForce RTX 4090 GPU. We set $k = 8$ for ALOI-100, HandWritten, and Hdigit, and $k = 4$ for YouTubeFace10; k is fixed across missing rates within a dataset. All K-means runs use $n_{\text{init}} = 10$.

Evaluation metrics. We use Accuracy (ACC), Normalized Mutual Information (NMI), and Adjusted Rand Index (ARI). Larger values indicate better clustering. Tables report the mean and sample standard deviation over the three training seeds.

4.2 Comparison with Existing Methods

Table 2 reports clustering performance on the four datasets. ReSA obtains the highest mean in all 36 dataset–missing-rate– metric cells. The advantage remains clear at high missing rates. At MR=0.7, relative to the strongest compared method for each metric, ReSA improves ACC, NMI, and ARI by 12.92, 6.24, and 17.99 percentage points on ALOI-100, and by 8.62, 6.61, and 11.67 points on HandWritten. On the Hdigit benchmark, where clustering accuracy is already high, ReSA still reaches 97.30/92.71/94.10 ACC/NMI/ARI at MR=0.7, improving over the corresponding strongest competitor by 0.69, 1.49, and 1.46 points. On YouTubeFace10, ReSA obtains 80.99/84.61/75.45, 80.51/84.33/75.35, and 79.94/84.69/74.91 at MR=0.3, 0.5, and 0.7, respectively. At MR=0.5, the gains over the strongest compared method are 6.07, 3.91, and 6.71 points for ACC, NMI, and ARI. These results show that ReSA outperforms the compared methods across different dataset scales, numbers of views, and missing rates.

4.3 Model Analysis

Ablation study. We ablate the relation representation, direct soft-assignment estimation, and global–view consensus, and compare each variant with the complete model. Table 3 reports mean ACC over the three missing rates for each dataset.

Table 2: Clustering performance (%) on four datasets. MR is the proportion of samples with at least one missing view. Results are mean ± sample standard deviation over seeds 0/1/2; all methods use the same observation mask for each dataset and MR. Higher is better; bold and underline indicate the best and second-best results, respectively.

Dataset	Method	MR=0.3			MR=0.5			MR=0.7		
		ACC↑	NMI↑	ARI↑	ACC↑	NMI↑	ARI↑	ACC↑	NMI↑	ARI↑
ALOI-100	COMPLETER	16.51±0.90	50.46±0.74	11.25±1.17	16.80±0.67	47.33±0.62	11.26±0.37	16.46±0.78	45.90±1.27	9.48±0.72
	DSIMVC	35.99±0.64	63.04±0.47	26.09±0.67	33.68±3.32	60.03±1.15	23.18±1.81	31.74±1.00	57.11±0.37	20.77±0.74
	CPSPAN	68.71±3.19	83.83±1.75	58.36±3.78	64.94±1.01	81.91±0.70	53.40±0.91	64.44±3.00	81.66±0.95	52.90±2.85
	DVIMC	1.07±0.12	0.30±0.52	0.00±0.00	1.26±0.45	0.96±1.66	0.07±0.13	1.60±0.52	3.42±3.79	0.22±0.22
	MVP	82.39±0.72	90.04±0.18	74.24±1.31	79.21±1.35	88.04±0.62	71.10±1.85	76.67±0.89	85.36±0.22	66.04±0.35
	FreeCSL	87.37±0.89	93.15±0.14	83.20±1.03	81.41±1.57	89.56±0.31	75.88±1.33	72.18±0.86	84.34±0.77	64.85±1.76
	PMIMC	75.43±1.71	88.53±0.78	67.39±3.35	71.65±1.59	86.70±0.64	63.11±0.98	71.93±1.71	86.24±0.29	62.11±2.41
	DIMVC-HIA	66.12±1.67	80.06±0.36	57.06±1.08	55.85±1.53	75.31±0.57	48.22±0.88	40.42±3.16	67.66±0.94	34.77±1.20
	GAVIM	12.19±1.57	25.45±1.38	5.17±1.18	12.33±1.65	24.85±1.74	4.95±1.15	12.41±1.16	24.13±1.67	5.01±1.03
	ReSA	**93.57±0.31**	**95.12±0.24**	**89.50±0.23**	**91.96±1.72**	**94.21±0.42**	**87.36±1.71**	**89.59±0.69**	**92.48±0.07**	**84.03±0.69**
HandWritten	COMPLETER	51.02±4.28	51.56±2.41	36.72±3.21	40.42±3.47	39.66±2.47	17.62±3.52	28.52±0.53	28.25±1.83	11.51±1.41
	DSIMVC	82.88±0.91	78.98±1.67	72.92±2.02	78.88±4.53	74.70±2.81	67.27±4.07	77.53±4.83	71.83±2.02	64.08±3.81
	CPSPAN	87.12±6.46	81.41±3.96	77.50±6.84	84.80±6.51	80.96±4.23	76.27±6.61	85.68±8.46	80.54±4.03	76.21±7.85
	DVIMC	85.88±0.75	85.27±0.73	80.19±0.39	42.27±30.13	51.67±24.86	32.44±30.37	14.32±4.80	7.66±9.94	2.94±4.88
	MVP	86.22±6.04	83.66±1.93	79.53±4.98	88.50±6.07	83.33±1.34	80.80±4.14	82.68±7.22	78.83±3.12	73.95±6.38
	FreeCSL	83.27±1.01	86.58±0.62	79.51±1.53	85.95±2.05	86.33±0.79	81.37±0.82	83.28±1.70	81.64±1.70	75.40±2.58
	PMIMC	89.23±1.86	81.75±1.88	78.65±3.04	82.20±6.93	78.06±4.43	71.44±7.15	85.60±7.15	80.46±3.22	75.99±6.19
	DIMVC-HIA	86.13±3.87	79.67±3.95	74.83±6.28	80.33±3.65	73.75±3.09	67.49±4.16	80.47±0.79	73.68±1.25	67.77±1.13
	GAVIM	82.05±10.96	82.90±3.98	76.71±9.16	84.90±13.56	83.03±3.94	78.83±10.16	80.22±3.45	79.19±1.17	73.03±1.71
	ReSA	**95.57±0.14**	**90.34±0.41**	**90.42±0.30**	**95.07±0.24**	**89.52±0.51**	**89.41±0.50**	**94.30±0.15**	**88.25±0.29**	**87.88±0.31**
Hdigit	COMPLETER	96.97±0.24	93.28±0.31	93.44±0.46	96.51±1.30	92.05±2.06	92.37±2.82	96.23±0.80	90.57±1.12	91.83±1.67
	DSIMVC	98.39±0.10	95.38±0.25	96.47±0.21	97.83±0.15	94.03±0.34	95.24±0.31	96.61±0.15	91.22±0.36	92.64±0.32
	CPSPAN	86.84±8.18	87.51±3.31	83.90±7.04	87.83±6.83	86.44±2.60	83.39±6.32	86.93±7.32	85.82±2.98	82.19±6.80
	DVIMC	94.06±7.26	93.31±3.09	91.97±7.23	97.46±0.24	93.31±0.48	94.47±0.51	96.25±0.40	90.43±0.82	91.87±0.84
	MVP	96.68±0.30	91.39±0.61	92.77±0.65	90.98±7.27	85.71±4.42	84.80±8.36	71.37±0.14	69.57±1.83	62.92±1.37
	FreeCSL	89.96±7.55	93.05±2.32	88.90±6.82	88.32±8.41	90.58±3.27	86.83±7.64	86.49±7.36	85.92±1.73	82.72±5.66
	PMIMC	96.13±0.21	90.59±0.45	91.62±0.44	96.16±0.38	90.60±0.70	91.70±0.79	95.83±0.20	89.85±0.41	90.99±0.43
	DIMVC-HIA	98.15±0.04	94.76±0.08	95.93±0.09	95.93±2.19	90.21±3.85	91.34±4.39	91.31±3.92	82.62±5.45	82.24±7.43
	GAVIM	69.69±1.50	67.52±2.94	59.52±3.24	65.09±7.21	61.92±9.44	53.66±10.76	69.90±1.92	65.44±5.55	58.41±5.97
	ReSA	**98.72±0.11**	**96.19±0.28**	**97.18±0.24**	**98.38±0.06**	**95.33±0.14**	**96.44±0.14**	**97.30±0.03**	**92.71±0.11**	**94.10±0.06**
YouTubeFace10	COMPLETER	44.71±2.25	49.30±3.64	20.28±3.61	40.40±4.15	43.59±2.72	15.22±0.96	37.91±2.78	41.89±6.55	11.79±7.36
	DIMVC	71.80±2.30	71.48±1.43	61.90±1.33	74.06±4.33	70.13±2.28	63.76±3.14	77.27±4.25	70.16±3.96	64.54±6.44
	DSIMVC	68.76±2.59	72.11±2.74	58.04±4.68	71.25±5.71	74.84±2.00	61.70±4.34	70.32±3.96	74.30±2.06	60.87±3.90
	CPSPAN	59.06±5.11	70.72±5.06	51.76±5.92	67.16±3.41	73.01±3.62	57.20±5.01	68.37±2.50	75.33±2.22	60.01±1.82
	DCP	62.05±0.96	69.53±3.89	36.48±3.70	64.28±0.75	70.84±0.55	38.86±1.15	62.61±4.85	70.76±2.73	40.44±2.75
	DVIMC	35.01±33.44	27.71±48.00	24.51±42.45	15.70±0.00	0.00±0.00	0.00±0.00	35.09±33.58	25.77±44.63	22.70±39.32
	MVP	66.02±3.10	69.43±3.74	57.39±3.30	69.46±11.79	70.90±7.23	57.68±11.90	72.17±10.20	72.08±7.66	60.52±6.79
	FreeCSL	75.05±5.42	80.78±4.25	70.01±6.21	74.44±2.73	80.42±1.85	68.64±2.41	76.89±2.74	82.92±2.64	72.50±2.20
	PMIMC	73.34±5.57	78.33±1.62	68.17±4.37	70.34±2.66	76.72±0.72	64.48±1.21	76.70±2.80	80.12±2.66	70.34±3.02
	DIMVC-HIA	71.11±3.49	76.00±1.84	62.20±3.21	65.36±3.50	71.29±1.80	55.70±2.23	67.57±5.08	72.60±3.28	57.37±3.92
	GAVIM	75.96±6.42	76.71±3.58	68.00±4.35	73.88±5.64	76.26±5.80	66.37±7.32	70.42±12.08	72.40±6.62	60.71±12.42
	GLGC	51.16±4.95	58.38±2.29	35.60±3.58	49.58±6.58	61.81±3.26	38.64±4.14	50.47±2.13	60.65±4.03	37.73±4.88
	ReSA	**80.99±3.60**	**84.61±3.09**	**75.45±5.08**	**80.51±0.85**	**84.33±1.01**	**75.35±2.22**	**79.94±3.17**	**84.69±3.09**	**74.91±5.14**

Table 3: Component ablation of ReSA. All variants and Full ReSA are run independently under the same ablation protocol; dataset columns report mean ACC (%) over the three missing rates. Higher is better; bold and underline indicate the best and second-best results, respectively.

Variant	Relation representation	Direct soft assignment	Global assignment	View-specific assignment	ALOI-100	HandWritten	Hdigit	YouTubeFace10
w/o Relation Representation	×	✓	✓	✓	89.95	94.98	98.13	75.41
w/o Direct Soft-Assignment Estimation	✓	×	✓	✓	91.53	94.66	98.13	75.30
w/o View-specific Assignment	✓	–	✓	×	88.91	94.26	98.16	75.71
w/o Global Assignment	✓	✓	×	✓	91.01	94.98	95.32	75.30
Full ReSA	✓	✓	✓	✓	**91.70**	**95.29**	**98.39**	**78.99**

Relation representation. The complete ReSA uses H for reference retrieval and relation-weight computation. To assess this design, we remove the dedicated relation representation after the projection head and use the normalized latent representation Z for the same retrieval and weighting operations, denoted *w/o Relation Representation*. This variant has lower ACC than the full model on all four datasets; on YouTubeFace10, mean ACC decreases from 78.99 to 75.41.

Direct soft-assignment estimation. To assess direct assignment estimation, we keep the references, relation weights, and target-view assignment map C_t fixed, but replace direct aggregation of the

reference assignments with the feature-first route that first recovers the target latent representation and then computes $C_t(\widehat{z}_i^{(t)})$, denoted *w/o Direct Soft-Assignment Estimation*. Its ACC is lower than that of the complete ReSA on all four datasets.

Global–view consensus. We construct *w/o View-specific Assignment*, which keeps only the global assignment induced by completed representations, and *w/o Global Assignment*, which keeps only the view-specific soft assignments. The complete ReSA obtains the highest ACC on all four datasets.

Further analysis of direct soft-assignment estimation. We keep the trained model, the target-view assignment map C_t, the references, the relation weights, and the downstream clustering process fixed. The feature-first estimator recovers a target latent representation with the shared relation and then computes $C_t(\widehat{z}_i^{(t)})$; the direct estimator aggregates the observed target-view reference assignments according to Eq. (11). During evaluation only, we use the complete target view to construct a reference soft assignment. This information is not used for training or missing-information estimation. The mean squared Hellinger distance decreases from 0.05694 for the feature-first estimator to 0.05039 for direct estimation, with a smaller assignment discrepancy in every evaluation setting. A one-sided paired Wilcoxon signed-rank test gives $p = 3.8 \times 10^{-6}$. Thus, under identical references and weights, directly aggregating target-view reference assignments more closely estimates the target view cluster membership of a missing sample.

Missing-view reconstruction. Clustering metrics evaluate the effect of recovered representations on the final clusters, but do not directly show whether target-view content remains in those representations. We therefore pass only the recovered target latent representation to the corresponding target decoder; soft assignments and final consensus outputs are never used for reconstruction. We evaluate Carl (Espinosa-Duró et al., 2013) and Multi-Modal CelebA-HQ (Xia et al., 2021). Carl has Gray, Infrared, and Thermal modalities; each modality is treated as a missing target, and the macro-average over the three directions at $\rho = 0.5$ is reported with mean and standard deviation over seeds 0/1/2. Multi-Modal CelebA-HQ has RGB and Sketch views; both directions are evaluated at $\rho = 0.1$ with seed 2 and 3,000 target samples per direction.

Following the DMVG protocol (Wen et al., 2024), DMVG500 and DMVG1000 use no data augmentation and train for 500 and 1,000 epochs, respectively, whereas DA-DMVG500 uses data augmentation and trains for 500 epochs. We report RMSE, NMSE, PSNR, and SSIM. Lower RMSE and NMSE are better, while higher PSNR and SSIM are better.

Table 4: Missing-view content reconstruction results. Carl: $\rho = 0.5$, macro average over three target modalities, mean±std over seeds 0/1/2. CelebA-HQ: $\rho = 0.1$, seed 2, with 3,000 targets per direction. RMSE/NMSE are lower-is-better; PSNR/SSIM are higher-is-better. "–" denotes unavailable SSIM.

Setting	Method	RMSE ↓	NMSE ↓	PSNR ↑	SSIM ↑
Carl, $\rho = .5$	DMVG500	0.1426±0.0006	0.3514±0.0048	17.033±0.028	–
	DA-DMVG500	0.1381±0.0006	0.3307±0.0040	17.324±0.018	–
	DMVG1000	0.1410±0.0004	0.3439±0.0035	17.145±0.028	–
	ReSA (recovery)	**0.1297±0.0002**	**0.2944±0.0017**	**17.915±0.002**	–
CelebA-HQ: RGB	DMVG500	0.3415	1.4115	9.333	0.1777
	DMVG1000	0.3419	1.4152	9.322	0.1798
	ReSA (recovery)	**0.2524**	**0.7713**	**11.958**	**0.4444**
CelebA-HQ: Sketch	DMVG500	0.3019	1.2300	10.402	0.2959
	DMVG1000	0.3033	1.2409	10.363	0.2921
	ReSA (recovery)	**0.2600**	**0.9124**	**11.699**	**0.5275**

As shown in Table 4, ReSA obtains the lowest RMSE and NMSE and the highest PSNR on Carl at $\rho = 0.5$. On Multi-Modal CelebA-HQ, it likewise obtains lower RMSE and NMSE and higher PSNR and SSIM when RGB or Sketch is the target view. These results indicate that the latent representation recovered using the learned relation retains content that can be decoded by the target-view decoder, providing complementary evidence for the representation recovery branch.

Figure 2 gives a qualitative reconstruction for a fixed Carl sample.

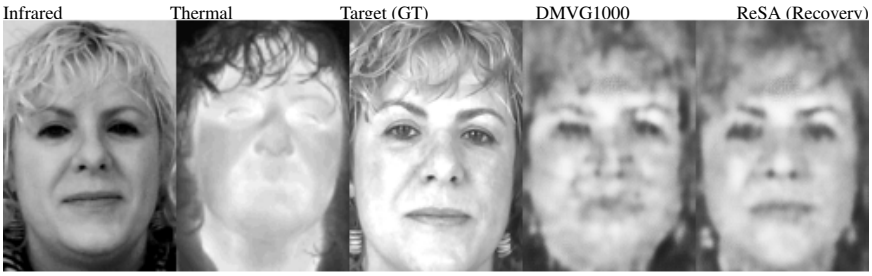


Figure 2: Qualitative reconstruction on Carl with Gray treated as the missing target view at $\rho = 0.5$ and seed 0. The displayed sample (ID 2) is the first example under the fixed sample ordering and was not selected by reconstruction error. All columns correspond to the same instance.

5 Conclusion

We proposed ReSA, an incomplete multi-view clustering method based on representation recovery and direct soft-assignment estimation. The cross-view relation-learning module builds relation representations from view-specific latent representations. CCL and NDC train these representations for cross-view reference retrieval and relation-weight computation. For a sample missing a target view, the relation-guided estimation module selects references and computes their weights, then reuses the same references and weights to aggregate target-view latent representations and soft assignments separately. The resulting estimates recover the missing representation and directly estimate its target-view soft assignment. Finally, global–view consensus combines global clustering information from completed multi-view representations with view-specific membership information from observed or estimated soft assignments. Experiments on multiple IMVC benchmarks and missing rates show that ReSA outperforms the compared methods. Ablations and further analyses support the effectiveness of its components and direct soft-assignment estimation. Missing-view reconstruction further shows that the recovered latent representations retain decodable target-view content.

Reproducibility Statement

The manuscript specifies the data construction, network architecture, training parameters, missing-information estimator, and consensus backend. Anonymous reproducibility materials will provide code, configurations, data and mask identifiers, and training logs. Masked target entries are excluded from representation learning, relation learning, reference selection, relation-weight computation, and soft-assignment estimation.

Ethics Statement

This study introduces no new human-subject data collection and uses established benchmark features and images. Dataset licenses, provenance, and possible demographic limitations of the source data will accompany the reproducibility materials. Clustering outputs may reflect biases in the source features and should not be interpreted as validated sensitive attributes or directly deployed decisions.

AI Use Disclosure

AI was used solely for language polishing of the following text.

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*, 38:11221–11229, 2024. doi: 10.1609/aaai.v38i10.29000.

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*, 40:20289–20297, 2026. doi: 10.1609/aaai.v40i24.39117.

Yuzhuo Dai, Jiaqi Jin, Zhibin Dong, Siwei Wang, Xinwang Liu, En Zhu, Xihong Yang, Xinbiao Gan, and Yu Feng. Imputation-Free and Alignment-Free: Incomplete Multi-View Clustering Driven by Consensus Semantic Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5071–5081, 2025.

Yiming Du, Ziyu Wang, Jian Li, Rui Ning, and Lusi Li. Deep Incomplete Multi-View Clustering via Hierarchical Imputation and Alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 40:20941–20949, 2026. doi: 10.1609/aaai.v40i25.39235.

Virginia Espinosa-Duró, Marcos Faundez-Zanuy, and Jiří Mekyska. A New Face Database Simultaneously Acquired in Visible, Near-Infrared and Thermal Spectrums. *Cognitive Computation*, 5(1):119–135, 2013. doi: 10.1007/s12559-012-9163-2.

Xin Gao and Jian Pu. Deep Incomplete Multi-View Learning via Cyclic Permutation of VAEs. In *International Conference on Learning Representations*, 2025. https://openreview.net/forum?id=s4MwstmB8o.

Hongqing He, Jie Xu, Wenyuan Yang, Yonghua Zhu, Guoqiu Wen, and Xiaofeng Zhu. Global-Graph Guided and Local-Graph Weighted Contrastive Learning for Unified Clustering on Incomplete and Noise Multi-View Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Menglei Hu and Songcan Chen. Doubly Aligned Incomplete Multi-view Clustering. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pp. 2262–2268, 2018. doi: 10.24963/ijcai.2018/313.

Menglei Hu and Songcan Chen. One-Pass Incomplete Multi-View Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1):3838–3845, 2019. doi: 10.1609/aaai.v33i01.33013838.

Dong Huang, Chang-Dong Wang, and Jian-Huang Lai. Fast Multi-View Clustering via Ensembles: Towards Scalability, Superiority, and Simplicity. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11388–11402, 2023. doi: 10.1109/TKDE.2023.3236698.

Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep Incomplete Multi-View Clustering with Cross-View Partial Sample and Prototype Alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11600–11609, 2023.

Harold W. Kuhn. The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955.

Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete Multi-View Clustering via Prototype-Based Imputation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 3911–3919, 2023. doi: 10.24963/ijcai.2023/435.

Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual Contrastive Prediction for Incomplete Multi-View Representation Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4447–4461, 2023. doi: 10.1109/TPAMI.2022.3197238.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. COMPLETER: Incomplete Multi-View Clustering via Contrastive Prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11174–11183, 2021.

Weixiang Shao, Lifang He, and Philip S. Yu. Multiple Incomplete Views Clustering via Weighted Nonnegative Matrix Factorization with $L_{2,1}$ Regularization. In *Machine Learning and Knowledge Discovery in Databases*, 9284:318–334, 2015. doi: 10.1007/978-3-319-23528-8_20.

Huayi Tang and Yong Liu. Deep Safe Incomplete Multi-View Clustering: Theorem and Algorithm. In *Proceedings of the 39th International Conference on Machine Learning*, 162:21090–21110, 2022.

Jingjing Tang, Qingqing Yi, Saiji Fu, and Yingjie Tian. Incomplete Multi-View Learning: Review, Analysis, and Prospects. *Applied Soft Computing*, 153:111278, 2024. doi: 10.1016/j.asoc.2024.111278.

Xinxin Wang, Yongshan Zhang, and Yicong Zhou. Learn Multi-task Anchor: Joint View Imputation and Label Generation for Incomplete Multi-view Clustering. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 6451–6459, 2025a. doi: 10.24963/ijcai.2025/718.

Youqing Wang, Yu Cao, Jinlu Wang, Xiang Xu, Jiapu Wang, Tengfei Liu, Junbin Gao, and Jipeng Guo. Aligning Collaborative View Recovery and Tensorial Subspace Learning via Latent Representation for Incomplete Multi-View Clustering. In *International Conference on Learning Representations*, 2026.

Ziyu Wang, Yiming Du, Yao Wang, Rui Ning, and Lusi Li. Deep Incomplete Multi-view Clustering via Multi-level Imputation and Contrastive Alignment. *Neural Networks*, 181:106851, 2025b. doi: 10.1016/j.neunet.2024.106851.

Jie Wen, Shijie Deng, Waikeung Wong, Guoqing Chao, Chao Huang, Lunke Fei, and Yong Xu. Diffusion-Based Missing-View Generation with the Application on Incomplete Multi-View Clustering. In *Proceedings of the 41st International Conference on Machine Learning*, 235:52762–52778, 2024.

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

Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang, and Jinxing Li. A Survey on Incomplete Multi-View Clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2):1136–1149, 2023b. doi: 10.1109/TSMC.2022.3192635.

Weihao Xia, Yujiu Yang, Jing-Hao Xue, and Baoyuan Wu. Towards Open-World Text-Guided Face Image Generation and Manipulation. *arXiv preprint arXiv:2104.08910*, 2021.

Gehui Xu, Jie Wen, Chengliang Liu, Bing Hu, Yicheng Liu, Lunke Fei, and Wei Wang. Deep Variational Incomplete Multi-View Clustering: Exploring Shared Clustering Structures. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 38:16147–16155, 2024. doi: 10.1609/aaai.v38i14.29548.

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*, 36:8761–8769, 2022. doi: 10.1609/aaai.v36i8.20856.

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. doi: 10.1109/TIP.2025.3529378.

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*, 39(21):22650–22658, 2025. doi: 10.1609/aaai.v39i21.34424.