Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

LRGR: Self-Supervised Incomplete Multi-View Clustering via Local Refinement and Global Realignment

Yanwanyu Xi[1], **Xiao Zheng**[2*], **Chang Tang**[3], **Xingchen Hu**[4], **Yuanyuan Liu**[1], **Jun-Jie Huang**[5] and **Xinwang Liu**[5]
[1]School of Computer Science, China University of Geosciences, Wuhan, China
[2]School of Computer Science, Hubei University of Technology, China
[3]School of Software Engineering, Huazhong University of Science and Technology, China
[4]College of Systems Engineering, National University of Defense Technology, China
[5]College of Computer Science and Technology, National University of Defense Technology, China
{xywy, tangchang, liuyy}@cug.edu.cn, {zhengxiao, jjhuang, xinwangliu}@nudt.edu.cn,
xhu4@ualberta.ca.

Abstract

Incomplete Multi-View Clustering (IMVC) aims to explore comprehensive representations from multiple views with missing samples. Recent studies have revealed that IMVC methods benefit from Graph Convolutional Network (GCN) in achieving robust feature imputation and effective representation learning. Despite these notable improvements, GCN imputation methods often cause a distribution shift between the imputed and original representations, particularly when the neighbors of the imputed nodes are assigned to different groups. Moreover, GCN learning methods tend to produce homogeneous imputed representations, which blur cluster boundaries and hinder effective discriminative clustering. To remedy these challenges, the Local Refinement and Global Realignment (LRGR) Self-supervised model is proposed for incomplete multi-view clustering, which includes two stages. In the first stage, a local imputed refinement module is designed to enhance the versatility of imputed representations through cross-view contrastive learning guided by view-specific prototypes. In the second stage, a global realignment module is introduced to achieve semantic consistency across views, alleviating distribution shifts by leveraging pseudo-labels and their corresponding confidence scores as guidance. Experiments on five widely used multi-view datasets demonstrate the competitiveness and superiority of our method compared to state-of-the-art approaches.

1 Introduction

Deep Multi-view Learning methods have demonstrated unprecedented success in domains such as pattern recognition, computer vision [He *et al.*, 2023], and bioinformatics [He *et al.*, 2024; Zhu *et al.*, 2024], owing to their superior feature extraction capabilities and multi-task adaptability. Among the downstream tasks, Incomplete Multi-View Clustering (IMVC) has emerged as a critical challenge, as it aims to uncover cluster structures in scenarios where data is partially missing across multiple views.

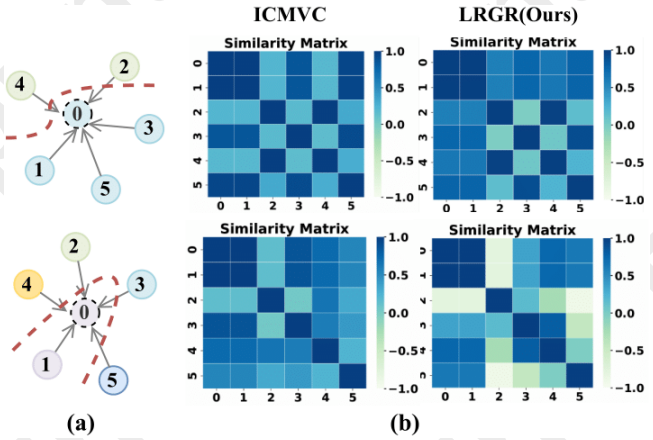


Figure 1: **(a) The distribution shift** arises when neighbors from different clusters are involved. In this example, Node 0 represents the missing sample, while Nodes 1–5 are its neighboring nodes selected from the constructed adjacency matrix. It can be observed that some of these neighbors belong to different clusters, which contributes to the distribution shift issue. **(b) The homogeneous representation** inherent in GCNs. Specifically, the embedded representation learned by ICMVC shows similarity between the imputed node and its neighbor nodes, even when they belong to different clusters. Different from the ICMVC, we observe the similarity matrix obtained by our LRGR achieves better discriminative ability and explores more accurate cluster structure. The imputed representation shows similarity with neighbors in the same cluster and reflects difference with neighbors in the different clusters.

Existing deep IMVC methods can be broadly categorized into imputation-based and imputation-free approaches. Imputation-free methods mainly focus on learning complementary fused representations or aligning different compo-

*Corresponding author.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

nents of representations across views, such as DIMVC [Xu *et al.*, 2022] and APADC [Xu *et al.*, 2023]. On the other hand, imputation-based methods concentrate on completing missing multi-view data by utilizing cross-view similarity relationships or prior knowledge. A critical challenge for these methods is to achieve accurate imputation, which is essential for restoring the underlying data distribution. Existing approaches, such as Completer [Lin *et al.*, 2021], DCP [Lin *et al.*, 2022], SURE [Yang *et al.*, 2022], RPCIC [Yuan *et al.*, 2024], CPSPAN [Jin *et al.*, 2023] and ProImp [Li *et al.*, 2023a], focus on exploring the relationships between views, samples, or prototypes to achieve accurate imputation. However, these methods often face challenges due to insufficient exploration of prior information, resulting in imprecise imputations and blurred cluster boundaries. To address this challenge, several GCN-based IMVC [Feng *et al.*, 2024; Pu *et al.*, 2024] methods have been developed. Due to the inherent message passing mechanisms of GCNs, these methods provide an automatic and efficient approach to handle missing data by aggregating the representations of neighborhood nodes, using the structure information. As a representative GCN-based method, ICMVC [Chao *et al.*, 2024] constructs the adjacency graph based on the multi-view consistency assumption and addresses missing values using GCN.

Although the aforementioned approaches have made significant progress, their reliance on cross-view similarity leads to distribution shifts, particularly when the identified similar neighbors belong to different clusters. This issue becomes challenging at high missing rates, as maintaining cross-view consistency becomes increasingly difficult. Additionally, the message passing mechanism of GCN aggregates neighbors of each node through the constructed adjacency matrix. As the number of propagation steps increases, the node embeddings become progressively similar to those of their neighbors. To better illustrate above two issues, we empirically analyze and illustrate them on the CUB dataset under the 50% and 70% missing rates. As shown in Fig. 1, the first column displays a subgraph of CUB, highlighting the imputed sample and its neighboring nodes selected from the adjacency matrix. Nodes from the same cluster are annotated in the same color. We visualize the similarity matrix of the imputed node and its corresponding neighbors, which is obtained using different GCN-based methods.

To address the distribution shift caused by inaccurate neighbors and the homogeneous representation inherent in the GCN, we propose a novel self-supervised local refinement and global realignment (LRGR) model for incomplete multi-view clustering. Specifically, we introduce the global semantic realignment module to generate consistent pseudo-labels and compute confidence scores, enabling semantic-level contrastive learning to mitigate the impact of incorrect neighbor aggregation for missing samples. To tackle the second issue, we design a local imputed refinement module that preserves the view versatility of the imputed representations, guided by view-specific prototypes. Overall, the contributions of this paper can be summarized as follows:

- From the aspect of data imputation in GCN, we observe the potential distribution shift issue caused by the inaccurate neighbors and propose a self-supervised realignment module to alleviate this shift problem, named global realignment;
- From the aspect of representation learning in GCN, we observe the underlying similar imputed representations and propose a local refined module to emphasize the versatility, named local refinement;
- Extensive experimental results verify the effectiveness of LRGR and show considerable improvement over ten state-of-the-art methods on five multi-view datasets.

2 Related Work

IMVC is an important branch of the MVC field that has attracted widespread research. Classic IMVC methods can be roughly categorized into three types: kernel-learning-based methods [Liu *et al.*, 2019a; Liu *et al.*, 2019b; Li *et al.*, 2023b], matrix-factorization-based methods [Hu and Chen, 2018], and graph-based methods [Wang *et al.*, 2022b; Wen *et al.*, 2023; Li *et al.*, 2022]. However, their shallow architectures often struggle to effectively handle complex datasets. In response, DL-based IMVC methods have emerged, and can be further subdivided into Autoencoder-based methods [Lin *et al.*, 2021; Lin *et al.*, 2022; Xu *et al.*, 2023], Graph Convolutional Network (GCN)-based methods [Chao *et al.*, 2024; Pu *et al.*, 2024] and Generative Adversarial Networks (GANs)-based methods [Wang *et al.*, 2022a; Zhang *et al.*, 2020]. In addition, to address the semantic alignment problem, several methods such as SPICE [Niu *et al.*, 2022] and PLP-L/G [Feng *et al.*, 2024] have been proposed. These approaches treat pseudo-labels as reliable supervision signals and align network predictions using KL divergence or cross-entropy loss. In contrast, our proposed LRGR framework focuses on resolving the distribution inconsistency of imputed samples across views. Among existing GCN-based IMVC methods, ICMVC is the most closely related to our work. While both approaches utilize GCNs for view imputation, their focuses differ significantly. ICMVC primarily exploits complementary information from high-confidence samples, aiming to enhance representation by aggregating reliable neighborhood information. In contrast, our proposed method focuses on mitigating the homogeneous representation issue inherent in GCN by leveraging view-specific information and addressing the distribution shift caused by pseudo-label guided.

3 Method

3.1 Notations and Overview

We define the incomplete multi-view data including N instances with v views as $\mathbf{X} = \{\mathbf{X}^{(0)}, ..., \mathbf{X}^{(v)}\}$, where $\mathbf{X}^{(v)} \in \mathbb{R}^{N^v \times d_v}$. The latent representations obtained by GCN are defined as $\mathbf{Z}^{(v)} = \{\mathbf{Z}_c^{(v)}, \mathbf{Z}_p^{(v)}\}$, where $\mathbf{Z}_c^{(v)} \in \mathbb{R}^{N_c^v \times d}$ and $\mathbf{Z}_p^{(v)} \in \mathbb{R}^{N_p^v \times d}$ denote the complete and imputed feature matrix of the v-th view, N_c^v and N_p^v are the number of complete, imputed samples of the v-th view, d is the latent dimension. And then we construct the indicator vector $P^{(v)}$,

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

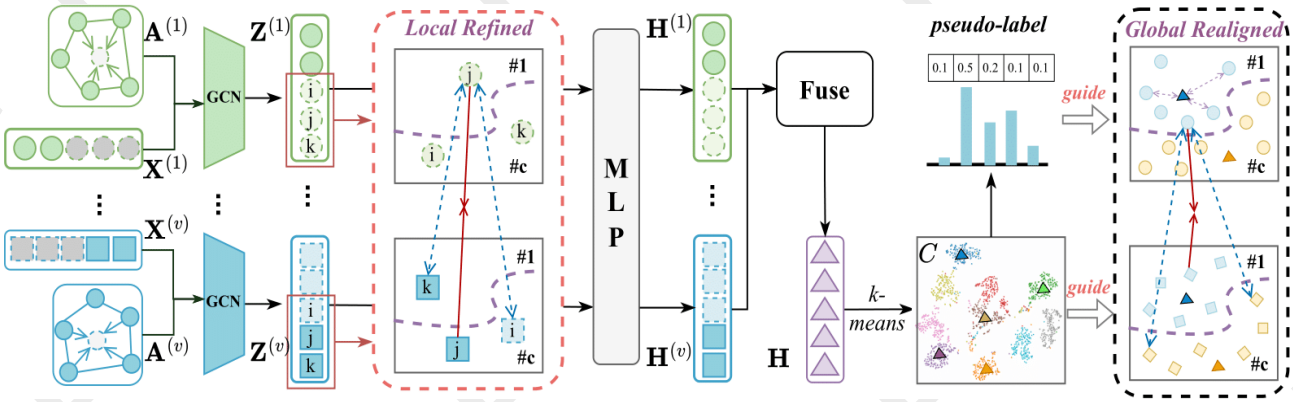


Figure 2: Overview of the proposed LRGR method. We first construct the view-specific adjacency matrix $\mathbf{A}^{(v)}$ by k-nearest neighbor. Then we impute the view-specific adjacency matrix $\mathbf{A}^{(v)}$ using the cross-view similarity transfer. After that, we send the feature matrix $\mathbf{X}^{(v)}$ and adjacency matrix $\mathbf{A}^{(v)}$ to GCN, to get the imputed representation $\mathbf{Z}^{(v)}$. i) **The local imputed refinement module**. To address the issue of homogeneous imputed representations caused by the message-passing mechanism, we leverage view-specific prototypes and contrastive learning to enhance the versatility of the representations. ii) **The global semantic realignment module**. To address the distribution shift problem caused by inaccurate neighbors, we use pseudo-labels and compute their confidence scores to preserve consistency at the semantic level.

where $p_i^{(v)} = 1$ if the i-th sample is missing in v-th view, otherwise $p_i^{(v)} = 0$.

In this section, we propose a novel self-supervised local refinement and global realignment (LRGR) model for IMVC. As illustrated in Fig. 2, LRGR mainly consists of two components, $i.e.$ a local imputed refined module to prevent homogeneous representations and a global self-supervised representation module to alleviate the distribution shift issue. Next, we will introduce the two modules and the final loss function in detail.

3.2 Missing View Imputation

To fully leverage structural information for missing data imputation by GCN, we first construct the view-specific adjacency matrix based on the view consistency assumption.

Missing Adjacency Matrix Handling

To better capture the data structure, we compute the similarity matrix of each view, denoted as $\{\mathbf{S}^{(v)}\}_{v=1}^{V}$ by normalized cosine function. For the missing samples, their corresponding similarity relationships cannot be directly computed. To address this, following [Chao *et al.*, 2024], we transfer the cross-view similarity relationships to the missing samples based on the consistency assumption. As shown in Fig. 3, the column corresponding to the missing sample in the similarity matrix is set to zero, and the missing row is imputed using the cross-view similarity. Mathematically, we first constructed the similarity matrix $\mathbf{S}^{(v)}$ by cosine similarities.

$$\mathbf{S}^{(v)} = \frac{\mathbf{X}^{(v)} \cdot \mathbf{X}^{(v)^T}}{||\mathbf{X}^{(v)}|| \cdot ||\mathbf{X}^{(v)}||}, \tag{1}$$

where $\mathbf{S}^{(v)} \in \mathbb{R}^{N \times N}$ denotes the similarity matrix for each view. Next, we use the indicator vector $P^{(v)}$ to select the existing cross-view similarity values for the similarity transfer operation:

$$\mathbf{S}^{(v)}[P^{(v)}, :] = \frac{1}{m} \sum\nolimits_{u \neq v} \mathbf{S}^{(u)}[P^{(v)}, :], \tag{2}$$

where $\mathbf{S}^{(v)}[P^{(v)}, :]$ selects the missing rows of the v-th similarities matrix, $\mathbf{S}^{(u)}[P^{(v)}, :]$ selects the existing rows of cross-view similarities matrix. After the missing similarity transfer, we construct the adjacency matrix $\mathbf{A} = \{\mathbf{A}^{(1)}, ..., \mathbf{A}^{(v)}\}$, where $a_{i,j}^{(v)} = \mathbf{A}^{(v)}[i,j] = 1$ if there is an edge between $x_i^{(v)}$ and $x_j^{(v)}$, otherwise $a_{i,j}^{(v)} = 0$, based on k-nearest neighbor. Then, to stabilize the training of the GCN, the adjacency matrix is normalized through calculating $\tilde{D}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{D}^{-\frac{1}{2}}$, where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is the self-connection, $\tilde{D} = \sum_j \tilde{a}_{ij}$ is the degree matrix.

GCN-based Instance Imputation

The core idea of GCN imputation is to leverage local structural information to iteratively update the representation of each node by aggregating features from its neighboring nodes. Specifically, after obtaining the complete adjacency matrix for each view, we feed both the adjacency matrix $\mathbf{A}^{(v)}$ and the feature matrix $\mathbf{X}^{(v)}$ into a three-layer GCN encoder. The node representations of each view $\mathbf{Z}^{(v)}$ can be updated by the graph convolution operation of the m-th layer, as described below:

$$\mathbf{Z}_{(m)}^{(v)} = \phi(\tilde{D}^{-\frac{1}{2}} \tilde{\mathbf{A}}^{(v)} \tilde{D}^{-\frac{1}{2}} \mathbf{Z}_{(m-1)}^{(v)} W_{(m)}^{(v)}), \quad \mathbf{Z}_{(0)}^{(v)} = \mathbf{X}^{(v)}, \tag{3}$$

where $\mathbf{Z}_{(m)}^{(v)}$ is the output embeddings of the m-th layer, $W_{(m)}^{(v)}$ indicates the trainable parameters in the m-th layer of v-th graph encoder. After that, we adopt a shared MLP layer $\mathbf{H}^{(v)} = MLP(\mathbf{Z}^{(v)}; \theta_{shared})$. to capture consistent patterns. Then, we utilize following reconstruction loss:

$$\mathcal{L}_{rec} = \frac{\sum_{i=1}^{v} (\tilde{\mathbf{X}}^{(i)} - \mathbf{X}^{(i)})}{N}, \tag{4}$$

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

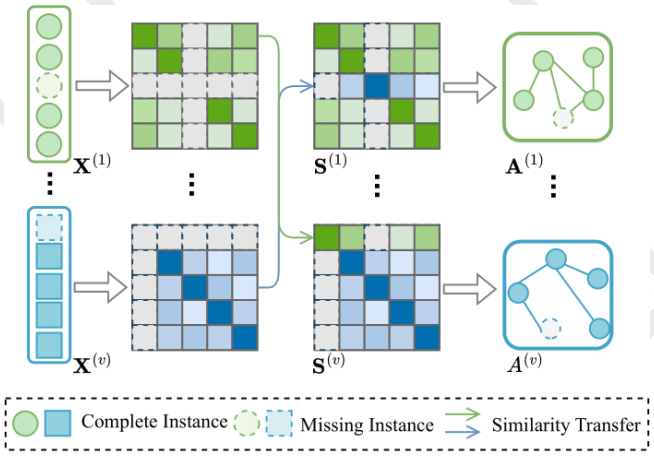


Figure 3: Similarity matrix imputation and adjacency matrix construction. In this step, we perform similarity matrix imputation using cross-view transfer. Specifically, the column corresponding to the missing sample in the similarity matrix is set to 0, and the corresponding row is imputed using the cross-view similarity.

where N denotes the number of the sample, $\tilde{\mathbf{X}}^{(i)} = \mathbf{D}^{(v)}(\mathbf{H}^{(v)}, \theta^{(v)})$ is the reconstructed representation obtained by view-specific decoders.

3.3 Local Imputed Refinement Module

Existing imputation methods rely on the view consistency assumption, which is typically satisfied at the cost of view versatility. Furthermore, the inherent message-passing mechanism in GCNs often increases the similarity between neighboring nodes [Xia *et al.*, 2022]. For missing samples, this exacerbates the issue of reduced versatility, resulting in a decreased discriminability of the imputed samples. To preserve the diversity of imputed samples, we leverage view-specific prototypes $\mathbf{C}^{(v)} \in \mathbb{R}^{k\times d}$ and contrastive learning to promote greater dissimilarity between imputed samples. Specifically, we select the imputed samples from each view $\mathbf{Z}_p^{(i)} \in \mathbb{R}^{N_p^i \times d}$ and their corresponding counterparts in cross-view $\mathbf{Z}_p^{(j)} \in \mathbb{R}^{N_p^i \times d}$ to form positive and negative sample pairs. We then adopt the following contrastive loss, which maximizes the similarity of assignment probabilities between the same samples across views, while simultaneously maximizing the difference in assignment probabilities between different instances.

$$\mathcal{L}_{loc} = -\frac{1}{N}\sum_{m}^{v}\sum_{n=m+1}^{v}\sum_{i}^{N}(\log\frac{e^{s(p_i^{(m)},p_i^{(n)})/\tau}}{\sum_{j=1}^{N}e^{s(p_i^{(m)},p_j^{(n)})/\tau}}), \tag{5}$$

$$p_{ik}^{(v)} = \frac{(1+||z_i^{(v)}-c_k^{(v)}||^2)^{-1}}{\sum_k(1+||z_i^{(v)}-c_k^{(v)}||^2)^{-1}}, \tag{6}$$

where $z_i^{(v)}$ denotes the representation of the i-th sample in the v-th view, $c_k^{(v)}$ represents the k-th prototypes in the v-th view, $p_{ik}^{(v)}$ denotes the probability of the i-th sample assigned to k-th cluster in v-th view, N denotes the number of samples, $s(\cdot,\cdot)$ is the cosine similarity, τ is the temperature parameter.

3.4 Global Semantic Realignment Module

Adaptive Fusion Block

To better extract the common pattern and evaluate the performance, we adopt an adaptive fusion block consisting of two convolutional layers. First, we concatenate the latent representations $\mathbf{H}^{(v)}$ from each view along the channel dimension:

$$\mathbf{H}_{con} = Concat(\mathbf{H}^{(1)}, ..., \mathbf{H}^{(v)}), \tag{7}$$

where $\mathbf{H}_{con} \in \mathbb{R}^{[N,view,d]}$. Then we pass the concatenated feature matrix $\mathbf{H}_{con}$ through a two-layer convolutional network. The first convolutional layer applies a 1D convolution operation with n kernels along the channel dimension, expanding the number of channels from the number of views to n. In this paper, n is set as 32.

$$\mathbf{H}_{con}^{'} = Conv1d_{(0)}(\mathbf{H}_{con}, W_{(0)}, b_{(0)}). \tag{8}$$

The second convolutional layer also uses a 1D convolution operation, a single convolutional kernel is applied across the feature channel dimension, which reduces the number of feature channels from n to 1.

$$\mathbf{H} = Conv1d_{(1)}(\mathbf{H}_{con}^{'}, W_{(1)}, b_{(1)}), \tag{9}$$

where $\mathbf{H} \in \mathbb{R}^{[N,1,d]}$ means the fused common representation, $W_{(.)}$ and $b_{(.)}$ are trainable parameters.

Semantic-wise Realignment

The local Refinement module enhances the diversity of imputed representations. However, this also means that it pushes samples away without considering their semantic categories, potentially causing a distribution shift. To alleviate this issue, we achieve the global realignment by preserving semantic consistency. Inspired by [Liu *et al.*, 2024], we compute the semantic pseudo-labels $l \in \mathbb{R}^N$ and global prototypes $\mathbf{C} \in \mathbb{R}^{[k,d]}$ by performing k-means clustering on the fused representation $\mathbf{H}$. To evaluate the confidence of the pseudo-labels, we calculate the t-student distribution between the view-specific embeddings $\mathbf{H}^{(v)}$ and the global prototypes $\mathbf{C}$, using the result as the confidence score for the pseudo-labels.

$$q_{ik}^{(v)} = \frac{(1+\left\|h_i^{(v)}-c_k\right\|^2)^{-1}}{\sum_k(1+\left\|h_i^{(v)}-c_k\right\|^2)^{-1}}, \tag{10}$$

where $q_{ik}^{(v)}$ denotes the probability of the i-th view assigned to the k-th cluster in v-th view. Finally, the semantic alignment loss is computed using semantic pseudo-labels and corresponding confidence score. For samples with high confidence, the loss is designed to reinforce the model's certainty in its pseudo-labels, encouraging samples with the same labels to be grouped together while pushing those with different labels apart.

$$\mathcal{L}_{glob} = -\frac{1}{N}\log\sum_{i=1}^{N}\frac{\sum_{j=1}\mathbb{I}_{[l_i=l_j=c]}\exp((q_{ic}^{(v)},q_{jc}^{(v)})/\tau)}{\sum_{k=1}^{N}\mathbb{I}_{[l_i\neq l_k]}\exp(s(q_i^{(v)},q_k^{(v)})/\tau)}, \tag{11}$$

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Missing rates	Method	COIL20			CUB			YouTubeFace10			DHA			UCI		
		ACC	NMI	F-score	ACC	NMI	F-score	ACC	NMI	F-score	ACC	NMI	F-score	ACC	NMI	F-score
Complete	DPML(2020)	68.40	77.92	68.40	76.00	74.38	74.96	73.57	77.78	74.23	72.88	77.69	73.47	72.15	68.33	72.25
	UIMC(2021)	46.71	60.85	37.76	75.20	76.41	68.58	O/M	O/M	O/M	75.90	**85.18**	70.09	83.54	83.88	79.01
	DSIMVC(2022)	62.78	68.75	61.89	66.17	58.97	65.86	72.10	71.58	72.38	63.33	58.83	63.18	81.80	78.80	81.68
	DIMVC(2022)	63.82	79.20	61.99	67.83	67.90	65.48	72.33	76.10	69.88	61.08	73.42	61.48	92.80	86.40	92.82
	IMVC-CBG(2022)	69.18	79.26	63.08	74.69	74.00	67.71	76.30	80.66	74.01	76.05	84.32	69.12	75.83	72.39	68.47
	ProImp(2023)	72.22	79.49	71.76	79.00	76.10	78.46	52.58	58.41	52.85	66.46	80.00	64.10	80.40	77.55	80.17
	DCP(2023)	68.61	77.75	63.86	71.67	74.72	70.00	67.24	70.57	69.48	NaN	NaN	NaN	78.60	78.66	76.20
	DSIMVC++(2023)	63.19	74.97	63.31	71.93	66.36	71.68	73.94	75.50	74.15	63.77	74.89	62.31	81.70	80.33	81.54
	RPCIC(2024)	72.78	81.82	71.34	75.50	72.54	74.32	74.06	79.15	72.85	71.22	77.43	70.43	92.15	85.88	92.20
	ICMVC(2024)	72.36	82.83	72.37	84.33	76.16	84.24	71.03	72.06	69.61	72.46	79.46	71.45	82.20	79.62	82.18
	LRGR(ours)	**76.32**	**84.25**	**76.01**	**84.50**	**79.85**	**84.33**	**78.37**	**82.08**	**76.42**	**75.78**	80.54	**75.73**	**94.15**	**88.18**	**94.18**
Incomplete	DPML(2020)	71.67	78.27	71.78	59.67	58.92	57.39	74.38	79.18	74.47	60.66	71.81	59.19	64.25	59.10	61.10
	UIMC(2021)	47.92	59.78	35.78	55.98	52.01	41.20	O/M	O/M	O/M	48.22	55.19	24.71	68.61	56.54	52.61
	DSIMVC(2022)	57.50	67.19	55.77	56.17	50.44	56.22	71.03	74.81	73.74	59.67	56.11	59.71	78.90	75.44	78.46
	DIMVC(2022)	63.61	79.08	62.51	47.96	54.94	44.39	72.28	77.42	72.14	60.17	74.72	60.09	82.20	65.11	82.16
	IMVC-CBG(2022)	65.30	75.52	59.73	61.24	57.36	49.92	70.44	72.31	61.69	60.71	71.21	47.80	68.10	61.03	50.74
	ProImp(2023)	27.01	39.69	25.02	67.17	**68.79**	65.22	53.82	62.24	52.02	60.25	**75.30**	57.19	77.95	71.98	77.59
	DCP(2023)	59.93	73.02	59.31	57.17	52.58	61.90	56.50	62.43	58.22	NaN	NaN	NaN	77.75	75.22	76.38
	DSIMVC++(2023)	56.46	68.60	53.52	63.33	55.62	63.02	68.39	71.37	69.36	50.84	64.88	49.61	80.20	73.44	80.00
	RPCIC(2024)	62.36	74.92	62.24	64.50	64.24	62.62	74.26	74.76	73.64	39.54	47.13	47.27	84.10	76.85	84.22
	ICMVC(2024)	**75.28**	**82.93**	73.47	62.17	59.42	62.32	77.29	78.49	77.24	58.34	67.87	58.39	81.35	77.58	81.10
	LRGR(ours)	73.89	76.72	**73.64**	**67.67**	65.26	**66.47**	**84.68**	**84.79**	**84.38**	**62.73**	71.26	**62.22**	**87.70**	**77.66**	**87.74**

Table 1: Experiments on Five Data Sets. The Best And Second-Best Results are Highlighted with **Bold** and Underline Respectively

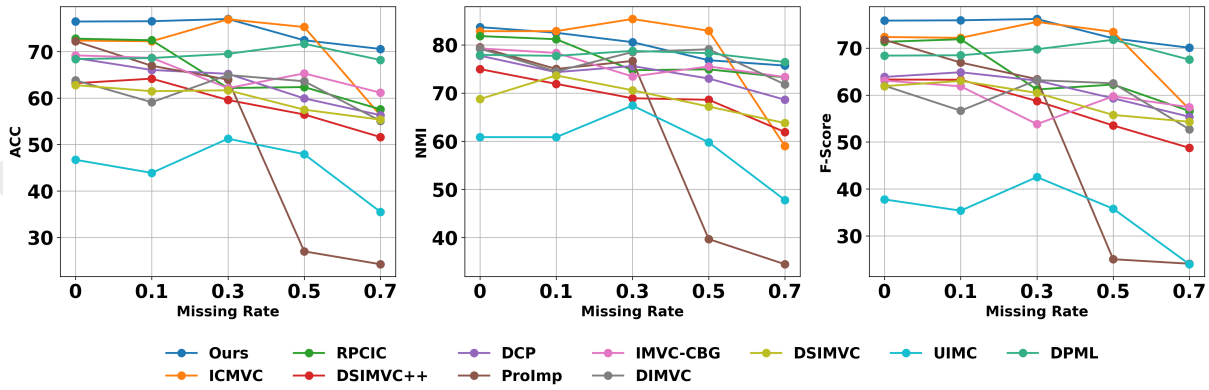


Figure 4: Clustering result on the COIL20 dataset under 0.1, 0.3, 0.5, and 0.7 missing rates.

where $q_{ic}^{(v)}$ is the probability of the i-th sample belongs to the c-th cluster in v-th view, $s(\cdot,\cdot)$ denotes the cosine similarity and l_i represents the pseudo-label of the i-th sample.

3.5 Loss Function

Finally, we combine the loss functions of the reconstruction loss, local refinement loss and global realignment loss into an overall loss function, which is expressed as follows:

$$\mathcal{L} = \mathcal{L}_{rec} + \alpha\mathcal{L}_{loc} + \beta\mathcal{L}_{glob}, \tag{12}$$

where $\mathcal{L}_{rec}$, $\mathcal{L}_{loc}$ and $\mathcal{L}_{glob}$ are defined in Eq. (4), Eq. (5) and Eq. (11), α and β are the trade-off parameters, which balance the importance between the local imputed refinement loss and the global semantic realignment loss.

4 Experiment

In this section, we evaluate the proposed LRGR method on five widely-used multi-view datasets compared with 10 state-of-the-art baselines. First, we present the basic information of experimental settings, like dataset and evaluation metrics. In Sec. 4.2, we give a detailed analysis between our LRGR and comparison methods. Then in Sec. 4.3, we conduct comprehensive ablation studies to demonstrate the effectiveness of LRGR. Finally, we give visualization results.

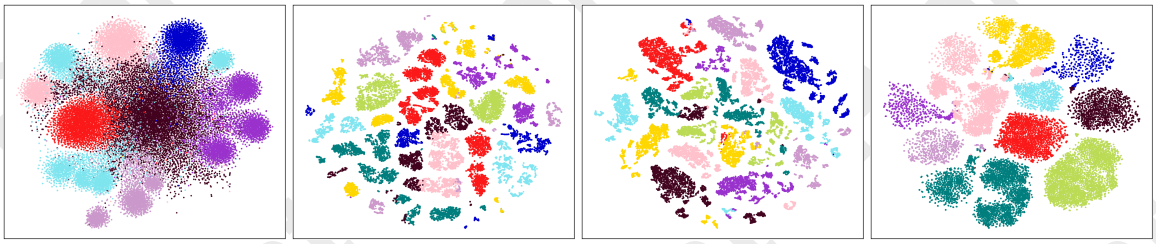
4.1 Experimental Settings

We evaluate our proposed model on five widely used datasets. Specifically, CUB [Wah *et al.*, 2011] with 600 samples, 2 views. DHA [Lin *et al.*, 2012] with 383 samples, 2 views. COIL20 [Wan *et al.*, 2021] with 1440 samples, 3 views. YTF10 [Wolf *et al.*, 2011] with 38654 samples, 4 views, and UCI [Asuncion *et al.*, 2007] with 2000 samples, 3 views. To assess performance in an incomplete data setting, we manually mask data at various missing rates. Specifically, for a dataset with v views, we randomly select m samples as missing and remove $1 \sim v-1$ views from each. The missing rate η is defined as $\eta = \frac{m}{n}$, where n is the number of samples. We implement our method using PyTorch 1.12 and conduct experiments on a standard Ubuntu 20.04.6 OS with an NVIDIA GTX 1060 GPU. The activation function is the RELU [Glo-

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Loss			CUB			DHA			COIL20			UCI		
$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{loc}}$	$\mathcal{L}_{\text{glb}}$	ACC	NMI	FS	ACC	NMI	FS	ACC	NMI	FS	ACC	NMI	FS
✓			65.33	62.27	64.65	61.70	71.68	60.96	71.67	75.86	71.60	70.80	70.06	70.43
✓		✓	65.67	64.28	64.93	63.34	71.30	63.77	71.39	76.61	70.97	74.20	71.53	73.78
✓	✓		66.17	65.05	64.51	62.73	71.48	62.06	72.22	76.44	72.05	86.90	76.53	86.90
✓	✓	✓	67.67	65.26	66.47	62.73	71.26	62.22	73.26	76.24	72.79	87.45	77.21	87.50

Table 2: Ablation studies with three losses on four datasets under 50% missing rate, where ✓ denotes the loss is adopted.



(a) IMVC-CBG (2022) (b) RPCIC (2024) (c) ICMVC (2024) (d) LRGR (Ours)

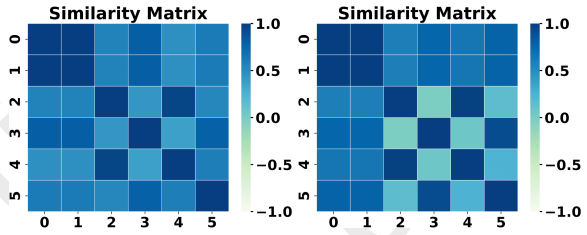
Figure 5: The t-SNE visualization results on YouTubeFace10 dataset with missing rate 0.5.

rot *et al.*, 2011]. Adam [Kingma, 2014] is chosen as the optimizer with the 0.0005 and 0.0001 learning rate in the pre and align training process. In pretrain stage, we update the network only with reconstruction loss in Eq.(4). And in the second stage, we update the network with Eq.(12). For comparison methods, we use the recommended parameters and network structures to ensure the best performance. For evaluation metrics, we utilize three metrics including clustering accuracy (ACC) [Cai *et al.*, 2005], normalized mutual information (NMI) [Knops *et al.*, 2006] and F-score[Powers, 2008] to evaluate performance.

4.2 Comparisons with State of the Arts

To verify the effectiveness of our model, we choose ten state-of-the-art IMVC baselines, including DPML [Zhang *et al.*, 2020], UIMC [Fang *et al.*, 2021], DSIMVC [Tang and Liu, 2022], DIMVC [Xu *et al.*, 2022], IMVC_CBG [Wang *et al.*, 2022b], ProImp [Lin *et al.*, 2022], DCP [Lin *et al.*, 2022], DSIMVC++ [Yan *et al.*, 2023], RPCIC [Sun *et al.*, 2024] and ICMVC [Chao *et al.*, 2024].

We first evaluate LRGR and comparisons under the Complete situation (with the missing rate of 0%) and Incomplete (with the missing rate of 50%). Table 1 demonstrates the clustering performance on ACC, NMI and F-Score. It can be observed that our LRGR significantly outperforms the state-of-the-art methods on all datasets with the missing rate of 0% and on CUB, YouTubeFace10, DHA and UCI datasets with the missing rate of 50%. Specifically, LRGR achieves a 1.42% (82.08% v.s. 80.66%) and 5.89%(84.38% v.s. 78.49%) improvement on NMI under the Complete and Incomplete scenarios on the YouTubeFace10 dataset, compared with the second best result. This increasing performance shows the effectiveness of the proposed local imputation refinement module and global semantic realignment module. To better prove the robustness of proposed LRGR, we illustrate the clustering results by increasing the missing rate from 10% to 70% with a gap of 20% on the COIL20 dataset. As shown in the Fig. 4, our LRGR substantially achieves the comparable results on three metrics on the COIL20 dataset. On the other hand, the robustness demonstrate the effectiveness of our imputation mechanism and also the global semantic realignment module.



(a) No Prototype (b) With Prototype

Figure 6: The ablation results on the CUB dataset under 50% missing rate. (a) The similarity heatmap without the view-specific prototypes in local imputed refinement module. (b) The similarity heatmap with the view-specific prototypes in local imputed refinement module.

4.3 Ablation Studies

In this section, to better understand the LRGR, we conduct a series of ablation studies. We first study and visualize the effectiveness of local imputed refinement module. Then we study the effectiveness of shared and adaptive fusion module. Finally, we conduct experiments with different combination of loss items in LRGR.

Effectiveness of Local Refinement

To demonstrate the model's ability to capture versatile information guided by view-specific prototypes, we visualize the similarity between an imputed node and its selected neighboring nodes on the CUB dataset, under a 50% missing rate,

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

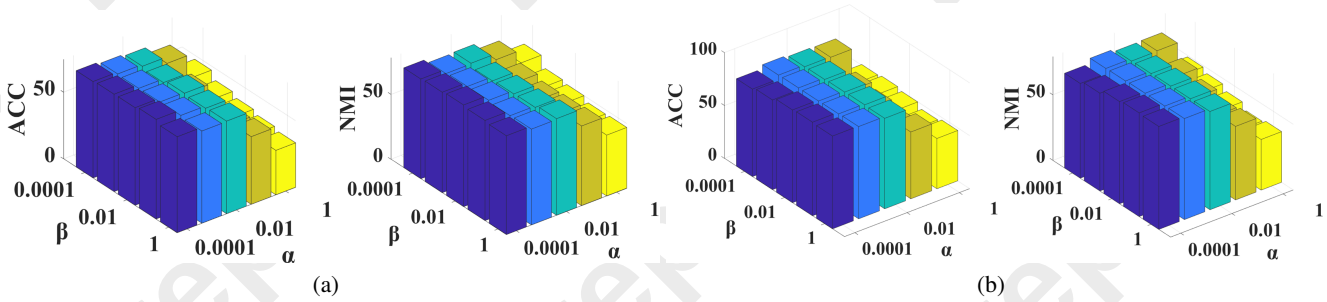


Figure 7: Parameter sensitivity analysis. The clustering performance of LRGR under different trade-off coeficient α and β on (a) COIL20 datatset and (b) UCI dataset under 50% missing rate.

without the guidance of prototypes in the local refinement module. As shown in Fig. 6 (a), we observe that the learned imputed representation retains some discriminative power, compared with Fig. 6 (b), but does not exhibit the same degree of distinction as in Fig. 6 (b). Such results demonstrate the effectiveness and superiority of the proposed local refinement module, particularly its capability to capture versatile information and avoid homogeneous representations.

Datasets	CUB		DHA		UCI		COIL20	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
w/o AFB	63.67	59.83	59.83	72.55	75.10	73.89	73.33	79.05
w/o MLP	60.00	57.54	56.73	71.18	78.70	69.13	62.92	67.84
LRGR	67.17	61.77	61.49	70.71	86.70	77.16	75.07	82.53

Table 3: Ablation studies without MLP layer and AFB on CUB, DHA, COIL20 and UCI datasets under 50% missing rate.

Effectiveness of Adaptive Fusion and MLP

To verify the effectiveness of the proposed adaptive fusion module, we perform ablation studies on four datasets with a 50% missing rate, where the fusion module is removed and the representations from all views are fused via an average weighted sum. To verify the effectiveness of the shared MLP layer, we conduct the ablation studies on four datasets with a 50% missing rate, where the MLP layer is removed.

Effectiveness of Loss Term

To further verify the effectiveness of each module in our proposed model, we conduct ablation experiments with different loss combinations on the UCI, COIL20 and CUB datasets under the 50% missing rate. The result of ablation study is presented in Table 2. Concretely speaking, while the reconstruction loss term $\mathcal{L}_{\text{rec}}$ can achieve comparable results on several datasets, it fails to learn discriminative representations and cannot prevent the emergence of homogeneous representations. In contrast, the proposed local imputed refinement loss effectively maintains view versatility, guided by view-specific prototypes.

4.4 Hyperparameters Sensitivity Analysis

In this study, we investigate the trade-off coefficients α and β in the loss function. We follow a common analysis strategy between local imputed refinement loss and global realignment loss. Specifically, we set the α and β varies from $[0.0001, 0.001, 0.01, 0.1, 1]$. As shown in Fig. 7, when α is too large, the global realignment loss becomes dominant, leading to a homogeneous representation. Conversely, when β is too large, the model tends to overly prioritize versatility at the cost of learning meaningful representations, thus failing to maintain semantic consistency. For proposed LRGR, the trade-off coefficient α exhibits robustness in [0.0001 0.01], the trade-off coefficient β exhibits robustness in [0.0001, 1].

4.5 Visualizations

To more effectively assess the performance of the proposed LRGR, we visualize the clustering results using t-SNE. The results for our method, along with four recent baseline methods, are presented on the YouTubeFace10 dataset at a 50% missing rate, as shown in Fig. 5. From the figure, it is clear that IMVC-CBG struggles to distinguish between clusters, with blurred inter-cluster boundaries and undefined intra-cluster structures. For ICMVC, and RPCIC, although they show relatively clear inter-cluster boundaries, there is some overlap between clusters. In contrast, LRGR demonstrates clear cluster boundaries and compact intra-cluster structures, indicating its ability to learn a well-defined clustering structure even in the presence of missing data.

5 Conclusion

In this paper, we propose a novel Local Imputed Refinement and Global Semantic Realignment (LRGR) model to address the challenges of similar imputed representations and distribution shifts. Unlike existing methods that primarily focus on shared patterns across views, our Local Imputed Refinement module explicitly preserves the diversity information in the imputed representations, guided by view-specific prototypes. Furthermore, to alleviate the distribution shift caused by inaccurate neighbors, we ensure semantic consistency through the pseudo-labels and their corresponding confidence scores. Extensive experiments demonstrate the effectiveness and superiority of the proposed LRGR model.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Acknowledgements

The work was supported in part by the National Natural Science Foundation of China under grants 62476258, and in part by the Natural Science Foundation of Hubei Province under grant 2025AFA113, and in part by the Key Laboratory of Target Cognition and Application Technology (No. 2023-CXPT-LC-005), and in part by the Fundamental Research Funds for National Universities, China University of Geosciences (Wuhan) under grant 2024XLB6.

References

[Asuncion *et al.*, 2007] Arthur Asuncion, David Newman, et al. Uci machine learning repository, 2007.

[Cai *et al.*, 2005] Deng Cai, Xiaofei He, and Jiawei Han. Document clustering using locality preserving indexing. *IEEE Transactions on Knowledge and Data Engineering*, 17(12):1624–1637, 2005.

[Chao *et al.*, 2024] Guoqing Chao, Yi Jiang, and Dianhui Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 11221–11229, 2024.

[Fang *et al.*, 2021] Xiang Fang, Yuchong Hu, Pan Zhou, and Dapeng Oliver Wu. Unbalanced incomplete multi-view clustering via the scheme of view evolution: Weak views are meat; strong views do eat. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(4):913–927, 2021.

[Feng *et al.*, 2024] Cong Feng, Ao Li, Haoyue Xu, Hailu Yang, and Xinwang Liu. Deep incomplete multiview clustering via local and global pseudo-label propagation. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[Glorot *et al.*, 2011] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.

[He *et al.*, 2023] Xiao He, Chang Tang, Xin Zou, and Wei Zhang. Multispectral object detection via cross-modal conflict-aware learning. In *Proceedings of ACM International Conference on Multimedia*, pages 1465–1474, 2023.

[He *et al.*, 2024] Xiao He, Chang Tang, Xinwang Liu, Chuankun Li, Shan An, and Zhenglai Li. Heterogeneous graph guided contrastive learning for spatially resolved transcriptomics data. In *Proceedings of the ACM International Conference on Multimedia*, pages 8287–8295, 2024.

[Hu and Chen, 2018] Menglei Hu and Songcan Chen. Doubly aligned incomplete multi-view clustering. *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2262–2268, 2018.

[Jin *et al.*, 2023] 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*, pages 11600–11609, 2023.

[Kingma, 2014] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Knops *et al.*, 2006] Zeger F Knops, JB Antoine Maintz, Max A Viergever, and Josien PW Pluim. Normalized mutual information based registration using k-means clustering and shading correction. *Medical Image Analysis*, 10(3):432–439, 2006.

[Li *et al.*, 2022] Zhenglai Li, Chang Tang, Xiao Zheng, Xinwang Liu, Wei Zhang, and En Zhu. High-order correlation preserved incomplete multi-view subspace clustering. *IEEE Transactions on Image Processing*, 31:2067–2080, 2022.

[Li *et al.*, 2023a] Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3911–3919, 2023.

[Li *et al.*, 2023b] Zhenglai Li, Chang Tang, Xiao Zheng, Zhiguo Wan, Kun Sun, Wei Zhang, and Xinzhong Zhu. Mutual structure learning for multiple kernel clustering. *Information Sciences*, 647:119445, 2023.

[Lin *et al.*, 2012] Yan-Ching Lin, Min-Chun Hu, Wen-Huang Cheng, Yung-Huan Hsieh, and Hong-Ming Chen. Human action recognition and retrieval using sole depth information. In *Proceedings of the ACM international conference on Multimedia*, pages 1053–1056, 2012.

[Lin *et al.*, 2021] 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*, pages 11174–11183, 2021.

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

[Liu *et al.*, 2019a] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Chang Tang, En Zhu, Jianping Yin, and Wen Gao. Efficient and effective incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4392–4399, 2019.

[Liu *et al.*, 2019b] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, En Zhu, Tongliang Liu, Marius Kloft, Dinggang Shen, Jianping Yin, and Wen Gao. Multiple kernel k-means with incomplete kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(5):1191–1204, 2019.

[Liu *et al.*, 2024] Sihang Liu, Wenming Cao, Ruigang Fu, Kaixiang Yang, and Zhiwen Yu. Rpsc: robust pseudo-labeling for semantic clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 14008–14016, 2024.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

[Niu *et al.*, 2022] Chuang Niu, Hongming Shan, and Ge Wang. Spice: Semantic pseudo-labeling for image clustering. *IEEE Transactions on Image Processing*, 31:7264–7278, 2022.

[Powers, 2008] David M. W. Powers. Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation. 2008.

[Pu *et al.*, 2024] Jingyu Pu, Chenhang Cui, Xinyue Chen, Yazhou Ren, Xiaorong Pu, Zhifeng Hao, S Yu Philip, and Lifang He. Adaptive feature imputation with latent graph for deep incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 14633–14641, 2024.

[Sun *et al.*, 2024] Yuan Sun, Yang Qin, Yongxiang Li, Dezhong Peng, Xi Peng, and Peng Hu. Robust multi-view clustering with noisy correspondence. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Tang and Liu, 2022] Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International Conference on Machine Learning*, pages 21090–21110. PMLR, 2022.

[Wah *et al.*, 2011] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

[Wan *et al.*, 2021] Zhibin Wan, Changqing Zhang, Pengfei Zhu, and Qinghua Hu. Multi-view information-bottleneck representation learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 10085–10092, 2021.

[Wang *et al.*, 2022a] Qianqian Wang, Zhiqiang Tao, Wei Xia, Quanxue Gao, Xiaochun Cao, and Licheng Jiao. Adversarial multiview clustering networks with adaptive fusion. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[Wang *et al.*, 2022b] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9776–9785, 2022.

[Wen *et al.*, 2023] Yi Wen, Siwei Wang, Qing Liao, Weixuan Liang, Ke Liang, Xinhang Wan, and Xinwang Liu. Unpaired multi-view graph clustering with cross-view structure matching. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[Wolf *et al.*, 2011] Lior Wolf, Tal Hassner, and Itay Maoz. Face recognition in unconstrained videos with matched background similarity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 529–534. IEEE, 2011.

[Xia *et al.*, 2022] Jun Xia, Lirong Wu, Ge Wang, Jintao Chen, and Stan Z Li. Progcl: Rethinking hard negative mining in graph contrastive learning. *Proceedings of the International Conference on Machine Learning, PMLR*, pages 162:24332–24346, 2022.

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

[Xu *et al.*, 2023] Jie Xu, Chao Li, Liang Peng, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 32:1354–1366, 2023.

[Yan *et al.*, 2023] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19863–19872, 2023.

[Yang *et al.*, 2022] Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1055–1069, 2022.

[Yuan *et al.*, 2024] Honglin Yuan, Shiyun Lai, Xingfeng Li, Jian Dai, Yuan Sun, and Zhenwen Ren. Robust prototype completion for incomplete multi-view clustering. In *Proceedings of the ACM International Conference on Multimedia*, pages 10402–10411, 2024.

[Zhang *et al.*, 2020] Changqing Zhang, Yajie Cui, Zongbo Han, Joey Tianyi Zhou, Huazhu Fu, and Qinghua Hu. Deep partial multi-view learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2402–2415, 2020.

[Zhu *et al.*, 2024] Yanran Zhu, Xiao He, Chang Tang, Xinwang Liu, Yuanyuan Liu, and Kunlun He. Multi-view adaptive fusion network for spatially resolved transcriptomics data clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2024.