Sample Weighted Incomplete Multimodal Clustering Based on Graph Coarsening Label Extraction

Zhenjiao Liu[1], Xue Xiao[1], Yao Chen[1], Jiao Xue[1], Shubin Ma[2], Liang Zhao[2] *

[1]Inspur Cloud Information Technology Co., Ltd., Jinan, Shandong, China
[2] Dalian University of Technology, Dalian, Liaoning, China

Abstract

Multimodal data is typically collected through heterogeneous sensors and processing pipelines. However, due to variations in acquisition environments, device capabilities, and feature extraction methods, such data often suffers from incompleteness and inconsistent quality across modalities. To address these challenges, prior studies have explored modality selection and data completion strategies to improve information fusion. Nevertheless, these approaches face two main limitations: (1) they struggle to simultaneously ensure computational efficiency for large-scale graph data and maintain structural and semantic consistency across heterogeneous modality graphs; and (2) most of them operate at the modality level and fail to capture fine-grained, sample-specific quality variations. To overcome these issues, we propose a novel clustering framework, Sample Weighted Incomplete Multimodal Clustering Based on Graph Coarsening Label Extraction (IMC-GCSW). The proposed method introduces a graph coarsening-based label extraction strategy. It significantly reduces the computational cost of multimodal graph processing, while preserving key node information and local topological structures. Furthermore, a quality-aware sample weighting strategy is designed to enable fine-grained modeling of modality-specific data quality, allowing the model to dynamically suppress the influence of low-quality modalities on individual samples. Experiments on both general-purpose datasets and the Fructus Aurantii Disease and Pest Datasets demonstrate that the proposed method exhibits superior performance and strong adaptability in handling multimodal data with incompleteness and quality inconsistency.

Extended version —
https://github.com/Autism-mm/IMC-GCSW-.git

Introduction

In real-world conditions, multimodal data is always from different collecting ways and transportation, and therefore multimodal clustering (Zhao et al. 2025c,a, 2024; Ma et al. 2025) methods can provide more reliable clustering results than the single-modality clustering methods by integrating information from different modalities. However, in real-world data processing practice, the multimodal data is usually incomplete (Wang et al. 2024b; Xing et al. 2024; Chen et al. 2022), which is caused by various reasons such as the difficulty of collection, the errors in transportation and the device fault.

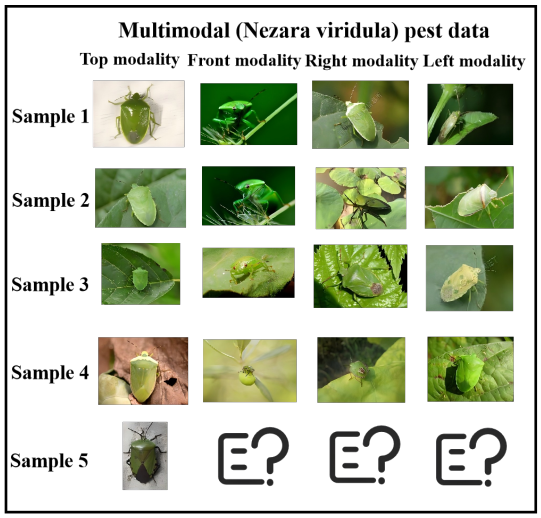

Figure 1: Example graph of incomplete data.

Figure 1 provides a toy example illustrating the clustering of incomplete multimodal data (Wang et al. 2022; Zhao et al. 2025b; Liu et al. 2023; Zhao et al. 2020) for Nezara viridula pest identification. The data is collected from four camera perspectives: top modality, front modality, right modality, and left modality. During the data collection process, each sample may lack information from one or more modalities due to various reasons, as indicated by the question mark (?) in the figure.

To address the issue of incomplete multimodal data caused by sensor failures, missing entries, or insufficient data collection, researchers have developed various methods tailored to real-world scenarios. Existing incomplete multimodal clustering approaches can be roughly categorized into four main types: **Modality-ignorant methods** (Chao, Jiang, and Chu 2024; Liu et al. 2020; Xu et al. 2024), which discard missing modalities and perform clustering based solely on the available modalities. For example, Hu et al. (Hu and Chen 2019) proposed OPIMC, which employed weighted matrix factorization and a one-pass learning strategy to effi-

*Corresponding author. Email: liangzhao@dlut.edu.cn

ciently cluster large-scale incomplete multimodal data without requiring modality imputation. **Modality-completion methods** (Wang et al. 2018; Lin et al. 2021; Pu et al. 2024), which first impute missing modalities—using statistical inference, pattern recognition, or machine learning models such as K-Means and autoencoders—and then conduct clustering. For example, Liu et al. (Liu et al. 2020) proposed EE-IMVC, which constructed a consensus clustering matrix to guide modality completion, avoiding costly kernel recovery. By integrating low-rank regularization, the method achieved efficient and effective clustering on large-scale incomplete multimodal data. **Graph-based methods** (Li et al. 2022a; Wen et al. 2020; Dong et al. 2024), which model the relationships between modalities and samples through graph structures, employing techniques like graph partitioning or regularization to handle missing modalities. Chen et al. (Chen et al. 2022) proposed ASR, which constructed augmented sparse representations to jointly model inter-sample and inter-modality relationships, enabling robust clustering without modality imputation. Zhao et al. (Zhao et al. 2023) proposed an unrestricted anchor graph-based GCN model, which captured global structure via graph convolution and effectively clustered highly incomplete multimodal data with enhanced robustness. **Deep learning-based methods** (Wei et al. 2020; Xu et al. 2022; Dai et al. 2025), which utilize deep architectures such as autoencoders and generative models to learn supplementary information or directly perform clustering. Lin et al. (Lin et al. 2022) proposed the Dual Contrastive Prediction (DCP) model, which uses a dual contrastive strategy to enhance both cross-modality consistency and intra-modality discrimination. Without explicitly imputing missing modalities, DCP can learn robust representations under complex missing patterns, improving clustering performance.

Although these methods alleviate the clustering challenges caused by incomplete multimodal data from various perspectives, they still suffer from two major limitations: (1) most approaches operate at the modality level, making it difficult to capture fine-grained (Meng et al. 2022; Benny and Wolf 2020; Yu et al. 2023a; Peng et al. 2024), sample-specific quality variations; and (2) they struggle to simultaneously achieve computational efficiency on large-scale graph data while preserving the structural and semantic consistency across heterogeneous modality graphs (Zhang et al. 2019; Wang et al. 2019; Hu et al. 2020; Zhao et al. 2021).

To capture sample-specific variations while balancing semantic consistency and computational efficiency, we propose a Sample-Weighted Incomplete Multimodal Clustering method based on Graph Coarsening Label Extraction (IMC-GCSW). Specifically, we design a graph coarsening-based label extraction strategy that leverages a clustering separation module to convert inter-cluster similarity into average similarity. A cross-modality voting mechanism is then employed to identify connected components, enabling consistent label association across modalities. This approach not only preserves critical node information in the original graph structure but also significantly improves computational efficiency in large-scale scenarios. Moreover, the proposed quality-aware sample weighting strategy adaptively assigns weights based on the quality of each sample, which helps reduce the influence of low-quality samples on clustering. Experimental results show that our method performs effectively in handling incomplete multimodal data.

- We design a cross-modality graph coarsening label association extraction strategy, which not only significantly improves computational efficiency but also preserves critical node information in the original graph structure. By minimizing supernode similarity, the method derives a unified clustering indicator matrix across modalities, effectively eliminating the need for time-consuming post-clustering procedures adopted in previous approaches.
- We design a sample-level weight learning mechanism that identifies quality differences among samples and assigns adaptive weights accordingly. This strategy effectively reduces the impact of low-quality data and mitigates clustering deviations, thereby enhancing robustness and reliability.
- We propose an efficiently optimizable objective function. Compared with five baseline methods, our approach achieves higher clustering efficiency on five datasets. Its effectiveness is further validated on the Fructus Aurantii Disease and Pest Datasets.

The Proposed Method

Suppose that an incomplete multimodal dataset is denoted as $\mathbf{X} = \{\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(v)}\}_{v=1}^{V} \in \mathcal{R}^{n \times d^{(v)}}$, $x_{i,j}^{(v)}$ represents the $\langle i, j \rangle$th element of the vth modality. As for dataset $\mathbf{A}$, $|\mathbf{A}|$ represents cardinal number. The method first pads original values in incomplete samples, using $\mathcal{W}^{(v)} = (\mathbf{V}^{(')}, \mathbf{E}^{(v)}, \mathbf{W}^{(v)})$ to represent the undirected weighted graph of the vth sample, and $\mathbf{V}^{(')}$ is the set of nodes, $\mathbf{E}^{(v)}$ is the set of edge, $\mathbf{W}^{(v)} \in \mathcal{R}^{n \times n}$ is the matrix of similarity, $\mathbf{D}^{(v)} \in R^{n \times n}$ is the degree matrix of $\mathbf{W}^{(v)}$, n is the amount of samples, and $d_{i,i}^{(v)} = \sum_{j=1}^{n} w_{ij}^{(v)}$, $\mathbf{L}^{(v)} = \mathbf{D}^{(v)} - \mathbf{W}^{(v)}$ is the Laplacian matrix of $\mathbf{W}^{(v)}$. Besides, $\delta(c, n) \in \mathcal{R}^{n}$ represents a vector, where the cth element is one and the other elements are zeros. Let $x \in \mathcal{Z}_{[1,n]}$ denotes an integer satisfying $1 < x < n$, where $x \in \mathcal{Z}_{[1,n]}$ represents the set of integers within the interval $(1, n)$. As for scala x and scala y, $x \propto y$ represents that x and y are positive correlation. For two scalar variables x and y, the notation $x \propto y$ indicates a positive correlation between them. Figure 2 illustrates the overall architecture of the proposed IMC-GCSW model. In this paper, incompleteness refers to the absence of entire samples.

Graph Coarsening-based Label Extraction Strategy

To effectively reduce the computational overhead of processing incomplete multimodal data and eliminate irrelevant redundancy, we design a graph coarsening-based label extraction strategy, which consists of a supernode similarity minimization module and a cross-modality nearest-neighbor voting module.

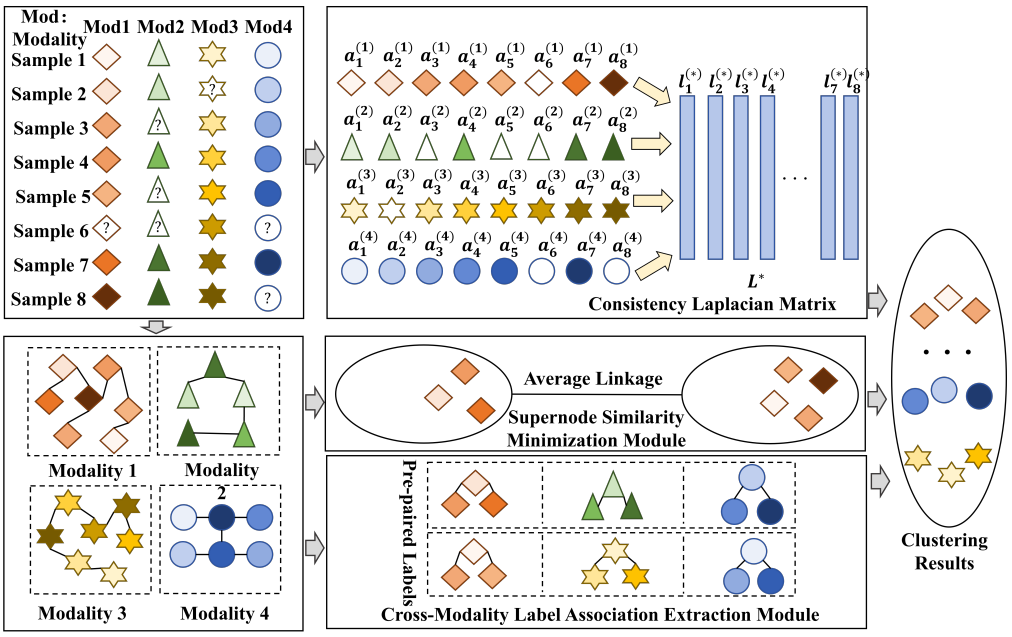

Figure 2: The flowchart of the IMC-GCSW algorithm. Different shapes represent different modalities.

Supernode Similarity Minimization Module The supernode similarity minimization module adopts a hierarchical clustering approach, which constructs a multi-level structure through iterative merging or splitting operations. This method typically treats each data point as an initial cluster and gradually merges similar or nearby clusters to form larger ones, thereby uncovering the underlying organizational patterns of the data.

Let $\mathbf{A}_k$ denote the set of all samples in the kth cluster, and $\bar{\mathbf{A}}_k$ represent the set of all samples excluding those in $\mathbf{A}_k$. The similarity between $\mathbf{A_k}$ and $\bar{\mathbf{A}}_k$ is defined by the following equation,

$$\begin{aligned} &\min_{\{\mathbf{A}_k\}_c} \sum_{k=1}^{c} \frac{\sum_{i\in\mathbf{A}_k}\sum_{j\in\bar{\mathbf{A}}_k}(\mathbf{w}_{i,j}^{(v)}+\mathbf{w}_{j,i}^{(v)})}{2|\mathbf{A}_k|\cdot|\bar{\mathbf{A}}_k|}, \\ &\text{s.t.} \quad \bigcup_{k=1}^{c}\mathbf{A}_k=\mathbf{A}, \forall r,d\in\mathcal{Z}_{[1,c]}, \mathbf{A}_r\cap\mathbf{A}_d=\emptyset. \end{aligned} \tag{1}$$

To better simplify Eq. 1, an indicator matrix $\mathbf{Y} \in \{0,1\}_{n\times c}$ with the constraint $\mathbf{Y1}=\mathbf{1}$ is introduced to replace the set operator $\{\mathbf{A}_k\}_{k=1}^{c}$. Considering the formulation under the multimodal setting, Eq. 1 can be rewritten accordingly

$$\min_{\mathbf{Y}\in\{0,1\}^{n\times c},\ \mathbf{Y1}=\mathbf{1}} \sum_{k=1}^{c} \frac{\left\|\mathbf{L}^{(v)}\mathbf{y}_k\right\|_1}{2\left\|\mathbf{y}_k\right\|_0\left(n-\left\|\mathbf{y}_k\right\|_0\right)}. \tag{2}$$

In multimodal learning, different modalities are often derived from distinct data sources, and each Laplacian matrix characterizes the internal structure of its modality. To unify structural information across modalities, a consensus Laplacian matrix $\mathbf{L}^*$ is introduced. This facilitates robust and efficient learning. Accordingly, Eq. 2 can be reformulated as

$$\min_{\mathbf{Y}\in\{0,1\}^{n\times c},\ \mathbf{Y1}=\mathbf{1}} \sum_{k=1}^{c} \frac{\left\|\mathbf{L}^{*}\mathbf{y}_k\right\|_1}{2\left\|\mathbf{y}_k\right\|_0\left(n-\left\|\mathbf{y}_k\right\|_0\right)}. \tag{3}$$

Cross-modality Nearest-neighbor Voting Module To better capture the consistency across modalities, we first define a coefficient matrix $\mathbf{E}^{(v)} \in \mathcal{R}^{n\times n}$, which characterizes the nearest-neighbor relationships in the v-th modality, as follows,

$$e_{i,j}^{(v)} = \begin{cases} 1, & \text{if } i=\bar{\mathbf{w}}_j^{(v)} \text{ or } j=\bar{\mathbf{w}}_i^{(v)} \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

Here, $e_{i,j}^{(v)}$ denotes the (i,j)-th element of $\mathbf{E}^{(v)}$, where $i=\bar{\mathbf{w}}_j^{(v)}$ indicates that node i is the nearest neighbor of node j in $\mathbf{W}^{(v)}$.To address the inconsistency of neighborhood structures for each sample across modalities, a cross-modality voting module is proposed to unify nearest-neighbor selection, as follows,

$$b_{i,j} = \begin{cases} 1, & \text{if } \sum_{v=1}^{V} e_{i,j}^{(v)} > \left\lfloor \frac{V}{2} \right\rfloor \\ 0, & \text{otherwise} \end{cases}. \tag{5}$$

If samples i and j appear as mutual nearest neighbors in more than half of the modalities ($e_{i,j}^{(s)}=1$), they are regarded as final nearest neighbors, and we set $b_{i,j}=1$. An adjacency matrix $\mathbf{B}$ is then constructed to define the graph $\mathcal{B}$, upon which the Tarjan algorithm is applied to identify connected components (Tarjan 1972). Based on this, a pairwise label matrix $\mathbf{T}$ is generated,

$$t_{i,j} = \begin{cases} 1, & \text{if } (i,j) \in \text{same component in } \mathcal{B} \\ 0, & \text{otherwise} \end{cases}. \quad (6)$$

Here t_{ij} is the (i,j)-th element of $\mathbf{T}$, which aggregates consistent nearest-neighbor information from the cross-modal contexts. In this way, the key local features of each modality can be preserved. Additionally, through $\mathbf{T}$, a pairwise label constraint is added to $\mathbf{Y}$, and the improvement is as follows,

$$\begin{aligned} &\min_{\mathbf{Y}} \sum_{k=1}^{c} \frac{\|\mathbf{L}^* \mathbf{y}_k\|_1}{2\|\mathbf{y}_k\|_0 (n - \|\mathbf{y}_k\|_0)} \\ &\text{s.t.} \quad \mathbf{Y} \in \{0,1\}_{n \times c}, \quad \mathbf{Y1} = \mathbf{1}, \quad \mathbf{Y}\mathbf{Y}^T \circ \mathbf{T} = \mathbf{T} \end{aligned} \quad (7)$$

Here, $\mathbf{Y}\mathbf{Y}^T \circ \mathbf{T} = \mathbf{T}$ ensures that $t_{i,j} = 1 \Rightarrow \mathbf{y}^i = \mathbf{y}^j$. Meanwhile, $\mathbf{Y}\mathbf{Y}^T \circ \mathbf{T} = \mathbf{T}$ also effectively aggregates consistent information across modalities, further enabling the fusion of multimodal structural information.

Quality-Aware Sample Weighting Strategy

To better extract features across modalities, a sample-level dynamic weighting strategy is introduced, where each sample is assigned a weight (Yu et al. 2023b) according to its quality. This enables a more fine-grained and accurate consensus representation. The weighting function is defined as follows,

$$\begin{aligned} &\min_{w_i^*, l_i^*, \mathbf{F}} \sum_{v=1}^{V} \left\{ \sum_{i=1}^{n} a_i^{(v)} \left\| \mathbf{w}_i^* - \mathbf{w}_i^{(v)} \right\|_F^2 \right\} + \lambda_1 \operatorname{Tr}\left(\mathbf{F}^T \mathbf{L}^* \mathbf{F}\right) \\ &\text{s.t.} \quad \forall v \in \mathcal{Z}_{[1,V]}, a_i^{(v)} \geq 0, \sum_{v=1}^{V} a_i^{(v)} = 1, \mathbf{F}^T \mathbf{F} = \mathbf{I}. \end{aligned} \quad (8)$$

Here, $a_i^{(v)}$ denotes the weight of the *i*-th sample in the *v*-th modality, $\mathbf{L}^*$ is the consensus Laplacian matrix, $\mathbf{I}$ is the identity matrix and $\mathbf{W}^*$ is the consensus similarity matrix. The second term, $Tr\left(\mathbf{F}^T \mathbf{L}^* \mathbf{F}\right)$, serves as a Laplacian regularization term that enforces the orthogonality of $\mathbf{F}$, thereby encouraging similar samples to be clustered together.

Compared to conventional *k*-means, hierarchical clustering demonstrates superior performance in our setting. Its ability to capture nested structures further enables a more fine-grained understanding of the clustering relations. The objective function is formulated as follows,

$$\begin{aligned} &\min_{\mathbf{Y},\, a_i^{(v)},\, \mathbf{L}^*,\, \mathbf{W}^*,\, \mathbf{F}} \sum_{v=1}^{V} \left\{ \sum_{k=1}^{c} \frac{\|\mathbf{L}^* \mathbf{y}_k\|_1}{2\|\mathbf{y}_k\|_0 (n - \|\mathbf{y}_k\|_0)} \right. \\ &\qquad \left. + \sum_{i=1}^{n} a_i^{(v)} \left\| \mathbf{w}_i^* - \mathbf{w}_i^{(v)} \right\|_F^2 \right\} + \lambda_1 Tr\left(\mathbf{F}^T \mathbf{L}^* \mathbf{F}\right) \\ &\text{s.t. } \mathbf{Y} \in \{0,1\}_{n \times c}, \quad \mathbf{Y1} = \mathbf{1}, \quad \forall v \in \mathcal{Z}_{[1,V]}, \\ &\quad a_i^{(v)} \geq 0, \quad \sum_{v=1}^{V} a_i^{(v)} = 1, \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}. \end{aligned} \quad (9)$$

Optimization

Due to the presence of multiple unknown variables, Eq.9 does not admit a closed-form solution. To address this, an iterative optimization strategy is employed, and the original model is reformulated as the following equivalent optimization problem,

$$\begin{aligned} &\min_{\mathbf{Y},\, a_i^{(v)},\, \mathbf{W}^*,\, \mathbf{F}} \sum_{v=1}^{V} \left\{ \sum_{k=1}^{c} \frac{\|(\operatorname{diag}(\mathbf{W}^*\mathbf{1}) - \mathbf{W}^*)\, \mathbf{y}_k\|_1}{2\|\mathbf{y}_k\|_0 (n - \|\mathbf{y}_k\|_0)} \right. \\ &\qquad \left. + \sum_{i=1}^{n} a_i^{(v)} \left\| \mathbf{w}_i^* - \mathbf{w}_i^{(v)} \right\|^2 \right\} + \frac{\lambda_1}{2} \sum_{i,j=1}^{n} \|\mathbf{f}_i - \mathbf{f}_j\|^2 w_{ij}^* \\ &\text{s.t.} \quad \mathbf{Y} \in \{0,1\}^{n \times c}, \quad \mathbf{Y1} = \mathbf{1}, \quad \mathbf{F}^\top \mathbf{F} = \mathbf{I}, \\ &\quad a_i^{(v)} \geq 0, \quad \sum_{v=1}^{V} a_i^{(v)} = 1. \end{aligned} \quad (10)$$

To solve the optimization problem in Eq.10, the Alternating Direction Method of Multipliers (ADMM) is adopted to obtain a locally optimal solution. To enable separability, several auxiliary variables are introduced, reformulating Eq.10 as follows,

$$\begin{aligned} &\min_{\mathbf{Y},\mathbf{J}^*,\mathbf{O}^*,\mathbf{W}^*,\mathbf{G}^*,\mathbf{F},a_i^{(v)}} \sum_{v=1}^{V} \left\{ \sum_{k=1}^{c} \frac{\left\|\bar{\mathbf{J}}\mathbf{y}_k\right\|_1}{2\|\mathbf{y}_k\|_0 (n - \|\mathbf{y}_k\|_0)} \right. \\ &\qquad \left. + \sum_{i=1}^{n} a_i^{(v)} \left\| \mathbf{o}_i^* - \mathbf{w}_i^{(v)} \right\|^2 \right\} + \frac{\lambda_1}{2} \sum_{i,j=1}^{n} \|\mathbf{f}_i - \mathbf{f}_j\|^2 g_{ij}^* \\ &\text{s.t.} \quad \mathbf{J}^* = \mathbf{W}^*, \quad \mathbf{O}^* = \mathbf{W}^*, \quad \mathbf{G}^* = \mathbf{W}^*, \\ &\quad \mathbf{Y} \in \{0,1\}^{n \times c}, \quad \mathbf{Y1} = \mathbf{1}, \quad \mathbf{F}^\top \mathbf{F} = \mathbf{I}, \\ &\quad \forall v \in \mathcal{Z}_{[1,V]}, \quad a_i^{(v)} \geq 0, \quad \sum_{v=1}^{V} a_i^{(v)} = 1. \end{aligned} \quad (11)$$

Here, $\bar{\mathbf{J}} = \operatorname{diag}(\mathbf{J}^*\mathbf{1}) - \mathbf{J}^*$. The augmented Lagrangian function corresponding to Eq.11 is formulated as follows,

$$\begin{aligned} L = &\sum_{v=1}^{V} \left\{ \sum_{k=1}^{c} \frac{\|\bar{\mathbf{J}}\mathbf{y}_k\|_1}{2\|\mathbf{y}_k\|_0 (n - \|\mathbf{y}_k\|_0)} \right. \\ &\left. + \sum_{i=1}^{n} a_i^{(v)} \|\mathbf{o}_i^* - \mathbf{w}_i^{(v)}\|^2 \right\} + \frac{\lambda_1}{2} \sum_{i,j=1}^{n} \|\mathbf{f}_i - \mathbf{f}_j\|^2 g_{ij}^* \\ &+ \frac{\mu}{2} \left(\left\| \mathbf{W}^* - \mathbf{J}^* + \frac{\mathbf{C}_1}{\mu} \right\|_F^2 + \left\| \mathbf{W}^* - \mathbf{O}^* + \frac{\mathbf{C}_2}{\mu} \right\|_F^2 \right. \\ &\left. + \left\| \mathbf{W}^* - \mathbf{G}^* + \frac{\mathbf{C}_3}{\mu} \right\|_F^2 \right). \end{aligned} \quad (12)$$

To solve the objective function in Eq. 12, an alternating optimization algorithm is employed. The objective involves twelve unknown variables: $\mathbf{W}^*$, $a_i^{(v)}$, $\mathbf{L}^*$, $\mathbf{J}^*$, $\mathbf{O}^*$, $\mathbf{G}^*$, $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$, μ, $\mathbf{F}$ and $\mathbf{Y}$. At each iteration, one variable is updated while fixing the others, and the process is repeated until convergence. Due to the complexity of the optimization, detailed update steps are provided in the appendix.

Dataset	**Rate**	HCP_IMSC	DAIMC	UEAF	IMG	BSV	Ours	Mask-IMVC	IMC-MCL	Ours
SensIT300	0.1	59.66	56.55	61.00	57.51	50.33	**61.33**	**65.68**	60.16	61.33
SensIT300	0.5	36.33	37.66	46.00	38.70	39.33	**47.67**	37.58	**54.00**	47.67
SensIT300	0.9	35.34	35.44	36.33	36.00	36.53	**36.83**	**42.99**	41.93	36.83
Statlog	0.1	55.75	47.79	38.61	**56.52**	46.58	47.63	**56.46**	46.75	47.63
Statlog	0.5	33.67	21.39	24.24	28.53	15.36	**37.53**	29.91	**42.76**	37.53
Statlog	0.9	16.13	15.91	15.57	15.93	14.96	**16.19**	33.68	**38.73**	16.19
ORL	0.1	82.25	55.66	64.50	34.55	29.37	**83.07**	57.78	33.95	**83.07**
ORL	0.5	31.50	16.41	31.75	20.07	25.75	**32.00**	**37.80**	24.98	32.00
ORL	0.9	16.50	16.41	13.50	14.11	9.61	**16.54**	**39.61**	24.00	16.54
Yale	0.1	56.06	55.35	57.47	47.30	28.36	**57.58**	47.82	52.24	**57.58**
Yale	0.5	33.27	20.40	33.93	25.73	29.33	**37.33**	34.81	**45.94**	37.33
Yale	0.9	20.60	16.16	15.15	14.84	20.60	**21.21**	**33.55**	27.64	21.21

Table 1: Mean ACCs (%) of different methods on the SensIT300, Statlog, ORL and Yale datasets.

Dataset	**Rate**	HCP_IMSC	DAIMC	UEAF	IMG	BSV	Ours	Mask-IMVC	IMC-MCL	Ours
SensIT300	0.1	44.13	45.19	47.02	45.30	47.39	**50.82**	50.68	18.40	**50.82**
SensIT300	0.5	38.17	42.64	44.15	42.17	38.08	**44.41**	36.06	17.87	**44.41**
SensIT300	0.9	32.12	31.44	33.47	30.20	33.53	**33.84**	**38.36**	16.61	33.84
Statlog	0.1	44.74	36.06	32.45	**47.10**	38.96	36.37	**58.38**	8.52	36.37
Statlog	0.5	24.07	18.74	19.56	24.72	24.71	**24.86**	**32.45**	7.71	24.86
Statlog	0.9	20.02	23.48	15.34	22.18	24.17	**25.35**	**30.79**	9.24	25.35
ORL	0.1	78.05	41.72	47.13	23.09	11.62	**79.86**	55.56	1.58	**79.86**
ORL	0.5	12.84	2.54	11.52	7.235	10.10	**14.34**	**33.32**	1.81	14.34
ORL	0.9	2.27	2.24	2.29	2.25	2.38	**2.49**	**34.53**	1.29	2.49
Yale	0.1	46.68	39.10	46.40	25.22	17.77	**47.94**	45.75	11.47	**47.94**
Yale	0.5	31.27	7.85	12.88	12.53	13.09	**32.80**	30.66	8.39	**32.80**
Yale	0.9	7.68	10.82	11.27	10.24	11.32	**11.93**	**28.82**	3.33	11.93

Table 2: Mean F1s (%) of different methods on the SensIT300, Statlog, ORL and Yale datasets.

Experiment

Dataset Description

Experiments are conducted on four public datasets (ORL (Wang et al. 2016), Yale, SensIT300, and Statlog (Zhou et al. 2015)) and two private datasets, namely Fructus Aurantii Disease Datasets (FADD) and Fructus Aurantii Pest Datasets (FAPD). Details of all datasets are summarized in Table 3.

Datasets	ORL	Yale	SensIT300	Statlog	FADD	FAPD
Dim 1	4096	4096	50	9	1000	1000
Dim 2	3304	4096	50	10	256	256
Samples	400	165	300	2310	1036	1532
Classes	40	15	3	7	7	10

Table 3: Introduction to datasets.

In the private datasets, the text modality has a dimensionality of 256, while the image (Li et al. 2025a, 2024; Zou et al. 2025) modality is 1000-dimensional. The Fructus Aurantii Disease Dataset includes 7 disease categories, while the Pest Dataset contains 10 pest categories.

To validate the effectiveness of the proposed method, IMC-GCSW is compared against seven representative incomplete multimodal fusion methods, including DAIMC (Hu and Chen 2018), UEAF (Wen et al. 2019), BSV (Xu et al. 2020), IMG (Zhao et al. 2016), HCP-IMSC (Li et al. 2022b), Mask-IMVC (Li et al. 2025b) and IMC-MCL (Yin et al. 2025).

Experimental Setting

In the experiments, missing rates of 0.1, 0.5, and 0.9 are applied to each dataset to simulate varying levels of incompleteness in multimodal data. All experiments are conducted on a Windows platform equipped with an Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz and 16GB of RAM, using MATLAB 2021b for implementation.

In the experiments, two commonly used metrics are employed to evaluate clustering performance: Accuracy (ACC) (Yang and Yi 2008) and F-score.

Experimental Results and Analysis

In this experiment, the best results for each metric are shown in bold, and the second-best are underlined. Tables 1 and 2 show the ACC and F-score of IMC-GCSW and baseline methods on four datasets. To test model robustness, experiments were done under low (0.1), medium (0.5), and high (0.9) missing rates. Notably, Mask-IMVC and IMC-MCL are deep learning-based methods designed for incomplete multimodal clustering, which are different from traditional approaches. The details are as follows:

Method	ACC (%)	F-score (%)	Time (s)
IMC-GCSW	**57.78**	**56.52**	**160.57**
HCP-IMSC	55.29	54.41	236.79

Table 4: Analysis of two hypergraph-based multimodal clustering methods on the Caltech101-7 dataset with 90% incomplete samples in each modality.

- IMC-GCSW achieves the best overall performance among all evaluated machine learning methods. As shown in Tables 1 and 2, IMC-GCSW consistently outperforms other incomplete multimodal clustering algorithms in terms of both ACC and F-score. On the ORL and Yale datasets, it achieves the highest accuracy and F-score across all levels of incompleteness.
- Graph-based methods generally outperform traditional approaches. In particular, the unimodal method BSV often performs worse than multimodal methods, indicating that leveraging cross-modal complementarity enhances clustering accuracy and generalization. For example, graph-based models such as HCP-IMSC and the proposed IMC-GCSW show superior performance compared to matrix factorization-based methods.
- Mask-IMVC and IMC-MCL are deep learning-based methods designed for incomplete multimodal clustering. Compared to these approaches, IMC-GCSW achieves either the best or second-best performance under low missing rates. Notably, on the ORL dataset, it significantly outperforms Mask-IMVC and IMC-MCL in terms of ACC and F-score, demonstrating its ability to effectively leverage complete modality information to enhance clustering on incomplete data.
- However, the performance of IMC-GCSW deteriorates at high missing rates, as the aggregation of supernodes to extract a consistent clustering indicator matrix heavily relies on complete modality features. In contrast, deep learning methods can discard missing modalities and perform unimodal clustering directly, resulting in stable performance at missing rates of 0.5 and 0.9. In some cases, these methods even outperform at a 0.9 missing rate compared to 0.5, which contradicts the fundamental principle of multimodal clustering—leveraging complementary information across modalities.

Effectiveness Evaluation and Convergence Analysis

Table 4 compares the performance of two hypergraph-based incomplete multimodal clustering methods. It can be observed that the proposed method exhibits clear advantages in terms of clustering metrics, runtime efficiency, and model complexity. To further verify its generalization and robustness, we conduct comparative experiments on alternative datasets (Caltech101-7), ensuring a more comprehensive and fair evaluation.

IMC-GCSW is a typical optimization algorithm based on the Augmented Lagrangian Method (ALM). In this section, we experimentally validate its convergence behavior. Figure 3 illustrates the convergence curves of IMC-GCSW on four

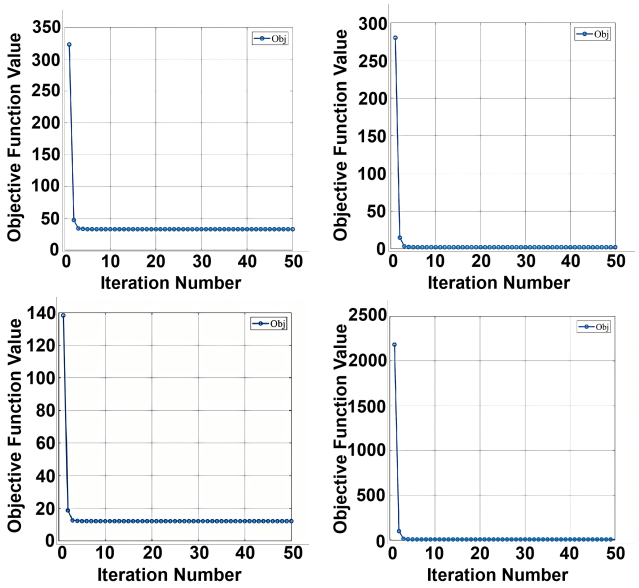

Figure 3: From left to right and top to bottom, the convergence curves correspond to the ORL, SensIT300, Yale, and Statlog datasets, respectively.

datasets with a missing rate of 0.5. As the number of iterations increases, the objective value shows a general downward trend, with a significant drop observed in the early stages. These results, shown in Figure 3, demonstrate that the proposed optimization strategy exhibits stable convergence and is capable of reaching a locally optimal solution within a reasonable number of iterations.

Application Analysis

To assess the practical effectiveness of the proposed model and its performance in real-world scenarios, we conducted experiments on the FADD and FAPD datasets. Tables 6 and 7 present the sample composition of these two datasets, respectively. Two representative baselines, UGCF (Wang et al. 2024a) and HCP-IMSC, are selected for comparison. UGCF restores missing values by leveraging intra-modal correlations while preserving local structures, and adopts a similarity graph topology similar to that in IMC-GCSW. HCP-IMSC integrates multimodal affinity matrices via a self-weighting mechanism to obtain a unified representation, effectively mitigating modality quality inconsistencies.

- Table 5 reports the clustering performance of IMC-GCSW and the baseline methods on the FADD and FAPD incomplete multimodal datasets. As shown in the results, IMC-GCSW consistently achieves the best performance across all evaluation metrics (NMI, ACC, and F-score), demonstrating its robustness and effectiveness in practical multimodal recognition scenarios.
- The superior performance achieved in our experiments can be attributed to two main factors. First, in real-world scenarios, the missing rate is typically not excessively high, as severe incompleteness is often considered a sensor malfunction that can be addressed through mainte-

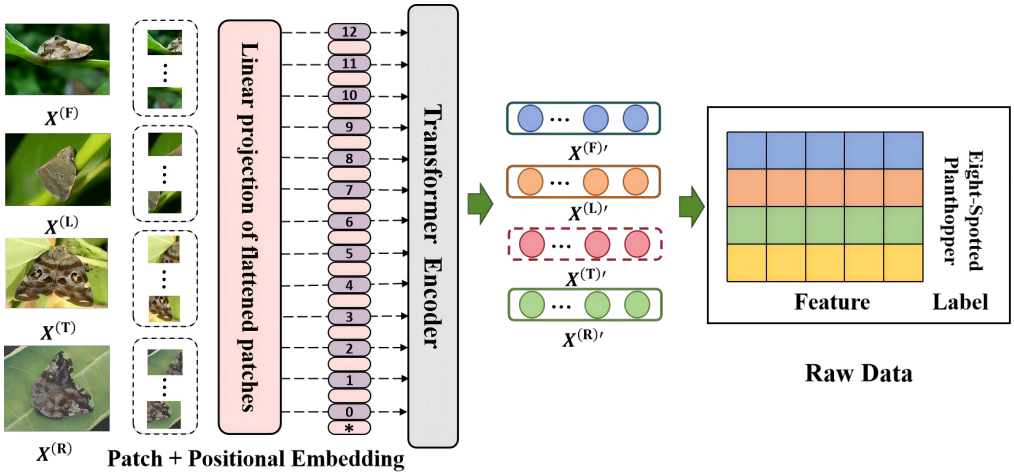

Figure 4: Feature extraction network architecture on the Fructus Aurantii disease and pest datasets, where the input modalities exhibit naturally missing patterns. (e.g., Eight-Spotted Planthopper).

Method	Citrus Disease Dataset			Citrus Pest Dataset		
	NMI	ACC	F-score	NMI	ACC	F-score
UGCF	57.18	52.82	56.72	63.91	59.71	54.72
HCP-IMSC	54.95	50.47	54.83	66.74	73.89	66.97
IMC-GCSW	**66.87**	**62.32**	**67.24**	**75.84**	**82.74**	**76.61**

Table 5: Clustering performance of different algorithms on the incomplete Fructus Aurantii Disease and Pest Datasets.

Disease Name	**Class-Id**	**Available Images**
Scab	0	95
Canker	1	121
Fruit Cracking	2	169
Green Mold	3	225
Black Stem-End Rot	4	104
Russeting	5	175
Sooty Mold	6	147

Table 6: Disease category and the number of available images per class.

Pest Name	**Class-Id**	**Available Images**
Eight-Spotted Planthopper	0	64
White Planthopper	1	397
Flat Slug Moth	2	169
Lime Swallowtail	3	269
Green Stink Bug	4	157
Black Cicada	5	197
Green Chafer	6	169
Rough Shield Bug	7	57
Malignant Leaf Beetle	8	25
Great Mormon	9	28

Table 7: Pest categories and the number of available images per class.

nance. Second, IMC-GCSW exhibits strong adaptability and robustness to real-world data distributions.

Figure 4 illustrates the architecture of the feature extraction network. Each sample in the pest dataset contains four modality-specific input feature maps captured from different viewpoints: front (F), left (L), top (T), and right (R). These multimodal images are designed for specific pest types, with each modality providing distinct visual information that facilitates a more comprehensive representation of pest characteristics.

During preprocessing, each image is divided into 16×16 patches, where each patch is flattened into a one-dimensional vector. This transformation converts the original 2D image data into a format suitable for standard machine learning models. The resulting feature tokens are then processed and aggregated by a Transformer encoder, and subsequently used for label prediction.

Conclusion

This paper proposes a sample-weighted clustering method based on graph coarsening and label extraction to address quality inconsistency in incomplete multimodal data. The method assigns weights to individual samples and introduces a cross-modality nearest-neighbor voting mechanism to capture the data structure. Experiments on benchmark datasets and the incomplete Fructus Aurantii Disease Dataset and Fructus Aurantii Pest Dataset demonstrate the superiority of the proposed method in clustering performance and efficiency. Future work will focus on enhancing its robustness under high missing rates.

Acknowledgments

This work is supported by the National Natural Science Foundation of China(62572095), the Fundamental Research Funds for the Central Universities (DUTZD25216), the Science and Technology Project of Liaoning Province(2024JH2/102600027), the Science and Technology Project of Dalian City(2024JJ12GX025, 2023JJ12SN029 and 2023JJ11CG005) and the Shandong Province Technology Innovation Guidance Plan (Central Guidance for Local Scientific and Technological Development Project: Research and Industrialization of Artificial Intelligence Large Model Service Platform) (YDZX2024088).

References

Benny, Y.; and Wolf, L. 2020. Onegan: Simultaneous unsupervised learning of conditional image generation, foreground segmentation, and fine-grained clustering. In *European Conference on Computer Vision*, 514–530. Springer.

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

Chen, J.; Yang, S.; Peng, X.; Peng, D.; and Wang, Z. 2022. Augmented sparse representation for incomplete multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 35(3): 4058–4071.

Dai, Y.; Jin, J.; Dong, Z.; Wang, S.; Liu, X.; Zhu, E.; Yang, X.; Gan, X.; and Feng, Y. 2025. Imputation-free and Alignment-free: Incomplete Multi-view Clustering Driven by Consensus Semantic Learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 5071–5081.

Dong, Z.; Jin, J.; Xiao, Y.; Xiao, B.; Wang, S.; Liu, X.; and Zhu, E. 2024. Subgraph propagation and contrastive calibration for incomplete multiview data clustering. *IEEE Transactions on Neural Networks and Learning Systems*.

Hu, M.; and Chen, S. 2018. Doubly aligned incomplete multi-view clustering. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, 2262–2268. AAAI Press.

Hu, M.; and Chen, S. 2019. One-pass incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 3838–3845.

Hu, Z.; Dong, Y.; Wang, K.; and Sun, Y. 2020. Heterogeneous graph transformer. In *Proceedings of the web conference 2020*, 2704–2710.

Li, X.; Liu, J.; Chen, Z.; Zou, Y.; Ma, L.; Fan, X.; and Liu, R. 2024. Contourlet residual for prompt learning enhanced infrared image super-resolution. In *European Conference on Computer Vision*, 270–288. Springer.

Li, X.; Wang, Z.; Zou, Y.; Chen, Z.; Ma, J.; Jiang, Z.; Ma, L.; and Liu, J. 2025a. Difiisr: A diffusion model with gradient guidance for infrared image super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 7534–7544.

Li, X.-L.; Chen, M.-S.; Wang, C.-D.; and Lai, J.-H. 2022a. Refining graph structure for incomplete multi-view clustering. *IEEE transactions on neural networks and learning systems*, 35(2): 2300–2313.

Li, Z.; Shi, Y.; He, X.; and Tang, C. 2025b. Mask-informed Deep Contrastive Incomplete Multi-view Clustering. *IEEE Transactions on Circuits and Systems for Video Technology*.

Li, Z.; Tang, C.; Zheng, X.; Liu, X.; Zhang, W.; and Zhu, E. 2022b. High-order correlation preserved incomplete multi-view subspace clustering. *IEEE Transactions on Image Processing*, 31: 2067–2080.

Lin, Y.; Gou, Y.; Liu, X.; Bai, J.; Lv, J.; and Peng, X. 2022. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4447–4461.

Lin, Y.; Gou, Y.; Liu, Z.; Li, B.; Lv, J.; and Peng, X. 2021. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11174–11183.

Liu, C.; Wen, J.; Wu, Z.; Luo, X.; Huang, C.; and Xu, Y. 2023. Information recovery-driven deep incomplete multi-view clustering network. *IEEE Transactions on Neural Networks and Learning Systems*.

Liu, X.; Li, M.; Tang, C.; Xia, J.; Xiong, J.; Liu, L.; Kloft, M.; and Zhu, E. 2020. Efficient and effective regularized incomplete multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence*, 43(8): 2634–2646.

Ma, S.; Zhao, L.; Lu, M.; Guo, Y.; and Xu, B. 2025. Consistency-Aware Padding for Incomplete Multi-Modal Alignment Clustering Based on Self-Repellent Greedy Anchor Search. In Kwok, J., ed., *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, 5923–5931. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Meng, M.; Wu, Z.; Liang, T.; Yu, J.; and Wu, J. 2022. Exploring fine-grained cluster structure knowledge for unsupervised domain adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(8): 5481–5494.

Peng, C.; Kang, K.; Chen, Y.; Kang, Z.; Chen, C.; and Cheng, Q. 2024. Fine-grained essential tensor learning for robust multi-view spectral clustering. *IEEE Transactions on Image Processing*, 33: 3145–3160.

Pu, J.; Cui, C.; Chen, X.; Ren, Y.; Pu, X.; Hao, Z.; Yu, P. S.; and He, L. 2024. Adaptive feature imputation with latent graph for deep incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 14633–14641.

Tarjan, R. 1972. Depth-first search and linear graph algorithms. *SIAM journal on computing*, 1(2): 146–160.

Wang, H.; Wang, Q.; Miao, Q.; and Ma, X. 2024a. Joint learning of data recovering and graph contrastive denoising for incomplete multi-view clustering. *Information Fusion*, 104: 102155.

Wang, Q.; Ding, Z.; Tao, Z.; Gao, Q.; and Fu, Y. 2018. Partial multi-view clustering via consistent GAN. In *2018 IEEE International Conference on Data Mining (ICDM)*, 1290–1295. IEEE.

Wang, S.; Lu, J.; Gu, X.; Weyori, B. A.; and Yang, J.-y. 2016. Unsupervised discriminant canonical correlation analysis based on spectral clustering. *Neurocomputing*, 171: 425–433.

Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; and Yu, P. S. 2019. Heterogeneous graph attention network. In *The world wide web conference*, 2022–2032.

Wang, Y.; Chang, D.; Fu, Z.; Wen, J.; and Zhao, Y. 2022. Incomplete multiview clustering via cross-view relation transfer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1): 367–378.

Wang, Z.; Li, L.; Ning, X.; Tan, W.; Liu, Y.; and Song, H. 2024b. Incomplete multi-view clustering via structure exploration and missing-view inference. *Information Fusion*, 103: 102123.

Wei, S.; Wang, J.; Yu, G.; Domeniconi, C.; and Zhang, X. 2020. Deep incomplete multi-view multiple clusterings. In *2020 IEEE International Conference on Data Mining (ICDM)*, 651–660. IEEE.

Wen, J.; Yan, K.; Zhang, Z.; Xu, Y.; Wang, J.; Fei, L.; and Zhang, B. 2020. Adaptive graph completion based incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 23: 2493–2504.

Wen, J.; Zhang, Z.; Xu, Y.; Zhang, B.; Fei, L.; and Liu, H. 2019. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 5393–5400.

Xing, L.; Song, Y.; Chen, B.; Yu, C.; and Qin, J. 2024. Incomplete multi-view clustering via correntropy and complement consensus learning. *IEEE Transactions on Multimedia*, 26: 8063–8076.

Xu, G.; Wen, J.; Liu, C.; Hu, B.; Liu, Y.; Fei, L.; and Wang, W. 2024. Deep variational incomplete multi-view clustering: Exploring shared clustering structures. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 16147–16155.

Xu, J.; Li, C.; Ren, Y.; Peng, L.; Mo, Y.; Shi, X.; and Zhu, X. 2022. Deep incomplete multi-view clustering via mining cluster complementarity. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 8761–8769.

Xu, X.; Li, K.; Xu, C.; and He, S. 2020. Gdface: Gated deformation for multi-view face image synthesis. In *Proceedings of the AAAI Conference on artificial intelligence*, volume 34, 12532–12540.

Yang, C.; and Yi, Z. 2008. Document clustering using locality preserving indexing and support vector machines. *Soft Computing*, 12(7): 677–683.

Yin, J.; Wang, P.; Sun, S.; and Zheng, Z. 2025. Incomplete Multi-View Clustering via Multi-Level Contrastive Learning. *IEEE Transactions on Knowledge and Data Engineering*.

Yu, S. L.; Liu, Q.; Wang, F.; Yu, Y.; and Chen, E. 2023a. Federated news recommendation with fine-grained interpolation and dynamic clustering. In *Proceedings of the 32nd ACM international conference on information and knowledge management*, 3073–3082.

Yu, X.; Liu, H.; Lin, Y.; Liu, N.; and Sun, S. 2023b. Sample-level weights learning for multi-view clustering on spectral rotation. *Information Sciences*, 619: 38–51.

Zhang, C.; Song, D.; Huang, C.; Swami, A.; and Chawla, N. V. 2019. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 793–803.

Zhao, H.; Liu, H.; Fu, Y.; et al. 2016. Incomplete multimodal visual data grouping. In *IJCAI*, volume 16, 2392–2398.

Zhao, J.; Wang, X.; Shi, C.; Hu, B.; Song, G.; and Ye, Y. 2021. Heterogeneous graph structure learning for graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 4697–4705.

Zhao, L.; Huang, P.; Chen, T.; Fu, C.; Hu, Q.; and Zhang, Y. 2024. Multi-sentence complementarily generation for text-to-image synthesis. *IEEE Transactions on Multimedia*, 26: 8323–8332.

Zhao, L.; Wang, X.; Liu, Z.; Wang, Z.; and Chen, Z. 2025a. Learnable Graph Guided Deep Multi-view Representation Learning via Information Bottleneck. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(4): 3303–3314.

Zhao, L.; Wang, Z.; Wang, X.; Chen, Z.; and Xu, B. 2025b. Incomplete and Unpaired Multi-View Graph Clustering with Cross-View Feature Fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 22786–22794.

Zhao, L.; Wang, Z.; Yuan, Y.; and Ding, F. 2023. Unrestricted anchor graph based gcn for incomplete multi-view clustering. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Zhao, L.; Xie, Q.; Li, Z.; Wu, S.; and Yang, Y. 2025c. Dynamic Graph Guided Progressive Partial View-Aligned Clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36(5): 9370–9382.

Zhao, L.; Yang, T.; Zhang, J.; Chen, Z.; Yang, Y.; and Wang, Z. J. 2020. Co-learning non-negative correlated and uncorrelated features for multi-view data. *IEEE Transactions on Neural Networks and Learning Systems*, 32(4): 1486–1496.

Zhou, D.; He, J.; Candan, K. S.; and Davulcu, H. 2015. Muvir: Multi-view rare category detection. In *IJCAI*, 4098–4104.

Zou, Y.; Chen, Z.; Zhang, Z.; Li, X.; Ma, L.; Liu, J.; Wang, P.; and Zhang, Y. 2025. Contourlet refinement gate framework for thermal spectrum distribution regularized infrared image super-resolution. *International Journal of Computer Vision*.