Multi-View Clustering with Granularity-Aware Pseudo Supervision

Jie Yang[1], Cheng-You Lu[2], Zhongli Wang[3], Hsiang-Ting Chen[4], Guang-Kui Xu[5], Chenglong Zhang[6], Shuting Dong[7], Xinyan Liang[8], Bingbing Jiang[3*]

[1]The University of Sydney
[2]University of Technology Sydney
[3]Hangzhou Normal University
[4]The University of Adelaide
[5]Xi'an Jiaotong University
[6]Nanjing University
[7]Jinan University
[8]Shanxi University

jie.yang1@sydney.edu.au, cheng-you.lu@student.uts.edu.au, wangzhongli1@stu.hznu.edu.cn, tim.chen@adelaide.edu.au, guangkuixu@mail.xjtu.edu.cn, clzhang123@163.com, dongst@stu.jnu.edu.cn, liangxinyan48@163.com, jiangbb@hznu.edu.cn

Abstract

Modern multi-view clustering (MVC) is dominated by two paradigms: multi-view fusion and pseudo-label-guided learning. Pseudo-labeling methods can suffer from confirmation bias; their reliance on a fixed-granularity supervision from an initial clustering can cause learned embeddings to drift from the data's true structure and lose discriminative power. Conversely, fusion methods excel at integrating information but often struggle to robustly differentiate between high-quality and noisy views, which can obscure final cluster boundaries and degrade performance. To address these complementary challenges, we propose GAPS (Granularity-Aware Pseudo Supervision), a novel MVC framework. GAPS introduces a granularity-aware supervision mechanism that generates a full hierarchy of pseudo-labels, enabling the selection of a supervision level that best aligns with the data's intrinsic multi-scale structure. Furthermore, to ensure a high-quality supervisory signal, it incorporates a reliability-aware view selection strategy using a novel Separation-Compactness Index (SCI) to identify and leverage the most informative view for pseudo-label generation. This dual approach ensures the supervisory signal is both structurally adaptive and derived from the most reliable source, leading to highly effective final representations. Extensive experiments on synthetic and real-world datasets demonstrate the effectiveness and superiority of GAPS over other competitors.

Introduction

Multi-view clustering (MVC) leverages diverse data modalities, such as images, text, or sensor signals, to uncover latent group structures more effectively than single-view approaches (Jiang et al. 2025a; Wang et al. 2025d; Wen et al. 2023; Yang and Lin 2022). By integrating complementary information, MVC can reveal more meaningful patterns (Liang et al. 2025a; Zhang et al. 2023; Guo et al. 2024). However, a central challenge in MVC is learning cluster-discriminative representations from these heterogeneous sources without access to ground-truth labels (Chen et al. 2020; Wang et al. 2025b; Liu, Chen, and Yue 2025).

Many MVC methods that rely on unsupervised fusion or latent space alignment often produce representations with limited cluster separability because they lack a direct clustering objective (Liang et al. 2025b; Liu et al. 2024a; Zhang et al. 2024a). For example, some methods perform consensus graph fusion by learning view-specific spectral embedding and then enforcing a low-rank tensor constraint on their Gram matrices to find a shared structure (Li et al. 2021; Chen, Wang, and Lai 2023; Wang et al. 2025a; Wu et al. 2025). While effective at structural fusion, this process is not guided by an explicit cluster-label signal, making the learned graph vulnerable to noise present in the initial features (Liu et al. 2024b; Zhang et al. 2025b). Other methods adopt an ensemble paradigm, creating multiple base partitions from random view groups which are then fused in a multi-stage process (Huang, Wang, and Lai 2023; Gan et al. 2025; Wu et al. 2024). These strategies, however, lack a cluster-driven refinement loop where intermediate cluster assignments actively supervise and enhance the feature embedding (Jiang et al. 2025b; Wang et al. 2025e). Similarly, one-step frameworks that jointly optimize for diverse representations and final partitions (Wan et al. 2025) still separate the core tasks of feature learning and clustering within their objective, missing the guidance that a direct supervisory signal can provide.

To incorporate cluster-awareness, recent works explored pseudo-labeling, where preliminary cluster assignments guide the embedding process (Yang et al. 2025). However, a critical flaw pervades these methods: they employ a fixed pseudo-label granularity, typically matching the final number of clusters. For instance, some frameworks first learn a latent pseudo-label matrix of size $n \times K$ (where n and K are the number of samples and final clusters, respectively) and then use this fixed matrix to guide regression on the original views (Cai et al. 2024; Chen et al. 2025a; Sun et al. 2025).

*The corresponding author

More advanced deep models use such pseudo-labels for contrastive pair refinement, removing false negatives to improve feature alignment (Hu et al. 2025; Li et al. 2023). Even in this advanced case, the pseudo labels are generated for a predetermined number of clusters and remain fixed throughout training. Likewise, methods based on collective matrix factorization incorporate a pseudo-label constraint derived from an initial clustering of each view into a preset number of groups (Wang et al. 2022; Cui et al. 2025). Such a rigid "one-size-fits-all" assumption is fundamentally restrictive. Real-world data may possess a natural structure better captured by a coarser or finer granularity than the final partition; coarse labels can enforce global consistency, while fine-grained labels preserve subtle local distinctions. Relying on a single, fixed resolution thus inherently constrains model performance and generalization.

Another under-explored challenge is the principled selection of the most informative view to guide the clustering process. While many approaches learn or assign weights to quantify each view's contribution, these fusion strategies remain vulnerable to the influence of poor-quality or less informative views. For example, some methods assign explicit weights to each view based on their contribution to minimizing intra-cluster distances (Xu, Wang, and Lai 2016; Chen et al. 2025b). Others propose auto-weighted frameworks where weights are learned adaptively as part of a joint optimization (Nie et al. 2018; Wang et al. 2025c). Further work develops implicit weight learning, where a view's importance is inferred from its loss value via a re-weighted optimization scheme (Nie et al. 2023). In all these cases, the view contribution is inferred indirectly as a byproduct of the main learning objective. This passive approach may still assign non-negligible weight to noisy views, potentially corrupting the consensus representation. A more principled alternative is to directly evaluate the clustering quality of each view with an internal validity metric (Yang et al. 2017, 2021). This allows for the proactive identification of the view that best captures the intrinsic cluster structure, ensuring that subsequent learning is guided by the most reliable signal available(Sun et al. 2024).

In this work, we propose GAPS (Granularity-Aware Pseudo Supervision), a novel multi-view clustering framework that directly addresses these fundamental limitations. First, GAPS incorporates a reliability-aware view selection strategy: each view is clustered, and a novel internal clustering validity index (SCI) is introduced to identify the view offering the most reliable structure. This view anchors the generation of pseudo-labels, ensuring that supervision is based on the strongest available signal. Second, GAPS introduces a granularity-aware supervision mechanism that decouples the pseudo-label resolution from the final output. By generating a full hierarchy of pseudo-labels (from fine to coarse) for the best view, GAPS enables the flexible selection of a supervision level that is most aligned with the data's intrinsic structure. Finally, these components are integrated into an efficient, shallow architecture where an SCI-weighted fused distance matrix informs the training of a multi-layer perceptron (MLP) under the guidance of the selected pseudo-labels, yielding the final cluster partition. Figure 1 illustrates the basic framework of GAPS, and the main contributions of GAPS are summarized as follows:

- We propose a principled mechanism for identifying and exploiting the most informative view for pseudo-label generation, addressing view heterogeneity and quality issues in multi-view scenarios.
- We introduce a flexible multi-view clustering framework that aligns pseudo-label supervision with the data's intrinsic granularity, enabling more robust and expressive representation learning.
- We integrate these processes into a simple, interpretable, and effective MVC framework, whose superiority in both flexible supervision and reliable view selection is validated by extensive experiments compared with state-of-the-art competitors on synthetic and real-world datasets.

Methodology

Reliability-Aware View Selection

Given multi-view data $\{\mathbf{V}^{(p)}\}_{p=1}^{P}$, for each view p and an integrated view $\mathbf{V}^{*}$, we compute the pairwise distance matrix $\mathbf{D}^{(p)}$ (or $\mathbf{D}^{*}$, the average of all $\mathbf{D}^{(p)}$) and employ an adjacency-constrained hierarchical clustering (CHC) mechanism originating from Torque Clustering (Yang and Lin 2024a, 2025, 2024b). We adopt CHC for each view because its constrained merging mechanism not only corrects outlier misclassification common in traditional hierarchical methods (e.g., average-linkage, Ward-linkage), but also produces high-purity pseudo labels at multiple granularities. For any current set of clusters $\mathbb{C}$, the nearest neighbor of cluster $C_i \in \mathbb{C}$ is defined as:

$$\mathrm{NN}_{\mathbb{C}}(i) = \arg \min_{C_j \in \mathbb{C} \setminus \{C_i\}} d(C_i, C_j), \tag{1}$$

where $d(C_i, C_j)$ denotes the distance between clusters C_i and C_j, measured via the single-linkage criterion. Clusters C_i and C_j are eligible for merging if $j = \mathrm{NN}_{\mathbb{C}}(i)$ and $|C_i| \leq |C_j|$. The corresponding adjacency matrix $\mathcal{A}$ is constructed by

$$\mathcal{A}_{ij} = \begin{cases} 1, \text{if } j = \mathrm{NN}_{\mathbb{C}}(i),\ |C_i| \leq |C_j| \\ 0, \text{otherwise.} \end{cases} \tag{2}$$

where C_i and C_j are the i-th and j-th clusters, respectively, and $|C_i|$ denotes the number of samples in C_i. To obtain K clusters, the K-1 links with the largest merging costs are removed, where the merging cost is defined as:

$$\mathrm{Cost}(C_i, C_j) = d(C_i, C_j)^2 \times |C_i| \times |C_j|. \tag{3}$$

For each view (including the fused view), we denote the resulting K-partition as $\mathcal{P}^{(p)}$. The separation-compactness index (SCI) for a partition $\mathcal{P}$ is computed as:

$$\mathrm{SCI}(\mathcal{P}) = \frac{\min\limits_{u \neq v} \min\limits_{x \in S_u,\, y \in S_v} \mathrm{dist}(x, y)}{\max\limits_{w} \dfrac{1}{|S_w|(|S_w| - 1)} \sum\limits_{a, b \in S_w, a \neq b} \mathrm{dist}(a, b)}, \tag{4}$$

where S_u is the set of samples in cluster u, and $dist(\cdot, \cdot)$ denotes the pairwise sample distance, for which we use the

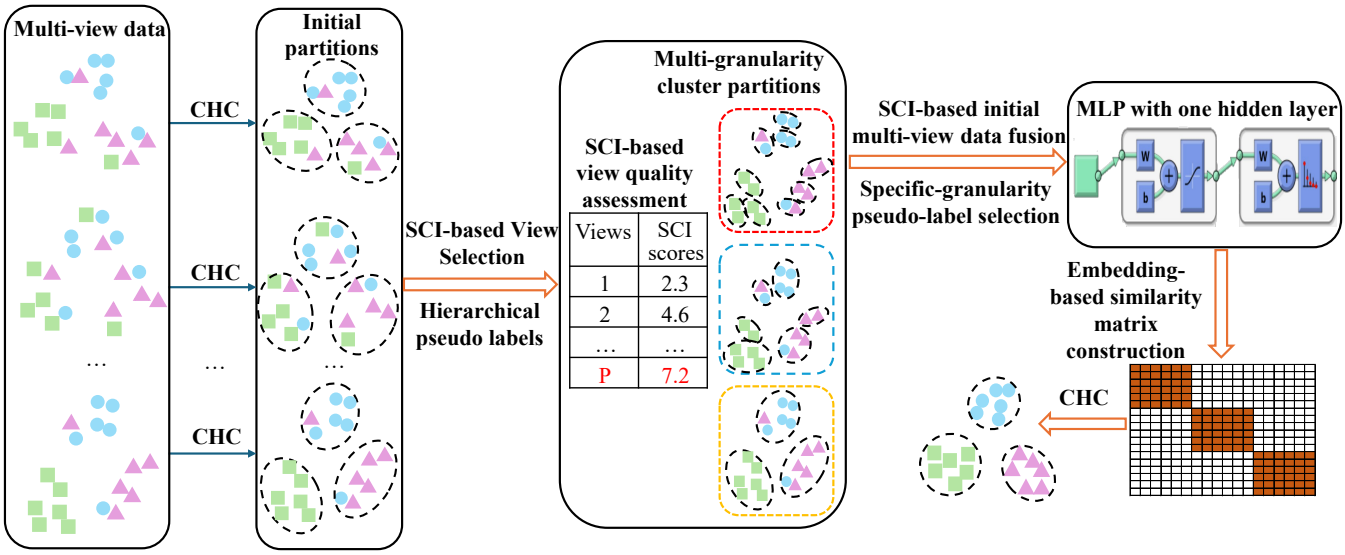


Figure 1: The proposed GAPS framework. It first identifies the most informative view by evaluating initial CHC-based partitions of each view with the SCI validity metric. A shallow MLP then performs pseudo-supervised representation learning, mapping an SCI-weighted fusion of all views to a specific granularity of pseudo labels selected from a hierarchy of the best view. Finally, the learned embedding is used to construct a refined similarity matrix, on which CHC is performed to provide the final partition.

cosine metric. The numerator measures the minimal inter-cluster distance, and the denominator measures the maximal intra-cluster dispersion. We employ the SCI metric to assess the quality of each view, as its boundary-sensitive and centroid-free nature makes it particularly effective in identifying datasets with non-convex cluster structures. In contrast, traditional internal validity indices such as the Silhouette score (Rousseeuw 1987) are primarily suited to convex clusters and offer limited robustness in complex scenarios. The most cluster-informative view is selected as:

$$p^{\dagger} = \arg\max_{p} \mathrm{SCI}(\mathcal{P}^{(p)}). \tag{5}$$

The superior inter-cluster separation and intra-cluster compactness of the optimal view $p^{\dagger}$ signify a high degree of clusterability, thereby providing a robust foundation for the subsequent generation of multi-granularity pseudo labels.

Granularity-Aware Pseudo-label Extraction

For the selected best view $p^{\dagger}$, we recursively run CHC without fixing the number of clusters, generating $m_{p^{\dagger}}$ hierarchy levels. For each level g, the pseudo-label vector is:

$$\mathbf{l}_g \in \{1, \ldots, K_g\}^n, \quad 1 \le g \le m_{p^{\dagger}}, \tag{6}$$

where $\mathbf{l}_g$ is the $n \times 1$ cluster assignment vector for all samples at the granularity level g, and K_g is the number of clusters at the level g.

The specific granularity level of the best view $p^{\dagger}$ is determined by the termination parameter $\lambda \in (0, 1]$:

$$k_{p^{\dagger}} = \max\left(1, \left\lfloor \lambda m_{p^{\dagger}} + 0.5 \right\rfloor\right), \tag{7}$$

where $m_{p^{\dagger}}$ is the total number of hierarchy levels for $p^{\dagger}$, and $\lfloor \cdot \rfloor$ denotes rounding to the nearest integer. We then select the pseudo-label vector $\mathbf{l}_{k_{p^{\dagger}}}$ as the supervision signal for representation learning. Our framework departs from fixed-resolution supervision by generating a full hierarchy of multi-granularity pseudo-labels via hierarchical clustering. A single granularity parameter λ navigates this hierarchy to select a supervisory signal whose resolution aligns with the intrinsic multiscale topology of data. The granularity parameter λ governs a critical trade-off: lower λ values favor fine-grained pseudo labels that can capture subtle boundaries missed by fixed-granularity methods but risk noise sensitivity, while higher λ values yield coarse, robust labels that may over-merge distinct sub-groups. The optimal choice is therefore a data-dependent balance between structural fidelity and noise robustness.

Pseudo-Label Guided Embedding Learning

We construct a fused distance matrix as an SCI-weighted combination of the individual view distance matrices:

$$\mathbf{F} = \sum_{p=1}^{P} w^{(p)} D^{(p)}, \quad \text{where } w^{(p)} = \frac{\mathrm{SCI}^{(p)}}{\sum_{q=1}^{P} \mathrm{SCI}^{(q)}}. \tag{8}$$

Here, $\mathrm{SCI}^{(p)}$ measures the clustering reliability of each view, and the normalized SCI score $w^{(p)}$ serves as an adaptive weight, allowing more reliable views to contribute more to the initial fusion. We employ distance-matrix-based fusion rather than direct feature concatenation for two reasons: (i) it overcomes inconsistencies from heterogeneous feature spaces, and (ii) it yields a unified similarity structure well-suited for downstream clustering. By weighting views according to their SCI scores, this approach also suppresses the effects of uninformative views.

A shallow neural network with one hidden layer of width q is trained to predict the pseudo-labels:

$$\mathbf{z}_i = \sigma(\mathbf{f}_i \mathbf{W}_1 + \mathbf{b}_1), \tag{9}$$

where $\sigma(\cdot)$ is a nonlinear activation (e.g., tanh), $\mathbf{f}_i$ is the feature vector for sample i, $\mathbf{W}_1$ and $\mathbf{b}_1$ are weights and biases for the hidden layer, and $\mathbf{z}_i$ is the hidden representation.

$$\hat{\mathbf{y}}_i = \text{softmax}(\mathbf{z}_i \mathbf{W}_2 + \mathbf{b}_2), \tag{10}$$

where $\mathbf{W}_2$ and $\mathbf{b}_2$ are weights and biases for the output layer, and $\hat{\mathbf{y}}_i$ is the predicted label distribution.

The network is trained by minimizing the cross-entropy loss:

$$\mathcal{J} = -\frac{1}{n} \sum_{i=1}^{n} \sum_{c=1}^{K_{k_{p^\dagger}}} \mathbf{1}(l_{k_{p^\dagger},i} = c) \log \hat{y}_{i,c}, \tag{11}$$

where $l_{k_{p^\dagger},i}$ is the assigned pseudo-label for sample i, and $K_{k_{p^\dagger}}$ is the number of clusters at the selected granularity.

The learned hidden representation $\mathbf{Z} \in \mathbb{R}^{n \times q}$ is then used to construct a k-nearest neighbor similarity matrix:

$$\mathcal{S}_{ij} = \begin{cases} \exp\left(-\frac{d(\mathbf{z}_i, \mathbf{z}_j)^2}{\tau^2}\right), & \text{if } \mathbf{z}_j \in \mathcal{N}_i^{(k)} \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

where $\mathcal{N}_i^{(k)}$ denotes the k-nearest neighbors of sample i, and

$$\tau^2 = \frac{1}{nk} \sum_{i=1}^{n} \sum_{\mathbf{z}_j \in \mathcal{N}_i^{(k)}} d(\mathbf{z}_i, \mathbf{z}_j)^2 \tag{13}$$

is the average squared distance among all k-NN pairs (Zhang et al. 2012).

CHC is finally performed on $1 - \mathcal{S}_{ij}$ as the pairwise distance, cutting at K clusters to produce the final partition. The framework and pseudo-code of GAPS are shown in Figure 1 and Algorithm 1, respectively.

Computational Complexity Analysis

The computational complexity of GAPS is primarily determined by the steps involving distance computation and hierarchical clustering. For each of the P views and the fused view, pairwise distance matrix calculation (Step 2) and CHC clustering (Step 3) requires $\mathcal{O}(Pn^2)$. Step 6, which recursively applies CHC on the selected view, requires $\mathcal{O}(n^2)$ cost. The construction of the kNN kernel and final clustering (Steps 12-13) together also require up to $\mathcal{O}(n^2)$. All other steps, including SCI computation, pseudo-label selection, feature concatenation, and shallow MLP training (Steps 4-11), are linear or near-linear with respect to n and thus negligible in the total cost. Overall, GAPS has a total complexity of $\mathcal{O}(Pn^2)$ in standard settings, which may be reduced to $\mathcal{O}(Pn \log n)$ if efficient kd-tree or approximate kNN acceleration is used in low-dimensional scenarios (Fei et al. 2025b,a).

Algorithm 1: Granularity-Aware Pseudo Supervision.

Input: Multi-view data $\{\mathbf{V}^{(p)}\}_{p=1}^{P}$, number of clusters K, granularity parameter λ, neighbor count k;

1: **for** each view $p = 1, \cdots, P$ and fused view:
2: Compute the distance matrix;
3: Run CHC and cut tree at K clusters by Eqs. (1)-(3), yielding the cluster partition;
4: Compute SCI index for each partition by Eq. (4);
5: **end for**
6: Identify the best view $p^\dagger$ using Eq. (5);
7: Based on $D^{(p^\dagger)}$, recursively run CHC for full hierarchy;
8: Record pseudo labels $\mathbf{l}_g$ for each level g by Eq. (6);
9: Compute the granularity level $k_{p^\dagger}$ by Eq. (7);
10: Select $\mathbf{l}_{k_{p^\dagger}}$ as the unique pseudo-label vector;
11: Construct the initial representation into $\mathbf{F}$ by Eq. (8);
12: Train MLP using $(\mathbf{F}, \mathbf{l}_{k_{p^\dagger}})$ to learn $\mathbf{Z}$ by Eqs. (9)-(11);
13: Build $\mathcal{S}$ using Eqs. (12) and (13);
14: Run CHC on $1 - \mathcal{S}_{ij}$ to cut at K clusters.

Output: Final cluster assignments for all samples.

Experiment Analysis

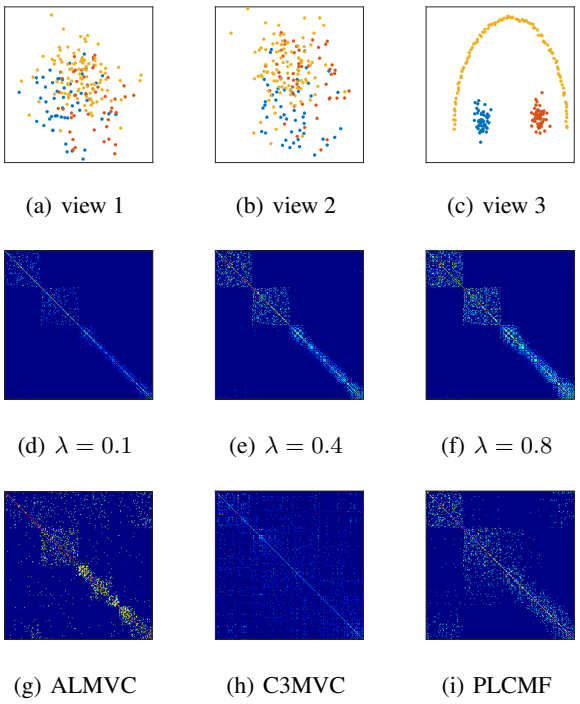
This section provides comprehensive experiments on synthetic data and six real-world datasets and compares the proposed GAPS with state-of-the-art competitors. The synthetic data contains three views, as shown in Figures 2(a)–(c), and the details of real-world datasets are summarized in Table **??**.

Experimental Settings

We compare the proposed GAPS with nine MVC methods, including three pseudo-label guided methods, i.e., Pseudo-Label Collective Matrix Factorization for Multi-View Clustering (**PLCMF**) (Wang et al. 2022), Continual Multi-View Clustering with Consistent Anchor Guidance (**C3MVC**) (Zhang et al. 2024b), and Anchor Learning for Multi-View Clustering (**ALMVC**) (Ma et al. 2024), six fusion-based or latent space alignment-based methods, i.e., Graph-based Multi-View Clustering (**GMC**) (Wang, Yang, and Liu 2020), Multi-view Subspace Clustering on Topological Manifold (**TMMSC**) (Hu et al. 2025), Simple one-step Multi-View Clustering (**SONIC**) (Xia et al. 2025), View Variation and View Heredity for Multi-View Clustering (**V3H**) (Fang et al. 2021), Multi-View Clustering With Incremental Instances and Views (**MVCIIV**) (Zhang et al. 2025a), and Weighted Multi-View Spectral Clustering (**WMSC**) (Zong et al. 2018). All methods are evaluated using two widely used external metrics: Accuracy (ACC) and Normalized Mutual Information (NMI) (Strehl and Ghosh 2002). GAPS is configured with an MLP containing 100 tanh hidden units (i.e., $q = 100$) and a softmax output, optimized using cross-entropy loss and scaled conjugate-gradient (learning rate = 0.01). Parameters for each baseline are tuned according to their original settings to ensure a fair comparison.

Datasets	Views (features)	Samples	Clusters
UCI	3(64/76/216)	2000	10
COIL-20	3(1024/3394/6750)	1440	20
HW	2(76/240)	2000	10
ORL	3(4096/3304/6750)	400	40
UMIST	3(30/30/30)	575	20
CMU-PIE	3(30/30/30)	2856	68

Table 1: Details of real-world multi-view datasets



(a) view 1 (b) view 2 (c) view 3

(d) $\lambda = 0.1$ (e) $\lambda = 0.4$ (f) $\lambda = 0.8$

(g) ALMVC (h) C3MVC (i) PLCMF

Figure 2: Results on the synthetic data, in which (a)-(c) show the original data distributions of three views, (d)-(f) shows the similarity matrices learned by GAPS with different granularity pseudo labels (i.e., $\lambda = 0.1, 0.4, 0.8$), (g)-(i) are the results of ALMVC, C3MVC, and PLCMF, respectively.

Experiments on Synthetic Dataset

To intuitively highlight the advantages of GAPS, we generate a synthetic dataset comprising three heterogeneous views, as depicted in Figures 2(a)-(c). Unlike the methods (i.e., ALMVC, C3MVC, and PLCMF) that rely on fixed-granularity pseudo labels, GAPS leverages the granularity-aware pseudo-label supervision strategy to align the multi-scale intrinsic structures of data in different views. This strategy enables the GAPS to learn more discriminative representations. As demonstrated in Figures 2(d)-(f), by varying the number of granularities (i.e., setting λ to 0.1, 0.4, and 0.8 to span from fine to coarse granularities), we observe that coarse-granularity supervision ($\lambda = 0.8$) leads to clearer and more well-separated cluster structures in the learned representations. Moreover, the SCI-based internal evaluation criterion in GAPS facilitates the selection of the most informative view (i.e., view 3), effectively mitigating the adverse impacts of less informative views during the view fusion process. In contrast, Figures 2(g)-(i) present the results of ALMVC, C3MVC, and PLCMF, respectively. Specifically, ALMVC is restricted to fixed-granularity pseudo labels and fails to select informative views. C3MVC fuses pseudo labels in a uniform manner across all views yet ignores view heterogeneities and granularities, which results in over-segmented structures when views vary in quality. PLCMF likewise uses fixed-granularity pseudo labels and lacks specific mechanisms to identify or filter out poor views, leading to even noisier results.

Methods	UCI	COIL20	HW	ORL	UMIST	CMU-PIE	Average
GMC	.8495	.7910	.8300	.6325	.5217	.7048	.7216
TMMSC	.9024	.8042	.9105	.7825	.5160	.7953	.7852
SONIC	.7680	.6778	.7585	.5950	.4452	.4741	.6188
V3H	.9051	.6012	.8669	.7412	.5294	.7231	.7278
MVCIIV	.9100	.8472	.9520	.8375	.6609	.8764	.8473
WMSC	.8410	.8465	.8335	.8300	.4539	.6590	.7440
C3MVC	.8595	.7569	.8000	.7575	.5026	.6572	.7223
PLCMF	.8780	.6660	.9270	.6775	.4957	.6537	.7163
ALMVC	.9085	.5785	.8720	.6250	.3461	.5473	.6462
GAPS	**.9645**	**.9979**	**.9705**	**.8550**	**.8835**	**1**	**.9452**

Table 2: ACC on Real-world Datasets.

Methods	UCI	COIL20	HW	ORL	UMIST	CMU-PIE	Average
GMC	.9013	.9410	.8767	.8590	.7373	.8892	.8674
TMMSC	.8885	.9190	.9032	.7800	.7223	.9072	.8534
SONIC	.7499	.7969	.7673	.7751	.6498	.6645	.7339
V3H	.8118	.7639	.7425	.8633	.6859	.8667	.7890
MVCIIV	.8336	.9548	.9126	.9029	.8416	.9710	.9028
WMSC	.8839	.9486	.8772	.8985	.7015	.8571	.8611
C3MVC	.7604	.8265	.7012	.9214	.7240	.8027	.7894
PLCMF	.8043	.7612	.8509	.8227	.6678	.7815	.7814
ALMVC	.8319	.7151	.7836	.8034	.4702	.7269	.7219
GAPS	**.9236**	**.9971**	**.9329**	**.9309**	**.9393**	**1**	**.9540**

Table 3: NMI on Real-world Datasets.

Experiments on Real-world Datasets

The ACC and NMI results of GAPS and compared methods across six datasets are summarized in Table **??** and Table **??**, respectively. We can observe that GAPS outperforms other competitors across all datasets, with average improvements of 22.29% in ACC and 16.46% in NMI over the best pseudo-label driven method (i.e., C3MVC), and 9.79% in ACC and 5.12% in NMI over the best fusion-based MVC methods (i.e., MVCIIV). This consistent superiority can be primarily attributed to GAPS's dual innovations: the granularity-aware pseudo-label supervision and the adaptive view selection via the SCI criterion. By dynamically aligning pseudo-label guidance to the intrinsic granularity of each view, GAPS can learn discriminative representations from data. The use of the SCI-based internal criterion enables GAPS to automatically identify and further exploit the most informative view for representation learning, thereby reducing the adverse impacts of poor views on multi-view fusion. In contrast, the pseudo-label-driven

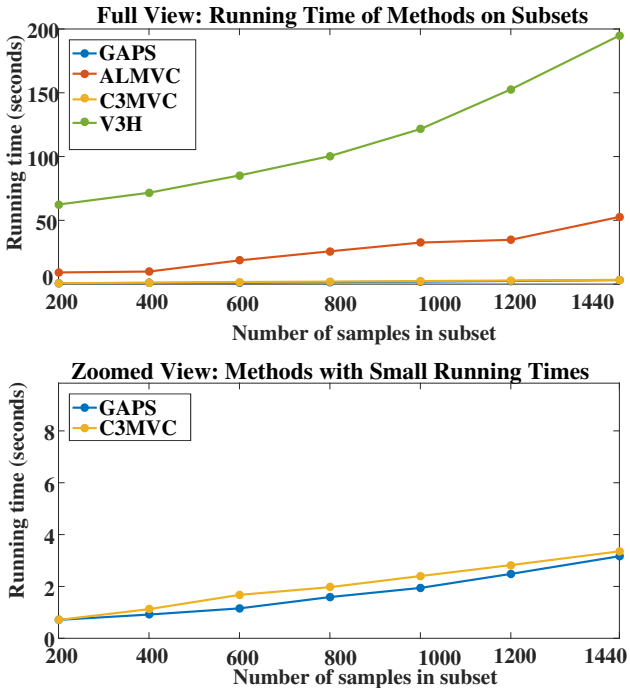


Figure 3: Runtime of GAPS and other methods.

methods such as ALMVC, C3MVC, and PLCMF rely solely on fixed-granularity pseudo labels and apply uniform aggregation across all views. This lack of flexibility and absence of informative-view selection make them susceptible to uninformative views, especially in the presence of heterogeneous structures. Meanwhile, the fusion-based methods like MVCIIV and TMMSC require explicit view alignments and involve extensive spectral computations, limiting their scalability and robustness when dealing with diverse multi-view data. As a result, these competitors cannot fully explore the complementary information embedded in heterogeneous views. By the joint use of granularity-aware supervision and automatic selection of cluster-informative views, GAPS maximizes the utility of multi-view data, yielding superior clustering performance. Finally, it is worth noting that the superiority of GAPS is particularly prominent on COIL20, UMIST, and CMU-PIE. We attribute this significant advantage to the use of CHC instead of the K-means or spectral clustering used by most competitors. The constrained merging mechanism of CHC is more effective at identifying the sparse, non-convex structures often found in such image data, thereby reducing misclassifications along cluster boundaries. In comparison, partitioning methods like K-means struggle with these non-convex shapes, while spectral clustering often incorrectly connects clusters with unclear boundaries, degrading their performance.

Figure 3 presents the runtime comparison of GAPS against several representative methods on the COIL20 dataset, with the number of samples ranging from 200 to 1440. Notably, GAPS consistently demonstrates the lowest runtime across all data sizes. This efficiency can be attributed to the streamlined, shallow clustering architecture of

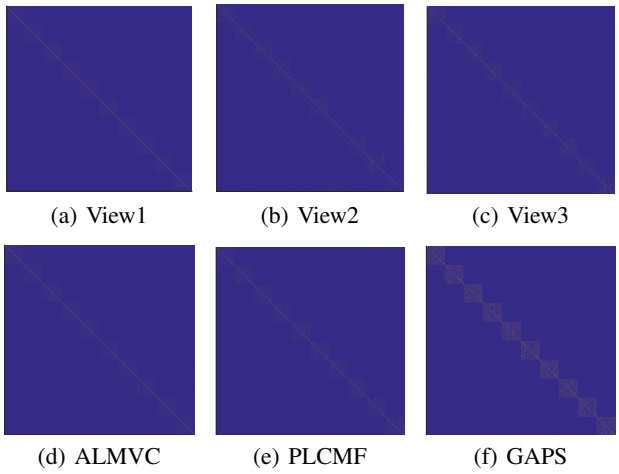


Figure 4: Visualization of learned similarity matrices for GAPS and other pseudo-label-driven methods on UCI.

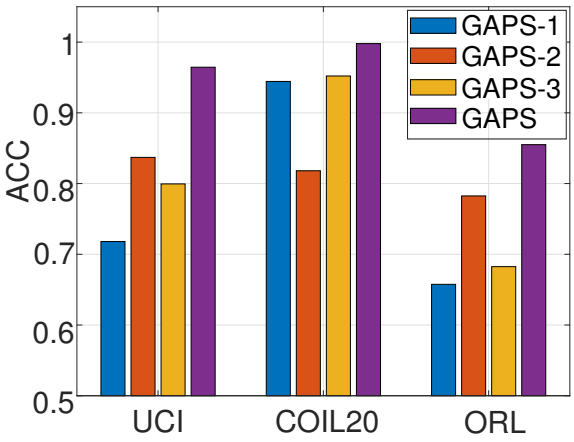
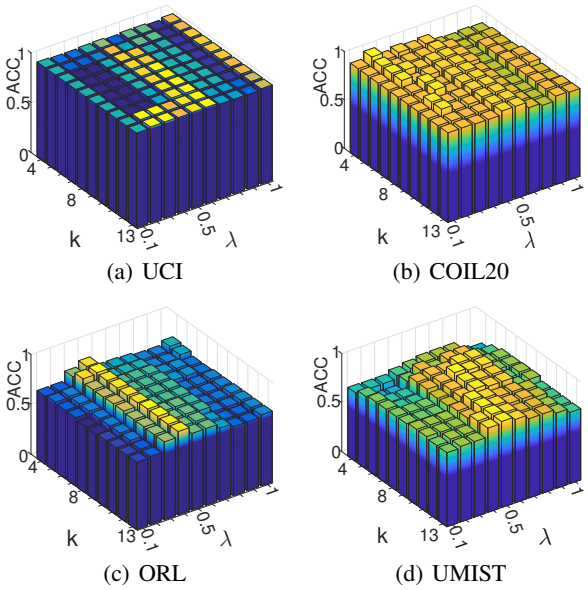


Figure 5: Comparison of GAPS with its variants.

GAPS, which avoids complex cross-view alignments found in compared methods. Instead, GAPS leverages fast pseudo-label supervision, efficient SCI-based selection, and single-pass fusion, enabling robust and scalable multi-view clustering with minimal computational costs.

Visualization

To visually highlight the effectiveness and superiority of the granularity-aware supervision and the reliability-aware view selection mechanism designed for GAPS, we compare the similarity matrix learned by GAPS to those of two pseudo-label-driven methods, i.e., ALMVC and PLCMF, on the UCI dataset. As shown in Figure 4, the first row (a)-(c) displays the similarity matrices of the original three views, while the second row (d)-(f) illustrates the matrices generated by ALMVC, PLCMF, and GAPS, respectively. It can be observed that the matrix learned by GAPS exhibits a significantly cleaner block-diagonal structure compared to the



(a) UCI (b) COIL20

(c) ORL (d) UMIST

Figure 6: Performance of GAPS with different parameters.

original views, indicating more compact intra-cluster similarity and greater inter-cluster separation. While ALMVC attempts to use a reliable reference view, its passive fusion strategy still allows noise from inferior views to contaminate the final representation, resulting in indistinct cluster boundaries. Conversely, PLCMF produces a sparser matrix but struggles to preserve the integrity of certain clusters, i.e., a direct consequence of its reliance on a fixed granularity for pseudo-labels, restricting its ability to adaptively capture finer or coarser clustering structures of data. This visualization demonstrates that the adaptive combination of view selection and granularity-aware supervision can facilitate learning a more effective representation for clustering.

Ablation Study

To further elucidate the contribution of each component within the proposed GAPS framework, we conducted a series of ablation studies. Specifically, GAPS-1 replaces the SCI index with the classic Silhouette index for best-view selection. GAPS-2 substitutes the CHC algorithm with the conventional Ward linkage method for both multi-granularity pseudo-label generation and final clustering. In GAPS-3, no selection criterion is applied; instead, multi-granularity pseudo-labels are generated by concatenating all views to guide representation learning. As illustrated in Figure 5, the ACC scores of GAPS-1, GAPS-2, and GAPS-3 across the three datasets are consistently lower than those achieved by the complete GAPS framework, highlighting the essential role of each proposed mechanism in achieving superior clustering performance.

Parameter Sensitivity Analysis

The proposed GAPS involves two parameters: λ, which controls the granularity level (from fine to coarse) of the pseudo labels selected from the most informative view, and k, which defines the number of nearest neighbors for constructing the similarity matrix from embedding. Figure 6 illustrates the ACC scores of GAPS across these parameters on the UMIST, COIL20, ORL, and UCI datasets, revealing overall stable performance. We find that the optimal λ varies significantly across datasets (e.g., 0.2 for COIL20, 0.4 for ORL, 0.6 for UCI, and 0.7 for UMIST), further highlighting the importance of granularity-aware supervision. Specifically, the finer granularity (corresponding to a smaller λ) enables the capture of subtle cluster structures but might increase the sensitivity to noise, whereas the coarser granularity (corresponding to a larger λ) can enhance the robustness to noise at the cost of merging distinct sub-clusters. Meanwhile, GAPS demonstrates notable insensitivity to k. This is because the pseudo-supervised embedding refinement module can project data into a more separable latent space, thereby mitigating the severity of this trade-off and ensuring the final partitioning is robust across a wide range of neighborhood definitions. These findings suggest that the selection of λ should consider the inherent multiscale characteristics of data, achieving promising performance.

Conclusion

In this paper, we introduced GAPS, a novel framework designed to address fundamental limitations in multi-view clustering. We identified that existing methods are often constrained by rigid pseudo-label granularity and are susceptible to noise from suboptimal view integration. Our solution successfully tackled these issues through two primary contributions: the introduction of a pseudo-supervision mechanism with flexible, controllable granularity, and a principled, reliability-driven strategy for selecting the most informative view. Our comprehensive experimental evaluation confirmed the efficacy of this approach, showing that GAPS not only achieves state-of-the-art clustering accuracy on multiple benchmarks but also maintains remarkable computational efficiency due to its simple design. The results underscore the importance of aligning supervision signals with the inherent structure of the data. Future work could explore automating the selection of the granularity parameter to develop a fully adaptive framework. For example, investigating the clusterability of the learned hidden representation $\mathbf{Z}$ using SCI may provide a principled way to automatically determine λ. Furthermore, extending the principle of granularity-aware supervision to other unsupervised or semi-supervised learning paradigms presents a promising research avenue.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (No. 62306171) and the Scientific Research Fund of Zhejiang Provincial Education Department.

References

Cai, R.; Chen, H.; Mi, Y.; Luo, C.; Horng, S.-J.; and Li, T. 2024. Multi-view clustering via pseudo-label guide learn-

ing and latent graph structure recovery. *Pattern Recognition*, 151: 110420.

Chen, M.; Wang, C.; and Lai, J. 2023. Low-rank tensor based proximity learning for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(5): 5076–5090.

Chen, M.; Zhu, X.; Lin, J.; and Wang, C. 2025a. Contrastive multiview attribute graph clustering with adaptive encoders. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4): 7184–7195.

Chen, M.-S.; Huang, L.; Wang, C.-D.; and Huang, D. 2020. Multi-view clustering in latent embedding space. In *Proceedings of the AAAI conference on artificial intelligence*, 3513–3520.

Chen, Y.; Wang, H.; Peng, J.; and Wang, Y. 2025b. Anchor Learning with Potential Cluster Constraints for Multi-view Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 15939–15947.

Cui, J.; Wu, X.; Zhang, H.; Dong, C.; and Wen, J. 2025. Structure-guided deep multi-view clustering. *Information Fusion*, 103461.

Fang, X.; Hu, Y.; Zhou, P.; and Wu, D. O. 2021. V3H: View variation and view heredity for incomplete multiview clustering. *IEEE Transactions on Artificial Intelligence*, 1(3): 233–247.

Fei, Z.; Ma, Y.; Zhao, J.; Wang, B.; and Yang, J. 2025a. KNEG-CL: Unveiling data patterns using a k-nearest neighbor evolutionary graph for efficient clustering. *Information Sciences*, 690: 121602.

Fei, Z.; Zhai, H.; Yang, J.; Wang, B.; and Ma, Y. 2025b. Discovering generalized clusters with adaptive mixture density-based clustering. *Knowledge-Based Systems*, 314: 113250.

Gan, Y.; You, Y.; Huang, J.; Xiang, S.; Tang, C.; Hu, W.; and An, S. 2025. Multi-view clustering via multi-stage fusion. *IEEE Transactions on Multimedia*, 24: 2461–2472.

Guo, Q.; Liang, X.; Qian, Y.; Cui, Z.; and Wen, J. 2024. A Progressive Skip Reasoning Fusion Method for Multi-Modal Classification. In *Proceedings of the ACM International Conference on Multimedia*, 429–437.

Hu, S.; Zhang, C.; Zou, G.; Lou, Z.; and Ye, Y. 2025. Deep multiview clustering by pseudo-label guided contrastive learning and dual correlation learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2): 3646–3658.

Huang, D.; Wang, C.-D.; and Lai, J.-H. 2023. Fast multi-view clustering via ensembles: Towards scalability, superiority, and simplicity. *IEEE Transactions on Knowledge and Data Engineering*, 35(11): 11388–11402.

Jiang, B.; Zhang, C.; Liang, X.; Zhou, P.; Yang, J.; Wu, X.; Guan, J.; Ding, W.; and Sheng, W. 2025a. Collaborative Similarity Fusion and Consistency Recovery for Incomplete Multi-view Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 17617–17625.

Jiang, B.; Zhang, C.; Wang, Z.; Liang, X.; Zhou, P.; Du, L.; Zhang, Q.; Ding, W.; and Liu, Y. 2025b. Scalable fuzzy clustering with collaborative structure learning and preservation. *IEEE Transactions on Fuzzy Systems*, 33(9): 3047–3060.

Li, X.; Sun, Y.; Sun, Q.; Ren, Z.; and Sun, Y. 2023. Cross-view graph matching guided anchor alignment for incomplete multi-view clustering. *Information Fusion*, 100: 101941.

Li, Z.; Tang, C.; Liu, X.; Zheng, X.; Zhang, W.; and Zhu, E. 2021. Consensus graph learning for multi-view clustering. *IEEE Transactions on Multimedia*, 24: 2461–2472.

Liang, X.; Li, S.; Guo, Q.; Qian, Y.; Jiang, B.; Luo, T.; and Du, L. 2025a. Evolutionary Multi-View Classification via Eliminating Individual Fitness Bias. In *Proceedings of the Annual Conference on Neural Information Processing Systems*.

Liang, X.; Wang, S.; Qian, Y.; Guo, Q.; Du, L.; Jiang, B.; Luo, T.; and Li, F. 2025b. Trusted Multi-View Classification with Expert Knowledge Constraints. In *Proceedings of the International Conference on Machine Learning*, 37409–37426.

Liu, S.; Wang, S.; Liang, K.; Zhang, J.; Dong, Z.; Liu, T.; Zhu, E.; He, K.; and Liu, X. 2024a. Alleviate anchor-shift: Explore blind spots with cross-view reconstruction for incomplete multi-view clustering. 87509–87531.

Liu, S.; Zhang, J.; Wen, Y.; Yang, X.; Wang, S.; Zhang, Y.; Zhu, E.; Tang, C.; Zhao, L.; and Liu, X. 2024b. Sample-level cross-view similarity learning for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, 14017–14025.

Liu, W.; Chen, Y.; and Yue, X. 2025. Enhancing Multi-View Classification Reliability with Adaptive Rejection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 18969–18977.

Ma, H.; Wang, S.; Yu, S.; Liu, S.; Huang, J.; Wu, H.; Liu, X.; and Zhu, E. 2024. Automatic and aligned anchor learning strategy for multi-view clustering. In *Proceedings of the ACM International Conference on Multimedia*, 5045–5054.

Nie, F.; Cai, G.; Li, J.; and Li, X. 2018. Auto-weighted multi-view learning for image clustering and semi-supervised classification. *IEEE Transactions on Image Processing*, 27(3): 1501–1511.

Nie, F.; Shi, S.; Li, J.; and Li, X. 2023. Implicit weight learning for multi-view clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8): 4223–4236.

Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20: 53–65.

Strehl, A.; and Ghosh, J. 2002. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3(12): 583–617.

Sun, Y.; Li, Y.; Ren, Z.; Duan, G.; Peng, D.; and Hu, P. 2025. ROLL: Robust Noisy Pseudo-label Learning for Multi-View Clustering with Noisy Correspondence. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 30732–30741.

Sun, Y.; Qin, Y.; Li, Y.; Peng, D.; Peng, X.; and Hu, P. 2024. Robust multi-view clustering with noisy correspondence. *IEEE Transactions on Knowledge and Data Engineering*, 36(12): 9150–9162.

Wan, X.; Liu, J.; Gan, X.; Liu, X.; Wang, S.; Wen, Y.; Wan, T.; and Zhu, E. 2025. One-step multi-view clustering with diverse representation. *IEEE Transactions on Neural Networks and Learning Systems*, 36(3): 5774–5786.

Wang, B.; Wang, J.; Zeng, C.; and Chen, M. 2025a. Locality-driven flexible consensus graph learning for multi-view clustering. *Pattern Recognition*, 111853.

Wang, B.; Zeng, C.; Chen, M.; and Li, X. 2025b. Towards Learnable Anchor for Deep Multi-View Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 21044–21052.

Wang, D.; Han, S.; Wang, Q.; He, L.; Tian, Y.; and Gao, X. 2022. Pseudo-label guided collective matrix factorization for multiview clustering. *IEEE Transactions on Cybernetics*, 52(9): 8681–8691.

Wang, H.; Yang, Y.; and Liu, B. 2020. GMC: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6): 1116–1129.

Wang, S.; Liu, X.; Liao, Q.; Wen, Y.; Zhu, E.; and He, K. 2025c. Scalable multi-view graph clustering with cross-view corresponding anchor alignment. *IEEE Transactions on Knowledge and Data Engineering*, 37(5): 2932–2945.

Wang, Z.; Li, X.; Sun, Y.; Sun, Q.; Sun, Y.; Ling, H.; Dai, J.; and Ren, Z. 2025d. TPCH: tensor-interacted projection and cooperative hashing for multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 21420–21428.

Wang, Z.; Yang, J.; Guan, J.; Zhang, C.; Liang, X.; Jiang, B.; and Sheng, W. 2025e. Enhanced Density Peak Clustering for High-Dimensional Data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 21411–21419.

Wen, J.; Zhang, Z.; Fei, L.; Zhang, B.; Xu, Y.; Zhang, Z.; and Li, J. 2023. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2): 1136–1149.

Wu, D.; Wang, P.; Lu, J.; Hu, Z.; Zhang, H.; and Nie, F. 2025. Triangle Topology Enhancement for Multi-view Graph Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 37(7): 4338–4348.

Wu, D.; Yang, Z.; Lu, J.; Xu, J.; Xu, X.; and Nie, F. 2024. EBMGC-GNF: Efficient Balanced Multi-view Graph Clustering via Good Neighbor Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 7878–7892.

Xia, X.; Huang, D.; Yang, C.; He, C.; and Wang, C. 2025. Simple One-Step Multi-View Clustering With Fast Similarity and Cluster Structure Learning. *IEEE Signal Processing Letters*, 32: 1850–1854.

Xu, Y.; Wang, C.; and Lai, J. 2016. Weighted multi-view clustering with feature selection. *Pattern Recognition*, 53: 25–35.

Yang, J.; Chen, W.; Liu, F.; Zhou, P.; Wang, Z.; Liang, X.; and Jiang, B. 2025. Multi-view clustering via multi-granularity ensemble. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 6794–6802.

Yang, J.; and Lin, C.-T. 2022. Multi-view adjacency-constrained hierarchical clustering. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(4): 1126–1138.

Yang, J.; and Lin, C.-T. 2024a. Enhanced Adjacency-Constrained Hierarchical Clustering Using Fine-Grained Pseudo Labels. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(3): 2481–2492.

Yang, J.; and Lin, C.-T. 2024b. Toward autonomous distributed clustering. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(2): 2065–2072.

Yang, J.; and Lin, C.-T. 2025. Autonomous clustering by fast find of mass and distance peaks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7): 5336–5349.

Yang, J.; Ma, Y.; Zhang, X.; Li, S.; and Zhang, Y. 2017. An initialization method based on hybrid distance for k-means algorithm. *Neural computation*, 29(11): 3094–3117.

Yang, J.; Wang, Y.-K.; Yao, X.; and Lin, C.-T. 2021. Adaptive initialization method for K-means algorithm. *Frontiers in Artificial Intelligence*, 4: 740817.

Zhang, C.; Fang, Y.; Liang, X.; Wu, X.; Jiang, B.; et al. 2024a. Efficient multi-view unsupervised feature selection with adaptive structure learning and inference. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, 5443–5452.

Zhang, C.; Jiang, B.; Wang, Z.; Yang, J.; Lu, Y.; Wu, X.; and Sheng, W. 2023. Efficient multi-view semi-supervised feature selection. *Information Sciences*, 649: 119675.

Zhang, C.; Wang, Z.; Jia, X.; Li, Z.; Chen, C.; and Li, H. 2025a. Multi-view Clustering with Incremental Instances and Views. *IEEE Transactions on Image Processing*, 34: 4203–4214.

Zhang, C.; Xu, D.; Jia, X.; Chen, C.; and Li, H. 2024b. Continual multi-view clustering with consistent anchor guidance. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 5434–5442.

Zhang, P.; Pan, Y.; Wang, S.; Yu, S.; Xu, H.; Zhu, E.; Liu, X.; and Tsang, I. 2025b. Max-Mahalanobis Anchors Guidance for Multi-View Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 22488–22496.

Zhang, W.; Wang, X.; Zhao, D.; and Tang, X. 2012. Graph degree linkage: Agglomerative clustering on a directed graph. In *European conference on computer vision*, 428–441. Springer.

Zong, L.; Zhang, X.; Liu, X.; and Yu, H. 2018. Weighted multi-view spectral clustering based on spectral perturbation. In *Proceedings of the AAAI conference on Artificial Intelligence*, 4621–4628.