DMCAR-MVC: Disentangled Mixture-of-Experts with Context-Aware Routing for Multi-View Clustering

Baili Xiao[1], Ke Liang[1], Jiaqi Jin[1], Jun Wang[1], Yinbo Xu[2*], Siwei Wang[3*], En Zhu[1*]

[1]College of Computer Science and Technology,National University of Defense Technology
[2]College of Intelligence Science and Technology,National University of Defense Technology
[3]Intelligent Game and Decision Lab
{xiaobaili,jinjiaqi,xuyinbo12,enzhu}@nudt.edu.cn

Abstract

Multi-View Clustering (MVC) aims to enhance clustering performance by integrating multi-source complementary information. However, existing deep MVC methods face inherent challenges in balancing the learning of shared consensus representations with the preservation of view-specific information: independent encoders hinder effective cross-view collaboration, while a single shared encoder tends to sacrifice representation diversity. Although the recently introduced Mixture-of-Experts (MoE) model offers a novel approach to facilitating view collaboration, its flattened expert pool design often leads to entanglement between shared and specific information, and its routing mechanism limits collaboration potential by neglecting cross-view context. To address these challenges, this paper proposes a novel deep multi-view clustering framework—Decoupled Mixture-of-Experts with Context-Aware Routing for Multi-View Clustering (DMCAR-MVC). At its core is an innovative Decoupled MoE (D-MoE) architecture. We establish a public expert pool to learn cross-view shared representations while equipping each view with an independent private expert pool to capture its unique information, thereby structurally enforcing the decoupling of shared and specific representations. Building on this, we further design a Context-Aware Hierarchical Routing (CAHR) mechanism. When routing for the public expert pool, this mechanism introduces a global context vector to guide expert selection, enabling more efficient and globally informed cross-view collaboration. Finally, to optimize the model, we adopt a multi-level contrastive learning paradigm: on one hand, a cross-view alignment loss ensures semantic consistency in shared representations; on the other, an orthogonality constraint is imposed to further enhance separability between shared and specific representations. Extensive experiments on multiple benchmark datasets demonstrate that DMCAR-MVC significantly outperforms state-of-the-art methods across key clustering metrics. Additionally, comprehensive ablation studies thoroughly validate the effectiveness and necessity of each proposed component.

Introduction

Multi-view clustering (MVC)(Gu et al. 2024b,a; Hu et al. 2024, 2025; Wang et al. 2025b; Yan, Yin, and Wen 2025; Zhou et al. 2025; Guan et al. 2025), a fundamental task in unsupervised learning(Liu et al. 2019, 2021a; Zhou et al. 2021), aims to achieve more accurate and reliable data partitions by effectively integrating complementary information from heterogeneous views, a scenario prevalent across diverse domains such as bioinformatics (Mitra, Saha, and Hasanuzzaman 2020), medical image analysis (Chen, Wang, and Liang 2023), and document analysis (Zhang et al. 2019; Neeraj and Maurya 2020). While early methods based on graphs, matrix factorization, or subspace learning (Zhang et al. 2020; Zhu et al. 2021; Liu et al. 2021b) achieved some success, their typical reliance on linear model assumptions often proves insufficient for capturing the complex nonlinear structures inherent in real-world data. Consequently, Deep Multi-View Clustering (DMVC)(Yan et al. 2021; Dong et al. 2023) has emerged as the dominant paradigm, leveraging the powerful representation learning capabilities of neural networks. Nevertheless, a core dilemma persists within DMVC: effectively balancing the twin goals of inter-view collaboration and view-specificity preservation. Existing DMVC methods primarily follow two mainstream paradigms that struggle with this trade-off. The first, an Independent Encoder approach(Ke et al. 2021; Chen et al. 2021; Yan et al. 2023), equips each view with a separate encoder. While this excels at preserving the unique statistical properties of each view, it defers cross-view collaboration until a late fusion stage, resulting in insufficient information interaction and limiting the quality of the learned shared representation. In contrast, the Shared Encoder paradigm (Xu et al. 2022b; Ke et al. 2024) enforces the use of a single encoder across all views. This approach promotes consistency through parameter sharing, yet its shortcoming is a tendency to overlook the inherent heterogeneity among views, potentially amplifying the negative impact of noisy views and thereby compromising representation diversity and robustness.

To alleviate the aforementioned collaboration-specificity tension, DMVC-CE(Zhang et al. 2025) innovatively introduces a Mixture-of-Experts (MoE) framework(Jacobs et al. 1991), attempting to dynamically process inputs from different views through a shared set of expert networks. A gating network dynamically selects multiple experts for each data sample, thereby capturing diverse complementary information across views. This approach undoubtedly opens new directions in the MVC field, but its design still suffers from two fundamental flaws that require urgent resolution. First,

*Corresponding author.

DMVC-CE has the issue of Representation Entanglement. DMVC-CE employs a flattened shared expert pool design, where experts dedicated to learning cross-view commonalities are intermixed with those capturing view-specific characteristics. The optimization process lacks an explicit mechanism to disentangle these two types of information, often resulting in learned representations that entangle shared and view-specific features. The second is the context-agnostic routing mechanism. When selecting experts for a given view, the routing network's decision relies solely on that view's input, entirely ignoring valuable contextual information from other views. This "isolated decision-making" routing strategy is suboptimal, as it misses opportunities for deeper collaboration by leveraging cross-view cues to guide expert selection.

To systematically address the aforementioned challenges, this paper proposes a novel deep multi-view clustering framework—Disentangled Mixture-of-Experts with Context-Aware Routing for Multi-View Clustering (DMCAR-MVC). Unlike existing methods that attempt to process all information within a single flattened expert pool, we advocate for explicit architectural disentanglement at the model level. The core of DMCAR is an innovative Decoupled Mixture-of-Experts (D-MoE) architecture. This framework structurally enforces the separation of shared and view-specific information by establishing a common expert pool shared across all views to learn cross-view representations, while equipping each view with an independent private expert pool to capture its unique characteristics. Building upon this disentangled architecture, we further design a Context-Aware Hierarchical Routing (CAHR) mechanism to deepen inter-view collaboration. When selecting experts for the shared pool, this mechanism dynamically aggregates a global context vector from all views and uses it to guide routing decisions. This approach ensures that the selection of shared experts transcends the limitations of any single view, incorporating a global perspective to enable more efficient and intelligent cross-view collaboration. Finally, to jointly optimize the entire framework, we introduce a multi-level contrastive learning paradigm. This paradigm not only ensures semantic consistency of shared representations through cross-view alignment but also actively imposes orthogonality constraints to maximize the separability between shared and view-specific representations.

The main contributions of this paper can be summarized as follows:

- We propose a novel Disentangled Mixture-of-Experts (D-MoE) architecture, which effectively resolves the entanglement of shared and specific information in existing methods at the model structural level by introducing separate shared and private expert pools.
- We design a Context-Aware Hierarchical Routing (CAHR) mechanism that leverages global contextual information aggregated from all views to guide the selection of shared experts, significantly enhancing the efficiency and depth of cross-view collaboration.
- We introduce a carefully crafted multi-level contrastive learning objective, which not only ensures consistency of shared representations through alignment loss but also actively promotes the disentanglement of shared and specific representations via orthogonality constraints.
- Through extensive experiments on multiple benchmark datasets, we comprehensively demonstrate the superior performance of the proposed DMCAR framework over existing state-of-the-art methods.

Related Work

Deep learning has become the dominant paradigm for Multi-View Clustering (MVC), yet existing methods grapple with balancing inter-view collaboration and view-specificity (Feng et al. 2025; Dong et al. 2025b,a; Wan et al. 2024). Early approaches fall into two main categories: the independent encoder paradigm, which preserves view-specific features but suffers from insufficient cross-view interaction, and the shared encoder paradigm, such as DMIM (Mao et al. 2021), which promotes consistency via parameter sharing but risks suppressing view heterogeneity and is vulnerable to noisy data. To transcend these limitations, recent work has leveraged contrastive learning to enhance representation consistency. For instance, DealMVC (Yang et al. 2023) aligns feature similarity with a pseudo-label graph, while CONAN (Ke et al. 2021) uses information bottleneck theory to align view-specific representations. Concurrently, efforts to better preserve view-specific information have emerged. MFLVC (Xu et al. 2022b) uses multi-level feature learning, CAMVC (Du, Zheng, and Xu 2024) employs attention to weigh features, SCMVC (Wu et al. 2024) uses self-weighting to manage view reliability, and DUA-nets (Geng et al. 2021) learns sample-wise weights. However, these methods typically rely on implicit guidance via loss functions, lacking an explicit architectural mechanism to decouple shared and view-specific information—the core problem we address.

Our approach is inspired by the Mixture-of-Experts (MoE) architecture, a conditional computation paradigm with origins in early machine learning (Jacobs et al. 1991) that has seen a major resurgence. By using sparsely activated gating, MoE enables massive model scaling without a proportional increase in computational cost, making it a cornerstone of modern large language models(Chen et al. 2025; Yang et al. 2025) like GShard (Lepikhin et al. 2020), Switch Transformer (Fedus, Zoph, and Shazeer 2022), GLam (Du et al. 2022), Llama-MoE (Zhu et al. 2024), and DeepSeek-MoE (Dai et al. 2024). In the context of MVC, DMVC-CE (Zhang et al. 2025) made the pioneering effort to apply MoE to facilitate cross-view collaboration. While innovative, its design suffers from fundamental limitations: a "flat" expert pool inherently entangles shared and view-specific representations, and its gating network's reliance on single-view inputs results in a context-deficient routing mechanism that limits collaborative potential. These unresolved challenges directly motivate our DMCAR-MVC framework, which introduces a Decoupled MoE (D-MoE) architecture and Context-Aware Hierarchical Routing (CAHR) to explicitly address these issues and advance the application of MoE in multi-view clustering.

Methodology

This chapter provides a detailed exposition of our proposed Decoupled Mixture-of-Experts with Context-Aware Routing for Multi-View Clustering (DMCAR-MVC) framework. We first outline the model's overall architecture and optimization objectives. Subsequently, we delve into the core Decoupled Mixture-of-Experts (D-MoE) structure, explaining how its design separates the learning of view-shared and view-specific information. Next, we focus on the Context-Aware Hierarchical Routing (CAHR) mechanism, engineered to enhance cross-view collaboration. We then introduce the multi-level contrastive learning objectives employed for the joint optimization of the entire framework. Finally, we describe how the learned decoupled representations are utilized for data reconstruction and ultimate clustering.

Overall Framework

The overall design of our proposed DMCAR-MVC framework is illustrated in the figure 1. Given a multi-view dataset $\mathcal{X} = {\mathbf{X}^{(v)} \in \mathbb{R}^{N \times d_v}}_{v=1}^{V}$ containing V views, where N is the number of samples and d_v is the feature dimension of the v-th view, the core objective of our method is to learn a set of high-quality disentangled latent representations $\mathbf{H}$. This representation consists of two components: a shared representation $\mathbf{H}_s \in \mathbb{R}^{N \times d_s}$ aligned across all views, and a set of view-specific representations ${\mathbf{H}_p^{(v)} \in \mathbb{R}^{N \times d_p}}_{v=1}^{V}$ that capture the unique information of each view. To achieve this goal, we design a unified optimization function $\mathcal{L}_{\text{total}}$, which is composed of three key weighted loss terms:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda_1 \mathcal{L}_{\text{align}} + \lambda_2 \mathcal{L}_{\text{disentangle}}, \tag{1}$$

where, $\mathcal{L}_{\text{recon}}$ is a reconstruction loss designed to ensure that the learned representations retain the primary information of the original data. $\mathcal{L}_{\text{align}}$ is a cross-view alignment loss aimed at guaranteeing that the shared representations extracted from different views maintain semantic consistency. $\mathcal{L}_{\text{disentangle}}$ is a disentanglement loss used to enhance the separability between shared and specific representations, ensuring they learn complementary rather than redundant information. λ_1 and λ_2 are hyperparameters that balance the importance of each loss term.

Disentangled MoE Architecture

To fundamentally address the prevalent issue of representation entanglement in existing methods, we designed a novel Disentangled Mixture-of-Experts (D-MoE) architecture. This architecture enforces a structural separation between shared and specific information learning pathways through two parallel processing streams: the shared representation learning stream and the specific representation learning stream. The shared representation learning stream is responsible for extracting consistent information common across all views. To achieve this, we establish a shared expert pool $\mathcal{E}_s = \{\mathcal{E}_{s,m}\}_{m=1}^{M_s}$ containing M_s expert networks. This expert pool is common to all views, and data from any view $\mathbf{X}^{(v)}$ is routed here to learn its corresponding shared representation $\mathbf{H}_s^{(v)}$. Unlike traditional MoE, the routing process here is guided by our innovative Context-Aware Hierarchical Routing (CAHR) mechanism, which will be detailed in next section. For an input $\mathbf{X}^{(v)}$ from view v, the generation of its shared representation can be formalized as:

$$\begin{aligned}\mathbf{H}_s^{(v)} &= \text{MoE}_s(\mathbf{X}^{(v)} | \mathbf{C}_{\text{global}}) \\ &= \sum_{m=1}^{M_s} G_{s,m}(\mathbf{X}^{(v)}, \mathbf{C}_{\text{global}}) \cdot \mathcal{E}_{s,m}(\mathbf{X}^{(v)}),\end{aligned} \tag{2}$$

where, $G_{s,m}$ represents the weights generated by the context-aware gating network, and $\mathbf{C}_{\text{global}}$ is the global context that aggregates information from all views.

In parallel, the view-specific representation learning stream aims to capture the unique statistical properties and structural information of each individual view. For each view v, we equip it with an independent private expert pool $\mathcal{E}_p^{(v)} = {\mathcal{E}_{p,m}^{(v)}}_{m=1}^{M_p}$, consisting of M_p experts. The data $\mathbf{X}^{(v)}$ of view v is exclusively fed into its dedicated private expert pool $\mathcal{E}_p^{(v)}$ to extract the view-specific representation $\mathbf{H}_p^{(v)}$. Here, the routing mechanism is conventional, with decisions relying solely on the input of the current view, thereby ensuring the view-specific nature of the learned information. The computation process is as follows:

$$\mathbf{H}_p^{(v)} = \text{MoE}_p^{(v)}(\mathbf{X}^{(v)}) = \sum_{m=1}^{M_p} G_{p,m}^{(v)}(\mathbf{X}^{(v)}) \cdot \mathcal{E}_{p,m}^{(v)}(\mathbf{X}^{(v)}), \tag{3}$$

where, $G_{p,m}^{(v)}$ represents the output of the private gating network for view v. Through this structural design, DMCAR-MVC achieves physical isolation of shared and specific representations during the encoding phase, laying the foundation for subsequent precise alignment and decoupled optimization.

Context-Aware Hierarchical Routing

To address the limitations of traditional MoE routing mechanisms in multi-view scenarios—specifically their "isolated decision-making" and "narrow perspective" issues—we designed a Context-Aware Hierarchical Routing (CAHR) mechanism. This mechanism represents one of the core innovations of our approach, specifically aimed at enhancing the routing capability of the shared expert pool to enable more globally informed decisions, thereby fostering deeper collaboration across views. The implementation of CAHR involves two key steps. First is the construction of a global context vector. For each sample i in the input batch, instead of feeding it directly into the gating network, we first pass it through a lightweight shared encoder ϕ_{light} to extract a preliminary, low-dimensional embedding for each view $\mathbf{x}_i^{(v)}$. These embeddings from all views are then aggregated (in this work, we use average pooling) to construct a compact yet comprehensive global context vector $\mathbf{c}_{\text{global},i}$ for the sample:

$$\mathbf{c}_{\text{global},i} = \frac{1}{V} \sum_{v=1}^{V} \phi_{\text{light}}(\mathbf{x}_i^{(v)}). \tag{4}$$

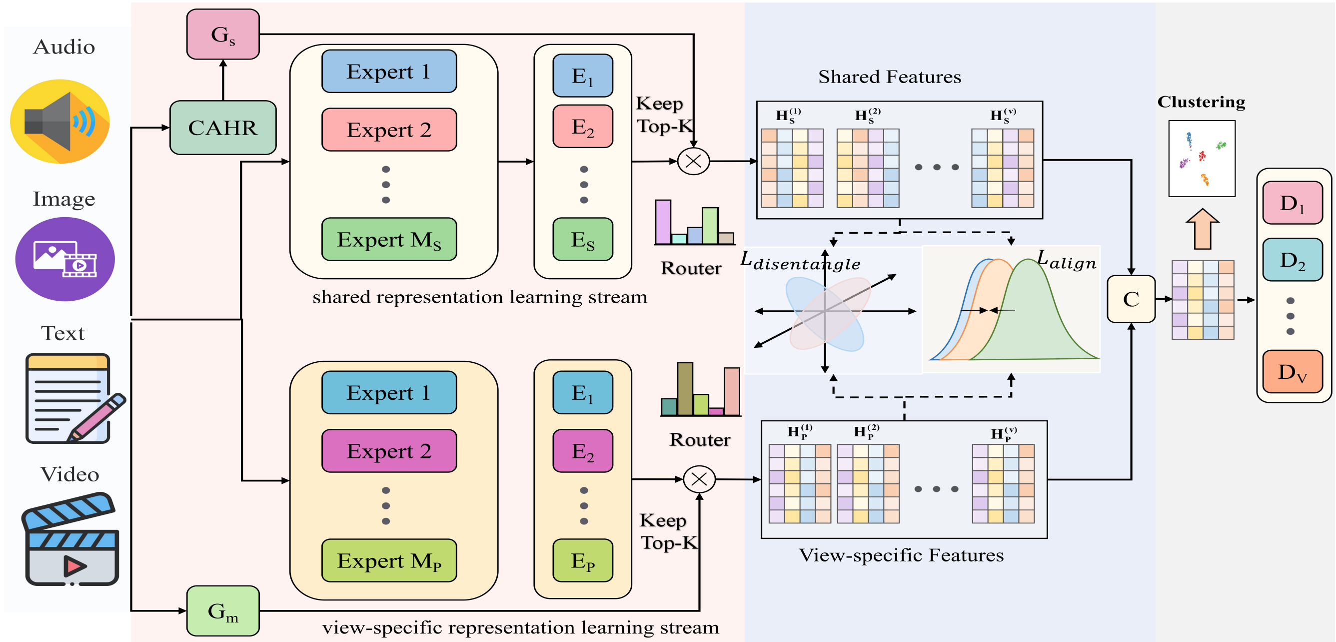


Figure 1: An overview of our proposed DMCAR framework.

This context vector $\mathbf{c}_{\text{global},i}$ can be regarded as a holistic "snapshot" of sample i across all views, providing valuable cross-view information for routing decisions.

Next is the context-aware gating network. When selecting experts from the shared expert pool, its gating network G_s no longer takes as input only the sample representation $\mathbf{x}_i^{(v)}$ from a single view, but rather the concatenation of that representation with its corresponding global context vector $\mathbf{c}_{\text{global},i}$. This approach enables the gating network to simultaneously "see" both the local information of the current view and the global context from other views when making decisions, thereby improving its judgment. Specifically, the routing logits are computed as follows:

$$\text{logits}_i^{(v)} = \mathbf{W}_g \cdot \text{ReLU}([\mathbf{x}_i^{(v)}, \mathbf{c}_{\text{global},i}]), \tag{5}$$

where, $[\cdot,\cdot]$ denotes the concatenation operation, and $\mathbf{W}_g$ is the learnable weight matrix of the gating network. The computed logits are then processed through a standard Softmax function and a Top-K selection mechanism to obtain the final weights $G_{s,m}(\mathbf{x}_i^{(v)}, \mathbf{c}_{\text{global},i})$ assigned to each shared expert. This design enables the expert selection process to transcend the limitations of a single view, achieving more efficient and intelligent cross-view collaboration.

Multi-level Contrastive Learning

To effectively optimize the D-MoE architecture and learn high-quality disentangled representations, we introduce a carefully designed multi-level contrastive learning paradigm comprising two complementary objective functions. The first objective is cross-view contrastive alignment for shared representations ($\mathcal{L}_{\text{align}}$). The core idea is that for the same data sample, its shared representations $\mathbf{h}_{s,i}^{(u)}$ and $\mathbf{h}_{s,i}^{(v)}$ extracted under different views should be semantically similar. We employ the widely used InfoNCE loss to achieve this goal. Specifically, for any sample i, its shared representations under views u and v form a positive pair, while its representations paired with those of any other sample j ($j \neq i$) constitute negative pairs. By maximizing the similarity between positive pairs and minimizing the similarity with all negative pairs, the model is encouraged to learn a view-invariant shared semantic space. The alignment loss for the view pair (u, v) is defined as:

$$\begin{aligned}
\mathcal{L}_{\text{align}}^{(u,v)} &= -\frac{1}{N}\sum_{i=1}^{N} \log \frac{P_i}{Q_i} \\
P_i &= \exp(\text{sim}(\mathbf{h}_{s,i}^{(u)}, \mathbf{h}_{s,i}^{(v)})/\tau) \\
Q_i &= \sum_{j=1, j\neq i}^{N} \exp(\text{sim}(\mathbf{h}_{s,i}^{(u)}, \mathbf{h}_{s,j}^{(v)})/\tau) + P_i
\end{aligned} \tag{6}$$

where, $\text{sim}(\cdot,\cdot)$ denotes cosine similarity, and τ is the temperature hyperparameter. The total alignment loss $\mathcal{L}_{\text{align}}$ is the average of the losses across all view pairs.

The second objective is the contrastive decoupling between shared and specific representations ($\mathcal{L}_{\text{disentangle}}$). To ensure that the private expert pool learns information complementary to, rather than overlapping with, the public expert pool, we explicitly impose an orthogonality constraint to maximize the divergence between the two representations. This is achieved by minimizing the squared cosine similarity between the shared representation $\mathbf{h}_{s,i}^{(v)}$ and the specific representation $\mathbf{h}_{p,i}^{(v)}$ under the same view. This loss function directly penalizes any linear correlation between the two representations, thereby driving them to occupy orthogonal subspaces in the representation space. The formal definition is as follows:

$$\mathcal{L}_{\text{disentangle}} = \frac{1}{V \cdot N} \sum_{v=1}^{V} \sum_{i=1}^{N} \left(\text{sim}(\mathbf{h}_{s,i}^{(v)}, \mathbf{h}_{p,i}^{(v)}) \right)^2. \quad (7)$$

This simple orthogonality loss complements our D-MoE architecture, working together to ensure effective decoupling of the final representations.

Reconstruction and Clustering

After extracting and optimizing disentangled representations through the D-MoE architecture using contrastive learning objectives, we complete the entire process through two steps: reconstruction and clustering. To ensure that the learned latent representations $\mathbf{H}_s^{(v)}$ and $\mathbf{H}_p^{(v)}$ retain sufficient original information and avoid model collapse, we introduce a reconstruction task as a regularization mechanism. Specifically, for each view v, we concatenate the learned shared and specific representations along the feature dimension to form a complete view representation $[\mathbf{H}_s^{(v)}, \mathbf{H}_p^{(v)}]$. This concatenated representation is then fed into a view-specific decoder $\mathcal{D}^{(v)}$ to reconstruct the original input data $\hat{\mathbf{X}}^{(v)} = \mathcal{D}^{(v)}([\mathbf{H}_s^{(v)}, \mathbf{H}_p^{(v)}])$. We employ mean squared error (MSE) to quantify reconstruction quality, with the total reconstruction loss $\mathcal{L}_{\text{recon}}$ being the average of reconstruction errors across all views:

$$\mathcal{L}_{\text{recon}} = \frac{1}{V} \sum_{v=1}^{V} ||\mathbf{X}^{(v)} - \hat{\mathbf{X}}^{(v)}||_F^2. \quad (8)$$

After the model training converges, we proceed to the final clustering stage. First, to obtain a unified shared representation robust across all views, we average the aligned shared representations $\mathbf{H}_s^{(v)}$ from all views, yielding $\mathbf{H}_s = \frac{1}{V} \sum_{v=1}^{V} \mathbf{H}_s^{(v)}$. Next, we concatenate this unified shared representation with the view-specific representations ${\mathbf{H}_p^{(v)}}_{v=1}^{V}$ to construct the final clustering representation $\mathbf{H}_{\text{cluster}}$, which incorporates both consistent and diverse information:

$$\mathbf{H}_{\text{cluster}} = [\mathbf{H}_s, \mathbf{H}_p^{(1)}, \dots, \mathbf{H}_p^{(V)}]. \quad (9)$$

Finally, we apply the standard k-means algorithm to this comprehensive representation $\mathbf{H}_{\text{cluster}}$ to obtain the final sample clustering results. This approach fully leverages the advantages of the disentangled representations learned by our model, thereby achieving superior clustering performance.The pseudocode of our DMCAR-MVC algorithm framework is presented in Algorithm 1.

Experiment

Experiment Setup

Datasets. To comprehensively evaluate the performance of our proposed method, we carefully selected six benchmark datasets widely used in related fields for experimental validation: ALOI-100, HW-6V, UCI-digit, BDGP, CiteSeer, and Cora. Table 1 provides a concise summary of the key statistical information for these datasets,including the number of samples (N), clusters (C), views (V), and feature dimensionality (D).

Algorithm 1: The Proposed DMCAR-MVC Algorithm

while not reaching the maximum training epochs **do**
    For each sample i, construct its global context vector $\mathbf{c}_{\text{global},i}$ using Eq. (4).
    for $v = 1$ to V **do**
        Generate the shared representation $\mathbf{H}_s^{(v)}$ via Eq. (2).
        Generate the view-specific representation $\mathbf{H}_p^{(v)}$ Eq. (3).
        Form $[\mathbf{H}_s^{(v)}, \mathbf{H}_p^{(v)}]$ and reconstruct $\hat{\mathbf{X}}^{(v)}$.
    end for
    Calculate $\mathcal{L}_{\text{align}}$ via Eq. (6).
    Calculate $\mathcal{L}_{\text{disentangle}}$ via Eq. (7).
    Calculate $\mathcal{L}_{\text{recon}}$ via Eq. (8).
    Compute $\mathcal{L}_{\text{total}}$ in Eq. (1).
    Update all network parameters by backpropagating the total loss $\mathcal{L}_{\text{total}}$.
end while
Generate the final clustering representation $\mathbf{H}_{\text{cluster}}$ using Eq. (9).
Apply the k-means algorithm on $\mathbf{H}_{\text{cluster}}$.
return The final cluster assignments.

Dataset	N	C	V	D
HW-6V	2000	10	6	216/76/64/6/240/47
UCI-digit	2000	10	3	64/76/216
ALOI-100	10800	100	4	77/13/64/125
BDGP	2500	5	3	1000/500/250
Cora	2708	7	4	2078/1433/2708/2708
CiteSeer	3312	6	4	3703/3312/3312/3312

Table 1: Multi-view datasets used in experiments.

Implementation Details. Our proposed DMCAR-MVC model was implemented using the PyTorch 1.12 framework(Tejani et al. 2019).All experiments were conducted on a computing platform equipped with the Windows 11 operating system, 64GB of RAM, an Intel Core i9-14900KF CPU, and an NVIDIA GeForce RTX 3090 GPU. For model training, we employed the Adam optimizer(Kingma and Ba 2015) with its default parameter settings and an initial learning rate set to 1×10^{-4}. A batch size of 128 was utilized throughout the training process. The model were trained for a total of 150 epochs.

Baseline Methods. To demonstrate the performance of our proposed DMCAR-MVC method, we compare it with 10 classical and state-of-the-art multi-view clustering methods.Traditional methods: **LMVSC**(Kang et al. 2020),and **BMVC**(Zhang et al. 2019).Deep multi-view clustering methods:**AE2-Nets**(Zhang, Liu, and Fu 2019), **SDMVC**(Xu et al. 2022a),**MFLVC**(Xu et al. 2022b),**DSMVC**(Tang and Liu 2022),**DEMVC**(Ren et al. 2021),**SDSNE**(Liu et al. 2022),**ACCMVC**(Yan et al. 2024),**DMAC**(Wang et al. 2025a).

Dataset	LMVSC	BMVC	AE2-Nets	SDMVC	MFLVC	DSMVC	DEMVC	SDSNE	ACCMVC	DMAC	DMCAR-MVC(Ours)
	ACC(%)										
HW-6V	85.10	75.50	87.82	49.45	86.20	<u>95.25</u>	27.45	87.60	50.25	81.90	**96.80**
UCI	89.35	85.45	80.57	63.00	<u>92.00</u>	85.45	62.30	84.55	39.95	68.65	**92.95**
ALOI	42.40	63.54	26.90	–	23.12	14.89	–	O/M	**86.29**	73.01	<u>80.36</u>
BDGP	48.44	40.08	41.45	34.04	33.12	53.08	25.80	21.96	<u>96.64</u>	84.16	**98.32**
Cora	36.99	30.43	22.15	31.06	31.02	28.88	30.54	<u>44.24</u>	26.22	27.07	**45.86**
CiteSeer	<u>43.62</u>	28.08	38.29	41.33	25.09	35.05	23.31	36.99	25.63	21.53	**66.66**
	NMI(%)										
HW-6V	80.56	77.52	80.18	51.90	85.53	<u>91.19</u>	29.39	87.75	62.60	75.57	**92.15**
UCI	61.07	65.84	69.51	64.29	85.40	80.67	60.54	**89.05**	36.65	59.82	<u>86.51</u>
ALOI	55.82	76.99	49.47	–	67.88	40.57	–	O/M	**93.93**	85.69	<u>92.03</u>
BDGP	23.54	15.57	19.80	8.67	8.56	31.99	5.60	2.37	<u>89.92</u>	76.14	**94.94**
Cora	16.87	9.26	0.79	5.12	12.97	8.14	6.34	**34.35**	13.70	8.69	<u>25.55</u>
CiteSeer	20.50	10.96	16.77	19.00	4.08	9.75	3.05	<u>24.32</u>	4.22	0.81	**49.58**
	Fscore(%)										
HW-6V	77.23	71.13	79.35	42.10	85.82	<u>95.27</u>	17.66	87.83	80.12	82.14	**96.68**
UCI	81.28	78.72	67.65	62.77	<u>92.05</u>	86.09	61.36	82.35	81.54	67.40	**92.95**
ALOI	31.68	50.98	24.59	–	12.02	0.36	–	O/M	23.66	<u>71.51</u>	**79.88**
BDGP	36.42	32.85	33.11	31.47	26.38	53.29	18.73	10.67	81.76	<u>83.13</u>	**98.32**
Cora	26.69	24.64	22.55	17.96	32.41	30.14	27.66	35.13	26.36	<u>26.94</u>	**49.86**
CiteSeer	32.24	23.36	31.12	<u>38.73</u>	25.50	36.62	20.54	23.05	23.02	19.52	**50.97**

Table 2: Clustering algorithms are compared on benchmark datasets using evaluation metrics including ACC, NMI, and Fscore. The best performance is highlighted in bold, while the second-best is shown in <u>underline</u>, “OM” denotes out-of-memory failure.

Comparisons with State of the Arts

Table 2 summarizes the comparative performance of the proposed DMCAR-MVC method against ten baseline algorithms on 6 benchmark datasets with diverse scales,evaluated using accuracy (ACC), normalized mutual information (NMI), and F-score.The results demonstrate that DMCAR-MVC method consistently achieves superior performance compared to the other methods. Specific observations are detailed below:

In nearly all test scenarios, DMCAR-MVC achieved either the best performance (highlighted in bold) or the second-best performance (underlined), significantly outperforming all comparative methods. Particularly on the HW-6V and BDGP datasets, our model set new state-of-the-art records across all three core metrics, strongly demonstrating its robust clustering capabilities. For instance, on the BDGP dataset, DMCAR-MVC reached 98.32% in both ACC and F-score, marking a significant improvement over the second-best method. Notably, our model exhibited exceptional advantages on the Cora and CiteSeer datasets, which are renowned for their complex graph structures. On CiteSeer, DMCAR-MVC achieved a remarkable ACC score of 66.66%, far surpassing the second-best result of 43.62%, underscoring our method’s effectiveness in handling data with intricate inter-view dependencies. Although our model yielded second-best results on a few individual metrics (e.g., ACC on ALOI-100 and NMI on Cora), its overall performance remained highly competitive, with leading results in other metrics.

Ablation Studies

1) w/o D-MoE: This variant degrades our designed Disentangled Mixture-of-Experts (D-MoE) architecture into a flattened expert pool, where all experts are shared across views without distinguishing between common and private experts. 2) w/o CAHR: This variant removes the Context-Aware Hierarchical Routing (CAHR) mechanism, reverting to a traditional routing approach that relies solely on the current view’s information for decision-making. 3) w/o $\mathcal{L}_{\text{align}}$: This variant eliminates the cross-view contrastive alignment loss term during optimization. 4) w/o $\mathcal{L}_{\text{disentangle}}$: This variant removes the disentanglement loss term designed to enhance representation orthogonality.

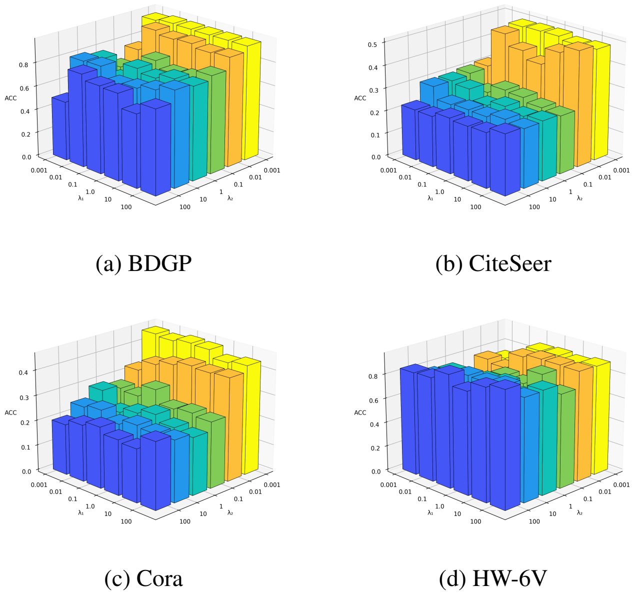
Method	ACC(%)	NMI(%)	Fscore(%)
w/o D-MoE	89.52	83.15	89.33
w/o CAHR	92.17	86.88	92.04
w/o $\mathcal{L}_{\text{align}}$	88.36	81.90	88.15
w/o $\mathcal{L}_{\text{disentangle}}$	94.65	89.04	94.58
DMCAR-MVC (Full)	**96.80**	**92.15**	**96.68**

Table 3: Ablation study results of different components of DMCAR-MVC on the HW-6V dataset. The best performance is indicated in bold.

Method	ACC(%)	NMI(%)	Fscore(%)
w/o D-MoE	91.87	86.44	91.86
w/o CAHR	94.51	89.26	94.51
w/o $\mathcal{L}_{\text{align}}$	90.23	84.19	90.22
w/o $\mathcal{L}_{\text{disentangle}}$	96.14	91.80	96.14
DMCAR-MVC (Full)	**98.32**	**94.94**	**98.32**

Table 4: Ablation study results of different components of DMCAR-MVC on the BDGP dataset. The best performance is indicated in bold.

The experimental results, as shown in Tables 3 and 4, clearly demonstrate the contribution of each component to the model’s final performance. First, the complete DMCAR-MVC model achieved the best performance across all metrics, confirming the superiority of our overall design. When



(a) BDGP (b) CiteSeer

(c) Cora (d) HW-6V

Figure 2: Hyperparameter Sensitivity Analysis of λ_1 and λ_2 on the Cora, BDGP, CiteSeer, and HW-6V Datasets

we removed either the D-MoE architecture (w/o D-MoE) or the alignment loss (w/o $\mathcal{L}_{\text{align}}$), the model exhibited the most significant performance drop. This aligns with our expectations, as D-MoE serves as the structural foundation for representation disentanglement, while $\mathcal{L}_{\text{align}}$ is the core driver ensuring cross-view consistency in shared representations. The absence of either fundamentally undermines the model's primary learning objective.

Second, removing the CAHR mechanism (w/o CAHR) also led to a noticeable performance decline, strongly validating the value of incorporating global contextual information into routing decisions. This result indicates that enabling expert selection to transcend single-view limitations significantly enhances deeper cross-view collaboration.When CAHR is removed, the routing mechanism reverts to a conventional, context-independent decision-making mode. In this mode, the gating network of each view makes judgments solely based on its own inputs, completely disregarding valuable information from other views.

Finally, eliminating the $\mathcal{L}_{\text{disentangle}}$ loss (w/o $\mathcal{L}_{\text{disentangle}}$) similarly resulted in performance degradation, demonstrating that while D-MoE structurally separates information flow, an explicit orthogonality constraint further refines shared and view-specific representations, maximizing their informational divergence. In summary, these ablation study results comprehensively and robustly verify that each proposed component is indispensable. They work synergistically, complementing one another to ultimately achieve the outstanding performance of the DMCAR-MVC framework.

Parameter Sensitivity Analysis

According to Eq. 1, λ_1 and λ_2 control the weights of the cross-view alignment loss $\mathcal{L}_{\text{align}}$ and the disentangling loss $\mathcal{L}_{\text{disentangle}}$, respectively. To analyze their sensitivity and verify the rationality of the D-MoE architecture, we perform a grid search over $0.001, 0.01, 0.1, 1.0, 10, 100$ for both hyperparameters on the Cora, BDGP, CiteSeer, and HW-6V datasets, as shown in Fig. 2.

On Cora, BDGP, and CiteSeer, the clustering accuracy (ACC) is notably sensitive to λ_2: the best or near-best performance is achieved when $\lambda_2 \in \{0.001, 0.01\}$, while larger values lead to a clear degradation. This is consistent with our ablation results and indicates that, given D-MoE already enforces a strong "hard-constraint" disentanglement through the shared/private expert pools, $\mathcal{L}_{\text{disentangle}}$ is more appropriate as a low-weight "soft constraint". An excessively strong orthogonality penalty can instead interfere with the core representations required for reconstruction and alignment.

By contrast, the model exhibits strong robustness to the weight λ_1 of the cross-view alignment loss $\mathcal{L}_{\text{align}}$. On Cora, BDGP, and CiteSeer, ACC varies only mildly when λ_1 changes over a wide range, suggesting that learning consistent shared representations is the main driving force of the framework. Thanks to the synergy between the context-aware routing (CAHR) mechanism and the contrastive learning objective, the model can stably converge to high-quality shared representations as long as a non-zero alignment signal exists (i.e., $\lambda_1 > 0$), without the need for precise tuning of its magnitude. The HW-6V dataset exhibits even higher robustness: performance remains near-optimal across the entire search range of both λ_1 and λ_2, which we attribute to the information redundancy brought by multiple views, thereby reducing the dependence on specific loss-weight settings. Taken together, we adopt $\lambda_1 = 1.0$ and $\lambda_2 = 0.01$ as the default hyperparameter configuration, which empirically confirms the rationality and efficiency of our model design.

Conclusions

In this paper, we propose a deep multi-view clustering framework, DMCAR-MVC, to address the core challenge of decoupling shared representations from view-specific ones. The framework is built upon a Disentangled Mixture-of-Experts (D-MoE) architecture that structurally separates the learning pathways of shared and view-specific information, thereby mitigating representation entanglement common in existing methods. On top of D-MoE, we introduce a Context-Aware Hierarchical Routing (CAHR) mechanism that aggregates a global context vector from all views to guide shared-expert selection, enabling more globally informed and deeper cross-view collaboration. The entire model is trained under a multi-level contrastive learning paradigm, which enforces semantic consistency of shared representations via cross-view alignment while enhancing the separability between shared and specific representations through orthogonality constraints. Extensive experiments on six challenging benchmark datasets show that DMCAR-MVC consistently outperforms state-of-the-art methods across major clustering metrics, and ablation studies confirm the necessity of each component and the superiority of the learned disentangled representations.

Acknowledgments

This work is supported by the National Key R&D Program of China under Grant No.2022ZD0209103, the National Natural Science Foundation of China (project no. 62325604, 62276271, 62406329, 62476280,62476281, 62441618,62506371), National Natural Science Foundation of China Joint Found under Grant No. U24A20323.

References

Chen, B.; Rouditchenko, A.; Duarte, K.; Kuehne, H.; Thomas, S.; Boggust, A.; Panda, R.; Kingsbury, B.; Feris, R.; Harwath, D.; Glass, J.; Picheny, M.; and Chang, S.-F. 2021. Multimodal Clustering Networks for Self-supervised Learning from Unlabeled Videos. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 7992–8001.

Chen, W.; Wang, H.; and Liang, C. 2023. Deep multi-view contrastive learning for cancer subtype identification. *Briefings in Bioinformatics*, 24(5): bbad282–bbad282.

Chen, X.; Zhou, S.; Liang, K.; Yuan, D.; Chen, H.; Sun, X.; Meng, L.; and Liu, X. 2025. Putting on the Thinking Hats: A Survey on Chain of Thought Fine-tuning from the Perspective of Human Reasoning Mechanism. *arXiv preprint arXiv:2510.13170*.

Dai, D.; Deng, C.; Zhao, C.; Xu, R.; Gao, H.; Chen, D.; Li, J.; Zeng, W.; Yu, X.; Wu, Y.; Xie, Z.; Li, Y.; Huang, P.; Luo, F.; Ruan, C.; Sui, Z.; and Liang, W. 2024. Deepseekmoe: Towards ultimate expert specialization in mixture-of-experts language models. *arXiv preprint arXiv:2401.06066*.

Dong, Z.; Hu, D.; Jin, J.; Wang, S.; Liu, X.; and Zhu, E. 2025a. Selective Cross-View Topology for Deep Incomplete Multi-View Clustering. *IEEE Transactions on Image Processing*, 34: 4792–4805.

Dong, Z.; Liu, M.; Wang, S.; Liang, K.; Zhang, Y.; Liu, S.; Jin, J.; Liu, X.; and Zhu, E. 2025b. Enhanced then Progressive Fusion with View Graph for Multi-View Clustering. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15518–15527.

Dong, Z.; Wang, S.; Jin, J.; Liu, X.; and Zhu, E. 2023. Cross-view Topology Based Consistent and Complementary Information for Deep Multi-view Clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 19440–19451.

Du, N.; Huang, Y.; Dai, A. M.; Tong, S.; Lepikhin, D.; Xu, Y.; Krikun, M.; Zhou, Y.; Yu, A. W.; Firat, O.; Zoph, B.; Fedus, L.; Bosma, M. P.; Zhou, Z.; Wang, T.; Wang, E.; Webster, K.; Pellat, M.; Robinson, K.; Meier-Hellstern, K.; Duke, T.; Dixon, L.; Zhang, K.; Le, Q.; Wu, Y.; Chen, Z.; and Cui, C. 2022. Glam: Efficient scaling of language models with mixture-of-experts. 5547–5569.

Du, T.; Zheng, W.; and Xu, X. 2024. Composite attention mechanism network for deep contrastive multi-view clustering. *Neural Networks*, 176: 106361–106361.

Fedus, W.; Zoph, B.; and Shazeer, N. 2022. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120): 1–39.

Feng, Y.; Liang, W.; Wan, X.; Liu, J.; Liu, S.; Qu, Q.; Guan, R.; Xu, H.; and Liu, X. 2025. Incremental Nyström-based Multiple Kernel Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 16613–16621.

Geng, Y.; Han, Z.; Zhang, C.; and Hu, Q. 2021. Uncertainty-aware multi-view representation learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(9): 7545–7553.

Gu, Z.; Feng, S.; Li, Z.; Yuan, J.; and Liu, J. 2024a. NOODLE: Joint Cross-View Discrepancy Discovery and High-Order Correlation Detection for Multi-View Subspace Clustering. *ACM Transactions on Knowledge Discovery from Data*, 18(6): 1–23.

Gu, Z.; Feng, S.; Yuan, J.; and Li, X. 2024b. Consensus representation-driven structured graph learning for multi-view clustering. *Applied Intelligence*, 54(17): 8545–8562.

Guan, R.; Liu, T.; Tu, W.; Tang, C.; Luo, W.; and Liu, X. 2025. Sampling Enhanced Contrastive Multi-View Remote Sensing Data Clustering with Long-Short Range Information Mining. *IEEE Transactions on Knowledge and Data Engineering*, 1–15.

Hu, S.; Fan, J.; Zou, G.; and Ye, Y. 2025. Multi-aspect Self-guided Deep Information Bottleneck for Multi-modal Clustering. 39(16): 17314–17322.

Hu, S.; Zhang, C.; Zou, G.; Lou, Z.; and Ye, Y. 2024. Deep Multiview Clustering by Pseudo-Label Guided Contrastive Learning and Dual Correlation Learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Jacobs, R. A.; Jordan, M. I.; Nowlan, S. J.; and Hinton, G. E. 1991. Adaptive mixtures of local experts. *Neural Computation*, 3(1): 79–87.

Kang, Z.; Zhou, W.; Zhao, Z.; Shao, J.; Han, M.; and Xu, Z. 2020. Large-Scale Multi-View Subspace Clustering in Linear Time. In *national conference on artificial intelligence*, volume 34, 4412–4419. Association for the Advancement of Artificial Intelligence (AAAI).

Ke, G.; Hong, Z.; Zeng, Z.; Liu, Z.; Sun, Y.; and Xie, Y. 2021. CONAN: contrastive fusion networks for multi-view clustering. 653–660.

Ke, G.; Wang, B.; Wang, X.; and He, S. 2024. Rethinking Multi-view Representation Learning via Distilled Disentangling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 26774–26783.

Kingma, D. P.; and Ba, J. L. 2015. Adam: A Method for Stochastic Optimization. In *international conference on learning representations*. Ithaca, NYarXiv.org.

Lepikhin, D.; Lee, H.; Xu, Y.; Chen, D.; Firat, O.; Huang, Y.; Krikun, M.; Shazeer, N.; and Chen, Z. 2020. GShard: Scaling Giant Models with Conditional Computation and Automatic Sharding. *arXiv: Computation and Language*.

Liu, C.; Liao, Z.; Ma, Y.; and Zhan, K. 2022. Stationary diffusion state neural estimation for multiview clustering. volume 36, 7542–7549.

Liu, J.; Liu, X.; Xiong, J.; Liao, Q.; Zhou, S.; Wang, S.; and Yang, Y. 2021a. Optimal Neighborhood Multiple Kernel

Clustering with Adaptive Local Kernels. *IEEE Transactions on Knowledge and Data Engineering*, 1–1.

Liu, X.; Kloft, M.; Guo, X.; He, L.; Liu, J.; and Yang, Y. 2021b. Multiview Subspace Clustering via Co-Training Robust Data Representation. *IEEE Transactions on Neural Networks*, 1–13.

Liu, X.; Zhu, X.; Li, M.; Wang, L.; Tang, C.; Yin, J.; Shen, D.; Wang, H.; and Gao, W. 2019. Late Fusion Incomplete Multi-View Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10): 2410–2423.

Mao, Y.; Yan, X.; Guo, Q.; and Ye, Y. 2021. Deep mutual information maximin for cross-modal clustering. volume 35, 8893–8901.

Mitra, S.; Saha, S.; and Hasanuzzaman, M. 2020. Multi-view clustering for multi-omics data using unified embedding. *Scientific Reports*, 10(1): 13654–13654.

Neeraj, N. K.; and Maurya, V. K. 2020. A REVIEW ON MACHINE LEARNING (FEATURE SELECTION, CLASSIFICATION AND CLUSTERING) APPROACHES OF BIG DATA MINING INDIFFERENT AREA OF RESEARCH. *Journal of Critical Reviews*, 7(19): 2610–2626.

Ren, Y.; Xu, Z.; Pan, L.; Zhu, C.; Xu, J.; and Li, G. 2021. Deep Embedded Multi-view Clustering with Collaborative Training. *Information Sciences*.

Tang, H.; and Liu, Y. 2022. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase. 202–211.

Tejani, A.; Fang, L.; Antiga, L.; Gross, S.; Steiner, B.; Bradbury, J.; Chintala, S.; Desmaison, A.; Massa, F.; Paszke, A.; Gimelshein, N.; Lin, Z.; Raison, M.; Lerer, A.; Yang, E.; Chanan, G.; Killeen, T.; Bai, J.; Chilamkurthy, S.; Kopf, A.; and DeVito, Z. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *neural information processing systems*, 32: 8026–8037.

Wan, X.; Liu, J.; Gan, X.; Liu, X.; Wang, S.; Wen, Y.; Wan, T.; and Zhu, E. 2024. One-Step Multi-View Clustering With Diverse Representation. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13.

Wang, B.; Zeng, C.; Chen, M.; and Li, X. 2025a. Towards Learnable Anchor for Deep Multi-View Clustering. volume 39, 21044–21052.

Wang, Z.; Li, X.; Sun, Y.; Sun, Q.; Sun, Y.; Ling, H.; Dai, J.; and Ren, Z. 2025b. TPCH: tensor-interacted projection and cooperative hashing for multi-view clustering. volume 39, 21420–21428.

Wu, S.; Zheng, Y.; Ren*, Y.; He, J.; Pu, X.; Huang, S.; Hao, Z.; and He, L. 2024. Self-Weighted Contrastive Fusion for Deep Multi-View Clustering. *IEEE Transactions on Multimedia*.

Xu, J.; Ren, Y.; Tang, H.; Yang, Z.; Pan, L.; Pu, X.; Philip, S. Y.; and He, L. 2022a. Self-supervised discriminative feature learning for deep multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*.

Xu, J.; Tang, H.; Ren, Y.; Peng, L.; Zhu, X.; and He, L. 2022b. Multi-level feature learning for contrastive multi-view clustering. 16051–16060.

Yan, W.; Zhang, Y.; Lv, C.; Tang, C.; Yue, G.; Liao, L.; and Lin, W. 2023. Gcfagg: Global and cross-view feature aggregation for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19863–19872.

Yan, W.; Zhang, Y.; Tang, C.; Zhou, W.; and Lin, W. 2024. Anchor-Sharing and Clusterwise Contrastive Network for Multiview Representation Learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Yan, X.; Ye, Y.; Yu, H.; Hu, S.; and Mao, Y. 2021. Deep multi-view learning methods: a review. *Neurocomputing*, 448: 106–129.

Yan, X.; Yin, J.; and Wen, J. 2025. Incomplete Multi-View Multi-label Learning via Disentangled Representation and Label Semantic Embedding. 30722–30731.

Yang, X.; Jiaqi, J.; Wang, S.; Liang, K.; Liu, Y.; Wen, Y.; Liu, S.; Zhou, S.; Liu, X.; and Zhu, E. 2023. Dealmvc: Dual contrastive calibration for multi-view clustering. 337–346.

Yang, X.; Jing, H.; Zhang, Z.; Wang, J.; Niu, H.; Wang, S.; Lu, Y.; Wang, J.; Yin, D.; Liu, X.; et al. 2025. Darec: A disentangled alignment framework for large language model and recommender system. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, 904–917. IEEE.

Zhang, C.; Liu, Y.; and Fu, H. 2019. AE2-Nets: Autoencoder in Autoencoder Networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2577–2585. IEEE.

Zhang, P.; Liu, X.; Xiong, J.; Zhou, S.; Zhao, W.; Zhu, E.; and Cai, Z. 2020. Consensus One-step Multi-view Subspace Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 1–1.

Zhang, Y.; Cai, J.; Wu, Z.; Wang, P.; and Ng, S.-K. 2025. Mixture of Experts as Representation Learner for Deep Multi-View Clustering. volume 39, 22704–22713.

Zhang, Z.; Liu, L.; Shen, F.; Shen, H. T.; and Shao, L. 2019. Binary Multi-View Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7): 1774–1782.

Zhou, P.; Du, L.; Liu, X.; Shen, Y.-D.; Fan, M.; and Li, X. 2021. Self-Paced Clustering Ensemble. *IEEE Transactions on Neural Networks and Learning Systems*, 32(4): 1–15.

Zhou, T.; Dong, Z.; Wang, S.; Liang, K.; Li, M.; Liu, X.; Zhu, E.; and Dong, X. 2025. DPFMVC: Dynamic Progressive Fusion for Multi-view Clustering. In *Proceedings of the 33rd ACM International Conference on Multimedia*, MM '25, 1102–1111. New York, NY, USA: Association for Computing Machinery. ISBN 9798400720352.

Zhu, E.; Liu, X.; Tang, C.; Li, Z.; Zheng, X.; Zhang, W.; and Yue, G. 2021. Consensus Graph Learning for Multi-view Clustering. *IEEE Transactions on Multimedia*, 1–1.

Zhu, T.; Qu, X.; Dong, D.; Ruan, J.; Tong, J.; He, C.; and Cheng, Y. 2024. LLaMA-MoE: Building Mixture-of-Experts from LLaMA with Continual Pre-training. *arXiv preprint arXiv:2406.16554*.