

Enhanced then Progressive Fusion with View Graph for Multi-View Clustering

Zhibin Dong¹, Meng Liu¹, Siwei Wang^{2,*}, Ke Liang¹, Yi Zhang¹, Suyuan Liu¹,
 Jiaqi Jin¹, Xinwang Liu^{1,*}, En Zhu^{1,*}

¹National University of Defense Technology, Changsha, China

²Intelligent Game and Decision Lab, Beijing, China

{dzb20, wangsiwei13, jinjiaqi }@nudt.edu.cn

Abstract

Multi-view clustering aims to improve clustering accuracy by effectively integrating complementary information from multiple perspectives. However, existing methods often encounter challenges such as feature conflicts between views and insufficient enhancement of individual view features, which hinder clustering performance. To address these challenges, we propose a novel framework, EPFMVC, which integrates feature enhancement with progressive fusion to more effectively align multi-view data. Specifically, we introduce two key innovations: (1) a Feature Channel Attention Encoder (FCAencoder), which adaptively enhances the most discriminative features in each view, and (2) a View Graph-based Progressive Fusion Mechanism, which constructs a view graph using optimal transport (OT) distance to progressively fuse similar views while minimizing inter-view conflicts. By leveraging multi-head attention, the fusion process gradually integrates complementary information, ensuring more consistent and robust shared representations. These innovations enable superior representation learning and effective fusion across views. Extensive experiments demonstrate that our method significantly outperforms state-of-the-art techniques, achieving notable improvements in multi-view clustering tasks across various datasets and evaluation metrics.

1. Introduction

In recent years, multi-view clustering[1, 7, 10, 24, 28, 35] has attracted significant attention across a wide range of domains, including computer vision[41], natural language processing[16], bioinformatics[23], and social network analysis[8]. The central goal of multi-view clustering is to effectively integrate information from different views to improve clustering accuracy. Unlike single-view methods, which rely on a single representation of the

data, multi-view clustering leverages complementary information from multiple perspectives[47], enhancing the ability to capture the underlying structure and diversity within the data. With the rise of deep learning, deep multi-view clustering[8, 21, 26, 38, 48, 49] has emerged as a powerful framework capable of learning complex, high-dimensional representations. These methods exploit shared patterns across views, enabling more effective fusion of information and yielding improved clustering results.

Deep multi-view clustering approaches can be broadly categorized into three primary methodologies: reconstruction-based methods [11, 34, 39, 42, 46], graph-based methods [6, 14, 18, 33, 37], and subspace-based methods [2, 27, 53]. Reconstruction-based methods typically utilize autoencoders to learn a shared latent space that enables the reconstruction of the original data or its representations [32]. These methods focus on aligning the different views by learning a common feature representation, thereby enabling effective clustering. Graph-based methods, on the other hand, construct topological representations of the data by defining relationships between samples, often through similarity matrices or graphs. Leveraging graph neural networks (GNNs) or spectral clustering techniques, these methods aim to learn shared representations across views while preserving the inherent structural and proximity information during the clustering process [40]. Lastly, subspace-based methods focus on discovering a common subspace across views in which the data points from different perspectives can be aligned. Techniques such as canonical correlation analysis (CCA) and subspace clustering are often employed to minimize view-specific biases, ultimately enhancing clustering performance [36]. By utilizing deep neural networks, these methods capture complex, non-linear relationships between views, improving both data representation and clustering effectiveness.

Despite the advancements made by these approaches, several challenges persist in deep multi-view clustering. A major issue lies in the direct fusion of information from

*Corresponding author

multiple views, which can often lead to conflicts, especially when views exhibit divergent feature distributions or discrepancies in sample relationships [12]. Without a careful strategy for addressing these conflicts, naive fusion can result in suboptimal clustering outcomes. Furthermore, many existing methods neglect the crucial step of enhancing the feature channels of individual views prior to fusion. Selective enhancement of feature channels is pivotal for improving the separability of data, as certain features may be more relevant to the clustering task than others. When feature enhancement is inadequately performed, the learned representations risk losing crucial discriminative information, thereby degrading the overall clustering performance. To address these challenges, we propose a novel framework that integrates feature enhancement and progressive fusion to tackle both intra-view and inter-view issues. As illustrated in Figure 1, our approach begins by enhancing the feature channels of each view using the SENet architecture [9]. This step strengthens prominent features while suppressing weaker ones, allowing for a more informative representation of each view. Following feature enhancement, we apply a multi-head self-attention mechanism [30] to map the enhanced features into a shared latent space, generating the latent representations for each view. In this latent space, we compute the optimal transport (OT) distance between the view representations to construct a view graph. This graph serves as the foundation for multi-head attention fusion, which aggregates similar views based on the view graph to obtain a unified, shared representation across all views. Additionally, we introduce two key optimization modules to further refine the clustering process: Intra-view Similarity Enhancement and Cross-view Feature Alignment. The Intra-view Similarity Enhancement module improves intra-cluster similarity in the latent space using contrastive learning, thereby ensuring that clusters are more compact and distinct. Meanwhile, the Cross-view Feature Alignment module aligns the latent representation of each view with the shared representation, ensuring consistency and coherence across all views. This dual optimization ensures that both intra-view relationships and inter-view consistency are effectively preserved, leading to improved clustering performance.

Our contributions can be summarized in three key aspects:

- We introduce the Feature Channel Attention Encoder (FCAencoder), a novel approach that enhances specific feature channels and employs multi-head self-attention to encode the latent space.
- We propose a pioneering progressive fusion mechanism based on a view graph, which constructs a view graph in the latent space and progressively fuses similar view representations. This novel mechanism effectively addresses the issue of feature conflicts, which often arise when di-

rectly fusing views with differing characteristics.

- Extensive experimental results validate the effectiveness of our method, demonstrating substantial improvements in clustering accuracy across various datasets, surpassing the performance of existing state-of-the-art techniques.

2. Related Work

2.1. Deep Multi-View Clustering

In recent years, deep multi-view clustering has gained increasing attention within the machine learning community. These approaches leverage the powerful representation learning capabilities of deep learning to learn shared representations in a latent space [13, 19, 39]. Graph-based methods often employ the topological structure of the samples to guide the learning of precise representations. For instance, Ren et al. [25] captured the topological information between samples by learning a dynamic weighted graph across views, enabling accurate clustering. Subspace-based methods focus on leveraging the consistency of multi-view data in subspace representations while adding constraints to achieve consistent representations across views. Chen et al. [1] proposed a fast self-guided multi-view subspace clustering algorithm that incorporates constraints within the subspace to build a hierarchical learning approach, from global to local, thus learning both consistency and complementarity across views. Reconstruction-based methods aim to learn latent representations for multiple views by reconstructing their features or topological structures. For example, Xu et al. [42] employed a self-supervised approach to obtain pseudo-labels by reconstructing view features, thus building a unified target distribution for multi-view discriminative feature learning. However, most of these methods fail to enhance relevant features directly from the original feature channels or address the feature gap that may exist during the fusion of multiple views. Instead, they often employ direct fusion strategies without considering the inherent discrepancies between views, which may hinder the learning of robust clustering representations.

2.2. Contrastive Learning

Recent years have witnessed the increasing adoption of contrastive learning in the domain of multi-view representation learning [3, 5, 20, 22, 31, 52]. Most of these approaches leverage contrastive learning to maximize the mutual information between different views, thereby capturing high-quality clustering representations. Such strategies have demonstrated notable effectiveness in aligning multi-view data. For instance, Xu et al. [43] exploit contrastive learning to achieve two consistency objectives, effectively mining the shared semantics across all views and yielding improved clustering performance. Similarly, Cui et al. [4] introduce a dual contrastive loss scheme: a dynamic clus-

tering diffusion loss that maximizes inter-cluster distance, coupled with a reliable neighbor-guided positive alignment loss to enhance intra-cluster compactness. Lu et al. [21] propose a decoupled contrastive multi-view clustering approach based on higher-order random walks, performing inter-view and intra-view contrastive learning in distinct embedding spaces to derive robust clustering representations.

3. Method

3.1. Preliminaries

In this section, we present our multi-view clustering framework, which effectively integrates feature representation and robust clustering through four key components: Feature Channel Attention Encoding Module, Progressive Fusion based on View Graph, Intra-View Similar Feature Enhancement Module, and Cross-View Feature Alignment Module. We represent the multi-view data as $\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(V)}\}$, where $\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$ represents the feature matrix of view v , containing n samples and $d^{(v)}$ dimensions, for $v = 1, 2, \dots, V$. Each component uniquely contributes to enhancing information exchange, reducing redundancy, and maximizing mutual information across views.

3.2. Feature Channel Attention Encoding Module

First, our framework employs the Squeeze-and-Excitation[9] (SE) Module to enhance the importance of specific feature channels in each view. Given an input feature matrix from view v , $\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$, the SE module extracts channel-wise statistics through global average pooling:

$$s_c^{(v)} = \frac{1}{n} \sum_{i=1}^n X_{i,c}^{(v)}, \quad c = 1, 2, \dots, d^{(v)}. \quad (1)$$

The channel-wise descriptors are then passed through two fully connected layers, using ReLU and Sigmoid activation functions to obtain channel weights:

$$\mathbf{u}^{(v)} = \sigma(\delta(s^{(v)} \cdot \mathbf{W}_1^{(v)}) \cdot \mathbf{W}_2^{(v)}), \quad (2)$$

where $\mathbf{W}_1^{(v)} \in \mathbb{R}^{d^{(v)} \times d^{(v)}/r}$ and $\mathbf{W}_2^{(v)} \in \mathbb{R}^{d^{(v)}/r \times d^{(v)}}$ are learnable parameters, δ represents the ReLU activation, and σ is the Sigmoid activation function. r is the scaling factor. Finally, the recalibrated feature representation is obtained by scaling the input features with the computed channel weights:

$$\mathbf{X}^{(v)'} = \mathbf{X}^{(v)} \odot \mathbf{u}^{(v)}, \quad (3)$$

where $\mathbf{u}^{(v)}$ is broadcasted across all samples. By selectively emphasizing the useful channels, the SE module effectively reduces redundancy while retaining valuable information. After enhancing the feature channels using the

Squeeze-and-Excitation (SE) mechanism, the feature matrix of each view $\mathbf{X}^{(v)}$ is transformed into $\mathbf{X}^{(v)'}$. To capture the intra-view dependencies and facilitate meaningful information exchange, we apply a multi-head self-attention mechanism to each view-specific representation $\mathbf{X}^{(v)'}$. This allows for a better understanding of feature relations within each view. For a given view v , the attention mechanism starts by calculating unique query $\mathbf{Q}^{(v)} = \mathbf{X}^{(v)'} \mathbf{W}_Q^{(v)}$, key $\mathbf{K}^{(v)} = \mathbf{X}^{(v)'} \mathbf{W}_K^{(v)}$, and value $\mathbf{V}^{(v)} = \mathbf{X}^{(v)'} \mathbf{W}_V^{(v)}$ matrices. Where $\mathbf{W}_Q^{(v)}$, $\mathbf{W}_K^{(v)}$, $\mathbf{W}_V^{(v)}$ are learnable parameter matrices for the view v . d_k represents the reduced dimensionality for the self-attention operation. After computing the query, key, and value matrices, the attention scores are calculated to quantify the relationship between different features within the same view. The attention score is computed by taking the dot product between the query and key matrices, followed by a softmax operation to normalize the scores:

$$\text{Attention}(\mathbf{Q}^{(v)}, \mathbf{K}^{(v)}) = \text{softmax} \left(\frac{\mathbf{Q}^{(v)} \mathbf{K}^{(v)T}}{\sqrt{d_k}} \right), \quad (4)$$

where the denominator $\sqrt{d_k}$ is a scaling factor that prevents the dot products from becoming excessively large, thus stabilizing the gradient during training. The resulting attention scores are then used to weight the value matrix:

$$\mathbf{A}^{(v)} = \text{Attention}(\mathbf{Q}^{(v)}, \mathbf{K}^{(v)}) \mathbf{V}^{(v)}, \quad (5)$$

where $\mathbf{A}^{(v)} \in \mathbb{R}^{n \times d_k}$ represents the attention-weighted feature representation of view v . The self-attention mechanism is extended by applying multiple heads to learn different aspects of the feature interactions. The outputs of all heads are concatenated to form the multi-head representation:

$$\text{MultiHead}(\mathbf{X}^{(v)'}) = \text{Concat}(\mathbf{A}_1^{(v)}, \mathbf{A}_2^{(v)}, \dots, \mathbf{A}_h^{(v)}) \mathbf{W}_{\text{out}}, \quad (6)$$

where $\mathbf{A}_i^{(v)}$ represents the output of the i -th attention head, and $\mathbf{W}_{\text{out}}$ is a learnable projection matrix that ensures the final output has the same dimensionality as the original feature representation. To maintain gradient flow and prevent vanishing or exploding gradients, we employ a residual connection that adds the original input $\mathbf{X}^{(v)'}$ to the output of the multi-head attention layer, followed by layer normalization:

$$\mathbf{Z}_{\text{attn}}^{(v)} = \text{LayerNorm}(\text{MultiHead}(\mathbf{X}^{(v)'}) + \mathbf{X}^{(v)'}). \quad (7)$$

The output $\mathbf{Z}_{\text{attn}}^{(v)}$ is then passed through a feed-forward network (FFN) to introduce further non-linearity and enhance the feature representation:

$$\mathbf{Z}^{(v)} = \text{LayerNorm}(\text{FFN}(\mathbf{Z}_{\text{attn}}^{(v)}) + \mathbf{Z}_{\text{attn}}^{(v)}). \quad (8)$$

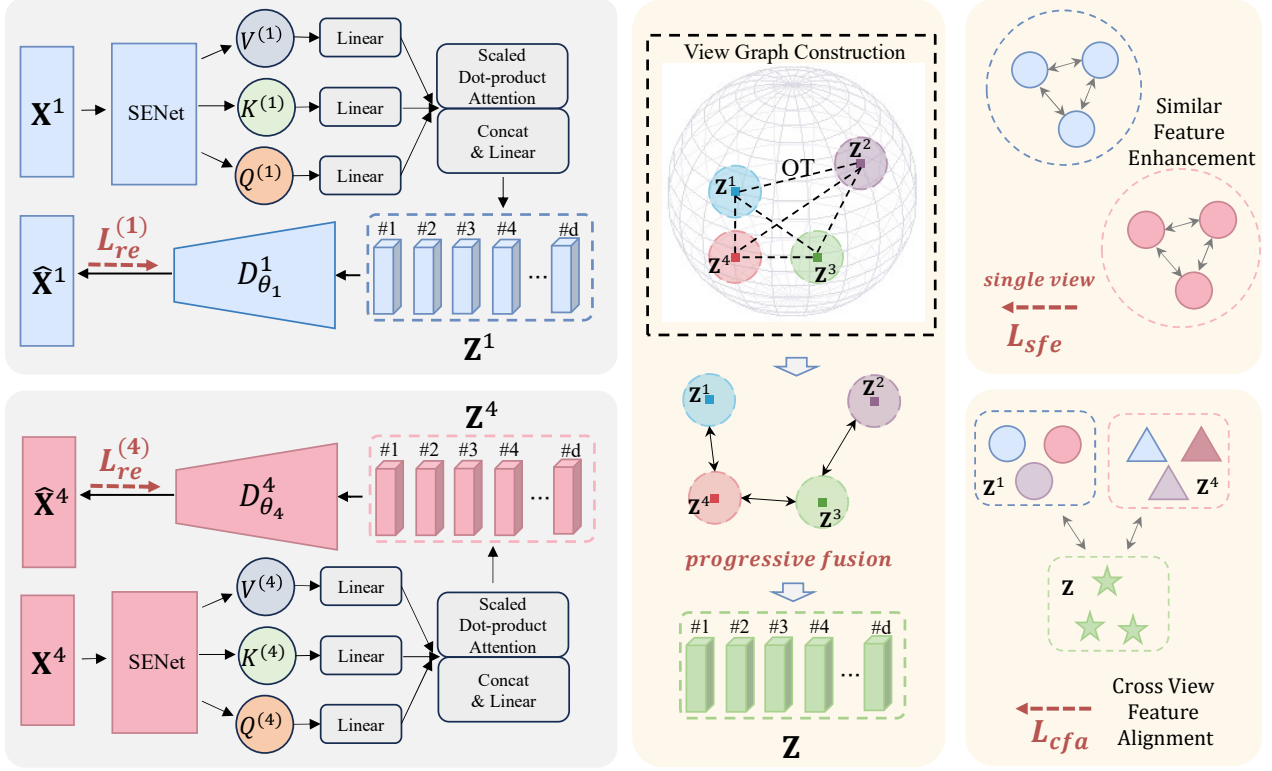


Figure 1. Framework of EPFMVC. It consists of four modules: Feature Channel Attention Encoding Module, Progressive Fusion based on View Graph, Intra-View Similar Feature Enhancement Module, and Cross-View Feature Alignment Module.

The final output representation for each view, denoted as $\mathbf{Z}^{(v)}$, is obtained through a sequence of attention-based operations, layer normalization, residual connections, and feed-forward transformations. Then we use the decoder to reconstruct $\mathbf{Z}^{(v)}$ on each view to obtain $\mathbf{X}^{(v)}$, and the reconstruction loss before and after reconstruction is $\mathcal{L}_{rec}$.

3.3. Progressive Fusion based on View Graph

After obtaining the individual view-specific representations $\mathbf{Z}^{(v)}$ for each view v , we leverage a Cross-View Attention mechanism to progressively fuse the information from multiple views. We compute pairwise similarities between views using Optimal Transport (OT) distance to build the graph. For views v and u , the OT distance is defined as:

$$d(\mathbf{Z}^{(v)}, \mathbf{Z}^{(u)}) = \text{OT}(\mathbf{Z}^{(v)}, \mathbf{Z}^{(u)}), \quad (9)$$

where $\text{OT}(\mathbf{Z}^{(v)}, \mathbf{Z}^{(u)})$ represents the optimal transport distance between views v and u , which measures the discrepancy between their feature representations. We store these pairwise distances in a distance matrix $\mathbf{D}$ of size $V \times V$, where V is the number of views. This distance matrix helps in identifying the most similar views for progressive fusion.

For each view-specific representation $\mathbf{Z}^{(v)}$, we compute unique Query $\mathbf{Q}_z^{(v)} = \mathbf{Z}^{(v)} \mathbf{W}_{qz}^{(v)}$, Key $\mathbf{K}_z^{(v)} = \mathbf{Z}^{(v)} \mathbf{W}_{kz}^{(v)}$, and Value $\mathbf{V}_z^{(v)} = \mathbf{Z}^{(v)} \mathbf{W}_{vz}^{(v)}$ projections to capture the attention-specific features for each view. These matrices transform the original representations $\mathbf{Z}^{(v)}$ into different spaces suitable for calculating attention scores and aggregating information. Using the distance matrix $\mathbf{D}$, we progressively fuse the most similar views by identifying, for each view v , the closest view u :

$$u = \arg \min_{k \neq v} d(\mathbf{Z}^{(v)}, \mathbf{Z}^{(k)}). \quad (10)$$

This process ensures that the fusion is conducted between pairs of views that have the highest similarity, thereby maximizing the consistency of the fused representation while minimizing potential conflicts. Once the closest views are identified, we calculate the cross-view attention between the current view $\mathbf{Z}^{(v)}$ and its closest view $\mathbf{Z}^{(u)}$. The resulting attention scores are then used to compute the weighted combination of the value representation:

$$\mathbf{C}^{(v)} = \text{Attention}(\mathbf{Q}_z^{(v)}, \mathbf{K}_z^{(u)}) \mathbf{V}_z^{(u)}, \quad (11)$$

where $\mathbf{C}^{(v)}$ is the cross-view attention output for view v , effectively capturing the most relevant information from its

closest view u . After computing the cross-view attention outputs $\mathbf{C}^{(v)}$ for each view v , we concatenate these outputs to form a fused representation:

$$\mathbf{C} = \text{Concat}(\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \dots, \mathbf{C}^{(V)}), \quad (12)$$

where $\mathbf{C}$ represents the concatenated representation from all views. This concatenated representation is then passed through a fully connected layer to project it back to the original dimensionality:

$$\mathbf{Z} = \mathbf{W}_{\text{fuse}} \mathbf{C}, \quad (13)$$

where $\mathbf{W}_{\text{fuse}}$ is a learnable parameter. The final output representation $\mathbf{Z}$ is obtained through a stepwise progressive fusion of the view-specific representations $\mathbf{Z}^{(v)}$ using cross-view attention.

3.4. Intra-View Similar Feature Enhancement Module

After obtaining the view-specific representations $\mathbf{Z}^{(v)}$ for each view v , the next step is to further enhance the feature representations within each individual view. This enhancement is achieved through a contrastive learning approach, where similar samples within a view are considered as positive pairs, and their representations are brought closer in the feature space.

We define a set of positive sample pairs $\mathcal{P}^{(v)}$ for each view by using k-nearest neighbors (KNN) to find the top k most similar samples for each instance. The set of positive pairs is denoted by:

$$\mathcal{P}^{(v)} = \{(i, j) \mid i \in [1, n], j \in \text{KNN}(i, k)\}. \quad (14)$$

To enhance the similarity of positive sample pairs within each view, we adopt a contrastive learning objective. Specifically, for each view v , we aim to minimize the distance between the representations of positive pairs $(i, j) \in \mathcal{P}^{(v)}$. The intra-view contrastive loss $\ell_{sfe}^{(v)}$ for view v is defined as:

$$\ell_{sfe}^{(v)} = -\frac{1}{|\mathcal{P}^{(v)}|} \sum_{(i,j) \in \mathcal{P}^{(v)}} \log \frac{\exp(\text{Sim}(\mathbf{z}_i^{(v)}, \mathbf{z}_j^{(v)})/\tau_s)}{\sum_{u=1}^n \exp(\text{Sim}(\mathbf{z}_i^{(v)}, \mathbf{z}_u^{(v)})/\tau_s)}, \quad (15)$$

where τ_s is a temperature hyperparameter that controls the sharpness of the similarity distribution. $\text{Sim}(\mathbf{z}_i^{(v)}, \mathbf{z}_j^{(v)})$ is a similarity function, typically the cosine similarity, defined as:

$$\text{Sim}(\mathbf{z}_i^{(v)}, \mathbf{z}_j^{(v)}) = \frac{\mathbf{z}_i^{(v)} \cdot \mathbf{z}_j^{(v)}}{\|\mathbf{z}_i^{(v)}\| \|\mathbf{z}_j^{(v)}\|}. \quad (16)$$

The overall intra-view similar feature enhancement loss ℓ_{sfe} is given by:

$$\mathcal{L}_{sfe} = \sum_{v=1}^V \ell_{sfe}^{(v)}. \quad (17)$$

This contrastive loss aims to maximize the similarity between positive sample pairs while ensuring that the similarity between non-positive pairs remains relatively lower. The intra-view similar feature enhancement module employs contrastive learning to ensure that the representations within each view are consistent.

3.5. Cross-View Feature Alignment Module

The final step in our multi-view clustering framework is to align the feature representations obtained from each view $\mathbf{Z}^{(v)}$ with the progressively fused common representation $\mathbf{Z}$ to ensure consistency across different views. To align the representations across different views, we employ a contrastive learning objective that minimizes the distance between the representations of positive pairs across views. Specifically, the overall cross-view alignment loss can be expressed as follows,

$$\mathcal{L}_{cfa} = -\frac{1}{n} \sum_{v=1}^V \sum_{i=1}^n \log \frac{\exp(\text{Sim}(\mathbf{z}_i^{(v)}, \mathbf{z}_i)/\tau_c)}{\sum_{u=1}^n \exp(\text{Sim}(\mathbf{z}_i^{(v)}, \mathbf{z}_u)/\tau_c)}. \quad (18)$$

This alignment process encourages the representation of each view to align with the common representation obtained through progressive fusion, thereby reducing discrepancies between the views and ensuring that similar samples are represented similarly across views. This loss enhances cross-view consistency and reduces feature gaps between views. The overall loss function for the entire network can be expressed as follows:

$$\mathcal{L} = \mathcal{L}_{rec} + \gamma \mathcal{L}_{sfe} + \varepsilon \mathcal{L}_{cfa}, \quad (19)$$

where γ and ε represent the regularization coefficients. Our goal is to optimize the entire network by minimizing the loss $\mathcal{L}$. The detailed procedure of the algorithm is outlined in Algorithm 1.

4. Experiment

4.1. Datasets and Experimental Setting

In this section, we evaluate the performance of our proposed model, EPFMVC, against nine state-of-the-art clustering algorithms on seven benchmark datasets. Table 1 provides specific information about these datasets, including the number of samples, clusters, views, and feature dimensions per view. Our model uses two hyperparameters γ and ε , both ranging between [0.001, 1000], and a learning rate of 0.001. We implemented the model in PyTorch 1.7.1 on a desktop with an NVIDIA GeForce RTX 3080 GPU and 64GB RAM, utilizing the Adam optimizer with default settings.

Algorithm 1: The proposed EPFMVC

Input: $\{\mathbf{X}^{(i)}\}_{i=1}^V \in \mathbb{R}^{n \times d^{(v)}}$, number of clusters k , learning rate α_t
Output: Clustering results.

```

// Feature channel attention coding learning
1 for  $i = 1 \rightarrow V$  do
2   Calculate the Enhanced feature matrix  $\mathbf{X}^{(v)'}$  with Eq. (3);
3   Calculate the representation  $\mathbf{Z}^{(i)}$  with Eq. (8);
4 Obtain the view graph  $\mathbf{D}$  by optimizing Eq.(9);
  // Consistent feature learning
5 while not converged do
6   Update representation  $\mathbf{Z}^{(i)}$  by solving Eq. (19);
7   Update view graph  $\mathbf{D}$  by calculating Eq.(9);
8   Obtain the nearest view pair based on the view graph by solving Eq.(10);
9   Fuse the two adjacent views by calculating Eq.(11);
10  Calculate common representation  $\mathbf{Z}$  with Eq. (13);
11 Perform  $k$ -means on the common representation  $\mathbf{Z}$ .
```

Table 1. Multi-view datasets used in experiments.

Dataset	Samples	Clusters	Views	Dimensionality
NGs	500	5	3	2000/2000/2000
BBC	685	5	4	4659/4633/4665/4684
HandWritten	2000	10	6	216/76/64/6/240/47
UCI-digit	2000	10	3	64/76/216
Hdigit	10000	10	2	784/256
Wiki	2866	10	2	128/10
CIFAR10-All	60000	10	7	3072/5376/512/5376/5376/1239/5376

4.2. Compared Methods and Main Results

We compare our proposed method, EPFMVC, against nine state-of-the-art multi-view clustering algorithms on seven benchmark datasets. They are **BMVC**[51], **LMVSC**[15], **AE2-Nets**[50], **AdaGAE** [17], **SD-MVC**[44], **MFLVC**[43], **DSMVC** [29], **SEM** [45] and **MVCAN** [46].

And we evaluated the performance of EPFMVC against nine baseline algorithms using four widely adopted clustering metrics: ACC (Clustering Accuracy), NMI (Normalized Mutual Information), Purity, and F-Score. The detailed performance results are presented in Table 2. Based on these results, we draw the following conclusions: (1) Our proposed method outperforms the nine competing methods on all evaluation metrics across all datasets. Notably, on the BBC, UCI, and CIFAR10-All datasets, EPFMVC achieves significant performance improvements of 11.97%,

3.25%, and 4.5% over the second-best method, respectively. These results highlight the effectiveness of our feature enhancement and progressive fusion modules. (2) For multi-view datasets, particularly BBC and UCI-Digit, our method demonstrates substantial gains over other baselines. This improvement is primarily attributed to the progressive fusion module, which effectively reduces feature redundancy across multiple views. The benefit of our method is especially evident in the NMI metric, which shows the improved quality of cluster assignment consistency. Additionally, on the large-scale CIFAR10-All dataset, EPFMVC consistently outperforms the nine competing algorithms, demonstrating its scalability and robustness in handling large datasets.

4.3. Ablation Study of Enhancement and Alignment Module

To assess the contributions of the Feature Enhancement Module and the Cross-View Alignment Module, we conducted ablation studies on the NGS and BBC datasets, with detailed results provided in Table 3. The ablation results clearly demonstrate the importance of both modules in enhancing clustering performance. We observed that incorporating the feature enhancement module significantly improves the quality of the learned representations, leading to better clustering results. Similarly, the cross-view alignment module plays a critical role in reducing the discrepancies between views, further boosting the clustering performance. Notably, when both modules are combined, the model achieves substantial improvements over variants that use either module alone or rely solely on the reconstruction objective. This highlights the synergistic effect of enhancing features and aligning them across views, ultimately resulting in more robust and discriminative clustering representations.

4.4. Ablation Study of FCAencoder Module

To validate the effectiveness of the Feature Channel Attention Encoding Module, we compared the Feature Channel Attention Encoder (FCAencoder) with a standard autoencoder under the same framework. We evaluated the clustering performance of both encoders across seven datasets and using four different metrics. The specific results are shown in Table 4. The results demonstrate that the representations learned by the FCAencoder consistently outperform those learned by the autoencoder across all datasets. Notably, on the high-dimensional datasets NGs, BBC, and CIFAR10-All, the FCAencoder achieves substantial improvements in clustering accuracy, with gains of 47.8%, 6.43%, and 6.29%, respectively. These results strongly indicate the effectiveness of the feature channel attention mechanism in enhancing the clustering performance by focusing on beneficial features while suppressing irrele-

Table 2. A comparison of clustering algorithms on benchmark datasets is provided, including metrics such as ACC, NMI, Purity, and F-score. The highest-performing result is indicated in **bold**, while the second best is underlined. ‘O/M’ denotes the out-of-memory failure.

Datasets	Conventional Methods		Deep Methods							
	BMVC	LMVSC	AE ² -Nets	AdaGAE	SDMVC	MFLVC	DSMVC	SEM	MVCAN	EPFMVC
ACC(%)										
NGS	56.80	39.14	33.00	49.40	43.20	84.80	43.40	<u>94.00</u>	23.80	95.00
BBC	<u>75.04</u>	52.41	29.34	41.02	62.48	49.34	42.77	68.61	34.45	87.01
UCI-digit	85.45	89.35	80.57	45.25	63.00	<u>92.00</u>	85.45	91.90	89.15	95.25
HW	75.50	85.10	87.82	51.35	49.45	86.20	95.25	86.55	<u>96.10</u>	96.90
Hdigit	79.11	97.09	94.21	56.77	38.65	99.00	99.06	<u>99.83</u>	93.83	99.88
Wiki	45.71	57.80	45.33	34.65	24.08	48.08	<u>62.53</u>	52.23	55.23	63.19
CIFAR10-All	31.00	O/M	23.27	O/M	O/M	O/M	20.77	31.50	<u>32.44</u>	36.94
NMI(%)										
NGS	40.02	13.22	10.68	21.14	19.20	68.88	12.59	<u>83.33</u>	5.33	84.97
BBC	49.26	42.99	1.73	17.80	54.96	43.14	13.96	<u>61.26</u>	2.92	67.01
UCI-digit	82.52	83.27	69.51	47.77	64.29	<u>85.40</u>	80.67	85.31	82.70	90.31
HW	77.52	80.56	80.18	48.11	51.90	85.53	91.19	86.99	<u>91.46</u>	93.15
Hdigit	77.80	92.93	86.96	64.66	37.08	98.56	97.36	<u>99.42</u>	<u>86.75</u>	99.60
Wiki	39.85	54.20	43.04	30.87	11.78	41.88	<u>54.60</u>	51.01	53.44	55.10
CIFAR10-All	20.13	O/M	10.69	O/M	O/M	O/M	8.57	<u>22.26</u>	21.63	24.79
Purity(%)										
NGS	56.80	39.14	34.20	49.40	43.80	84.80	43.40	<u>94.00</u>	24.00	95.00
BBC	75.04	56.20	35.69	47.45	75.47	61.46	47.59	<u>79.56</u>	34.74	87.01
UCI-digit	85.45	89.35	80.57	49.15	67.05	<u>92.00</u>	85.40	91.90	89.15	95.25
HW	79.45	85.10	87.82	52.90	49.85	86.20	95.25	86.55	<u>96.10</u>	96.90
Hdigit	82.73	97.09	94.21	60.93	42.17	98.82	99.06	<u>99.83</u>	93.83	99.88
Wiki	48.08	63.04	52.41	40.65	27.11	50.87	63.33	59.42	<u>63.43</u>	65.25
CIFAR10-All	32.17	O/M	24.07	O/M	O/M	O/M	21.80	<u>34.55</u>	34.31	37.84
F-score(%)										
NGS	46.53	27.89	35.15	49.09	33.11	84.56	42.86	<u>93.98</u>	13.94	94.98
BBC	62.62	44.02	27.03	42.43	65.55	53.32	46.34	<u>70.38</u>	20.25	87.19
UCI-digit	78.72	81.28	67.65	45.80	62.77	<u>92.05</u>	86.09	91.93	89.27	95.26
HW	71.13	77.22	79.35	52.97	42.10	85.82	95.27	84.88	<u>96.11</u>	96.90
Hdigit	75.66	94.32	88.99	54.74	35.78	98.82	99.06	<u>99.83</u>	93.83	99.88
Wiki	38.79	50.11	39.90	34.22	24.09	4.48	<u>61.35</u>	<u>54.06</u>	55.02	62.74
CIFAR10-All	20.89	O/M	16.23	O/M	O/M	O/M	20.69	31.08	<u>32.24</u>	36.86

Table 3. Ablation study on NGs. ✓ denotes EPFMVC with the component.

Components			Metrics(%)			
$\mathcal{L}_{rec}$	$\mathcal{L}_{sfe}$	$\mathcal{L}_{cfa}$	ACC	NMI	Purity	F-score
✓			38.60	15.80	40.20	37.14
✓	✓		73.20	46.83	73.20	73.52
✓		✓	92.60	79.78	92.60	92.62
✓	✓	✓	95.00	84.97	95.00	94.98

vant ones, thereby leading to more discriminative representations for downstream tasks.

4.5. Ablation of Progresssive Fusion

To validate the effectiveness of the Progressive Fusion Mechanism based on View Graph, we compared its clustering performance with that of a Direct Concatenation Fusion approach within our proposed framework. The results are presented in Figure 2. Specifically, we conducted experiments on three datasets—BBC, HW, and CIFAR10-All—to compare the two fusion strategies. The results indicate

Table 4. FCAencoder comparative experiments are conducted on 7 datasets.

Metrics(%)	Strategies	Datasets						
		NGS	BBC	UCI-digit	HW	Hdigit	Wiki	CIFAR10-All
ACC	Autoencoder	47.20	80.58	93.55	81.65	99.46	59.39	30.65
	FCAencoder	95.00	87.01	95.25	96.90	99.88	63.19	36.94
NMI	Autoencoder	38.52	62.67	87.30	81.64	98.31	53.07	20.44
	FCAencoder	84.97	67.01	90.31	93.15	99.60	55.10	24.79
Purity	Autoencoder	49.60	80.58	93.55	81.65	99.46	62.67	31.87
	FCAencoder	95.00	87.01	95.25	96.90	99.88	65.25	37.84
F-Score	Autoencoder	46.12	81.58	93.63	80.83	99.46	61.07	30.89
	FCAencoder	94.98	87.19	95.26	96.90	99.88	62.74	36.86

that the progressive fusion mechanism consistently outperforms direct concatenation fusion across all three datasets. In terms of ACC, the progressive fusion approach achieves improvements of 19.86%, 9.35%, and 3.04% on the BBC, HW, and CIFAR10-All datasets, respectively. These findings highlight the advantage of the progressive fusion strategy for clustering tasks. Conceptually, features from more dissimilar views are more challenging to fuse effectively due to the increased feature gap. By progressively fusing the views, our approach mitigates this challenge by inte-

grating the most similar views first, which leads to more effective information aggregation. Additionally, we visualized the similarity matrices of the final fused clustering representations for both fusion strategies, as shown in Figure 3. The similarity matrices show that after progressive fusion, the intra-cluster similarity is significantly enhanced, resulting in clearer block-like structures. This visual evidence further supports the effectiveness of the progressive fusion mechanism in enhancing cluster representation.

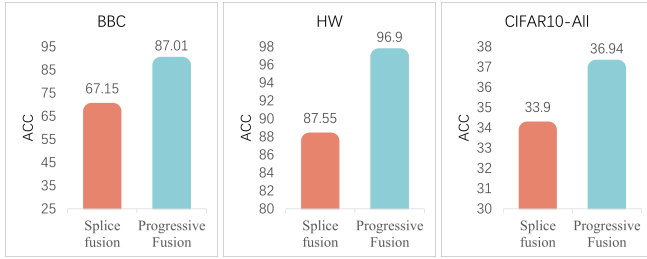


Figure 2. Comparison results of progressive fusion and splicing fusion on ACC.

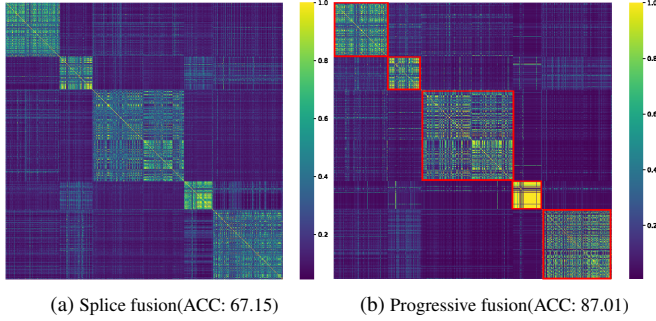


Figure 3. Comparison of representation similarity graphs between splice fusion and progressive fusion on BBC.

4.6. Sensitivity Analysis and Visualization

We also conducted a parameter sensitivity analysis of our model, with the detailed results presented in Figure 4. Our model involves two hyperparameters, γ and ε , which we varied over a range of values: [0.001, 0.01, 0.1, 1, 10, 100, 1000]. The results demonstrate that, as the values of the two hyperparameters change, the clustering performance on the UCI-Digit and HW datasets exhibits minor fluctuations. However, the variations in performance remain within a relatively small range, demonstrating the robustness of our method with respect to hyperparameter selection. Additionally, we visualized the original features and the learned feature representations of the UCI-Digit dataset using t-SNE, as shown in Figure 5. The visualization reveals that the learned representations exhibit a high degree of intra-cluster compactness and clear inter-cluster separation, indicating strong discriminative power.

These findings further support the effectiveness of our proposed method in learning high-quality, well-separated feature representations for clustering tasks.

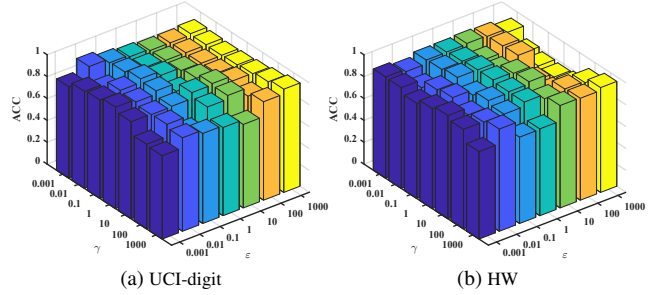


Figure 4. Sensitivity analysis of γ and ε applied to the UCI-digit and HW datasets.

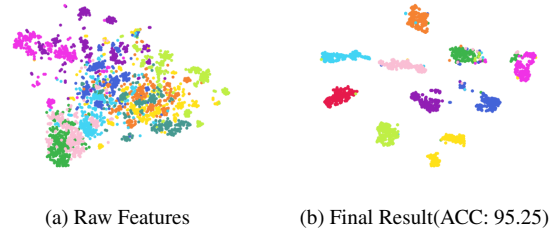


Figure 5. Visualization of the original features and final results on the UCI-digit dataset.

5. Conclusion

In this paper, we propose a novel multi-view clustering framework that combines feature enhancement with progressive fusion to address key challenges in multi-view data integration. Specifically, the framework utilizes a Feature Channel Attention Encoder to enhance discriminative features and a View Graph-based Progressive Fusion Mechanism to align views in a shared latent space. Extensive experiments demonstrate that our method outperforms existing state-of-the-art techniques, achieving superior clustering accuracy across various datasets. We believe our approach provides new insights for the multi-view clustering community, opening up avenues for further research and practical applications.

6. Acknowledgment

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, 62476281, 62441618), National Natural Science Foundation of China Joint Found under Grant No. U24A20323.

References

- [1] Zhe Chen, Xiao-Jun Wu, Tianyang Xu, and Josef Kittler. Fast self-guided multi-view subspace clustering. *IEEE Transactions on Image Processing*, 32:6514–6525, 2023. [1](#), [2](#)
- [2] Chenhang Cui, Yazhou Ren, Jingyu Pu, Xiaorong Pu, and Lifang He. Deep multi-view subspace clustering with anchor graph. *arXiv preprint arXiv:2305.06939*, 2023. [1](#)
- [3] Jinrong Cui, Zhipeng He, Qiong Huang, Yulu Fu, Yuting Li, and Jie Wen. Structure-aware contrastive hashing for unsupervised cross-modal retrieval. *Neural Networks*, 174:106211, 2024. [2](#)
- [4] Jinrong Cui, Yuting Li, Han Huang, and Jie Wen. Dual contrast-driven deep multi-view clustering. *IEEE Transactions on Image Processing*, 33:4753–4764, 2024. [2](#)
- [5] Jinrong Cui, Yazi Xie, Chengliang Liu, Qiong Huang, Mu Li, and Jie Wen. Deep dual incomplete multi-view multi-label classification via label semantic-guided contrastive learning. *Neural Networks*, 180:106674, 2024. [2](#)
- [6] Zhibin Dong, Jiaqi Jin, Yuyang Xiao, Siwei Wang, Xinzhong Zhu, Xinwang Liu, and En Zhu. Iterative deep structural graph contrast clustering for multiview raw data. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. [1](#)
- [7] Zhibin Dong, Siwei Wang, Jiaqi Jin, Xinwang Liu, and En Zhu. Cross-view topology based consistent and complementary information for deep multi-view clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19440–19451, 2023. [1](#)
- [8] Uno Fang, Man Li, Jianxin Li, Longxiang Gao, Tao Jia, and Yanchun Zhang. A comprehensive survey on multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12350–12368, 2023. [1](#)
- [9] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. [2](#), [3](#)
- [10] Shizhe Hu, Guoliang Zou, Chaoyang Zhang, Zhengzheng Lou, Ruilin Geng, and Yangdong Ye. Joint contrastive triple-learning for deep multi-view clustering. *Information Processing & Management*, 60(3):103284, 2023. [1](#)
- [11] Dong Huang, Xiaozhi Deng, Ding-Hua Chen, Zihao Wen, Weijun Sun, Chang-Dong Wang, and Jian-Huang Lai. Deep clustering with hybrid-grained contrastive and discriminative learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024. [1](#)
- [12] Haojian Huang, Chuanyu Qin, Zhe Liu, Kaijing Ma, Jin Chen, Han Fang, Chao Ban, Hao Sun, and Zhongjiang He. Trusted unified feature-neighborhood dynamics for multi-view classification. *arXiv preprint arXiv:2409.00755*, 2024. [2](#)
- [13] Shudong Huang, Zhao Kang, and Zenglin Xu. Auto-weighted multi-view clustering via deep matrix decomposition. *Pattern Recognition*, 97:107015, 2020. [2](#)
- [14] Yan-Di Huang, Guang-Yu Zhang, Dong Huang, Chang-Dong Wang, Yang Liu, and Enbo Huang. Confidence-oriented contrastive graph clustering. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2024. [1](#)
- [15] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4412–4419, 2020. [6](#)
- [16] Junlong Ke, Zichen Wen, Yechenhao Yang, Chenhang Cui, Yazhou Ren, Xiaorong Pu, and Lifang He. Integrating vision-language semantic graphs in multi-view clustering. *IJCAI*, 2024. [1](#)
- [17] Xuelong Li, Hongyuan Zhang, and Rui Zhang. Adaptive graph auto-encoder for general data clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9725–9732, 2022. [6](#)
- [18] Xingfeng Li, Zhenwen Ren, Quansen Sun, and Zhi Xu. Auto-weighted tensor Schatten p-norm for robust multi-view graph clustering. *Pattern Recognition*, 134:109083, 2023. [1](#)
- [19] Zhaoyang Li, Qianqian Wang, Zhiqiang Tao, Quanxue Gao, Zhaohua Yang, et al. Deep adversarial multi-view clustering network. In *IJCAI*, pages 2952–2958, 2019. [2](#)
- [20] Fangfei Lin, Bing Bai, Kun Bai, Yazhou Ren, Peng Zhao, and Zenglin Xu. Contrastive multi-view hyperbolic hierarchical clustering. *arXiv preprint arXiv:2205.02618*, 2022. [2](#)
- [21] Yiding Lu, Yijie Lin, Mouxing Yang, Dezhong Peng, Peng Hu, and Xi Peng. Decoupled contrastive multi-view clustering with high-order random walks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 14193–14201, 2024. [1](#), [3](#)
- [22] Caixuan Luo, Jie Xu, Yazhou Ren, Junbo Ma, and Xiaofeng Zhu. Simple contrastive multi-view clustering with data-level fusion. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 4697–4705, 2024. [2](#)
- [23] Sayantan Mitra, Sriparna Saha, and Mohammed Hasanuzzaman. Multi-view clustering for multi-omics data using unified embedding. *Scientific reports*, 10(1):13654, 2020. [1](#)
- [24] Yalan Qin, Zhenjun Tang, Hanzhou Wu, and Guorui Feng. Flexible tensor learning for multi-view clustering with markov chain. *IEEE Transactions on Knowledge and Data Engineering*, 2023. [1](#)
- [25] Yazhou Ren, Jingyu Pu, Chenhang Cui, Yan Zheng, Xinyue Chen, Xiaorong Pu, and Lifang He. Dynamic weighted graph fusion for deep multi-view clustering. *IJCAI*, 2024. [2](#)
- [26] Yazhou Ren, Jingyu Pu, Zhimeng Yang, Jie Xu, Guofeng Li, Xiaorong Pu, S Yu Philip, and Lifang He. Deep clustering: A comprehensive survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2024. [1](#)
- [27] Long Shi, Lei Cao, Jun Wang, and Badong Chen. Enhanced latent multi-view subspace clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024. [1](#)
- [28] Yuan Sun, Yang Qin, Yongxiang Li, Dezhong Peng, Xi Peng, and Peng Hu. Robust multi-view clustering with noisy correspondence. *IEEE Transactions on Knowledge and Data Engineering*, 2024. [1](#)

- [29] Huayi Tang and Yong Liu. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 202–211, 2022. 6
- [30] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 2
- [31] Zhibin Wan, Changqing Zhang, Pengfei Zhu, and Qinghua Hu. Multi-view information-bottleneck representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10085–10092, 2021. 2
- [32] Fangdi Wang, Jiaqi Jin, Zhibin Dong, Xihong Yang, Yu Feng, Xinwang Liu, Xinzong Zhu, Siwei Wang, Tianrui Liu, and En Zhu. View gap matters: Cross-view topology and information decoupling for multi-view clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 8431–8440, 2024. 1
- [33] Huibing Wang, Mingze Yao, Guangqi Jiang, Zetian Mi, and Xianping Fu. Graph-collaborated auto-encoder hashing for multiview binary clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 1
- [34] Shiping Wang, Xincan Lin, Zihan Fang, Shide Du, and Guobao Xiao. Contrastive consensus graph learning for multi-view clustering. *IEEE/CAA Journal of Automatica Sinica*, 9(11):2027–2030, 2022. 1
- [35] Siwei Wang, Xinwang Liu, Suyuan Liu, Wenxuan Tu, and En Zhu. Scalable and structural multi-view graph clustering with adaptive anchor fusion. *IEEE Transactions on Image Processing*, 33:4627–4639, 2024. 1
- [36] Yinuo Wang, Yu Guo, Zheng Wang, and Fei Wang. Joint learning of latent subspace and structured graph for multi-view clustering. *Pattern Recognition*, 154:110592, 2024. 1
- [37] Zichen Wen, Tianyi Wu, Yazhou Ren, Yawen Ling, Chenhang Cui, Xiaorong Pu, and Lifang He. Dual-optimized adaptive graph reconstruction for multi-view graph clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 1819–1828, 2024. 1
- [38] Song Wu, Yan Zheng, Yazhou Ren, Jing He, Xiaorong Pu, Shudong Huang, Zhifeng Hao, and Lifang He. Self-weighted contrastive fusion for deep multi-view clustering. *IEEE Transactions on Multimedia*, 2024. 1
- [39] Shunxin Xiao, Shide Du, Zhaoliang Chen, Yunhe Zhang, and Shiping Wang. Dual fusion-propagation graph neural network for multi-view clustering. *IEEE Transactions on Multimedia*, 25:9203–9215, 2023. 1, 2
- [40] Shunxin Xiao, Jiacheng Li, Jielong Lu, Sujia Huang, Bao Zeng, and Shiping Wang. Graph neural networks for multi-view learning: a taxonomic review. *Artificial Intelligence Review*, 57(12):341, 2024. 1
- [41] Yuan Xie, Bingqian Lin, Yanyun Qu, Cuihua Li, Wensheng Zhang, Lizhuang Ma, Yonggang Wen, and Dacheng Tao. Joint deep multi-view learning for image clustering. *IEEE Transactions on Knowledge and Data Engineering*, 33(11):3594–3606, 2020. 1
- [42] Jie Xu, Yazhou Ren, Huayi Tang, Zhimeng Yang, Lili Pan, Yang Yang, Xiaorong Pu, S Yu Philip, and Lifang He. Self-supervised discriminative feature learning for deep multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2022. 1, 2
- [43] Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-level feature learning for contrastive multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16051–16060, 2022. 2, 6
- [44] Jie Xu, Yazhou Ren, Huayi Tang, Zhimeng Yang, Lili Pan, Yang Yang, Xiaorong Pu, Philip S. Yu, and Lifang He. Self-supervised discriminative feature learning for deep multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):7470–7482, 2023. 6
- [45] Jie Xu, Shuo Chen, Yazhou Ren, Xiaoshuang Shi, Hengtao Shen, Gang Niu, and Xiaofeng Zhu. Self-weighted contrastive learning among multiple views for mitigating representation degeneration. *Advances in Neural Information Processing Systems*, 36, 2024. 6
- [46] Jie Xu, Yazhou Ren, Xiaolong Wang, Lei Feng, Zheng Zhang, Gang Niu, and Xiaofeng Zhu. Investigating and mitigating the side effects of noisy views for self-supervised clustering algorithms in practical multi-view scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22957–22966, 2024. 1, 6
- [47] Weiqing Yan, Yuanyang Zhang, Chenlei Lv, Chang Tang, Guanghui Yue, Liang Liao, and Weisi Lin. Gcfagg: Global and cross-view feature aggregation for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19863–19872, 2023. 1
- [48] Xihong Yang, Jin Jiaqi, Siwei Wang, Ke Liang, Yue Liu, Yi Wen, Suyuan Liu, Sihang Zhou, Xinwang Liu, and En Zhu. Dealmvc: Dual contrastive calibration for multi-view clustering. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 337–346, 2023. 1
- [49] Zhijian Yang, Junhao Wen, Ahmed Abdulkadir, Yuhuan Cui, Guray Erus, Elizabeth Mamourian, Randa Melhem, Dhiyaa Srinivasan, Sindhuja T Govindarajan, Jiong Chen, et al. Gene-sgan: discovering disease subtypes with imaging and genetic signatures via multi-view weakly-supervised deep clustering. *Nature communications*, 15(1):354, 2024. 1
- [50] Changqing Zhang, Yeqing Liu, and Huazhu Fu. Ae2-nets: Autoencoder in autoencoder networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 6
- [51] Zheng Zhang, Li Liu, Fumin Shen, Heng Tao Shen, and Ling Shao. Binary multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1774–1782, 2019. 6
- [52] Yan Zheng, Song Wu, Junyu Lin, Yazhou Ren, Jing He, Xiaorong Pu, and Lifang He. Cross-view contrastive fusion for enhanced molecular property prediction. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 2024. 2
- [53] Pengfei Zhu, Xinjie Yao, Yu Wang, Binyuan Hui, Dawei Du, and Qinghua Hu. Multiview deep subspace clustering networks. *IEEE Transactions on Cybernetics*, 2024. 1