Dynamic Weighted Graph Fusion for Deep Multi-View Clustering

Yazhou Ren[1,2]*, Jingyu Pu[1], Chenhang Cui[1], Yan Zheng[1],
Xinyue Chen[1], Xiaorong Pu[1,2], Lifang He[3]
[1]School of Computer Science and Engineering,University of Electronic Science and Technology of China
[2]Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China
[3]Department of Computer Science and Engineering, Lehigh University, Bethlehem, USA
{yazhou.ren, puxiaor}@uestc.edu.cn, pujingyu0105@163.com,
{chenhangcui, yan9zheng9, martinachen2580}@gmail.com, lih319@lehigh.edu

Abstract

By exploring complex graph information hidden in data from multiple views, multi-view clustering based on graph neural network significantly enhances the clustering performance and has drawn increasing attention in recent years. Although considerable progress has been made, most existing GNN based MVC models merely consider the explicit presence of graph structure in raw data and ignore that latent graphs of different views also provide specific information for the clustering task. We propose dynamic weighted graph fusion for deep multi-view clustering (DFMVC) to address this issue. Specifically, DFMVC learns embedded features via deep autoencoders and then constructs latent graphs for each individual view. Then, it concatenates the embedded features of all views to form a global feature to leverage complementary information, as well as generates a fusion graph via combining all latent graphs to accurately capture the topological information among samples. Based on the informative fusion graph and global features, the graph convolution module is adopted to derive a representation with global comprehensive information, which is further used to generate pseudo-label information. In a self-supervised manner, such information guides each view to dynamically learn discriminative features and latent graphs. Extensive experimental results demonstrate the efficacy of DFMVC.

1 Introduction

Clustering represents a fundamental task in unsupervised learning and has been widely utilized across various domains, such as biology [Gönen and Margolin, 2014], psychology [Burr *et al.*, 2022], information retrieval [Bruno and Marchand-Maillet, 2009], etc. Conventional clustering methods typically work with single-view data [Peng *et al.*, 2019]. However, with the advance of multimedia technology, real data are frequently collected from multiple sources or described from different perspectives, known as multi-view data. Numerous studies [Wang *et al.*, 2023; Wang *et al.*, 2022a; Zhang *et al.*, 2015; Han *et al.*, 2022; Cui *et al.*, 2024; Wen *et al.*, 2024] have shown that multi-view data contains complementary information that could help the model to learn more comprehensive and more expressive representations. In the literature, multi-view clustering (MVC) has been extensively studied at both shallow and deep levels. Shallow MVC approaches include 1) co-training approach [Nie *et al.*, 2020; Hu *et al.*, 2020], which seeks to maximize the mutual agreement across multiple views; 2) multi-kernel approach [Gönen and Margolin, 2014; Lu *et al.*, 2014], which regards different kernels as multiple views and linearly or non-linearly integrates them; 3) subspace clustering approach [Li *et al.*, 2019a; Zheng *et al.*, 2020], which aims at exploring shared representations of multiple views; and 4) graph-based approach [Tang *et al.*, 2020; Lin and Kang, 2022], which attempts to deploy the relations between different instances with certain distance metrics or similarity. Deep MVC approaches include 1) deep embedded clustering (DEC) based approach [Xu *et al.*, 2023], which adopts autoencoders to learn low-dimensional embeddings [Yang *et al.*, 2021; Li *et al.*, 2020]; 2) generative adversarial network (GAN) based approach [Zhou and Shen, 2020; Li *et al.*, 2019b], which uses adversarial training to capture data distribution; and 3) graph neural network (GNN) based approach [Fan *et al.*, 2020; Xia *et al.*, 2021], which applies GNN to take full advantage of the characteristics of multi-view graph data. In this study, we focus on DEC and GNN based MVC, as it allows to exploit the principles of consistency and complementarity as well as the topological information hidden in multi-view data.

Most existing graph-based MVC methods [Chen *et al.*, 2022] often require explicit graph data as input, real-world data often lacks the original graph structure. Therefore, graph-based MVC methods [Wang *et al.*, 2022a] with unknown graph structure is emerging in recent years, which consider graph structures in multi-view data by constructing local or global graphs. However, most existing methods suffer from the drawback that the constructed graphs are not clustering task-oriented. They also generally ignore the local structures of all views, and the constructed graph structure information is not properly utilized.

To address the above issues, in this paper we propose **d**ynamic weighted graph **f**usion for deep **m**ulti-**v**iew

*Corresponding author.

clustering (DFMVC). DFMVC firstly learns the embedded features via deep autoencoders and constructs local latent graphs for multiple views seperately. Then, the embedded features of all views are concatenated to create a global feature to leverage complementary information and a fusion graph is generated via fusing all latent graphs to capture the topological information among samples.

The proposed DFMVC incorporates global features and the fusion graph as inputs to the graph convolution module, which serves to generate pseudo-label information for self-supervision. This valuable information acts as a supervisory signal for all views, facilitating the learning of embedded features and latent graphs. To ensure the continual refinement of the pseudo-label information, these updated embedded features and latent graphs are cyclically utilized. Furthermore, in order to enhance fusion process and enhance the quality of the fused graph, a meticulous weighted graph fusion scheme is devised. This scheme effectively discerns and accounts for the distinct contributions made by each individual view.

In general, DFMVC effectively enhances the clustering performance by comprehensively considering both the latent sample attributes and the intrinsic topological structure of all views. Moreover, by refining the fusion graph dynamically, even if the learned graph structure is not suitable for clustering at the beginning. With the aid of all views' complementing information, it will be steadily optimized. The main contributions of this work are as follows:

- We propose a novel deep MVC framework by exploiting dynamic graph fusion. It utilizes both global features, offering complementary information, and latent graphs, which provide topological information, to enhance deep multi-view clustering capabilities.
- In a self-supervised manner, we dynamically update the fusion graph and pseudo-label information such that constructed graph structure favors clustering task.
- We design a weighted graph fusion scheme that adaptively evaluates the weight of each view to enhance the quality of graph fusion. Extensive experiments on real-world multi-view data demonstrate the superior performance of our proposed model.

2 Related Work

2.1 Multi-View Clusteirng

By utilizing the complimentary information among multi-view data, multi-view clustering can enhance the clustering performance. The core problem is how to make full use of the information among different views and give full play to each view's advantages while reducing each view's limitations to achieve accurate and robust clustering.

Bickel et al. [Bickel and Scheffer, 2004] used the strategy of stitching to connect features of multiple views into a new feature space in which traditional clustering algorithms are executed to achieve better clustering performance, leading to a series of works [Zheng *et al.*, 2020; Yu *et al.*, 2021]. However, this type of method ignores the complementary information hidden in multi-view data. To this end, MVC methods based on representation learning emerged, which can be further divided into multi-view graph clustering [Kang *et al.*, 2020], multi-view subspace clustering [Zheng *et al.*, 2020], and nonnegative matrix factorization (NMF) based multi-view clustering [Ren *et al.*, 2020].

Nowadays, MVC has also become a hot research topic in deep learning, and numerous deep MVC models [Xu *et al.*, 2021b; Lin *et al.*, 2022] have been proposed. Using autoencoders to learn low-dimensional feature representations for multi-view clustering [Chen *et al.*, 2020; Xu *et al.*, 2021a] allows for adequate learning of representative latent features, and these studies have yielded impressive results.

2.2 Graph-based Deep Multi-View Clustering

Graph neural networks (GNNs) are widely used in unsupervised learning due to their powerful representation capabilities on graph-structured data, and many graph-based deep MVC models have emerged [Tang *et al.*, 2020; Li *et al.*, 2021; Wang *et al.*, 2022b]. The graph convolution network (GCN), an important branch of GNN, which can considerably enhance the effectiveness of the clustering algorithm by extracting data from the properties of the graph and the nodes, attracts people's increasing attention [Ren *et al.*, 2022; Du *et al.*, 2023; Ren *et al.*, 2024].

Many researchers have considered the explicit presence of graph structure in multi-view data, and constructed local and global graphs to perform multi-graph learning and consensus clustering [Zheng *et al.*, 2021; Hao *et al.*, 2021]. For example, Wang et al. [Wang *et al.*, 2022a] used k-nearest neighbors (kNN) to construct the attribute graph structure for multi-view source data and then obtained attention weighted fusion graph based on graph encoders and attention network to complete the graph reconstruction and clustering tasks. Ma et al. [Ma *et al.*, 2022] construct multiple learned graphs using multi-view data. This approach aims to effectively capture the intricate local manifold structures that are concealed within each view. By generating high-quality graph structures and consensus graphs, these components mutually influence and guide one another, leading to iterative refinement and improvement in the overall analysis process.

Although the above methods improve the MVC performance by constructing multi-view graphs and fusing graph structure, they typically ignore the latent graphs of different views. To address this, Fan et al. [Fan *et al.*, 2020] pioneered the application of the Graph Convolutional Network (GCN) architecture to tackle a challenging multi-view clustering task. This novel approach incorporates a GCN encoder along with multiple graph decoders, enabling the effective encoding of multi-view attribute graphs into a low dimensional feature space. By leveraging this framework, the encoded representations capture essential information from diverse views, facilitating comprehensive and robust clustering of the data.

Xia et al. [Xia *et al.*, 2021] used the Euler transform to extract features as view descriptors, and then performed representation learning and clustering by constructing a multi-view self-supervised graph convolutional clustering network. However, the graphs in existing GCN-based MVC methods are typically fixed, which makes the predefined/constructed graphs a major determinant of clustering performance.

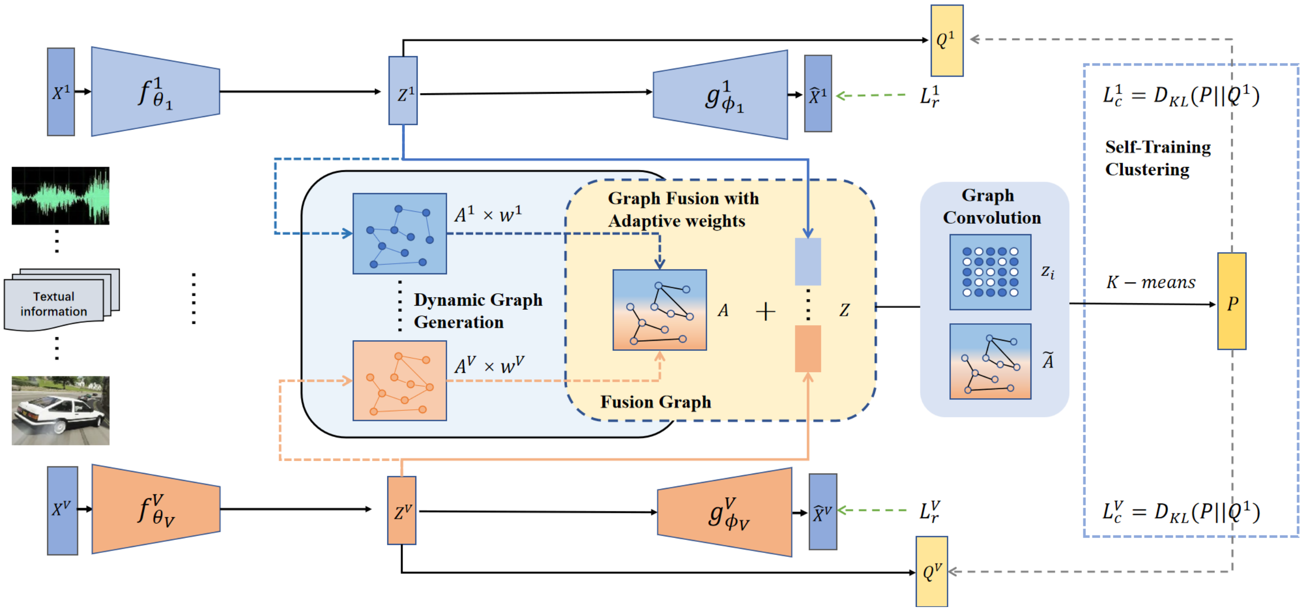

Figure 1: The framework of DFMVC. For the v-th view, X^v denotes the input data, Z^v means the learned features, A^v is the latent adjacency graph generated from Z^v, and Q^v is the cluster assignment distribution. A, Z, and P denote the weighted fusion graph, global features, and unified target distribution respectively.

In conclusion, although a great deal of GCN-based MVC models have made encouraging progress, they still have inherent shortcomings. First, since the graph information is usually not given, GCN models that require explicit graph data as input are not practical in real-world applications. Second, the graphs in GCN models are typically fixed, which may not favor the clustering task. By contrast, our proposed model dynamically constructs the graph data by embedding the features of multi-view data, which is suitable for MVC problem without explicit graphs. In addition, we design a dynamic weighted graph fusion strategy to address the graph fixation issue, further improving the practicality and robustness of the MVC model.

3 Proposed Method

Given a multi-view dataset $X = \{X^v \in R^{N \times d_v}\}_{v=1}^{V}$, where $X^v = [x_1^v; x_2^v; \dots; x_N^v]$ denotes the v-th view's data, N is the number of instances, V is the number of total views, and d_v is the dimension of the v-th view. The goal of dynamic weighted graph fusion for deep multi-view clustering (DFMVC) is to partition the instances into K clusters. Figure 1 provides an overview of our proposed DFMVC framework.

3.1 Dynamic Graph Generation

To learn the representative latent features, we use the deep autoencoder [Rumelhart *et al.*, 1986] to project the raw data of each view to a low-dimensional space. For the v-th view, $f^v_{\theta^v}$ and $g^v_{\phi^v}$ denote the encoder and decoder respectively, where θ^v and ϕ^v are learnable parameters. The embedded feature can then be indicated as

$$z_i^v = f^v_{\theta^v}(x_i^v), \tag{1}$$

where $z_i^v \in R^l$ denotes the l-dimensional embedded feature of x_i^v (l is the same for all views). After that, z_i^v is decoded as $\hat{x}_i^v$ through the decoder $g^v_{\phi^v}$, and the reconstruction loss of all the views can be formulated as

$$L_r = \sum_{v=1}^{V} L_r^v = \sum_{v=1}^{V} \sum_{i=1}^{N} ||x_i^v - g^v_{\phi^v}(f^v_{\theta^v}(x_i^v))||^2. \tag{2}$$

To obtain expressive graph structure representation, we dynamically generate a graph in each view's latent space. It is mainly because when composing fixed graphs based on raw data [Wang *et al.*, 2021; Tang *et al.*, 2020], graphs constructed cannot be adjusted according to the clustering results. On the contrary, as the features extracted by the deep autoencoder are more representative than the raw features, we construct graphs in the latent space. In addition, during the dynamic update process, the fusion of graphs and the acquisition of pseudo-label information are carried out alternately and promote each other. Furthermore, unlike [Cheng *et al.*, 2021; Ren *et al.*, 2022] that necessitate explicit graph data as input, our approach introduces the capability to dynamically construct graphs within the latent feature space. This advantageous characteristic allows us to overcome the limitations posed by a scarcity of real-world data with graph structures.

Therefore, we choose to generate graph information in the latent feature space and update it dynamically.

Specifically, we define the latent graph $G^v = \{Z^v, A^v\}$ for the v-th view, where $Z^v = [z_1^v; z_2^v; \dots; z_N^v]$ denotes the node properties of graph G^v. A^v is the adjacency matrix and A_{ij}^v denotes the edge between the samples i and j in the v-th view. Samples i and j are connected if $A_{ij}^v = 1$ and disconnected otherwise. We utilize the kNN graph [Peterson, 2009] to construct A^v:

$$A_{ij}^v = \begin{cases} 1 & j \in N_i^v \\ 0 & \text{otherwise}, \end{cases} \tag{3}$$

where N_i^v is the set of k nearest neighbors of z_i^v in the v-th view according to the Euclidean distance in the latent space.

3.2 Graph Fusion with Adaptive Weight

Our motivations for designing the weighted dynamic graph fusion mechanism are:

- The latent graphs of different views contain the discriminative and complementary information. Therefore, fusing these multi-view graphs into a more informative and robust global graph will be beneficial to the clustering task.
- Due to the noisy issue and the incompleteness of raw features, the generated graphs in certain views may not correctly reflect the actual topological structure among samples. Thus it is essential to adaptively assign different weights to the graphs with different qualities.

Building upon the reasons mentioned above, we approach dynamic graph fusion as follows:

$$A = \min_A \sum_{v=1}^{V} w_v ||A - A^v||_F^2. \tag{4}$$

Inspired by [Kang *et al.*, 2020], we use the inverse distance of fusion graph A and each latent graph A^v to obtain w_v. In our model we assign the value of the exponential parameter of the inverse distance weighting method to 2 and then w_v can be adaptively computed as

$$w_v = \frac{1}{||A - A^v||_F}. \tag{5}$$

With the above weighting approach, a latent graph with high quality (which is close to the fusion graph) will be assigned a high weight. By this way, the influence of graphs reflecting clear topological structure and complementary information will be enhanced and correspondingly the graphs containing unclear information will play an inconsequential role in the cluster assignment.

Based on the obtained graph weights, in each iteration, the proposed DFMVC updates the fusion graph colmun-wisely through the following precedures:

$$\min_{A(:,i)} \sum_{v=1}^{V} w_v ||A(:,i) - A^v(:,i)||^2. \tag{6}$$

By computing the derivative of Eq. (6) and setting it to zero, it yields

$$A(:,i) = \frac{\sum_v w_v A^v(:,i)}{\sum_v w_v}. \tag{7}$$

The graph A is a fusion of $[A^1, ..., A^v]$ which is supposed to capture the ground-truth sample similarity hidden in the multi-view data. $[A^1, ..., A^v]$ are generated by embeddings from each view, they may not be optimal for the clustering result at first. At the beginning of the training, latent graphs from different views are given the equal weight. Then, the graph that is close to the fusion graph should be assigned a large weight. A relative favourable fusion graph can be learned through weight update. Through the dynamic graph fusion, our method progressively obtain more comprehensive information and clearer clustering structure from the latent features as training process forwards. Meanwhile, by assigning smaller weights to the less reliable views, the negative affect of noise graphs is reduced effectively and thus the clustering performance can be further promoted.

3.3 Graph Convolution

To enhance the clustering results with both latent features and topological information, we further utilize the GCN structure to obtain better sample representations based on the fusion graph and global features.

Concretely, as the embedded features of different views are learned independently and each view can provide specific and complementary information, we concatenate the embedded features across all the views to generate the global feature $z_i = \left[z_i^1, z_i^2, \ldots, z_i^V\right] \in R^{1 \times \sum_{v=1}^V d_v}$. Let $Z = [z_1; z_2; \ldots; z_N] \in R^{N \times \sum_{v=1}^V d_v}$ denote all the global features. To leverage the graph structure information, we introduce a two-layer GCN module for dynamic fusion graph. By aggregating the multi-order neighbourhood information with stacking layers, this GCN module guides the learned common representation maintaining the adjacent relationship in each view. As the GCN module directly performs graph convolution on the global feature and dynamic fusion graph, the refined representation $\tilde{z}$ is obtained by

$$\tilde{z}_i = \sum_{h=0}^{2} \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}\right)^h z_i, \tag{8}$$

where $\tilde{A} = I_n + A$, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. Here, the number of layers of the graph convolution is always set to 2. Through this GCN module, the dynamic fusion graph term seeks to find the underlying relationships from samples that make the cluster structure explicit. Compared with other existing GCN-based methods, our adopted simple yet effective GCN module has better generalization capability.

3.4 Self-Training Clustering

Inspired by a popular single-view deep embedded clustering (DEC) method [Xie *et al.*, 2016] that applies Student's t-distribution [Van der Maaten and Hinton, 2008] for self-training, we construct clustering layer $c_{\mu^v}^v$, where μ^v represents the learnable cluster centroids. For the v-th view, let $Q^v = [q_1^v; q_2^v; \ldots; q_N^v] \in R^{N \times K}$ denote the cluster assignments of all samples, where q_{ij}^v is the probability that the i-th sample belongs to the j-th cluster. q_{ij}^v can be considered as a soft cluster assignment and is obtained by:

$$q_{ij}^v = c_{\mu^v}^v\left(z_i^v\right) = \frac{(1 + \left\|z_i^v - \mu_j^v\right\|^2)^{-1}}{\sum_j (1 + \left\|z_i^v - \mu_j^v\right\|^2)^{-1}}. \tag{9}$$

After obtaining the refined global features $\tilde{z}$ via graph convolution module (Section 3.3), we adopt K-means [Mac-

Queen, 1967] to compute the cluster centroids c_j:

$$\min_{c_1,c_2,\ldots,c_K} \sum_{i=1}^{N}\sum_{j=1}^{K} \left\| \tilde{z}_i - c_j \right\|^2 . \tag{10}$$

Then, the soft pseudo assignment t_{ij} between each global embedding and each cluster centroid with Student's t-distribution is defined as

$$t_{ij} = \frac{(1 + \|\tilde{z}_i - c_j\|^2)^{-1}}{\sum_j (1 + \|\tilde{z}_i - c_j\|^2)^{-1}}. \tag{11}$$

In order to enhance the discriminability of the pseudo soft assignments, the global target distribution P can be computed by

$$p_{ij} = \frac{(t_{ij}^2 / \sum_i t_{ij})}{\sum_j (t_{ij}^2 / \sum_i t_{ij})}. \tag{12}$$

The clustering loss for the v-th view is determined by the KL divergence between the unified target distribution P and the cluster assignment distribution Q^v:

$$L_c^v = D_{KL}\left(P \| Q^v\right) = \sum_{i=1}^{N}\sum_{j=1}^{k} p_{ij} \log \frac{p_{ij}}{q_{ij}^v}. \tag{13}$$

To learn an accurate assignment for clustering, it is necessary to introduce an integrated objective to guide the learning process. With the aim of achieving this goal, we undertake a joint optimization approach, simultaneously optimizing the deep autoencoder embedding and clustering learning. Thus, the overall objective function of DFMVC is defined as:

$$L^v = (1-\tau) L_r^v + \tau L_c^v, \tag{14}$$

where $0 < \tau < 1$ is a trade-off coefficient controlling the degree of distorting embedded space, and is always set as $\tau = 0.5$ for all experiments (which means L_r^v and L_c^v are equally important). Minimizing KL divergence between Q and P makes the distribution of Q sharper and mine the information in different views. Global fusion can iteratively lead all views to learn more comprehensive information. After obtaining soft cluster assignments from multiple views, the highly confident predictions will guide the training process, which can also prevent the disruption caused by a few erroneous predictions. When DFMVC stops, the final clustering prediction can be obtained by:

$$y_i = \arg\max_j \left(\frac{1}{V} \sum_{v=1}^{V} q_{ij}^v \right). \tag{15}$$

3.5 Optimization

The detailed optimization procedure is described in Algorithm 1.

First, we pretrain autoencoders $f_{\theta^v}^v$ and $g_{\phi^v}^v$ of all views to obtain θ^v and ϕ^v by minimizing the reconstruction loss in Eq. (2). After that, we use K-means to initialize the cluster centroids of each view, and w_t is initialized to $\frac{1}{V}$. Then the fusion graph is obtained by Eq. (7). In the training process, with the advantage of dynamic latent graph fusion in discovering clearer graph structure representation and global information, the target distribution P obtained by fusion graph can better improve the cluster distribution of data and robustness with adaptive weight w_v. In addition, the weight of of each latent graph is updated iteratively during the training process.

Algorithm 1 Dynamic weighted graph fusion for deep multi-view clustering (DFMVC)

```
Input: multi-view dataset $X$, cluster number $K$.
Initialization: $\{w_v\}_{v=1}^V \leftarrow \frac{1}{V}$.
Obtain $\{\theta^v, \phi^v, \mu^v, A^v\}_{v=1}^V$ by pretraining autoencoders.
while not reach the maximum iterations $T_1$ do
    Update $P$ according to Eqs. (10) - (12).
    while not reach the maximum iterations $T_2$ do
        Update $\{\theta^v, \phi^v\}_{v=1}^V$ according to Eq. (14).
    end while
    Update $\{w_v\}_{v=1}^V$ according to Eq. (5).
    Update $A$ according to Eq. (7).
end while
Output: Cluster assignment $y$ according to Eq. (15).
```

4 Experiments

4.1 Experimental Setup

Datasets The following three real-world multi-view datasets are tested in our experiments. **BDGP** [Cai *et al.*, 2012] contains 2500 samples of 5 different types of drosophila embryos. Each sample has two views corresponding to the visual and textual features. **Fashion-MV** [Xiao *et al.*, 2017] collects images from 10 categories, i.e., t-shirt, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, and ankle boot. Following the literature [Xu *et al.*, 2021b], we use 30000 samples to construct Fashion-MV. Three viewpoints of a single occurrence are represented by each of the three photographs picked from the same category. **Handwritten Numerals (HW)** encompasses 2000 samples distributed across 10 classes, each representing numerals ranging from 0 to 9. Each category comprises six distinct visual views. The details of datasets are shown in Table 1.

Comparison Methods The comparison methods include three single-view clustering methods: K-means [MacQueen, 1967], SC (spectral clustering [Ng *et al.*, 2001]), DEC (deep embedded clustering [Xie *et al.*, 2016]), and seven state-of-the-art MVC methods: GFSC (multi-graph fusion for multi-view spectral clustering [Kang *et al.*, 2020]), GMC (graph-based multi-view clustering [Wang *et al.*, 2019]), SAMVC (self-paced and auto-weighted multi-view clustering [Ren *et al.*, 2020]), DEMVC (deep embedded multi-view clustering

Dataset	Sample	View	Dimension	Class
BDGP	2500	2	[1750, 79]	5
Fashion-MV	10000	3	[784, 784, 784]	10
HW	2000	6	[216, 76, 64, 6, 240, 47]	10

Table 1: Statistics of experimental datasets.

Datasets	BDGP			Fashion-MV			HW		
Methods	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
K-means (1967)[MacQueen, 1967]	43.24	56.94	26.04	70.93	65.61	56.89	75.45	78.58	66.72
SC (2001) [Ng *et al.*, 2001]	51.72	58.91	31.56	53.54	57.72	42.61	77.69	86.91	75.26
DEC (2016) [Xie *et al.*, 2016]	94.78	86.62	87.02	67.07	72.34	62.91	81.13	82.61	74.25
GMC (2019) [Wang *et al.*, 2019]	73.40	72.44	62.88	79.38	89.60	72.10	88.20	90.73	85.40
SAMVC (2020) [Ren *et al.*, 2020]	51.31	45.15	19.60	62.86	68.78	56.65	76.37	84.41	73.87
GFSC (2020) [Kang *et al.*, 2020]	93.14	86.51	94.74	-	-	-	86.56	83.66	90.02
DEMVC (2021) [Xu *et al.*, 2021a]	92.78	83.31	82.64	78.64	90.61	77.93	67.69	70.61	58.86
SIMVC (2021) [Trosten *et al.*, 2021]	70.40	54.51	72.62	82.50	83.93	84.51	64.10	82.10	66.58
COMVC (2021) [Trosten *et al.*, 2021]	80.22	67.01	80.32	85.73	86.44	83.31	73.90	83.44	72.77
SDMVC (2023) [Xu *et al.*, 2023]	97.89	93.41	94.85	86.26	92.15	84.05	97.18	94.44	93.93
Ours	**98.63**	**95.90**	**96.64**	**95.75**	**92.20**	**91.31**	**98.01**	**95.38**	**95.61**

Table 2: Results of all methods on three original datasets. In each column, the best result is shown in boldface and the second-best result is underlined.

with collaborative training [Xu *et al.*, 2021a]), SIMVC & COMVC (reconsidering representation alignment for multi-view clustering [Trosten *et al.*, 2021]), SDMVC (self-supervised discriminative feature learning for deep multi-view clustering [Xu *et al.*, 2023]).

Implementation Details Across all datasets utilized, we employ an identical fully connected (Fc) autoencoder structure. Specifically, for each view, the encoder structure comprises the following layers: Input - Fc300 - Fc1000 - Fc2000 - Fc10. The decoders mirror the respective encoders of their corresponding views. To facilitate pre-training, all autoencoders undergo 3000 epochs. For our method, the trade-off coefficient τ is set at 0.5, while the kNN graph algorithm employs a value of 30 for the number of neighbors (k). It is worth noting that our approach necessitates constructing a graph encompassing all nodes, with the batch size equal to the total number of instances (N). For the comparing methods, we use the source codes provided by the authors with suggested parameter settings.

4.2 Clustering Results

Comparison with Baselines The quantitative comparison of DFMVC and baseline models is shown in Table 2 (raw datasets). The best result is shown in bold in each column, and the second-best result is underlined. We can observe from Table 2 that DFMVC achieves the best across all multi-view datasets on three evaluation methods. Especially on Fashion-MV, DFMVC makes more remarkable improvements than existing methods. This is mainly because by dynamically updating the weighted fusion graph, DFMVC can effectively capture the explicit global information and avoid the influence of low-quality natural features in one view, such as noise or incompletion.

The graph-based shallow multi-view clustering methods achieve better performance, indicating that graph structure information can make beneficial contribution to clustering.

Ablation Studies To further verify the contribution of each module in the proposed method, we conduct ablation studies on the Eq. (4) and Eq. (8), which represent dynamically updating fusion graph and training with graph structure respectively. Table 3 shows the clustering results with different variants included, where **w/o** means that variant is removed from the method. Concretely, (A) is optimized without graph structure, which achieves the basic purpose of learning the consistency. (B) is optimized with fusion graph but without updating dynamically.

From the results, (B) performs better than (A) indicating that fusion graph structure contains more accurate global information to enhance the latent representation. Besides, we can also find that the full version of the proposed model (C) performs better than (B). The reason is that with the dynamic weighted graph fusion mechanism, our method obtains a clearer and more robust graph structure to mine the global information accurately.

We also conduct ablation studies on the noisy dataset. The experiment shows the excellent performance on noisy datasets illustrates the effectiveness of dynamic update process to deal with the incompleteness of raw features (half the feature values are zero in one view). The specific experimental results are presented in the appendix.

Visualization of Learning Process We visualize the learning process of embedded features on BDGP via t-SNE [Van der Maaten and Hinton, 2008]. The color denotes the label of each sample. After the pretraining of autoencoders is finished, the embedded features of the two views of BDGP are shown in Figure 2. At the beginning of finetuning stage ($T = 0$), the embedded features are non-separable. However, as the training process forwards. As the embedded features progressively separate, the clustering structures become more clearer, coinciding with the gradual dispersion of their respective centroids, which visually demonstrates the effectiveness of the proposed DFMVC.

Parameter Sensitivity Analysis The main hyper-parameters of the proposed DFMVC are the trade-off coefficient τ in Eq. (14) and the number of neighbors k in constructing the kNN graph. We test the average clustering performance of twenty independent runs on BDGP dataset and HW dataset. From Figure 3, we can observe that assigning k with a too small value will lead the graph become too sparse and thus lost some of graph structure information. As a consequence, with less information can be aggregated, the representation capability of the GCN module will be

	Variants	BDGP			Fashion-MV			HW		
	Evaluation Measures	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
(A)	**w/o** graph & dynamic update	71.78	65.44	59.85	70.58	77.40	65.60	83.37	75.12	69.69
(B)	**w/o** dynamic update	97.62	92.52	94.20	92.74	88.28	85.75	94.75	89.50	88.56
(C)	DFMVC	98.63	95.90	96.64	95.75	92.20	91.31	98.01	95.38	95.61

Table 3: Ablation studies on dynamic graph update process.

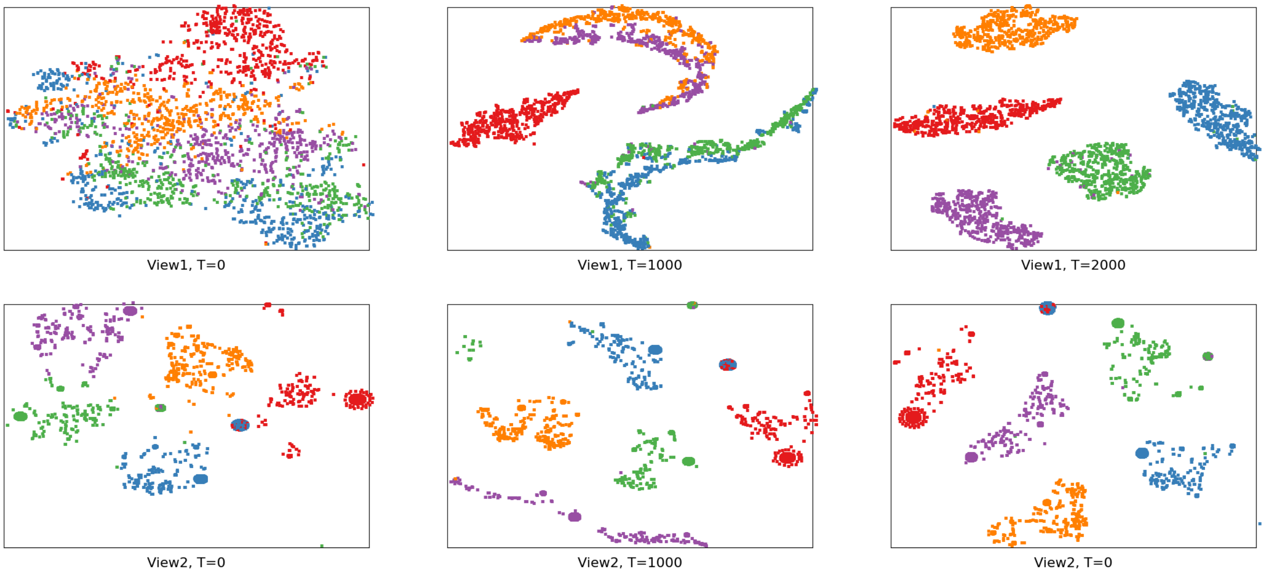

Figure 2: Visualization of the refined latent features via t-SNE of different views when training on BDGP dataset.

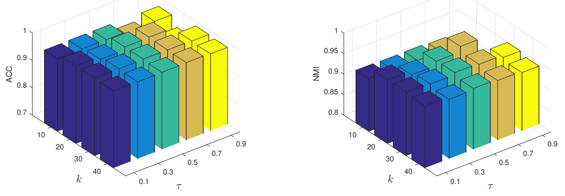

Figure 3: Clustering performance w.r.t. different parameter values on BDGP dataset and HW dataset.

weaken. However, that does not mean an extremely larger k is better, since it will bring more time complexity to the composition and may introduce unnecessary noise. Besides, it is shown that the too large and too small values of the trade-off parameter τ also negatively impact on the clustering performance. This indicates that both the reconstruction loss and clustering loss in Eq. (14) are important for the training of DFMVC. The suggested default parameter values are $k = 30$ and $\tau = 0.5$.

5 Conclusion

In this paper, we propose a dynamic weighted graph fusion framework for deep multi-view clustering (DFMVC), which effectively leverages both global features and latent graphs to enhance the clustering performance. Specifically, DFMVC firstly learns embedded features via deep autoencoders and then constructs latent graphs for each individual view. After that, it captures the complementary node attributes and topological information from different views by concatenating the embedded features of all views as well as fusing all the latent graphs. Extensive experiments on multiple real-world multi-view datasets and their noisy versions demonstrate the superior effectiveness and robustness of the proposed model. Since the computing complexity is a major issue of graph-based models, it is interesting to explore the deep multi-view graph clustering on large-scale datasets to further enhance our model's efficiency by utilizing techniques such as anchor graphs or encoding meaningful graph-related information.

Acknowledgments

This work is supported in part by Shenzhen Science and Technology Program under grants JCYJ20230807115959041 and JCYJ20230807120010021. Lifang He is partially supported by the NSF grants (MRI-2215789, IIS-1909879, IIS-2319451), NIH grant under R21EY034179, and Lehigh's grants under Accelerator and CORE.

References

[Bickel and Scheffer, 2004] Steffen Bickel and Tobias Scheffer. Multi-view clustering. In *ICDM*, pages 19–26, 2004.

[Bruno and Marchand-Maillet, 2009] Eric Bruno and Stéphane Marchand-Maillet. Multiview clustering: a late fusion approach using latent models. In *SIGIR*, pages 736–737, 2009.

[Burr *et al.*, 2022] Viv Burr, Nigel King, and Mark Heckmann. The qualitative analysis of repertory grid data: Interpretive clustering. *Qualitative Research in Psychology*, 19(3):678–702, 2022.

[Cai *et al.*, 2012] Xiao Cai, Hua Wang, Heng Huang, and Chris Ding. Joint stage recognition and anatomical annotation of drosophila gene expression patterns. *Bioinformatics*, 28(12):i16–i24, 2012.

[Chen *et al.*, 2020] Man-Sheng Chen, Ling Huang, Chang-Dong Wang, and Dong Huang. Multi-view clustering in latent embedding space. In *AAAI*, pages 3513–3520, 2020.

[Chen *et al.*, 2022] Jianpeng Chen, Yawen Ling, Jie Xu, Yazhou Ren, Shudong Huang, Xiaorong Pu, and Lifang He. Variational graph generator for multi-view graph clustering. *arXiv preprint arXiv:2210.07011*, 2022.

[Cheng *et al.*, 2021] Jiafeng Cheng, Qianqian Wang, Zhiqiang Tao, Deyan Xie, and Quanxue Gao. Multi-view attribute graph convolution networks for clustering. In *IJCAI*, pages 2973–2979, 2021.

[Cui *et al.*, 2024] Chenhang Cui, Yazhou Ren, Jingyu Pu, Jiawei Li, Xiaorong Pu, Tianyi Wu, Yutao Shi, and Lifang He. A novel approach for effective multi-view clustering with information-theoretic perspective. *Advances in Neural Information Processing Systems*, 36, 2024.

[Du *et al.*, 2023] Yangfan Du, Gui-Fu Lu, and Guangyan Ji. Robust and optimal neighborhood graph learning for multi-view clustering. *Information Sciences*, 631:429–448, 2023.

[Fan *et al.*, 2020] Shaohua Fan, Xiao Wang, Chuan Shi, Emiao Lu, Ken Lin, and Bai Wang. One2multi graph autoencoder for multi-view graph clustering. In *WWW*, pages 3070–3076, 2020.

[Gönen and Margolin, 2014] Mehmet Gönen and Adam A Margolin. Localized data fusion for kernel k-means clustering with application to cancer biology. In *NeurIPS*, pages 1–9, 2014.

[Han *et al.*, 2022] Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification with dynamic evidential fusion. *TPAMI*, 45(2):2551–2566, 2022.

[Hao *et al.*, 2021] Wenyu Hao, Shanmin Pang, and Zhikai Chen. Multi-view spectral clustering via common structure maximization of local and global representations. *Neural Networks*, 143:595–606, 2021.

[Hu *et al.*, 2020] Shizhe Hu, Xiaoqiang Yan, and Yangdong Ye. Dynamic auto-weighted multi-view co-clustering. *Pattern Recognition*, 99:107101, 2020.

[Kang *et al.*, 2020] Zhao Kang, Guoxin Shi, Shudong Huang, Wenyu Chen, Xiaorong Pu, Joey Tianyi Zhou, and Zenglin Xu. Multi-graph fusion for multi-view spectral clustering. *KBS*, 189:105102, 2020.

[Li *et al.*, 2019a] Ruihuang Li, Changqing Zhang, Huazhu Fu, Xi Peng, Joey Tianyi Zhou, and Qinghua Hu. Reciprocal multi-layer subspace learning for multi-view clustering. In *ICCV*, pages 8172–8180, 2019.

[Li *et al.*, 2019b] Zhaoyang Li, Qianqian Wang, Zhiqiang Tao, Quanxue Gao, and Zhaohua Yang. Deep adversarial multi-view clustering network. In *IJCAI*, pages 2952–2958, 2019.

[Li *et al.*, 2020] Xuelong Li, Rui Zhang, Qi Wang, and Hongyuan Zhang. Autoencoder constrained clustering with adaptive neighbors. *TNNLS*, 32(1):443–449, 2020.

[Li *et al.*, 2021] Lusi Li, Zhiqiang Wan, and Haibo He. Incomplete multi-view clustering with joint partition and graph learning. *TKDE*, pages 1–15, 2021.

[Lin and Kang, 2022] Zhiping Lin and Zhao Kang. Graph filter-based multi-view attributed graph clustering. In *IJCAI*, pages 2723–2729, 2022.

[Lin *et al.*, 2022] Fang-Yu Lin, Bing Bai, Kun Bai, Yazhou Ren, Peng Zhao, and Zenglin Xu. Contrastive multi-view hyperbolic hierarchical clustering. In *IJCAI*, 2022.

[Lu *et al.*, 2014] Yanting Lu, Liantao Wang, Jianfeng Lu, Jingyu Yang, and Chunhua Shen. Multiple kernel clustering based on centered kernel alignment. *Pattern Recognition*, 47:3656–3664, 2014.

[Ma *et al.*, 2022] Xuanlong Ma, Xueming Yan, Jingfa Liu, and Guo Zhong. Simultaneous multi-graph learning and clustering for multiview data. *Information Sciences*, 593:472–487, 2022.

[MacQueen, 1967] James MacQueen. Classification and analysis of multivariate observations. In *BSMSP*, pages 281–297, 1967.

[Ng *et al.*, 2001] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *NeurIPS*, pages 1–8, 2001.

[Nie *et al.*, 2020] Feiping Nie, Shaojun Shi, and Xuelong Li. Auto-weighted multi-view co-clustering via fast matrix factorization. *Pattern Recognition*, 102:107286, 2020.

[Peng *et al.*, 2019] Xi Peng, Hongyuan Zhu, Jiashi Feng, Chunhua Shen, Haixian Zhang, and Joey Tianyi Zhou. Deep clustering with sample-assignment invariance prior. *TNNLS*, 31(11):4857–4868, 2019.

[Peterson, 2009] Leif E Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.

[Ren *et al.*, 2020] Yazhou Ren, Shudong Huang, Peng Zhao, Minghao Han, and Zenglin Xu. Self-paced and auto-weighted multi-view clustering. *Neurocomputing*, 383:248–256, 2020.

[Ren *et al.*, 2022] Pengfei Ren, Haifeng Sun, Jiachang Hao, Jingyu Wang, Qi Qi, and Jianxin Liao. Mining multi-view information: A strong self-supervised framework for depth-based 3d hand pose and mesh estimation. In *CVPR*, pages 20555–20565, 2022.

[Ren *et al.*, 2024] Yazhou Ren, Xinyue Chen, Jie Xu, Jingyu Pu, Yonghao Huang, Xiaorong Pu, Ce Zhu, Xiaofeng Zhu, Zhifeng Hao, and Lifang He. A novel federated multi-view clustering method for unaligned and incomplete data fusion. *Information Fusion*, 108:102357, 2024.

[Rumelhart *et al.*, 1986] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[Tang *et al.*, 2020] Chang Tang, Xinwang Liu, Xinzhong Zhu, En Zhu, Zhigang Luo, Lizhe Wang, and Wen Gao. Cgd: Multi-view clustering via cross-view graph diffusion. In *AAAI*, pages 5924–5931, 2020.

[Trosten *et al.*, 2021] Daniel J Trosten, Sigurd Lokse, Robert Jenssen, and Michael Kampffmeyer. Reconsidering representation alignment for multi-view clustering. In *CVPR*, pages 1255–1265, 2021.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 9(11):2579–2605, 2008.

[Wang *et al.*, 2019] Hao Wang, Yan Yang, and Bing Liu. Gmc: Graph-based multi-view clustering. *TKDE*, 32(6):1116–1129, 2019.

[Wang *et al.*, 2021] Ru Wang, Lin Li, Xiaohui Tao, Xiao Dong, Peipei Wang, and Peiyu Liu. Trio-based collaborative multi-view graph clustering with multiple constraints. *IPM*, 58(3):102466, 2021.

[Wang *et al.*, 2022a] Ru Wang, Lin Li, Xiaohui Tao, Peipei Wang, and Peiyu Liu. Contrastive and attentive graph learning for multi-view clustering. *IPM*, 59(4):102967, 2022.

[Wang *et al.*, 2022b] Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu, Jiyuan Liu, Sihang Zhou, and En Zhu. Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph. In *CVPR*, pages 9776–9785, 2022.

[Wang *et al.*, 2023] Qianqian Wang, Zhiqiang Tao, Wei Xia, Quanxue Gao, Xiaochun Cao, and Licheng Jiao. Adversarial multiview clustering networks with adaptive fusion. *TNNLS*, 34:7635–7647, 2023.

[Wen *et al.*, 2024] Zichen Wen, Yawen Ling, Yazhou Ren, Tianyi Wu, Jianpeng Chen, Xiaorong Pu, Zhifeng Hao, and Lifang He. Homophily-related: Adaptive hybrid graph filter for multi-view graph clustering. In *AAAI*, pages 15841–15849, 2024.

[Xia *et al.*, 2021] Wei Xia, Qianqian Wang, Quanxue Gao, Xiangdong Zhang, and Xinbo Gao. Self-supervised graph convolutional network for multi-view clustering. *TMM*, pages 3182–3192, 2021.

[Xiao *et al.*, 2017] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

[Xie *et al.*, 2016] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *ICML*, pages 478–487, 2016.

[Xu *et al.*, 2021a] Jie Xu, Yazhou Ren, Guofeng Li, Lili Pan, Ce Zhu, and Zenglin Xu. Deep embedded multi-view clustering with collaborative training. *Information Sciences*, 573:279–290, 2021.

[Xu *et al.*, 2021b] Jie Xu, Yazhou Ren, Huayi Tang, Xiaorong Pu, Xiaofeng Zhu, Ming Zeng, and Lifang He. Multi-vae: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering. In *ICCV*, pages 9234–9243, 2021.

[Xu *et al.*, 2023] 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 & Data Engineering*, 35(07):7470–7482, 2023.

[Yang *et al.*, 2021] Lin Yang, Wentao Fan, and Nizar Bouguila. Deep clustering analysis via dual variational autoencoder with spherical latent embeddings. *TNNLS*, 2021.

[Yu *et al.*, 2021] Xiao Yu, Hui Liu, Yan Wu, and Caiming Zhang. Fine-grained similarity fusion for multi-view spectral clustering. *Information Sciences*, 568:350–368, 2021.

[Zhang *et al.*, 2015] Changqing Zhang, Huazhu Fu, Si Liu, Guangcan Liu, and Xiaochun Cao. Low-rank tensor constrained multiview subspace clustering. In *Proceedings of the IEEE international conference on computer vision*, pages 1582–1590, 2015.

[Zheng *et al.*, 2020] Qinghai Zheng, Jihua Zhu, Zhongyu Li, Shanmin Pang, Jun Wang, and Yaochen Li. Feature concatenation multi-view subspace clustering. *Neurocomputing*, 379:89–102, 2020.

[Zheng *et al.*, 2021] Qinghai Zheng, Jihua Zhu, Yuanyuan Ma, Zhongyu Li, and Zhiqiang Tian. Multi-view subspace clustering networks with local and global graph information. *Neurocomputing*, 449:15–23, 2021.

[Zhou and Shen, 2020] Runwu Zhou and Yi-Dong Shen. End-to-end adversarial-attention network for multi-modal clustering. In *CVPR*, pages 14619–14628, 2020.