Gated Variational Graph Autoencoders as Experts with Competition and Consensus for Multi-view Clustering

Zhaoliang Chen[1], William K. Cheung[1], Hong-Ning Dai[1], Byron Choi[1], Jiming Liu[1]

[1]Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China
chenzl23@outlook.com, {william, henrydai, choi, jiming}@comp.hkbu.edu.hk

Abstract

Multi-view clustering has been found useful to leverage diverse data sources for accurate and robust underlying data representations. It typically relies on effectively integrating the latent features from different views through allocating weights while simultaneously mining their specificity and consensus information. However, it remains open how to achieve a more fine-grained sample-level weight allocation for promoting view-specific information fusion and view-shared consensus. To address this problem, we propose a novel multi-expert learning framework named Gated Variational Graph AutoEncoder with Competition and Consensus (GVGAE-C^2). In particular, it employs multiple view-specific Variational Graph AutoEncoders (VGAEs) as experts to capture the latent features from their own views. Furthermore, we design a fine-grained structure-aware gating network, which dynamically computes sample-level weights based on the proposed structure-aware quality evaluation on each expert, thus facilitating competition among experts. Meanwhile, each expert is trained not only to study its assigned view's specificity features, but also explicitly encouraged to learn consensus-aware features across views. Extensive multi-view clustering experiments on benchmark datasets reveal that GVGAE-C^2 significantly outperforms state-of-the-art methods.

Introduction

Multi-view clustering, as an unsupervised learning method for multi-source data, has received increasing attention in recent years (Sun et al. 2025; Yu et al. 2024b; Li et al. 2024; Lu et al. 2023; Li et al. 2025b). Multi-view data refer to different features (e.g., visual and textual features) describing the same object or varied high-level representations (e.g., distinct feature descriptors for images) of the same sample. These data often encapsulate both consistent information shared across views and view-specific features involving individual knowledge. Aiming to develop multi-view integration strategies, existing studies have investigated methods that simultaneously merge view-specific information and extract consensus representations across different views (Li et al. 2022; Tang et al. 2022; Li et al. 2025a).

To achieve the integration of diverse latent features across different views, prior studies predominantly assign static weights (Wang et al. 2022; Wu et al. 2023) or trainable weights (Wang et al. 2023; Li, Li, and Wang 2020) to latent features or optimization targets at the *view level*. However, view-level weight assignment may overlook data distribution and lead to suboptimal fusion when samples exhibit view-specific heterogeneity, as shown in Figure 1(a). Thus, a *sample-level weight assignment* is more suitable for multi-view learning. A sample-level weighting method can assign weights to each sample rather than the whole view, thereby

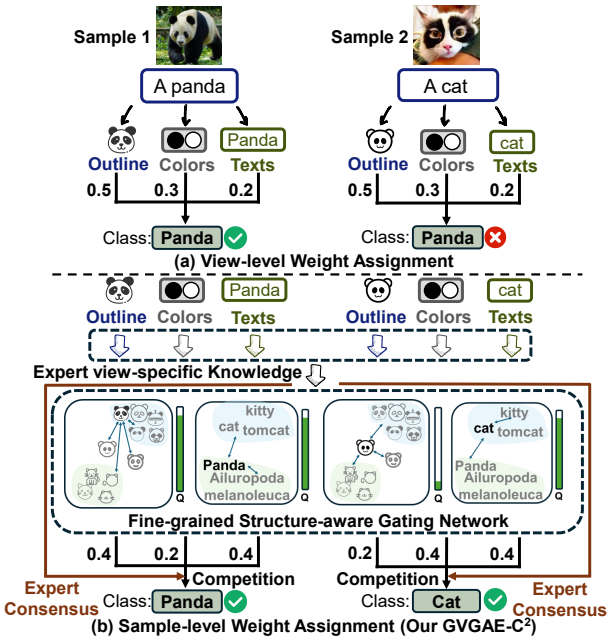

Figure 1: Comparison between *view-level* weight assignment and *sample-level* weight assignment. (a) Prior approaches provide global view-level weights for all samples, incorrectly grouping a black-eyed cat into the panda cluster owing to the high weight assigned to low-quality outline features. (b) Our method leverages the fine-grained structure-aware gating network to analyze the experts' feature quality, providing sample-level weights for competitive view-specific knowledge and achieving expert consensus, thereby correctly grouping the black-eyed cat into the cat cluster with a lower weight assigned to outline features.

allowing *fine-grained* feature integration with consideration of feature diversity between individual samples. It can be achieved by the Mixture of Experts (MoE) framework (Cai et al. 2025), which employs multiple experts to learn from diverse input data or probability distributions (Sutter, Daunhawer, and Vogt 2021; Wu et al. 2024; Hirt et al. 2024). The gating network of MoE can analyze each input from different domains, and then decides to activate the corresponding experts (specialists) with dynamic weights. The MoE function *aligns* with the core objective of fine-grained multi-view feature integration, i.e., assigning *sample-level* weights to the learned multi-view latent features.

Although early efforts have been made to introduce MoE and the gating network into multi-view learning (Zhang et al. 2025; Du et al. 2025), it poses several critical problems to directly applying gating networks in multi-view weight assignment. **Problem 1**: existing methods employ multiple structure-shared experts to learn a specific view, namely, a *one-to-many "view-experts"* framework, which adopts common view-level fusion strategies while fails to utilize the gating network to achieve *sample-level* weight allocation and expert competition. Moreover, this one-to-many "view-experts" structure may train multiple experts on the same view to learn redundant features, resulting in unnecessary computation. **Problem 2**: a critical goal of multi-view learning is to capture both view-specific uniqueness and cross-view consensus (Qin, Feng, and Zhang 2025; Wang et al. 2024a; Sun et al. 2024). Nevertheless, experts may converge to similar features due to the singular shared architecture, thus hindering the joint learning of both individual and consensus information. In summary, the investigation of learning both view-specific knowledge and consensus information among experts is still under-explored.

An effective way to solve **Problem 1** is constructing a *one-to-one "view-expert"* framework, where each subnetwork becomes an expert responsible for its individual view, providing own knowledge and judgement from its perspective. In this way, the model inherently promotes *fine-grained sample-level* expert/view weight assignment via the gating network, and also reduces the computational redundancy. However, a gating network cannot explicitly determine expert weights solely by assessing multi-view features since it only analyzes *domain-specific feature subsets* and activates the corresponding experts responsible for input features. By contrast, the multi-view learning model needs to analyze *all types of known features* simultaneously and assign weights according to a certain criterion. To this end, it is a necessity to enable the gating network to assess the quality of expert outputs (i.e., specialties) at the sample level and assign weights to experts accordingly, although it poses **Challenge 1**: *how to characterize the expert quality and build a new gating network accordingly.* Moreover, although high-quality experts are expected to receive higher weights, knowledge from relatively low-quality experts cannot be completely excluded because they may be complementary to other experts' with a performance contribution after feature integration. This leads to **Challenge 2**: *how to ensure that high-quality experts receive higher weights, while avoiding the complete ignorance of information from lower-quality experts.* Furthermore, to tackle **Problem 2**, experts' outputs need to be divided into: view-specific knowledge and consensus knowledge. Nevertheless, it is non-trivial to achieve this goal due to the identical network structures of experts. Therefore, it raises **Challenge 3**: *how to maintain the uniqueness of each expert's view-specific knowledge while also reaching consensus among experts.*

To address the above challenges, we propose a novel multi-view clustering framework dubbed Gated Variational Graph AutoEncoder with Competition and Consensus (GVGAE-C^2), where experts compete to output view-specific information while simultaneously reaching partial consensus. In a nutshell, Figure 1(b) elaborates on our model. To tackle **Challenge 1**, we meticulously design a fine-grained structure-aware gating network to evaluate the quality of expert outputs in an unsupervised manner. As shown in Figure 1(b), GVGAE-C^2 leverages the structure-aware gating network to yield different sample-level weights based on *structure quality*. Intuitively, neighboring samples tend to be closer while non-neighboring samples are to be more distant. This property can be effectively captured by the graph structure among samples, thereby making it easier for downstream clustering algorithms to delineate clusters. To this end, we employ the Variational Graph Autoencoder (VGAE) as the expert network. Motivated by contrastive learning (Pan and Kang 2021; Su et al. 2025), we design a fine-grained contrastive quality evaluation method based on the graph reconstruced by VGAE. For each expert, a higher quality indicates that output features can effectively aggregate similar neighbors and push dissimilar neighbors further apart. With the proposed quality evaluation approach, we can further solve **Challenge 2** via designing a novel expert allocation loss. For each sample, the first term of the loss aims to assign larger weights to high-quality experts, while the second term enforces that the global average weight allocated to each expert remains similar and avoids the model collapsing into a few high-quality experts. As for **Challenge 3**, we design a discrepancy loss that encourages maximal independence among view-specific information, while also maximizing the similarity among consensus representations. To accurately capture relationships between features, we project features onto a Hilbert space, taking advantage of its expressive power of high-dimensional representations. In summary, our contributions are highlighted as follows:

- We present a novel one-to-one "view-expert" framework with parallel VGAEs and fine-grained structure-aware gating networks. In this way, trainable weights are assigned to different experts based on structure-aware expert quality at the sample level, thereby integrating view-specific knowledge in an expert competition manner.
- We devise a new expert allocation loss combined with the expert quality evaluation, in which higher weights are assigned to high-quality experts at the sample level, while balancing the global average weights across all experts to avoid the collapse of experts.
- We design an optimization mechanism with the expert discrepancy loss to maximize the individual knowledge diversity and minimize the consensus feature variation

across experts, thus further enhancing the competition and consensus among experts.

Related Work

Multi-view clustering aims to improve clustering accuracy by extracting complementary and consensus representations from multiple heterogeneous data sources (Yu et al. 2024a; Lan et al. 2025; Yuan et al. 2025). Compared to single-view clustering, the core hypothesis aims to integrate consensus and complementary information across views for enhancing the robustness and discriminative features (Yu et al. 2024c; Li et al. 2023; Yu et al. 2025). Thus, how to assign weights to different views becomes a critical research problem in multi-view learning. On one hand, prior methods calculate fixed weights or average weights (Chen et al. 2025; Wu et al. 2023) for multi-view latent embeddings. On the other hand, trainable weight allocation enables the model to better adapt to complex data (Cui et al. 2024; Hu, Lou, and Ye 2022). These weights are generally optimized by trainable network parameters (Li, Li, and Wang 2020; Chen et al. 2023b) or traditional optimization methods (Nie, Li, and Li 2016; Chen, Wang, and Lai 2023). However, these models generally assign view-level weights to samples. There is limited research on sample-level weight assignment, while it is crucial to capture the heterogeneity among samples.

Mixture of Experts is a popular architecture designed to scale model capacity while maintaining computational efficiency, with inclusion of several experts and the gating network dynamically routing each input sample (Zhou et al. 2022; Cai et al. 2025). MoE has been a mature tool in multimodal learning, including multimodal VAE exploring both modality-specific and modality-invariant information (Shi et al. 2019; Qiu et al. 2025). As a key module, substantial studies have leveraged the gating network in different fields, including time series analysis (Chen et al. 2024), computer vision (Cao et al. 2023) and natural language processing (Du et al. 2022). Inspired by MoE architectures, some latest studies on multi-view clustering (Zhang et al. 2025; Du et al. 2025) have also developed multiple experts for learning multi-view features integration at the *view level*. These studies usually utilize several experts to learn a specific view, rather than training an expert dedicated to one view. This incomplete design may reduce the expert diversity because these experts share the same structure and input. The gating network only considers input features, resulting in a lack of feature quality analysis under multi-view scenarios. To overcome these limitations, we propose a fine-grained structure-aware gating network to provide a novel solution for sample-level multi-view weight assignment with the expert quality evaluation strategy.

Preliminary and Overview

A collection of multi-view data $\mathcal{X} = \{\mathbf{X}^{(v)} \in \mathbb{R}^{n\times d_v}\}_{v=1}^{V}$ with n samples include V views of different features. Correspondingly, V parallel VGAEs are built to work on these V views, where each VGAE is an expert in its own field. As shown in Figure 2, to utilize these experts to learn underlying multi-view features, we construct V graphs with adjacency matrices $\mathcal{A} = \{\mathbf{A}^{(v)} \in \mathbb{R}^{n\times n}\}_{v=1}^{V}$ based on feature similarities, where each sample is connected to k nearest neighbors. The model is a one-to-one "view-expert" framework capable of learning both individual knowledge and consensus information on each view, which evaluates the quality of the learned features through a quality evaluation function. Next, the model consolidates different experts' individual knowledge through a fine-grained structure-aware gating network, with subsequent integration of consensus information to yield the unique representation, which is then leveraged to perform downstream clustering tasks.

The Proposed Model

Variational Graph Autoencoder as Experts

We build experts based on VGAE (Kipf and Welling 2016), where each VGAE is regarded as a specific expert for a single view. Next, we elaborate on the structure of an individual expert network.

Encoder The 1st layer of the v-th expert encoder is

$$\mathbf{H}^{(v)} = \mathrm{ReLU}\left(\mathrm{GCL}^{(v)}\left(\hat{\mathbf{A}}^{(v)}, \mathbf{X}^{(v)}|\mathbf{W}_0^{(v)}\right)\right), \quad (1)$$

where $\mathrm{GCL}^{(v)}(\hat{\mathbf{A}}^{(v)}, \mathbf{X}^{(v)}|\mathbf{W}_0^{(v)}) = \hat{\mathbf{A}}^{(v)}\mathbf{X}^{(v)}\mathbf{W}_0^{(v)}$ is the graph convolutional layer, and $\hat{\mathbf{A}}^{(v)} = (\tilde{\mathbf{D}}^{(v)})^{-\frac{1}{2}}\tilde{\mathbf{A}}^{(v)}(\tilde{\mathbf{D}}^{(v)})^{-\frac{1}{2}}$ is the renormalized adjacency matrix with self-loop connections. Notably, $\tilde{\mathbf{A}}^{(v)} = \mathbf{A}^{(v)} + \mathbf{I}$ and $[\tilde{\mathbf{D}}^{(v)}]_{ii} = \sum_j [\tilde{\mathbf{A}}^{(v)}]_{ij}$. The 2nd layer of the encoder is

$$\boldsymbol{\mu}^{(v)} = \mathrm{GCL}_{\boldsymbol{\mu}}^{(v)}\left(\hat{\mathbf{A}}^{(v)}, \mathbf{H}^{(v)}|\mathbf{W}_{\boldsymbol{\mu}}^{(v)}\right), \quad (2)$$

$$\log\boldsymbol{\sigma}^{(v)} = \mathrm{GCL}_{\boldsymbol{\sigma}}^{(v)}\left(\hat{\mathbf{A}}^{(v)}, \mathbf{H}^{(v)}|\mathbf{W}_{\boldsymbol{\sigma}}^{(v)}\right), \quad (3)$$

where $\boldsymbol{\mu}^{(v)}$ refers to the matrix of mean vectors, and $\log\boldsymbol{\sigma}^{(v)}$ is derived from another graph convolutional layer. With encoders, the stochastic latent variable $\mathbf{Z}^{(v)}$ is obtained by $q(\mathbf{Z}^{(v)}|\mathbf{X}^{(v)}, \mathbf{A}^{(v)}) = \prod_{i=1}^{n} q(\mathbf{z}_i^{(v)}|\mathbf{X}^{(v)}, \mathbf{A}^{(v)})$, where $q(\mathbf{z}_i^{(v)}|\mathbf{X}^{(v)}, \mathbf{A}^{(v)}) = \mathcal{N}(\mathbf{z}_i^{(v)}|\boldsymbol{\mu}_i^{(v)}, \mathrm{diag}((\boldsymbol{\sigma}_i^{(v)})^2))$. We use the reparameterization trick to resample latent variable $\mathbf{Z}^{(v)}$ from $\boldsymbol{\mu}^{(v)}$ and $\boldsymbol{\sigma}^{(v)}$, i.e., $\mathbf{Z}^{(v)} = \boldsymbol{\mu}^{(v)} + \boldsymbol{\sigma}^{(v)} \odot \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, 1)$ is the Gaussian noise.

Decoder With $\mathbf{Z}^{(v)}$, the decoder is defined as $p(\mathbf{A}^{(v)}|\mathbf{Z}^{(v)}) = \prod_{i=1}^{n}\prod_{j=1}^{n} p([\mathbf{A}^{(v)}]_{ij}|\mathbf{z}_i^{(v)}, \mathbf{z}_j^{(v)})$, where

$$p\left([\mathbf{A}^{(v)}]_{ij} = 1|\mathbf{z}_i^{(v)}, \mathbf{z}_j^{(v)}\right) = \varphi(\mathbf{A}_{\mathrm{rec}}^{(v)}) = \varphi(\mathbf{z}_i^{(v)}{\mathbf{z}_j^{(v)}}^T). \quad (4)$$

Herein, $\varphi(\cdot)$ is the sigmoid activation function. The decoder aims to reconstruct the adjacency matrix from $\mathbf{Z}^{(v)}$.

Loss The graph reconstruction loss of the v-th expert is

$$\mathcal{L}_{\mathrm{rec}}^{(v)} = \mathbb{E}_{q(\mathbf{Z}^{(v)}|\mathbf{X}^{(v)}, \mathbf{A}^{(v)})}[\log p(\mathbf{A}^{(v)}|\mathbf{Z}^{(v)})]. \quad (5)$$

Meanwhile, to make the empirical distribution approximate the prior distribution, we use the Kullback-Leibler divergence function to minimize the distance between the following distributions, i.e.,

$$\mathcal{L}_{\mathrm{kl}}^{(v)} = -\mathrm{KL}\left[q(\mathbf{Z}^{(v)}|\mathbf{X}^{(v)}, \mathbf{A}^{(v)})\|p(\mathbf{Z}^{(v)})\right], \quad (6)$$

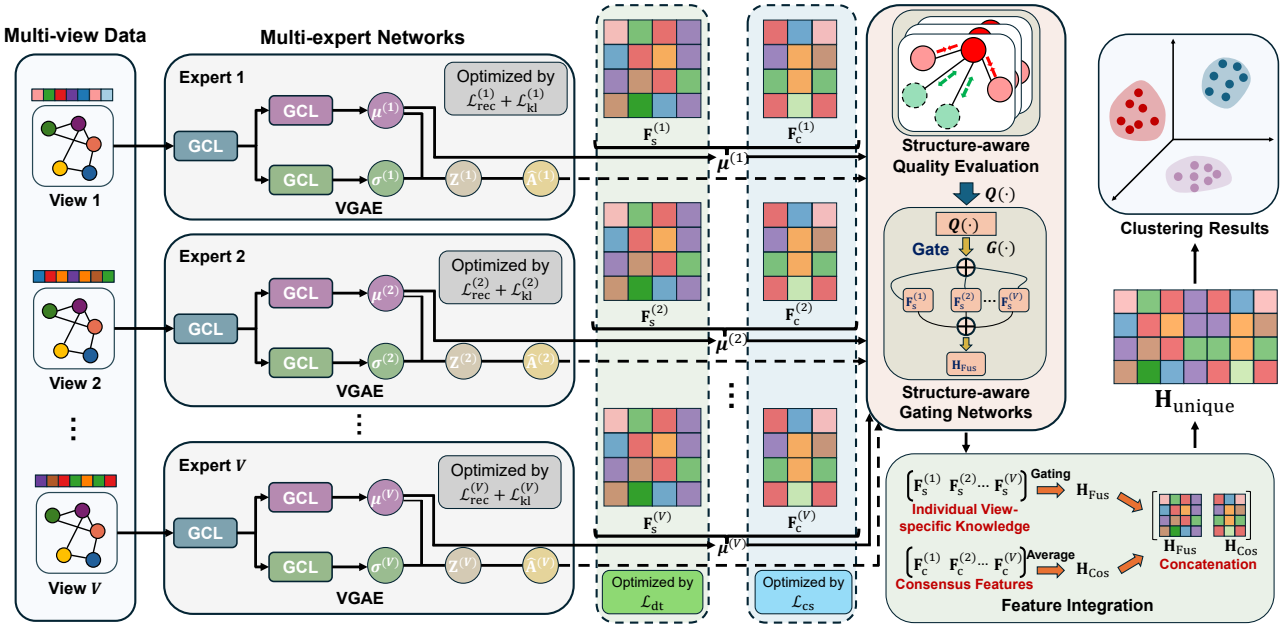

Figure 2: Framework of the proposed GVGAE-C^2.

where a Gaussian prior distribution $p(\mathbf{Z}^{(v)}) = \prod_i p(\mathbf{z}_i^{(v)}) = \prod_i \mathcal{N}(\mathbf{z}_i^{(v)}|0,\mathbf{I})$ is adopted. Thus, the Kullback-Leibler divergence loss can be rewritten as $\mathcal{L}_{\text{kl}}^{(v)} = \frac{1}{2n^2}\sum_{i,k=1}^{n,d}([\boldsymbol{\mu}^{(v)}]_{ik}^2 + [\boldsymbol{\sigma}^{(v)}]_{ik}^2 - 2\log[\boldsymbol{\sigma}^{(v)}]_{ik} - 1)$.

Expert Competition and Consensus

Each expert aims to learn two parts of latent embeddings from their own view: 1) view-shared consensus information and 2) competitive individual knowledge. To achieve this goal, we divide the latent embeddings $\boldsymbol{\mu}^{(v)}$ learned by each expert into two submatrices, i.e., $\boldsymbol{\mu}^{(v)} = \text{Concat}([\mathbf{F}_s^{(v)}, \mathbf{F}_c^{(v)}])$, where $\mathbf{F}_s^{(v)} \in \mathbb{R}^{n\times d_s}$ denotes the expert-specific knowledge and $\mathbf{F}_c^{(v)} \in \mathbb{R}^{n\times d_c}$ contains the consensus information across experts, thereby $\boldsymbol{\mu}^{(v)} \in \mathbb{R}^{n\times(d_s+d_c)}$. With these two submatrices, we build the following expert competition and consensus mechanism, which is realized by the structure-aware gating network and the expert discrepancy loss.

Fine-grained Structure-aware Gating Network To achieve the fine-grained structure-aware gating network, we design a structure-aware quality evaluation method. For the node i, we define the contrastive quality of features learned by the v-th expert as

$$[Q(\mathbf{A}_{\text{rec}}^{(v)})]_i = \frac{1}{k}\sum_{j\in\mathcal{N}_k(i)} -\log\frac{\exp([\mathbf{A}_{\text{rec}}^{(v)}]_{ij})}{\sum_{k\neq i}\exp([\mathbf{A}_{\text{rec}}^{(v)}]_{ik})}, \quad (7)$$

where $\mathbf{A}_{\text{rec}}^{(v)}$ is the adjacency matrix reconstructed by the v-th expert, $\mathcal{N}_k(i)$ indicates the k nearest neighbors of the node i. A smaller value of $[Q(\mathbf{A}_{\text{rec}}^{(v)})]_i$ indicates a higher quality of features, since a high-quality representation should make neighbors of node i closer and push non-neighbors apart. Next, we design the structure-aware gating network as

$$\begin{aligned} R\left(\mathbf{A}_{\text{rec}}\right) = &\text{Softmax}\left(Q(\mathbf{A}_{\text{rec}})\mathbf{W}_{\text{gate}}\right. \\ &\left. + \epsilon\cdot\text{Softplus}\left(Q(\mathbf{A}_{\text{rec}})\mathbf{W}_{\text{noise}}\right)\right), \end{aligned} \quad (8)$$

where $Q(\mathbf{A}_{\text{rec}}) = 1-[Q(\mathbf{A}_{\text{rec}}^{(1)}); Q(\mathbf{A}_{\text{rec}}^{(2)}); \cdots; Q(\mathbf{A}_{\text{rec}}^{(V)})] \in \mathbb{R}^{n\times V}$ is the quality matrix of all experts, $\mathbf{W}_{\text{gate}} \in \mathbb{R}^{V\times V}$ and $\mathbf{W}_{\text{noise}} \in \mathbb{R}^{V\times V}$ are trainable weights. To highlight high-quality experts, we perform min-max normalization on $Q(\mathbf{A}_{\text{rec}})$. With $R\left(\mathbf{A}_{\text{rec}}\right)$, the gating function is defined as:

$$G\left(\mathbf{A}_{\text{rec}}\right) = \text{Softmax}\left(R\left(\mathbf{A}_{\text{rec}}\right)\right) \in \mathbb{R}^{n\times V}. \quad (9)$$

Finally, the sample-level fusion of the expert-specific features is conducted by $[\mathbf{H}_{\text{Fus}}]_i = \sum_{v=1}^{V}[G\left(\mathbf{A}_{\text{rec}}\right)]_{[i,v]} \odot [\mathbf{F}_s^{(v)}]_i$, where V is the number of experts. With the fine-grained structure-aware gating network, the model can adaptively assign lower weights to experts with inferior output quality.

Although structure-aware gating networks may cause the model to always favor a few high-quality experts, low-quality experts may also contain complementary information. The ignorance of their contributions can result in the aggregated features failing to fully exploit the information from different views. To avoid the multi-expert model collapsing into one with only a few high-quality experts, we design a new expert allocation loss. Meanwhile, to achieve expert competition, we also assume that high-quality experts should be assigned higher weights. Thus, we define the ex-

pert allocation loss as:

$$\mathcal{L}_{\text{alloc}}(\mathbf{A}_{\text{rec}}) = \frac{1}{n}\sum_{i=1}^{n}\exp(-\sum_{v=1}^{V}[G(\mathbf{A}_{\text{rec}})]_{iv}[Q(\mathbf{A}_{\text{rec}})]_{iv}) + \pi \text{CV}(\text{I}(\mathbf{A}_{\text{rec}}))^2, \tag{10}$$

where $\text{I}(\mathbf{A}_{\text{rec}})$ evaluates the importance of experts via $\text{I}(\mathbf{A}_{\text{rec}}) = \sum_{\mathbf{a}_{\text{rec}} \in \mathbf{A}_{\text{rec}}} G(\mathbf{a}_{\text{rec}})$. In Eq. (10), π is the balance parameter and $\text{CV}(\cdot)$ indicates the coefficient of variation.

Optimization on Expert Discrepancy Next, we elaborate on the optimization target w.r.t. expert feature discrepancy. We assume that consensus features $\{\mathbf{F}_c^{(v)}\}_{v=1}^V$ learned by different experts should be similar in the high-dimensional space, thereby providing a richer geometric structure for capturing complex feature dependencies. Thus, we develop a consensus loss $\mathcal{L}_{cs}$ to encourage all experts to collaborate in learning an approximate feature in high-dimensional Hilbert space. We adopt the Hilbert-Schmidt Independence Criterion (HSIC) (Gretton et al. 2005) for its ability to measure nonlinear dependencies between high-dimensional features derived from variables $\mathbf{A}$ and $\mathbf{B}$. HSIC is defined as $\text{HSIC}(\mathbf{A}, \mathbf{B}, \mathcal{F}, \mathcal{R}) = \|\mathbf{C}_{\mathbf{AB}}\|_{\text{HS}}^2$, where $\mathcal{F}$ and $\mathcal{R}$ are Reproducing Kernel Hilbert Spaces (RKHS) associated with kernel functions $k^{(a)} : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ and $k^{(b)} : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$. $\mathbf{C}_{\mathbf{AB}}$ indicates the cross-covariance operator $\mathbf{C}_{\mathbf{AB}} : \mathcal{R} \to \mathcal{F}$. Given any consensus feature pair $(\mathbf{F}_c^{(a)}, \mathbf{F}_c^{(b)})$ from experts a and b, the empirical HSIC is:

$$\text{HSIC}\left(\mathbf{F}_c^{(a)}, \mathbf{F}_c^{(b)}\right) = \frac{\text{Tr}\left(\mathbf{K}^{(a)}\mathbf{W}\mathbf{K}^{(b)}\mathbf{W}\right)}{(n-1)^2}, \tag{11}$$

where $[\mathbf{K}^{(a)}]_{ij} = k^{(a)}([\mathbf{F}_c^{(a)}]_i, [\mathbf{F}_c^{(a)}]_j)$ and $[\mathbf{K}^{(b)}]_{ij} = k^{(b)}([\mathbf{F}_c^{(b)}]_i, [\mathbf{F}_c^{(b)}]_j)$ are kernel matrices. $\mathbf{W} = \mathbf{I}_n - \frac{1}{n}(\mathbf{1}\mathbf{1}^T)$ is the centering matrix, where $\mathbf{I}_n$ is the identity matrix and $\mathbf{1}$ is the one vector. In this paper, we use the Gaussian kernel defined as $k([\mathbf{F}_c]_i, [\mathbf{F}_c]_j) = \exp(-\frac{\|[\mathbf{F}_c]_i - [\mathbf{F}_c]_j\|^2}{2\xi^2})$, where ξ is the bandwidth hyperparameter. Particularly, $\text{HSIC}(\mathbf{F}_c^{(a)}, \mathbf{F}_c^{(b)}) = 0$ reveals that $\mathbf{F}_c^{(a)}$ and $\mathbf{F}_c^{(b)}$ are independent and orthogonal, i.e., $\mathbf{F}_c^{(a)} \perp \mathbf{F}_c^{(b)}$. Thus, we need to maximize $\text{HSIC}(\mathbf{F}_c^{(a)}, \mathbf{F}_c^{(b)})$ to get similar consensus features $\mathbf{F}_c^{(a)}$ and $\mathbf{F}_c^{(b)}$ in RKHS for any a, b.

Based on HSIC, we next define the expert consensus loss function to minimize the independency between consensus features generated by different experts. We first define the Normalized HSIC (NHSIC) between consensus features of experts a and b as $\text{NHSIC}(\mathbf{F}_c^{(a)}, \mathbf{F}_c^{(b)}) = \frac{\text{HSIC}\left(\mathbf{F}_c^{(a)}, \mathbf{F}_c^{(b)}\right)}{\sqrt{\text{HSIC}\left(\mathbf{F}_c^{(a)}, \mathbf{F}_c^{(a)}\right) \cdot \text{HSIC}\left(\mathbf{F}_c^{(b)}, \mathbf{F}_c^{(b)}\right)}}$, where $\text{NHSIC}(\mathbf{F}_c^{(a)}, \mathbf{F}_c^{(b)})$ ranges in $[0, 1]$. Thus, we define the consensus loss as

$$\mathcal{L}_{cs} = \sum_{a=1}^{V}\sum_{b=(a+1)}^{V}\left(1 - \text{NHSIC}\left(\mathbf{F}_c^{(a)}, \mathbf{F}_c^{(b)}\right)\right). \tag{12}$$

Therefore, we can minimize $\mathcal{L}_{cs}$ to obtain similar consensus features $\{\mathbf{F}_c^{(v)}\}_{v=1}^V$ across all experts. Because the learned consensus features should be similar after optimization, we can directly obtain the unique consensus features aggregated from all experts with an average weighted sum operation, i.e., $\mathbf{H}_{\text{Cos}} = \frac{1}{V}\sum_{v=1}^{V}\mathbf{F}_c^{(v)}$.

Moreover, to enhance the diversity of experts, we assume that expert-specific knowledge $\{\mathbf{F}_s^{(v)}\}_{v=1}^V$ are independent. Hence, we also minimize the expert diversity loss as

$$\mathcal{L}_{dt} = \sum_{a=1}^{V}\sum_{b=(a+1)}^{V}\text{NHSIC}\left(\mathbf{F}_s^{(a)}, \mathbf{F}_s^{(b)}\right). \tag{13}$$

Expert Aggregation Eventually, we concatenate the learned features from expert-specific knowledge fusion and expert consensus fusion to achieve the unique embeddings of nodes. Namely,

$$\mathbf{H}_{\text{unique}} = \text{Softmax}\left(\text{Concat}([\mathbf{H}_{\text{Fus}}, \mathbf{H}_{\text{Cos}}])\right). \tag{14}$$

Optimization Target

The proposed model aims to optimize the following loss:

$$\mathcal{L} = \sum_{v=1}^{V}(\mathcal{L}_{\text{rec}}^{(v)} + \mathcal{L}_{\text{kl}}^{(v)}) + \lambda\mathcal{L}_{\text{alloc}} + \gamma(\mathcal{L}_{cs} + \mathcal{L}_{dt}), \tag{15}$$

where λ and γ are balance hyperparameters. The 1st term is the VGAE training loss for promoting expert knowledge learning, the 2nd term is the expert allocation loss and the 3rd term is the expert discrepancy loss. With the learned node features $\mathbf{H}_{\text{unique}}$, we conduct the downstream clustering via the K-Means algorithm.

Experimental Analysis

Experimental Settings

Datasets. We leverage six multi-view datasets to comprehensively evaluate GVGAE-C^2 and other baselines from different domains, including image data (ALOI, MNIST, Caltech101-20), text-based data (WebKB) and multivariate time series data (Wafer and LP1).

Compared Baselines. To validate the superiority of the proposed GVGAE-C^2, we compare it with the following multi-view clustering baselines: K-means, DSMVC (Tang and Liu 2022), FastMICE (Huang, Wang, and Lai 2023), FSMSC (Chen et al. 2023c), CVCL (Chen et al. 2023a), SMVAGC-SF (Wang et al. 2024b), RCAGL (Liu et al. 2024) and DMVC-CE (Zhang et al. 2025).

Experimental Results

Baseline Comparison. Table 1 reports the performance (ACC, Purity and F-score) comparison between our GVGAE-C^2 and baselines. All models are run five times and the average results are recorded. From experiments, we can observe that GVGAE-C^2 consistently outperforms all baselines in most metrics and datasets, including MoE-based multi-view clustering model DMVC-CE, thereby

Methods	Metrics	ALOI	MNIST	Caltech101-20	WebKB	Wafer	LP1
K-means	ACC	49.62 (1.57)	24.95 (1.45)	32.28 (1.46)	51.78 (9.37)	52.74 (1.24)	37.95 (4.52)
	Purity	51.34 (0.95)	32.16 (1.79)	60.18 (0.85)	58.65 (9.52)	**89.32 (0.00)**	42.50 (4.14)
	F-score	42.55 (1.12)	21.89 (1.39)	27.14 (1.28)	42.73 (5.27)	62.03 (4.31)	28.47 (2.44)
DSMVC	ACC	87.58 (2.41)	24.44 (1.91)	32.74 (0.98)	33.87 (3.59)	54.27 (2.35)	50.04 (2.41)
	Purity	87.78 (2.75)	30.68 (1.45)	65.54 (1.07)	52.87 (3.35)	88.32 (0.00)	63.64 (2.58)
	F-score	85.78 (1.37)	21.77 (1.28)	35.78 (0.90)	30.89 (2.09)	62.98 (2.33)	47.72 (4.97)
FastMICE	ACC	73.24 (6.48)	81.64 (4.23)	42.37 (2.05)	55.13 (3.59)	53.32 (3.55)	67.95 (7.25)
	Purity	74.49 (6.10)	**87.00 (2.08)**	76.63 (1.32)	67.70 (1.92)	**89.32 (0.00)**	70.68 (3.54)
	F-score	59.56 (8.10)	78.01 (3.70)	**60.57 (3.03)**	55.26 (3.12)	62.10 (3.74)	56.95 (3.63)
FSMSC	ACC	68.80 (4.87)	56.63 (0.19)	44.57 (0.07)	53.48 (0.00)	56.04 (0.00)	37.05 (3.82)
	Purity	71.92 (3.21)	67.69 (0.16)	**78.80 (0.07)**	67.83 (0.00)	**89.32 (0.00)**	40.68 (1.24)
	F-score	57.91 (3.31)	53.37 (0.19)	80.16 (0.11)	56.56 (0.00)	63.01 (0.00)	28.17 (1.62)
CVCL	ACC	88.77 (0.25)	78.90 (0.00)	33.36 (0.16)	44.60 (1.07)	65.98 (0.00)	43.18 (0.00)
	Purity	80.53 (0.41)	83.00 (0.00)	33.36 (0.16)	68.85 (1.36)	**89.32 (0.00)**	43.18 (0.00)
	F-score	83.70 (0.37)	70.00 (0.00)	27.98 (0.13)	39.15 (0.62)	68.22 (0.00)	40.59 (0.00)
SMVAGC-SF	ACC	88.00 (7.13)	83.60 (4.68)	37.33 (2.95)	42.23 (1.53)	55.38 (2.31)	39.32 (3.65)
	Purity	89.04 (5.72)	85.21 (2.91)	65.74 (0.91)	51.17 (0.83)	**89.32 (0.00)**	42.50 (2.06)
	F-score	83.70 (5.71)	75.84 (3.58)	32.97 (2.61)	38.49 (1.49)	60.14 (3.51)	29.32 (1.26)
RCAGL	ACC	54.22 (0.00)	65.35 (0.00)	48.16 (0.00)	60.75 (0.00)	44.96 (3.17)	38.64 (0.00)
	Purity	87.49 (0.00)	66.10 (0.00)	69.28 (0.00)	64.15 (0.00)	87.14 (4.26)	59.09 (0.00)
	F-score	79.06 (0.00)	49.24 (0.00)	57.42 (0.00)	49.15 (0.00)	44.50 (2.37)	37.60 (0.00)
DMVC-CE	ACC	90.44 (0.92)	76.78 (7.04)	44.32 (2.34)	32.87 (0.70)	65.71 (4.91)	40.23 (1.16)
	Purity	89.34 (1.24)	80.46 (6.14)	70.63 (3.18)	49.48 (1.89)	**89.32 (0.00)**	30.19 (1.43)
	F-score	89.79 (2.23)	77.82 (7.03)	50.22 (4.13)	35.08 (1.12)	71.71 (2.89)	20.81 (0.65)
GVGAE-C^2	ACC	**91.99 (3.64)**	**86.00 (0.13)**	**48.44 (3.13)**	**63.48 (3.22)**	**68.29 (0.00)**	**75.91 (4.09)**
	Purity	**92.12 (3.38)**	86.00 (0.13)	75.22 (0.97)	**72.43 (1.19)**	**89.32 (0.00)**	**78.64 (2.08)**
	F-score	**91.47 (4.95)**	**86.90 (0.14)**	53.25 (3.00)	**60.69 (2.51)**	**73.49 (0.00)**	**74.58 (6.00)**

Table 1: Clustering performance comparison (%) with the best metrics in **bold** and the second best metrics underlined.

Datasets	Original	w/o QE	w/o GN	w/o QE & GN
Wafer	68.29	63.02	64.55	62.95
LP1	75.91	66.81	69.09	59.09

Table 2: Ablation study (ACC%) w.r.t. structure-aware Quality Evaluation (QE) and Gating Network (GN).

demonstrating significant performance gains. This validates that sample-level weight assignment based on the fine-grained structure-aware gating network, expert competition as well as consensus mechanisms, achieves remarkable performance improvements compared to existing models.

Ablation Study. We conduct an ablation study to evaluate the impact of the fine-grained structure-aware gating network. When the structure-aware quality evaluation method is removed, the input to the gating network is replaced by node features. When the gating network is removed, the model directly assigns weights based on quality scores. When the entire module is removed, the model integrates expert features with trainable view-level weights. As reported by Table 2, ablation results reveal the decreased accuracy when removing the structure-aware quality evaluation, the gating network or both of them. This result validates the effectiveness of the proposed module, demonstrating the necessity of the fine-grained sample-level weight assignment.

Parameter Sensitivity. Figure 3 demonstrates the parameter sensitivity analysis w.r.t. two critical hyperparameters: λ for the expert allocation loss and γ for the expert discrepancy loss. The experimental results clearly demonstrate that excessively large hyperparameter values can result in degraded performance in most cases. In general, GVGAE-C^2 achieves satisfactory performance when γ is set between 0.1 and 1, whereas yields better results when λ does not exceed 1. These results suggest that suitable expert assignment and discrepancy losses contribute to improved clustering performance. However, excessively high λ and γ values may lead the model to overlook the expert VGAE loss, thereby resulting in the performance decline.

View-specific and Consensus Information Visualization. To visualize the high-dimensional consensus and view-specific structures captured by experts, we transform these features into affinity graphs (i.e., $\mathbf{A} = \mathbf{F}\mathbf{F}^T$). As shown in

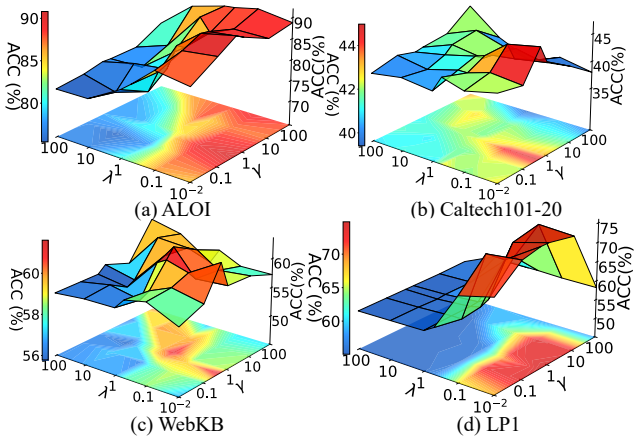

Figure 3: Parameter sensitivity (ACC%) w.r.t. λ and γ.

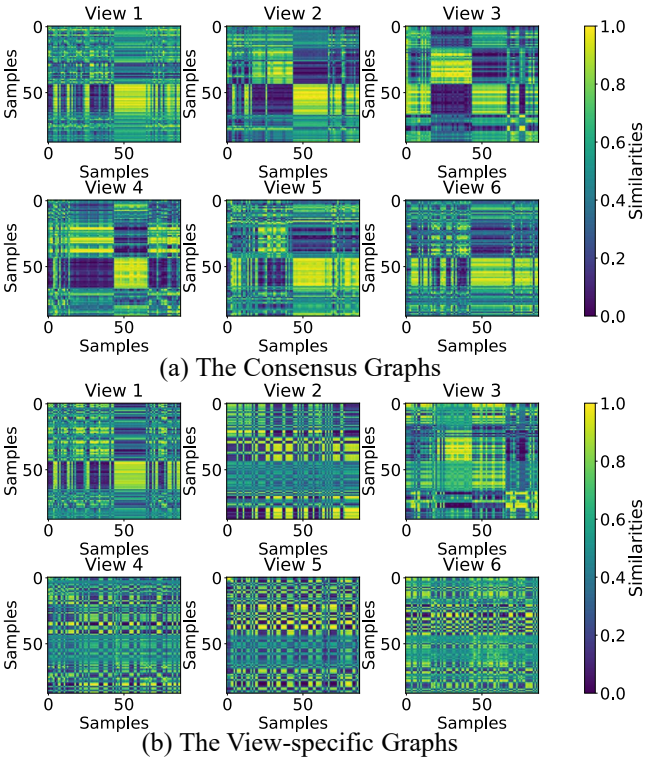

Figure 4: Visualization of consensus graphs and view-specific graphs on LP1 dataset.

Figure 4, although the consensus information learned by different experts exhibits some minor differences, it is overall similar. In contrast, the view-specific knowledge captured by different experts shows notable distinctions. This validates that our discrepancy optimization mechanism effectively explores the consensus among experts while simultaneously enabling each expert to study diverse individual knowledge.

Expert Quality Visualization. Figure 5 visualizes the feature quality learned by experts. These two representative datasets reveal two distinct quality distributions: (1) For the

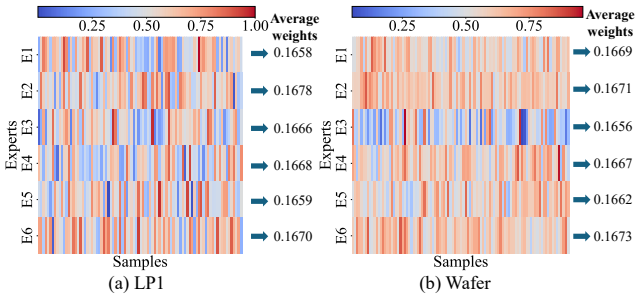

Figure 5: Visualization of expert quality and global average expert weights on LP1 and Wafer datasets.

Datasets	w/o E1	w/o E2	w/o E3	w/o E4	w/o E5	w/o E6
LP1	71.36	68.64	66.36	72.95	74.55	72.27
Wafer	56.58	55.81	62.24	54.79	53.85	55.21

Table 3: Clustering ACC% without a specific expert.

LP1 dataset, the quality vectors of experts vary across different samples; (2) For the Wafer dataset, the 3rd expert shows relatively lower quality on most samples compared to others. Nevertheless, in either case, due to the proposed expert allocation loss, low-quality experts are prevented from being completely excluded. The final total average weights assigned to each expert are generally similar, with lower-quality experts receiving relatively lower total weights. Furthermore, we conducted ablation studies on the experts. As shown in Table 3, removing the experts can result in the performance decline. Particularly, on the Wafer dataset, removing the expert with overall lower quality (E3) also leads to a decrease in model performance, though the drop is less significant than when other experts are removed. This indicates that the model still benefits from the complementary features of low-quality experts. Simply removing such experts would also cause notable decrease in performance, which demonstrates the importance of the expert allocation loss.

Conclusion

To achieve the fine-grained sample-level weight assignment for integrating multi-view complementary and consensus features, we propose a novel model named GVGAE-C^2, which employs VGAEs as experts to explore individual knowledge and consensus information across views. In particular, we meticulously design a fine-grained structure-aware gating network, which dynamically assigns sample-level weights to different experts for individual knowledge integration. Furthermore, we develop an expert competition-consensus mechanism to encourage diversity in individual knowledge while minimizing consensus feature divergence. Extensive experiments on six multi-view datasets demonstrate that GVGAE-C^2 achieves state-of-the-art clustering performance, validating the effectiveness of our approach.

Acknowledgments

This work is supported by the Hong Kong Research Grant Council (with Grant No. RIF R2002-20F), and the Seed Funding for Collaborative Research Grants of HKBU (with Grant No. RC-SFCRG/23-24/R2/SCI/06).

References

Cai, W.; Jiang, J.; Wang, F.; Tang, J.; Kim, S.; and Huang, J. 2025. A Survey on Mixture of Experts in Large Language Models. *IEEE Transactions on Knowledge and Data Engineering*, 37(7): 3896–3915.

Cao, B.; Sun, Y.; Zhu, P.; and Hu, Q. 2023. Multi-modal gated mixture of local-to-global experts for dynamic image fusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 23555–23564.

Chen, J.; Mao, H.; Woo, W. L.; and Peng, X. 2023a. Deep multiview clustering by contrasting cluster assignments. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 16752–16761.

Chen, M.; Wang, C.; and Lai, J. 2023. Low-Rank Tensor Based Proximity Learning for Multi-View Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(5): 5076–5090.

Chen, M.-S.; Zhu, X.-R.; Lin, J.-Q.; and Wang, C.-D. 2025. Contrastive Multiview Attribute Graph Clustering With Adaptive Encoders. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4): 7184–7195.

Chen, P.; ZHANG, Y.; Cheng, Y.; Shu, Y.; Wang, Y.; Wen, Q.; Yang, B.; and Guo, C. 2024. Pathformer: Multi-scale Transformers with Adaptive Pathways for Time Series Forecasting. In *The Twelfth International Conference on Learning Representations*.

Chen, Z.; Fu, L.; Yao, J.; Guo, W.; Plant, C.; and Wang, S. 2023b. Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, 95: 109–119.

Chen, Z.; Wu, X.-J.; Xu, T.; and Kittler, J. 2023c. Fast Self-Guided Multi-View Subspace Clustering. *IEEE Transactions on Image Processing*, 32: 6514–6525.

Cui, J.; Li, Y.; Huang, H.; and Wen, J. 2024. Dual Contrast-Driven Deep Multi-View Clustering. *IEEE Transactions on Image Processing*, 33: 4753–4764.

Du, L.; Shi, Y.; Chen, Y.; Li, F.; Zhou, P.; Duan, L.; and Qian, Y. 2025. A dual mixture-of-experts framework for multi-view K-means clustering with view balance and regional sparsity. *Information Fusion*, 103449.

Du, N.; Huang, Y.; Dai, A. M.; Tong, S.; Lepikhin, D.; Xu, Y.; Krikun, M.; Zhou, Y.; Yu, A. W.; Firat, O.; et al. 2022. Glam: Efficient scaling of language models with mixture-of-experts. In *International Conference on Machine Learning*, 5547–5569.

Gretton, A.; Bousquet, O.; Smola, A.; and Schölkopf, B. 2005. Measuring statistical dependence with Hilbert-Schmidt norms. In *International Conference on Algorithmic Learning Theory*, 63–77.

Hirt, M.; Campolo, D.; Leong, V.; and Ortega, J. 2024. Learning multi-modal generative models with permutation-invariant encoders and tighter variational objectives. *Transactions on Machine Learning Research*.

Hu, S.; Lou, Z.; and Ye, Y. 2022. View-Wise Versus Cluster-Wise Weight: Which Is Better for Multi-View Clustering? *IEEE Transactions on Image Processing*, 31: 58–71.

Huang, D.; Wang, C.-D.; and Lai, J.-H. 2023. Fast Multi-View Clustering Via Ensembles: Towards Scalability, Superiority, and Simplicity. *IEEE Transactions on Knowledge and Data Engineering*, 35(11): 11388–11402.

Kipf, T. N.; and Welling, M. 2016. Variational Graph Auto-Encoders. arXiv:1611.07308.

Lan, Y.; Xu, S.; Su, C.; Ye, R.; Peng, D.; and Sun, Y. 2025. Multi-view Hashing Classification. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 2122–2130.

Li, H.; Guo, Y.; You, J.; You, X.; and Ren, Z. 2025a. Graph Proxy Fusion: Consensus Graph Intermediated Multi-View Local Information Fusion Clustering. *IEEE Transactions on Multimedia*, 27: 1736–1747.

Li, S.; Li, W.; and Wang, W. 2020. Co-GCN for Multi-View Semi-Supervised Learning. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 4691–4698.

Li, X.; Pan, Y.; Sun, Y.; Sun, Q.; Sun, Y.; Tsang, I. W.; and Ren, Z. 2025b. Incomplete Multi-View Clustering With Paired and Balanced Dynamic Anchor Learning. *IEEE Transactions on Multimedia*, 27: 1486–1497.

Li, X.; Pan, Y.; Sun, Y.; Sun, Q.; Tsang, I. W.; and Ren, Z. 2024. Fast Unpaired Multi-view Clustering. 4488–4496.

Li, X.; Sun, Y.; Sun, Q.; and Ren, Z. 2023. Consensus Cluster Center Guided Latent Multi-Kernel Clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(6): 2864–2876.

Li, Z.; Tang, C.; Liu, X.; Zheng, X.; Zhang, W.; and Zhu, E. 2022. Consensus Graph Learning for Multi-View Clustering. *IEEE Transactions on Multimedia*, 24: 2461–2472.

Liu, S.; Liao, Q.; Wang, S.; Liu, X.; and Zhu, E. 2024. Robust and Consistent Anchor Graph Learning for Multi-View Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 36(8): 4207–4219.

Lu, Z.; Nie, F.; Wang, R.; and Li, X. 2023. A Differentiable Perspective for Multi-View Spectral Clustering With Flexible Extension. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6): 7087–7098.

Nie, F.; Li, J.; and Li, X. 2016. Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, 1881–1887.

Pan, E.; and Kang, Z. 2021. Multi-view contrastive graph clustering. *Advances in Neural Information Processing Systems*, 34: 2148–2159.

Qin, Y.; Feng, G.; and Zhang, X. 2025. Scalable One-Pass Incomplete Multi-View Clustering by Aligning Anchors. In

Proceedings of the 39th AAAI Conference on Artificial Intelligence, volume 39, 20042–20050.

Qiu, P.; Zhu, W.; Kumar, S.; Chen, X.; Yang, J.; Sun, X.; Razi, A.; Wang, Y.; and Sotiras, A. 2025. Multimodal Variational Autoencoder: A Barycentric View. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, volume 39, 20060–20068.

Shi, Y.; N, S.; Paige, B.; and Torr, P. 2019. Variational Mixture-of-Experts Autoencoders for Multi-Modal Deep Generative Models. In *Advances in Neural Information Processing Systems*, volume 32.

Su, C.; Zheng, H.; Peng, D.; and Wang, X. 2025. DiCA: Disambiguated Contrastive Alignment for Cross-Modal Retrieval with Partial Labels. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, volume 39, 20610–20618.

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

Sun, Y.; Qin, Y.; Li, Y.; Peng, D.; Peng, X.; and Hu, P. 2024. Robust Multi-View Clustering With Noisy Correspondence. *IEEE Transactions on Knowledge and Data Engineering*, 36(12): 9150–9162.

Sutter, T. M.; Daunhawer, I.; and Vogt, J. E. 2021. Generalized Multimodal ELBO. In *Proceedings of the 9th International Conference on Learning Representations*.

Tang, C.; Zheng, X.; Liu, X.; Zhang, W.; Zhang, J.; Xiong, J.; and Wang, L. 2022. Cross-View Locality Preserved Diversity and Consensus Learning for Multi-View Unsupervised Feature Selection. *IEEE Transactions on Knowledge and Data Engineering*, 34(10): 4705–4716.

Tang, H.; and Liu, Y. 2022. 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*, 202–211.

Wang, Q.; Tao, Z.; Gao, Q.; and Jiao, L. 2024a. Multi-View Subspace Clustering via Structured Multi-Pathway Network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5): 7244–7250.

Wang, Q.; Tao, Z.; Xia, W.; Gao, Q.; Cao, X.; and Jiao, L. 2023. Adversarial Multiview Clustering Networks With Adaptive Fusion. *IEEE Transactions on Neural Networks and Learning Systems*, 34(10): 7635–7647.

Wang, S.; Chen, Z.; Du, S.; and Lin, Z. 2022. Learning Deep Sparse Regularizers With Applications to Multi-View Clustering and Semi-Supervised Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9): 5042–5055.

Wang, S.; Liu, X.; Liu, S.; Tu, W.; and Zhu, E. 2024b. Scalable and Structural Multi-View Graph Clustering With Adaptive Anchor Fusion. *IEEE Transactions on Image Processing*, 33: 4627–4639.

Wu, X.; Huang, S.; Wang, W.; Ma, S.; Dong, L.; and Wei, F. 2024. Multi-head mixture-of-experts. *Advances in Neural Information Processing Systems*, 37: 94073–94096.

Wu, Z.; Lin, X.; Lin, Z.; Chen, Z.; Bai, Y.; and Wang, S. 2023. Interpretable graph convolutional network for multi-view semi-supervised learning. *IEEE Transactions on Multimedia*, 25: 8593–8606.

Yu, S.; Dong, Z.; Wang, S.; Wan, X.; et al. 2024a. Towards resource-friendly, extensible and stable incomplete multi-view clustering. In *Proceedings of the 41st International Conference on Machine Learning*, 57415–57440.

Yu, S.; Liu, S.; Wang, S.; Tang, C.; Luo, Z.; Liu, X.; and Zhu, E. 2025. Sparse Low-Rank Multi-View Subspace Clustering With Consensus Anchors and Unified Bipartite Graph. *IEEE Transactions on Neural Networks and Learning Systems*, 36(1): 1438–1452.

Yu, S.; Wang, S.; Dong, Z.; Tu, W.; Liu, S.; Lv, Z.; Li, P.; Wang, M.; and Zhu, E. 2024b. A Non-parametric Graph Clustering Framework for Multi-View Data. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, 16558–16567.

Yu, S.; Wang, S.; Wen, Y.; et al. 2024c. How to Construct Corresponding Anchors for Incomplete Multiview Clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4): 2845–2860.

Yuan, H.; Sun, Y.; Zhou, F.; Wen, J.; Yuan, S.; You, X.; and Ren, Z. 2025. Prototype Matching Learning for Incomplete Multi-View Clustering. *IEEE Transactions on Image Processing*, 34: 828–841.

Zhang, Y.; Cai, J.; Wu, Z.; Wang, P.; and Ng, S.-K. 2025. Mixture of Experts as Representation Learner for Deep Multi-View Clustering. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, volume 39, 22704–22713.

Zhou, Y.; Lei, T.; Liu, H.; Du, N.; Huang, Y.; Zhao, V.; Dai, A. M.; Le, Q. V.; Laudon, J.; et al. 2022. Mixture-of-experts with expert choice routing. *Advances in Neural Information Processing Systems*, 35: 7103–7114.