From Static to Active: Knowledge-Aware Node State Selection in Multi-view Graph Learning

Weiran Liao[1], Jielong Lu[2], Yuhong Chen[3], Shide Du[1], Hongrong Chen[1], Shiping Wang[1]*

[1] College of Computer and Data Science, Fuzhou University, Fuzhou, China
[2] College of Computer Science and Technology, Zhejiang University, Hangzhou, China
[3] Key Laboratory of Multimedia Trusted Perception and Efficient Computing Ministry of Education of China, Xiamen University, Xiamen, China
wrliao777@163.com, jielonglu2022@163.com, yhchen2320@163.com, dushidems@gmail.com, Hongrongchen2023@163.com, shipingwangphd@163.com

Abstract

Multimedia technologies leverage multi-source to alleviate real-world data incompleteness, providing a versatile platform for multi-view learning. Among existing research, graph-based multi-view learning has achieved notable success. However, prior studies always immerse in comprehensive collaboration across all views and nodes to pursue consistency and complementary, which ignore the negative contribution of nodes from low-quality views. To overcome the above limitation, we explore node behavior selection in multi-view dynamic modeling and propose a knowledge-aware multi-view state space model. Specifically, nodes autonomously select either activation sequences or static sequences according to their current knowledge. In the former, we design the mask-based attention mechanism to capture the dynamics of node behaviors. In the latter, we construct a history pool and simulate synaptic signals to regulate the behavioral distribution of nodes. Moreover, the proposed model provides a directional inter-view diffusion equation that selectively propagates information to alleviate interference from low-quality nodes across views. Extensive experiments demonstrate that the proposed model outperforms baselines on multiple benchmarks and achieves significant performance improvement.

Introduction

The maturation of multimedia technologies have generated large amounts of heterogeneous real-world data. Such data typically integrates multiple perspectives and modalities to represent the same entity, commonly named multi-view data. Owing to its strong ability to mine and integrate heterogeneous information, multi-view learning has been widely used in machine learning (Wang et al. 2022a,b; Wang, Zhang, and Zhou 2025), data mining (Wang et al. 2021; Kou et al. 2024; Wang, Zhang, and Zhou 2025), and recommender system (Lin et al. 2023; Li et al. 2024). In recent years, with its outstanding capability in capturing local and global features, graph-based multi-view learning (Chen et al. 2022; Sun et al. 2024; Wu et al. 2024) has gained widespread popularity.

*Corresponding author.

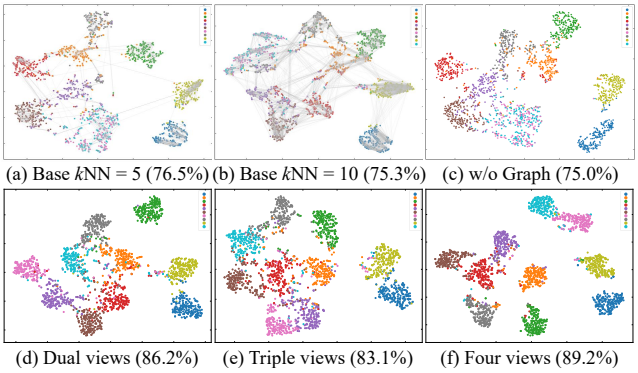


Figure 1: Comparison of classification visualization between graph-based and non-graph methods on the dataset HW for single-view and multi-view representations.

Most existing graph-based multi-view learning (Yang et al. 2022; Wang et al. 2024; Yang et al. 2024) focus on enriching intra-view contextual structures to enhance view representation, independently. However, richer intra-view contextual interactions often involve low-quality nodes in joint optimization, which potentially degrade the overall quality of the view. Fig. 1(a) and Fig. 1(b) illustrate that higher information density within a view strengthens intra-class connectivity, and it also introduces more heterogeneous edges, potentially contaminating same-class nodes. Furthermore, the absence of contextual guidance for view updates leads to excellent handling of features for certain categories, as shown in Fig. 1(c). Although numerous studies have attempted to optimize intra-view data dependencies to alleviate the influence of heterogeneous edges (Chen et al. 2023a,b; Lu et al. 2024a), these methods still depend on static contextual structures and joint training of all nodes to explore consistency and complementarity.

The emergence of Mamba (Gu and Dao 2024) as a new paradigm of State Space Models (SSMs) has demonstrated excellent performance across various fields (Liu et al. 2025; Shaker et al. 2025; He et al. 2025), and offers a novel perspective for advancing dynamic modeling in multi-view

learning. Relevant studies preliminarily explore the application of SSMs in multi-view learning (Zhu et al. 2025b,a), which enhances dynamic modeling within individual views. Nevertheless, these methods still inevitably involve all views in a joint optimization process. As shown in Fig. 1(d), Fig. 1(e) and Fig. 1(f), multi-view collaborative training provides more effective feature representations compared to single-view learning (*e.g.*, Fig. 1(a)), but the inclusion of low-quality views may degrade overall performance (*e.g.*, Fig. 1(e)). Although existing methods (Xu et al. 2020; Zhang et al. 2021; Xu et al. 2023) have achieved some success in optimizing shared representations via attention-weighted fusion and feature concatenation, they remain ground in collaborative fusion over all nodes across views and lack fine-grained, node-level interactions among views. Overall, existing multi-view graph methods face two key limitations: 1) reliance on manually constructed static intra-view structures, which ignore the autonomy of individual nodes; 2) immersion in collaborative training among all view nodes, which inevitably involves low-quality view representations.

In response to these challenges, we break away from traditional multi-view learning that relies on collaborative optimization over all view nodes, instead of model state aggregation, and updates as a node-level autonomous behavior. Building on this perspective, we propose a Knowledge-Aware Multi-view State Space Model (KAMSSM) that interprets node states as intrinsic knowledge representations and achieves dynamic interactions both intra- and inter-view. Specifically, within each view, we integrate a customized environment network with the Gumbel-Softmax distribution, allowing nodes to decide whether to engage in communication. The participating nodes form the activation sequence, while the others remain in the static sequence. Unlike previous graph-based multi-view methods, the activation sequence leverages node states to facilitate node-level contextual dependency mining. Nodes within the activation sequence are further classified into source nodes, receiver nodes, and standard nodes, based on their interaction behavior (Node behavior illustration is provided in Appendix A). Source nodes send messages, receiver nodes collect messages, and standard nodes perform both roles. Accordingly, we propose a mask-based linear attention mechanism to enable state updates for different nodes based on their performed actions, and employ a smoothness loss (Xing et al. 2024) to regularize the node state within the activation sequence. The static sequence designs a history pool and simulates synaptic signals to provide historical guidance for behavior selection and avoid potential key nodes from remaining silent by dynamically adjusting the Gumbel-Softmax distribution.

Furthermore, we introduce a directional inter-view diffusion equation that enables fine-grained, selective state exchange across views. This approach facilitates sufficient information flow while suppressing interference from low-quality views, promoting dynamic balance and effective collaboration across views. The key contribution of this paper can be summarized as:

- We propose a knowledge-aware multi-view state space model, treating node states as knowledge and enabling node-level dynamic interactions intra- and inter-view.
- Redefine intra-view node serialization through the introduction of activation and static sequences to enable behavior-driven state aggregation and update.
- We provide a directional inter-view diffusion equation for selective and efficient cross-view interaction, enabling fine-grained state flow and suppressing interference from low-quality view.

Related Works

In this section, we briefly review the existing graph-based multi-view learning and state space models.

Graph-based Multi-view Learning

Given the powerful ability to capture both local and global information, graph structures have attracted increasing multi-view interest (Xia et al. 2021; Jiang and Liu 2022; Li et al. 2023; Lu et al. 2025). The propagation formula for Graph Neural Networks (GNNs) can be expressed as:

$$\boldsymbol{h}_i^k = \sigma\left(\mathbf{W}^k \cdot \text{Agg}\left(\{\boldsymbol{h}_j^{k-1} \mid j \in \mathcal{N}(i)\}\right) + \boldsymbol{b}^k\right), \quad (1)$$

where $\boldsymbol{h}_i^k$ is the feature of node i at layer k, $\mathcal{N}(i)$ is its neighbor set, $\mathbf{W}^k$ and $\boldsymbol{b}^k$ are the corresponding learnable weights and biases, and $\sigma(\cdot)$ is a nonlinear activation.

Recent advances in graph-based multi-view learning (Cheng et al. 2021; Wen et al. 2023; Lin et al. 2025) have explored various strategies for capturing view-specific structures and enhancing cross-view consistency. To improve structural alignment, some studies constructed subspace anchor graphs and bipartite graphs, to capture hidden features and global structural relationships, leveraging GCNs for embedding learning (Cui et al. 2023; Fu et al. 2024). Several works have focused on topological refinement. (Chen et al. 2023b) employed topology sparsification for view propagation and fusion, while (Chen et al. 2023a) optimized multi-view dependencies via kNN and kFN strategies. In addition, (Wu et al. 2023) combined reconstruction error with Laplacian embedding to model inter-view independence and consistency. (Lu et al. 2024a) and (Zhuang et al. 2024). leveraged energy-based equations and Laplacian smoothing theory, to guide inter-view diffusion and co-optimize graph structure with consistency-aware objectives.

Despite graph-based multi-view learning has made significant progress in capturing data dependencies, it still lacks node-level modeling capabilities and fails to dynamically adapt to changes in node states.

State Space Models

With its strong dynamic modeling capability, State Space Models (SSMs) have been widely applied across various fields (Park et al. 2024; Gao, Qi, and Chen 2024; Zhu et al. 2024). Specifically, SSMs define linear Ordinary Differential Equation (ODE) (Aoki 2013), which maps the input sequence $\mathbf{x}(t) \in \mathbb{R}^L$ to response the output $y(t) \in \mathbb{R}^L$ by incorporating the hidden state $\mathbf{h}(t) \in \mathbb{R}^{N \times L}$, given as

$$\begin{cases} \mathbf{h}'(t) = \mathbf{A}\mathbf{h}(t) + \mathbf{B}\mathbf{x}(t), \\ \mathbf{y}(t) = \mathbf{C}\mathbf{h}(t), \end{cases} \quad (2)$$

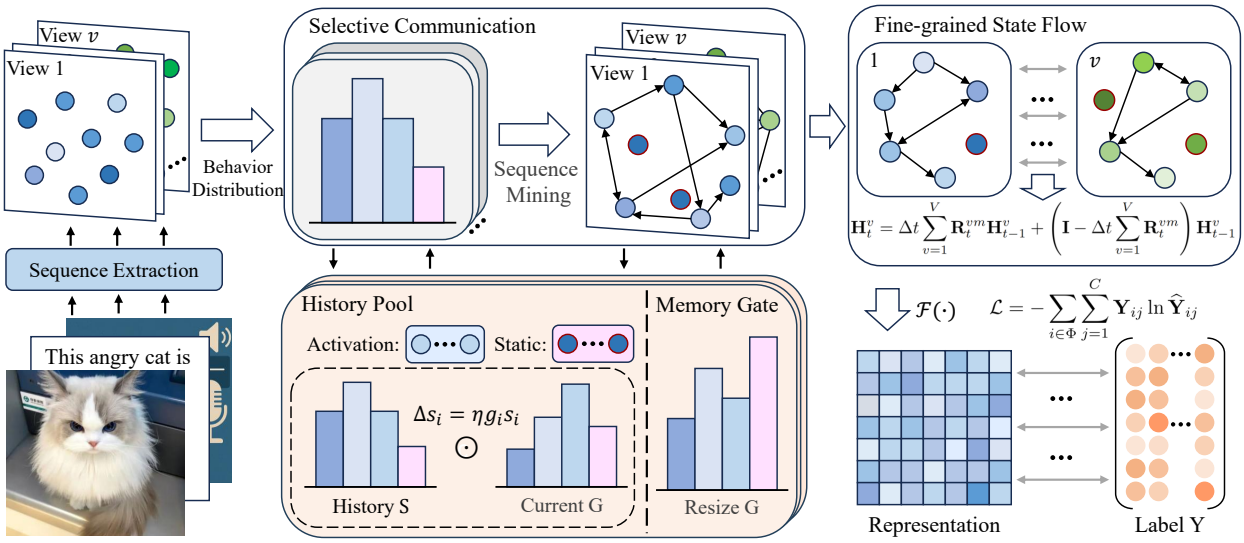


Figure 2: An overview of the proposed knowledge-aware multi-view state space model framework.

where $\mathbf{A} \in \mathbb{R}^{N \times N}$ can be seen as an evolution parameter, $\mathbf{B} \in \mathbb{R}^N$ and $\mathbf{C} \in \mathbb{R}^N$ are the projection parameters. After parameter discretization (Gu et al. 2020), the above formulation can be expressed as:

$$\begin{cases} \mathbf{h}(t) = \bar{\mathbf{A}}\mathbf{h}(t-1) + \bar{\mathbf{B}}\mathbf{x}(t), \\ \mathbf{y}(t) = \mathbf{C}\mathbf{h}(t) + \mathbf{D}\mathbf{x}(t), \end{cases} \tag{3}$$

where

$$\begin{aligned} \bar{\mathbf{A}} &= \exp(\Delta\mathbf{A}), \\ \bar{\mathbf{B}} &= (\Delta\mathbf{A})^{-1}\left(\exp(\Delta\mathbf{A} - \mathbf{I})\right) \cdot \Delta\mathbf{B}. \end{aligned} \tag{4}$$

Building on this foundation, (Gu, Goel, and Re 2022; Gu et al. 2022) propose a structured state space model by applying low-rank correction and diagonalization to the state matrix, which simplifies computation and improves efficiency. The emergence of Mamba as a new paradigm in state space models, which introduces a novel selection mechanism (Gu and Dao 2024) and gains wide attention across various research fields due to its strong dynamic modeling capabilities. (Behrouz and Hashemi 2024) propose the Graph Mamba framework, which encodes subgraph sequences and adopts bidirectional scan, revealing the promise of SSMs for graph-structured data. (Zhu et al. 2025b) incorporate Mamba into multi-view learning to adaptively select and enhance view representations. TMFN leverages Mamba to implement a selective fusion mechanism for trusted integration of multi-view data (Zhu et al. 2025a).

Although the Mamba offers new insights into dynamic modeling for multi-view learning, existing methods often focus on intra-view dynamics while overlooking fine-grained interactions across views.

Methodology

In this section, we propose a Knowledge-Aware Multi-view State Space Model (KAMSSM), which is grounded in node-level modeling and designed to optimize fine-grained interactions both intra- and inter-view. Fig. 2 represents the network structure of the proposed method.

Notation and Problem Definition

In multi-view classification scenarios, we are given a dataset $\mathcal{D} = \{\mathcal{X}, \mathcal{Y}\}$, where $\mathcal{X} = \{\mathbf{X}^v \in \mathbb{R}^{N \times D_v}\}_{v=1}^V$ represents the view-specific feature, and $\boldsymbol{y}_i \in \mathcal{Y}$ denotes the label associated with each sample. Notation details are provided in Appendix B. The objective of multi-view learning is to exploit the underlying consistency and complementarity among views with limited labeled samples and improve the prediction of unseen labels. Considering the different feature dimensions across views, we map the features from each view into a shared space, where the features are represented as $\mathbf{H}_0^v$.

Selective Intra-view Communication

KAMSSM formulates node states as knowledge representations that guide selective information exchange. At each moment, nodes autonomously enter either the activation or static sequence.

Adaptive Sequence Assignment Nodes from different views are encoded through the environment encoders into low-dimensional embeddings that serve as their behavior distributions:

$$\mathbf{P}_t^v = Env(\mathbf{H}_t^v), \tag{5}$$

where $Env(\cdot)$ denotes the environment network of the current view, and the details are provided in Appendix B. Subsequently, the Gumbel-Softmax estimator is applied to evaluate the Gumbel-Softmax distribution of nodes,

$$\boldsymbol{g}_i^v = \frac{\exp\left((\log(\boldsymbol{p}_i^v) + \boldsymbol{n}_i^v)/\tau\right)}{\sum_{j=1}^{c'} \exp\left((\log(\boldsymbol{p}_j^v) + \boldsymbol{n}_j^v)/\tau\right)}, \tag{6}$$

where $\mathbf{G}_t^v = \{\boldsymbol{g}_1^v, \cdots, \boldsymbol{g}_n^v\} \in \mathbb{R}^{N \times c'}$, c' is the number of behavior categories, $\boldsymbol{n}_i^v \sim Gumbel(0, 1)$ is the indepen-

dently distributed Gumbel noise, and τ is a temperature parameter that controls the smoothness of the distribution.

Through hard sampling on the Gumbel-Softmax distribution, all nodes are assigned into activation sequences $\hat{\mathbf{H}}_t^v \in \mathbb{R}^{\hat{N}_v \times d}$ and static sequences $\bar{\mathbf{H}}_t^v \in \mathbb{R}^{\bar{N}_v \times d}$, where $\hat{\mathbf{H}}_t^v \cup \bar{\mathbf{H}}_t^v = \mathbf{H}_t^v$, and $\hat{\mathbf{H}}_t^v \cap \bar{\mathbf{H}}_t^v \neq \varnothing$. We then compute the representations for both sequences.

Activation Sequences Nodes within the activation sequence are categorized as source, receiver, or standard nodes, based on their communication behavior. Specifically, we design a mask-based attention mechanism to capture intra-sequence contextual dependencies, enabling nodes to communicate selectively, given by

$$\mathbf{A}_t^v = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{\mathbf{d}_k}}\right) \odot \mathbf{M}^v, \tag{7}$$

where $\odot$ denotes Hadamard product, $\mathbf{Q} = \mathbf{W}_t^Q \mathbf{H}_t^v$, $\mathbf{K} = \mathbf{W}_t^K \mathbf{H}_t^v$, and $\boldsymbol{d}_k$ is the degree of $\mathbf{K}$. $\mathbf{M}^v \in \mathbb{R}^{N \times N}$ is a mask matrix to control the behavior of different node types, which can be described as:

$$\mathbf{M}^v = \begin{cases} \mathbf{M}_{ij}^v = 1, & \text{if } n_i \text{ receives messages,} \\ \mathbf{M}_{ij}^v = 0, & \text{if } n_i \text{ does not receive messages,} \\ \mathbf{M}_{ji}^v = 1, & \text{if } n_i \text{ sends messages,} \\ \mathbf{M}_{ji}^v = 0, & \text{if } n_i \text{ does not send messages,} \end{cases}$$

where n_i represents the i-th node for any $j \in \{1, \ldots, N\}$. However, Eq. (7) inevitably involves a computational complexity of $\mathcal{O}(\mathcal{N}^2)$, which causes significant extra computation when dealing with large-scale data.

Therefore, we reorder matrix operations to reduce time complexity, and construct matrices $\tilde{\mathbf{K}}_t^v$ and $\tilde{\mathbf{Q}}_t^v$ matrices like (Wu et al. 2022). Similarly, the corresponding mask matrix $\mathbf{M}^v$ will be separated into $\mathbf{M}\mathbf{Q}^v$ and $\mathbf{M}\mathbf{K}^v$ to constrain the communication among nodes, given by

$$\mathbf{M}\mathbf{K}_{ij}^v = \begin{cases} 1, & \text{if } n_i \text{ sends messages,} \\ 0, & \text{if } n_i \text{ does not send messages,} \end{cases}$$

$$\mathbf{M}\mathbf{Q}_{ij}^v = \begin{cases} 1, & \text{if } n_i \text{ receives messages,} \\ 0, & \text{if } n_i \text{ does not receive messages,} \end{cases}$$

for any $j \in \{1, \ldots, d\}$. Then the masked Key and Query matrices are denoted as $\hat{\mathbf{K}}_t^v$, $\hat{\mathbf{Q}}_t^v$, respectively. The state update equation under the activation sequence can be expressed as follows:

$$\hat{\mathbf{H}}_t^v = \mathbf{D}^{-1}[\hat{\mathbf{Q}}_{t-1}^v (\hat{\mathbf{K}}_{t-1}^{v\top} \hat{\mathbf{H}}_{t-1}^v)], \tag{8}$$

where $\mathbf{D}$ is the degree of attention score.

Notably, the graph transformers tend to assign higher attention scores to nodes with similar embeddings, which keeps neighboring node embeddings smooth (Xing et al. 2024). Accordingly, we optimize the smoothness of embeddings learned from the activation sequence, defined as

$$\mathcal{L}_{sm} = \frac{1}{V} \sum_{v=1}^{V} \|\hat{\mathbf{H}}_t^v - \hat{\mathbf{H}}_{t-1}^v\|_F, \tag{9}$$

where V refers to the number of views, and a small $\mathcal{L}_{sm}$ indicates the activation sequence effectively identifies informative nodes and aggregates information.

Static Sequences The node that is not required to perform communication is assigned to the static sequence, which alleviates redundant information exchange. To avoid key nodes being trapped in static sequences and enhance temporal modeling, the static sequence leverages the history pool to track past behaviors $\boldsymbol{b}_i^v$ and Gumbel-Softmax distributions $\boldsymbol{s}_i^v$. In particular, we treat the current and historical Gumbel-Softmax distributions of a node as two neurons and simulate synaptic signaling to adjust its distribution, where the signal strength is enhanced when both neurons are simultaneously activated,

$$\Delta \boldsymbol{g}_i^v = \eta \boldsymbol{g}_i^v \boldsymbol{s}_i^v, \quad \hat{\boldsymbol{g}}_i^v = \boldsymbol{g}_i^v + \Delta \boldsymbol{g}_i^v, \tag{10}$$

where η denotes the learning rate of the signal. And the adjusted Gumbel-Softmax distribution can be expressed as:

$$\hat{\mathbf{G}}_t^v = \mathbf{G}_t^v + \eta \mathbf{G}_t^v \odot \mathbf{S}^v \odot \mathbf{B}^v. \tag{11}$$

The memory activation module will re-select the nodes within the static sequences. And the state update equation under the static sequence can be expressed as follows:

$$\bar{\mathbf{H}}_t^v = \bar{\mathbf{B}}_{t-1}^v \bar{\mathbf{H}}_{t-1}^v, \tag{12}$$

where $\bar{\mathbf{B}}_{t-1}^v$ is used to train networks on static sequences, such as an MLP or a linear model.

Fine-grained Inter-view Interaction

Each view is updated independently through an activation sequence and a static sequence, which can be formulated as:

$$\mathbf{H}_t^v = \mathbf{D}^{-1}[\hat{\mathbf{Q}}_t^v (\hat{\mathbf{K}}_t^{v\top} \hat{\mathbf{H}}_t^v)] + \mathbf{B}_t^v \bar{\mathbf{H}}_t^v. \tag{13}$$

Algorithm 1: Training procedure of KAMSSM.

Require: Multi-view data $\{\mathbf{X}^v\}_{v=1}^V$, semi-supervised information $\mathbf{Y}$, hyperparameters τ, η and Δt.
Ensure: The predicted label matrix $\tilde{\mathbf{Y}}$ for unlabeled data.

```
Initialize {W^v, b^v}_{v=1}^V of the network;
while not converged do
    for v = 1 → V do
        Construct the initial Gumbel distribution by Eq. (6);
    end for
    for t = 1 → T do
        for v = 1 → V do
            Compute the activation state Ĥ_t^(v) by Eq. (8),
            Calculate the L_sm through Eq. (9),
            Update Gumbel distribution by Eq. (11),
            Compute the static state H̄_t^(v) by Eq. (12);
        end for
        Compute inter-view state diffusion by Eq. (15);
    end for
    Calculate Z by Eq. (16),
    Calculate Ŷ and L_ce by Eq. (17),
    Calculate the golbal L by Eq.(18)
    Optimize {W^(v), b^(v)}_{v=1}^V and {W, b} of the networks with backward propagation;
end while
return The predicted label matrix Ŷ.
```

Existing multi-view methods leverage attention weight fusion or feature concatenation to construct shared representations, but the lack of fine-grained interaction will lead to interference from low-quality nodes, causing information contamination.

To this end, we propose a directional diffusion equation that guides the transition of node states across different views via diffusion coefficients, given as

$$\frac{\partial \mathbf{H}_t^v(i)}{\partial t} = \sum_{m=1}^{V} \mathbf{R}_t^{vm} \left(\mathbf{H}_{t-1}^m(i) - \mathbf{H}_{t-1}^v(i) \right), \tag{14}$$

where $\mathbf{R}_t^{vm}$ denotes the diffusion coefficient to describe state flow from view m to view v for node i. Then we rely on the explicit Euler method with step size Δt to solve the differential equation as follows:

$$\mathbf{H}_t^v = \Delta t \sum_{v=1}^{V} \mathbf{R}_t^{vm} \mathbf{H}_{t-1}^v + \left(\mathbf{I} - \Delta t \sum_{v=1}^{V} \mathbf{R}_t^{vm} \right) \mathbf{H}_{t-1}^v. \tag{15}$$

Training Strategy

After K times alternating updates, we fuse the representation from all views, the final representation is given as:

$$\mathbf{Z} = \mathcal{F}(\mathbf{H}_t^1, \ldots, \mathbf{H}_t^v). \tag{16}$$

Subsequently, we employ an MLP to map the fused representation to class prediction probabilities, and define a loss function based on the cross-entropy errors:

$$\hat{\mathbf{Y}} = \text{MLP}(\mathbf{Z}), \quad \mathcal{L}_{ce} = -\sum_{i \in \Phi} \sum_{j=1}^{C} \mathbf{Y}_{ij} \ln \hat{\mathbf{Y}}_{ij}. \tag{17}$$

where Φ is the set of samples with labels. The algorithmic loss consists of a smoothness loss and a cross-entropy loss, expressed as:

$$\mathcal{L} = \mathcal{L}_{ce} + \mathcal{L}_{sm}. \tag{18}$$

The algorithm complexity is $\mathcal{O}(Nd^2)$, with a detailed analysis provided in Appendix B, and the specific algorithmic workflow is outlined in Algorithm 1.

Experiment

In this section, we construct a series of experiments to evaluate the effectiveness of KAMSSM and answer several key questions:

- **Q1**: How does KAMSSM perform on multi-view semi-supervised node classification tasks?
- **Q2**: Does KAMSSM efficiently perform state updates by incorporating node-level knowledge awareness both intra- and inter-view?

Experimental Setup

Our model is implemented in PyTorch and trained on an Intel i5-12600KF CPU with an RTX 3090 GPU (24GB cache). All experiments are conducted with a 10% supervision rate, and the detailed model parameters for each dataset are provided in Appendix C.

Datasets	# Samples	# Views	# Classes	Data types
BDGP	2,500	2	5	Object image
Caltech	1,474	6	7	Object image
Flickr	12,154	2	7	Object image
HW	2,000	6	10	Digit image
OutScene	2,688	4	8	Object image
WebKB	203	3	4	Web Text
NoisyMNIST	70,000	2	10	Digit image
NUSWIDE	20,000	2	8	Object image

Table 1: A brief description of the tested datasets.

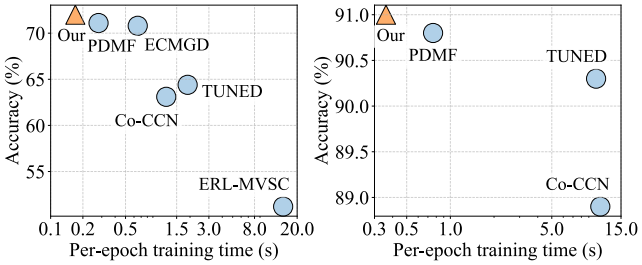
Datasets We evaluate the KAMSSM performance on six real-world multi-view datasets and further analyze its computational efficiency under two large-scale datasets. Table 1 presents a brief summary of these datasets; more details are given in Appendix C.

Baseline Methods KAMSSM is compared with representative non-graph and graph-based multi-view learning. Non-graph methods include ERL-MVSC (Huang et al. 2021), DSRL (Wang et al. 2022b) and PDMF (Xu et al. 2023), while graph-based methods include Co-GCN (Li, Li, and Wang 2020) LGCN-FF (Chen et al. 2023b), GEGCN (Lu et al. 2024b), ECMGD (Lu et al. 2024a) and TUNED (Huang et al. 2025).

Comparison to SOTA (Q1)

In this section, we evaluate the performance of the proposed model on semi-supervised node classification tasks. Then, we further assess its generalization, parameter sensitivity, and convergence.

Semi-supervised Classification Table 2 presents the node classification results, where KAMSSM achieves superior performance compared to the baseline algorithms across most datasets. In particular, it achieves notable improvements on the datasets Flickr and WebKB, outperforming the second-best method in accuracy by 2.1% and 6.8%, respectively. The only exception is Caltech, where KAMSSM slightly trails DSRL but exhibits more stable performance with lower variance.

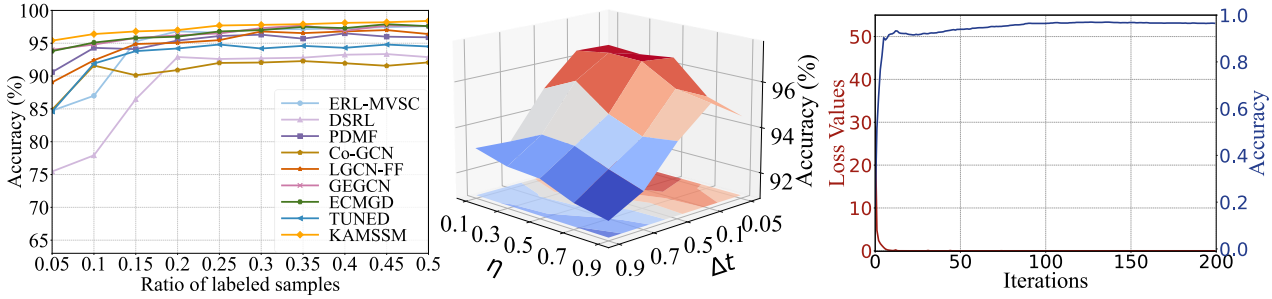


(a) NUSWIDE (b) NoisyMNIST

Figure 3: Running time and performance comparison on large-scale datasets (OOM methods not included).

Classification		w/o Graph Multi-view Learning			Graph-based Multi-view Learning					
Dataset	Metrics	ERL-MVSC	DSRL	PDMF	Co-GCN	LGCN-FF	GEGCN	ECMGD	TUNED	KAMSSM
BDGP	ACC	93.5 (0.8)	98.0 (1.7)	98.1 (0.4)	94.6 (1.7)	98.1 (0.2)	95.6 (0.7)	97.9 (0.2)	97.3 (0.2)	**98.6 (0.2)**
	F1	93.5 (0.8)	98.0 (1.7)	98.1 (0.4)	94.6 (1.7)	98.0 (0.2)	95.6 (0.7)	97.9 (0.2)	97.3 (0.2)	**98.6 (0.2)**
Caltech	ACC	93.3 (1.2)	**95.2 (2.8)**	92.2(1.0)	90.5 (3.5)	91.7 (1.2)	93.6 (0.3)	93.8 (0.1)	90.2 (1.2)	95.1 (0.3)
	F1	75.6 (4.9)	**81.1 (10.5)**	69.2 (4.1)	65.0 (10.6)	63.7 (4.6)	70.0 (1.9)	76.7(0.9)	60.5 (4.9)	80.7 (0.7)
Flickr	ACC	59.2 (0.5)	67.4 (8.3)	71.3 (0.3)	61.2 (2.6)	46.2 (4.1)	64.3 (0.1)	68.1 (0.1)	69.4 (0.4)	**73.4 (0.2)**
	F1	59.0 (0.5)	67.2 (8.5)	71.2 (0.2)	61.1 (2.4)	43.6 (5.3)	64.3 (0.1)	67.5 (0.2)	68.7 (1.1)	**73.2 (0.3)**
HW	ACC	87.0 (0.4)	77.9 (0.9)	94.3 (0.6)	91.6 (2.7)	92.4 (7.1)	94.8 (0.2)	95.1 (0.2)	91.9 (0.7)	**96.4 (0.1)**
	F1	92.7 (0.5)	87.5 (0.3)	94.4 (0.6)	86.9 (0.0)	92.4 (7.2)	94.8 (0.3)	95.1 (0.2)	92.1 (0.7)	**96.4 (0.1)**
OutScene	ACC	68.8 (1.4)	44.7 (0.8)	75.5 (0.6)	71.0 (2.1)	61.1 (11.0)	77.6 (0.3)	78.8 (0.2)	75.0 (1.5)	**80.1 (0.4)**
	F1	69.2 (1.4)	42.1 (2.9)	75.9 (0.6)	71.3 (2.0)	57.9 (15.6)	77.9 (0.3)	78.8 (0.2)	74.9 (1.7)	**80.2 (0.3)**
WebKB	ACC	65.9 (8.8)	60.4 (10.4)	72.2 (3.2)	71.2 (7.9)	78.6 (2.1)	75.2 (1.2)	77.8 (0.6)	74.5 (6.5)	**84.6 (0.9)**
	F1	38.0 (17.1)	27.2 (11.1)	37.9 (2.4)	39.6 (4.1)	38.2 (1.5)	34.5 (1.5)	39.4 (0.7)	41.8 (4.3)	**45.9 (0.6)**

Table 2: Average ACC and F1 with standard deviation (5 trials). The best highlighted in bold and the second-best underlined.



(a) HW accuracy vs. supervision ratio. (b) HW performance w.r.t. η and Δt. (c) KAMSSM loss and accuracy on HW.

Figure 4: Experimental results on the dataset HW w.r.t. supervision ratios, parameter sensitivity, and convergence curves.

To further assess the scalability of KAMSSM, we extend the evaluation to large-scale datasets. As shown in Fig. 3, the proposed model maintains high computational efficiency and strong predictive performance as the data size increases.

Training Size Set We further conduct an in-depth experimental analysis of KAMSSM under different supervision rates. Fig. 4(a) presents the classification performance of KAMSSM and baseline algorithms on the dataset HW across supervision rates ranging from 0.05 to 0.5. It is evident that KAMSSM consistently outperforms in all proportions of categorized information.

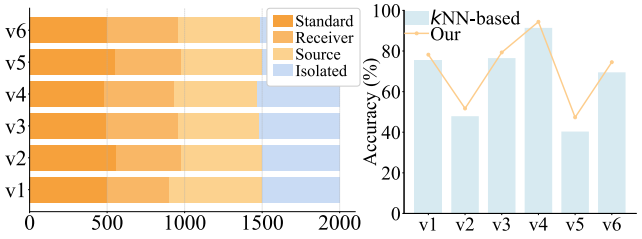
Parameter Sensitivity Analysis From Fig. 4(b), we observe that KAMSSM achieves optimal performance on the dataset HW with a smaller diffusion coefficient Δt and a moderate signal learning rate η on the HW. The parameter sensitivity results on other datasets are provided in Appendix C for comprehensive reference.

Convergence Analysis Then, as shown in Fig. 4(c), KAMSSM exhibits stable convergence within 100 epochs on the dataset HW. Additional convergence experiments on the other datasets are presented in Appendix C.

Selective Collaboration (Q2)

This section presents an empirical analysis of how KAMSSM performs knowledge-aware state updates among nodes.

Dynamic Intra-view Interaction Fig. 5(a) shows the distribution of node types within each view on the dataset HW. Combined with Fig. 5(b), we observe that lower-quality



(a) Node Type Distribution across views. (b) Per-view performance.

Figure 5: Proportions of node behavior and performance per view on the dataset HW.

(a) PDMF (b) GEGCN (c) ECMGD (d) TUNED (e) KAMSSM

Figure 6: t-SNE visualizations based on the fused representations of the dataset HW.

views have a higher proportion of standard nodes, reflecting more active intra-view interactions. And Fig. 5(b) illustrates that KAMSSM improves view quality more effectively than static kNN-based contextual structures through selective intra-view communication.

Visualization Analysis Fig. 6 presents the node classification visualization results on the HW dataset for the proposed model and the baselines. The results indicate that our model effectively enhances inter-class separability and intra-class compactness. More visualizations on the dataset HW are given in Appendix C.

Stability Analysis To verify the advantage of dynamic modeling under perturbation scenarios, we inject Gaussian noise $\mathcal{N}(0, \sigma^2)$ into each view. As shown in Fig. 7, as noise variance increases, all methods show performance drops, while KAMSSM consistently leads, demonstrating its ability to mitigate low-quality features through knowledge-guided node-level decisions. Further node attack experiments on additional datasets are provided in Appendix C.

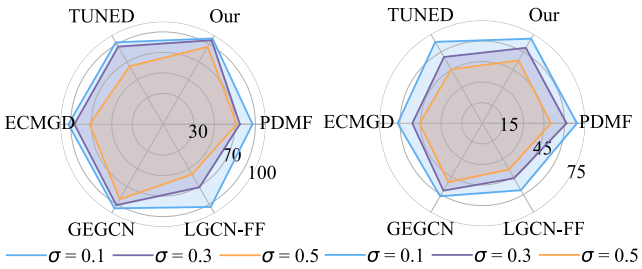


Figure 7: Performance comparison under node attacks on the datasets HW and Flickr.

Ablation Study (Q1 & Q2)

In this section, we evaluate the contribution of each component in the proposed model. KAMSSM enables dynamic node interactions within and across views, guided by knowledge. The former is primarily driven by node behaviors through activation and static sequences. In the activation sequences (AS), we replace the dynamic directional data dependency with static graphs constructed by kNN with k=10. In the static sequences (SS), we remove the historical pool and synaptic signals, and the Gumbel-Softmax distribution at each step depends only on the current node state. To verify the necessity of fine-grained inter-view interaction (FI), we adopt an attention-based weighted fusion across views while preserving intra-view modeling. From Table 3, we observe that all three modules contribute significantly to enhancing dynamic view modeling. Notably, although the activation and static sequences significantly enhance the model, the absence of directional diffusion-based view interactions often results in degraded view quality, particularly on the datasets Flickr, OutScene, and WebKB.

Dataset	w/o AS	w/o SS	w/o FI	Our
BDGP	95.6 (0.3)	97.9 (0.2)	97.7 (0.7)	**98.6 (0.2)**
Caltech	92.3 (0.3)	93.9 (0.5)	93.3 (0.5)	**95.1 (0.3)**
Flickr	64.5 (0.5)	71.1 (0.2)	71.9 (0.5)	**73.4 (0.2)**
HW	94.7 (0.5)	95.8 (0.5)	95.6 (0.5)	**96.4 (0.1)**
OutScene	74.9 (0.9)	76.9 (0.9)	77.5 (0.8)	**80.1 (0.4)**
WebKB	72.8 (4.0)	78.1 (9.8)	82.0 (1.0)	**84.6 (0.9)**

Table 3: Performance comparisons among model variants on the above six datasets.

Conclusion

In this paper, we move beyond previous graph-based multi-view learning that relies on collaborative optimization over all view nodes, and propose a knowledge-aware state space model that enables node-level selective interactions both intra- and inter-views. The proposed model redefines intra-view node serialization through activation and static sequences to enable node behavior-driven state aggregation and update, avoiding the involvement of all nodes in contextual interactions. In addition, we define a directional inter-view diffusion equation to achieve fine-grained dynamic interaction among views and prevent contamination from low-quality view. Experimental evaluations on various multi-view datasets confirm the superiority of the proposed model, as well as its efficiency in conducting node-level state updates within and across views. Moving forward, we aim to broadly explore neural state space models and neural ODEs for explaining multi-view learning, while also investigating more lightweight model designs.

Acknowledgments

This work is in part supported by the National Natural Science Foundation of China under Grants U25A20527 and 62276065, and the Fujian Provincial Natural Science Foundation of China under Grant 2024J01510026.

References

Aoki, M. 2013. *State space modeling of time series*. Springer Science & Business Media.

Behrouz, A.; and Hashemi, F. 2024. Graph mamba: Towards learning on graphs with state space models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 119–130.

Chen, M.; Wang, C.; Huang, D.; Lai, J.; and Yu, P. S. 2022. Efficient orthogonal multi-view subspace clustering. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 127–135.

Chen, Y.; Wu, Z.; Chen, Z.; Dong, M.; and Wang, S. 2023a. Joint learning of feature and topology for multi-view graph convolutional network. *Neural Networks*, 168: 161–170.

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.

Cheng, J.; Wang, Q.; Tao, Z.; Xie, D.; and Gao, Q. 2021. Multi-view attribute graph convolution networks for clustering. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence*, 2973–2979.

Cui, C.; Ren, Y.; Pu, J.; Pu, X.; and He, L. 2023. Deep multi-view subspace clustering with anchor graph. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence*, 3577–3585.

Fu, Y.; Li, Y.; Huang, Q.; Cui, J.; and Wen, J. 2024. Anchor graph network for incomplete multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36: 3708–3719.

Gao, Z.; Qi, Z.; and Chen, T. 2024. Mambas: Maneuvering Analog Multi-User Beamforming using an Array of Subarrays in mmWave Networks. In *Proceedings of the 30th International Conference on Mobile Computing and Networking*, 694–708.

Gu, A.; and Dao, T. 2024. Mamba: Linear-time sequence modeling with selective state spaces. In *Proceedings of the 1st Conference on Language Modeling*, 1–32.

Gu, A.; Dao, T.; Ermon, S.; Rudra, A.; and Ré, C. 2020. Hippo: Recurrent memory with optimal polynomial projections. In *Proceedings of the 33rd Advances in Neural Information Processing Systems*, 1474–1487.

Gu, A.; Goel, K.; Gupta, A.; and Ré, C. 2022. On the parameterization and initialization of diagonal state space models. In *Proceedings of the 35th Advances in Neural Information Processing Systems*, 35971–35983.

Gu, A.; Goel, K.; and Re, C. 2022. Efficiently Modeling Long Sequences with Structured State Spaces. In *Proceedings of the 10th International Conference on Learning Representations*, 1–15.

He, X.; Liang, H.; Peng, B.; Xie, W.; Khan, M. H.; Song, S.; and Yu, Z. 2025. MSAmba: Exploring Multimodal Sentiment Analysis with State Space Models. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, 1309–1317.

Huang, A.; Wang, Z.; Zheng, Y.; Zhao, T.; and Lin, C.-W. 2021. Embedding regularizer learning for multi-view semi-supervised classification. *IEEE Transactions on Image Processing*, 30: 6997–7011.

Huang, H.; Qin, C.; Liu, Z.; Ma, K.; Chen, J.; Fang, H.; Ban, C.; Sun, H.; and He, Z. 2025. Trusted unified feature-neighborhood dynamics for multi-view classification. In *Proceedings of the 39th AAAI conference on Artificial Intelligence*, 17413–17421.

Jiang, Z.; and Liu, X. 2022. Adaptive KNN and graph-based auto-weighted multi-view consensus spectral learning. *Information Sciences*, 609: 1132–1146.

Kou, Z.; Wang, J.; Tang, J.; Jia, Y.; Shi, B.; and Geng, X. 2024. Exploiting Multi-Label Correlation in Label Distribution Learning. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, 4326–4334.

Li, R.; Chen, M.; Ding, X.; Wang, C.; Xie, S.; Liu, S.; Chen, M.; and Guizani, M. 2024. Periodic Prompt on Dynamic Heterogeneous Graph for Next Basket Recommendation. In *Proceedings of the 2nd International Conference on Data Mining*, 747–752.

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

Lin, R.; Li, J.; Du, S.; Wang, S.; and Zhang, L. 2025. OIMGC-Net: Optimization-inspired Interpretable Multi-view Graph Clustering Network. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 1957–1966.

Lin, Z.; Tan, Y.; Zhan, Y.; Liu, W.; Wang, F.; Chen, C.; Wang, S.; and Yang, C. 2023. Contrastive intra-and inter-modality generation for enhancing incomplete multimedia recommendation. In *Proceedings of the 31st ACM International Conference on Multimedia*, 6234–6242.

Liu, J.; Han, J.; Liu, L.; Aviles-Rivero, A. I.; Jiang, C.; Liu, Z.; and Wang, H. 2025. Mamba4D: Efficient 4D Point Cloud Video Understanding with Disentangled Spatial-Temporal State Space Models. In *Proceedings of the 42nd IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17626–17636.

Lu, J.; Wu, Z.; Chen, Z.; Cai, Z.; and Wang, S. 2024a. Towards multi-view consistent graph diffusion. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 186–195.

Lu, J.; Wu, Z.; Yu, J.; Shen, Q.; Bu, J.; and Wang, H. 2025. Where Views Meet Curves: Virtual Anchors for Hyperbolic Multi-View Graph Diffusion. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 2131–2140.

Lu, J.; Wu, Z.; Zhong, L.; Chen, Z.; Zhao, H.; and Wang, S. 2024b. Generative essential graph convolutional network for multi-view semi-supervised classification. *IEEE Transactions on Multimedia*, 26: 7987–7999.

Park, J.; Park, J.; Xiong, Z.; Lee, N.; Cho, J.; Oymak, S.; Lee, K.; and Papailiopoulos, D. 2024. Can Mamba Learn How To Learn? A Comparative Study on In-Context Learning Tasks. In *Proceedings of the 41st International Conference on Machine Learning*, 39793–39812.

Shaker, A.; Wasim, S. T.; Khan, S.; Gall, J.; and Khan, F. S. 2025. GroupMamba: Efficient Group-Based Visual State Space Model. In *Proceedings of the 42nd IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14912–14922.

Sun, Y.; Li, X.; Sun, Q.; Zhang, M.-L.; and Ren, Z. 2024. Improved weighted tensor schatten p-norm for fast multi-view graph clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 1427–1436.

Wang, J.; Feng, S.; Lyu, G.; and Yuan, J. 2024. Surer: Structure-adaptive unified graph neural network for multi-view clustering. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, 15520–15527.

Wang, Q.; Ding, Z.; Tao, Z.; Gao, Q.; and Fu, Y. 2021. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Transactions on Image Processing*, 30: 1771–1783.

Wang, Q.; Tao, Z.; Gao, Q.; and Jiao, L. 2022a. Multi-View Subspace Clustering via Structured Multi-Pathway Network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5): 7244–7250.

Wang, S.; Chen, Z.; Du, S.; and Lin, Z. 2022b. 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, X.; Zhang, Y.; and Zhou, Y. 2025. Highly efficient rotation-invariant spectral embedding for scalable incomplete multi-view clustering. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, 21312–21320.

Wen, Y.; Liu, S.; Wan, X.; Wang, S.; Liang, K.; Liu, X.; Yang, X.; and Zhang, P. 2023. Efficient multi-view graph clustering with local and global structure preservation. In *Proceedings of the 31st ACM International Conference on Multimedia*, 3021–3030.

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

Wu, Q.; Zhao, W.; Li, Z.; Wipf, D. P.; and Yan, J. 2022. Nodeformer: A scalable graph structure learning transformer for node classification. In *Proceedings of the 35th Advances in Neural Information Processing Systems*, 27387–27401.

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.

Xia, W.; Wang, Q.; Gao, Q.; Zhang, X.; and Gao, X. 2021. Self-supervised graph convolutional network for multi-view clustering. *IEEE Transactions on Multimedia*, 24: 3182–3192.

Xing, Y.; Wang, X.; Li, Y.; Huang, H.; and Shi, C. 2024. Less is more: on the over-globalizing problem in graph transformers. In *Proceedings of the 41st International Conference on Machine Learning*, 1–17.

Xu, C.; Zhao, W.; Zhao, J.; Guan, Z.; Yang, Y.; Chen, L.; and Song, X. 2023. Progressive deep multi-view comprehensive representation learning. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, 10557–10565.

Xu, J.; Li, W.; Liu, X.; Zhang, D.; Liu, J.; and Han, J. 2020. Deep embedded complementary and interactive information for multi-view classification. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 6494–6501.

Yang, B.; Zhang, X.; Li, Z.; Nie, F.; and Wang, F. 2022. Efficient multi-view K-means clustering with multiple anchor graphs. *IEEE Transactions on Knowledge and Data Engineering*, 35(7): 6887–6900.

Yang, X.; Zhu, T.; Wu, D.; Wang, P.; Liu, Y.; and Nie, F. 2024. Bidirectional fusion with cross-view graph filter for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 36(11): 5675–5680.

Zhang, C.; Wang, S.; Liu, J.; Zhou, S.; Zhang, P.; Liu, X.; Zhu, E.; and Zhang, C. 2021. Multi-view clustering via deep matrix factorization and partition alignment. In *Proceedings of the 29th ACM International Conference on Multimedia*, 4156–4164.

Zhu, J.; Zou, X.; Liu, L.; Huang, Z.; Zhang, Y.; Tang, C.; and Dai, L.-R. 2025a. Trusted Mamba Contrastive Network for Multi-View Clustering. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1–5.

Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; and Wang, X. 2024. Vision mamba: efficient visual representation learning with bidirectional state space model. In *Proceedings of the 41st International Conference on Machine Learning*, 62429–62442.

Zhu, Y.; Zheng, X.; He, X.; Zou, X.; Wang, P.; Tang, C.; Liu, X.; and He, K. 2025b. BMCST: Balanced multi-view clustering for spatially resolved transcriptomics with mamba-driven dynamic feature refinement. *Information Fusion*, 103425.

Zhuang, S.; Huang, S.; Huang, W.; Chen, Y.; Wu, Z.; and Liu, X. 2024. Enhancing Multi-view Graph Neural Network with Cross-view Confluent Message Passing. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 10065–10074.