Boosting Knowledge Transfer and Retention with Brain-inspired Multi-View Incremental Learning

Yuhong Chen[1,2], **Zihan Fang**[3], **Huifeng Yin**[4], **Yujie Wu**[5], **Qi Xu**[6], **Lei Deng**[4], **Shiping Wang**[3*] and **Mingkun Xu**[1*]

[1]Guangdong Institute of Intelligence Science and Technology
[2]Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University
[3]College of Computer and Data Science, Fuzhou University
[4]CBICR, Department of Precision Instrument, Tsinghua University
[5]Department of Computing, The Hong Kong Polytechnic University
[6]School of Computer Science and Technology, Dalian University of Technology
yhchen2320@163.com, shipingwangphd@163.com, xumingkun@gdiist.cn

Abstract

Traditional multi-view learning models are primarily designed for static datasets with fixed views. However, in dynamic incremental view environments, this approach inevitably leads to view forgetting, where the introduction of new views weakens previously acquired knowledge. In contrast, the human brain exhibits remarkable memory retention and knowledge transfer capabilities when receiving objects described from different perspectives, with past experience further supplementing new insights. Inspired by underlying neural processing mechanisms, we propose a novel view incremental learning framework named Hebbian View Orthogonal Projection (HVOP). HVOP constructs a knowledge transfer space, where gradient updates are projected onto the orthogonal complement of historical representations, thereby mitigating interference between old and new views. By further incorporating recursive lateral connections and Hebbian learning rules, the proposed model imparts brain-like dynamic adaptability to the learning process, enhancing knowledge transfer and integration, thereby enabling stable knowledge transfer under evolving views. We validate HVOP on node classification tasks, demonstrating its superior performance in both knowledge retention and transfer compared to traditional methods. The results highlight the efficacy of biologically inspired mechanisms in mitigating the view forgetting phenomenon.

1 Introduction

Across many real-world applications, such as medical imaging [Weng *et al.*, 2024], social network analysis [Su *et al.*, 2026; Wu *et al.*, 2023b], and recommendation systems [Li *et al.*, 2024b], data are often observed as incremental views, where sensors or modalities become available over time rather than simultaneously. However, most existing multi-view learning methods [Zhao *et al.*, 2017; Xu *et al.*, 2024] are designed under a static setting and encounter two fundamental challenges in this incremental scenario. On one hand, they typically require retraining from scratch whenever a new view is introduced, leading to prohibitive computational overhead. On the other hand, when trained incrementally, they suffer from catastrophic forgetting, failing to retain the knowledge acquired from earlier views.

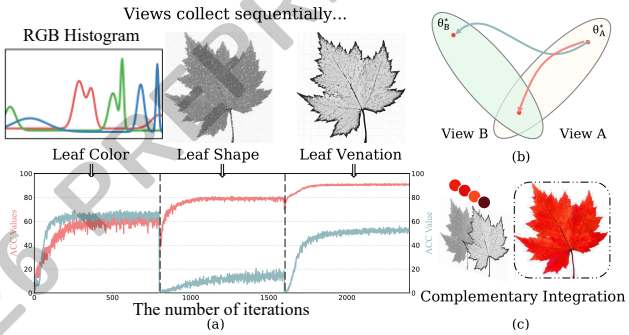

Figure 1: (a) Traditional multi-view methods suffer from view forgetting in the view increment setting. Leaf color, shape and venation views are sequentially entered into the model for learning, and the performance of the model depends only on the quality of the current view. (b) The updated direction of weights for the arrival of the new view, where the blue curve represents the updated direction of the traditional method and the red curve represents the desired updated direction. (c) The graph of view complementarity, where the combination of leaf shape, venation, and color produces the most complete representation of the data.

As illustrated in Figure 1(a), static multi-view models degrade significantly under view-incremental settings (blue curve). This degradation arises because knowledge learned from previous views cannot be effectively transferred to facilitate learning on newly arriving views. Consequently, the

*Corresponding Authors.

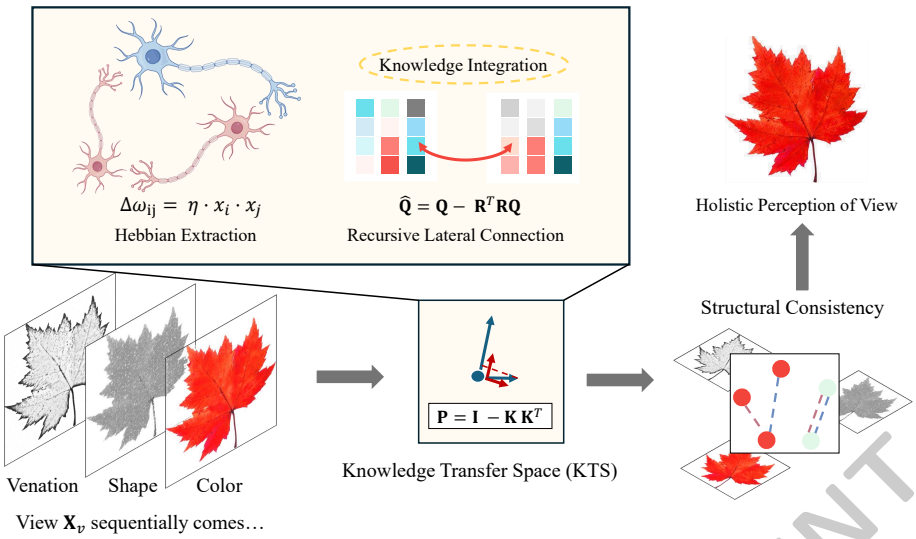

Figure 2: An overview of the proposed framework HVOP. HVOP refines a knowledge transfer space through recursive lateral connections and Hebbian learning, aiming at the transfer of view knowledge, while capturing the topological consistency of the multi-view structure, thus realizing holistic cognition in scenarios with incremental views.

model becomes overly dependent on the quality of the current view, limiting its ability to integrate information across views and to generalize. This phenomenon can be further understood from an optimization perspective, as shown in Figure 1(b). Each ellipse corresponds to an optimal parameter region associated with a particular view space. After learning view A, the model converges to an optimal region θ_A^*. When a new view B arrives, the optimization process is driven solely by gradients from view B, causing the parameter updates to drift toward θ_B^* and overwrite the knowledge in θ_A^*. Ideally, the optimization trajectory should be constrained by previously learned views, enabling convergence to a parameter region that jointly accommodates multiple view spaces (red curve). The inability to achieve this coordinated update highlights the intrinsic rigidity of existing incremental multi-view learning paradigms. Moreover, as shown in Figure 1(c), effectively modeling the consistency and complementarity between old and new views requires strong knowledge integration rather than isolated view-wise adaptation.

Interestingly, the human brain exhibits a markedly different learning behavior in analogous settings. It can continuously incorporate new sensory information while preserving previously acquired knowledge, a capability supported by dynamic neural interactions and memory consolidation mechanisms [Wang *et al.*, 2025; Chiel and Beer, 1997; Ito *et al.*, 2022]. Neuroscientific studies suggest that this flexibility is enabled by the brain's ability to reorganize synaptic connections and functional networks in response to new stimuli, while maintaining stable representations for existing memories [Deco *et al.*, 2011; Hassabis *et al.*, 2017]. In particular, the hippocampus facilitates the coexistence of plasticity and stability by forming partially decoupled yet interacting memory pathways, allowing new information to be integrated without destructive interference [Kumaran *et al.*, 2016; Scoville and Milner, 1957; Squire, 1992]. Furthermore, multisensory integration has been shown to enhance memory retention, indicating that coordinated processing across modalities plays a key role in robust learning [Stein and Stanford, 2008].

Motivated by these observations, we propose a brain-inspired framework for incremental multi-view learning, termed Hebbian View Orthogonal Projection (HVOP), as illustrated in Figure 2. HVOP draws inspiration from neural mechanisms to explicitly address knowledge retention and transfer under evolving view spaces. By introducing lateral connections and Hebbian learning dynamics, the framework approximates an orthogonal projection process that constrains parameter updates, thereby mitigating information loss when new views are introduced. This design enables previously learned view knowledge to actively guide optimization in subsequent learning stages, effectively mimicking the human brain's capacity to integrate and preserve information across multiple sensory channels. The effectiveness of this method has been verified through a large number of experiments in various multi-view benchmark tests. In these experiments, the performance of HVOP was mostly superior to that of traditional and state-of-the-art multi-view learning methods. We demonstrate that biologically inspired dynamic subspace tracking enables orthogonal learning in view-incremental settings. Specifically, the framework employs mechanisms analogous to:

- **Lateral connections in neural circuits:** These are simulated in our network architecture to facilitate the integration of information across different views, enhancing the model's ability to maintain a coherent representation of the data [Harris and Mrsic-Flogel, 2013].
- **Synaptic plasticity via Hebbian learning:** This is used to adapt the synaptic weights, ensuring that the connections are strengthened or weakened according to their utility in task performance [Feldman, 2012;

Gerstner *et al.*, 2018]. This adjustment supports the mechanism of orthogonal projection.

- **Memory retention and transfer in the hippocampus:** We mimic this aspect by introducing a specific Knowledge Transfer Space within our model. This is akin to regions in the hippocampus that are responsible for encoding new memories and consolidating long-term ones, respectively [Cohen, 1993].

2 Related Work

Multi-view learning [Wang *et al.*, 2021; Yao *et al.*, 2024; Zhuang *et al.*, 2025a; Xu *et al.*, 2025] integrates and encodes information from multiple sources to derive low-dimensional representations that capture both consistency and complementarity across views [Wu *et al.*, 2023a; Zhuang *et al.*, 2024; Wu *et al.*, 2026]. In contrast, multi-view incremental learning, or multi-view continual learning [Yan *et al.*, 2024; Li *et al.*, 2024a] primarily focuses on mitigating catastrophic forgetting by regularizing weight updates. Research addressing catastrophic forgetting typically centers on two key strategies: knowledge retention and knowledge transfer. Knowledge retention methods aim to preserve information from old tasks during the training of new ones, often through memory mechanisms or regularization techniques [Babakniya *et al.*, 2024; Chen *et al.*, 2021]. Conversely, knowledge transfer methods seek to leverage knowledge from prior tasks to facilitate learning new tasks; common approaches include using outputs from old tasks as guidance via knowledge distillation [Kang *et al.*, 2023; Kumari *et al.*, 2022; Zhou and Cao, 2021]. Overall, achieving an optimal balance between retaining old knowledge and acquiring new knowledge remains challenging. While recent work explores multi-view incremental learning in scenarios with increasing views [Chen *et al.*, 2025b], its emphasis on restricting weight updates may hinder effective learning of new views.

3 Methodology

For multi-view incremental learning, two objectives are essential: (i) preserving knowledge acquired from previous views, and (ii) effectively integrating information from newly arriving views. Learning from new views should therefore avoid interfering with representations formed from earlier ones. Inspired by biological processes in the human brain, the proposed Hebbian View Orthogonal Projection (HVOP) aims to replicate the adaptability to dynamic data observed in natural neural systems. Based on the principles of synaptic plasticity and memory consolidation in biological networks, HVOP integrates and preserves knowledge across dynamic view flows.

3.1 Preliminaries

Notations. Denote given multi-view feature matrices as $\mathcal{X} = \{\mathbf{X}_v\}_{v=1}^{V}$, where $\mathbf{X}_v = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \in \mathbb{R}^{N \times d_v}$ is the data from the v-th view for any $v \in \{1, \cdots, V\}$. Here, N is the number of samples, V is the number of views, and d_v denotes the distinct dimensionality of the feature matrices for each view. Additionally, c represents the number of classes. To quantify forgetting, we use $\mathcal{F}_{i,j}$ to denote the test accuracy on view j after finishing training on view i.

Definition 3.1 (View Forgetting). Given a sequence of view data $\mathcal{X} = \{\mathbf{X}_v\}_{v=1}^{V}$, after learning the v view features $\mathbf{X}_v$, the performance of the learned view v is tested, represented as $\mathcal{F}_{v,v}$. When the learning of the last view $\mathbf{X}_V$ is completed, the performance of the model significantly decreases in the previously tested views, shown as $\mathcal{F}_{v,V} < \mathcal{F}_{v,v}$.

3.2 Model Initialization for Incremental Views

For the existing view data, we use graph convolutional neural networks [Kipf and Welling, 2016; Chen *et al.*, 2023a; Chen *et al.*, 2025a; Zhuang *et al.*, 2025b] and fully connected layers to capture view-specific knowledge.

Upon the arrival of view $\mathbf{X}_v$, we build a kNN graph and obtain the corresponding adjacency $\mathbf{A}_v \in \mathbb{R}^{N \times N}$. We adopt a single-layer GCN with parameters shared across views:

$$\mathbf{Z}_v = \mathcal{GC}(\mathbf{A}_v, \mathbf{X}_v) = \sigma\big(\hat{\mathbf{A}}_v \mathbf{X}_v \mathbf{W}\big). \tag{1}$$

Here $\sigma(\cdot)$ denotes the activation function (ReLU in our implementation), $\widetilde{\mathbf{A}}_v = \mathbf{A}_v + \mathbf{I}$ adds self-loops, and $\mathbf{D}_v^e$ is the corresponding diagonal degree matrix used for normalization. And all views are aligned to a unified feature dimension before applying the shared transformation. After that, we designed a fully connected layer for implementing the classification, which can be represented as

$$\mathbf{H}_v = \mathcal{FC}_{\mathbf{W}_f}(\mathbf{Z}_v), \tag{2}$$

where $\mathbf{H}_v \in \mathbb{R}^{N \times c}$ defaults to the final predicted output.

Previous work has shown that the above models perform well in static multi-view learning scenarios. However, dynamic view-level learning has been neglected, where the model should be reused when posing the newly captured views. Therefore, the question is *how can incremental view learning be achieved without affecting the performance of models trained on historical views?* In order to preserve the knowledge of the old views, we need to impose a restriction on the direction of the gradient update for the next training.

3.3 Orthogonal Projection for Knowledge Transfer

This section defines the learning objective for preventing view forgetting and presents how it can be satisfied through orthogonal constraints. We first state a desideratum that prevents view forgetting (Eq. 3), then translate it into an orthogonal projection operator (Eq. 4), and finally realize this projection online via biologically inspired subspace tracking (Eqs. 6–9).

Assume the input vectors $\mathbf{x}_{old} \in \mathbb{R}^n$ from prior views, span the subspace $\mathbf{X}$. we explicitly formulate the desired condition that new-view updates must preserve the response to old-view inputs. Considering that changes in view information affect the updating of weights, which leads to view forgetting, in order to preserve the knowledge of the old view, we need to impose restrictions on the direction of the gradient $\Delta\mathbf{W}^P$ update for the next view, which makes it orthogonal to the subspace $\mathbf{X}$, satisfactory that $\Delta\mathbf{W}^P\mathbf{x}_{old} = 0$, which $\mathbf{x}_{old}$ denotes the past input vectors. This ensures that the gradient update of the weights does not affect the integration of knowledge of past views, demonstrated as

$$(\mathbf{W} + \Delta\mathbf{W}^P)\mathbf{x}_{old} = \mathbf{W}\mathbf{x}_{old}. \tag{3}$$

Eq. 3 therefore serves as the fundamental learning objective that subsequent projection techniques are designed to satisfy. Intuitively, Eq. 3 enforces that the model's response on historical inputs remains unchanged after learning a new view, *i.e.*, the update must not alter previously formed representations.

To address this, we specifically designate the principal subspace of $\mathbf{X}$ as the **Knowledge Transfer Space (KTS)**. The KTS is pivotal in assimilating crucial information which summarizes dominant historical representations for constraining future updates. Unlike task-based methods such as GPM or OWM, which compute static principal directions from accumulated gradients, KTS serves as a reference subspace that captures dominant representations from previous views, against which new updates are constrained. In other words, KTS acts as a running summary of dominant historical representations, providing the subspace to be protected when learning subsequent views.

If we calculate a projection matrix $\mathbf{P}$ to the subspace orthogonal to the KTS, then gradients can be projected as $\Delta\mathbf{W}^P = \Delta\mathbf{W}\mathbf{P}^T$. Inspired by [Saha *et al.*, 2021], for the gradient update of the fully connected layer, it can be adjust by subtracting its projection onto the space spanned by the top k principal component. If $\mathbf{K} \in \mathbb{R}^{d\times k}$ contains the top-k principal directions of historical representations, then the gradient for the classifier can be projected onto the orthogonal complement:

$$\nabla_{\mathbf{W}}\mathcal{L} \leftarrow (\mathbf{I} - \mathbf{K}\mathbf{K}^T)\nabla_{\mathbf{W}}\mathcal{L} = \mathbf{P}\nabla_{\mathbf{W}}\mathcal{L}. \tag{4}$$

Here, $\nabla_{\mathbf{W}}\mathcal{L}$ is the gradient of loss with respect to weight $\mathbf{W}$, $\mathbf{K}\mathbf{K}^T$ is the projection matrix to the principal subspace of the input feature. Thus, this effectively filters out components of the gradient that align with the principal directions, potentially reducing interference from predominant features that are already well-represented in the model. However, explicitly estimating $\mathbf{K}$ by SVD is costly and does not naturally adapt to streaming views.

3.4 Recursive Lateral Connections for Online Projection

This mechanism serves as an online approximation of the orthogonal projection defined in Sec. 3.3. Since explicitly estimating $\mathbf{K}$ by SVD is impractical under streaming views, we replace the static principal basis with an online subspace tracker, denoted by $\mathbf{R}$. Inspired by recursive lateral connections in biological neural systems, we introduce a recurrent projection mechanism to integrate new knowledge while suppressing interference with previously acquired representations.

Given an input representation $\mathbf{Q}$, we first project it into a neural subspace via $\mathbf{O} = \mathbf{R}\mathbf{Q}$, where $\mathbf{R}$ encodes the interaction between prior and incoming knowledge. A recurrent feedback term is then computed as

$$\mathbf{Q}^- = -\mathbf{R}^T\mathbf{O}, \tag{5}$$

which acts to remove components aligned with previously learned directions. Combining the original and recurrent representations yields

$$\hat{\mathbf{Q}} = \mathbf{Q} + \mathbf{Q}^- = (\mathbf{I} - \mathbf{R}^\top\mathbf{R})\mathbf{Q}. \tag{6}$$

Notably, this mechanism does not introduce a new objective; it provides an efficient realization of the orthogonal projection required by Eq. 3 in an online manner. This operation can be interpreted as a soft orthogonal projection onto the complement of the subspace spanned by prior knowledge.

In this way, the new view knowledge in the KTS can be further integrated and enhanced to provide richer information for global cognition. Here, the $\mathbf{Q}$ will change when each view is added, and will spontaneously integrate the knowledge left over from the previous step of the model, represented as

$$\mathbf{Q}_v = \sigma\big(\hat{\mathbf{A}}_v\mathbf{X}_v\mathbf{W}\big) + \alpha\mathbf{Z}_{v-1}. \tag{7}$$

Among them, $\mathbf{Q}_v$ refers to the representation we fuse after new view v arrives, which is composed of the representation of the current view and the knowledge of the previous view.

To alleviate the view forgetting phenomenon, we introduce an orthogonal projection mechanism inspired by regularization-based methods to impose restrictions on weight updates. In this way, the knowledge update brought by the new view will not conflict or interfere with the already solidly mastered knowledge system of the old view, thus facilitating a more efficient and robust learning process. From Eq. (6), we can conclude that the projection matrix $\mathbf{P}^{'}$ as

$$\mathbf{P}^{'} = \mathbf{I} - \mathbf{R}^T\mathbf{R}. \tag{8}$$

We find a striking similarity between Eq. (8) and the orthogonal projection mechanism in Eq. (4). Thus, as long as the recursive lateral connections $\mathbf{R}$ are equal or similar to the principal component matrix $\mathbf{K}$, the model can realize orthogonal projections, which inspires us to employ Hebbian learning to dynamically extract the principal component matrix. Oja rule, as an improvement of Hebbian learning, gradually approximates the principal direction of the data by adjusting the weight matrix. Its update formula is:

$$\mathbf{R}_{t+1} = \mathbf{R}_t + \eta(x_t y_t^T - y_t y_t^T\mathbf{R}_t), \tag{9}$$

where $\mathbf{R}_t$ denotes the dynamic principal matrix, η refers to learning rate, x_t is the input vector and y_t indicates the output vector computed from the current weight matrix and inputs.

With this rule, the data is projected and normalized to gradually approach its principal direction, ultimately causing the weight matrix to approximate the principal components of the input data. This adaptive mechanism enables online subspace regularization and mitigates view forgetting without resorting to explicit SVD-based decomposition. In this way, we leverage the biological mechanisms of recursive lateral connections and Hebbian learning to achieve orthogonal projection and facilitate knowledge transfer within a view incremental framework. This makes HVOP fundamentally different from PCA/SVD-based orthogonal methods that require storing past data or recomputing subspaces offline.

To further ensure the structural consistency of the model when dealing with different views, we design the following loss function to ensure that the graph structural information can be preserved in the incremental learning of views: $\mathcal{L}_{RE} = \frac{1}{2}\sum_{t=1}^{|\mathcal{T}|}\|\mathbf{A}_t - \sigma\left(\mathbf{Q}_{|\mathcal{T}|}\cdot\mathbf{Q}_{|\mathcal{T}|}^T\right)\|_F^2$, where $\mathbf{A}_t$ denotes the adjacency in the past view information and $|\mathcal{T}|$ is the

Datasets	Views	Samples	Features	Classes	Data Types
Animals	2	10,158	4,096	50	Images
Flower17	7	1,360	7	17	Images
Iaprtc12	2	7,855	100	6	Images & Text
NGs	3	500	2,000	5	Newsgroup document
NoisyMNIST	2	15,000	784	10	Images
YaleB_Extended	5	2,424	1,024	38	Images

Table 1: A brief description of the test multi-view datasets.

current training number of views. Despite the fact that the view data is in a constant state of flux, the intrinsic topology of the graph data shows surprising stability. With such constraints, we can capture the consistent representation between view information very well. Besides, for semi-supervised node classification, we calculate CrossEntropy and update parameters in HVOP, as follows $\mathcal{L}_{CE} = -\sum_{i\in\Omega}\sum_{j=1}^{c}\hat{y}_{ij}\ln y_{ij}$, where Ω refers to the set of labeled samples, $\hat{y}_{ij}$ usually denotes to one-hot encoding format and c is the amount of classes. Our approach constantly transfers and retains knowledge in the view incremental setting, while approaching a more comprehensive overall perception. We analyze the stability–plasticity trade-off (retention vs. new-view adaptation) and hyper-parameter sensitivity in Sec. 4.

4 Experiments

To validate the effectiveness of the proposed method, we have designed the experimental section intended to answer the following key evaluation questions (EQs):

- **EQ1** Does HVOP achieve desired performance for multi-view classification task?
- **EQ2** Does HVOP alleviate view forgetting in view incremental learning through orthogonal projection?
- **EQ3** Does HVOP capture the association between old and new views when new ones add up, and does it achieve global knowledge integration across views?
- **EQ4** How does HVOP behave in terms of training convergence, robustness to different view orders, and sensitivity to key hyper-parameters?

4.1 Experimental Setup

Datasets. We adopt six multi-view graph datasets widely used in different domains to evaluate the performance of HVOP compared to state-of-the-art baselines. Table 1 illustrates the details of all six datasets.

Compared Methods. Besides, we compare the proposed method with classical and state-of-the-art baseline models, including GAT [Veličković *et al.*, 2018], SI [Zenke *et al.*, 2017], MAS [et al., 2018], LGCNFF [Chen *et al.*, 2023b], DUANet [Geng *et al.*, 2021], RCML [Xu *et al.*, 2024] and MVCIL [Li *et al.*, 2024a], MVIL [Chen *et al.*, 2025b].

Evaluation Metric. Four well-known classification evaluation metrics are applied to our experiments, including Accuracy (ACC), Precision (P), Recall (R), and Macro-averaged F1 Score (MAF1). Experimental rounds of the proposed method are 800 per view under 10% randomly labeled data and the average value is taken by repeating the run 3 times.

4.2 Excellent Perception: Comparison to SOTA (EQ1)

We evaluate the effectiveness of our method on the multi-view classification task, the results are summarized in Table 2. We first compare it with the conventional static multi-view methods, including DUANet, LGCNFF and RCML, which are designed to simultaneously access all complete views, effectively leveraging inter-view correlations. DUANet excels at integrating reliable evidence by assessing the confidence level of each view. LGCNFF and RCML achieve desired performance by capturing both consistent and complementary information from multi-view data. These methods excel in scenarios where datasets consist of a fixed set of high-quality views, enabling effective exploitation of the available information. However, as the number and variety of views increment, static multi-view learning may fail to perform optimally in such more intricate scenarios.

In dynamic multi-view incremental environments, we incorporate classical baseline GAT into the view incremental framework. And observe that GAT struggles to retain information from previous views, with its performance largely dependent on the quality of the most recent view data rather than the integration of all views. We further introduce continual learning methods that leverage synaptic plasticity, SI and MAS, which aim to capture overall information through plasticity mechanisms. While these methods effectively leverage biological properties to maintain consistency across views, they encounter challenges when handling large category sets. Additionally, we compare our approach to multi-view class-incremental learning (MVCIL) methods. MVCIL demonstrates strong performance in handling datasets with multiple views, largely due to its effective retention of previously learned knowledge. However, it still faces limitations when dealing with low-quality views. Overall, the proposed HVOP reinforces the ability to extract maximal information from each view while seamlessly and efficiently integrating the continuously expanding view data.

4.3 Stability in Knowledge Retention: Demonstration of View Forgetting Relief (EQ2)

To verify that HVOP excels at memorizing and integrating knowledge from previous views, we tested it on inductive reasoning in Figure 3. We assessed the performance of both GCN and HVOP under conditions where views were incrementally introduced, focusing specifically on their ability to retain and utilize knowledge from earlier views. The results revealed that GCN exhibited a more significant deterioration in view performance as new views were added. In contrast, HVOP demonstrated a more gradual decline, with instances of stabilizing performance, indicating its resilience to the introduction of new information. Even as new views are continuously incorporated, HVOP is able to preserve and integrate knowledge from previous views. This implies that the HVOP model possesses robust memory retention capabilities, enabling it to maintain a stable understanding of previous views while nurturing the development of a more holistic cognition.

Datasets	Metric	DUANet	LGCNFF	RCML	GAT	SI	MAS	MVCIL	MVIL	HVOP
Animals	ACC	68.29 (0.46)	74.15 (1.32)	82.11 (0.17)	39.20 (0.36)	45.86 (0.14)	46.29 (0.16)	67.03 (0.03)	**84.42 (0.10)**	**84.73 (0.12)**
	P	66.63 (0.69)	69.07 (1.71)	78.43 (0.44)	34.18 (1.33)	39.92 (0.18)	39.89 (0.28)	64.86 (0.10)	**81.59 (0.31)**	**82.53 (0.26)**
	R	61.54 (0.45)	65.49 (1.46)	76.03 (0.18)	32.45 (0.44)	38.15 (0.04)	38.59 (0.18)	60.53 (0.03)	**78.10 (0.19)**	**77.93 (0.03)**
	MAF1	61.43 (0.55)	65.20 (1.59)	76.08 (0.20)	31.53 (0.69)	37.03 (0.13)	37.19 (0.02)	61.05 (0.00)	**78.32 (0.14)**	**78.43 (0.03)**
Flower17	ACC	58.24 (1.24)	45.35 (4.20)	30.42 (1.92)	17.78 (1.56)	44.81 (0.20)	47.10 (0.20)	40.93 (1.63)	**60.54 (0.72)**	**69.34 (0.42)**
	P	60.22 (1.75)	55.44 (6.80)	35.06 (4.12)	10.53 (3.75)	43.26 (0.22)	45.82 (0.33)	42.01 (1.62)	**61.98 (1.01)**	**70.29 (0.39)**
	R	58.24 (1.24)	45.32 (4.20)	30.42 (1.92)	17.78 (1.56)	44.81 (0.20)	47.10 (0.20)	40.93 (1.63)	**60.54 (0.72)**	**69.34 (0.42)**
	MAF1	57.31 (1.13)	39.75 (5.54)	25.06 (2.52)	10.19 (2.77)	42.86 (0.22)	45.61 (0.31)	38.53 (2.18)	**60.22 (0.70)**	**69.10 (0.56)**
Iaprtc12	ACC	37.53 (0.88)	57.84 (1.27)	52.36 (1.34)	34.98 (0.12)	57.00 (0.10)	57.01 (0.10)	43.12 (0.22)	**63.67 (0.22)**	**64.59 (0.30)**
	P	42.69 (1.21)	60.18 (1.47)	**66.83 (0.77)**	54.52 (0.18)	57.94 (0.09)	57.95 (0.09)	42.10 (0.32)	**65.54 (0.21)**	64.65 (0.41)
	R	37.29 (0.89)	58.87 (1.41)	51.65 (1.64)	31.08 (0.24)	57.81 (0.10)	57.82 (0.10)	44.18 (0.19)	**64.72 (0.21)**	**66.64 (0.24)**
	MAF1	37.83 (0.97)	59.28 (1.39)	55.22 (1.48)	27.69 (0.33)	57.85 (0.10)	57.86 (0.10)	42.14 (0.21)	**64.50 (0.21)**	**65.23 (0.35)**
NGs	ACC	32.62 (2.66)	90.65 (1.27)	81.53 (0.58)	72.52 (4.74)	79.11 (1.33)	83.56 (0.67)	58.89 (7.11)	**95.04 (0.64)**	**95.19 (0.10)**
	P	57.28 (4.22)	92.02 (0.66)	87.01 (0.29)	75.00 (4.66)	82.56 (0.07)	84.13 (0.62)	60.38 (8.56)	**95.16 (0.60)**	**95.40 (0.08)**
	R	32.62 (2.66)	90.67 (1.27)	81.53 (0.58)	72.52 (4.74)	79.11 (1.33)	83.56 (0.67)	58.89 (7.11)	**95.04 (0.64)**	**95.19 (0.10)**
	MAF1	25.84 (3.66)	90.57 (1.23)	82.32 (0.56)	72.50 (5.12)	79.31 (1.30)	83.60 (0.64)	58.81 (7.67)	**95.05 (0.64)**	**95.21 (0.11)**
NoisyMNIST	ACC	73.28 (2.86)	89.78 (0.48)	86.81 (0.11)	73.43 (0.61)	82.74 (2.10)	90.02 (0.29)	90.24 (0.62)	**91.43 (0.64)**	**90.79 (0.47)**
	P	72.85 (4.46)	89.75 (0.47)	86.89 (0.08)	74.02 (0.87)	85.44 (0.97)	90.20 (0.04)	90.26 (0.46)	**91.55 (0.49)**	**90.64 (0.50)**
	R	72.92 (2.81)	89.58 (0.49)	86.37 (0.11)	72.89 (0.63)	82.45 (2.07)	89.80 (0.30)	90.04 (0.47)	**91.29 (0.62)**	**90.60 (0.64)**
	MAF1	72.14 (3.87)	89.58 (0.49)	86.30 (0.12)	72.43 (0.71)	82.25 (1.96)	89.83 (0.28)	90.00 (0.48)	**91.30 (0.62)**	**90.59 (0.72)**
YaleB_Extended	ACC	**66.57 (1.81)**	34.01 (1.43)	62.59 (0.57)	31.19 (0.82)	32.76 (0.25)	32.86 (0.16)	64.16 (0.82)	65.85 (0.51)	**64.53 (0.55)**
	P	**74.48 (1.91)**	48.32 (6.29)	**81.25 (0.84)**	60.14 (4.80)	36.12 (0.33)	36.98 (0.45)	67.45 (0.97)	68.78 (0.51)	67.88 (0.23)
	R	**66.54 (1.79)**	33.99 (1.40)	62.61 (0.56)	31.21 (0.86)	32.83 (0.25)	32.92 (0.15)	64.24 (0.82)	65.90 (0.51)	**64.59 (0.55)**
	MAF1	**67.62 (1.82)**	33.24 (3.38)	67.22 (0.39)	34.01 (1.87)	33.51 (0.23)	33.85 (0.00)	64.86 (0.81)	66.45 (0.51)	**65.15 (0.51)**

Table 2: Multi-view classification performance with various algorithms, where the best results are highlighted in **orange** and the second-best results are highlighted in **blue**.

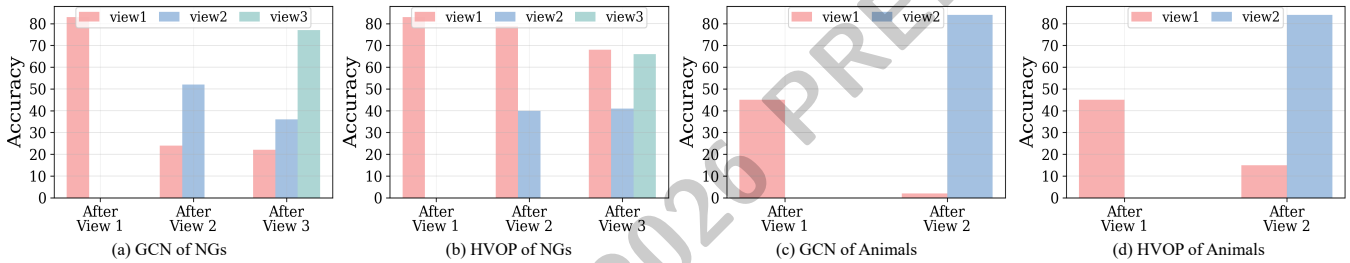

Figure 3: The phenomenon of view forgetting for HVOP vs. GCN in scenarios with incremental views.

4.4 Mitigating Forgetting: Verification of Knowledge Transfer Ability (EQ3)

We plot the change in performance as the views continue to increase, and add the results of training with only one view as a comparison, on the Figure 4. We observed that the performance of single-view learning is inferior to that of HVOP, indicating that HVOP effectively facilitates knowledge transfer and integration between views during the learning process, thereby enhancing overall cognitive ability. By integrating both coherent and complementary information from multiple views, model gains a more comprehensive understanding of data, resulting in more accurate and robust classifications. Furthermore, we designed the ablation experiment by removing the orthogonal projection module to visualize its importance. The experimental results show that the performance is significantly affected by the fluctuations in the quality of the new view data, showing an unstable behavior. While our HVOP not only underscores the model's exceptional capacity for knowledge accumulation, but also demonstrates its ability to continuously absorb and utilize new knowledge from additional views while preserving previously information.

4.5 Convergence Insights (EQ4)

In Figure 5(a), we show the loss convergence of the NGs dataset under both GCN and HVOP methods. The GCN's loss exhibits sharp fluctuations when new view data is introduced, suggesting that it struggles with effective knowledge transfer and integration across different views. This variability in loss indicates that the GCN model may not be robust to changes in input data, potentially affecting its performance and reliability in real-world applications where data variability is common. In contrast, the HVOP model demonstrates a smoother loss trajectory, indicating that changes in view data have a less significant impact on its performance. This suggests that the HVOP model is better equipped to integrate associations between old and new data, promoting a more stable and continuous progression in overall cognition.

The comparative analysis of loss reduction between the HVOP and GCN models across two different datasets highlights HVOP's superior ability to handle data variability and integrate new information smoothly. This stability is crucial for applications requiring reliable performance in dynamic settings, suggesting that the HVOP framework could be a preferable choice in scenarios where continuous learning and adaptation to new data are required.

4.6 View Order Analysis (EQ4)

In Figure 5(b), we present the impact of different sequential combinations of views on HVOP in Dataset Flower17. Unsurprisingly, all variations of view permutations lead to improvements in the final overall cognition, indicating that HVOP is not overly sensitive to the specific order in which views are introduced. Although different view orders induce distinct learning trajectories and intermediate representations, they consistently converge to a similar performance level, revealing a shared trend of steadily enhanced cognitive ability. This behavior suggests that the proposed orthogonal projection and knowledge transfer mechanisms enable robust integration of multi-view information under diverse streaming orders. In future work, we will further investigate how view sequencing interacts with view quality and complementarity, and how adaptive ordering strategies may further improve overall cognition.

4.7 Parameter Sensitivity Analysis (EQ4)

In this subsection, we performed the sensitivity analysis on the parameters k and hidden, where k denotes the number of neighbors chosen by kNN and hidden denotes the number of hidden layer dimensions in Figure 6.

Impact of k It can be observed that the performance of Animals, NGs and NoisyMNIST increases with the increase of k value. This could suggest that a larger k-value might facilitate better performance. However, the performance of others showed an upward trend with the decrease of k value. For instance, the Flower17's high number of views or YaleB_Extended's relatively small sample size might necessitate a focus on fewer, more relevant neighbors to optimize performance. A lower k value, in this context, could reduce noise and prioritize the most pertinent data points, thereby leading to improved performance outcomes.

Impact of hidden d Three datasets, Iaprtc12, NoisyMNIST, and YaleB_ Extended, showed an upward trend in performance as d increased, while the other datasets showed the opposite trend. Datasets characterized by complex, high-dimensional features, and a substantial number of multi-view settings, generally exhibit enhanced performance with larger hidden values. However, beyond a certain threshold, the performance improvements may plateau. Consequently, the optimal hidden value can differ significantly depending on the intrinsic properties of various datasets, including feature size, class count, and the number of views.

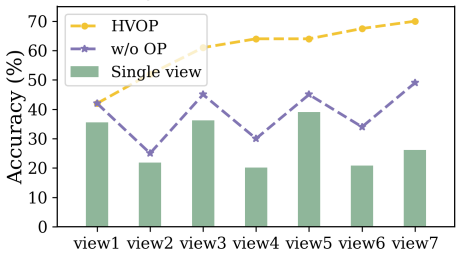

Figure 4: A comparison of HVOP streaming input and single-view learning performance.

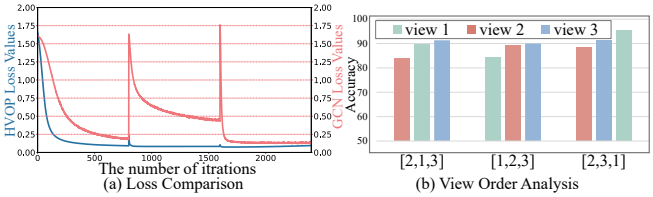

Figure 5: (a) Comparison of the loss reduction between HVOP and GCN in learning NGs dataset. (b) The overall performance of learning from NGs datasets in different view orders.

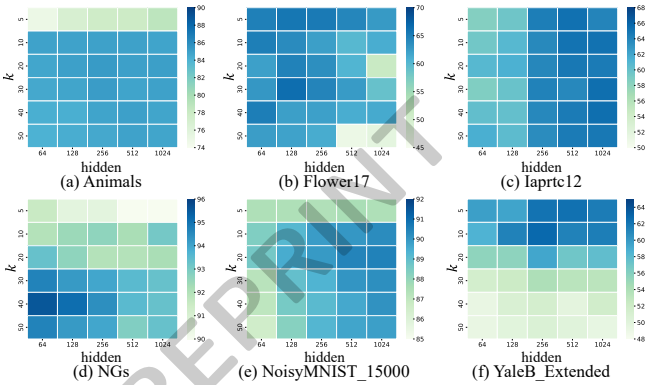

Figure 6: Parameter sensitivity analysis of HVOP on the test dataset with respect to the k and hidden.

5 Discussions and Conclusions

The traditional multi-view learning frameworks are often limited by their ability to handle static multi-view data, making it difficult to adapt to the dynamic growth and changes in real-world view data. In response, this study proposes a Hebbian-based View Orthogonal Projection framework, called HVOP to overcome this challenge. HVOP emphasizes view incremental learning, aiming to maintain the stability of the overall cognitive structure while focusing on the in-depth understanding of a single view in each learning task. We provide an empirical analysis of view forgetting and formulate view transfer learning as a problem setting. This concept hypothesizes that by effectively mitigating view forgetting, it is possible to achieve smooth transfer and integration of knowledge across views. HVOP method cleverly integrates recursive lateral connection mechanisms with Hebbian learning principles. This combination not only facilitates efficient knowledge extraction across views but ensures deep integration and continual optimization of both new and old view information within the neural network. Experiments exhibited that HVOP significantly enhances the model's adaptability and generalization ability in handling dynamic multi-view data.

Acknowledgments

This work was supported in part by the Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project under Grant No.2025ZD0215500 and National Natural Science Foundation of China (NSFC) under Grant No.62506084.

References

[Babakniya *et al.*, 2024] Sara Babakniya, Zalan Fabian, Chaoyang He, Mahdi Soltanolkotabi, and Salman Avestimehr. A data-free approach to mitigate catastrophic forgetting in federated class incremental learning for vision tasks. *Advances in Neural Information Processing Systems*, 36, 2024.

[Chen *et al.*, 2021] Pei-Hung Chen, Wei Wei, Cho-Jui Hsieh, and Bo Dai. Overcoming catastrophic forgetting by bayesian generative regularization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 1760–1770, 2021.

[Chen *et al.*, 2023a] Yuhong Chen, Zhihao Wu, Zhaoliang Chen, Mianxiong Dong, and Shiping Wang. Joint learning of feature and topology for multi-view graph convolutional network. *Neural Networks*, 168:161–170, 2023.

[Chen *et al.*, 2023b] Zhaoliang Chen, Lele Fu, Jie Yao, Wenzhong Guo, Claudia Plant, and Shiping Wang. Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, 95:109–119, 2023.

[Chen *et al.*, 2025a] Yuhong Chen, Fuhai Chen, Zhihao Wu, Zhaoliang Chen, Zhiling Cai, Yanchao Tan, and Shiping Wang. Heterogeneous graph embedding with dual edge differentiation. *Neural Networks*, 183:106965, 2025.

[Chen *et al.*, 2025b] Yuhong Chen, Ailin Song, Huifeng Yin, Shuai Zhong, Fuhai Chen, Qi Xu, Shiping Wang, and Mingkun Xu. Multi-view incremental learning with structured hebbian plasticity for enhanced fusion efficiency. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

[Chiel and Beer, 1997] Hillel J Chiel and Randall D Beer. The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in neurosciences*, 20(12):553–557, 1997.

[Cohen, 1993] NJ Cohen. *Memory, amnesia and the hippocampal system*. MIT Press, 1993.

[Deco *et al.*, 2011] Gustavo Deco, Viktor K Jirsa, and Anthony R McIntosh. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature reviews neuroscience*, 12(1):43–56, 2011.

[et al., 2018] Aljundi et al. Memory aware synapses: Learning what (not) to forget. In *ECCV*, pages 139–154, 2018.

[Feldman, 2012] Daniel E Feldman. The spike-timing dependence of plasticity. *Neuron*, 75(4):556–571, 2012.

[Geng *et al.*, 2021] Yu Geng, Zongbo Han, Changqing Zhang, and Qinghua Hu. Uncertainty-aware multi-view representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7545–7553, 2021.

[Gerstner *et al.*, 2018] Wulfram Gerstner, Marco Lehmann, Vasiliki Liakoni, Dane Corneil, and Johanni Brea. Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. *Frontiers in Neural Circuits*, 12:53, 2018.

[Harris and Mrsic-Flogel, 2013] Kenneth D Harris and Thomas D Mrsic-Flogel. Cortical connectivity and sensory coding. *Nature*, 503(7474):51–58, 2013.

[Hassabis *et al.*, 2017] Demis Hassabis, Dharshan Kumaran, Christopher Summerfield, and Matthew Botvinick. Neuroscience-inspired artificial intelligence. *Neuron*, 95(2):245–258, 2017.

[Ito *et al.*, 2022] Takuya Ito, Guangyu Robert Yang, Patryk Laurent, Douglas H Schultz, and Michael W Cole. Constructing neural network models from brain data reveals representational transformations linked to adaptive behavior. *Nature Communications*, 13(1):673, 2022.

[Kang *et al.*, 2023] Mengxue Kang, Jinpeng Zhang, Jinming Zhang, Xiashuang Wang, Yang Chen, Zhe Ma, and Xuhui Huang. Alleviating catastrophic forgetting of incremental object detection via within-class and between-class knowledge distillation. In *Proceedings of the International Conference on Computer Vision*, pages 18848–18858, 2023.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2016.

[Kumaran *et al.*, 2016] Dharshan Kumaran, Demis Hassabis, and James L McClelland. What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends in Cognitive Sciences*, 20(7):512–534, 2016.

[Kumari *et al.*, 2022] Lilly Kumari, Shengjie Wang, Tianyi Zhou, and Jeff A Bilmes. Retrospective adversarial replay for continual learning. *Advances in neural information processing systems*, 35:28530–28544, 2022.

[Li *et al.*, 2024a] Depeng Li, Tianqi Wang, Junwei Chen, Kenji Kawaguchi, Cheng Lian, and Zhigang Zeng. Multi-view class incremental learning. *Information Fusion*, 102:102021, 2024.

[Li *et al.*, 2024b] Yang Li, Qi'Ao Zhao, Chen Lin, Jinsong Su, and Zhilin Zhang. Who to align with: Feedback-oriented multi-modal alignment in recommendation systems. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 667–676, 2024.

[Saha *et al.*, 2021] Gobinda Saha, Isha Garg, and Kaushik Roy. Gradient projection memory for continual learning. In *International Conference on Learning Representations*, 2021.

[Scoville and Milner, 1957] William Beecher Scoville and Brenda Milner. Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1):11, 1957.

[Squire, 1992] Larry R Squire. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review*, 99(2):195, 1992.

[Stein and Stanford, 2008] Barry E Stein and Terrence R Stanford. Multisensory integration: current issues from

the perspective of the single neuron. *Nature Reviews Neuroscience*, 9(4):255–266, 2008.

[Su *et al.*, 2026] Haiyao Su, Yuhong Chen, Jielong Lu, and Shiping Wang. Position-aware and degree-adaptive graph neural networks for multi-view representation learning. *Knowledge-Based Systems*, page 115997, 2026.

[Veličković *et al.*, 2018] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

[Wang *et al.*, 2021] Shiping Wang, Zhaoliang Chen, Shide Du, and Zhouchen Lin. 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, 2021.

[Wang *et al.*, 2025] Yusong Wang, Xuanye Fang, Huifeng Yin, Dongyuan Li, Guoqi Li, Qi Xu, Yi Xu, Shuai Zhong, and Mingkun Xu. BIG-FUSION: brain-inspired global-local context fusion framework for multimodal emotion recognition in conversations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1574–1582, 2025.

[Weng *et al.*, 2024] Ying Weng, Yiming Zhang, Wenxin Wang, and Tom Dening. Semi-supervised information fusion for medical image analysis: Recent progress and future perspectives. *Information Fusion*, page 102263, 2024.

[Wu *et al.*, 2023a] Zhihao Wu, Xincan Lin, Zhenghong Lin, Zhaoliang Chen, Yang Bai, and Shiping Wang. Interpretable graph convolutional network for multi-view semi-supervised learning. *IEEE Transactions on Multimedia*, 25:8593–8606, 2023.

[Wu *et al.*, 2023b] Zhihao Wu, Zhao Zhang, and Jicong Fan. Graph convolutional kernel machine versus graph convolutional networks. In *Advances in Neural Information Processing Systems*, 2023.

[Wu *et al.*, 2026] Zhihao Wu, Jielong Lu, Zihan Fang, Jinyu Cai, Guangyong Chen, Jiajun Bu, and Haishuai Wang. Unifying multi-view knowledge for graph learning via model collaboration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 27010–27018, 2026.

[Xu *et al.*, 2024] Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable conflictive multi-view learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16129–16137, 2024.

[Xu *et al.*, 2025] Zhiyong Xu, Yuhong Chen, Ying Zou, Shide Du, Yan Chen, and Shiping Wang. Mbgcn: Multi-view block-wise graph convolutional networks on large-scale graphs. *IEEE Transactions on Image Processing*, 34:8285–8300, 2025.

[Yan *et al.*, 2024] Xiaoqiang Yan, Yingtao Gan, Yiqiao Mao, Yangdong Ye, and Hui Yu. Live and learn: continual action clustering with incremental views. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

[Yao *et al.*, 2024] Dingling Yao, Danru Xu, Sébastien Lachapelle, Sara Magliacane, Perouz Taslakian, Georg Martius, Julius von Kügelgen, and Francesco Locatello. Multi-view causal representation learning with partial observability. In *The Twelfth International Conference on Learning Representations*, 2024.

[Zenke *et al.*, 2017] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *Proceedings of the International Conference on Machine Learning*, pages 3987–3995, 2017.

[Zhao *et al.*, 2017] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017.

[Zhou and Cao, 2021] Fan Zhou and Chengtai Cao. Overcoming catastrophic forgetting in graph neural networks with experience replay. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4714–4722, 2021.

[Zhuang *et al.*, 2024] Shuman Zhuang, Sujia Huang, Wei Huang, Yuhong Chen, Zhihao Wu, and Ximeng Liu. Enhancing multi-view graph neural network with cross-view confluent message passing. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 10065–10074, 2024.

[Zhuang *et al.*, 2025a] Shuman Zhuang, Zhihao Wu, Yuhong Chen, Zihan Fang, Jiali Yin, and Ximeng Liu. Strategy-architecture synergy: A multi-view graph contrastive paradigm for consistent representations. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 7291–7299, 2025.

[Zhuang *et al.*, 2025b] Shuman Zhuang, Zhihao Wu, Zhaoliang Chen, Hong-Ning Dai, and Ximeng Liu. Refine then classify: Robust graph neural networks with reliable neighborhood contrastive refinement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 13473–13482, 2025.