Multi-View Alignment and Denoising via Center-Guided Spectral Diffusion

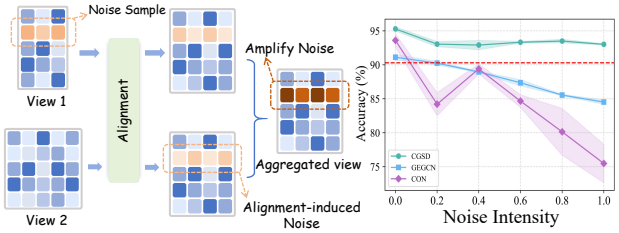
Jiayuan Wang[1] , **Jie Lian**[1] , **Yongquan Shi**[1] , **Jielong Lu**[2] , **Zhiyuan Lai**[1] , **Shiping Wang**[1*]
[1]College of Computer and Data Science, Fuzhou University, Fuzhou, China
[2]Zhejiang Key Laboratory of Accessible Perception and Intelligent Systems, College of Computer Science and Technology, Zhejiang University, Zhejiang, China
wwwangjiayuan@gmail.com, linj2680157@gmail.com, losparksayoji@outlook.com, jielonglu2022@163.com, zhiyuanlai001@163.com, shipingwangphd@163.com

Abstract

Multi-view learning aims to enhance performance by integrating information from multiple sources. While different views offer complementary perspectives, extracting consistent and discriminative representations remains a significant challenge due to discrepancies in representation and presence of noise. Existing methods typically separate the denoising and alignment processes, mapping the denoised heterogeneous views to a shared subspace for alignment. This means that noise may propagate or even amplify during the alignment stage, ultimately leading to suboptimal solutions. To address this issue, we propose center-guided spectral diffusion, which replaces traditional alignment with generative modeling. This method avoids noise propagation in multi-view alignment process and prevents multi-view features be aligned to the noise subspace during fusion process. Specifically, we first performs unconditional diffusion in both the feature space and a low-rank spectral space to learn a stable consensus center anchor. This anchor is then used to condition a guided generative diffusion process, enabling the model to generate more consistent and realistic sample representations. By combining conditional and unconditional diffusion, the proposed method alleviates the noise amplification problem commonly found in traditional alignment methods. Experimental results show that the proposed method outperforms state-of-the-art methods on several datasets.

1 Introduction

Multi-view data is prevalent in a wide range of real-world applications, including multi-camera vision systems and cross-lingual text modeling [Xie *et al.*, 2020; Santosh *et al.*, 2020; Liu *et al.*, 2024]. While different views offer complementary observations of the same underlying structure from diverse perspectives, challenges arise due to differences in representation, noise interference, and structural misalignment across views [Fang *et al.*, 2024; Lu *et al.*, 2025; Liu *et al.*, 2025a]. A central problem in multi-view learning is how to extract consistent and discriminative representations from such heterogeneous and partially redundant information [Tian *et al.*, 2023; Peng *et al.*, 2024]. In recent years, researchers have attempted to model similarity matrices to uncover potential relationships among samples [Chen *et al.*, 2023c; Du *et al.*, 2024; Wu *et al.*, 2025]. While this graph-based approach has achieved impressive results, the presence of extensive noise has exacerbated the heterogeneity between views, making denoising particularly important. Mainstream methods typically treat view denoising and alignment as two separate sequential processes. In this separation-based paradigm, each view is first denoised, and then aligned by mapping heterogeneous modalities to a shared subspace via a neural network [Geng *et al.*, 2021; Xu *et al.*, 2023; Lu *et al.*, 2024a]. Research shows that denoising and alignment operations in multi-view domains help models obtain more generalizable representations [Zhang *et al.*, 2020; Lian *et al.*, 2026].



(a) Traditional alignment methods (b) Impact of noise view

Figure 1: (a) Traditional methods introduce noise during the alignment process and then amplify it during the fusion stage. (b) The impact of adding noise to a single view and then fusion the view using various baseline model. Here, GEGCN learns the consistent representations through neural network; CON denotes the concatenation method. The red dashed line indicates the performance of the deleted noisy view on the baseline model.

This separation-based paradigm has achieved tremendous success. However, we have identified a limitation: *Anomaly features in low-quality views may distort cross-view alignment, which introduces additional noise into the alignment and fusion process, leading to denoising failure.* Recently research has shown that noisy samples tend to have larger gradient norms during training, causing the alignment process to

*Corresponding author.

focus excessively on these noisy samples [Huang *et al.*, 2021; Peng *et al.*, 2022], which may even steer the view alignment toward the noise subspace. As illustrated in Figure 1(a), the distorted embeddings of low-quality views propagate to other views during alignment through parameter-sharing mechanisms or interaction between views. The contaminated interactions force the representations of other views to shift toward noise, thereby disrupting the original semantic structure. In end-to-end training, the polluted representations serve as input for the next training iteration, further exacerbating information distortion. We conduct experiments to simulate the impact of different noise intensities on view alignment, as illustrated in Figure 1(b). Specifically, noise is only added to the first view, and we then evaluate the performance of different alignment methods. The experimental results show that as the noise intensity in a single view increases, traditional methods continuously contaminate the remaining views during the alignment process. The performance may even deteriorate to the point where abandoning the view will yield better results.

To address the above challenges, we propose Center Guided Spectral Diffusion (CGSD). Unlike previous methods such as direct mapping, we use a generative process to replace alignment to achieve synergistic optimisation of denoising and alignment. Specifically, we first model an unconditional diffusion stochastic differential equations system in both the feature space and the low-rank spectral space, which is used to iteratively update a consensus center anchor within multi-view. This consensus center anchor is then used as a input to a conditional diffusion model. Since anchors are consensus representations distilled from multi-view, conditional diffusion will be guided to generate aligned samples. After the model captures the underlying data distribution, in order to balance the truncation error and the number of network evaluations, we employ an Heun's second-order method as the solver for the differential equations and utilize Bayesian inference during the sampling phase to guide the generation of aligned samples. The contributions of can be summarized as follows:

- **Generative models replace alignment:** The CGSD framework is proposed to effectively mitigate noise amplification in multi-view data fusion by replacing traditional alignment with generative process.
- **Center-guided mechanism is designed:** The multi-view consensus anchor points are extracted through an unconditional diffusion process and then applied to conditional diffusion, guiding the alignment of multi-view.
- **Extensive experiments on real-world datasets:** The experimental results show that the proposed model achieves state-of-the-art results on multiple datasets.

2 Related Work

In this section, we will briefly introduce multi-view learning and score-based generative diffusion models.

2.1 Graph-based Multi-view Learning

The main objective of multi-view learning is to utilise complementary information from multiple perspectives or multi-source data to improve the performance and applicability of machine learning models. In recent years, with the advancement of graph learning methods, multi-view learning tasks have achieved significant breakthroughs [Lian *et al.*, 2024; Liu *et al.*, 2025b; Wang *et al.*, 2025]. [Chen *et al.*, 2023b] proposed to jointly learn feature and graph fusion for multi-view data through a two-stage, which integrated a feature fusion network and a learnable graph convolutional network with a novel differentiable shrinkage activation function to capture discriminative node relationships. [Lu *et al.*, 2024b] proposed to address multi-view semi-supervised classification by introducing a generative essential graph convolutional network that unified multi-graph consistency extraction, graph refinement, and task-specific optimization, leveraging a learnable threshold shrinkage function to obtain effective graph embeddings. [Chen *et al.*, 2023a; Lin *et al.*, 2025] focused on the optimisation and denoising of multi-view graph structures. Research shows that, compared with traditional methods, graph-based multi-view learning can achieve better results.

Notations	Descriptions
V	The number of views.
N	The number of total training samples.
d_v	The dimensions of the v-th view feature.
λ	Control the hyperparameters that guide the intensity.
$\{\mathbf{X}_v\}_{v=1}^V$	$\mathbf{X}_v \in \mathbb{R}^{N\times d_v}$ is the v-view training samples.
$\{\mathbf{A}_v\}_{v=1}^V$	$\mathbf{A}_v \in \mathbb{R}^{N\times N}$ is the v-view similarity matrix.
$\mathbf{\Phi}$	The spectrum space of $\mathbf{A}$.
$\mathbf{U}$	The truncated spectrum space of $\mathbf{A}$.
$\hat{\mathbf{X}}$	The aligned feature space.
$\hat{\mathbf{U}}$	The aligned truncated spectral space.
$\sigma(t)$	Control the scaling factor for different time steps.
$\epsilon_\phi(\cdot)$	Spectral space diffusion denoising neural network.
$s_\theta(\cdot)$	Feature space diffusion denoising neural network.
$\hat{\epsilon}_\phi(\cdot)$	Guided spectral alignment neural network.
$\hat{s}_\theta(\cdot)$	Guided feature alignment neural network.
$\mathcal{L}_{\mathrm{DM}}$	Unconditional gradient matching optimisation objective.
$\mathcal{L}_{\mathrm{CDM}}$	Conditional gradient matching optimisation objective.

Table 1: Symbols and their descriptions.

2.2 Generation Diffusion Model

Diffusion models are widely used due to their powerful ability to learn data distributions [Ho *et al.*, 2020; Dhariwal and Nichol, 2021; Chen *et al.*, 2024; Liu *et al.*, 2023]. In recent years, relevant advances in the field of graph generation have been continuously explored, with related work mainly focusing on graph-level tasks. [Jo *et al.*, 2022] used a continuous-time diffusion framework, enabling effective graph generation that satisfied structural dependencies and permutation invariance. However, running full-rank diffusion on the adjacency matrix causes fatal errors in the graph structure connections. [Luo *et al.*, 2023] applied low-rank diffusion SDEs in the graph spectrum space to efficiently learn graph topology generation, achieving excellent theoretical guarantees and state-of-the-art performance. In terms of nodes-level tasks, [Li *et al.*, 2024] enhanced latent space learning with discriminative content and a content-preservation mechanism, significantly improving anomaly detection performance while maintaining low time and space complexity. Although diffu-

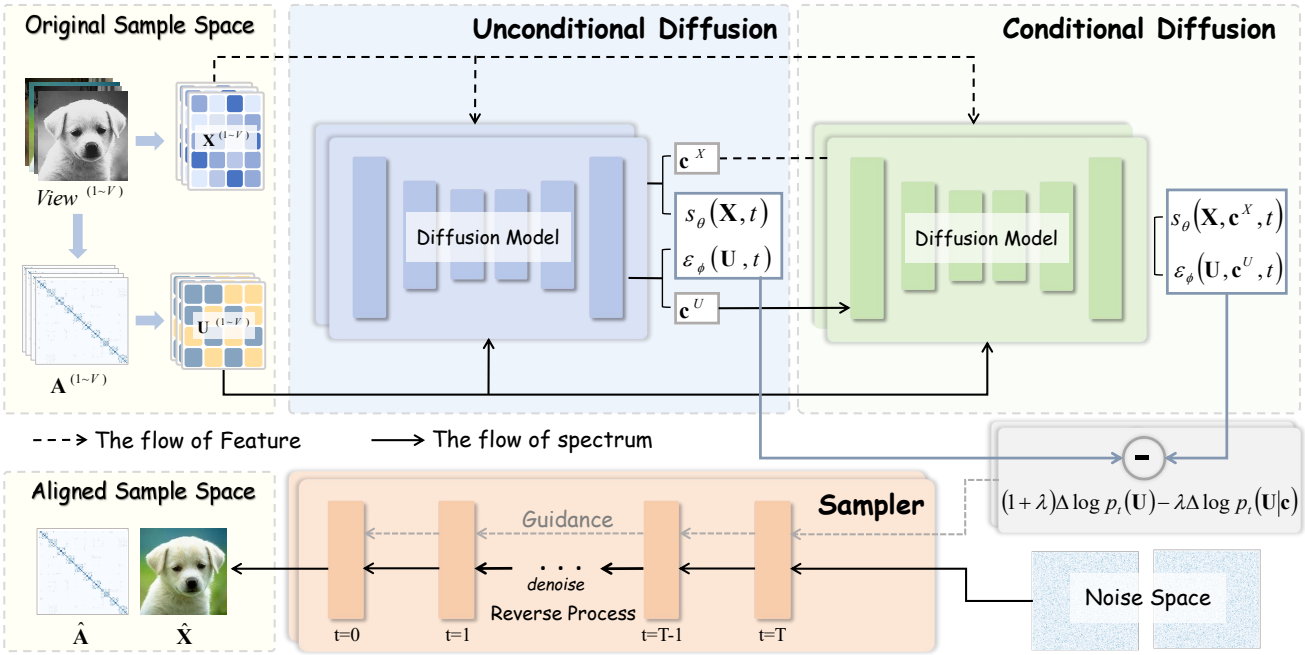


Figure 2: The overall architecture of the proposed **CGSD** framework. The model consists of three main stages: **1) Unconditional diffusion**, with the main purpose of learning the underlying distribution of data and stable consensus center anchor; **2) Conditional Diffusion**, where the learned anchor serves as conditions to guide the generation of denoised and aligned representations; **3) Sampling Stage**: To integrate the gradients of conditional and unconditional distributions.

sion models have achieved remarkable success in many fields, their extension to multi-view learning remains challenging due to the inherent information redundancy and noise from complementary multi-source inputs.

3 Preliminary

In this section, we first introduce some basic rules for using symbols in Table 1. Since diffusion models have limited applications in multi-view scenes, we also provide a comprehensive introduction to the basic theory of diffusion models.

3.1 Score-based Generative Diffusion Model

Diffusion models establish a unique bidirectional mapping mechanism between a complex data distribution $\mathcal{D}$ and a priori distribution $\mathcal{S}$ through a Stochastic Differential Equation (SDE) framework. Unlike traditional generative models (*e.g.*, VAE and GAN) that focus only on the unidirectional mapping from $\mathcal{S}$ to $\mathcal{D}$, the diffusion model considers the bidirectional relationship between $\mathcal{D}$ and S at the same time. After sufficient training, the model can reconstruct the data from normally distributed random Gaussian noise.

[Karras *et al.*, 2022] extended the SDE of [Song *et al.*, 2021] to a more general form:

$$d\mathbf{x}_{\pm} = \underbrace{-\dot{\sigma}(t)\sigma(t)\nabla_{\mathbf{x}}\log p(\mathbf{x})}_{\text{Probability flow ODE}}\, dt\pm \underbrace{\beta(t)\sigma(t)^2\nabla_{\mathbf{x}}\log p(\mathbf{x})\, dt + \sqrt{2\beta(t)}\sigma(t)d\omega_t}_{\text{Langevin diffusion SDE}}, \quad (1)$$

where $p(\cdot)$ is the probability density function of $\mathbf{x}$, $\nabla_{\mathbf{x}}\log p(\mathbf{x})$ is the score function of $\mathbf{x}$, $d\mathbf{x}_+$ and $d\mathbf{x}_-$ represent the forward and reverse processes of the diffusion model, $\beta(t)$ denotes the relative rate at which the existing noise is replaced by the new noise, and $\dot{\sigma}(t)$ is the first-order derivative of $\sigma(t)$. When $\beta(t) = \dot{\sigma}(t)/\sigma(t)$, the SDE of [Song *et al.*, 2021] is recovered. Then the SDE for forward diffusion can be formalized as:

$$\mathbf{x}_0 \sim \mathcal{D}, \quad d\mathbf{x}_t = f(\mathbf{x}_t, t)dt + g(t)d\omega_t, \quad (2)$$

where $f(\cdot)$ is the drift function, $g(\cdot)$ is the diffusion function, and $d\omega_t$ is the differential of the Wiener process, representing Gaussian random noise. The reverse time process SDEs can be derived as:

$$d\mathbf{x}_t = -2\dot{\sigma}(t)\sigma(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x})\, dt + \sqrt{2\dot{\sigma}(t)\sigma(t)}\, d\bar{\omega}_t. \quad (3)$$

In order to learn the unknown score function to match the noise in the inverse process, the following matching scores need to be minimized:

$$\mathcal{L}_{\mathrm{DM}} = \mathbb{E}_{\mathbf{x}_0\sim p(\mathbf{x})}\mathbb{E}_{\mathbf{x}_t\sim p(\mathbf{x}_t|\mathbf{x}_0)}\left\|\boldsymbol{\epsilon}_\theta - \nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)\right\|_2^2, \quad (4)$$

where $\boldsymbol{\epsilon}_\theta$ is the neural network that learns the probability density gradient score matching. Ideally, it will guide the data towards the real data distribution $\mathcal{D}$.

4 Methodology

This section details the methodology proposed in this study, and the architecture of the model is illustrated in Figure 2.

4.1 Multi-view Denoising Spectral Diffusion

Denote a set of views $\mathcal{X}$ as $\{\mathbf{X}^1, \mathbf{X}^2, \cdots, \mathbf{X}^V\}$. Each view can be represented as a graph $\mathcal{G}_v = (\mathcal{V}^{(v)}, \mathcal{E}^{(v)}, \mathbf{A}^{(v)}, \mathbf{X}^{(v)})$, where $\mathcal{V}^{(v)}$ is the set of nodes, $\mathcal{E}^{(v)}$ is the set of edges, $\mathbf{A}^{(v)} \in \mathbb{R}^{N\times N}$ and $\mathbf{X}^{(v)} \in \mathbb{R}^{N\times d_v}$, N is the total number of nodes and d_v indicates the feature dimension. We aim for the diffusion model to learn the underlying distribution and recover more discriminative data during the sampling process. Specifically, we train two score-based neural networks, including spectral space diffusion $\epsilon_\phi(\cdot)$ and feature space diffusion $s_\theta(\cdot)$, to generate high-fidelity denoised samples by running a reversed time SDEs.

Applying full-rank diffusion directly on sparse adjacency matrices can lead to significant errors, spectral-domain diffusion offers improved fidelity in reconstruction [Luo *et al.*, 2023]. However, unlike this work, our diffusion operation relies on eigenvectors rather than eigenvalues to obtain purer and more discriminative information. Specifically, let the spectral decomposition of $\mathbf{A}$ be given by $\mathbf{\Phi\Lambda\Phi}^{\mathrm{T}}$, where $\mathbf{\Phi}$ is the orthogonal of eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. Noise removal using truncated eigenvectors is an effective technique grounded in graph frequency domain analysis. This approach offers several advantages, including a solid theoretical foundation, effective noise suppression, and the ability to capture global structural information [Bruna *et al.*, 2014]. In our method, the diffusion process emphasizes the low-frequency components of the eigenvectors and achieves noise reduction by reconstructing the first k components of the truncated eigenvectors. For clarity, we denote these truncated vectors as $\mathbf{U}$. The forward spectral diffusion is then described by the following system of SDEs:

$$d\mathbf{X}_t^v = f^X(\mathbf{X}_t^v, t)\, dt + \sigma_X(t)\, d\omega_t^X, \tag{5}$$

$$d\mathbf{U}_t^v = f^U(\mathbf{U}_t^v, t)\, dt + \sigma_U(t)\, d\omega_t^U, \tag{6}$$

where $\mathbf{X}_t^v$ and $\mathbf{U}_t^v$ represent the evolutionary states of the feature vectors and eigenvectors, v is the view index, $f^X(\cdot, t) : \mathbb{R}^{N\times d_v} \mapsto \mathbb{R}^{N\times d_v}$ and $f^U(\cdot, t) : \mathbb{R}^{N\times N} \mapsto \mathbb{R}^{N\times N}$ denote the drift terms, $\sigma_X(t)$ and $\sigma_U(t)$ denote the corresponding diffusion coefficients. The stochastic processes ω_t^X and ω_t^U are standard Wiener processes.

Forward diffusion perturbs the original data with progressively stronger Gaussian noise under a prescribed drift $f^X(\cdot, t)$. Now we define the reverse time SDEs as follows:

$$d\mathbf{X}_t^v = -2\dot{\sigma}_X(t)\sigma_X(t)\nabla_X \log p_t(\mathbf{X}^v)\, dt + \sqrt{2\dot{\sigma}_X(t)\sigma_X(t)}\, d\bar{\omega}_t^X, \tag{7}$$

$$d\mathbf{U}_t^v = -2\dot{\sigma}_U(t)\sigma_U(t)\nabla_U \log p_t(\mathbf{U}^v)\, dt + \sqrt{2\dot{\sigma}_U(t)\sigma_U(t)}\, d\bar{\omega}_t^U, \tag{8}$$

where $p_t(\mathbf{X}^v)$ and $p_t(\mathbf{U}^v)$ denote the marginal probability density functions of $\mathbf{X}^v$ and $\mathbf{U}^v$ over time, $\bar{\omega}$ is reverse-time Wiener process, and the score matching functions $\nabla_X \log p_t(\mathbf{X}^v)$ and $\nabla_U \log p_t(\mathbf{U}^v)$ are used to train the neural network to learn the gradient of the data distribution.

4.2 Learning Center Anchor via Spectral Diffusion

As mentioned in previous sections, existing methods amplify noise during alignment, leading to failed denoising or even negative effects. To address this issue, we design a module that dynamically learns data centers to align multi-view representations and mitigate noise interference in views, ensuring that the model accurately captures the intrinsic relationships between them.

We begin by following the work of [Arazo *et al.*, 2019], which highlights that noise samples often deviate from the true data distribution. Specifically, we compute the similarity between the original representation and its reconstructed counterpart, which can be formalized as follows:

$$\omega_i^{(v)} = \frac{\exp\left(s_i^{(v)}/\tau\right)}{\sum \exp\left(s_j^{(v)}/\tau\right)}, \text{where } s_i^{(v)} = \frac{\mathbf{X}_i^{(v)} \cdot \hat{\mathbf{X}}_i^{(v)}}{\|\mathbf{X}_i^{(v)}\|\|\hat{\mathbf{X}}_i^{(v)}\|}, \tag{9}$$

where τ is the temperature coefficient that controls the smoothness of the weights, and s_i represents the cosine similarity between the original representation $\mathbf{X}$ and the current reconstructed representation $\hat{\mathbf{X}}$. After calculating the weights using Equation (9), we use it to reduce the impact of noisy samples on the data center. Specifically, we initialize the value as the representation mean to anchor the initial data center (that is, $\boldsymbol{c}^{(v)} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_i^{(v)}$) and iteratively refine it through the training process of the unconditional diffusion:

$$\boldsymbol{c}_{next}^{(v)} = \text{LayerNorm}\left(\sum_{i=1}^{N} \omega_i^{(v)} \hat{\mathbf{X}}_i^{(v)} + \boldsymbol{c}_{cur}^{(v)}\right), \tag{10}$$

$$\boldsymbol{c}_{\text{final}} = \frac{1}{V}\sum_{v=1}^{V} \boldsymbol{c}^{(v)}, \tag{11}$$

where $\boldsymbol{c}^{(v)}$ represents the representation center of view v, and $\boldsymbol{c}_{\text{final}}$ represents the global center of multi-view data. Iterative updates of the anchor enable the diffusion model to align representations into a consistent space during the reverse denoising process. In practice, we substitute $\mathbf{X}$ and $\mathbf{U}$ into the above equation to obtain the corresponding representations $\boldsymbol{c}_{\text{final}}^X$ and $\boldsymbol{c}_{\text{final}}^U$. In subsequent steps, these two consensus center anchors are used as conditional inputs to the diffusion model to guide its generation process.

4.3 Center-Guided Multi-view Alignment

Inspired by the Classifier-Free Guidance [Ho and Salimans, 2021], we use the consensus center anchor obtained from unconditional diffusion model as conditional inputs to the conditional diffusion model. Based on Bayes' theorem, we construct an adjustable guidance mechanism by integrating gradient information from both conditional and unconditional probabilities during the sampling process:

$$\tilde{s}_\theta(\mathbf{X}_t, t) = (1+\lambda)s_\theta(\mathbf{X}_t, \boldsymbol{c}_{\text{final}}^X, t) - \lambda s_\theta(\mathbf{X}_t, t), \tag{12}$$

$$\tilde{\epsilon}_\phi(\mathbf{U}_t, t) = (1+\lambda)\epsilon_\phi(\mathbf{U}_t, \boldsymbol{c}_{\text{final}}^U, t) - \lambda \epsilon_\phi(\mathbf{U}_t, t), \tag{13}$$

where $\lambda \geq 0$ is a hyperparameter that modulates the influence of the conditional guidance, $\boldsymbol{c}_{\text{final}}^X$ is the joint anchor of the

representation of the multi-view features, and $\boldsymbol{c}_{\text{final}}^{U}$ denotes the joint anchor of the multi-view spectral vector.

The traditional Euler method has a local truncation error of $O(h^2)$ relative to the step size h when solving ordinary differential equations, while high-order Runge-Kutta methods require multiple evaluations of the neural network s_θ and ϵ_ϕ [Süli and Mayers, 2003; Karras *et al.*, 2022]. In contrast, Heun's second-order method [Butcher, 2016] provides a favourable trade-off between truncation error and the number of neural network function evaluations [Jolicoeur-Martineau *et al.*, 2021]. Following the guidance framework defined by Equations (12) and (13), the integration process can be approximated using the trapezoidal rule:

$$\mathbf{X}_{t-1} = \mathbf{X}_t - \frac{1}{2}\left(\frac{\tilde{s}_\theta(\mathbf{X}_t, t)}{\sigma(t)} + \frac{\tilde{s}_\theta(\mathbf{X}^*_{t-1}, t-1)}{\sigma(t-1)}\right), \quad (14)$$

$$\mathbf{U}_{t-1} = \mathbf{U}_t - \frac{1}{2}\left(\underbrace{\frac{\tilde{\epsilon}_\phi(\mathbf{U}_t, t)}{\sigma(t)}}_{\text{Predictor}} + \underbrace{\frac{\tilde{\epsilon}_\phi(\mathbf{U}^*_{t-1}, t-1)}{\sigma(t-1)}}_{\text{Corrector}}\right), \quad (15)$$

where $\mathbf{X}^*_{t-1}$ and $\mathbf{U}^*_{t-1}$ are intermediate predicted values given by the first-order Euler formula. For the specific derivation process of Equations (14) and (15), please refer to **Appendix**. By predicting the noise in the data through the above system, we can both recover the underlying data distribution and guide the data toward alignment.

4.4 Training and Inference

Training Stage. During the training stage, the consensus center anchor is first obtained by training the unconditional diffusion model using loss (4). Subsequently, the conditional diffusion model is then iteratively optimized according to the following loss functions:

$$\mathcal{L}^X_{\text{CDM}} = \mathbb{E}_{\mathbf{X}_0\sim p(\mathbf{X})}\mathbb{E}_{\mathbf{X}_t\sim p(\mathbf{X}_t|\mathbf{X}_0,\boldsymbol{c})} \left\| \boldsymbol{s}^X_\theta - \nabla_{\mathbf{X}_t}\log p_t\left(\mathbf{X}_t|\boldsymbol{c}^X_{\text{final}}\right)\right\|_2^2, \quad (16)$$

$$\mathcal{L}^U_{\text{CDM}} = \mathbb{E}_{\mathbf{U}_0\sim p(\mathbf{U})}\mathbb{E}_{\mathbf{U}_t\sim p(\mathbf{U}_t|\mathbf{U}_0,\boldsymbol{c})} \left\| \epsilon^U_\phi - \nabla_{\mathbf{U}_t}\log p_t\left(\mathbf{U}_t|\boldsymbol{c}^U_{\text{final}}\right)\right\|_2^2. \quad (17)$$

Inference Stage. During the inference Stage, the data is reconstructed based on Equations (16) and (17) to obtain the aligned and denoised single-view outputs, denoted as $\hat{\mathbf{X}}$ and $\hat{\mathbf{U}}$. In addition, once $\hat{\mathbf{U}}$ is generated through the spectral diffusion process, it is reintroduced into the decomposition equation to reconstruct $\hat{\mathbf{A}}$ as $\hat{\mathbf{U}}\boldsymbol{\Lambda}\hat{\mathbf{U}}^{\mathrm{T}}$. For a more detailed algorithm flowchart, please refer to **Appendix**.

4.5 Theoretical Analysis

Proposition 1. *Let $\mathbf{A} \in \mathbb{R}^{N\times N}$ be the original data matrix, and $\mathbf{U} \in \mathbb{R}^{N\times k}$ be the matrix consisting of the top-k eigenvectors of $\mathbf{A}$, where $k \ll N$. For two diffusion processes, the upper bound of the generation error for the truncated eigenvector diffusion is strictly tighter than that of the full matrix diffusion, satisfying:*

$$\mathbb{E}\|\mathbf{U} - \widehat{\mathbf{U}}\|^2 \leqslant O\left(e^{Nk}\right), \quad (18)$$

$$\mathbb{E}\|\mathbf{A} - \widehat{\mathbf{A}}^{\text{full}}\|^2 \leqslant O\left(e^{N^2}\right). \quad (19)$$

Remark: Proposition 1 demonstrates that running diffusion on truncated eigenvectors yields tighter error bounds. Furthermore, as the sample size increases and k becomes negligible relative to N, the advantage in error bounds is further amplified. Consequently, our method exhibits significant advantages on large-scale datasets. The proof can be found in **Appendix**.

Complexity. The overall computational complexity of the diffusion phase is $O((4R+1)\cdot V\cdot N\cdot D^2)$. Upon simplification, the complexity exhibits a linear relation. A detailed analysis of the complexity is provided in **Appendix**.

5 Experiments

In this section, we first introduce the dataset and experimental setup, and then evaluate the performance of the proposed model. In addition, we conduct a series of experiments to further demonstrate the effectiveness of the model.

5.1 Datasets and Compared Methods

Datasets. We evaluate the proposed method on eight different real-world datasets and three large-scale datasets, including 100leaves, Animals, ALOI, Flower17, GRAZ02, Scene15, MNIST, NUS-WIDE, NUSWIDEOBJ, NoisyMNIST, and YTF. Among these, NUSWIDEOBJ, NoisyMNIST, and YTF are classified as large-scale datasets. Further details are provided in **Appendix**.

Compared Methods. We compare the proposed method with state-of-the-art baseline models, including Co-GCN [Li *et al.*, 2020], DSRL [Wang *et al.*, 2021], ECMGD [Lu *et al.*, 2024a], RCML [Xu *et al.*, 2024], LGCN-FF [Chen *et al.*, 2023b], and GEGCN [Lu *et al.*, 2024b]. Here LGCNFF and GEGCN are graph-learning models, and a detailed description of the methodology can be found in **Appendix**.

5.2 Experimental Setting

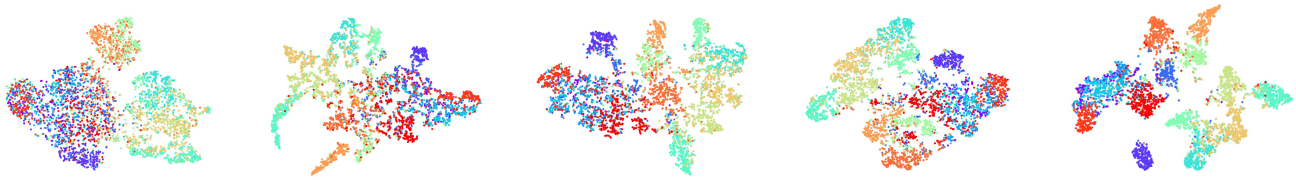
To mitigate randomness, each experiment repeats five times, and both the mean and variance are reported. For classification tasks, we follow the standard data partitioning protocols and conduct experiments using a fixed learning rate of 0.001 over 1000 training epochs. The experiments employ two authoritative evaluation metrics: Accuracy and F1-Score. More detailed experimental settings can be found in **Appendix**.

5.3 Classification on Multi-view Datasets

Performance. Graph convolutional networks are widely used for node classification tasks [Kipf and Welling, 2017; Li *et al.*, 2018; Xu *et al.*, 2019]. By feeding the aligned and denoised views obtained through CGSD into the GCN, we evaluate our node classification performance. The experimental results presented in Table 2 demonstrate that the proposed algorithm consistently outperforms other baseline methods across most datasets. Among them, LGCN-FF and GEGCN are denoising alignment algorithms. Specifically, it achieves notable accuracy improvements of 4.01%, 10.81%, and 2.61% on the 100leaves, Flower17, and Scene15 datasets, respectively. However, the GRAZ02 dataset contains too few samples to fully demonstrate the algorithm's advantages. Furthermore, due to the linear time complexity of CGSD

Model	Metrics	100leaves	Animals	ALOI	Flower17	GRAZ02	Scene15	MNIST	NUS-WIDE
Co-GCN	ACC	30.53 ± 5.01	73.00 ± 5.60	94.27 ± 0.72	37.58 ± 0.90	40.54 ± 2.65	58.70 ± 1.21	92.00 ± 0.51	41.74 ± 1.56
	F1	29.12 ± 4.91	73.72 ± 1.62	95.19 ± 0.87	33.67 ± 1.04	38.91 ± 1.54	56.72 ± 0.93	91.93 ± 0.53	39.73 ± 0.94
DSRL	ACC	67.91 ± 1.62	80.02 ± 2.37	91.67 ± 0.30	43.26 ± 5.22	48.13 ± 1.00	61.80 ± 0.91	92.91 ± 0.23	43.02 ± 1.38
	F1	64.42 ± 1.84	75.30 ± 2.84	91.72 ± 0.31	41.26 ± 4.80	48.74 ± 1.00	60.55 ± 0.84	92.82 ± 0.20	38.33 ± 2.33
ECMGD	ACC	85.50 ± 0.31	**81.84 ± 0.19**	**94.32 ± 0.72**	**61.63 ± 0.38**	60.58 ± 0.21	75.30 ± 0.42	88.31 ± 0.11	**47.64 ± 0.80**
	F1	85.02 ± 0.41	75.26 ± 0.27	**95.36 ± 0.72**	**61.15 ± 0.18**	60.27 ± 0.51	73.85 ± 0.32	88.12 ± 0.12	**46.67 ± 0.04**
RCML	ACC	52.01 ± 0.70	81.79 ± 0.06	83.44 ± 1.41	38.31 ± 2.42	42.39 ± 0.25	**76.19 ± 0.77**	85.10 ± 0.04	40.65 ± 0.17
	F1	43.92 ± 0.96	**75.58 ± 0.09**	83.63 ± 1.57	34.92 ± 2.47	38.30 ± 0.31	**75.03 ± 0.63**	84.66 ± 0.05	39.40 ± 0.18
LGCN-FF	ACC	62.13 ± 0.66	72.34 ± 6.00	93.14 ± 1.39	50.52 ± 0.00	49.58 ± 2.58	57.98 ± 0.38	90.40 ± 2.14	44.18 ± 0.59
	F1	61.67 ± 0.53	66.26 ± 6.84	93.16 ± 1.38	50.64 ± 0.00	43.68 ± 1.00	55.39 ± 1.42	88.90 ± 2.11	41.23 ± 1.04
GEGCN	ACC	**91.11 ± 0.55**	79.52 ± 0.22	90.71 ± 0.41	61.47 ± 0.74	**61.64 ± 0.50**	72.98 ± 0.44	**93.33 ± 0.19**	42.85 ± 1.28
	F1	**91.00 ± 0.50**	69.93 ± 0.34	90.75 ± 0.40	59.92 ± 0.84	**61.52 ± 0.33**	71.85 ± 0.56	**93.32 ± 0.20**	42.72 ± 0.98
Ours	ACC	**95.12 ± 0.35**	**85.23 ± 0.13**	**96.39 ± 0.80**	**72.44 ± 2.18**	**60.64 ± 1.09**	**78.80 ± 0.91**	**93.76 ± 0.23**	**49.09 ± 1.08**
	F1	**95.02 ± 0.36**	**77.96 ± 0.24**	**96.06 ± 0.81**	**72.63 ± 2.09**	**60.45 ± 1.34**	**77.89 ± 0.86**	**93.65 ± 0.25**	**48.85 ± 2.06**

Table 2: Classification results (mean% and standard deviation%) of all methods on multi-view datasets, using 10% of labeled samples for supervision. The best results are highlighted in green, while the second-best are highlighted in brown



(a) Co-GCN (b) LGCN-FF (c) GEGCN (d) ECMGD (e) Ours

Figure 3: T-sne visualization of DSRL, LGCN-FF, GEGCN, ECMGD, and CGSD on dataset Scene15.

diffusion process, the method is well-suited for large-scale datasets. As shown in Table 3, CGSD maintains competitive performance on large-scale multi-view datasets, achieving state-of-the-art performance on all three datasets.

Visualization. Figure 3 shows the t-SNE visualization results of the classification performance of different contrast methods on the Scene15 dataset. Observing the chart, compared to other models, we can see that CGSD significantly improves the inter-class separability within this dataset, indicating that our approach extracts more discriminative features. For visualization of all compared algorithms, please refer to **Appendix**.

5.4 Visualisation of Denoising Results

Figure 4 visualizes the original adjacency matrix, alongside the results of two advanced graph learning methods and our proposed denoising method. Notably, our method produces the most distinct diagonal block structure, indicating that the learned similarity matrix focuses on intra-class relationships with a marked reduction in heterogeneous connections. This structure offers a significant advantage for learning node embeddings via graph convolutional networks, as reflected in our superior performance on node classification tasks. The effectiveness of our method stems from the use of a diffusion model, which integrates informative signals from multi-view while simultaneously denoising truncated eigenvector.

5.5 Time Efficiency Analysis

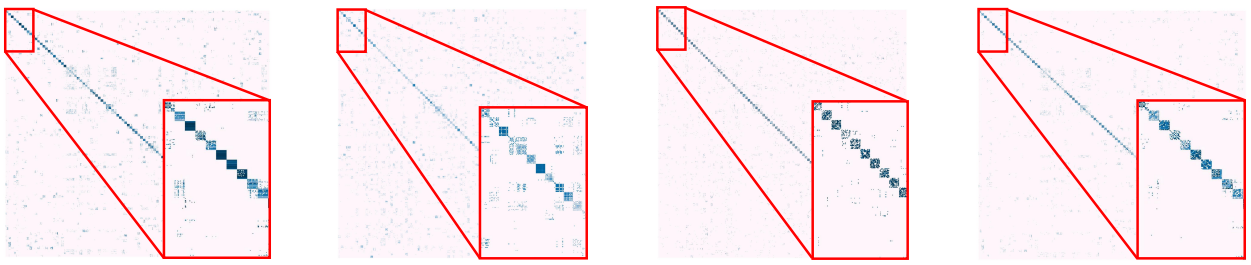
We conduct a comprehensive evaluation of the computational efficiency and performance of various algorithms on multi-view data, as shown in Figure 5. The results demonstrate that our proposed method not only achieves higher accuracy but also requires significantly less training time compared to most competing approaches. This improvement is largely due to our linear-time complexity diffusion process and advanced view fusion.

5.6 Ablation Study

To validate the effectiveness of CGSD, we conduct ablation studies, the results of which are illustrated in Figure 6. We independently remove the alignment operations within the feature and spectral spaces to isolate the impact of these alignment mechanisms. It can be observed that the ablated variants exhibit significant performance differences across the majority of datasets. This further indicates that the data sampled following our diffusion process possesses superior discriminability compared to the original data distribution. Detailed results of the ablation studies for the remaining datasets are provided in **Appendix**.

5.7 Parameter Sensitivity

We conduct experiments based on Equations (12) and (13) using various values of λ, with the detailed results presented in Figure 7. Specifically, we report node classification performance for $\lambda = 0$ (representing purely unconditional diffusion) and for λ values ranging from 0.1 to 1.0, capturing

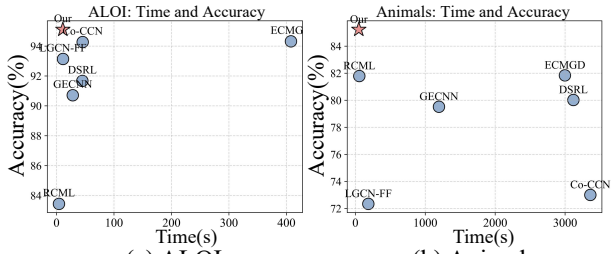


(a) Original (b) LGCN-FF (c) GEGCN (d) Ours

Figure 4: Visualization of the original, LGCN-FF-generated, GEGCN-learned, and our denoised adjacency matrices. Darker colors represent higher values; red boxes highlight our method's advantages.

Model	Metrics	NUSWIDEOBJ	NoisyMNIST	YTF
Co-GCN	ACC	26.55 ± 1.59	87.85 ± 1.87	**98.77 ± 0.01**
	F1	**15.22 ± 2.03**	87.59 ± 1.76	**98.76 ± 0.02**
DSRL	ACC	OOM	88.24 ± 0.01	OOM
	F1	OOM	88.52 ± 0.00	OOM
ECMGD	ACC	OOM	89.53 ± 0.52	OOM
	F1	OOM	89.31 ± 0.54	OOM
RCML	ACC	**34.05 ± 0.15**	88.40 ± 0.07	78.62 ± 0.52
	F1	11.84 ± 0.13	88.05 ± 0.08	76.58 ± 0.47
LGCN-FF	ACC	14.24 ± 1.24	88.68 ± 0.57	OOM
	F1	2.43 ± 1.27	85.56 ± 0.49	OOM
GEGCN	ACC	OOM	**93.63 ± 0.57**	OOM
	F1	OOM	**93.52 ± 0.54**	OOM
Ours	ACC	**38.22 ± 0.21**	**97.80 ± 0.36**	**99.86 ± 0.00**
	F1	**23.62 ± 0.16**	**97.78 ± 0.37**	**99.79 ± 0.01**

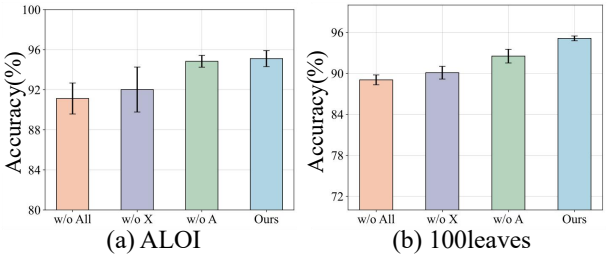
Table 3: Classification performance on large-scale datasets. The best results are highlighted in green, while the second-best are highlighted in brown, where 'OOM' denotes Out-of-Memory.



(a) ALOI (b) Animals

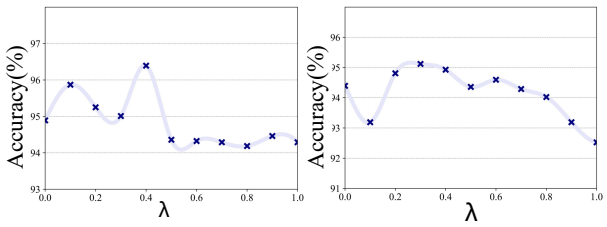
Figure 5: A comparsion of the runtime (s) and accuracy (%) of the algorithms on the ALOI and Animals datasets.

different levels of guidance strength. The results show that both datasets achieve optimal performance when λ is approximately 0.4, suggesting that moderate guidance effectively facilitates the alignment process by enabling the model to extract complementary information from multi-view. In contrast, excessive guidance leads to a decline in performance. Detailed results of the parameter sensitivity for the remaining datasets are shown in **Appendix**.



(a) ALOI (b) 100leaves

Figure 6: Performance comparison between different model variants on the ALOI and 100leaves datasets.



(a) ALOI (b) 100leaves

Figure 7: Performance of ALOI and 100leaves under different λ control quantities.

6 Conclusion and Future Work

This paper proposes a framework called CGSD, whose main innovation is to replace the alignment process with a generative approach. Unlike existing multi-view learning methods, we learn the underlying distribution of samples to generate aligned samples that conform to the distribution of multi-view data. This method effectively alleviates the problem of noise amplification in current alignment methods. Experiments on multiple real-world datasets validate the superiority of CGSD. In the future, we plan to further refine CGSD through several promising directions, such as developing adaptive guidance mechanisms for generation and investigating more stable sampling strategies, thereby extending it to increasingly complex real-world scenarios.

Acknowledgements

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

[Arazo *et al.*, 2019] Eric Arazo, Diego Ortego, Paul Albert, Noel O'Connor, and Kevin McGuinness. Unsupervised label noise modeling and loss correction. In *International Conference on Machine Learning*, pages 312–321, 2019.

[Bruna *et al.*, 2014] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann Lecun. Spectral networks and deep locally connected networks on graphs. In *International Conference on Learning Representations*, pages 1–14, 2014.

[Butcher, 2016] John Charles Butcher. *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2016.

[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.*, 2023c] Zhaoliang Chen, Zhihao Wu, Shiping Wang, and Wenzhong Guo. Dual low-rank graph autoencoder for semantic and topological networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4191–4198, 2023.

[Chen *et al.*, 2024] Huanran Chen, Yinpeng Dong, Zhengyi Wang, Xiao Yang, Chengqi Duan, Hang Su, and Jun Zhu. Robust classification via a single diffusion model. In *International Conference on Machine Learning*, pages 1–13, 2024.

[Dhariwal and Nichol, 2021] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, pages 8780–8794, 2021.

[Du *et al.*, 2024] Shide Du, Zhiling Cai, Zhihao Wu, Yueyang Pi, and Shiping Wang. Umcgl: Universal multi-view consensus graph learning with consistency and diversity. *IEEE Transactions on Image Processing*, 33:3399–3412, 2024.

[Fang *et al.*, 2024] Zihan Fang, Shide Du, Yuhong Chen, and Shiping Wang. Beyond the known: Ambiguity-aware multi-view learning. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 8518–8526, 2024.

[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*, pages 7545–7553, 2021.

[Ho and Salimans, 2021] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *Advances in Neural Information Processing Systems*, pages 1–8, 2021.

[Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. pages 6840–6851, 2020.

[Huang *et al.*, 2021] Zhenyu Huang, Guocheng Niu, Xiao Liu, Wenbiao Ding, Xinyan Xiao, Hua Wu, and Xi Peng. Learning with noisy correspondence for cross-modal matching. volume 34, pages 29406–29419, 2021.

[Jo *et al.*, 2022] Jaehyeong Jo, Seul Lee, and Sung Ju Hwang. Score-based generative modeling of graphs via the system of stochastic differential equations. In *International Conference on Machine Learning*, pages 10362–10383, 2022.

[Jolicoeur-Martineau *et al.*, 2021] Alexia Jolicoeur-Martineau, Ke Li, Rémi Piché-Taillefer, Tal Kachman, and Ioannis Mitliagkas. Gotta go fast with score-based generative models. In *The Symbiosis of Deep Learning and Differential Equations*, pages 1–6, 2021.

[Karras *et al.*, 2022] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, pages 26565–26577, 2022.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, pages 1–13, 2017.

[Li *et al.*, 2018] Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3538–3545, 2018.

[Li *et al.*, 2020] Shu Li, Wen-Tao Li, and Wei Wang. Co-gcn for multi-view semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4691–4698, 2020.

[Li *et al.*, 2024] Jinghan Li, Yuan Gao, Jinda Lu, Junfeng Fang, Congcong Wen, Hui Lin, and Xiang Wang. Diffgad: A diffusion-based unsupervised graph anomaly detector. In *The Thirteenth International Conference on Learning Representations*, pages 1–14, 2024.

[Lian *et al.*, 2024] Jie Lian, Xuzheng Wang, Xincan Lin, Zhihao Wu, Shiping Wang, and Wenzhong Guo. Graph anomaly detection via multi-view discriminative awareness learning. *IEEE Transactions on Network Science and Engineering*, 11:6623–6635, 2024.

[Lian *et al.*, 2026] Jie Lian, Zhihao Wu, Jielong Lu, Jiajun Yu, Qianqian Shen, and Haishuai Wang. Beyond local patterns: Multiscale inconsistency learning for graph anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 15216–15224, 2026.

[Lin *et al.*, 2025] Weihong Lin, Zhaoliang Chen, Yuhong Chen, and Shiping Wang. Heterogeneous graph neural

network with adaptive relation reconstruction. *Neural Networks*, 187:107313–107323, 2025.

[Liu *et al.*, 2023] Xihui Liu, Dong Huk Park, Samaneh Azadi, Gong Zhang, Arman Chopikyan, Yuxiao Hu, Humphrey Shi, Anna Rohrbach, and Trevor Darrell. More control for free! image synthesis with semantic diffusion guidance. In *Proceedings of the Winter Conference on Applications of Computer Vision*, pages 289–299, 2023.

[Liu *et al.*, 2024] Suyuan Liu, Qing Liao, Siwei Wang, Xinwang Liu, and En Zhu. Robust and consistent anchor graph learning for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 36:4207–4219, 2024.

[Liu *et al.*, 2025a] Chengliang Liu, Jie Wen, Yong Xu, Bob Zhang, Liqiang Nie, and Min Zhang. Reliable representation learning for incomplete multi-view missing multi-label classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47:4940–4956, 2025.

[Liu *et al.*, 2025b] Lu Liu, Yongquan Shi, Yueyang Pi, Wenzhong Guo, and Shiping Wang. Efficient multi-view graph convolutional networks via local aggregation and global propagation. *Expert Systems with Applications*, 266:126131, 2025.

[Lu *et al.*, 2024a] Jielong Lu, Zhihao Wu, Zhaoliang Chen, Zhiling Cai, and Shiping Wang. Towards multi-view consistent graph diffusion. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 186–195, 2024.

[Lu *et al.*, 2024b] Jielong Lu, Zhihao Wu, Luying Zhong, Zhaoliang Chen, Hong Zhao, and Shiping Wang. Generative essential graph convolutional network for multi-view semi-supervised classification. *IEEE Transactions on Multimedia*, 26:7987–7999, 2024.

[Lu *et al.*, 2025] Mingfei Lu, Qi Zhang, and Badong Chen. Divergence-guided disentanglement of view-common and view-unique representations for multi-view data. *Information Fusion*, 114:102661, 2025.

[Luo *et al.*, 2023] Tianze Luo, Zhanfeng Mo, and Sinno Jialin Pan. Fast graph generation via spectral diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46:3496–3508, 2023.

[Peng *et al.*, 2022] Xiaokang Peng, Yake Wei, Andong Deng, Dong Wang, and Di Hu. Balanced multimodal learning via on-the-fly gradient modulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8238–8247, 2022.

[Peng *et al.*, 2024] HU Peng, ZHEN Liangli, PENG Xi, et al. Deep supervised multi-view learning with graph priors [j]. *IEEE Transactions on Image Processing*, 33:123–133, 2024.

[Santosh *et al.*, 2020] Tokala Santosh, Avirup Saha, and Niloy Ganguly. Mvl: Multi-view learning for news recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1873–1876, 2020.

[Song *et al.*, 2021] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, pages 1–12, 2021.

[Süli and Mayers, 2003] Endre Süli and David F Mayers. *An introduction to numerical analysis*. Cambridge University Press, 2003.

[Tian *et al.*, 2023] Xudong Tian, Zhizhong Zhang, Cong Wang, Wensheng Zhang, Yanyun Qu, Lizhuang Ma, Zongze Wu, Yuan Xie, and Dacheng Tao. Variational distillation for multi-view learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46:4551–4566, 2023.

[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*, 44:5042–5055, 2021.

[Wang *et al.*, 2025] Xuzheng Wang, Shiyang Lan, Zhihao Wu, Wenzhong Guo, and Shiping Wang. Multi-view representation learning with decoupled private and shared propagation. *Knowledge-Based Systems*, 310:112956, 2025.

[Wu *et al.*, 2025] Yuze Wu, Shiyang Lan, Zhiling Cai, Mingjian Fu, Jinbo Li, and Shiping Wang. Schg: Spectral clustering-guided hypergraph neural networks for multi-view semi-supervised learning. *Expert Systems with Applications*, 277:127242, 2025.

[Xie *et al.*, 2020] Yuan Xie, Bingqian Lin, Yanyun Qu, Cuihua Li, Wensheng Zhang, Lizhuang Ma, Yonggang Wen, and Dacheng Tao. Joint deep multi-view learning for image clustering. *IEEE Transactions on Knowledge and Data Engineering*, 33:3594–3606, 2020.

[Xu *et al.*, 2019] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, pages 1–13, 2019.

[Xu *et al.*, 2023] Cai Xu, Wei Zhao, Jinglong Zhao, Ziyu Guan, Yaming Yang, Long Chen, and Xiangyu Song. Progressive deep multi-view comprehensive representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10557–10565, 2023.

[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*, pages 16129–16137, 2024.

[Zhang *et al.*, 2020] Changqing Zhang, Yajie Cui, Zongbo Han, Joey Tianyi Zhou, Huazhu Fu, and Qinghua Hu. Deep partial multi-view learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44:2402–2415, 2020.