Graph Meets Deep Unfolding: An Interpretable Mutual-benefit Multi-view Learning Network

Renjie Lin[1], **Hongzhi He**[2], **Yilin Wu**[3], **Shide Du**[3], **Le Zhang**[1*]
[1] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2] Maynooth International Engineering College, Fuzhou University, Fuzhou, China
[3] College of Computer and Data Science, Fuzhou University, Fuzhou, China
renjielin1993@gmail.com, hongzhihe1858@gmail.com, dushidems@gmail.com, yinnwu510@gmail.com, lezhang@uestc.edu.cn

Abstract

Significant efforts have been focused on enhancing the utilization of multiple node features and topological structures in multi-view graph learning through explicit model-driven and implicit deep learning-based methodologies. The former excels in embedding prior knowledge, thereby offering theoretical interpretability but is limited in application flexibility due to manual parameter selection. In contrast, the latter leverages automatic differentiation, providing greater flexibility but lacking theoretical interpretability due to their opaque nature. Motivated by these observations, we propose an interpretable deep unfolding network for mutual-benefit multi-view graph learning, aiming to combine the strengths of both approaches. Specifically, we employ the Alternating Direction Method of Multipliers (ADMM) to solve a multi-view graph learning model with sparse and low-rank constraints. This solution is then integrated into deep unfolding networks to enhance interpretability. Furthermore, we convert optimization conditions into implicit losses and utilize automatic differentiation to update parameters, reducing the need for manual tuning and increasing flexibility. This integration optimizes multi-view learning for a graph representation that balances interpretability and flexibility. Empirical evaluations on six diverse datasets demonstrate the effectiveness and superiority of the proposed method over state-of-the-art approaches.

Code — https://github.com/rl227/IMML-Net

Introduction

Multi-modal artificial intelligence has revolutionized various fields, driving advancements in areas such as data analysis (Cheng et al. 2024; Niyogi, Lindquist, and Maiti 2024; Kou et al. 2024), machine learning (Gu and Zhu 2024; Hu et al. 2024b), and network science (Wang et al. 2023; Luo et al. 2025). In this context, multi-view graph learning has gained significant attention due to its ability to leverage multiple topological structures to improve data representation (Yang et al. 2024; Hu et al. 2024a). Traditional methods for multi-view graph learning can be broadly categorized into explicit model-based and implicit deep-based approaches.

Explicit model-based methods (Wang et al. 2021; Yu et al. 2024) excel in integrating structured prior knowledge into graph learning through precise mathematical formulations. This approach provides robust theory-level interpretability by transparently common graph patterns through regularization mechanisms (Wang et al. 2022a). For instance, multi-view sparse learning (Wang et al. 2022a) leverages such constraints to identify critical cross-view graph structures while suppressing redundant information through sparse optimization. However, these methods demand extensive manual parameter tuning for graph weight optimization, significantly limiting their application-level flexibility due to computational burdens, expertise requirements, and poor adaptability across diverse datasets or tasks. Conversely, ***implicit deep-based methods*** (Liu, Tong, and Chen 2023; Li et al. 2024) leverage the automatic differentiation capabilities of deep architectures to achieve exceptional application-level flexibility. This enables dynamic adaptation to complex multi-view dependencies through joint optimization and neural architecture reconfiguration, exemplified by multi-view networks (Lin et al. 2023) that process heterogeneous graph data while autonomously adjusting parameters across views. Such inherent adaptability facilitates deployment in evolving real-world scenarios. Nevertheless, the opaqueness of deep architectures fundamentally obstructs interpretability, complicating mechanistic understanding of graph structure generation and limiting explanatory transparency. Despite resolving manual tuning limitations, this black-box nature remains a core trade-off against the mathematical clarity offered by model-based alternatives.

Deep unfolding networks (Gregor and LeCun 2010; Zheng et al. 2023; Joukovsky, Eldar, and Deligiannis 2024; Fang et al. 2024b; Du et al. 2025), a novel neural architecture, effectively alleviate interpretability issues while achieving strong results across multiple tasks. These methods map optimization processes of explicit model-based methods to learnable implicit deep networks via end-to-end training. This approach uniquely balances interpretability from model-based frameworks with the flexibility of deep learning for automatic tuning. Consequently, we propose to translate multi-view graph optimizations into equivalent deep networks to construct interpretable yet flexible graph learning architectures leveraging automatic differentiation.

*Corresponding author.

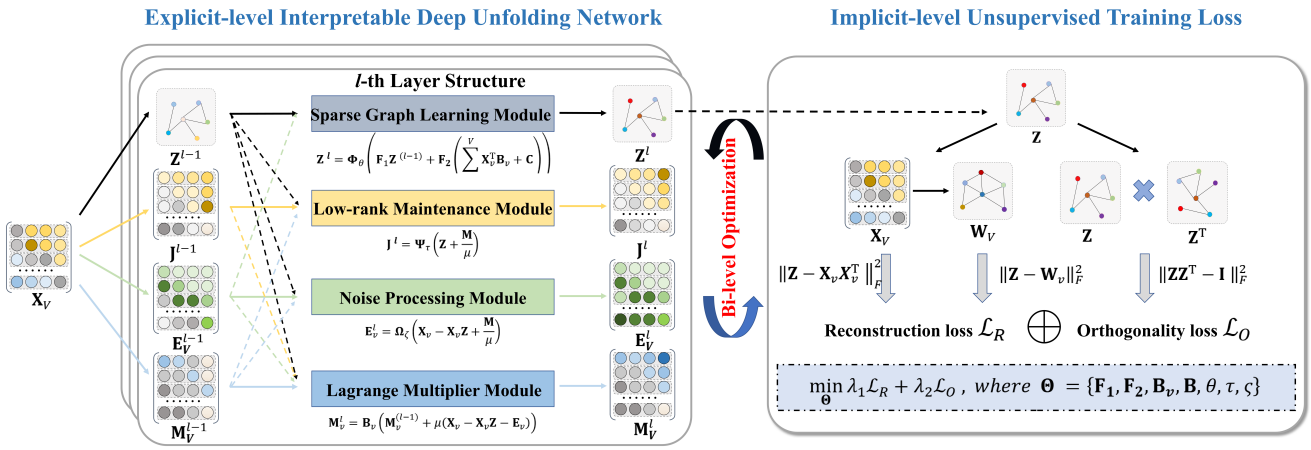

Figure 1: An overview of the proposed method.

Inspired by these insights, we propose an interpretable deep unfolding network for mutual-benefit multi-view graph learning (IMML-Net), which fuses the strengths of explicit model-based and implicit deep-based approaches. Our framework balances **theory-level interpretability** and **application-level flexibility** to overcome existing limitations. First, we utilize the Alternating Direction Method of Multipliers (ADMM) optimizer to effectively solve the multi-view graph learning model with sparse and low-rank constraints. This iterative solution of mathematical support is incorporated as prior knowledge into the construction of explicit-level deep unfolding networks, thereby enhancing the theory-level interpretability. Second, we convert certain conditions into implicit-level losses and employ the automatic differentiation rules of deep learning to update parameters. This reduces the need for manual parameter adjustment, thereby improving the application-level adaptability. This dual-path integration performs mutual-benefit optimization, bridging the method gap for explainable yet adaptable graph learning systems. The proposed framework can be illustrated as Fig. 1, and the main contributions can be summarized as follows:

- We propose an interpretable deep unfolding network for mutual-benefit multi-view graph learning, combining the strengths of explicit model-based and implicit deep-based methodologies.
- We employ ADMM to solve multi-view graph learning with sparse and low-rank constraints, integrating it into deep unfolding networks for better interpretability and converting specific conditions into implicit losses updated via automatic differentiation for flexibility.
- Empirical evaluations across diverse datasets show the effectiveness and superiority of the proposed method compared to state-of-the-art approaches.

Related Work

In this section, the relevant work details and summary are as follows. The proposed method is closely related to multi-view graph learning and deep unfolding networks.

Multi-view Graph Learning

Multi-view graph learning aims to integrate node and topological information from multiple sources to enhance task performance (Wang, Zhang, and Zhou 2025b; Wang et al. 2025; Ji and Feng 2025a; Wang, Zhang, and Zhou 2025a). Two common approaches are Explicit model-based and implicit deep-based methods. **Explicit model-based methods** perform joint graph learning of features and structures by constructing optimization problems with embedded prior knowledge for better interpretability (Luo et al. 2024; Chen et al. 2025). For example, Li *et al.* (Li et al. 2021) used flexible regression residuals, Wang *et al.* (Wang et al. 2022c) built anchor correspondence, and Ji *et al.* (Ji and Feng 2025b) incorporated tensor regularization. Although these excellent methods can learn an interpretable multi-view graph with theoretical support, they are limited by the time-consuming manual parameter adjustment and extensive empirical verification required. This limits their adaptability in different datasets or tasks, lacking a certain application-level flexibility. Unlike the above methods, **implicit deep-based methods** utilize deep networks and automatic differentiation mechanisms to integrate nodes and topology for graph learning. Lin *et al.* (Lin and Kang 2021) implemented node-sampling anchor matrices, Zhang *et al.* (Zhang et al. 2023) designed pathway-dropping architectures, and Du *et al.* (Du et al. 2024) proposed consensus-generative graph fusion. Although these methods have learned a multi-view graph with good task performance, their black-box nature unfortunately impairs the theoretical interpretability of the decision-making processes and model behaviors involved in learning these graphs.

Interpretable Deep Unfolding Networks

Deep unfolding networks (DUNs) transformatively integrate optimization algorithms' theoretical strengths with deep learning's adaptive nature by recasting iterative optimization steps as neural network layers. This significantly enhances efficiency in solving complex problems (Fan et al. 2024; Hu et al. 2024c), ensuring flexibility through end-to-end parameter learning while maintaining explicit methods' in-

Notations	Descriptions
V, C	The number of views and classes.
N	The number of total training samples.
D_v	The dimensions of the v-th view feature.
$\{\mathbf{X}_v\}_{v=1}^V$	$\mathbf{X}_v \in \mathbb{R}^{D_v \times N}$ is the v-view training samples.
$\{\mathbf{E}_v\}_{v=1}^V$	$\mathbf{E}_v \in \mathbb{R}^{D_v \times N}$ is the v-th noise matrix.
$\{\mathbf{M}_v\}_{v=1}^V$	$\mathbf{M}_v \in \mathbb{R}^{D_v \times N}$ is the view-specific Lagrange multiplier.
$\mathbf{Z}$	$\mathbf{Z} \in \mathbb{R}^{N \times N}$ is the learned graph.
$\mathbf{J}$	$\mathbf{J} \in \mathbb{R}^{N \times N}$ is the intermediate variable matrix.
$\mathbf{M}$	$\mathbf{M} \in \mathbb{R}^{N \times N}$ is the common-variable Lagrange multiplier.

Table 1: Essential notations and descriptions.

terpretability. Applications demonstrate broad impact: Ning *et al.* developed model-guided DUNs for interpretable image super-resolution (Ning et al. 2021); Fu *et al.* created model-driven structures for JPEG artifact removal (Fu et al. 2022); Zhou *et al.* integrated physical mechanisms into pansharpening tasks (Zhou et al. 2023); Du *et al.* enhanced trustworthiness in open-set learning (Du et al. 2023); Fang *et al.* balanced interpretation and extensibility in representation learning (Fang et al. 2024a); and Huang *et al.* designed exclusion mechanisms for reflection removal (Huang et al. 2025). Similarly, our work utilizes deep unfolding networks as a bridge between explicit model-based methods and implicit deep-based methods to construct a multi-view learning network that is compatible with theory-level interpretability and application-level flexibility. To our knowledge, we have innovatively adopted deep unfolding networks as a bridge to inject prior knowledge into the first instance of an equivalent deep learning method, endowing interpretable multi-view graph learning with adaptive flexibility.

Proposed Model

In this section, we propose a multi-view deep unfolding network for graph learning. First, we design an interpretable deep unfolding network based on ADMM iteration for multi-view graph learning, integrating prior-knowledge, such as sparsity and low-rank constraints. Second, we construct an implicit-level unsupervised training loss and utilize it to update the associated parameters by transforming the specific multi-view attribute conditions. This overcomes the search limitations of fixed values, offering a flexible graph learning space that enhances application adaptability. The key notations are summarized in Table 1.

Model Formulation and Optimization

Suppose that $\mathbf{X} \in \mathbb{R}^{D \times N}$ is the data point, where D features and N samples, facing three low-quality challenges (Liu et al. 2012): 1) Feature redundancy in shared spaces, 2) Weak global structure compactness hindering pattern extraction, 3) High noise/corruption disrupting data organization. According to (Liu et al. 2012), these issues motivate the decomposition $\mathbf{X} = \mathbf{DZ} + \mathbf{E}$ where $\mathbf{Z} \in \mathbb{R}^{N \times N}$ is a common representation matrix and $\mathbf{E} \in \mathbb{R}^{D \times N}$ an error matrix. During this data reconstruction process, it is crucial to capture important information, maintain a compact global structure, and eliminate noise errors. To this end, a constrained optimization problem is formulated, ensuring both the sparsity and low-rank nature of the representation $\mathbf{Z}$ while eliminating noise through $\mathbf{E}$ as

$$\min_{\mathbf{Z},\mathbf{E}} \alpha\|\mathbf{Z}\|_1 + \beta\|\mathbf{Z}\|_* + \gamma\|\mathbf{E}\|_{2,1}, \text{s.t., } \mathbf{X} = \mathbf{DZ} + \mathbf{E}, \tag{1}$$

where $\|\cdot\|_1$, $\|\cdot\|_*$, and $\|\cdot\|_{2,1}$ represent l_1-norm sparse, nuclear-norm low-rank, and $l_{2,1}$-norm row sparse terms, respectively. α, β, and γ are three trade-off parameters. Problem (1) aims to preserve vital global structures and significant local features in a common representation while reducing the impact of data errors.

In practice, multi-view scenarios are more prevalent, where models often face issues related to low-quality data across multiple views. This typically refers to certain modalities or views in the data being incomplete or degraded due to factors such as noise, missing values, sensor inaccuracies, and other imperfections. Such data often also exhibit characteristics like information redundancy, weak structure, or high levels of noise. Therefore, Problem (1) is transformed into the following optimization problem as

$$\min_{\mathbf{Z},\mathbf{E}_v} \alpha\|\mathbf{Z}\|_1 + \beta\|\mathbf{Z}\|_* + \sum_{v=1}^{V} \gamma\|\mathbf{E}_v\|_{2,1}, \text{s.t., } \mathbf{X}_v = \mathbf{X}_v\mathbf{Z} + \mathbf{E}_v. \tag{2}$$

Problem (2) learns a common low-rank representation $\mathbf{Z}$ to minimize redundancy while isolating view-specific noise $\mathbf{E}_v$, enabling unified graph mapping. On this basis, as a common practice for optimizing nuclear norm, we need to introduce an intermediate variable $\mathbf{J}$ into Problem (2) as

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{J},\mathbf{E}_v} \alpha\|\mathbf{Z}\|_1 + \beta\|\mathbf{J}\|_* + \sum_{v=1}^{V} \gamma\|\mathbf{E}_v\|_{2,1}, \\ \text{s.t., } \mathbf{X}_v = \mathbf{X}_v\mathbf{Z} + \mathbf{E}_v, \mathbf{Z} = \mathbf{J}. \end{aligned} \tag{3}$$

The Augmented Lagrange function can be reformulated

$$\begin{aligned} &\mathcal{L}(\mathbf{Z}, \mathbf{J}, \mathbf{E}_v, \mathbf{M}_v, \mathbf{M}) = \alpha\|\mathbf{Z}\|_1 + \beta\|\mathbf{J}\|_* + \sum_{v=1}^{V} \Big(\gamma\|\mathbf{E}_v\|_{2,1} + \\ &\langle \mathbf{M}_v, \mathbf{X}_v - \mathbf{X}_v\mathbf{Z} - \mathbf{E}_v\rangle + \frac{\mu}{2}\|\mathbf{X}_v - \mathbf{X}_v\mathbf{Z} - \mathbf{E}_v\|_F^2\Big) + \\ &\langle \mathbf{M}, \mathbf{Z} - \mathbf{J}\rangle + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{J}\|_F^2, \end{aligned} \tag{4}$$

where μ is a penalty parameter, and $\mathbf{M}_v$ and $\mathbf{M}$ are view-specific and common Lagrange multipliers, respectively. The final joint optimization problem can be presented as

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{J},\mathbf{E}_v,\mathbf{M}_v,\mathbf{M}} &\alpha\|\mathbf{Z}\|_1 + \beta\|\mathbf{J}\|_* + \sum_{v=1}^{V} \Big(\gamma\|\mathbf{E}_v\|_{2,1} + \frac{\mu}{2}\|\mathbf{X}_v - \\ &\mathbf{X}_v\mathbf{Z} - \mathbf{E}_v + \frac{\mathbf{M}_v}{\mu}\|_F^2 - \frac{1}{2\mu}\|\mathbf{M}_v\|_F^2\Big) + \\ &\frac{\mu}{2}\|\mathbf{Z} - \mathbf{J} + \frac{\mathbf{M}}{\mu}\|_F^2 - \frac{1}{2\mu}\|\mathbf{M}\|_F^2. \end{aligned} \tag{5}$$

Problem (5) formalizes common graph learning and view denoising. We solve this using ADMM, transforming each sub-variable into a sub-problem to be solved through alternating iterations until the optimal solution is found.

1) Optimization with Respect to $\mathbf{Z}$: Fix other variables, the objective problem for $\mathbf{Z}$ is

$$\min_{\mathbf{Z}} \alpha\|\mathbf{Z}\|_1 + \sum_{v=1}^{V} \big(\frac{\mu}{2}\|\mathbf{X}_v - \mathbf{X}_v\mathbf{Z} - \mathbf{E}_v + \frac{\mathbf{M}_v}{\mu}\|_F^2\big) + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{J} + \frac{\mathbf{M}}{\mu}\|_F^2. \tag{6}$$

The solution to this problem can be obtained using the iterative soft thresholding algorithm (ISTA) operator (Beck and Teboulle 2009). First, we take the derivative of the convex part of the optimization function corresponding to the above problem (6), we have

$$\nabla\mathcal{L}_1(\mathbf{Z}) = \mu\sum_{v=1}^{V}\mathbf{X}_v^\top\left(\mathbf{X}_v\mathbf{Z}-(\mathbf{X}_v-\mathbf{E}_v+\frac{\mathbf{M}_v}{\mu})\right)+\mu\left(\mathbf{Z}-(\mathbf{J}-\frac{\mathbf{M}}{\mu})\right), \tag{7}$$

where $\mathcal{L}_1(\cdot)$ denotes the convex part of the optimization function of Problem (6), and $\nabla\mathcal{L}_1(\cdot)$ is the derivative of $\mathcal{L}_1(\cdot)$. Then, following the solution approach of ISTA, we can obtain

$$\begin{aligned}\mathbf{Z}^{(l+1)} &= \mathbf{Prox}_{\frac{\alpha}{L}}\left(\mathbf{Z}^{(l)}-\frac{1}{L}\nabla\mathcal{L}_1(\mathbf{Z}^{(l)})\right)\\ &= \mathbf{Prox}_{\frac{\alpha}{L}}\left(\mathbf{Z}^{(l)}-\frac{1}{L}\left(\mu\sum_{v=1}^{V}\mathbf{X}_v^\top\left(\mathbf{X}_v\mathbf{Z}^{(l)}-\mathbf{B}_v\right)+\mu\left(\mathbf{Z}^{(l)}-\mathbf{C}\right)\right)\right),\end{aligned} \tag{8}$$

where L is the Lipschitz constant of $\nabla\mathcal{L}_1(\cdot)$, l denotes the l-th iteration, $\mathbf{B}_v = \mathbf{X}_v - \mathbf{E}_v + \frac{\mathbf{M}_v}{\mu}$, and $\mathbf{C} = \mathbf{J} - \frac{\mathbf{M}}{\mu}$. Specifically, given a matrix $\mathbf{A}\in\mathbb{R}^{N_1\times N_2}$ is a matrix of any dimension with a sparse threshold $\xi_1 \geq 0$, we have an ISTA operator as

$$\mathbf{Prox}_{\xi_1}(\mathbf{A}) = \mathbf{sign}(\mathbf{A})\odot\max(|\mathbf{A}|-\xi_1,0), \tag{9}$$

where $\mathbf{sign}(\mathbf{A})$ is a symbol matrix of $\mathbf{A}$, $\odot$ is the multiplication of two matrix element levels, and $|\cdot|$ is an absolute value operation.

2) Optimization with Respect to $\mathbf{J}$: Fix other variables, the optimization problem for $\mathbf{J}$ is

$$\min_{\mathbf{J}}\beta\|\mathbf{J}\|_* + \frac{\mu}{2}\|\mathbf{Z}-\mathbf{J}+\frac{\mathbf{M}}{\mu}\|_F^2. \tag{10}$$

The solution to this problem can be obtained using the singular value thresholding (SVT) operator (Liu et al. 2012) as

$$\mathbf{J}^{(l+1)} = \mathbf{SVT}_{\frac{\beta}{\mu}}\left(\mathbf{Z}+\frac{\mathbf{M}}{\mu}\right). \tag{11}$$

As for $\mathbf{SVT}(\cdot)$ method, we need to first introduce the following operations. Given a matrix $\mathbf{A}\in\mathbb{R}^{N_1\times N_2}$ of rank r, a singular value decomposition (SVD) can be utilized as

$$\mathbf{A} = \mathbf{U}\,\mathbf{diag}(\{\sigma_i\}_{1\leq i\leq r})\mathbf{V}, \tag{12}$$

where $\mathbf{U}\in\mathbb{R}^{N_1\times r}$ and $\mathbf{V}\in\mathbb{R}^{r\times N_2}$ are two matrices with orthonormal columns, and $\mathbf{diag}(\cdot)$ denotes a matrix where the diagonal elements have positive singular values $\{\sigma_i\}_{1\leq i\leq r}$ while the remaining elements are zero. Then, for a singular value threshold $\xi_2\geq 0$, a SVT operator as

$$\mathbf{SVT}_{\xi_2}(\mathbf{A}) = \mathbf{U}\,\mathbf{diag}(\{\max(\sigma_i-\xi_2,0)\}_{1\leq i\leq r})\mathbf{V}. \tag{13}$$

By substituting it into Eq. (11), the solution to sub-problem (10) can be obtained.

3) Optimization with Respect to $\mathbf{E}_v$: Similarly, we obtain the optimization problem for solving variable $\mathbf{E}_v$ as

$$\min_{\mathbf{E}_v}\gamma\|\mathbf{E}_v\|_{2,1} + \frac{\mu}{2}\|\mathbf{X}_v-\mathbf{X}_v\mathbf{Z}-\mathbf{E}_v+\frac{\mathbf{M}_v}{\mu}\|_F^2. \tag{14}$$

The solution to this problem can be obtained using the group ISTA (GISTA) (Liu, Lin, and Yu 2010) operator as

$$\mathbf{E}_v^{(l+1)} = \mathbf{Prog}_{\frac{\gamma}{\mu}}\left(\mathbf{X}_v-\mathbf{X}_v\mathbf{Z}+\frac{\mathbf{M}_v}{\mu}\right). \tag{15}$$

Specifically, given a matrix $\mathbf{A}\in\mathbb{R}^{N_1\times N_2}$ is a matrix of any dimension with a column sparse threshold $\xi_3\geq 0$, we have a GISTA operator as

$$\mathbf{Prog}_{\xi_3}(\mathbf{A}) = \frac{\max(\|\mathbf{A}_i\|_2-\xi_3,0)}{\|\mathbf{A}_i\|_2}\mathbf{A}_i, \tag{16}$$

where $\mathbf{A}_i$ is the i-th column of $\mathbf{A}$. Eq. (15) with (16) is the iteration solution of sub-problem (14).

4) Optimization with Respect to Lagrange Multiplier $\mathbf{M}_v$, $\mathbf{M}$: Finally, we need to fix the corresponding variables to solve the Lagrange multipliers $\mathbf{M}_v$ and $\mathbf{M}$ from the following optimization problem as

$$\begin{aligned}\min_{\mathbf{M}_v,\mathbf{M}}\sum_{v=1}^{V}\Big(\frac{\mu}{2}\|\mathbf{X}_v-\mathbf{X}_v\mathbf{Z}-\mathbf{E}_v+\frac{\mathbf{M}_v}{\mu}\|_F^2-\frac{1}{2\mu}\|\mathbf{M}_v\|_F^2\Big)+\\ \frac{\mu}{2}\|\mathbf{Z}-\mathbf{J}+\frac{\mathbf{M}}{\mu}\|_F^2-\frac{1}{2\mu}\|\mathbf{M}\|_F^2.\end{aligned} \tag{17}$$

The corresponding iteration form is

$$\mathbf{M}_v^{(l+1)} = \mathbf{M}_v^{(l)} + \mu(\mathbf{X}_v-\mathbf{X}_v\mathbf{Z}-\mathbf{E}_v), \tag{18}$$

$$\mathbf{M}^{(l+1)} = \mathbf{M}^{(l)} + \mu(\mathbf{Z}-\mathbf{J}), \tag{19}$$

where penalty parameter μ is obtained as

$$\mu^{(l+1)} = \min\left(\mu_{\max},\rho\mu^{(l)}\right),\ \text{where}$$

$$\rho = \begin{cases}\rho_0, \text{ if } \mu^{(l)}\cdot\max\left(\left\|\mathbf{Z}^{(l+1)}-\mathbf{Z}^{(l)}\right\|_F,\ \left\|\mathbf{J}^{(l+1)}-\mathbf{J}^{(l)}\right\|_F,\right.\\ \qquad\left.\left\|\mathbf{E}_v^{(l+1)}-\mathbf{E}_v^{(l)}\right\|_F\right)/\|\mathbf{X}_v\|_F<\varepsilon;\\ 1, \text{ otherwise.}\end{cases} \tag{20}$$

where ρ is a scale factor, ρ_0, $\mu_{\max}$, and ε are three predetermined constants. Therefore, we present the ADMM solution for the overall optimization Problem (5). Next, we will construct corresponding interpretable deep unfolding networks based on these iterations.

Explicit-level Interpretable Network Architecture

In this section, we construct an interpretable deep unfolding network from Problem 5's framework, unfolding iterative optimization into modules embedding sparsity and low-rank constraints. The explicit network integrates four modules: sparse graph learning, low-rank maintenance, noise processing, and Lagrange multiplier updates—collectively achieving unified graph representations. Each module corresponds to an ADMM variable solution, forming a neural cycle for interpretable learning. This structure allows us to build a deep network that is both interpretable.

1) Sparse Graph Learning Module (SparseModule): To address feature information redundancy, SparseModule focuses on extracting the most relevant and non-redundant features from the multi-view data. By enforcing sparsity in the learned graph representation $\mathbf{Z}$, it ensures that only the most critical connections between graph nodes are preserved, reducing redundancy of the shared space. Thus, the update for SparseModule is given by

$$\mathbf{Z}^{(l+1)} = \mathbf{\Phi}_\theta\left(\mathbf{F}_1\mathbf{Z}^{(l)}+\mathbf{F}_2\left(\sum_{v=1}^{V}\mathbf{X}_v^\top\mathbf{B}_v+\mathbf{C}\right)\right), \tag{21}$$

where $\mathbf{F}_1 \in \mathbb{R}^{N\times N} = \left(\mathbf{I} - \frac{\mu}{L}\sum_{v=1}^{V}\mathbf{X}_v^\top\mathbf{X}_v - \frac{\mu}{L}\mathbf{I}\right)$ and $\mathbf{F}_2 \in \mathbb{R}^{N\times N} = \frac{\mu}{L}\mathbf{I}$ are two learnable network layers, $\mathbf{I} \in \mathbb{R}^{N\times N}$ is a unit matrix, and $\mathbf{\Phi}_\theta(\cdot)$ denotes a deep network component mapping of ISTA operator $\mathbf{Prox}_{\frac{\alpha}{L}}(\cdot)$ due to the similarity as the description of (Beck and Teboulle 2009) in functionality,

$$\mathbf{\Phi}_\theta(\mathbf{A}) = \mathbf{ReLU}(\mathbf{A} - \theta) - \mathbf{ReLU}(-\mathbf{A} - \theta), \tag{22}$$

where θ is a learnable sparse threshold, which applies it to ensure that only the most informative features remain.

2) Low-rank Maintenance Module (LRModule): To improve global structure compactness, LRModule preserves the underlying low-rank structure $\mathbf{J}$ of the data. By focusing on the most significant patterns, this module compresses the global representation, enabling the extraction of core patterns. It ensures that the multi-view data is embedded in a compact manner, aiding in the identification of shared and important structures. Therefore, LRModule is updated as

$$\mathbf{J}^{(l+1)} = \mathbf{\Psi}_\tau\left(\mathbf{Z} + \frac{\mathbf{M}}{\mu}\right), \tag{23}$$

where $\mathbf{\Psi}_\tau(\cdot)$ denotes a deep network component mapping of SVT operator $\mathbf{SVT}_{\frac{\beta}{\mu}}(\cdot)$, which is consistent with the description in (Liu et al. 2012),

$$\mathbf{\Psi}_\tau(\mathbf{A}) = \mathbf{U}\,\mathbf{ReLU}(\mathbf{diag}(\{\sigma_i - \tau\}_{1\le i\le r}))\mathbf{V}, \tag{24}$$

where τ is a learnable singular value threshold. This update helps in maintaining a compressed global structure by retaining only the most significant components of the data.

3) Noise Processing Module (NoiseModule): To handle high levels of noise, NoiseModule separates and removes noise and view-specific errors from the data, ensuring that only clean and useful information is incorporated into the learned representation. By isolating noise in the form of the error matrix $\mathbf{E}_v$, it allows the model to better focus on the essential patterns within the data and avoid disruptions caused by noise or corruptions. Here, NoiseModule is presented as

$$\mathbf{E}_v^{(l+1)} = \mathbf{\Omega}_\zeta\Big(\mathbf{X}_v - \mathbf{X}_v\mathbf{Z} + \frac{\mathbf{M}_v}{\mu}\Big), \tag{25}$$

where $\mathbf{\Omega}_\zeta(\cdot)$ denotes a deep network component mapping of GISTA operator $\mathbf{Prog}_{\frac{\gamma}{\mu}}(\cdot)$, and the starting point of its function is same as (Liu, Lin, and Yu 2010),

$$\mathbf{\Omega}_\zeta(\mathbf{A}) = \frac{\mathbf{ReLU}(\|\mathbf{A}_i\|_2 - \zeta)}{\|\mathbf{A}_i\|_2}\mathbf{A}_i, \tag{26}$$

where ζ is a learnable column sparse threshold. This ensures that noise is progressively filtered out, contributing to a cleaner representation for all views.

4) Lagrange Multiplier Module (LagModule): Finally, LagModule ensures that the sparse and low-rank constraints are properly balanced while addressing the challenges posed by redundancy, weak structure, and noise. It maintains the connection between the sparse and low-rank components through updates of the Lagrange multipliers $\mathbf{M}_v$ and $\mathbf{M}$, ensuring convergence to an optimal solution that satisfies the constraints. The updates for LagModule are

$$\mathbf{M}_v^{(l+1)} = \mathbf{B}_v\left(\mathbf{M}_v^{(l)} + \mu(\mathbf{X}_v - \mathbf{X}_v\mathbf{Z} - \mathbf{E}_v)\right), \tag{27}$$

Algorithm 1: IMML-Net

Require: Multi-view data $\{\mathbf{X}_v\}_{v=1}^V$, layer number L, training epoch E.
Ensure: Graph representation $\mathbf{Z}$.

```
Initialize {W_v}_{v=1}^V, Z, and parameter set Θ;
for e = 1 → E do
    for l = 1 → L do
        Calculate Z^(l) by SparseModule (21);
        Update J^(l) by LRModule (23);
        Compute E_v^(l) by NoiseModule (25);
        Update M_v^(l) and M^(l) by LagModule (27)-(28);
    end for
    Update implicit-level training loss (31);
    Update learnable parameters Θ by backpropagation;
end for
return Graph representation Z.
```

$$\mathbf{M}^{(l+1)} = \mathbf{B}\left(\mathbf{M}^{(l)} + \mu(\mathbf{Z} - \mathbf{J})\right), \tag{28}$$

where $\mathbf{B}_v$ and $\mathbf{B}$ are two linear layers to provide additional flexibility, μ is converted from a complicated update ($\mu^{(l+1)} = \min\left(\mu_{\max}, \rho\mu^{(l)}\right)$) into a self-learning parameter for more conveniently adaptive punishment and adjustment.

Implicit-level Unsupervised Training Loss

In order to further leverage the application-level flexibility, we design an implicit unsupervised loss for training unfolding networks using automatic differentiation. By converting key conditions into loss functions, these components—reconstruction and orthogonality losses—enforce sparsity and low-rank representation learning while preserving essential information and cross-view consistency. The implicit training loss consists of two main components: reconstruction loss and orthogonality loss.

To ensure the online learning representation does not overly lose the original structure, we use reconstruction loss. This preserves feature semantics and topological structure within each view. Enforcing consistency between representation $\mathbf{Z}$, affinity matrix $\mathbf{W}_v$, and original feature matrix $\mathbf{X}_v$ helps capture key info across multi-view data. The reconstruction loss is expressed as

$$\mathcal{L}_R = \sum_{v=1}^{V}(\|\mathbf{Z} - \mathbf{W}_v\|_F^2 + \|\mathbf{Z} - \mathbf{X}_v\mathbf{X}_v^T\|_F^2), \tag{29}$$

where the first term ensures consistency between the representation and the affinity matrix, the second term ensures the representation reconstructs the original feature matrix.

Second, to maintain a structured and independent latent representation, orthogonality loss enforces that the columns of the representation are orthogonal to each other. his ensures that the learned representation remains independent across components, reducing redundancy and enhancing stability. The orthogonality loss is given as

$$\mathcal{L}_O = \|\mathbf{Z}\mathbf{Z}^T - \mathbf{I}\|_F^2, \tag{30}$$

where $\mathbf{I}$ is the identity matrix, ensuring the orthogonality of the representation.

Methods	ALOI			Caltech101-7			Caltech101-20			Notting-Hill			Scene15			Youtube		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
k-means	47.5	47.3	33.0	49.6	32.7	30.2	31.3	34.5	18.9	63.7	58.0	49.7	30.8	28.5	14.8	24.2	15.1	7.90
MCGC	52.4	52.5	25.9	53.6	42.9	38.4	54.6	50.5	38.8	52.3	60.1	41.3	42.3	41.9	24.7	30.0	17.4	8.80
GMC	64.9	61.8	32.9	69.2	60.6	59.4	45.6	38.5	32.8	31.2	55.2	62.2	38.1	41.6	29.1	21.7	17.1	5.40
LVMSC	69.5	72.5	59.2	37.3	61.9	73.3	25.0	44.9	67.5	72.6	67.5	60.9	40.6	40.3	24.8	20.1	15.80	3.10
CDMGC	47.1	49.7	52.6	76.3	60.5	62.3	61.4	50.7	38.9	63.9	68.8	72.9	34.4	40.2	19.1	20.6	16.2	6.30
FMVSC	57.1	71.1	51.0	61.5	57.0	64.0	62.5	43.9	63.6	71.3	67.7	69.3	45.7	45.8	25.8	36.1	25.4	13.5
OMSC	58.7	70.0	48.1	80.3	66.2	74.4	65.2	58.7	67.1	60.7	53.3	55.9	41.5	45.3	32.7	35.4	22.1	10.2
SMGC	64.6	18.8	48.9	82.1	68.7	80.2	69.4	56.4	74.5	60.4	51.7	55.7	42.5	44.3	23.0	35.4	20.2	9.85
RCAGL	54.2	67.8	50.1	80.1	66.3	72.8	70.3	61.9	72.6	75.5	60.1	63.2	41.9	54.8	31.7	34.1	17.4	12.6
DSRL-Net	67.1	65.7	56.6	79.4	70.2	74.9	68.5	65.4	75.5	65.3	56.4	62.8	43.3	44.6	26.0	32.6	20.7	3.38
SDSNE-Net	58.9	61.3	52.2	82.1	70.4	76.8	70.3	66.2	74.8	62.6	56.1	50.6	46.7	50.9	35.1	34.1	19.9	7.95
DSMVC-Net	74.3	48.2	52.4	80.6	73.2	80.1	71.2	67.2	70.9	69.6	53.9	59.5	48.9	42.8	34.1	33.7	**58.4**	10.3
CWCL-Net	68.7	66.6	57.9	84.9	73.3	82.7	73.1	65.5	73.3	73.9	65.7	64.3	56.0	42.3	27.5	**56.0**	21.7	12.4
IMML-Net	**76.5**	**80.8**	**67.4**	**86.7**	**75.6**	**85.5**	**74.9**	**69.9**	**79.2**	**89.5**	**80.8**	**81.9**	**58.3**	**58.1**	**38.9**	37.0	23.8	**14.7**

Table 2: Clustering performance (mean%) of all methods on six datasets. The best results are highlighted in bold.

To train the network, we use the joint training loss below to get the optimal parameters.

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_R + \lambda_2 \mathcal{L}_O, \tag{31}$$

where λ_1 and λ_2 are two trade-off parameters balancing the reconstruction and orthogonality losses. The parameters $\mathbf{\Theta} = \{\mathbf{\Theta}_R, \mathbf{\Theta}_O\} = \{\mathbf{F}_1, \mathbf{F}_2, \mathbf{B}_v, \mathbf{B}, \theta, \tau, \varsigma\}$ can be updated by the above training loss. By employing this implicit-level unsupervised training loss (31), we further enhance the framework's application-level flexibility while ensuring theory-level interpretability, achieving a mutually beneficial balance between explicit and implicit multi-view learning.

Summary and Complexity

1) Summary: Eqs. (21), (23), (25), (27), and (28) are the total explicit-level interpretable deep unfolding network architecture. This architecture combines the power of graph-based learning with deep unfolding techniques to create a structured, interpretable framework for multi-view learning. By embedding sparsity, low-rank constraints, and noise filtering into the iterative optimization process, the network ensures that the learned representations are both robust and efficient. In parallel, the implicit-level unsupervised training loss automatically updates network parameters, ensuring that the representations are consistent, compact, and independent across views. Together, these components provide a flexible yet interpretable solution to multi-view learning, balancing theoretical rigor with practical adaptability. Algorithm 1 presents the overall process of IMML-Net.

2) Complexity: The computational complexity of $\mathbf{Z}$ sub-block network costs $\mathcal{O}(n^3)$ with the SVD. It takes $\mathcal{O}(n^3)$ to compute $\mathbf{J}$ and $\mathbf{M}_2$ sub-block networks, and the complexity of $\mathbf{E}_v$ and $\mathbf{M}_{(1,v)}$ sub-block networks consumes $\mathcal{O}\sum_{v=1}^{V}(n^2 d_v)$. In summary, the overall complexity requires $\mathcal{O}(n^2(n + \sum_{v=1}^{V} d_v))$ for each training epoch.

Experiments

To verify the effectiveness of the subspace representation, we evaluate the proposed model on a mainstream task with six datasets and make a comparison with other state-of-the-art multi-view learning algorithms.

Datasets	# Samples	# Views	# Feature Dimensions	# Classes
ALOI	1,079	4	64/64/77/13	10
Caltech101-7	1,474	6	48/40/254/1,984/512/928	7
Caltech101-20	2,386	6	48/40/254/1,984/512/928	20
Notting-Hill	4,660	3	6,750/2,000/3,304	5
Scene15	4,485	3	1,800/1,180/1,240	15
Youtube	2,000	6	2,000/1,024/64/512/64/647	10

Table 3: A brief description of the tested datasets.

Experimental Setup

To validate our method, six publicly available datasets are employed: ALOI, Caltech101-7/20, Notting-Hill, Scene15, and YouTube (Details are presented in Table 3).

The compared multi-view clustering methods include: Explicit model-based methods: k-means, CDMGC (Huang et al. 2022), FMVSC (Wang et al. 2022d), OMSC (Chen et al. 2022), SMGC (Tan et al. 2023), and RCAGL (Liu et al. 2024); Implicit deep-based methods: DSRL-Net (Wang et al. 2022b), SDSNE-Net (Liu et al. 2022), DSMVC-Net (Tang and Liu 2023), and CWCL-Net (Yan et al. 2025).

Moreover, three well-known clustering evaluation metrics, including clustering accuracy (ACC), normalized mutual information (NMI), and adjusted rand index (ARI), are utilized for a quantitative comparison. The proposed method is implemented with PyTorch on a standard Ubuntu-16.04 operation system with a NVIDIA Tesla P100 GPU. The learning rate is set to be 0.001, and 200 epochs are uniformly trained. The trade-off parameters λ_1 and λ_2 are set as 1.0.

Experimental Results

To assess IMML-Net, we compare it with SOTA methods on six datasets (Table 2). The key observations can be made:

- **Overall Superiority of IMML-Net**: Achieves best performance across all datasets. Specifically, IMML-Net yields the highest ACC (e.g., 76.5% on ALOI, 86.7% on Caltech101-7, and 89.5% on Notting-Hill), NMI (e.g., 80.8% on ALOI and Notting-Hill), and ARI (e.g., 85.5% on Caltech101-7 and 81.9% on Notting-Hill), outperforming the best baseline by noticeable margins.

$\mathcal{L}_R$	$\mathcal{L}_O$	ALOI	Caltech101-7	Notting-Hill	Scene15	Youtube
		51.5 (0.0)	56.1 (0.0)	57.0 (0.0)	42.4 (1.5)	26.1 (1.3)
✓		71.8 (4.6)	80.8 (0.8)	79.4 (0.4)	50.9 (0.8)	30.5 (0.5)
	✓	72.8 (4.8)	81.8 (0.3)	83.8 (2.8)	53.7 (1.6)	26.9 (1.2)
✓	✓	**76.5 (4.0)**	**86.7 (1.2)**	**89.5 (1.1)**	**58.3 (1.4)**	**37.0 (1.9)**

Table 4: Ablation study of the proposed method.

- **Comparison with Explicit Methods**: IMML-Net consistently outperforms explicit models (e.g., FMVSC/S-MGC/RCAGL) despite their dataset-specific strengths, overcoming their generalization and flexibility limitations. By integrating sparse/low-rank constraints into unfolding networks and avoiding manual hyperparameter tuning, IMML-Net achieves superior adaptability.
- **Comparison with Implicit Deep Networks**: IMML-Net exceeds or matches deep baselines (e.g., CWCL-Net 84.9% ACC on Caltech101-7) across all datasets. The fusion of model-driven modules with implicit losses enhances both representation quality and interpretability.
- **Dataset-Wise Insights**: IMML-Net performs exceptionally well on datasets with complex modality combinations or noisy features like Notting-Hill and Scene15, showcasing its ability to denoise, compress, and align multi-view information through its modular architecture.
- **Summary**: Overall, the experimental results demonstrate the effectiveness and superiority of IMML-Net in clustering quality, robustness, and generalizability across diverse multi-view scenarios. Its bi-level interpretable design offers a promising approach for future multi-view representation learning models.

Ablation Study

We conduct an ablation study to evaluate the impact of the implicit-level losses in IMML-Net, namely the reconstruction loss ($\mathcal{L}_R$) and the orthogonality loss ($\mathcal{L}_O$), with results summarized in Table 4. Introducing either loss individually leads to notable performance gains, demonstrating their effectiveness in enhancing subspace representation. Notably, using $\mathcal{L}_O$ alone often yields better results than $\mathcal{L}_R$ alone, suggesting that the orthogonality constraint plays a more critical role in promoting discriminative clustering structures. The best performance is consistently achieved when both losses are applied jointly, highlighting their complementary nature and validating the necessity of the combined design in improving the robustness and expressiveness of the learned representations.

Parameter Sensitivity Analysis

To examine the robustness of IMML-Net with respect to the trade-off parameters, we analyze the impact of λ_1 and λ_2, which control the weights of the reconstruction loss and orthogonality loss, respectively. As shown in Fig. 2, we vary each parameter in the range $\{0.0001, 0.001, \cdots, 10.0\}$ while fixing the other. Performance remains stable across datasets for a wide range of values, indicating the proposed method is not overly sensitive to parameter choices. This robustness enhances its practicality in real-world applications. Furthermore, Fig. 3 illustrates the convergence behavior of

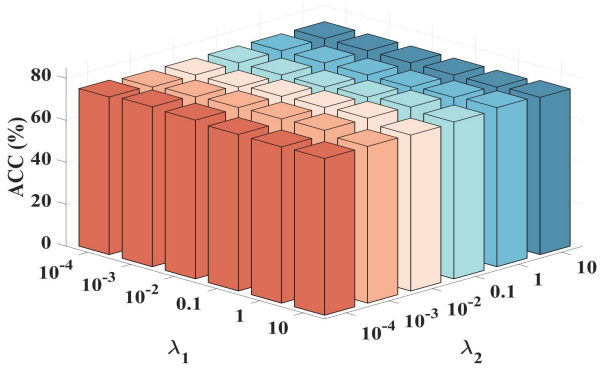

Figure 2: Parameter sensitivity of the proposed method with respect to λ_1 and λ_2 on Caltech101-20.

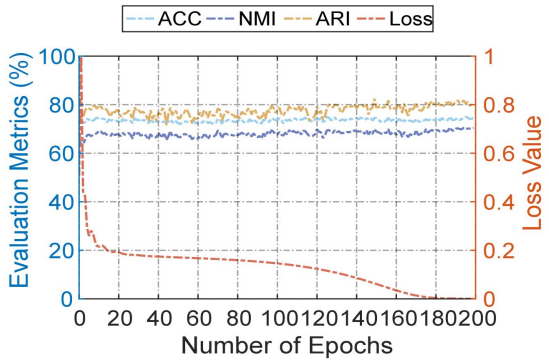

Figure 3: The curve of loss values and performance.

IMML-Net by plotting the training loss and performance on Caltech101-20. We observe that the loss values decrease rapidly during the early training stages and gradually stabilize, while the ACC, NMI, and ARI scores steadily increase and converge, demonstrating a strong correlation between the implicit-level training objective and clustering performance. In summary, IMML-Net exhibits favorable parameter robustness and stable optimization dynamics, confirming its practical applicability and training reliability.

Conclusion

In this paper, we presented an interpretable multi-view subspace representation learning framework from a bi-level optimization perspective. To overcome the existing problems in two mainstream multi-view subspace representation learning methods, we utilized bi-level optimization as a bridge to establish an interpretable network while preserving scalable solution spaces and good fitting ability. The proposed method constructed a forward propagation network by the proposed explicit-level problem, and then subspace representation learning was driven by two implicit-level losses. This bi-level optimization combines explicit and implicit problems to learn an interpretable latent subspace representation, outperforming current state-of-the-art multi-view learning methods. In future work, we will utilize the proposed framework as a guideline to construct a more task-specific bi-level (explicit and implicit-level) joint objective function for a broader range of application fields.

Acknowledgments

This work was supported by the Key Program for International Cooperation of Ministry of Science and Technology of China under Grant 2024YFE0100700.

References

Beck, A.; and Teboulle, M. 2009. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1): 183–202.

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

Chen, M.-S.; Lai, P.-Y.; Liao, D.-Z.; Wang, C.-D.; and Lai, J.-H. 2025. Graph Prompt Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7): 5794–5805.

Cheng, C.; Liu, W.; Fan, Z.; Feng, L.; and Jia, Z. 2024. A Novel Transformer Autoencoder for Multi-modal Emotion Recognition with Incomplete Data. *Neural Networks*, 172: 106111.

Du, S.; Cai, Z.; Wu, Z.; Pi, Y.; and Wang, S. 2024. UMCGL: Universal Multi-view Consensus Graph Learning with Consistency and Diversity. *IEEE Transactions on Image Processing*, 33: 3399–3412.

Du, S.; Fang, Z.; Lan, S.; Tan, Y.; Günther, M.; Wang, S.; and Guo, W. 2023. Bridging Trustworthiness and Open-World Learning: An Exploratory Neural Approach for Enhancing Interpretability, Generalization, and Robustness. In *Proceedings of the 31st ACM International Conference on Multimedia*, 8719–8729.

Du, S.; Wu, C.; Fang, Z.; Zhao, W.; Wu, Y.; Wang, C.; and Wang, S. 2025. LargeMvC-Net: Anchor-based Deep Unfolding Network for Large-scale Multi-view Clustering. In *Proceedings of the 33rd ACM International Ionference on Multimedia*, 1714–1723.

Fan, X.; Wang, X.; Gao, J.; Wang, J.; Luo, Z.; and Liu, R. 2024. Bi-level Learning of Task-Specific Decoders for Joint Registration and One-Shot Medical Image Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 11726–11735.

Fang, Z.; Du, S.; Cai, Z.; Lan, S.; Wu, C.; Tan, Y.; and Wang, S. 2024a. Representation Learning Meets Optimization-Derived Networks: From Single-View to Multi-view. *IEEE Transactions on Multimedia*, 26: 8889–8901.

Fang, Z.; Du, S.; Chen, Y.; and Wang, S. 2024b. Beyond the Known: Ambiguity-Aware Multi-view Learning. In *Proceedings of the 32nd ACM International Ionference on Multimedia*, 8518–8526.

Fu, X.; Wang, M.; Cao, X.; Ding, X.; and Zha, Z. 2022. A Model-Driven Deep Unfolding Method for JPEG Artifacts Removal. *IEEE Transactions on Neural Networks and Learning Systems*, 33(11): 6802–6816.

Gregor, K.; and LeCun, Y. 2010. Learning fast approximations of sparse coding. In *Proceedings of the 27th International Conference on Machine Learning*, 399–406.

Gu, S.; and Zhu, F. 2024. BAGAIL: Multi-Modal Imitation Learning from Imbalanced Demonstrations. *Neural Networks*, 174: 106251.

Hu, D.; Liu, S.; Wang, J.; Zhang, J.; Wang, S.; Hu, X.; Zhu, X.; Tang, C.; and Liu, X. 2024a. Reliable Attribute-missing Multi-view Clustering with Instance-level and feature-level Cooperative Imputation. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 1456–1466.

Hu, L.; Gao, L.; Liu, Z.; Pun, C.; and Feng, W. 2024b. COMMA: Co-articulated Multi-Modal Learning. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, 2238–2246.

Hu, Q.; Ma, J.; Gao, Y.; Jiang, J.; and Yuan, Y. 2024c. MAUN: Memory-Augmented Deep Unfolding Network for Hyperspectral Image Reconstruction. *IEEE/CAA Journal of Automatica Sinica*, 11(5): 1139–1150.

Huang, J.; Liu, T.; Chen, Z.; Liu, X.; Wang, M.; and Dragotti, P. L. 2025. A Lightweight Deep Exclusion Unfolding Network for Single Image Reflection Removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6): 4957–4973.

Huang, S.; Tsang, I.; Xu, Z.; and Lv, J. C. 2022. Measuring Diversity in Graph Learning: A Unified Framework for Structured Multi-view Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 34(12): 5869–5883.

Ji, J.; and Feng, S. 2025a. Anchors Crash Tensor: Efficient and Scalable Tensorial Multi-view Subspace Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4): 2660–2675.

Ji, J.; and Feng, S. 2025b. Anchors Crash Tensor: Efficient and Scalable Tensorial Multi-view Subspace Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4): 2660–2675.

Joukovsky, B.; Eldar, Y. C.; and Deligiannis, N. 2024. Interpretable Neural Networks for Video Separation: Deep Unfolding RPCA With Foreground Masking. *IEEE Transactions on Image Processing*, 33: 108–122.

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

Li, L.; Pan, Y.; Liu, J.; Liu, Y.; Liu, X.; Li, K.; Tsang, I. W.; and Li, K. 2024. BGAE: Auto-Encoding Multi-view Bipartite Graph Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 36(8): 3682–3696.

Li, Z.; Qiang, Q.; Zhang, B.; Wang, F.; and Nie, F. 2021. Flexible Multi-view Semi-Supervised Learning with Unified Graph. *Neural Networks*, 142: 92–104.

Lin, R.; Du, S.; Wang, S.; and Guo, W. 2023. Multi-Channel Augmented Graph Embedding Convolutional Network for Multi-view Clustering. *IEEE Transactions on Network Science and Engineering*, 10(4): 2239–2249.

Lin, Z.; and Kang, Z. 2021. Graph Filter-based Multi-view Attributed Graph Clustering. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*, 2723–2729.

Liu, C.; Liao, Z.; Ma, Y.; and Zhan, K. 2022. Stationary Diffusion State Neural Estimation for Multiview Clustering. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, 7542–7549.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2012. Robust Recovery of Subspace Structures by Low-Rank Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1): 171–184.

Liu, G.; Lin, Z.; and Yu, Y. 2010. Robust Subspace Segmentation by Low-Rank Representation. In *Proceedings of the 27th International Conference on Machine Learning*, 663–670.

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

Liu, Y.; Tong, S.; and Chen, Y. 2023. HMM-GDAN: Hybrid Multi-View and Multi-Scale Graph Duplex-Attention Networks for Drug Response Prediction in Cancer. *Neural Networks*, 167: 213–222.

Luo, R.; Huang, H.; Lee, I.; Xu, C.; Qi, J.; and Feng, X. 2025. FairGP: A Scalable and Fair Graph Transformer Using Graph Partitioning. In *Proceedings of The 39th Annual AAAI Conference on Artificial Intelligence*, 12319–12327.

Luo, R.; Huang, H.; Yu, S.; Han, Z.; He, E.; Zhang, X.; and Xia, F. 2024. FUGNN: Harmonizing Fairness and Utility Graph Neural Networks. In *Proceedings of the 30th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2072–2081.

Ning, Q.; Dong, W.; Shi, G.; Li, L.; and Li, X. 2021. Accurate and Lightweight Image Super-Resolution With Model-Guided Deep Unfolding Network. *IEEE Journal of Selected Topics in Signal Processing*, 15(2): 240–252.

Niyogi, P. G.; Lindquist, M. A.; and Maiti, T. 2024. A Tensor Based Varying-Coefficient Model for Multi-Modal Neuroimaging Data Analysis. *IEEE Transactions on Signal Processing*, 72: 1607–1619.

Tan, Y.; Liu, Y.; Huang, S.; Feng, W.; and Lv, J. 2023. Sample-level Multi-view Graph Clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 23966–23975.

Tang, H.; and Liu, Y. 2023. Deep Safe Multi-view Clustering: Reducing the Risk of Clustering Performance Degradation Caused by View Increase. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 202–211.

Wang, Q.; Ding, Z.; Tao, Z.; Gao, Q.; and Fu, Y. 2021. Generative Partial Multi-view Clustering with Adaptive Fusion and Cycle Consistency. *IEEE Transactions on Image Processing*, 30: 1771–1783.

Wang, Q.; Liu, R.; Chen, M.; and Li, X. 2022a. Robust Rank-Constrained Sparse Learning: A Graph-Based Framework for Single View and Multiview Clustering. *IEEE Tansactions on Cybernetics*, 52(10): 10228–10239.

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

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

Wang, S.; Liu, X.; Liao, Q.; Wen, Y.; Zhu, E.; and He, K. 2025. Scalable Multi-view Graph Clustering With Cross-View Corresponding Anchor Alignment. *IEEE Transactions on Knowledge and Data Engineering*, 37(5): 2932–2945.

Wang, S.; Liu, X.; Liu, S.; Jin, J.; Tu, W.; Zhu, X.; and Zhu, E. 2022c. Align Then Fusion: Generalized Large-scale Multi-view Clustering with Anchor Matching Correspondence. In *Advances in Neural Information Processing Systems*, 5882–5895.

Wang, S.; Liu, X.; Zhu, X.; Zhang, P.; Zhang, Y.; Gao, F.; and Zhu, E. 2022d. Fast Parameter-free Multi-view Subspace Clustering with Consensus Anchor Guidance. *IEEE Transactions on Image Processing*, 31: 556–568.

Wang, X.; Zhang, Y.; and Zhou, Y. 2025a. Highly Efficient Rotation-Invariant Spectral Embedding for Scalable Incomplete Multi-View Clustering. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, 21312–21320.

Wang, X.; Zhang, Y.; and Zhou, Y. 2025b. Multimodal Remote Sensing Image Clustering With Multiscale Spectral-Spatial Anchor Graphs. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–12.

Yan, W.; Zhang, Y.; Tang, C.; Zhou, W.; and Lin, W. 2025. Anchor-Sharing and Cluster-Wise Contrastive Network for Multiview Representation Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2): 3797–3807.

Yang, B.; Wu, J.; Zhang, X.; Zheng, X.; Nie, F.; and Chen, B. 2024. Discrete Correntropy-Based Multi-View Anchor-Graph Clustering. *Information Fusion*, 103: 102097.

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

Zhang, Z.; Wang, Q.; Tao, Z.; Gao, Q.; and Feng, W. 2023. Dropping Pathways Towards Deep Multi-view Graph Subspace Clustering Networks. In *Proceedings of the 31st ACM International Conference on Multimedia*, 3259–3267.

Zheng, Z.; Dai, W.; Xue, D.; Li, C.; Zou, J.; and Xiong, H. 2023. Hybrid ISTA: Unfolding ISTA With Convergence Guarantees Using Free-Form Deep Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3): 3226–3244.

Zhou, M.; Huang, J.; Zheng, N.; and Li, C. 2023. Learned Image Reasoning Prior Penetrates Deep Unfolding Network for Panchromatic and Multi-Spectral Image Fusion. In *Proceedings of the IEEE International Conference on Computer Vision*, 12364–12373.