

# OpenViewer: Openness-Aware Multi-View Learning

Shide Du<sup>1,2</sup>, Zihan Fang<sup>1,2</sup>, Yanchao Tan<sup>1,2</sup>, Changwei Wang<sup>3</sup>, Shiping Wang<sup>1,2</sup>, Wenzhong Guo<sup>1,2\*</sup>

<sup>1</sup> College of Computer and Data Science, Fuzhou University, Fuzhou, China

<sup>2</sup> Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou University, Fuzhou, China

<sup>3</sup> Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology, Jinan, China

dushidems@gmail.com, fzihan11@163.com, yctan@fzu.edu.cn, changweiwang@sdas.org, shipingwangphd@163.com, guowenzhong@fzu.edu.cn

## Abstract

Multi-view learning methods leverage multiple data sources to enhance perception by mining correlations across views, typically relying on predefined categories. However, deploying these models in real-world scenarios presents two primary openness challenges. **1) Lack of Interpretability:** The integration mechanisms of multi-view data in existing black-box models remain poorly explained; **2) Insufficient Generalization:** Most models are not adapted to multi-view scenarios involving unknown categories. To address these challenges, we propose OpenViewer, an openness-aware multi-view learning framework with theoretical support. This framework begins with a Pseudo-Unknown Sample Generation Mechanism to efficiently simulate open multi-view environments and previously adapt to potential unknown samples. Subsequently, we introduce an Expression-Enhanced Deep Unfolding Network to intuitively promote interpretability by systematically constructing functional prior-mapping modules and effectively providing a more transparent integration mechanism for multi-view data. Additionally, we establish a Perception-Augmented Open-Set Training Regime to significantly enhance generalization by precisely boosting confidences for known categories and carefully suppressing inappropriate confidences for unknown ones. Experimental results demonstrate that OpenViewer effectively addresses openness challenges while ensuring recognition performance for both known and unknown samples.

**Code** — <https://github.com/dushide/OpenViewer>

**Extended version** — <https://arxiv.org/abs/2412.12596>

## Introduction

Multi-view learning has emerged as a prominent area of artificial intelligence, focusing on leveraging diverse data sources to enhance perception (Tan et al. 2024; Yu et al. 2024b). This learning paradigm processes real-world objects from various extractors or sensors, exploiting correlations across multiple views to enhance performance in applications like computer vision (Ning et al. 2024), natural language processing (Song et al. 2024), large-scale language models (Guo et al. 2023), and more (Pei et al. 2023;

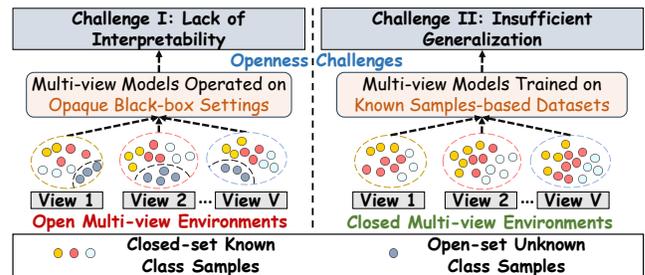


Figure 1: Two multi-view environments and challenges.

Ye and Li 2024). However, traditional multi-view methods, whether heuristic (Zhang et al. 2023; Xiao et al. 2024) or deep learning (Xu et al. 2023; Yang et al. 2024), typically operate under the assumption that all samples belong to known categories (Du et al. 2023). When deployed in real-world settings, these approaches encounter two significant openness challenges, as illustrated in Fig. 1. **Challenge I: Lack of Interpretability.** These black-box methods often lack of explanation in the integration process of multi-view data involving both known and unknown category samples. This opacity undermines their reliability in open scenarios. **Challenge II: Insufficient Generalization.** Trained on known samples-based datasets, existing multi-view closed-set methods fail to identify unknown categories during testing, frequently mislabeling them as known with unduly high confidences. Consequently, they struggle to generalize to multi-view environments containing unknown samples. This issue arises because the models are not preemptively adapted to the range of potential unknown categories.

To effectively address these challenges, we propose OpenViewer, an openness-aware multi-view learning framework designed for real-world environments, as outlined in Fig. 2. OpenViewer starts with a pseudo-unknown sample generation mechanism, allowing the model to efficiently simulate open multi-view environments and previously adapt to potential unknown samples. Grounded on ADMM iterative solutions with functionalized priors, we derive an interpretable multi-view feature expression-enhanced deep unfolding network, comprising redundancy removal, dictionary learning, noise processing, and complementarity fusion modules. The

\*Corresponding author.

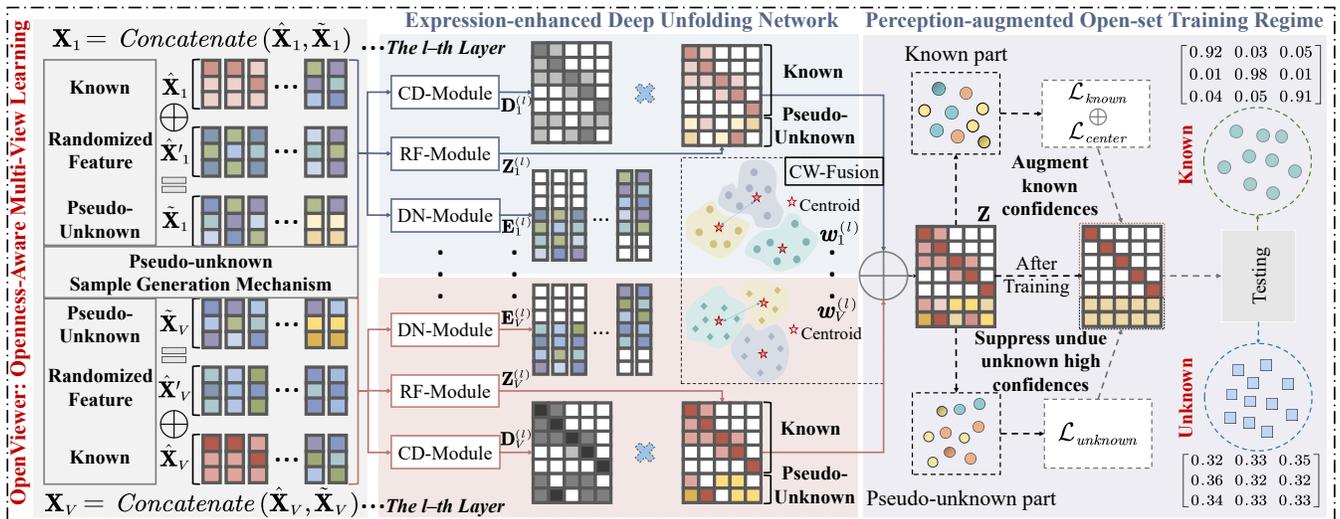


Figure 2: An overview of the proposed openness-aware multi-view learning framework (OpenViewer).

corresponding functions of each module are intuitively reflected in the prior-mapping optimization process, offering a more transparent integration mechanism. Additionally, we implement a multi-view sample perception-augmented open-set training regime to further boost confidences for known categories and suppress inappropriate confidences for unknown ones. This enables the model to dynamically perceive of known and unknown samples, thereby improving generalization. Finally, we present theoretical analysis and proof to substantiate OpenViewer’s ability to increase both interpretability and generalization. The main contributions of OpenViewer can be listed as follows:

- **Formulation of OpenViewer:** We propose OpenViewer, an openness-aware multi-view learning framework designed to tackle the challenges of interpretability and generalization, backed by theoretical guarantees.
- **Openness-aware models design:** We develop an interpretable expression-enhanced deep unfolding network, bolstered by a pseudo-unknown sample generation mechanism and a perception-augmented open-set training regime, to improve adaptation and generalization.
- **Extensive experiments on real-world datasets:** Experimental results validate OpenViewer’s effectiveness in addressing openness challenges, demonstrating superior recognition performance for both known and unknown.

## Related Work

**Two Multi-view Learning Methods.** 1) **Heuristic methods** leverage multi-view prior knowledge to formulate and iteratively solve joint optimization objectives, leading to optimal multi-view learning solutions. For example, Wan *et al.* (Wan *et al.* 2023) proposed an auto-weighted multi-view optimization problem for large-scale data. Yu *et al.* (Yu *et al.* 2024a) devised a non-parametric joint optimization functions to partition multi-view data; 2) **Deep learning methods** utilize network architectures to automate the optimiza-

tion of multi-view learning solutions and parameters. For example, Xiao *et al.* (Xiao *et al.* 2023) performed multi-view deep learning by the consistency and complementarity. Xu *et al.* (Xu *et al.* 2024b) introduced the view-specific encoders and product-of-experts approach to aggregate multi-view information. Further work on multi-view learning can be discovered in (Chen *et al.* 2020; Wang *et al.* 2022; Yang *et al.* 2022; Liu *et al.* 2024) (heuristic) and (Yang *et al.* 2021; Lin *et al.* 2023; Du *et al.* 2024; Wang *et al.* 2024) (deep learning).

**Interpretable Deep Unfolding Networks.** Deep unfolding networks, derived from iterative solutions that encapsulate domain-specific priors and functional knowledge, have achieved success while maintaining strong interpretability across multiple fields (Gregor and LeCun 2010; Bonet *et al.* 2022; Zheng *et al.* 2023; Joukovsky, Eldar, and Deligiannis 2024). Some notable works, for example, Fu *et al.* (Fu *et al.* 2022) designed a model-driven deep unfolding structure for JPEG artifacts removal. Li *et al.* (Li *et al.* 2023) displayed a low-rank deep unfolding network for hyperspectral anomaly detection. Wu *et al.* (Wu *et al.* 2024) constructed a deep unfolding network based on first-order optimization algorithms. Additional similar efforts in deep unfolding networks can be traced in (Zhou *et al.* 2023; Weerdt, Eldar, and Deligiannis 2024; Fang *et al.* 2024b).

**Open-set Learning.** Open-set learning seeks to extend the closed-set hypothesis by equipping models with the ability to distinguish known and unknown classes. For instance, Dhamija *et al.* (Dhamija, Günther, and Boulton 2018) introduced the negative classes for improving the efficiency of unknown rejection. Duan *et al.* (Duan *et al.* 2023) formulated a subgraph-subgraph contrast to open-set graph learning into a multi-scale contrastive network. Safaei *et al.* (Safaei *et al.* 2024) explored an entropic open-set active learning framework to select informative unknown samples. Related open-set learning methods can also be found in (Bendale and Boulton 2016; Du *et al.* 2023; Gou *et al.* 2024).

## Openness-Aware Multi-View Learning

In this section, we introduce the specific architecture of OpenViewer, including pseudo-unknown generation mechanism, expression-enhanced deep network and perception-augmented open-set regime. The dimensions and descriptions of notations are listed in Table 1 of **Extended Version**.

### Pseudo-unknown Sample Generation Mechanism

To better tackle open-set environments, inspired by Mixup (Zhang et al. 2018), we use a pseudo-unknown sample generation mechanism to simulate open multi-view environments and previously adapt to potential unknown samples. Specifically, a perturbation parameter  $\zeta \in [0, 1]$  is sampled from a  $\zeta \sim \text{Beta}(\omega, \omega)$  distribution. Based on this, we randomly select  $\hat{\mathbf{x}}_v^{(i)}$  and  $\hat{\mathbf{x}}_v^{(j)}$  from the  $v$ -th view original feature  $\hat{\mathbf{X}}_v$ , ensuring they belong to different categories. Then, the pseudo-unknown sample  $\tilde{\mathbf{x}}_v$  is generated as

$$\tilde{\mathbf{x}}_v = \zeta \hat{\mathbf{x}}_v^{(i)} + (1 - \zeta) \hat{\mathbf{x}}_v^{(j)}, \quad (1)$$

where  $\zeta$  determines the extent to which each original sample contributes to the features of the generated samples. Using Eq. (1), we generate a set of pseudo-unknown samples,  $\mathcal{D}_{generated}$ , and merge it with  $\mathcal{D}_{original}$  to prepare the model for adapting to unknown classes.

### Expression-enhanced Deep Unfolding Network

Subsequently, we design an interpretable expression-enhanced deep unfolding network to clarify the multi-view integration principle. We first abstract four multi-view functionalized priors as shown in Fig. 3, including: 1) **View-specific Redundancy** denotes the redundant similar features within each view; 2) **View-specific Consistency** indicates the dictionary coefficients, reflecting each representation's consistent contribution to the reconstruction of each view; 3) **View-specific Diversity** signifies the diverse noise information within each view; 4) **Cross-view Complementarity** refers to processed cross-view representations that can complementary, enhance and express each other. Following that, we first consider the three view-specific priors, and construct a generalized expression-enhanced optimization problem as

$$\min_{\mathbf{Z}_v, \mathbf{D}_v, \mathbf{E}_v} \sum_{v=1}^V \left( \mathcal{I}(\mathbf{X}_v, \mathbf{Z}_v, \mathbf{D}_v, \mathbf{E}_v) + \alpha \Omega(\mathbf{Z}_v) + \beta \Psi(\mathbf{D}_v) + \gamma \Phi(\mathbf{E}_v) \right), \quad (2)$$

where  $\alpha, \beta, \gamma$  are the regularization parameters, and generalized Problem (2) includes the above functionalized priors that can be further concretized as

$$\min_{\mathbf{Z}_v, \mathbf{D}_v, \mathbf{E}_v} \sum_{v=1}^V \left( \frac{1}{2} \|\mathbf{X}_v - \mathbf{Z}_v \mathbf{D}_v - \mathbf{E}_v\|_F^2 + \alpha \|\mathbf{Z}_v\|_1 + \frac{\beta}{2} \|\mathbf{D}_v\|_F^2 + \gamma \|\mathbf{E}_v\|_{2,1} \right). \quad (3)$$

Problem (3) aims to learn a *redundancy* free representation  $\mathbf{Z}_v$  using  $l_1$ -norm  $\|\cdot\|_1$ , while optimizing *consistency* dictionary coefficients  $\mathbf{D}_v$  and capturing *diversity* noise  $\mathbf{E}_v$  with the  $l_{2,1}$ -norm  $\|\cdot\|_{2,1}$ . So  $\mathbf{X}_v$  can be expressed as a linear combination  $\mathbf{Z}_v \mathbf{D}_v + \mathbf{E}_v$ . To optimize such a

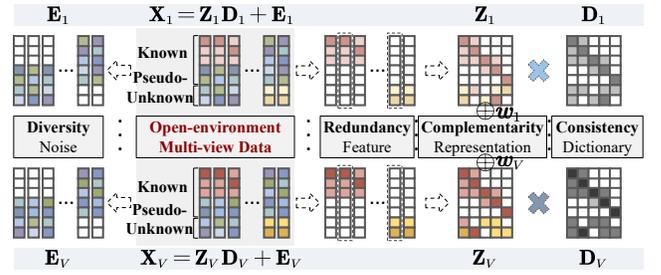


Figure 3: Four multi-view priors and their relationships.

mixed non-convex problem (3) consisting of smooth terms  $\mathcal{I}(\cdot), \Psi(\cdot)$  and non-smooth terms  $\Omega(\cdot), \Phi(\cdot)$ , ADMM (Boyd et al. 2011) is employed to decompose it into three sub-problems for solving. For  $\mathbf{Z}_v = \{\hat{\mathbf{Z}}_v, \tilde{\mathbf{Z}}_v\}$ ,  $\mathbf{E}_v = \{\hat{\mathbf{E}}_v, \tilde{\mathbf{E}}_v\}$ ,  $\mathbf{D}_v = \{\hat{\mathbf{D}}_v, \tilde{\mathbf{D}}_v\}$  sub-problems, where  $\hat{\mathbf{Z}}_v, \hat{\mathbf{E}}_v$ , and  $\hat{\mathbf{D}}_v$  are known, and  $\tilde{\mathbf{Z}}_v, \tilde{\mathbf{E}}_v$ , and  $\tilde{\mathbf{D}}_v$  are pseudo-unknown, we utilize proximal gradient descent method (Beck and Teboulle 2009) to solve  $\mathbf{Z}_v$  and  $\mathbf{E}_v$  variables, while  $\mathbf{D}_v$  variable has a closed-form solution, obtained as

$$\begin{aligned} \mathbf{Z}_v^{(l+1)} &\leftarrow \mathcal{S}_{\frac{\alpha}{L_{p_v}}} \left( \mathbf{Z}_v^{(l)} - \frac{1}{L_{p_v}} \nabla \mathcal{I}(\mathbf{Z}_v^{(l)}) \right), \\ \mathbf{E}_v^{(l+1)} &\leftarrow \mathcal{P}_{\frac{\gamma}{L_{p_v}}} \left( \mathbf{E}_v^{(l)} - \frac{1}{L_{p_v}} \nabla \mathcal{I}(\mathbf{E}_v^{(l)}) \right), \\ \mathbf{D}_v^{(l+1)} &\leftarrow \{ \nabla \mathcal{I}(\mathbf{D}_v^{(l)}) + \nabla \Psi(\mathbf{D}_v^{(l)}) = 0 \}, \end{aligned} \quad (4)$$

where  $\mathcal{S}_{\frac{\alpha}{L_{p_v}}}(\cdot)$  and  $\mathcal{P}_{\frac{\gamma}{L_{p_v}}}(\cdot)$  are the redundancy and diversity proximal operators, respectively.  $\nabla(\cdot)$  denotes the gradient of the current variable,  $L_{p_v}$  is the  $v$ -th Lipschitz constant of  $\nabla \mathcal{I}(\cdot)$ , and  $l$  is the current iteration. Subsequently, we expand the gradient-related notations, detailed as

$$\begin{cases} \mathbf{Z}_v^{(l+1)} &\leftarrow \mathcal{S}_{\frac{\alpha}{L_{p_v}}} \left( \mathbf{Z}_v^{(l)} - \frac{1}{L_{p_v}} (\mathbf{Z}_v^{(l)} \mathbf{D}_v^{(l)} (\mathbf{D}_v^{(l)\top})^{(l)} - \mathbf{X}_v (\mathbf{D}_v^{(l)\top})^{(l)} + \mathbf{E}_v^{(l)} (\mathbf{D}_v^{(l)\top})^{(l)}) \right), \\ &\Rightarrow \mathcal{S}_{\frac{\alpha}{L_{p_v}}} \left( \mathbf{Z}_v^{(l)} (\mathbf{I} - \frac{1}{L_{p_v}} \mathbf{D}_v^{(l)} (\mathbf{D}_v^{(l)\top})^{(l)}) + \frac{1}{L_{p_v}} (\mathbf{X}_v - \mathbf{E}_v^{(l)}) (\mathbf{D}_v^{(l)\top})^{(l)} \right), \\ \mathbf{D}_v^{(l+1)} &\leftarrow \left( (\mathbf{Z}_v^{(l+1)\top}) \mathbf{Z}_v^{(l+1)} + \beta \mathbf{I} \right)^{-1} (\mathbf{Z}_v^{(l+1)\top})^{(l+1)} (\mathbf{X}_v - \mathbf{E}_v^{(l)}), \\ \mathbf{E}_v^{(l+1)} &\leftarrow \mathcal{P}_{\frac{\gamma}{L_{p_v}}} \left( \mathbf{X}_v - \mathbf{Z}_v^{(l+1)} \mathbf{D}_v^{(l+1)} \right), \end{cases} \quad (5)$$

where  $\mathbf{I} \in \mathbb{R}^{C \times C}$  is an identity matrix. Thus far, we have used the ADMM optimizer to solve the corresponding sub-problems and derive iterative solutions for *redundancy*, *consistency*, *diversity* priors. Then, an inter-class discretion-guided weighting method is applied to account for the cross-view *complementary* prior. Intuitively, the closer the sample centroids are within a view, the less complementary information each view provides. Base on this, we dynamically perceive the centroid as  $\mathbf{o}_v^{(i)} = \frac{1}{|B_i|} \sum_{j: \mathbf{Y}_j=i} \mathbf{z}_v^{(j)}$  and then calculate the inter-class distances between all cen-

troids for each view, where  $|B_i|$  is the number of training instances in category  $i$ . To prevent the largest distance between two classes from overly influencing the measure of inter-class discrepancy, we focus on the minimum distance  $\mathbf{d}_v$  between any two categories, defined as  $\mathbf{d}_v = \min \left\{ \text{Dist} \left( \mathbf{o}_v^{(i)}, \mathbf{o}_v^{(j)} \right) \right\}, i, j \in \mathbf{Y}$  and  $i \neq j$ . Here,  $\text{Dist}(\cdot, \cdot)$  is the distance function, which in this work is the Euclidean distance. This strategy provides a balanced assessment of the complementary contribution of different views. The higher the complementarity information between views (*i.e.*, the greater the distance between centroids), the greater their assigned weights, denoted as

$$\mathbf{w}_v = \frac{\exp(-\bar{\mathbf{d}}_v)}{\sum_{v=1}^V \exp(-\bar{\mathbf{d}}_v)}, \bar{\mathbf{d}}_v = \mathbf{d}_v^{-1} / \|\mathbf{d}_v^{-1}\|_1, \sum_{v=1}^V \mathbf{w}_v = 1, \quad (6)$$

where  $\bar{\mathbf{d}}_v$  is obtained by normalizing the inverse of  $\mathbf{d}_v$  between centroids through  $\ell_1$ -norm. At last, the inter-class discretion-guided weights  $\{\mathbf{w}_v\}_{v=1}^V$  can be applied to perform complementary fusion  $\mathcal{F}(\cdot)$  as

$$\mathbf{Z}^{(l+1)} \leftarrow \sum_{v=1}^V \mathbf{w}_v^{(l+1)} \mathbf{Z}_v^{(l+1)}. \quad (7)$$

Based on solutions (4) and (7), the multi-view feature expression-enhanced deep unfolding network can be conceptualized as four interpretable prior-mapping modules by parameterizing alternative components (Zhou et al. 2023; Weerdt, Eldar, and Deligiannis 2024) as

$$\begin{cases} \text{RF-Module: } \mathbf{Z}_v^{(l+1)} & \leftarrow \mathcal{S}_{\theta_v^{(l)}} \left( \mathbf{Z}_v^{(l)} \mathbf{R} + (\mathbf{X}_v - \mathbf{E}_v^{(l)}) \right. \\ & \left. (\mathbf{D}_v^\top)^{(l)} \mathbf{U} \right), \\ \text{CD-Module: } \mathbf{D}_v^{(l+1)} & \leftarrow \mathbf{M} (\mathbf{Z}_v^\top)^{(l+1)} (\mathbf{X}_v - \mathbf{E}_v^{(l)}), \\ \text{DN-Module: } \mathbf{E}_v^{(l+1)} & \leftarrow \mathcal{P}_{\rho_v^{(l)}} \left( \mathbf{X}_v - \mathbf{Z}_v^{(l+1)} \mathbf{D}_v^{(l+1)} \right), \\ \text{CW-Fusion: } \mathbf{Z}^{(l+1)} & \leftarrow \sum_{v=1}^V \mathbf{w}_v^{(l+1)} \mathbf{Z}_v^{(l+1)}, \end{cases} \quad (8)$$

where  $\mathbf{R} = \mathbf{I} - \frac{1}{L_{p_v}} \mathbf{D}_v \mathbf{D}_v^\top$ ,  $\mathbf{U} = \frac{1}{L_{p_v}} \mathbf{I}$ , and  $\mathbf{M} = (\mathbf{Z}_v^\top \mathbf{Z}_v + \beta \mathbf{I})^{-1}$ . The learnable redundancy and diversity proximal operators  $\mathcal{S}_{\theta_v^{(l)}}(\cdot)$  and  $\mathcal{P}_{\rho_v^{(l)}}(\cdot)$  are the reparameterized versions of  $\mathcal{S}_{\frac{\alpha}{L_{p_v}}}(\cdot)$  and  $\mathcal{P}_{\frac{\gamma}{L_{p_v}}}(\cdot)$ , with learnable threshold parameters  $\theta_v$  and  $\rho_v$ , respectively. Moreover,  $\mathcal{S}_{\theta_v}(\mathbf{a}^{(ij)}) = \sigma(\mathbf{a}^{(ij)} - \theta_v) - \sigma(-\mathbf{a}^{(ij)} - \theta_v)$ , and  $\mathcal{P}_{\rho_v}(\mathbf{a}^{(i)}) = \frac{\sigma(\|\mathbf{a}^{(i)}\|_2 - \rho_v)}{\|\mathbf{a}^{(i)}\|_2} \mathbf{a}^{(i)}$ , if  $\rho_v < \|\mathbf{a}^{(i)}\|_2$ ; otherwise, 0.  $\mathbf{a}^{(ij)}$  is the element in the  $i$ -th row and  $j$ -th column of the matrix,  $\mathbf{a}^{(i)}$  is the  $i$ -th column of the matrix, and  $\sigma(\cdot)$  can be activation functions such as ReLU, SeLU and etc.

The constructed network, incorporating these modules, can engage in multi-view expression enhancement while integrating their functions into deep networks to maintain interpretability: 1) **Redundancy Free Representation Module (RF-Module)** introduces learnable layers and redundancy-free operators to reduce redundant features and retain the most critical view information  $\mathbf{Z}_v$ ; 2) **Consistency**

**Dictionary Learning Module (CD-Module)** captures dictionary coefficients  $\mathbf{D}_v$  within each view, denoting the consistent contribution of each  $\mathbf{Z}_v$  to the reconstruction of  $\mathbf{X}_v$ ; 3) **Diversity Noise Processing Module (DN-Module)** develops learnable diversity operators to eliminate irrelevant information  $\mathbf{E}_v$  caused by the noise or outliers; 4) **Complementarity Fusion Representation Module (CW-Fusion)** implements complementary weight fusion to integrate representations as  $\mathbf{Z}$  to differentiate between known and unknown. Unfolding network (8) is composed of  $L$  layers, with each layer corresponding to a single ADMM iteration. The interpretability is reflected in the optimization process: 1) For multi-view known parts, it enhances expression by processing noise and integrating complementary; 2) For multi-view unknown parts, it employs redundancy removal, noise processing, and adapts to a pseudo-unknown dictionary to highlight inappropriate unknown confidences. This enhanced expression provides a solid foundation for distinguishing between known and unknown, thereby boosting OpenViewer’s interpretability and trustworthiness.

### Perception-augmented Open-set Training Regime

The above interpretable network has performed feature-level integration and enhancement. Subsequently, we design a loss regime to further augment sample-level perception and improve the model’s generalization. For known samples, we first ensure the model’s ability to recognize them by applying a cross-entropy loss. Building on this, we promote the separation of all known classes by a distance margin term  $\max(\boldsymbol{\xi} - \|\hat{\mathbf{z}}^{(i)}\|_2, 0)^2$ , formalized as

$$\begin{aligned} \mathcal{L}_{known} &= -\frac{1}{N^o} \sum_{i=1}^{N^o} \sum_{c=1}^C \left( \hat{\mathbf{y}}_c^{(i)} \log P(c | \hat{\mathbf{z}}^{(i)}) \right) \\ &+ \sum_{i=1}^{N^o} \max(\boldsymbol{\xi} - \|\hat{\mathbf{z}}^{(i)}\|_2, 0)^2, \end{aligned} \quad (9)$$

where  $\hat{\mathbf{z}}^{(i)} \in \hat{\mathbf{Z}}$ ,  $P$  is the Softmax score, and  $\boldsymbol{\xi}$  is the distance margin. In this way, the feature vector is pushed out of the margin  $\boldsymbol{\xi}$  to make its norm as close to or greater than  $\boldsymbol{\xi}$  as possible, thereby augmenting the discrimination between known class samples. For the more critical unknown part, we aim to minimize the allocation of pseudo-unknown samples to known groups. Therefore, we employ a  $\ell_2$ -norm regularization term  $\|\tilde{\mathbf{z}}^{(i)}\|_2^2$  to ensure that OpenViewer suppresses excessive unknown high confidences, expressed as

$$\mathcal{L}_{unknown} = -\frac{1}{C} \sum_{i=1}^{N^e} \sum_{c=1}^C \left( \log P(c | \tilde{\mathbf{z}}^{(i)}) \right) + \sum_{i=1}^{N^e} \|\tilde{\mathbf{z}}^{(i)}\|_2^2, \quad (10)$$

where  $\tilde{\mathbf{z}}^{(i)} \in \tilde{\mathbf{Z}}$ , and  $P(c | \tilde{\mathbf{z}}^{(i)})$  is the probability that the model predicts the pseudo-unknown sample  $\tilde{\mathbf{z}}^{(i)}$  as belonging to category  $c$ . Loss (10) ensures that the model’s prediction confidence for each known class is average and penalizes pseudo-unknown that are close to known, thereby suppressing inappropriate confidences for unknown samples.

However, the above loss only increases the inter-class separability between known and unknown samples. To promote

intra-class compactness, we use the following center loss to further separate the feature vectors of different classes as

$$\mathcal{L}_{center} = \frac{1}{2} \sum_{i=1}^{N^o} \left\| \hat{\mathbf{z}}_c^{(i)} - \mathbf{c}^{\hat{\mathbf{y}}_c^{(i)}} \right\|_2^2, \quad (11)$$

where  $\mathbf{c}^{\hat{\mathbf{y}}_c^{(i)}}$  is the center vector corresponding to the  $i$ -th sample’s true label. Then, each category center is dynamically updated during training to better reflect the sample distribution of its corresponding category, described as

$$\Delta \mathbf{c}^{(j) \in C} = \frac{\sum_{i=1}^{N^o} \delta(\hat{\mathbf{y}}_c^{(i)} = j) \cdot (\mathbf{c}^{(j)} - \hat{\mathbf{z}}_c^{(i)})}{1 + \sum_{i=1}^{N^o} \delta(\hat{\mathbf{y}}_c^{(i)} = j)}, \quad (12)$$

where  $\Delta \mathbf{c}^{(j) \in C}$  the update amount for the center of category  $j$ ,  $\delta(\cdot)$  is an indicator function that takes the value 1 when the  $i$ -th sample belongs to category  $j$ ; otherwise, 0. Meanwhile, the center vectors of each class are adjusted by the calculated update amounts as  $\hat{\mathbf{c}}_{new}^{(i)} \leftarrow \hat{\mathbf{c}}_{old}^{(i)} - \Delta \mathbf{c}^{(j)}$ , where  $\hat{\mathbf{c}}_{new}^{(i)}$  is the updated center vectors, while  $\hat{\mathbf{c}}_{old}^{(i)}$  is the old center vectors.

At last, we train the unfolding network (8) by combining these losses to augment perception as

$$\mathcal{L}_{total} = \mathcal{L}_{known} + \lambda_1 \mathcal{L}_{unknown} + \lambda_2 \mathcal{L}_{center}, \quad (13)$$

where  $\lambda_1$  and  $\lambda_2$  are two trade-off parameters. The contribution of training regime (13) to OpenViewer’s generalization is twofold: 1) From a feature correspondence perspective, it ensures that known parts elicit a strong response, while undue confidences of pseudo-unknown are suppressed to a low response; 2) From an entropy perspective, it reduces the entropy of known to augment discrimination, while increasing the entropy of pseudo-unknown to ensure that they have low confidence in being classified as known, thereby further reinforcing the recognition of unknown. OpenViewer can be summarized as Algorithm 1 in **Appendix**.

## Main Theoretical Presentation and Analysis

**Theorem 1. (Interpretability Boundary)** *If each submodule is convergent, then the stacked deep unfolding network consisting of all modules is bounded.*

**Remark 1.** Supported by **Theorem 1**, the interpretable deep unfolding network (8) will be bounded regardless of the initial multi-view cases with known and pseudo-unknown, indicating that information from different views can be reasonably integrated and **interpreted** in mixed scenarios, thereby improving the trustworthiness.

**Theorem 2. (Generalization Support)** *For the fixed step size (i.e.,  $\eta_t = \eta$ ) as  $T \rightarrow \infty$ , and given the existing upper boundary  $\epsilon$ , the difference  $\mathcal{L}_T^* - \mathcal{L}^*$  generalizes to  $\frac{\eta \epsilon^2}{2}$  with a convergence rate  $\mathcal{O}(1/T)$ .*

**Remark 2.** **Theorem 2** theoretically ensures that OpenViewer maintains stable **generalization** by learning a true distribution within a convergence radius of  $\frac{\eta \epsilon^2}{2}$  and a convergence rate  $\mathcal{O}(1/T)$ , even when encountering unknown.

Datasets	# Samples	# Views	# Feature Dimensions	# Classes
Animals	10,158	2	4,096/4,096	50
AWA	30,475	6	2,688/2,000/252/ 2,000/2,000/2,000	50
NUSWIDEOBJ	30,000	5	65/226/145/74/129	31
VGGFace2-50	34,027	4	944/576/512/640	50
ESP-Game	11,032	2	100/100	7
NUSWIDE20k	20,000	2	100/100	8

Table 1: A brief description of the tested datasets.

**Complexity.** The time complexity of OpenViewer with  $L$  layers costs  $\mathcal{O}((NC(C + D_v + V) + C^2 D_v) L)$ , and the space complexity of OpenViewer denotes  $\mathcal{O}(N(D_v + C)V)$ . Additional proofs and details of the main theories and complexity can be found in **Appendix**.

## Experiments and Studies

### Datasets, Compared Methods, and Evaluation Metric.

We conduct experiments in challenging open-environment classification tasks under six well-known multi-view datasets. This includes two scenarios: 1) Animals, AWA, NUSWIDEOBJ, and VGGFace2-50 datasets contain different manual and deep features; 2) ESP-Game and NUSWIDE20k datasets include various vision and language features. The statistics of these datasets are summarized in Table 1 (details in **Appendix**). Moreover, to simulate the performance of OpenViewer in open-environment, we also utilize the concept of *openness* (Scheirer et al. 2012) to divide known and unknown categories of multi-view datasets. Meanwhile, the dataset is partitioned as follows: 10% of the known class samples are allocated for training, another 10% for validation, and the rest 80% for testing.

Due to the limited exploration of related open multi-view learning tasks, we drew on backbone networks from other different multi-view tasks as compared methods (details in **Appendix**), including: MvNNcor (Xu et al. 2020), TMC (Han et al. 2021), MMDynamics (Han et al. 2022), IMvGCN (Wu et al. 2023), LGCN-FF (Chen et al. 2023), ORLNet (Fang et al. 2024a), and RCML (Xu et al. 2024a).

To estimate recognition performance effectively, the Open-Set Classification Rate (OSCR) (Dhamija, Günther, and Boulton 2018) is adopted as metrics, consisting of Correct Classification Rate (CCR) and False Positive Rate (FPR).

**Experimental Setups.** OpenViewer is implemented using the PyTorch on an NVIDIA GeForce RTX 4080 GPU with 16GB of memory. We train OpenViewer for 100 epochs with a batch size of 50, a learning rate of 0.01,  $\xi = 5$ , and  $\lambda_1$  and  $\lambda_2$  selected from  $\{10^{-3}, 5 \times 10^{-3}, \dots, 10^0\}$ . The number of unfolding layers is set to  $L = 1$  as suggested in **Appendix** Fig. 3, balancing complexity and efficiency while preserving interpretable expression-enhanced capabilities. The ablation-models (**Appendix** Table 1) are OpenSViewer (w/o CD-Module and DN-Module) and OpenSDViewer (w/o DN-Module) using for self-verification.

**Experimental Results.** We present the overall OSCR curve results of all multi-view learning methods in classification under the condition of *openness* = 0.1, as shown in

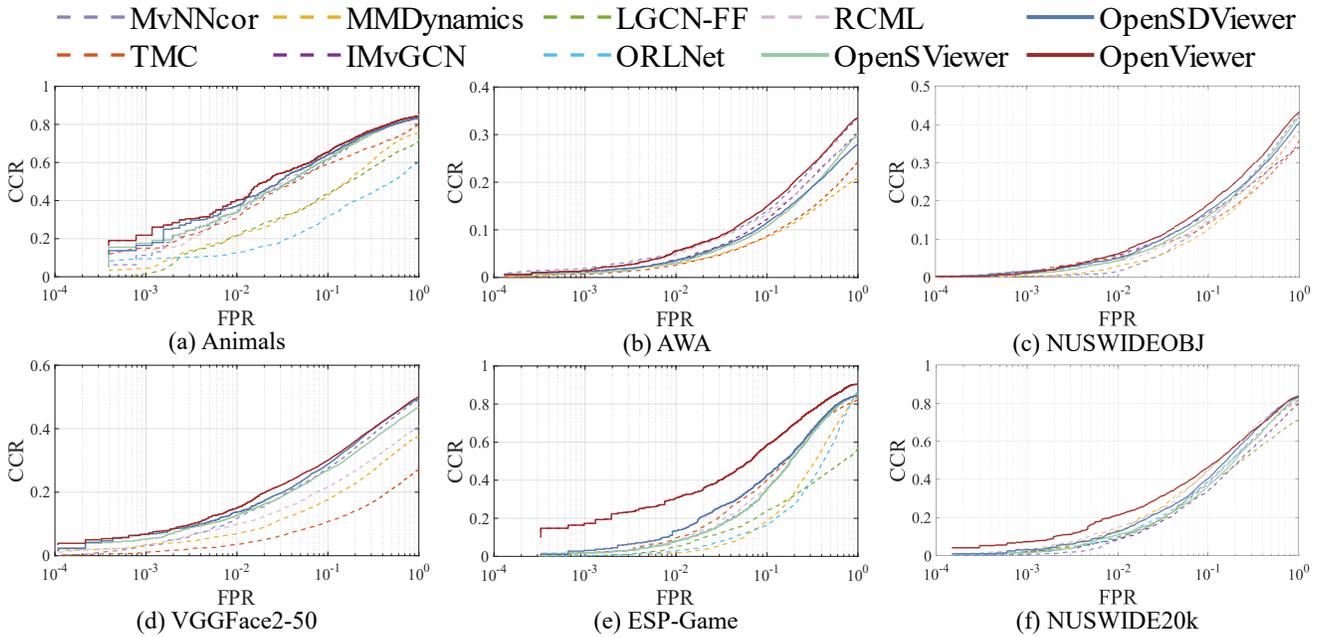


Figure 4: OSCR curves plotting the CCR over FPR on all test multi-view datasets for all compared methods.

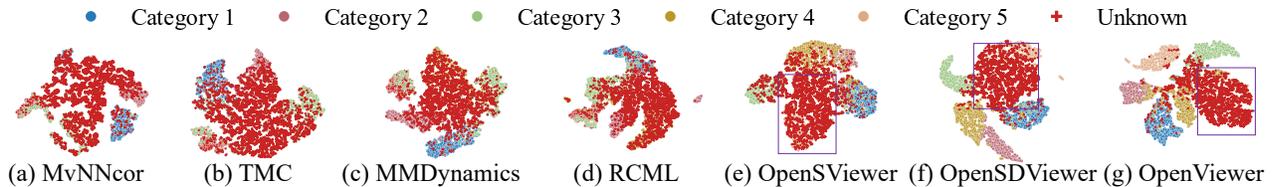


Figure 5: The t-SNE visualizations based on the fused representations of ESP-Game dataset.

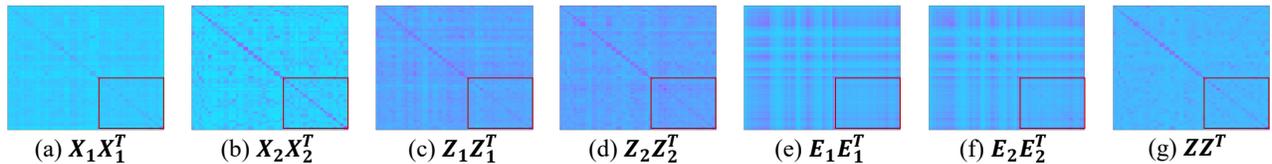


Figure 6: The heatmaps on feature, redundancy-free, noise and fusion matrices with  $w_1 = 0.5132$  and  $w_2 = 0.4868$  of Animals.

Fig. 4, from which we observe: Intuitively, OpenViewer (red solid line) outperforms other methods (colored dashed lines) across all cases, whether on multi-feature or multi-modal datasets. Although some multi-view methods, such as TMC and RCML, occasionally outperform OpenViewer in specific cases, it demonstrates more stable performance across all scenarios. Finally, OpenViewer’s effective balance across different FPR and CCR results in outstanding performance. On one hand, this may be attributed to the effective enhancement of both known and unknown expression through interpretable integration, such as the multi-feature expression enhancement seen with VGGFace2-50. On the other hand, the pseudo-unknown mechanism and perception-augmented loss contribute to significant confidence differentiation, as depicted in Fig. 6 (g) for Animals. Moreover, Fig. 5 (all in

Appendix Fig. 1) indicates that OpenViewer achieves the highest separation between different categories and minimal overlap between unknown and known categories. Additionally, Fig. 6 clearly demonstrates the impact of functionalized priors in OpenViewer. Original features exhibit complementary information but also contain redundancy, noise, and inappropriate confidences. Whereas OpenViewer effectively filters relevant features and noise, and suppresses inappropriate confidences for unknown (the red box). Ultimately, the complementary fusion effectively enhances expression with a clean diagonalized structure, revealing high response for known and low response for unknown.

**Ablation Study.** To verify that each module and loss term contributes to address openness challenges, we conduct ablation experiments. First, when we comprehensively exam-

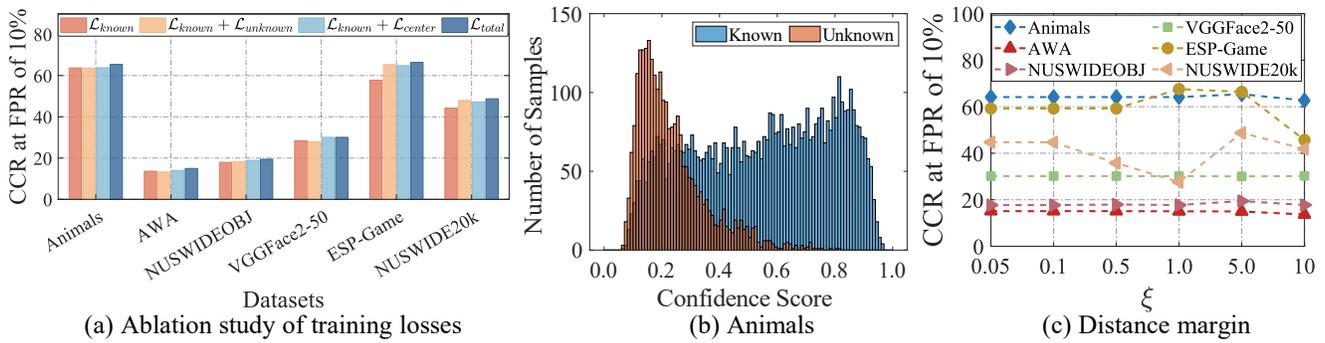


Figure 7: Ablation study of training losses, confidence score, and parameter sensitivity of  $\xi$ .

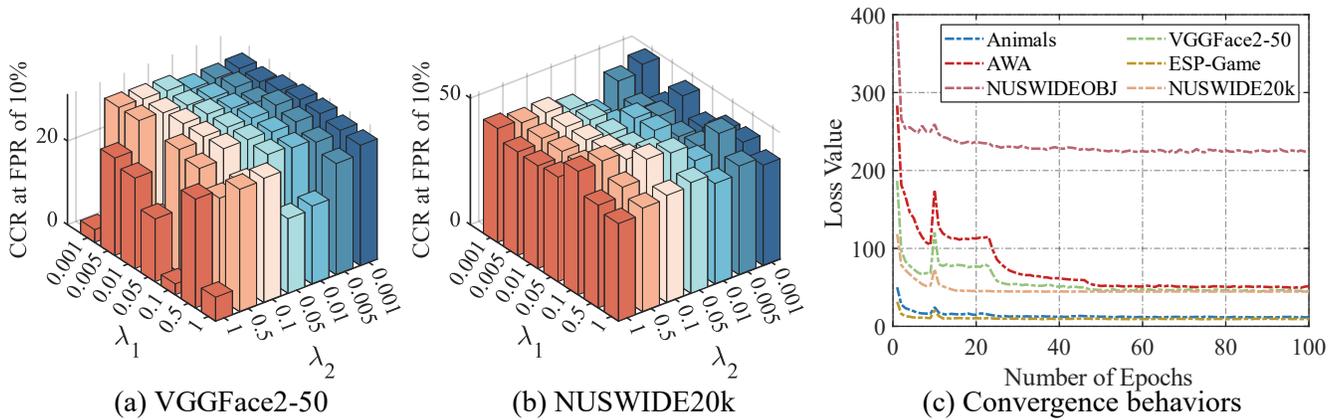


Figure 8: Parameter sensitivity of  $\lambda_1$ ,  $\lambda_2$ , and loss behaviors.

ine the variants OpenViewer, OpenSDViewer, and OpenViewer portrayed in Fig. 4 and Fig. 5 (Appendix Fig. 1), we can excitedly discover that the addition of each interpretable module promotes performance improvement and effective separation of samples. For example, after adding the diversity noise processing module (DN-Module), the performance of ESP-Game improved from 42.56% to 58.65%, with increased inter-class separability between known and unknown. This improvement can be attributed to the removal of multi-view noise, which enhances feature expression. Furthermore, Fig. 7 (a) highlights the generalized contributions of all loss components. From this, their combination can promote the model’s generalization by distinguishing between normal known and undue unknown confidences. Fig. 7 (b) depicts that training regime accentuates differences in confidence scores between known and unknown distributions of Animals, aiding in recognition. Specifically, the unknown loss enhances sample discrimination, while the center loss promotes intra-class compactness.

**Parameter Sensitivity Analysis.** First, Fig. 7 (c) reveals when the parameter  $\xi$  is set to 5, the feature range is adequate to effectively distinguish. However, when this value is exceeded, the wider feature range causes overlap among known classes, leading to a decline in overall performance. Second, Fig. 8 (a)-(b) (all in Appendix Fig. 4) illustrates

the parameter sensitivity of OpenViewer on two representative datasets in terms of  $\lambda_1$  and  $\lambda_2$  of loss (13). The model performance is generally robust in most cases, but it collapses when  $\lambda_2$  becomes too large, causing all samples to provide meaningless confidences across all classes, with no clear winning class. Finally, we showcase Fig. 8 (c) to elucidate the behaviors between loss values and training epochs. The curve displays that after 100 training epochs, the loss value stabilizes, indicating convergence and underscoring its stability, as depicted in **Theoretical Analysis**.

## Conclusion and Future Work

In this paper, we proposed OpenViewer to address the openness challenges in open-settings. OpenViewer began with a pseudo-unknown sample generation mechanism to previously adapt to unknown, followed by a multi-view expression-enhanced deep unfolding network to offer a more interpretable integration mechanism. Additionally, OpenViewer employed a perception-augmented open-set training regime to improve generalization between known and unknown classes. Extensive experiments on diverse multi-view datasets showed that OpenViewer outperformed existing methods in recognition while effectively tackling openness challenges. In future work, we will explore more sophisticated openness-aware circumstances based multi-view learning, including heterogeneous or incomplete data.

## Acknowledgments

This work is in part supported by the National Natural Science Foundation of China (Grant Nos. U21A20472 and 62276065), and the National Key Research and Development Plan of China (Grant No. 2021YFB3600503).

## 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.
- Bendale, A.; and Boulton, T. E. 2016. Towards open set deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1563–1572.
- Bonet, E. R.; Do, T. H.; Qin, X.; Hofman, J.; Manna, V. P. L.; Philips, W.; and Deligiannis, N. 2022. Explaining graph neural networks with topology-aware node selection: Application in air quality inference. *IEEE Transactions on Signal and Information Processing over Networks*, 8: 499–513.
- Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J.; et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1): 1–122.
- Chen, M.; Huang, L.; Wang, C.-D.; and Huang, D. 2020. Multi-view clustering in latent embedding space. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, 3513–3520.
- Chen, Z.; Fu, L.; Yao, J.; Guo, W.; Plant, C.; and Wang, S. 2023. Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, 95: 109–119.
- Dhamija, A. R.; Günther, M.; and Boulton, T. 2018. Reducing network agnostophobia. In *Advances in Neural Information Processing Systems*, 1–12.
- 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 Thirty-First ACM International Conference on Multimedia*, 8719–8729.
- Duan, J.; Wang, S.; Zhang, P.; Zhu, E.; Hu, J.; Jin, H.; Liu, Y.; and Dong, Z. 2023. Graph anomaly detection via multi-scale contrastive learning networks with augmented view. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, 7459–7467.
- 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 Thirty-Second ACM International Conference 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.
- Gou, Y.; Zhao, H.; Li, B.; Xiao, X.; and Peng, X. 2024. Test-Time degradation adaptation for open-set image restoration. In *Proceedings of the Forty-First International Conference on Machine Learning*, 1–11.
- Gregor, K.; and LeCun, Y. 2010. Learning fast approximations of sparse coding. In *Proceedings of the Twenty-Seventh International Conference on Machine Learning*, 399–406.
- Guo, Z.; Tang, Y.; Zhang, R.; Wang, D.; Wang, Z.; Zhao, B.; and Li, X. 2023. ViewRefer: Grasp the multi-view knowledge for 3D visual grounding. In *Proceedings of the IEEE International Conference on Computer Vision*, 15326–15337.
- Han, Z.; Yang, F.; Huang, J.; Zhang, C.; and Yao, J. 2022. Multimodal dynamics: Dynamical fusion for trustworthy multimodal classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 20707–20717.
- Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2021. Trusted multi-view classification. In *Proceedings of the Ninth International Conference on Learning Representations*, 1–11.
- 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.
- Li, C.; Zhang, B.; Hong, D.; Yao, J.; and Chanussot, J. 2023. LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–12.
- Lin, Y.; Gou, Y.; Liu, X.; Bai, J.; Lv, J.; and Peng, X. 2023. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4447–4461.
- Liu, S.; Zhang, J.; Wen, Y.; Yang, X.; Wang, S.; Zhang, Y.; Zhu, E.; Tang, C.; Zhao, L.; and Liu, X. 2024. Sample-level cross-view similarity learning for incomplete multi-view clustering. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, 14017–14025.
- Ning, X.; Yu, Z.; Li, L.; Li, W.; and Tiwari, P. 2024. DILF: Differentiable rendering-based multi-view image-language fusion for zero-shot 3D shape understanding. *Information Fusion*, 102: 102033.
- Pei, S.; Kou, Z.; Zhang, Q.; and Zhang, X. 2023. Few-shot low-resource knowledge graph completion with multi-view task representation generation. In *Proceedings of the Twenty-Ninth ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1862–1871.
- Safaei, B.; VS, V.; de Melo, C. M.; and Patel, V. M. 2024. Entropic open-set active learning. In *Proceedings of the Thirty-Eighth Thirty-Eighth AAAI Conference on Artificial Intelligence*, 4686–4694.

- Scheirer, W. J.; de Rezende Rocha, A.; Sapkota, A.; and Boulton, T. E. 2012. Toward open-set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7): 1757–1772.
- Song, S.; Zhao, S.; Wang, C.; Yan, T.; Li, S.; Mao, X.; and Wang, M. 2024. A dual-way enhanced framework from text matching point of view for multimodal entity linking. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, 19008–19016.
- Tan, Y.; Cai, H.; Huang, S.; Wei, S.; Yang, F.; and Lv, J. 2024. An effective augmented Lagrangian method for fine-grained multi-view optimization. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, 15258–15266.
- Wan, X.; Liu, X.; Liu, J.; Wang, S.; Wen, Y.; Liang, W.; Zhu, E.; Liu, Z.; and Zhou, L. 2023. Auto-weighted multi-view clustering for large-scale data. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, 10078–10086.
- Wang, S.; Liu, X.; Liu, S.; Jin, J.; Tu, W.; Zhu, X.; and Zhu, E. 2022. Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences. In *Advances in Neural Information Processing Systems*, 5882–5895.
- Wang, S.; Liu, X.; Liu, S.; Tu, W.; and Zhu, E. 2024. Scalable and structural multi-view graph clustering with adaptive anchor fusion. *IEEE Transactions on Image Processing*, 33: 4627–4639.
- Weerd, B. D.; Eldar, Y. C.; and Deligiannis, N. 2024. Deep unfolding transformers for sparse recovery of video. *IEEE Transactions on Signal Processing*, 72: 1782–1796.
- Wu, Z.; Lin, X.; Lin, Z.; Chen, Z.; Bai, Y.; and Wang, S. 2023. Interpretable graph convolutional network for multi-view semi-supervised learning. *IEEE Transactions on Multimedia*, 25: 8593–8606.
- Wu, Z.; Xiao, M.; Fang, C.; and Lin, Z. 2024. Designing universally-approximating deep neural networks: A first-order optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(9): 6231–6246.
- Xiao, S.; Du, S.; Chen, Z.; Zhang, Y.; and Wang, S. 2023. Dual fusion-propagation graph neural network for multi-view clustering. *IEEE Transactions on Multimedia*, 25: 9203–9215.
- Xiao, Y.; Zhang, J.; Liu, B.; Zhao, L.; Kong, X.; and Hao, Z. 2024. Multi-view maximum margin clustering with privileged information learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4): 2719–2733.
- Xu, C.; Si, J.; Guan, Z.; Zhao, W.; Wu, Y.; and Gao, X. 2024a. Reliable conflictive multi-view learning. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, 16129–16137.
- Xu, G.; Wen, J.; Liu, C.; Hu, B.; Liu, Y.; Fei, L.; and Wang, W. 2024b. Deep variational incomplete multi-view clustering: Exploring shared clustering structures. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, 16147–16155.
- Xu, J.; Chen, S.; Ren, Y.; Shi, X.; Shen, H.; Niu, G.; and Zhu, X. 2023. Self-weighted contrastive learning among multiple views for mitigating representation degeneration. In *Advances in Neural Information Processing Systems*, 1119–1131.
- Xu, J.; Li, W.; Liu, X.; Zhang, D.; Liu, J.; and Han, J. 2020. Deep embedded complementary and interactive information for multi-view classification. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, 6494–6501.
- Yang, C.; Wen, H.; Hooi, B.; and Zhou, L. 2024. CapMax: A framework for dynamic network representation learning from the view of multiuser communication. *IEEE Transactions on Neural Networks and Learning Systems*, 35(4): 4554–4566.
- Yang, M.; Li, Y.; Hu, P.; Bai, J.; Lv, J. C.; and Peng, X. 2022. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1): 1055–1069.
- Yang, M.; Li, Y.; Huang, Z.; Liu, Z.; Hu, P.; and Peng, X. 2021. Partially view-aligned representation learning with noise-robust contrastive loss. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1134–1143.
- Ye, F.; and Li, S. 2024. MileCut: A multi-view truncation framework for legal case retrieval. In *Proceedings of the Thirty-Third ACM on Web Conference*, 1341–1349.
- Yu, S.; Wang, S.; Dong, Z.; Tu, W.; Liu, S.; Lv, Z.; Li, P.; Wang, M.; and Zhu, E. 2024a. A non-parametric graph clustering framework for multi-view data. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, 16558–16567.
- Yu, S.; Wang, S.; Zhang, P.; Wang, M.; Wang, Z.; Liu, Z.; Fang, L.; Zhu, E.; and Liu, X. 2024b. DVSAI: Diverse view-shared anchors based incomplete multi-view clustering. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, 16568–16577.
- Zhang, H.; Cissé, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2018. Mixup: Beyond empirical risk minimization. In *Proceedings of the Sixth International Conference on Learning Representations*, 1–13.
- Zhang, P.; Wang, S.; Li, L.; Zhang, C.; Liu, X.; Zhu, E.; Liu, Z.; Zhou, L.; and Luo, L. 2023. Let the data choose: Flexible and diverse anchor graph fusion for scalable multi-view clustering. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, 11262–11269.
- 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.