Dynamic Multiple High-order Correlations Fusion with Noise Filtering for Incomplete Multi-view Noisy-label Learning

Kaixiang Wang*, Xiaojian Ding, Fan Yang
Nanjing University of Finance and Economics
kxwang@nufe.edu.cn, wjswsl@163.com, nufe_yf@163.com

Abstract

Multi-view multi-label data often suffers from incomplete feature views and label noise. This paper is the first to address both challenges simultaneously, rectifying critical deficiencies in existing methodologies that inadequately extract and fuse high-order structural correlations across views while lacking robust solutions to mitigate label noise. We introduce a dynamic multiple high-order correlations fusion with noise filtering, specifically designed for incomplete multi-view noisy-label learning. By capitalizing on a dynamic multi-hypergraph neural network, inspired by the principles of ensemble learning, we adeptly capture and integrate high-order correlations among samples from different views. The model's capability is further augmented through an innovative hypergraph fusion technique based on random walk theory, which empowers it to seamlessly amalgamate both structural and feature information. Moreover, we propose sophisticated noise-filtering matrices that are tightly embedded within the hypergraph neural network, devised to counteract the detrimental impact of label noise. Recognizing that label noise perturbs the data distribution in the label space, these filtering matrices exploit the distributional disparities between feature and label spaces. The high-order structural information derived from both domains underpins the learning and efficacy of the noise-filtering matrices. Empirical evaluations on benchmark datasets unequivocally demonstrate that our method significantly outperforms contemporary state-of-the-art techniques.

1 Introduction

The prevalence of multi-view and multi-label challenges has become increasingly prominent with the rise of diverse data modalities. For instance, users on social networking platforms can be characterized through modalities such as images, text, and audio, and may be associated with multiple labels, such as 'Symphony', 'Beethoven', 'Piano', and 'Classical school'. Accurately classifying such multi-view, multi-label data is crucial for comprehensive data analysis [Qian *et al.*, 2014; Zhu *et al.*, 2015].

The investigation of multi-view multi-label learning is more complex than addressing either multi-view [Wang *et al.*, 2022] or multi-label [Liu *et al.*, 2021] data independently, as it requires the simultaneous consideration of both the correlations among different data views and the intricate dependencies across multiple labels. In light of this complexity, researchers have increasingly shifted their attention to incomplete multi-view multi-label learning, recognizing that real-world applications frequently encounter incomplete data in both its feature representations and label annotations [Tan *et al.*, 2018; Li and Chen, 2021; Li *et al.*, 2024; Ou *et al.*, 2024; Liu *et al.*, 2023; Liu *et al.*, 2024]. Consequently, this learning framework has emerged as particularly well-suited for addressing practical, real-world challenges.

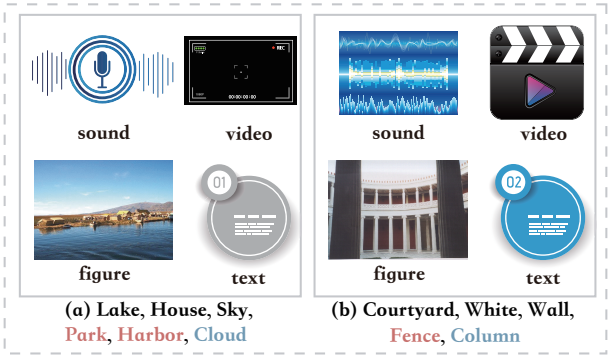

Figure 1. Multi-view multi-label datasets with missing and noisy labels: (a) Blurred objects are randomly mislabelled; (b) Similar objects are mislabelled with each other.

Despite significant progress in multi-view multi-label learning, a crucial issue is often overlooked: label noise. In today's data-driven world, errors in data collection, storage, and annotation are common, particularly with crowdsourced multi-view data, where human mistakes introduce incorrect labels, creating label noise. As datasets expand, the impact of label noise becomes more severe (Figure 1). This typically

* Corresponding Author.

arises from two sources: (1) ambiguous objects leading to random mislabeling (e.g., missing 'cloud' while incorrectly labeling 'harbor') and (2) confusion between similar objects (e.g., mistaking a 'column' for a 'fence'). The co-existence of label noise and missing labels presents a greater challenge.

In response to these challenges, incomplete multi-view noisy-label learning has emerged as a novel learning paradigm designed to address the complexities of real-world, large-scale multi-view and multi-label data with prevalent incomplete views and erroneous annotations. To tackle these issues, this paper introduces a novel approach: Dynamic Multiple High-Order Correlations Fusion with Noise Filtering for Incomplete Multi-View Noisy-Label Learning (DMHNF). The proposed model addresses two pivotal challenges: (1) the structural differences and correlations across multiple views, and (2) the detection and mitigation of label noise, which significantly hampers learning performance.

On one hand, in addressing the structural correlation differences across views, most existing studies emphasize integrating features and second-order correlations (e.g., graph structures [Hamilton *et al.*, 2017; Veličković *et al.*, 2017; Defferrard *et al.*, 2016]) between samples from different views. However, limited attention has been given to the exploration and integration of high-order correlations between samples across views. High-order correlations [Gao *et al.*, 2020] encapsulate the distinct distributional characteristics of each view, which become particularly critical when certain views are incomplete. In such cases, the high-order correlations present in the remaining views can provide crucial information to compensate for the missing data. On the other hand, regarding the challenge of detecting and mitigating label noise, most current multi-label learning approaches focus on leveraging label correlations to impute missing labels. However, these methods often fail to utilize high-order correlations between samples to assist in correcting label noise. Our approach addresses this gap by incorporating noise correction matrices within the hypergraph network, enabling the dynamic adjustment and correction of noisy labels. By leveraging high-order correlations in the feature space, we can also refine the label space, allowing for label noise correction from a distributional perspective. These integrated strategies enable the proposed method to attain both high accuracy and robustness. The key contributions of this study are as follows:

(1) We introduce the incomplete multi-view noisy-label learning paradigm, which concurrently addresses view incompleteness and label noise. This paradigm is well-suited for large-scale data, where multi-view descriptions are often incomplete and annotations frequently contain errors.

(2) We propose a novel dynamic multi-hypergraph fusion network that enhances multi-view data integration by concurrently fusing high-order structural and feature information. These two types of information are exploited in parallel, mutually reinforcing one another.

(3) Our method facilitates communication between the feature and label spaces via the dynamic hypergraph, leveraging high-order correlations in the feature space to correct label noise. By integrating noise filtering matrices with the hypergraph neural network, we reduce the detrimental effects of label noise from a distributional standpoint.

2 Related Work

Multi-view multi-label data is increasingly common in modern datasets, but as data volume grows, ensuring completeness and label accuracy remains a challenge. This has elevated incomplete multi-view weak-label learning as a key focus area in machine learning research. Tan et al. pioneered the field with their model iMVWL [Tan *et al.*, 2018], which jointly optimizes a shared subspace for incomplete views and a weak label classifier while learning local label correlations. This approach tackles the dual challenges of missing views and incomplete labels, laying the groundwork for subsequent research. Building upon these insights, Li et al. developed NAIM3L [Li and Chen, 2021], a concise yet powerful model that addresses non-aligned incomplete multi-view and multi-label learning. It leverages multi-view consensus and uncovers both global and local label structures, aligning individual labels across diverse views to manage non-aligned data more effectively. Given the impact of deep learning in this domain, Liu et al. introduced DICNet [Liu *et al.*, 2023], a deep instance-level contrastive network designed for doubly incomplete multi-view multi-label classification. This model employs autoencoders for end-to-end feature extraction and implements a contrastive learning strategy, combined with a multi-view weighted fusion module to boost classification accuracy. Lastly, Liu et al. developed a method SIP [Liu *et al.*, 2024] based on semantic invariance and prototype modeling. By applying the Information Bottleneck Theory, it compresses cross-view representations to maximize shared information, enhancing the accuracy of label prediction while modeling multi-label prototypes in latent space. Despite these advances, significant complexities remain when dealing with label noise in incomplete multi-view multi-label data, highlighting the need for further innovation in this rapidly evolving area of research.

Various methodologies [Cui *et al.*, 2020; Zhang *et al.*, 2019; Xie and Huang, 2022] have emerged to address the intricate scenarios characterized by the coexistence of label noise and missing labels. Chen et al. introduced a generalized noisy multi-label classification method [Chen *et al.*, 2024] utilizing a label embedding network that captures the intricate correlations between the feature and label spaces. This innovative method employs regularization on noisy predictions based on label correlations, yielding promising outcomes in scenarios marked by the presence of both label noise and missing labels. From this analysis, it is evident that the complexities associated with multi-view multi-label learning escalate significantly when both label noise and missing labels are present. Consequently, models must be adept at integrating incomplete representations across diverse views while concurrently managing label noise and missing labels with efficacy. The dynamic hypergraph network [Jiang *et al.*, 2019; Fu *et al.*, 2022; Zhou *et al.*, 2023] is suitable for discovering and utilizing the complex high-order correlations in the problem proposed in this paper. In this paper, we tackle this multifaceted challenge by designing a dynamic multi-hypergraph neural network supplemented by label filtering matrices, leveraging high-order correlations among samples to address these critical issues effectively.

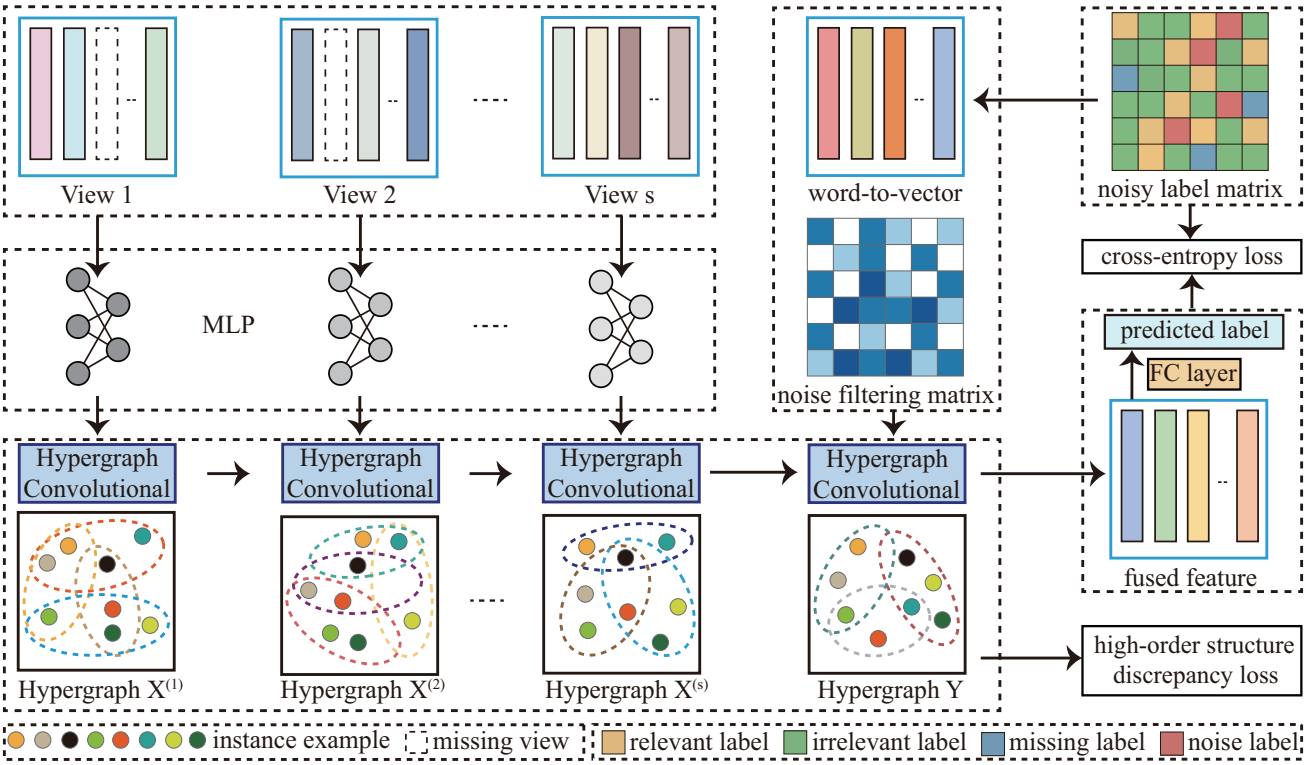

Figure 2. Framework of the proposed Dynamic Multiple High-order Correlations Fusion with Noise Filtering for Incomplete Multi-view Noisy-label Learning, consisting of two main modules: Dynamic Multi-hypergraph Neural Network and Hybrid Noise Filtering.

3 Proposed Approach

The dataset is denoted as $\mathcal{D} = \{\{X^{(q)} \in \mathbb{R}^{m\times d_q}\}_{q=1}^{s}, Y\}$, where $X^{(q)} \in \mathbb{R}^{m\times d_q}$ is the q-th view (with feature dimension d_q), $Y \in \{0,1\}^{m\times n}$ is the label matrix, $W \in \{0,1\}^{m\times s}$ indicates missing views, and m and n denote the numbers of samples and label classes, respectively. The objective is to learn a mapping function $f : X \to Y$ through which the label $\hat{y}$ of a test instance $\hat{x}$ can be accurately inferred. Our method addresses both incomplete data and label noise with two key contributions (Figure 2). First, we model high-order correlations across views, treating the label space as a special view, using a dynamic multi-hypergraph neural network. This network effectively integrates these correlations to enhance learning. Second, we design two label-filtering matrices that refine noisy labels and address missing information. These matrices, tightly integrated within the hypergraph neural network, mitigate the adverse effects of label noise while leveraging high-order structural information.

3.1 Dynamic Multi-hypergraph Neural Network

Building the multi-hypergraph neural network first requires constructing the corresponding hypergraphs in both the feature space and the label space. For the feature space hypergraph, we adopt a clustering-based approach to construct hyperedges, which in turn forms the hypergraph. Specifically, for each view, an independent hypergraph is built based on the data from that view. In this setup, each sample is treated as a vertex, with corresponding hyperedges generated for each vertex [Zhou *et al.*, 2006]. A hypergraph $G = (V, E, \omega)$ consists of vertices V, hyperedges E, and weights $\omega(e)$. It is represented by an incidence matrix H, where $h(v, e) = 1$ if and only if $v \in e$. Vertex and hyperedge degrees, $d(v)$ and $\delta(e)$, are computed based on H, enabling random walks on G to model relationships as a finite Markov chain. Following the method outlined in [Zhu *et al.*, 2017a], hyperedges in the feature space $H^{(q)}$ are constructed using the formula:

$$e_i^{(q)} = \{v_j \mid \theta(x_i, x_j) \geq 0.1\sigma_i^{(q)}\}, \quad i, j = 1, \ldots, m \tag{1}$$

where $\theta(x_i, x_j)$ represents the similarity between samples x_i and x_j, and $\sigma_i^{(q)}$ is the average similarity between x_i and all other samples. For the label space hypergraph, we construct two hypergraphs. In the first hypergraph, labels are treated as hyperedges, allowing the label matrix to be directly represented as a hypergraph incidence matrix $H^l = Y$. In the second hypergraph, H^c is constructed to contains semantic information of labels, hyperedges in which are constructed using the formula:

$$e_i^c = \{v_j \mid \omega(y_i, y_j) \geq 0.1\sigma_i^c\}, \quad i, j = 1, \ldots, m \tag{2}$$

where $\omega(y_i, y_j)$ denotes sum of the pairwise distances between the word-to-vector representations of corresponding relevant labels in the two label sets, and σ_i^c is the average similarity between y_i and all other samples.

This neural network is primarily used to fuse information from multiple views of the data, including both the structural

correlations information among the data and the feature information of the data. So the fusion occurs on two levels: the fusion of high-order structural information and the fusion of feature information from different views. In this study, we introduce a dynamic multi-hypergraph updating mechanism for the fusion of high-order structural information. This method aims to generate a new hypergraph by fusing hypergraph sequences from prior network iterations. Notably, here we do not directly construct a new hypergraph based on the preceding hypergraph sequence. Instead, we formulate the Laplacian matrix of the new hypergraph, and subsequently, build a hypergraph neural network based on this Laplacian matrix [Feng *et al.*, 2019]. The specific fusion process is delineated as follows.

We associate each hypergraph $(H^{(1)}, \cdots, H^{(s)})$ [Zhou *et al.*, 2006] with a natural random walk [Chitra and Raphael, 2019; Carletti *et al.*, 2020]. This association allows us to fuse the transition probabilities and stationary distributions of these hypergraphs to obtain the Laplacian matrix of the fused hypergraph. Let hypergraphs $H^{(1)}$ and $H^{(q)}$ representing the initial hypergraph structure and the hypergraph generated at the q-th layer, respectively.

Let T and Π represent the transition probability and stationary distribution matrix of the hypergraph random walk, respectively. The entries of T and Π are defined as follows:

$$t(u,v) = \sum_{e \in E} w(e) \frac{h(u,e)}{d(u)} \frac{h(v,e)}{\delta(e)}, \quad \pi(v) = \frac{d(v)}{vol(V)} \tag{3}$$

The corresponding matrices for $H^{(1)}$ and $H^{(q)}$ are denoted as T_1, Π_1 and T_q, Π_q. To elaborate on the fusion, the multiple hypergraph cut is explained through as follows:

$$\beta_i(u) = \frac{\alpha_i \pi_i(u)}{\sum_{i=1}^{q} \alpha_i \pi_i(u)}, \quad \sum_{i=1}^{q} \alpha_i = 1 \tag{4}$$

$$\alpha_i = (1-\alpha_1) \frac{cos(\pi_1, \pi_i)}{\sum_{i=1}^{q} cos(\pi_1, \pi_i)}, 2 \le i \le q \tag{5}$$

The parameter α_i is used to determine the relative importance of each hypergraph during fusion. The cosine similarity $(cos(\pi_i, \pi_j))$ between the stationary distributions π_i and π_j measures the degree of similarity between two hypergraphs. Notably, greater weight is assigned to the hypergraph more similar to the original, enhancing the overall convergence of the model.

The fused transition probabilities and stationary distribution are defined as:

$$t_q(u,v) = \sum_{i=1}^{q} \beta_i(u) t_i(u,v), \quad \pi_q(v) = \sum_{i=1}^{q} \alpha_i \pi_i(v) \tag{6}$$

Notably, the formulation of transition probability and stationary distribution does not follow a simple linear combination on each hypergraph. The hypergraph Laplacian correspond to hypergraph $H^{(q)}$ is denoted as $L^{(q)}$. The Laplacian matrix of the fused hypergraph is obtained using the equation:

$$L^{(q)} = \Pi - \frac{\Pi T + T^{\mathrm{T}} \Pi}{2} = \sum_{i=1}^{q} \alpha_i \Pi_i (I - T_i) \tag{7}$$

Subsequently, following the approach in [Feng *et al.*, 2019], a hyperedge convolutional layer $f(X, W, \Theta)$ is built as:

$$Y^{(q)} = \sigma\left((L^{(q)} - I) X^{(q)} \Theta^{q}\right) \tag{8}$$

Here, $X^{(q)}$ represents the hypergraph signal at the q-th layer, and σ denoting the nonlinear activation function. The obtained $Y^{(q)}$ is instrumental for subsequent learning endeavors.

In the aforementioned steps, we have completed the fusion of high-order structural correlations information across different views based on hypergraphs. The fusion of multi-view information based on the aforementioned hypergraph requires consideration of two key aspects. First, the dynamic integration of data, where information from different views is continuously incorporated into the model, much like the previous hypergraph fusion process. Second, the challenge of missing views, which is addressed by utilizing missing-view indicator matrix to facilitate data fusion. By leveraging the hypergraph correlations and multi-view information, the model effectively mitigates the adverse effects of missing views on its performance.

Here, $X^{(q)}$ represents the deep features extracted through a convolutional neural network. The hypergraph convolution based on feature space information can be expressed as:

$$X^{k+1} = \sigma\left((L^{(q)} - I)(W'_{:,q+1} X^{(q+1)}) \Theta_f^{k+1}\right) + \alpha_i X^k \tag{9}$$

where $L^{(q)}$ is laplacian matrix of the dynamic fused hypergraph of $(H^{(1)}, \cdots, H^{(q-1)})$. Θ_f^{k+1} are the learnable parameters during network training, and $\sigma(\cdot)$ is the nonlinear activation function. After the convolution operation, we obtain the image features fused by multi-view data, which capture the high-order structural information of the samples in the feature space.

3.2 Hybrid Noise Filtering

It is important to highlight that our method differs significantly from traditional incomplete multi-view multi-label approaches at both the data and model levels, with the key difference lying in the complexity of the label space. While conventional incomplete multi-view multi-label problems consider only missing labels in the label space, real-world applications commonly involve the coexistence of both missing and noisy labels (Figure 3).

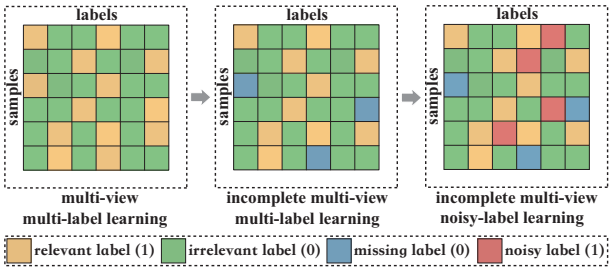

Figure 3. Label space of multi-view noisy-label learning.

Here, it is assumed that the noise in the label space is multi-label noise, meaning there exists a certain transition correlations between two labels, and this transition is independent of the sample features. However, the label space in multi-label settings is quite complex, so we aim to perform this transformation from the perspective of data distribution. These two noise correction matrices function alongside their corresponding hypergraph matrices in different label spaces, primarily to counteract the changes in distribution caused by label noise, ensuring accurate training of the classifier.

To address this, we construct two noise filtering matrices T^l and T^c. The purpose of these matrices is to transform feature representations to align with the label space, thereby achieving distribution alignment between the two.

$$\begin{aligned}\hat{Y}^{k+1} = \sigma(T^l D_v^{-1/2} H^l W D_e^{-1} (H^l)^\top D_v^{-1/2} (\hat{Y}^k) \Theta_l^k \\ + \mu T^c D_v^{-1/2} H^c W D_e^{-1} (H^c)^\top D_v^{-1/2} (\hat{Y}^k) \Theta_c^k)\end{aligned} \tag{10}$$

Next, we will explore how the label transformation matrix T (the probability of transitioning from y to $\tilde{y}$ can be expressed as $p(\tilde{y} \mid y)$) can enhance the robustness of the learned multi-label classification model against label noise, and the method for constructing this matrix.

Assume $y^{i*} = 1$ indicates that sample x truly possesses the i-th label, and $\tilde{y}^j = 1$ indicates that sample x has been labeled with the j-th label under noise interference. The transition between the two can be achieved through the label transformation matrix $T^*_{ji} = p(\tilde{y}^j = 1 \mid y^{i*} = 1)$, where the label transformation matrix $T \in \mathbb{R}^{n \times n}$ is an asymmetric matrix. Assume the classifier trained by the model predicts the true label probabilities as $\hat{p}(y^{i*} = 1 \mid \mathbf{x}, \theta)$. By utilizing the label transformation matrix T, the model is adjusted to align the predicted distribution with the label distribution of the noisy data.

$$\hat{p}(\tilde{y}^j = 1 \mid \mathbf{x}, \theta, T) = \sum_i t_{ji} \hat{p}(y^{i*} = 1 \mid \mathbf{x}, \theta) \tag{11}$$

The training objective of the multi-label classification model is to predict the true and accurate label set y^*, rather than the noisy label set $\tilde{y}$. After adding T to the model, another confusion matrix $\tilde{U}$ will be obtained:

$$\tilde{u}_{ij} := \frac{1}{|S_j|} \sum_{n \in S_j} \hat{p}(\tilde{y}^i = 1 \mid \mathbf{x}_n, \theta, T) \tag{12}$$

where S_j is the set of training samples with label $y^{j*} = 1$. The meaning of the confusion matrix is the probability that a sample belonging to label j is misclassified as label i. Therefore, if U can be made to become the identity matrix, it implies that the base model can predict the true labels in the training data.

According to Equation (11), we can obtain $\tilde{U} = TU$. It should be noted that in multi-view multi-label learning, each sample carries multiple labels, and the likelihood of the true U becoming an identity matrix is very low. Thus, we only use this idea to correct noisy labels. The model with the transformation matrix T constrains the confusion matrix $\tilde{U}$ to be consistent with the true distribution of noisy labels.

$$\tilde{u}_{ik} = \frac{1}{|S_k|} \sum_{n \in S_k} \hat{p}(\tilde{y}^i \mid x_n, y_n^{k*}) \to t^*_{ik} \Longrightarrow \tilde{U} \to T^* \tag{13}$$

By constraining $T = T^*$, this will force U to converge to the identity matrix. Therefore, using the model with the added transformation matrix T for training will reduce the probability of label misclassification and enhance the model's robustness against noise. From the above analysis, it is evident that the noise transformation matrix T is the core of the noise correction module. T can be approximately computed through the following two steps:

$$\bar{x}^i = \arg\max_{x \in X} \hat{p}(\tilde{y}^i = 1 \mid x), \quad T_{ij} = \hat{p}(\tilde{y}^j = 1 \mid \bar{x}^i) \tag{14}$$

3.3 Overall Loss Function

To train the entire network, we employ the cross-entropy loss function, defined as follows:

$$L_{ce} = \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij} \log(\hat{y}_{ij}) + (1 - y_{ij}) \log(1 - \hat{y}_{ij}) \tag{15}$$

To ensure consistency between the feature space and the label space after multi-view information fusion, we propose the following high-order structure discrepancy loss:

$$L_f(\hat{X}^f, \hat{Y}^{lc}) = \|\hat{X}^f - \hat{Y}^{lc}\|_F^2 \tag{16}$$

Additionally, we leverage the canonical Kullback-Leibler (KL) divergence to measure the alignment between the representations processed by the two noise-correcting matrices. These matrices derived from hypergraphs constructed based on the label space, focus on different aspects of the data.

$$L_{kl} = \sum_{i=1}^{m} \mathrm{KL}\left(\hat{y}^l_{ij} | 1 \le j \le n, \hat{y}^c_{ij} | 1 \le j \le n\right) \tag{17}$$

In summary, the overall loss function for the proposed model is composed of three key components:

$$L = L_{ce} + \beta L_f + \gamma L_{kl} \tag{18}$$

These loss terms guide training by aligning the feature and label spaces, improving classification accuracy and robustness in noisy and incomplete multi-view data.

4 Experiments

4.1 Data sets and Experimental settings

We evaluated the proposed DMHNF on five benchmark datasets: Corel5k [Duygulu *et al.*, 2002], Pascal07 [Everingham *et al.*, 2010], ESPGame [Von Ahn and Dabbish, 2004], IAPRTC12 [Guillaumin *et al.*, 2009], and Mirflickr [Huiskes and Lew, 2008], using six feature sets: GIST, HSV, DenseHue, DenseSift, RGB, and LAB. Evaluation used six metrics: ranking loss (RL), average precision (AP), Hamming loss (HL), area under the curve (AUC), OneError (OE), and Coverage (Cov). For clearer comparison, we adopted 1-RL, 1-HL, 1-OE, and 1-Cov, where higher values indicate better performance. Our approach was compared with state-of-the-art methods in multi-label learning (GLOCAL [Zhu *et al.*, 2017b], DM2L [Ma and Chen, 2021], CDMM [Xie and Huang, 2022]) and multi-view multi-label learning (LVSL [Zhao *et al.*, 2023], iMVWL [Tan *et al.*, 2018], NAIM3L [Li and Chen, 2021], DICNet [Liu *et al.*, 2023], SIP [Liu

Data	Metric	GLOCAL	CDMM	DM2L	LVSL	iMVWL	NAIM3L	DICNet	SIP	DMHNF
Corel5k	AP	$0.285_{0.004}$	$0.354_{0.004}$	$0.262_{0.005}$	$0.342_{0.004}$	$0.283_{0.008}$	$0.309_{0.004}$	$0.381_{0.004}$	$0.418_{0.009}$	$0.426_{0.004}$
	1-HL	$0.987_{0.000}$	$0.987_{0.000}$	$0.987_{0.000}$	$0.987_{0.000}$	$0.978_{0.000}$	$0.987_{0.000}$	$0.988_{0.000}$	$0.988_{0.000}$	$0.991_{0.000}$
	1-RL	$0.840_{0.003}$	$0.884_{0.003}$	$0.843_{0.002}$	$0.881_{0.003}$	$0.865_{0.005}$	$0.878_{0.002}$	$0.882_{0.004}$	$0.911_{0.003}$	$0.923_{0.004}$
	AUC	$0.843_{0.003}$	$0.888_{0.003}$	$0.845_{0.002}$	$0.884_{0.003}$	$0.868_{0.005}$	$0.881_{0.002}$	$0.884_{0.004}$	$0.913_{0.003}$	$0.927_{0.005}$
	1-OE	$0.327_{0.010}$	$0.410_{0.007}$	$0.295_{0.014}$	$0.391_{0.009}$	$0.311_{0.015}$	$0.350_{0.009}$	$0.468_{0.007}$	$0.489_{0.016}$	$0.497_{0.007}$
	1-Cov	$0.648_{0.006}$	$0.723_{0.007}$	$0.647_{0.005}$	$0.718_{0.006}$	$0.702_{0.008}$	$0.725_{0.005}$	$0.727_{0.011}$	$0.787_{0.009}$	$0.791_{0.011}$
Pascal07	AP	$0.496_{0.004}$	$0.508_{0.005}$	$0.471_{0.008}$	$0.504_{0.005}$	$0.437_{0.018}$	$0.488_{0.003}$	$0.505_{0.012}$	$0.555_{0.010}$	$0.567_{0.008}$
	1-HL	$0.927_{0.000}$	$0.931_{0.001}$	$0.928_{0.001}$	$0.930_{0.000}$	$0.882_{0.004}$	$0.928_{0.001}$	$0.929_{0.001}$	$0.931_{0.001}$	$0.943_{0.001}$
	1-RL	$0.767_{0.004}$	$0.812_{0.004}$	$0.761_{0.005}$	$0.806_{0.003}$	$0.736_{0.015}$	$0.783_{0.001}$	$0.783_{0.008}$	$0.830_{0.004}$	$0.833_{0.008}$
	AUC	$0.786_{0.003}$	$0.838_{0.003}$	$0.779_{0.004}$	$0.832_{0.002}$	$0.767_{0.015}$	$0.811_{0.001}$	$0.809_{0.006}$	$0.850_{0.005}$	$0.869_{0.006}$
	1-OE	$0.443_{0.005}$	$0.419_{0.008}$	$0.420_{0.011}$	$0.419_{0.008}$	$0.362_{0.023}$	$0.421_{0.006}$	$0.427_{0.015}$	$0.464_{0.018}$	$0.487_{0.006}$
	1-Cov	$0.703_{0.004}$	$0.759_{0.003}$	$0.692_{0.004}$	$0.751_{0.003}$	$0.677_{0.015}$	$0.727_{0.002}$	$0.731_{0.006}$	$0.783_{0.006}$	$0.791_{0.003}$
ESPGame	AP	$0.221_{0.002}$	$0.289_{0.003}$	$0.212_{0.002}$	$0.285_{0.003}$	$0.244_{0.005}$	$0.246_{0.002}$	$0.297_{0.002}$	$0.311_{0.004}$	$0.328_{0.006}$
	1-HL	$0.982_{0.000}$	$0.983_{0.000}$	$0.982_{0.000}$	$0.983_{0.000}$	$0.972_{0.000}$	$0.983_{0.000}$	$0.983_{0.000}$	$0.983_{0.000}$	$0.983_{0.003}$
	1-RL	$0.780_{0.004}$	$0.832_{0.001}$	$0.781_{0.001}$	$0.829_{0.001}$	$0.808_{0.002}$	$0.818_{0.002}$	$0.832_{0.001}$	$0.849_{0.002}$	$0.867_{0.001}$
	AUC	$0.784_{0.004}$	$0.836_{0.001}$	$0.785_{0.001}$	$0.833_{0.002}$	$0.813_{0.002}$	$0.824_{0.002}$	$0.836_{0.001}$	$0.853_{0.002}$	$0.857_{0.008}$
	1-OE	$0.317_{0.005}$	$0.396_{0.005}$	$0.294_{0.006}$	$0.389_{0.004}$	$0.343_{0.013}$	$0.339_{0.003}$	$0.439_{0.007}$	$0.455_{0.007}$	$0.462_{0.002}$
	1-Cov	$0.496_{0.006}$	$0.574_{0.004}$	$0.488_{0.003}$	$0.567_{0.005}$	$0.548_{0.004}$	$0.571_{0.003}$	$0.593_{0.003}$	$0.628_{0.005}$	$0.637_{0.006}$
IAPRTC12	AP	$0.256_{0.002}$	$0.305_{0.004}$	$0.234_{0.003}$	$0.304_{0.004}$	$0.237_{0.003}$	$0.261_{0.001}$	$0.323_{0.001}$	$0.331_{0.006}$	$0.348_{0.003}$
	1-HL	$0.980_{0.000}$	$0.981_{0.000}$	$0.980_{0.000}$	$0.981_{0.000}$	$0.969_{0.000}$	$0.980_{0.000}$	$0.981_{0.000}$	$0.980_{0.000}$	$0.982_{0.001}$
	1-RL	$0.825_{0.002}$	$0.862_{0.002}$	$0.823_{0.002}$	$0.861_{0.002}$	$0.833_{0.002}$	$0.848_{0.001}$	$0.873_{0.001}$	$0.885_{0.003}$	$0.896_{0.000}$
	AUC	$0.830_{0.001}$	$0.864_{0.002}$	$0.825_{0.001}$	$0.863_{0.001}$	$0.835_{0.001}$	$0.850_{0.001}$	$0.874_{0.000}$	$0.886_{0.002}$	$0.897_{0.007}$
	1-OE	$0.378_{0.007}$	$0.432_{0.008}$	$0.340_{0.006}$	$0.429_{0.009}$	$0.352_{0.008}$	$0.390_{0.005}$	$0.468_{0.002}$	$0.463_{0.009}$	$0.464_{0.005}$
	1-Cov	$0.534_{0.003}$	$0.597_{0.004}$	$0.529_{0.004}$	$0.597_{0.004}$	$0.564_{0.005}$	$0.592_{0.004}$	$0.649_{0.001}$	$0.675_{0.007}$	$0.686_{0.004}$
Mirflickr	AP	$0.537_{0.002}$	$0.570_{0.002}$	$0.514_{0.006}$	$0.553_{0.002}$	$0.490_{0.012}$	$0.551_{0.002}$	$0.589_{0.005}$	$0.614_{0.004}$	$0.629_{0.007}$
	1-HL	$0.874_{0.001}$	$0.886_{0.001}$	$0.878_{0.001}$	$0.885_{0.001}$	$0.839_{0.002}$	$0.882_{0.001}$	$0.888_{0.002}$	$0.891_{0.001}$	$0.896_{0.003}$
	1-RL	$0.832_{0.001}$	$0.856_{0.001}$	$0.831_{0.003}$	$0.856_{0.001}$	$0.803_{0.008}$	$0.844_{0.001}$	$0.863_{0.004}$	$0.877_{0.002}$	$0.879_{0.008}$
	AUC	$0.828_{0.001}$	$0.846_{0.001}$	$0.828_{0.003}$	$0.844_{0.001}$	$0.787_{0.012}$	$0.837_{0.001}$	$0.849_{0.004}$	$0.860_{0.003}$	$0.862_{0.005}$
	1-OE	$0.552_{0.005}$	$0.631_{0.004}$	$0.510_{0.008}$	$0.607_{0.004}$	$0.511_{0.022}$	$0.585_{0.003}$	$0.637_{0.007}$	$0.662_{0.008}$	$0.676_{0.010}$
	1-Cov	$0.605_{0.003}$	$0.640_{0.001}$	$0.604_{0.005}$	$0.636_{0.001}$	$0.572_{0.013}$	$0.631_{0.002}$	$0.652_{0.007}$	$0.678_{0.003}$	$0.685_{0.007}$

Table 1. Quantitative results on data sets with 50% missing-view and 50% missing-label rates (the bottom right digit is the standard deviation).

et al., 2024]). We designed two experimental scenarios using 70% of the dataset for training. In the first scenario, 50% of instances in each view were randomly deactivated, ensuring each sample retained at least one view, while 50% of both positive and negative labels were removed to simulate partial label absence. In the second scenario, label noise was introduced alongside view incompleteness by randomly adding and removing labels, simulating 20% false positives and 20% false negatives to reflect a realistic noisy label space.

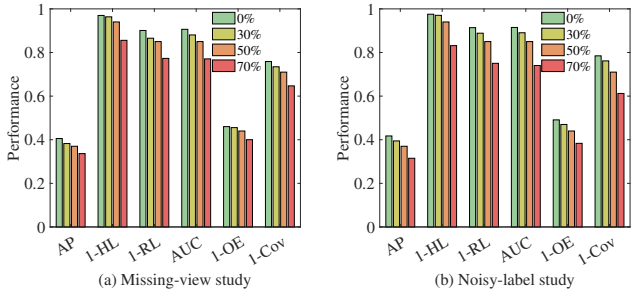

Figure 4. Performance under varying missing views and label noise.

4.2 Experimental Results and Analysis

Tables 1 and 2 present experimental results on datasets with missing-label and noisy-label conditions in incomplete multi-view data. Table 3 shows the ablation study, while Figure 4 illustrates performance trends with varying missing-view and noisy-label rates. Based on these findings, we conclude:

1. As shown in Tables 1 and 2, our method consistently outperforms existing approaches in handling missing views and noisy labels across all benchmark datasets. In particular, when compared with state-of-the-art multi-view multi-label learning models, as well as methods specifically designed to address missing modalities or incomplete labels, our approach achieves superior performance under both missing-view and label-noise conditions. These results underscore the effectiveness and robustness of our model.

2. Our method excels in scenarios where both missing labels and label noise coexist with incomplete views, outperforming methods addressing only missing labels. This emphasizes the method's robustness in complex settings, particularly when label noise accompanies missing data, showcasing its broad applicability and enhanced practical utility.

3. Analysis of Tables 1 and 2 reveals that missing and noisy labels together present a more challenging problem than missing labels alone. Despite lower proportions of miss-

Data	Metric	GLOCAL	CDMM	DM2L	LVSL	iMVWL	NAIM3L	DICNet	SIP	DMHNF
Corel5k	AP	$0.205_{0.004}$	$0.268_{0.005}$	$0.326_{0.010}$	$0.243_{0.009}$	$0.304_{0.007}$	$0.263_{0.004}$	$0.291_{0.003}$	$0.344_{0.008}$	$0.375_{0.004}$
	1-HL	$0.936_{0.000}$	$0.944_{0.002}$	$0.939_{0.000}$	$0.943_{0.000}$	$0.941_{0.000}$	$0.927_{0.000}$	$0.942_{0.000}$	$0.938_{0.000}$	$0.949_{0.000}$
	1-RL	$0.734_{0.002}$	$0.782_{0.003}$	$0.824_{0.004}$	$0.762_{0.009}$	$0.798_{0.004}$	$0.797_{0.002}$	$0.794_{0.003}$	$0.795_{0.004}$	$0.849_{0.009}$
	AUC	$0.735_{0.006}$	$0.772_{0.010}$	$0.821_{0.003}$	$0.763_{0.002}$	$0.797_{0.004}$	$0.786_{0.002}$	$0.793_{0.005}$	$0.782_{0.002}$	$0.843_{0.003}$
	1-OE	$0.241_{0.017}$	$0.292_{0.009}$	$0.371_{0.018}$	$0.269_{0.011}$	$0.353_{0.013}$	$0.277_{0.016}$	$0.316_{0.012}$	$0.423_{0.007}$	$0.441_{0.018}$
	1-Cov	$0.531_{0.007}$	$0.592_{0.010}$	$0.645_{0.006}$	$0.581_{0.012}$	$0.653_{0.008}$	$0.628_{0.011}$	$0.646_{0.005}$	$0.655_{0.003}$	$0.704_{0.012}$
Pascal07	AP	$0.446_{0.002}$	$0.465_{0.006}$	$0.478_{0.007}$	$0.423_{0.008}$	$0.468_{0.004}$	$0.399_{0.020}$	$0.453_{0.003}$	$0.469_{0.005}$	$0.510_{0.011}$
	1-HL	$0.862_{0.001}$	$0.885_{0.002}$	$0.891_{0.001}$	$0.887_{0.002}$	$0.887_{0.000}$	$0.834_{0.004}$	$0.863_{0.003}$	$0.874_{0.001}$	$0.891_{0.001}$
	1-RL	$0.674_{0.007}$	$0.708_{0.005}$	$0.754_{0.003}$	$0.697_{0.008}$	$0.746_{0.002}$	$0.663_{0.012}$	$0.707_{0.002}$	$0.709_{0.009}$	$0.753_{0.005}$
	AUC	$0.692_{0.008}$	$0.724_{0.002}$	$0.740_{0.003}$	$0.716_{0.004}$	$0.753_{0.001}$	$0.702_{0.020}$	$0.752_{0.003}$	$0.744_{0.002}$	$0.776_{0.004}$
	1-OE	$0.398_{0.009}$	$0.394_{0.004}$	$0.378_{0.010}$	$0.381_{0.009}$	$0.372_{0.007}$	$0.352_{0.019}$	$0.387_{0.005}$	$0.393_{0.017}$	$0.422_{0.012}$
	1-Cov	$0.611_{0.008}$	$0.643_{0.004}$	$0.677_{0.002}$	$0.622_{0.003}$	$0.684_{0.006}$	$0.603_{0.019}$	$0.668_{0.003}$	$0.660_{0.007}$	$0.714_{0.004}$
ESPGame	AP	$0.188_{0.003}$	$0.196_{0.005}$	$0.262_{0.002}$	$0.191_{0.004}$	$0.259_{0.001}$	$0.216_{0.007}$	$0.224_{0.002}$	$0.277_{0.003}$	$0.284_{0.006}$
	1-HL	$0.927_{0.001}$	$0.916_{0.002}$	$0.924_{0.000}$	$0.926_{0.000}$	$0.918_{0.001}$	$0.923_{0.002}$	$0.912_{0.000}$	$0.929_{0.000}$	$0.936_{0.000}$
	1-RL	$0.692_{0.002}$	$0.704_{0.005}$	$0.753_{0.001}$	$0.705_{0.003}$	$0.758_{0.002}$	$0.723_{0.004}$	$0.737_{0.003}$	$0.754_{0.001}$	$0.761_{0.002}$
	AUC	$0.705_{0.003}$	$0.702_{0.002}$	$0.756_{0.003}$	$0.708_{0.001}$	$0.751_{0.004}$	$0.733_{0.003}$	$0.747_{0.002}$	$0.752_{0.001}$	$0.771_{0.002}$
	1-OE	$0.236_{0.021}$	$0.282_{0.004}$	$0.354_{0.002}$	$0.266_{0.006}$	$0.347_{0.004}$	$0.309_{0.013}$	$0.305_{0.005}$	$0.398_{0.009}$	$0.406_{0.007}$
	1-Cov	$0.446_{0.003}$	$0.448_{0.008}$	$0.519_{0.001}$	$0.437_{0.003}$	$0.517_{0.006}$	$0.481_{0.002}$	$0.498_{0.004}$	$0.512_{0.003}$	$0.521_{0.005}$
IAPRTC12	AP	$0.207_{0.005}$	$0.231_{0.001}$	$0.271_{0.004}$	$0.216_{0.002}$	$0.275_{0.006}$	$0.214_{0.003}$	$0.230_{0.000}$	$0.275_{0.010}$	$0.291_{0.007}$
	1-HL	$0.926_{0.000}$	$0.933_{0.001}$	$0.936_{0.000}$	$0.937_{0.002}$	$0.934_{0.002}$	$0.927_{0.001}$	$0.938_{0.003}$	$0.937_{0.000}$	$0.945_{0.001}$
	1-RL	$0.723_{0.002}$	$0.746_{0.006}$	$0.781_{0.001}$	$0.743_{0.003}$	$0.773_{0.005}$	$0.758_{0.002}$	$0.769_{0.000}$	$0.785_{0.003}$	$0.793_{0.001}$
	AUC	$0.724_{0.004}$	$0.748_{0.001}$	$0.782_{0.005}$	$0.744_{0.002}$	$0.785_{0.006}$	$0.750_{0.002}$	$0.774_{0.003}$	$0.783_{0.000}$	$0.806_{0.002}$
	1-OE	$0.271_{0.008}$	$0.340_{0.030}$	$0.396_{0.004}$	$0.307_{0.002}$	$0.381_{0.009}$	$0.316_{0.006}$	$0.357_{0.003}$	$0.405_{0.005}$	$0.411_{0.001}$
	1-Cov	$0.472_{0.004}$	$0.481_{0.005}$	$0.538_{0.003}$	$0.479_{0.002}$	$0.535_{0.010}$	$0.507_{0.004}$	$0.534_{0.007}$	$0.588_{0.003}$	$0.605_{0.005}$
Mirflickr	AP	$0.457_{0.002}$	$0.489_{0.005}$	$0.527_{0.008}$	$0.460_{0.003}$	$0.502_{0.002}$	$0.441_{0.009}$	$0.491_{0.006}$	$0.537_{0.001}$	$0.563_{0.007}$
	1-HL	$0.784_{0.002}$	$0.805_{0.000}$	$0.812_{0.003}$	$0.817_{0.002}$	$0.811_{0.001}$	$0.774_{0.001}$	$0.811_{0.000}$	$0.813_{0.002}$	$0.826_{0.003}$
	1-RL	$0.736_{0.001}$	$0.741_{0.002}$	$0.767_{0.004}$	$0.749_{0.000}$	$0.776_{0.001}$	$0.720_{0.003}$	$0.767_{0.002}$	$0.771_{0.001}$	$0.792_{0.003}$
	AUC	$0.724_{0.002}$	$0.748_{0.003}$	$0.763_{0.001}$	$0.741_{0.004}$	$0.753_{0.002}$	$0.705_{0.001}$	$0.754_{0.002}$	$0.752_{0.003}$	$0.773_{0.004}$
	1-OE	$0.457_{0.007}$	$0.494_{0.006}$	$0.574_{0.008}$	$0.455_{0.003}$	$0.546_{0.002}$	$0.461_{0.005}$	$0.524_{0.004}$	$0.576_{0.001}$	$0.591_{0.006}$
	1-Cov	$0.535_{0.005}$	$0.548_{0.004}$	$0.573_{0.003}$	$0.547_{0.001}$	$0.573_{0.001}$	$0.515_{0.002}$	$0.574_{0.006}$	$0.593_{0.004}$	$0.612_{0.005}$

Table 2. Quantitative results on data sets with 40% missing-view and 40% noisy-label rates (the bottom right digit is the standard deviation).

ing/noisy labels in Table 2, overall performance is weaker, confirming that our method is more effective in handling the combined challenges of missing and noisy labels.

4. We conducted ablation experiments, showing that high-order correlations are essential. Replacing them with second-order correlations weakens their interaction with filtering matrices. Without these matrices, inaccurate high-order correlations degrade accuracy more than second-order ones, highlighting their complementary roles. Using two filtering matrices outperforms one due to the label space's complexity. A hypergraph based solely on word-to-vector representations results in sparse high-order correlations. Combining both approaches enhances accuracy, improving noise correction.

Method				Corel5k		Pascal07	
Backbone	M-HNN	T^l	T^c	AP	AUC	AP	AUC
✓				0.321	0.818	0.470	0.757
✓	✓			0.363	0.829	0.491	0.767
✓	✓	✓		0.368	0.837	0.498	0.771
✓	✓		✓	0.364	0.836	0.501	0.769
✓	✓	✓	✓	0.375	0.843	0.510	0.776

Table 3. Ablation results on two datasets with 40% missing-view rate and 40% noisy-label rate.

5 Conclusion

This paper presents a novel dynamic multi-hypergraph neural network model specifically designed for addressing the challenges of incomplete multi-view noisy-label learning. The model integrates multi-view data by capturing high-order correlations among samples, effectively leveraging both feature and structural information across different views. To address label noise, we propose label noise filtering matrices that corrects disrupted correlations in the label space by utilizing the intrinsic high-order correlations in the feature space. Extensive experiments demonstrate that our approach not only excels in incomplete multi-view multi-label learning scenarios, but also maintains robust performance in the more challenging setting of incomplete multi-view learning with noisy labels. These findings further underscore the effectiveness, adaptability, and broad applicability of the proposed model.

Moreover, this work introduces a novel problem formulation that closely mirrors the complexities of real-world multi-label learning scenarios, thereby offering a principled and practical solution to challenges arising from incomplete views and noisy annotations. We believe these contributions lay a solid groundwork for future research, encouraging further exploration and advancement of dynamic multi-hypergraph methods in diverse application domains.

Acknowledgements

This work is supported by the National Key Research and Development Program of China (No. 2022YFB3305504).

References

[Carletti *et al.*, 2020] Timoteo Carletti, Federico Battiston, Giulia Cencetti, and Duccio Fanelli. Random walks on hypergraphs. *Physical review E*, 101(2):022308, 2020.

[Chen *et al.*, 2024] Jia-Yao Chen, Shao-Yuan Li, Sheng-Jun Huang, Songcan Chen, Lei Wang, and Ming-Kun Xie. Unm: A universal approach for noisy multi-label learning. *IEEE Transactions on Knowledge and Data Engineering*, 36:4968–4980, 2024.

[Chitra and Raphael, 2019] Uthsav Chitra and Benjamin Raphael. Random walks on hypergraphs with edge-dependent vertex weights. In *International Conference on Machine Learning*, pages 1172–1181, 2019.

[Cui *et al.*, 2020] Zijun Cui, Yong Zhang, and Qiang Ji. Label error correction and generation through label relationships. In *AAAI Conference on Artificial Intelligence*, pages 3693–3700, 2020.

[Defferrard *et al.*, 2016] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, pages 3844–3852, 2016.

[Duygulu *et al.*, 2002] Pinar Duygulu, Kobus Barnard, Joao FG de Freitas, and David A Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *European Conference on Computer Vision*, pages 97–112, 2002.

[Everingham *et al.*, 2010] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88:303–338, 2010.

[Feng *et al.*, 2019] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *AAAI Conference on Artificial Intelligence*, pages 3558–3565, 2019.

[Fu *et al.*, 2022] Jun Fu, Chen Hou, Wei Zhou, Jiahua Xu, and Zhibo Chen. Adaptive hypergraph convolutional network for no-reference 360-degree image quality assessment. In *ACM International Conference on Multimedia*, pages 961–969, 2022.

[Gao *et al.*, 2020] Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2548–2566, 2020.

[Guillaumin *et al.*, 2009] Matthieu Guillaumin, Thomas Mensink, Jakob Verbeek, and Cordelia Schmid. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In *International Conference on Computer Vision*, pages 309–316, 2009.

[Hamilton *et al.*, 2017] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.

[Huiskes and Lew, 2008] Mark J Huiskes and Michael S Lew. The mir flickr retrieval evaluation. In *ACM International Conference on Multimedia Information Retrieval*, pages 39–43, 2008.

[Jiang *et al.*, 2019] Jianwen Jiang, Yuxuan Wei, Yifan Feng, Jingxuan Cao, and Yue Gao. Dynamic hypergraph neural networks. In *International Joint Conferences on Artificial Intelligence*, pages 2635–2641, 2019.

[Li and Chen, 2021] Xiang Li and Songcan Chen. A concise yet effective model for non-aligned incomplete multi-view and missing multi-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):5918–5932, 2021.

[Li *et al.*, 2024] Quanjiang Li, Tingjin Luo, Mingdie Jiang, Jiahui Liao, and Zhangqi Jiang. Deep incomplete multi-view network semi-supervised multi-label learning with unbiased loss. In *ACM International Conference on Multimedia*, pages 9048–9056, 2024.

[Liu *et al.*, 2021] Weiwei Liu, Haobo Wang, Xiaobo Shen, and Ivor W Tsang. The emerging trends of multi-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7955–7974, 2021.

[Liu *et al.*, 2023] Chengliang Liu, Jie Wen, Xiaoling Luo, Chao Huang, Zhihao Wu, and Yong Xu. Dicnet: Deep instance-level contrastive network for double incomplete multi-view multi-label classification. In *AAAI Conference on Artificial Intelligence*, pages 8807–8815, 2023.

[Liu *et al.*, 2024] Chengliang Liu, Gehui Xu, Jie Wen, Yabo Liu, Chao Huang, and Yong Xu. Partial multi-view multi-label classification via semantic invariance learning and prototype modeling. In *International Conference on Machine Learning*, pages 32253–32267, 2024.

[Ma and Chen, 2021] Zhongchen Ma and Songcan Chen. Expand globally, shrink locally: Discriminant multi-label learning with missing labels. *Pattern Recognition*, 111:107675, 2021.

[Ou *et al.*, 2024] Shilong Ou, Zhe Xue, Yawen Li, Meiyu Liang, Yuanqiang Cai, and Junjiang Wu. View-category interactive sharing transformer for incomplete multi-view multi-label learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27467–27476, 2024.

[Qian *et al.*, 2014] Buyue Qian, Xiang Wang, Jieping Ye, and Ian Davidson. A reconstruction error based framework for multi-label and multi-view learning. *IEEE Transactions on Knowledge and Data Engineering*, 27(3):594–607, 2014.

[Tan *et al.*, 2018] Qiaoyu Tan, Guoxian Yu, Carlotta Domeniconi, Jun Wang, and Zili Zhang. Incomplete multi-view weak-label learning. In *International Joint Conference on Artificial Intelligence*, pages 2703–2709, 2018.

[Veličković *et al.*, 2017] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[Von Ahn and Dabbish, 2004] Luis Von Ahn and Laura Dabbish. Labeling images with a computer game. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 319–326, 2004.

[Wang *et al.*, 2022] Qianqian Wang, Zhiqiang Tao, Quanxue Gao, and Licheng Jiao. Multi-view subspace clustering via structured multi-pathway network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5):7244–7250, 2022.

[Xie and Huang, 2022] Ming-Kun Xie and Sheng-Jun Huang. Ccmn: A general framework for learning with class-conditional multi-label noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):154–166, 2022.

[Zhang *et al.*, 2019] Changqing Zhang, Ziwei Yu, Huazhu Fu, Pengfei Zhu, Lei Chen, and Qinghua Hu. Hybrid noise-oriented multilabel learning. *IEEE Transactions on Cybernetics*, 50(6):2837–2850, 2019.

[Zhao *et al.*, 2023] Dawei Zhao, Qingwei Gao, Yixiang Lu, and Dong Sun. Non-aligned multi-view multi-label classification via learning view-specific labels. *IEEE Transactions on Multimedia*, 25:7235–7247, 2023.

[Zhou *et al.*, 2006] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems*, pages 1601–1608, 2006.

[Zhou *et al.*, 2023] Peng Zhou, Zongqian Wu, Xiangxiang Zeng, Guoqiu Wen, Junbo Ma, and Xiaofeng Zhu. Totally dynamic hypergraph neural network. In *International Joint Conference on Artificial Intelligence*, pages 2476–2483, 2023.

[Zhu *et al.*, 2015] Xiaofeng Zhu, Xuelong Li, and Shichao Zhang. Block-row sparse multiview multilabel learning for image classification. *IEEE Transactions on Cybernetics*, 46(2):450–461, 2015.

[Zhu *et al.*, 2017a] Xiaofeng Zhu, Yonghua Zhu, Shichao Zhang, Rongyao Hu, and Wei He. Adaptive hypergraph learning for unsupervised feature selection. In *International Joint Conference on Artificial Intelligence*, pages 3581–3587, 2017.

[Zhu *et al.*, 2017b] Yue Zhu, James T Kwok, and Zhi-Hua Zhou. Multi-label learning with global and local label correlation. *IEEE Transactions on Knowledge and Data Engineering*, 30(6):1081–1094, 2017.