Label Enhancement via Cross-View Fusion and Mixed Graph Propagation

Mengjiao Kai, **Chao Tan***, **Yanda Wang**†, **Juanna Zhai**, **Kang Wu**, **Ningkang Peng** and **Yanhui Gu**

School of Computer and Electronic Information/School of Artificial Intelligence, Nanjing Normal University, Nanjing, China
tutu_tanchao@163.com, yandawang@nnu.edu.cn

Abstract

Label Distribution Learning (LDL) effectively addresses label ambiguity by modeling the degree to which each label describes an instance. A key challenge in LDL is Label Enhancement (LE): recovering label distributions from logical labels. Existing LE methods typically treat logical labels as supervisory signals and learn a direct mapping from features to label distributions. However, they fail to fully exploit the rich information encoded in logical labels, limiting their performance. We propose CVMG (Cross-View Fusion and Mixed Graph-based Label Enhancement), a novel approach that addresses this limitation through two key innovations. First, we employ a cross-attention mechanism to integrate logical labels and features, leveraging their complementary information to generate enriched feature representations. Second, we construct a mixed dependency graph that captures both instance-level relationships from enhanced features and category-level dependencies from logical labels. Label distributions are then recovered through propagation over this graph. Extensive experiments on 13 real-world datasets demonstrate that CVMG significantly outperforms state-of-the-art methods, validating the effectiveness of our approach.

1 Introduction

Label ambiguity has become an important challenge in machine learning [Liu *et al.*, 2021], especially when dealing with complex real-world data that cannot be effectively represented by single or multiple labels. Traditional supervised learning methods, such as single-label learning (SLL) and multi-label learning (MLL) [Gibaja and Ventura, 2014; Moyano *et al.*, 2019; Zhang *et al.*, 2022], assign fixed labels to instances, but fail to capture the intricate relationships between labels, which are crucial in tasks like emotion recognition, medical diagnosis, and image captioning [Jia *et al.*, 2019]. Label Distribution Learning (LDL) addresses this by assigning a distribution of labels to each instance, enabling a more detailed representation of the data [Geng, 2016].

LDL requires instances with exact label distributions for the training process. However, obtaining label distributions is often costly and impractical, especially when dealing with large datasets, as most real-world datasets are labeled with logical labels [Xu *et al.*, 2019]. To solve this problem, label enhancement (LE) emerged as a promising solution. LE recovers label distributions exactly from existing logical labels. This process is vital for enabling LDL in practical applications, providing the necessary supervision for training models that can handle label distributions.

Despite the promising performance of previous LE methods, they still exhibit several limitations. First, most existing LE methods treat logical labels and features independently, failing to effectively integrate their complementary information. They overlook the fact that features and categories provide different perspectives on the same instance. Specifically, the more correlated two labels are, the closer the corresponding description degrees should be. Second, categories within the label typically exhibit semantic or co-occurrence-based relationships, which can provide valuable additional information for recovering label distributions. Specifically, the higher the correlation between two categories, the closer their corresponding description degrees should be. However, many methods fail to fully consider the interdependencies between categories during label distribution recovery, resulting in suboptimal accuracy in the recovered distributions.

To address the aforementioned challenges, we propose a novel label enhancement method—CVMG (Cross-View Fusion and Mixed Graph-based Label Enhancement). This method integrates logical labels and feature information using a cross-attention mechanism, which generates high-level feature representations. By combining complementary information from both perspectives, CVMG enhances feature representations, thereby improving the accuracy of label distribution recovery. Furthermore, CVMG constructs a mixed dependency graph that captures both instance-level and category-level dependencies. This graph structure enables efficient propagation of label information across instances and categories, addressing the issue of inadequate consideration of category dependencies in existing methods.

Our contributions are summarized as follows:

- We propose a novel label enhancement method, termed

*First corresponding author
†Second corresponding author

CVMG (Cross-View Fusion and Mixed Graph-based Label Enhancement), which leverages a cross-attention mechanism to fuse logical labels and feature information from two complementary views, thereby generating high-level feature representations.

- CVMG further constructs a mixed dependency graph that integrates instance-level dependencies induced by high-level features with category-level dependencies mined from logical labels, and performs label propagation on this mixed graph to improve the accuracy of label distribution recovery.
- Extensive experiments conducted on 13 benchmark datasets demonstrate the effectiveness and superiority of CVMG over existing state-of-the-art label enhancement methods.

2 Related Work

2.1 Label Enhancement

Label Enhancement is used to recover label distributions from logical labels, thereby facilitating more accurate Label Distribution Learning. In recent years, many LE algorithms have been proposed to address the problem of label distribution recovery. Fuzzy-based LE methods, such as fuzzy clustering-based LE [El Gayar *et al.*, 2006] and kernel-based LE [Jiang *et al.*, 2006], handle label uncertainty by constructing fuzzy membership degrees for each label class. However, it remains unclear how the fuzziness introduced by these methods can effectively enhance logical labels into accurate label distributions. Graph-based LE methods leverage the topological structure of graphs to capture relationships between instances and propagate label information. Label Propagation (LP)-based LE [Li *et al.*, 2015] and Manifold Learning (ML)-based LE [Hou *et al.*, 2016] model instance similarities through graph structures and propagate label information across similar instances, thereby enhancing logical labels and generating label distributions. Graph Laplacian-based LE (GLLE) [Xu *et al.*, 2019] further improves label distribution recovery by considering both instance-level and category-level dependencies but fails to fully exploit the category-related information embedded in logical labels. Label Enhancement with Sample Correlations (LESC) [Tang *et al.*, 2020] considers the global graph information based on the low-rank representation (LRR) assumption.

Recent research has shifted towards more advanced paradigms, aiming to model the complex relationship between features and logical labels more effectively. PLEML [Zhu *et al.*, 2021] generates privileged information via multi-label learning to recover label distributions more effectively. LEVI [Xu *et al.*, 2020] models label distributions as latent vectors and infers them through variational approximation. ConLE [Wang *et al.*, 2023] aligns features and logical labels in a shared embedding space using contrastive learning. LIB [Zheng *et al.*, 2023] further improves recovery by exploiting both label assignment and label gap information. In semi-supervised settings, SMLE [Ye *et al.*, 2025] refines label distributions for labeled and unlabeled data via kNN aggregation and label propagation, while leveraging local label correlations. Similarly, FLEM [Zhao *et al.*, 2025] jointly optimizes LE and multi-label learning in an end-to-end manner, reducing mismatches with model training. Together, these methods strengthen the connection between features and logical labels to advance label distribution recovery. While these methods have advanced the field of label distribution enhancement, they heavily depend on feature correlations and overlook the richness of information inherent in the logical labels themselves.

2.2 Attention Mechanism

The attention mechanism [Vaswani *et al.*, 2017] has become a central technique in machine learning, enabling models to focus on relevant parts of the input and thus improve overall performance. Self-attention-based models like the Transformer excel in capturing long-range dependencies [Li *et al.*, 2024], making them highly effective for sequence modeling. Given its powerful representation capabilities, the attention mechanism has also been widely applied to multi-label classification tasks [Lanchantin *et al.*, 2021; Dosovitskiy *et al.*, 2020]. [Hua *et al.*, 2020] proposed an attentional region extraction network to localize and re-coordinate attentional regions for consequent label inference. [Tan *et al.*, 2022] introduced a Transformer-driven semantic relation inference method using class activation mapping to locate regions of interest. Recent advancements in cross-attention include multimodal feature fusion, such as in the CAGRN method, where label embeddings are used as queries to interact with feature representations, thus explicitly modeling feature-label dependencies and enabling category-level relational modeling [Bi *et al.*, 2024]. Similarly, CAFF-DINO demonstrates the use of cross-attention in multi-spectral object detection, where it aligns and fuses features from different modalities using cross-attention, showing its utility in complementary information fusion [Helvig *et al.*, 2024]. Building on these ideas, we introduce a cross-attention mechanism for Label Enhancement (LE), integrating complementary information from logical labels and features, optimizing feature representation. This approach allows for the optimization of feature representations, thereby laying the foundation for more accurate label distribution recovery.

3 Methodology

3.1 Formulation of Label Enhancement

Let $\boldsymbol{X} = \left[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\right] \in \mathbb{R}^{q \times n}$ be the feature set of instances, where n is the number of instances and q is the feature dimension. The logical label vector of instance $\boldsymbol{x}_i$ is given by $\boldsymbol{y}_i = (y_{i,1}, y_{i,2}, \cdots, y_{i,c})^{\mathrm{T}} \in \{0,1\}^c$, where c is the number of labels. If the label $y_{i,k}$ is relevant to $\boldsymbol{x}_i$, then $y_{i,k} = 1$; otherwise $y_{i,k} = 0$. Let $d_{\boldsymbol{x}}^{y}$ denote the description degree of label y to feature $\boldsymbol{x}$. Then $\boldsymbol{d}_i = (d_{\boldsymbol{x}_i}^{y_{i,1}}, d_{\boldsymbol{x}_i}^{y_{i,2}}, \cdots, d_{\boldsymbol{x}_i}^{y_{i,c}})^{\mathrm{T}}$ represents the ground-truth label distribution of $\boldsymbol{x}_i$. Given a training set $\mathcal{S} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i), 1 \leq i \leq n\}$.

LE recovers the label distribution $\widehat{\boldsymbol{d}}_i$ of $\boldsymbol{x}_i$ from the logical label $\boldsymbol{y}_i$, converting $\mathcal{S}$ to the label distribution dataset $\widehat{\mathcal{D}} = \left\{(\boldsymbol{x}_i, \widehat{\boldsymbol{d}}_i), 1 \leq i \leq n\right\}$, satisfying $\widehat{d}_{\boldsymbol{x}_i}^{y_{i,j}} \in [0,1]$ and $\sum_{j=1}^{c} \widehat{d}_{\boldsymbol{x}_i}^{y_{i,j}} = 1$, while making the prediction label distribu-

tion dataset $\widehat{\mathcal{D}}$ as close as possible to the (unknown) true label distribution dataset $\mathcal{D} = \{(\boldsymbol{x}_i, \boldsymbol{d}_i)|1 \leq i \leq n\}$.

The proposed CVMG is illustrated in Figure 1.

3.2 Cross-View Fusion via Attention Mechanism

Logical labels and features essentially describe the same instance from complementary perspectives. Logical labels provide significant information about the relevance of the label to the instance, while features offer a comprehensive description of the instance. In label enhancement, by integrating information from both perspectives, the unique characteristics of each can be better utilized, leading to more accurate label distributions. Therefore, we introduce a cross-attention mechanism to fuse their complementary information.

The core idea behind attention mechanisms in deep learning is to dynamically assign different weights to different parts of the input, allowing the model to determine which parts are most crucial. In our model, the cross-attention mechanism is used to combine the distinct perspectives of logical labels and features. Specifically, the features are treated as queries, while the logical labels are used to generate keys and values:

$$h_i = \text{softmax}\left(\frac{W^Q(x_i) \cdot W^K(y_i)^\top}{\sqrt{d_k}}\right) W^V(y_i), \quad \forall i \in \{1, \ldots, n\}. \tag{1}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ represent the original feature and logical label of each instance, while $\boldsymbol{W}^Q$, $\boldsymbol{W}^K$, and $\boldsymbol{W}^V$ are the trainable weight matrices. $\boldsymbol{h}_i$ is the high-level feature representation generated by the cross-attention mechanism through the calculation of the correlation between features and labels, leading to the final high-level feature representation.

We train the cross-attention module using contrastive loss. Contrastive loss works by maximizing the similarity between samples of the same class and minimizing the similarity between samples of different classes, thereby encouraging the cross-attention module to learn more discriminative feature representations. We calculate the Jaccard similarity between the logical labels and set a threshold to determine whether a pair of samples belongs to the same class. The construction of the contrastive loss is as follows:

$$\mathcal{L}_{\text{con}} = -\frac{1}{n}\sum_{i=1}^{n}\frac{1}{|\mathcal{P}(i)|}\sum_{p\in\mathcal{P}(i)}\log\frac{\exp\left(\frac{\boldsymbol{h}_i\cdot\boldsymbol{h}_p}{\tau}\right)}{\sum_{a\in\mathcal{A}(i)}\exp\left(\frac{\boldsymbol{h}_i\cdot\boldsymbol{h}_a}{\tau}\right)} \tag{2}$$

where n denotes the number of instances, $\mathcal{P}(i)$ is the set of positive samples for h_i, and $\mathcal{A}(i)$ is the set of all contrastive samples for h_i, which includes both positive samples and negative samples from different classes. τ is the temperature parameter to control the softness.

The cross-attention module minimizes the contrastive loss to extract information from the original features that aligns with the logical labels. In this way, a high-level feature representation is generated for each instance, incorporating both logical label and feature information, enabling a more precise description of the instance. The resulting high-level features are then used for subsequent label enhancement tasks, further improving the model's performance.

3.3 Label Propagation Based on Mixed Dependency Graph

In the previous section, we introduced how to obtain the high-level feature representation $\boldsymbol{H} = \left[\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_n\right] \in \mathbb{R}^{q\times n}$ through the attention module. In this section, our goal is to recover the label distribution using the high-level features and logical labels. Two matrices can be constructed: the feature matrix $\boldsymbol{H}$ and the logical label matrix $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n] \in \{0,1\}^{c\times n}$. Using these two matrices as inputs, our objective is to obtain the label distribution matrix $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_n] \in [0,1]^{c\times n}$.

In the label enhancement task, the matrix $\boldsymbol{D}$ should satisfy the following two criteria:

Instance-level label distribution similarity assumption: Specifically, if two instances have similar feature representations, their corresponding label distributions (i.e., the i-th and j-th columns of $\boldsymbol{D}$) should also be similar. This assumption is derived from the smoothness assumption.

Category-level label distribution similarity assumption: Specifically, if two categories are highly correlated, such as having high co-occurrence rates, they are likely to co-occur in many instances. This means that the corresponding row vectors in $\boldsymbol{D}$ should be similar.

We model the constraint information in these criteria by constructing a mixed dependency graph and performing label propagation (LP) on this graph. In the following, we will provide a detailed explanation of how these criteria are integrated into a unified optimization framework.

Instance-level Label Distribution Dependency

At the instance level, we model label distribution dependencies by exploiting the similarity among instances in the high-level feature space. Under the smoothness assumption, the manifold structure of the feature space is expected to be preserved in the label space, such that instances with similar features tend to exhibit similar label distributions.

Traditional Label Propagation-based LE methods construct fully connected graphs, allowing labels to propagate among all instances. However, instances with low feature similarity often lack meaningful semantic relationships, and such weak connections may introduce noise. Therefore, we sparsify the instance-level similarity matrix $\boldsymbol{W}_H$ by preserving only truly similar instance connections, improving both robustness and computational efficiency. In this sparsification strategy, if two instances i and j have k-nearest neighbors, their similarity value $(w_H)_{ij}$ is computed based on their Gaussian RBF kernel similarity as follows, with NNk_H referring to the k nearest neighbor search within nodes of the high-level feature set.

$$(w_H)_{ij} = \begin{cases} \exp\left(-\dfrac{\|\boldsymbol{h}_i - \boldsymbol{h}_j\|_2^2}{2\sigma^2}\right), & \text{if } \boldsymbol{h}_j \in \text{NNk}_H(\boldsymbol{h}_i), \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

where $\sigma > 0$ is the width parameter for similarity calculation which is fixed to be 1 in this paper.

Category-level Label Distribution Dependency

At the category level, we focus on the correlation between categories. Specifically, the higher the correlation between

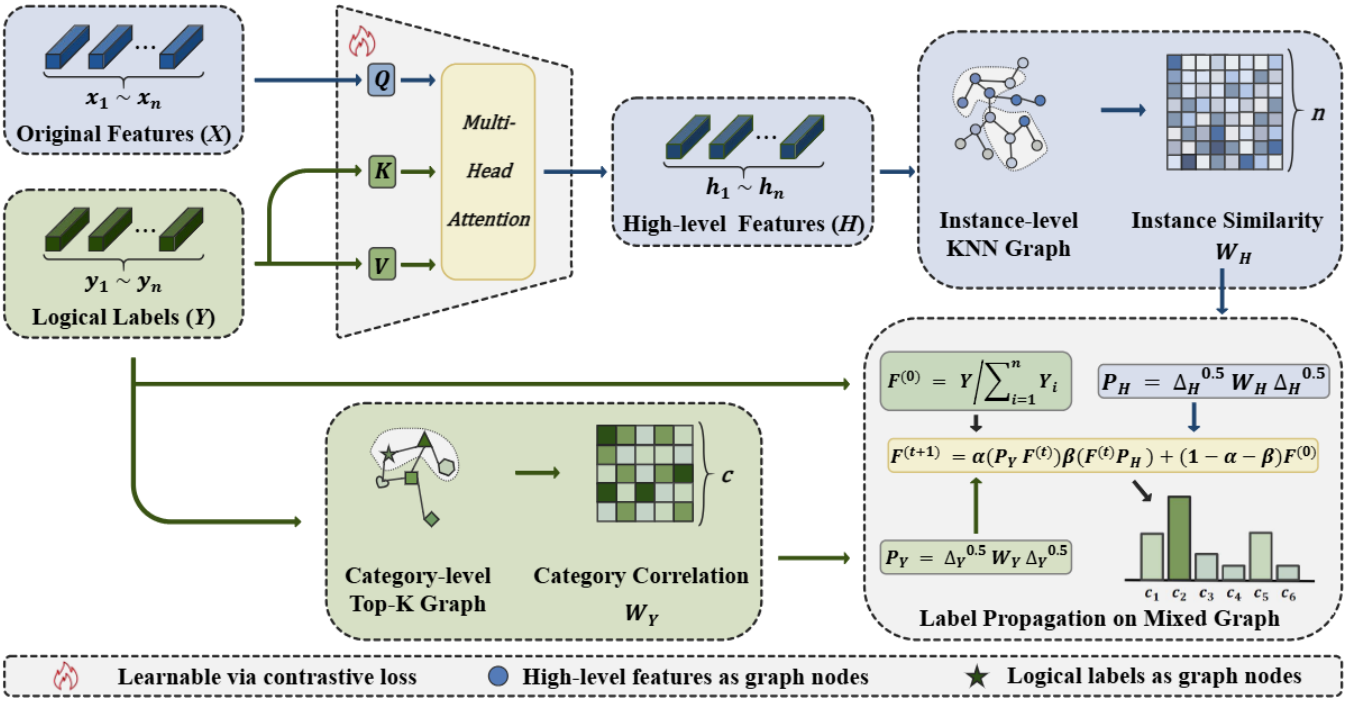

Figure 1: The overall framework of proposed CVMG. It consists of three key components: (1) Cross-view fusion using a cross-attention mechanism to generate high-level representations; (2) Instance-level KNN graph and category-level top-K graph construction to capture relationships; (3) Label propagation on a mixed dependency graph to recover label distributions.

two categories, the closer their semantic or categorical descriptions, and their corresponding label distributions should exhibit higher similarity. For instance, in facial expression datasets, happiness and surprise often appear together on the same face, especially when a person is extremely surprised, typically accompanied by a smile. This category-level similarity can be mined from known logical labels and serves as an effective guide in the label propagation process.

In category-level similarity modeling, we use Jaccard similarity to measure the correlation between categories. We first compute the intersection matrix $\boldsymbol{I} = (i_{ij}) \in \mathbb{N}^{c\times c}$ and the union matrix $\boldsymbol{U} = (u_{ij}) \in \mathbb{N}^{c\times c}$ from the label matrix $\boldsymbol{Y}$, where $\boldsymbol{I}$ represents the co-occurrence of categories and $\boldsymbol{U}$ represents their total occurrence, as defined by the following formulas, where u_i denotes the total count of category i across all instances.

$$\boldsymbol{I} = \boldsymbol{Y} \cdot \boldsymbol{Y}^{\mathrm{T}}, \quad u_i = \sum_{k=1}^{n} \boldsymbol{y}_{i,k}, \quad u_{i,j} = u_i + u_j. \tag{4}$$

Based on these matrices, the Jaccard similarity matrix $\boldsymbol{J} = (j_{ij}) \in \mathbb{R}^{c\times c}$ for each pair of categories is computed as:

$$j_{i,j} = \begin{cases} \dfrac{i_{i,j}}{u_{i,j}}, & \text{if } u_{i,j} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Considering that the similarity between categories may be influenced by noise, especially when the correlation between categories is weak, we retain only the TopK most relevant category pairs. This allows us to focus on the most semantically related pairs, thereby avoiding interference from weakly correlated category pairs. This approach not only reduces the impact of noise but also lowers computational complexity. The final category-level correlation matrix $\boldsymbol{W}_Y$ is shown below:

$$(w_Y)_{i,j} = \begin{cases} j_{i,j}, & \text{if } j \in \text{TopK}(i), \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

LE Using Mixed Dependency Graph

By combining the instance-level similarity matrix $\boldsymbol{W}_H$ and the category-level correlation matrix $\boldsymbol{W}_Y$, we construct a mixed dependency graph that incorporates both instance-level and category-level similarities. Label propagation is then performed on this mixed dependency graph. During the label propagation process, we construct the label propagation matrix $\boldsymbol{P}_H$ based on $\boldsymbol{W}_H$: $\boldsymbol{P}_H = (\boldsymbol{\Delta}_H)^{-1/2}\boldsymbol{W}_H(\boldsymbol{\Delta}_H)^{-1/2}$. Here, $\boldsymbol{\Delta}_H = \mathrm{diag}(\boldsymbol{\delta}_H(1), \ldots, \boldsymbol{\delta}_H(n))$ is a diagonal matrix, with each diagonal element d_i representing the sum of the elements in the i-th row of matrix $\boldsymbol{W}_H$, specifically: $\boldsymbol{\delta}_H(1) = \sum_{j=1}^{n}(w_H)_{ij}$. Similarly, the label propagation matrix $\boldsymbol{P}_Y$ constructed based on the category-level correlation matrix $\boldsymbol{W}_Y$, is given by $\boldsymbol{P}_Y = (\boldsymbol{\Delta}_Y)^{-1/2}\boldsymbol{W}_Y(\boldsymbol{\Delta}_Y)^{-1/2}$, where $\boldsymbol{\Delta}_Y = \mathrm{diag}(\boldsymbol{\delta}_Y(1), \ldots, \boldsymbol{\delta}_Y(c))$.

Let $\boldsymbol{F} = (f_{i,j}) \in \mathbb{R}^{c\times n}$ be a non-negative matrix, where each element $f_{i,j} \geq 0$ is proportional to the importance of the j-th label in the i-th instance. Based on the logical labels, the initial matrix $\boldsymbol{F}^{(0)}$ is instantiated as the normalized logical label matrix $\boldsymbol{Y}$, which serves as the initial "soft labels". The label propagation update rule combines both instance-level and label-level propagation. Specifically, in the t-th iteration,

the matrix $\boldsymbol{F}$ is updated using the label propagation matrices $\boldsymbol{P}_H$ and $\boldsymbol{P}_Y$ as follows:

$$\boldsymbol{F}^{(t+1)} = \alpha(\boldsymbol{P}_Y \boldsymbol{F}^{(t)}) + \beta(\boldsymbol{F}^{(t)} \boldsymbol{P}_H) + (1-\alpha-\beta)\mathbf{F}^{(0)} \quad (7)$$

where α and β are hyperparameters that control the label-level and instance-level propagation. For stability, weights are normalized to form a convex combination. After completing the iterative propagation, the matrix $\boldsymbol{F}^{\star}$ is obtained, and each column of $\boldsymbol{F}^{\star}$ is normalized to estimate the label distribution, as follows:

$$\widehat{d}_{\boldsymbol{x}_i}^{y_{i,j}} = \frac{f_{j,i}^{\star}}{\sum_{k=1}^{c} f_{k,i}^{\star}},\ 1 \le i \le n,\ 1 \le j \le c. \quad (8)$$

4 Experiments

4.1 Datasets

We evaluate our method using 13 real-world multi-label datasets, covering a variety of domains. Facial expression datasets SJAFFE and SBU-3DFE rate images on six basic emotions using a 5-level scale. The Natural Scene dataset consists of 2,000 images of natural scenes, and the Movie dataset contains ratings for 7,755 movies, collected from 478,656 unique users. Additionally, the Yeast datasets (from Yeast-alpha to Yeast-spoem) include data from biological experiments involving gene expression levels of budding yeast at different time points. A summary of these datasets is provided in Table 1.

Dataset	Examples (n)	Features (q)	Labels (c)
SJAFFE	213	243	6
SBU_3DFE	2,500	243	6
Natural_Scene	2,000	294	9
Movie	7,755	1,869	5
Yeast-alpha	2,465	24	18
Yeast-spo	2,465	24	6
Yeast-cdc	2,465	24	15
Yeast-diau	2,465	24	7
Yeast-dtt	2,465	24	4
Yeast-elu	2,465	24	14
Yeast-heat	2,465	24	6
Yeast-cold	2,465	24	4
Yeast_spoem	2,465	24	2

Table 1: Multilabel datasets with ground-truth label distributions used in experimental evaluation.

All datasets offer ground-truth label distributions, enabling label enhancement evaluations. For consistency with previous studies, we apply a binarization strategy to these datasets to ensure uniformity in the evaluation process.

4.2 Evaluation Metrics

We evaluate the recovery performance using six widely used metrics: Chebyshev distance (Cheb), Clark distance (Clark), Canberra metric (Canber), Kullback-Leibler divergence (KL), Cosine coefficient (Cosine) and Intersection similarity (Intersec). The first four are distance metrics, where smaller values indicate better performance, while the last two are similarity metrics, where larger values are preferred. The formulas for these metrics are summarized in Table 4.

4.3 Algorithm Comparison

We compare CVMG with six state-of-the-art algorithms in the field of label enhancement: LP [Li *et al.*, 2015], GLLE [Xu *et al.*, 2019], LESC [Tang *et al.*, 2020], LIB [Zheng *et al.*, 2023], LEVI [Xu *et al.*, 2020], ConLE [Wang *et al.*, 2023].

- LP adapts semi-supervised label propagation to LE, employing graph models to construct the label propagation matrix and generate label distributions.
- GLLE enhances label distributions in the feature space guided by the topological information.
- LESC uses low-rank representation to capture global sample relationships, improving label distribution recovery by predicting implicit correlations.
- LIB infers label distributions from logical labels using variational inference, focusing on relevant label information and ignoring irrelevant data.
- LEVI recovers label distributions through variational inference, incorporating multi-output regression for more accurate label prediction.
- ConLE integrates features and logical labels into a unified space via contrastive learning, generating high-level features for label distribution recovery.

4.4 Recovery Performance

Implementation Details. In the experiments, the cross-attention fusion module uses 8 attention heads to generate enhanced features, with batch size 256 and 50 epochs. Positive pairs are formed by Jaccard similarity (threshold 0.3) on logical labels. The contrastive learning temperature coefficient is set to 0.1, and the optimizer learning rate is 1×10^{-3}. Label propagation is performed on the enhanced features: the instance-level KNN graph connects each instance to its 20 nearest neighbors, and the category-level top-K graph connects each category to two-thirds of the total categories. During label propagation, α is selected from $[0.8, 0.85, \ldots, 1.0]$, and β is selected from $[0.1, 0.12, \ldots, 0.2]$. To ensure stability and convergence, when the sum of α and β exceeds 1, they are scaled to slightly below 1, maintaining positive weight for the initial label term and enabling shrinkage updates. The iterative propagation converges and terminates when $\|\boldsymbol{F}^{(t+1)} - \boldsymbol{F}^{(t)}\|_{\infty} < 10^{-5}$, where $\boldsymbol{F}^{(t)}$ denotes the predicted label distribution matrix at iteration t. For baseline methods, results are generated using their original settings.

Experimental Results. We evaluate our method on 13 real-world datasets and compare it with six benchmark LE algorithms, with detailed results in Table 2. The best performance for each dataset is highlighted in bold, and the last row shows the average ranking across all datasets. Our method consistently outperforms the benchmarks, achieving rankings of 1.38, 1.00, 1.08, 1.27, 1.35, and 1.00 across six metrics, demonstrating superior label distribution recovery. Moreover, the results are largely insensitive to the choice of trade-off hyperparameters, indicating stable performance without precise fine-tuning.

Metrics	Chebyshev ↓							Clark ↓						
Methods	LP	GLLE	LESC	LIB	LEVI	ConLE	Ours	LP	GLLE	LESC	LIB	LEVI	ConLE	Ours
SJAFFE	0.107	0.087	0.069	0.071	0.073	0.069	**0.068**	0.502	0.377	0.276	0.262	0.285	0.269	**0.245**
SBU_3DFE	0.123	0.126	0.122	0.094	0.095	**0.082**	0.091	0.580	0.391	0.378	0.297	0.304	0.297	**0.288**
Natural-Scene	0.275	0.335	0.344	0.314	0.324	0.314	**0.212**	2.482	2.460	2.464	2.452	2.454	2.450	**2.173**
Movie	0.161	0.122	0.121	0.107	0.109	0.097	**0.082**	0.913	0.569	0.564	0.517	0.548	0.463	**0.382**
Yeast-alpha	0.040	0.020	0.015	0.017	**0.013**	**0.013**	**0.013**	1.185	0.337	0.253	0.275	0.219	0.214	**0.203**
Yeast-spo	0.090	0.062	0.060	0.053	0.045	0.042	**0.041**	0.558	0.266	0.255	0.251	0.187	0.177	**0.173**
Yeast-cdc	0.042	0.022	0.019	0.017	0.015	0.014	**0.013**	1.014	0.306	0.251	0.242	0.209	0.178	**0.172**
Yeast-diau	0.099	0.053	0.042	0.049	0.033	0.031	**0.030**	0.788	0.296	0.224	0.273	0.191	0.175	**0.173**
Yeast-dtt	0.128	0.052	0.043	0.034	0.034	0.047	**0.026**	0.499	0.143	0.119	0.092	0.140	0.114	**0.071**
Yeast-elu	0.044	0.023	0.019	0.018	0.012	0.013	**0.013**	0.973	0.295	0.241	0.224	0.222	0.165	**0.153**
Yeast-heat	0.086	0.049	0.046	0.039	0.033	**0.031**	**0.031**	0.568	0.213	0.199	0.165	0.147	0.136	**0.131**
Yeast-spoem	0.163	0.088	0.087	0.069	**0.050**	0.067	0.053	0.272	0.132	0.129	0.104	0.098	0.097	**0.083**
Yeast-cold	0.137	0.066	0.056	0.054	0.051	0.044	**0.038**	0.503	0.176	0.152	0.146	0.140	0.119	**0.104**
Avg.Rank	6.54	6.08	4.73	3.85	3.04	2.38	**1.38**	7.00	5.92	4.77	3.58	3.54	2.19	**1.00**

Metrics	Canberra ↓							Kullback-Leibler ↓						
Methods	LP	GLLE	LESC	LIB	LEVI	ConLE	Ours	LP	GLLE	LESC	LIB	LEVI	ConLE	Ours
SJAFFE	1.064	0.781	0.561	0.531	0.587	0.545	**0.497**	0.077	0.050	0.029	0.027	0.031	0.028	**0.025**
SBU_3DFE	1.245	0.828	0.799	0.611	0.635	0.670	**0.587**	0.105	0.069	0.064	0.041	0.042	**0.039**	0.040
Natural-Scene	6.790	6.851	6.878	6.786	6.801	6.708	**5.608**	1.080	2.063	1.166	0.865	0.928	**0.757**	0.883
Movie	1.720	1.045	1.034	0.920	0.968	0.837	**0.695**	0.127	0.123	0.120	0.077	0.081	0.060	**0.041**
Yeast-alpha	4.544	1.134	0.846	0.893	0.732	0.696	**0.610**	0.121	0.013	0.008	0.009	0.006	0.005	**0.004**
Yeast-spo	2.131	0.548	0.533	0.454	0.372	0.353	**0.348**	0.084	0.029	0.027	0.019	0.014	0.013	**0.011**
Yeast-cdc	3.644	0.959	0.765	0.747	0.642	0.505	**0.499**	0.111	0.014	0.010	0.008	0.006	**0.004**	**0.004**
Yeast-diau	1.748	0.671	0.480	0.621	0.421	**0.365**	0.368	0.127	0.027	0.017	0.022	0.011	0.009	**0.009**
Yeast-dtt	0.941	0.248	0.206	0.158	0.247	0.199	**0.121**	0.103	0.013	0.009	0.005	0.005	0.009	**0.003**
Yeast-elu	3.381	0.902	0.707	0.670	0.674	0.480	**0.441**	0.109	0.013	0.009	0.008	0.005	0.004	**0.004**
Yeast-heat	1.293	0.430	0.401	0.327	0.295	0.268	**0.254**	0.089	0.017	0.016	0.011	0.008	0.007	**0.006**
Yeast-spoem	0.365	0.183	0.263	0.144	0.135	0.137	**0.115**	0.067	0.027	0.027	0.018	0.013	0.014	**0.011**
Yeast-cold	0.924	0.305	0.263	0.250	0.243	0.203	**0.179**	0.103	0.019	0.015	0.012	0.011	0.009	**0.007**
Avg.Rank	6.77	5.92	4.92	3.46	3.54	2.31	**1.08**	6.85	6.04	4.77	3.58	3.27	2.00	**1.27**

Metrics	Cosine ↑							Intersection ↑						
Methods	LP	GLLE	LESC	LIB	LEVI	ConLE	Ours	LP	GLLE	LESC	LIB	LEVI	ConLE	Ours
SJAFFE	0.941	0.958	0.973	0.974	0.970	**0.972**	0.973	0.837	0.872	0.905	0.909	0.899	0.907	**0.911**
SBU_3DFE	0.922	0.927	0.932	**0.958**	0.957	**0.958**	0.957	0.810	0.850	0.855	0.887	0.882	0.886	**0.889**
Natural-Scene	0.850	0.778	0.760	0.742	0.712	0.804	**0.856**	0.451	0.522	0.510	0.459	0.441	0.537	**0.692**
Movie	0.929	0.936	0.937	0.955	0.955	0.964	**0.974**	0.778	0.831	0.833	0.859	0.850	0.871	**0.893**
Yeast-alpha	0.911	0.987	0.992	0.992	0.995	0.995	**0.996**	0.774	0.938	0.953	0.951	0.960	0.961	**0.967**
Yeast-spo	0.939	0.974	0.975	0.982	0.988	**0.989**	**0.989**	0.819	0.909	0.912	0.925	0.940	0.942	**0.944**
Yeast-cdc	0.916	0.987	0.991	0.992	0.994	0.995	**0.996**	0.779	0.937	0.930	0.951	0.958	0.966	**0.968**
Yeast-diau	0.915	0.975	0.985	0.979	**0.990**	0.991	0.992	0.788	0.906	0.933	0.915	0.942	0.949	**0.950**
Yeast-dtt	0.921	0.988	0.992	0.995	0.995	0.992	**0.997**	0.786	0.939	0.949	0.961	0.958	0.950	**0.970**
Yeast-elu	0.918	0.987	0.991	0.992	0.996	0.996	**0.997**	0.782	0.936	0.949	0.952	0.959	0.966	**0.969**
Yeast-heat	0.932	0.984	0.986	0.990	0.992	0.993	**0.994**	0.808	0.929	0.934	0.946	0.952	0.956	**0.959**
Yeast-spoem	0.950	0.978	0.978	0.985	0.990	0.989	**0.991**	0.837	0.912	0.913	0.931	0.937	0.937	**0.946**
Yeast-cold	0.925	0.982	0.986	0.988	0.990	0.991	**0.994**	0.794	0.924	0.935	0.938	0.940	0.950	**0.956**
Avg.Rank	6.62	5.65	4.62	3.69	3.35	2.50	**1.35**	6.92	5.69	4.77	3.69	3.35	2.35	**1.00**

Table 2: Performance comparison of different methods across multiple datasets and metrics. The best results are highlighted in bold. ↓ indicates lower is better, ↑ indicates higher is better.

Ablation Studies. The proposed method consists of two main components: cross-view fusion via an attention mechanism and label propagation based on a mixed dependency graph. These two components are designed to address complementary aspects of the label enhancement problem. To better understand the contribution of each component and their respective roles in the overall framework, we conducted an ablation study by selectively removing or modifying individual modules.

Therefore, we first removed the part of CVMG responsible

Dataset	Cheb ↓	Clark ↓	KL ↓	Cos ↑	Intersec ↑
	0.013	**0.203**	**0.004**	**0.996**	**0.967**
Yeast-alpha	0.014	0.222	0.005	0.994	0.962
	0.015	0.247	0.006	0.993	0.957
	0.013	**0.172**	**0.004**	**0.996**	**0.968**
Yeast-cdc	0.015	0.195	0.005	0.994	0.962
	0.015	0.217	0.006	0.993	0.957
	0.038	**0.104**	**0.007**	**0.994**	**0.956**
Yeast-cold	0.043	0.116	0.008	0.992	0.950
	0.054	0.137	0.011	0.989	0.940
	0.030	**0.173**	**0.009**	**0.992**	**0.950**
Yeast-diau	0.035	0.200	0.012	0.988	0.940
	0.047	0.261	0.019	0.981	0.917
	0.026	**0.071**	**0.003**	**0.997**	**0.969**
Yeast-dtt	0.029	0.079	0.004	0.996	0.966
	0.034	0.095	0.005	0.994	0.959
	0.013	**0.153**	**0.004**	**0.997**	**0.965**
Yeast-elu	0.014	0.177	0.005	0.995	0.963
	0.016	0.200	0.006	0.994	0.957
	0.031	**0.131**	**0.006**	**0.994**	**0.959**
Yeast-heat	**0.031**	0.135	0.007	0.993	0.956
	0.033	0.138	0.007	0.992	0.955
	0.041	**0.173**	**0.011**	**0.989**	**0.944**
Yeast-spo	0.044	0.187	0.013	0.987	0.940
	0.045	0.191	0.014	0.987	0.937
	0.053	**0.083**	**0.011**	**0.991**	**0.946**
Yeast-spoem	0.054	0.087	0.012	0.990	0.943
	0.067	0.103	0.013	0.988	0.933
	0.091	**0.288**	**0.040**	**0.957**	**0.889**
SBU_3DFE	0.098	0.326	0.048	0.951	0.875
	0.093	0.339	0.047	0.952	0.874
	0.068	**0.245**	**0.025**	**0.973**	**0.911**
SJAFFE	0.077	0.275	0.030	0.969	0.903
	0.072	0.268	0.028	0.971	0.905
	0.212	**2.173**	**0.883**	**0.856**	**0.692**
Natural Scene	0.216	2.271	0.887	0.846	0.690
	0.306	2.446	0.826	0.766	0.474
	0.082	**0.382**	**0.041**	**0.974**	**0.893**
Movie	0.092	0.432	0.055	0.968	0.881
	0.098	0.509	0.067	0.962	0.865

Table 3: Ablation study comparing CVMG (full model, the top row) with two variants: MGLE (without cross-attention feature generation, the middle row) and CVLP (without mixed dependency graph label propagation, the bottom row). The best results are highlighted in bold.

for cross-view fusion, resulting in a comparative algorithm MGLE. Secondly, we removed the label propagation based on the mixed dependency graph and reduced it to label propagation on a fully connected graph based solely on feature similarity, resulting in the algorithm CVLP.

The results are shown in Table IV, comparing the complete CVMG framework (top row) with two ablation versions: MGLE (middle row) and CVLP (bottom row). This behavior is expected, as label enhancement is inherently influenced by both relationships among instances and dependencies among categories, and removing either component inevitably weakens the model's ability to recover accurate label distributions. The results show that removing the cross-attention feature generation leads to a significant performance drop. This degradation can be attributed to the absence of high-level features, causing the model to struggle with capturing the complex relationships between logical labels and features. High-level features play a crucial role in improving the accuracy of label distribution recovery. Similarly, removing the mixed dependency graph-based label propagation causes a sharp decline in the efficiency of label information propagation. Without this mechanism, label propagation becomes limited and fails to fully exploit the relationships between instances and categories. For datasets with very few labels (e.g., Yeast-spoem, c = 2), the category-level Top-K graph is simple and contributes modestly, yet still provides complementary guidance as a lightweight correlation regularizer, while the instance-level KNN graph dominates propagation. Overall, the ablation study highlights the critical role of cross-view fusion via attention mechanism and mixed dependency graph-based label propagation in label enhancement.

Measure	Formula
Chebyshev ↓	$\mathrm{Dis}_1(\boldsymbol{D}, \widehat{\boldsymbol{D}}) = \max_j \lvert d_j - \widehat{d}_j \rvert$
Clark ↓	$\mathrm{Dis}_2(\boldsymbol{D}, \widehat{\boldsymbol{D}}) = \sqrt{\sum_{j=1}^{c} \frac{(d_j - \widehat{d}_j)^2}{(d_j + \widehat{d}_j)^2}}$
Canberra ↓	$\mathrm{Dis}_3(\boldsymbol{D}, \widehat{\boldsymbol{D}}) = \sum_{j=1}^{c} \frac{\lvert d_j - \widehat{d}_j \rvert^2}{d_j + \widehat{d}_j}$
Kullback-Leibler ↓	$\mathrm{Dis}_4(\boldsymbol{D}, \widehat{\boldsymbol{D}}) = \sum_{j=1}^{c} d_j \ln \frac{d_j}{\widehat{d}_j}$
Cosine ↑	$\mathrm{Sim}_1(\boldsymbol{D}, \widehat{\boldsymbol{D}}) = \frac{\sum_{j=1}^{c} d_j \widehat{d}_j}{\sqrt{\sum_{j=1}^{c} d_j^2} \sqrt{\sum_{j=1}^{c} \widehat{d}_j^2}}$
Intersection ↑	$\mathrm{Sim}_2(\boldsymbol{D}, \widehat{\boldsymbol{D}}) = \sum_{j=1}^{c} \min(d_j, \widehat{d}_j)$

Table 4: Evaluation measures for label distribution comparison.

5 Conclusion

In this work, we propose CVMG (Cross-View Fusion and Mixed Graph-based Label Enhancement), a novel method for the label enhancement problem. By integrating logical labels and features through a cross-attention mechanism, CVMG is able to capture complementary information from different views and generate more informative high-level features. Based on these representations, the proposed method further constructs a mixed dependency graph that jointly models instance-level similarities and category-level dependencies, enabling effective label propagation. Experimental results on 13 real-world datasets show that CVMG consistently outperforms existing label enhancement methods.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 62476135, 62406145, and the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province under Grant 24KJB520015.

References

[Bi *et al.*, 2024] Haixia Bi, Honghao Chang, Xiaotian Wang, and Danfeng Hong. Cross-attention-driven adaptive graph relational network for multilabel remote sensing scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024.

[Dosovitskiy *et al.*, 2020] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[El Gayar *et al.*, 2006] Neamat El Gayar, Friedhelm Schwenker, and Günther Palm. A study of the robustness of knn classifiers trained using soft labels. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*, pages 67–80. Springer, 2006.

[Geng, 2016] Xin Geng. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1734–1748, 2016.

[Gibaja and Ventura, 2014] Eva Gibaja and Sebastián Ventura. Multi-label learning: a review of the state of the art and ongoing research. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(6):411–444, 2014.

[Helvig *et al.*, 2024] Kevin Helvig, Baptiste Abeloos, and Pauline Trouvé-Peloux. Caff-dino: Multi-spectral object detection transformers with cross-attention features fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3037–3046, 2024.

[Hou *et al.*, 2016] Peng Hou, Xin Geng, and Min-Ling Zhang. Multi-label manifold learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

[Hua *et al.*, 2020] Yuansheng Hua, Lichao Mou, and Xiao Xiang Zhu. Relation network for multilabel aerial image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7):4558–4572, 2020.

[Jia *et al.*, 2019] Xiuyi Jia, Xiang Zheng, Weiwei Li, Changqing Zhang, and Zechao Li. Facial emotion distribution learning by exploiting low-rank label correlations locally. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 9841–9850, 2019.

[Jiang *et al.*, 2006] Xiufeng Jiang, Zhang Yi, and Jian Cheng Lv. Fuzzy svm with a new fuzzy membership function. *Neural Computing & Applications*, 15(3):268–276, 2006.

[Lanchantin *et al.*, 2021] Jack Lanchantin, Tianlu Wang, Vicente Ordonez, and Yanjun Qi. General multi-label image classification with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16478–16488, 2021.

[Li *et al.*, 2015] Yu-Kun Li, Min-Ling Zhang, and Xin Geng. Leveraging implicit relative labeling-importance information for effective multi-label learning. In *2015 IEEE International Conference on Data Mining*, pages 251–260. IEEE, 2015.

[Li *et al.*, 2024] Chenyu Li, Bing Zhang, Danfeng Hong, Jun Zhou, Gemine Vivone, Shutao Li, and Jocelyn Chanussot. Casformer: Cascaded transformers for fusion-aware computational hyperspectral imaging. *Information Fusion*, 108:102408, 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.

[Moyano *et al.*, 2019] Jose M Moyano, Eva L Gibaja, Krzysztof J Cios, and Sebastián Ventura. An evolutionary approach to build ensembles of multi-label classifiers. *Information Fusion*, 50:168–180, 2019.

[Tan *et al.*, 2022] Xiaowei Tan, Zhifeng Xiao, Jianjun Zhu, Qiao Wan, Kai Wang, and Deren Li. Transformer-driven semantic relation inference for multilabel classification of high-resolution remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:1884–1901, 2022.

[Tang *et al.*, 2020] Haoyu Tang, Jihua Zhu, Qinghai Zheng, Jun Wang, Shanmin Pang, and Zhongyu Li. Label enhancement with sample correlations via low-rank representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5932–5939, 2020.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[Wang *et al.*, 2023] Yifei Wang, Yiyang Zhou, Jihua Zhu, Xinyuan Liu, Wenbiao Yan, and Zhiqiang Tian. Contrastive label enhancement. *arXiv preprint arXiv:2305.09500*, 2023.

[Xu *et al.*, 2019] Ning Xu, Yun-Peng Liu, and Xin Geng. Label enhancement for label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 33(4):1632–1643, 2019.

[Xu *et al.*, 2020] Ning Xu, Jun Shu, Yun-Peng Liu, and Xin Geng. Variational label enhancement. In *International conference on machine learning*, pages 10597–10606. PMLR, 2020.

[Ye *et al.*, 2025] Qianzhi Ye, Jia Zhang, Hanrui Wu, Tianlong Gu, CL Philip Chen, and Jinyi Long. Smle: Semi-supervised multi-label learning with label enhancement. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[Zhang *et al.*, 2022] Shu Zhang, Ran Xu, Caiming Xiong, and Chetan Ramaiah. Use all the labels: A hierarchical multi-label contrastive learning framework. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16660–16669, 2022.

[Zhao *et al.*, 2025] Xingyu Zhao, Yuexuan An, Ning Xu, Lei Qi, and Xin Geng. Interactive fusion label enhancement for multi-label learning. *ACM Transactions on Knowledge Discovery from Data*, 19(7):1–23, 2025.

[Zheng *et al.*, 2023] Qinghai Zheng, Jihua Zhu, and Haoyu Tang. Label information bottleneck for label enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7497–7506, 2023.

[Zhu *et al.*, 2021] Wenfang Zhu, Xiuyi Jia, and Weiwei Li. Privileged label enhancement with multi-label learning. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 2376–2382, 2021.