Dual Graph Disambiguation for Multi-Instance Partial-Label Learning

Zhen Zhu[1], Kai Tang[2], Songhe Feng[3*], Yixuan Tang[5*], Haobo Wang[2], Gengyu Lyu[4], Cheng Peng[2], Yining Sun[6]

[1]Beijing Jiaotong University,
[2]Zhejiang University,
[3]Tangshan Research Institute of Beijing Jiaotong University,
[4]Beijing University of Technology,
[5]National University of Singapore,
[6]Yixiaomo.Inc

hzzhuzhen@yeah.net, kai.t@zju.edu.cn, shfeng@bjtu.edu.cn, yixuan@comp.nus.edu.sg, wanghaobo@zju.edu.cn, lyugengyu@gmail.com, chengchng@zju.edu.cn, 2114944317@qq.com

Abstract

In multi-instance partial label learning (MIPL), each sample is a bag of multiple instances linked to a candidate label set containing one true and multiple false labels, yielding inexact supervision in both instance features and label space. However, existing works adopt decoupled approaches that focus exclusively on either instance-level feature fusion or label-level disambiguation, failing to fully exploit the intrinsic dependencies between these two spaces. To overcome this limitation, graph-based methods are widely recognized as a powerful paradigm in weakly supervised learning, yet their success hinges on reliable features—precisely what MIPL lacks due to instance-level noise. To bridge this gap, we propose DualG, a novel framework that simultaneously addresses feature learning and label disambiguation through dual-level graph propagation. Specifically, we construct dual relevance graphs at both the bag and instance levels. At the bag level, we build a similarity graph based on fused feature representations; at the instance level, we employ attention scores to filter out irrelevant instances and construct a reliable instance-level relevance graph. These complementary graphs enable our joint label disambiguation framework to simultaneously address inexact supervision signals in both instance space and label space. Experimental results on five benchmark datasets demonstrate that DualG outperforms existing MIPL and partial label learning methods, validating its effectiveness and superiority. Source code is available at

Code — https://github.com/hzzhuzhen/DualG

1 Introduction

Deep learning techniques have achieved remarkable success by leveraging large-scale labeled training data (Taye 2023; Dong, Wang, and Abbas 2021). However, in real-world applications, supervision signals often suffer from inexactness, where feature-label alignment is coarse-grained. This manifests in two complementary aspects: *instance-level inexactness*—where multiple instances form a single bag (Multi-Instance Learning, MIL) (Ilse, Tomczak, and Welling 2018; Cui et al. 2023)—and *label-level inexactness*—where a sample associates with multiple candidate labels (Partial Label Learning, PLL) (Hüllermeier and Beringer 2005). Critically, these two challenges frequently co-occur, giving rise to Multi-Instance Partial-Label Learning (MIPL). Medical imaging exemplifies this dual inexact supervision (Lu et al. 2021; Campanella et al. 2019): each diagnostic image (a "bag" of regions) typically receives a set of plausible diagnoses rather than a definitive label to reduce annotation burden. Despite its practical significance in weakly supervised learning (Zhou 2018; Han et al. 2018; Berthelot et al. 2019; Ishida et al. 2017; Wei et al. 2022; Xie and Huang 2021; Lyu et al. 2021), existing MIPL methods (Tang, Zhang, and Zhang 2024b, 2023) adopt decoupled approaches that focus exclusively on either instance-level feature fusion or label-level disambiguation, failing to exploit intrinsic dependencies between these two spaces and consequently suffering from error propagation.

Fortunately, graph-based disambiguation techniques present a promising direction, having widely recognized as a powerful paradigm in weakly supervised learning through their fundamental assumption that neighboring instances in feature space share consistent labels. However, their direct application to MIPL is fundamentally challenged by the fundamental conflict between required feature reliability and inherent instance ambiguity—where noisy instances distort neighborhood structures and compromise graph construction. This critical conflict explains why graph-based approaches remain largely unexplored in MIPL despite their success elsewhere, directly motivating our investigation into the feasibility of graph-learning in MIPL.

To bridge this gap, we propose a novel framework named **Dual G**raph Disambiguation for Multi-Instance Partial-Label Learning (dubbed DualG), which is motivated by the demonstrated success of graph learning techniques in the field of label disambiguation (Wang, Li, and Zhang 2019). DualG extends this powerful graph-learning paradigm to the more complex multi-instance setting of MIPL, tightly integrating representation learning with label disambiguation. Specifically, DualG constructs dual relevance graphs at both the bag and instance levels. At the bag level, a similarity

* Corresponding authors.

graph is built based on fused feature representations, effectively capturing global bag-to-bag relationships. Concurrently, at the instance level, attention scores are employed to globally filter out irrelevant instances, and an instance-level reliable relevance graph is constructed, modeling fine-grained instance-to-instance relationships within and potentially across bags. These complementary graphs enable a joint label disambiguation framework to simultaneously address inexact supervision signals in both the instance space and the label space.

We comprehensively evaluate DualG on various MIPL benchmarks and real-world datasets, which demonstrates that DualG outperforms state-of-the-art MIPL and PLL methods, validating its effectiveness and superiority in handling dual inexact supervision. Specifically, across 5 different benchmarks with 16 different settings, our method achieves average absolute performance gains of **1.63%** in ACC with the same or less computing resource.

2 Related Work

Multi-Instance Learning. Originating from drug activity prediction (Dietterich, Lathrop, and Lozano-Pérez 1997), multi-instance learning (MIL) has been widely applied in various domains, including text classification (Zhang et al. 2025; Zhou, Sun, and Li 2009) and image annotation (Zheng, Jiang, and Yao 2024; Jin et al. 2025; Zhang 2021). Modern deep MIL methods predominantly rely on attention mechanisms (Wang et al. 2022; Tan et al. 2023; Ilse, Tomczak, and Welling 2018). Although these methods achieve promising results under the assumption of precise bag-level labels, they face challenges when learning with ambiguous bag-level annotations.

Partial-Label Learning. Recent advances in Partial Label Learning (PLL) heavily rely on deep learning techniques (Wang et al. 2025). Among them, recognition-based strategies aim to identify the true label from the candidate label set (Lv et al. 2024; Tian et al. 2024; Yao et al. 2020). In contrast, average-based strategies assume equal contribution from all candidate labels and treat candidate label averages as predictions (Lv et al. 2023). Another line of research focuses on generative modeling, explicitly modeling partial label-ground-truth relationships to obtain theoretically grounded loss functions (Qiao, Xu, and Geng 2023; Feng et al. 2020; Feng and An 2019). Graph-based methods exploit the local manifold structure in the feature space to assist in label disambiguation, showing strong performance in PLL (Feng and An 2018; Zhang and Yu 2015; Zhang, Zhou, and Liu 2016). However, these approaches generally lack the capability to handle multi-instance bags, limiting their applicability in complex weakly supervised scenarios.

Multi-Instance Partial-Label Learning. Due to conventional MIL and PLL addressing only one aspect of inexact supervision (Tang et al. 2024; Yang, Tang, and Zhang 2024, 2025; Tang, Zhang, and Zhang 2024a), MIPL has recently attracted significant attention for handling dual forms of supervision uncertainty. Current approaches like MIPLGP (Tang, Zhang, and Zhang 2024b) introduce a Gaussian process-based algorithm, and DEMIPL (Tang, Zhang, and Zhang 2023) aggregates each multi-instance bag into a feature representation with a momentum-based disambiguation strategy. Both of them primarily rely on mapping instances to candidate label sets, yet fail to fully exploit the intrinsic relationships between label spaces and bag-level feature representations. Although graph-based disambiguation shows promise for conventional tasks, its application in MIPL is hindered by instance ambiguity. We demonstrate that dual-level graph modeling effectively resolves dual inexact supervision while maintaining robustness to instance-level ambiguity.

3 Methodology

3.1 Problem Formulation

In this study, we define a MIPL training dataset as $D = \{(\boldsymbol{X}_i, S_i) \mid 1 \leq i \leq m\}$, where D contains m multi-instance bags and their corresponding candidate label sets.

- $\boldsymbol{X}_i = \{\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, ..., \boldsymbol{x}_{i,n_i}\}$ contains n_i instances in d-dimensional space, i.e., $x_{i,j} \in \mathbb{R}^d$
- $S_i \subset Y$ is the associated candidate label set, where $Y = \{1, 2, 3, ..., K\}$ is the complete set of K classes. This candidate set S_i is guaranteed to contain exactly one ground-truth label, denoted as y_i^*, along with other false positive labels.

Instance-Level feature extraction. For a given multi-instance bag $\boldsymbol{X}_i = \{\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \ldots, \boldsymbol{x}_{i,n_i}\}$ containing n_i instances, the instance-level feature representation $\boldsymbol{H}_i = \{\boldsymbol{h}_{i,1}, \boldsymbol{h}_{i,2}, \ldots, \boldsymbol{h}_{i,n_i}\}$ is learned using a feature extractor $\phi(\cdot)$ that $\boldsymbol{h}_{i,j} = \phi(\boldsymbol{x}_{i,j})$. All features are mapped into a shared embedding space of dimensionality l.

Bag-Level Fusion Mechanism. Regarding bag-level fusion, we adopt the approach from (Tang, Zhang, and Zhang 2023) by employing an attention mechanism to determine instance contributions and derive aggregated feature representations. Specifically, we first calculate the relevance of each instance to all classes through instance-level attention.

$$\boldsymbol{\delta}_{i,j} = \tanh\left(\boldsymbol{W}_v^\top \boldsymbol{h}_{i,j} + \boldsymbol{b}_v\right) \odot \text{sigm}\left(\boldsymbol{W}_u^\top \boldsymbol{h}_{i,j} + \boldsymbol{b}_u\right) \quad (1)$$

Subsequently, these relevance scores are transformed into contribution weights for constructing the bag-level feature representation: $a_{i,j} = \frac{1}{1+\exp\{-\boldsymbol{W}^\top \boldsymbol{\delta}_{i,j}\}}$, where $\boldsymbol{W}^\top \in \mathbb{R}^{1\times K}$, $\boldsymbol{W}_v^\top$, $\boldsymbol{W}_u^\top \in \mathbb{R}^{K\times d'}$, and $\boldsymbol{b}_v$, $\boldsymbol{b}_u \in \mathbb{R}^K$ are parameters of linear models. At the end, we obtain the bag-level representation $\boldsymbol{z}_i = \frac{1}{\sum_{j=1}^{n_i} i,j} \sum_{j=1}^{n_i} a_{i,j} \boldsymbol{h}_{i,j}$.

3.2 Dual Graph Construction for Pseudo-Label

Graph-based disambiguation strategies have demonstrated effectiveness in label disambiguation tasks (Wang et al. 2019; Li, Lyu, and Feng 2021; Zhang and Fang 2020). However, these methods cannot be directly applied to MIPL tasks due to dual-level ambiguities requiring simultaneous resolution. To address this challenge, we propose dual relevance graphs at both bag and instance levels. This approach explicitly models dependencies between instances within bags

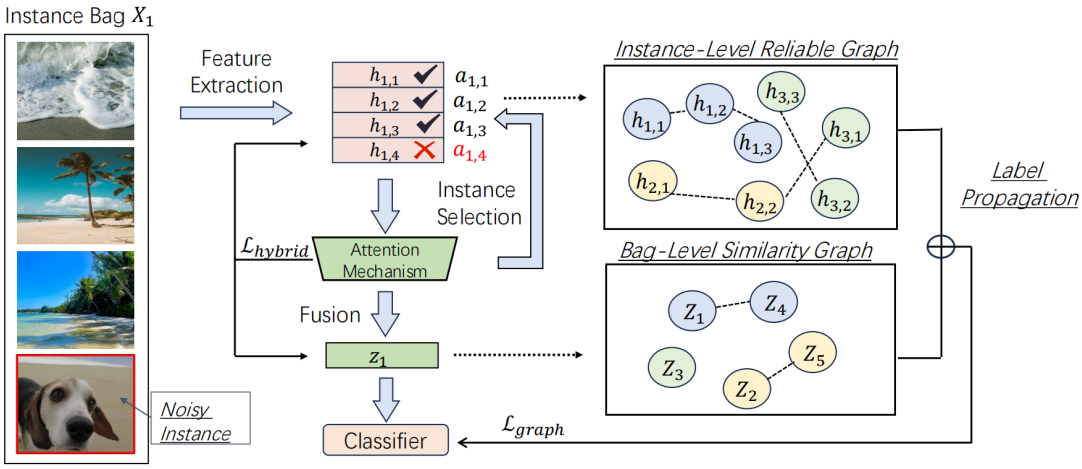


Figure 1: The framework of DualG.

and across bags, enabling precise resolution of dual inexact supervision signals to generate reliable pseudo-labels for model updating.

Bag-level Similarity Graph. To obtain the bag-level similarity graph $\boldsymbol{G}^{\text{bag}}$, we adopt *single kernel* (Iscen et al. 2017) as a similarity measure and compute a sparse instance affinity matrix $\boldsymbol{U} \in \mathbb{R}^{n\times n}$, where each element is defined as:

$$\boldsymbol{U}_{ij} = \begin{cases} \max\left(\boldsymbol{z}_i^\top \boldsymbol{z}_j, 0\right)^\rho, & \text{if } i \neq j \text{ and } j \in \mathcal{N}_K(\boldsymbol{z}_i), \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $\mathbf{z}_i$ and $\boldsymbol{z}_j$ denote the features of bags i and j, respectively. $\mathcal{N}(\boldsymbol{z}_i)$ denotes the set of K-nearest neighbors of $\mathbf{z}_i$. $\rho > 0$ controls the weight of similarity. The bag-level feature graph is obtained by symmetrization: $G^{\text{bag}} = \boldsymbol{U} + \boldsymbol{U}^\top$.

Instance-level Relevance Graph. However, constructing graph solely based on bag-level features tends to bring about suboptimal label disambiguation, as bags inevitably contain noisy instances in multi-instance learning. As a complementary strategy, we globally filter out low-confidence instances using attention scores, thereby modeling reliable instance-level relationships both within individual bags and across potentially multiple bags. Specifically, we filter out low-confidence sample as:

$$\mathcal{S}_{\text{clean}} = \{(i,j) | \text{Rank}_{(i,j)} \leq \lceil r \cdot \sum_{b=1}^{m} n_b \rceil\}, \tag{3}$$

where $\text{Rank}_{(i,j)}$ denotes the descending-order ranking of attention scores across all instances and $\lceil\cdot\rceil$ is the ceiling function. In practical applications, we set the hyperparameter r to 0.75. For $(i,j), (i'j') \in \mathcal{S}_{\text{clean}}$, we calculate similarity between key instances similarly:

$$\boldsymbol{V}_{(i,j)(i',j')} = \max\left(\boldsymbol{h}_{i,j}^\top \boldsymbol{h}_{i',j'}, 0\right)^\sigma \cdot \mathbf{1}_{(i,j)\neq(i',j')}, \tag{4}$$

where $\sigma > 0$ controls the weight of similarity. Subsequently, we obtain the instance-level graph $\boldsymbol{G}^{\text{ins}} = \boldsymbol{V} + \boldsymbol{V}^T$ through symmetrization.

Graph-Induced Label Confidence Estimation. We formulate label confidence estimation as an optimization problem that simultaneously preserves graph structure consistency and respects MIPL constraints. The objective function comprises four complementary components:

First, a **label consistency term** ensures alignment with current pseudo-label estimates: $\left\|\hat{\boldsymbol{Y}} - \boldsymbol{Y}'\right\|_F^2$, where $\hat{\boldsymbol{Y}} \in \mathbb{R}^{n\times k}$ is the predicted label matrix, $\boldsymbol{Y}'$ represents current pseudo-label estimates and more details can be found in following section, and $\|\cdot\|_F$ denotes the Frobenius norm.

Second, we incorporate **dual-level graph regularization** to enforce smoothness across both bag and instance structures. For the bag-level graph $\boldsymbol{G}^{\text{bag}}$ with Laplacian $\boldsymbol{L}_{\text{bag}} = \boldsymbol{I} - \boldsymbol{D}_{\text{bag}}^{-\frac{1}{2}} \boldsymbol{G}^{\text{bag}} \boldsymbol{D}_{\text{bag}}^{-\frac{1}{2}}$ (where $\boldsymbol{D}_{\text{bag}} = \text{diag}\left[d_1^{\text{bag}}, \ldots, d_n^{\text{bag}}\right]$ is the degree matrix), we apply Laplacian regularization:

$$R_{\text{bag}}(\hat{\boldsymbol{Y}}) = \text{tr}(\hat{\boldsymbol{Y}}^\top \boldsymbol{L}_{\text{bag}} \hat{\boldsymbol{Y}}), \tag{5}$$

which penalizes label discrepancies between similar bags. Similarly, we also calculate Laplacian regularization on $\boldsymbol{G}^{ins}$: $R_{\text{ins}}(\hat{\boldsymbol{Y}}) = \text{tr}(\hat{\boldsymbol{Y}}^\top \boldsymbol{L}_{\text{ins}} \hat{\boldsymbol{Y}})$.

Combining these components, our complete objective function becomes:

$$J(\hat{\boldsymbol{Y}}, \boldsymbol{Y'}) = \eta \left\|\hat{\boldsymbol{Y}} - \boldsymbol{Y'}\right\|_F^2 + \beta R_{\text{bag}}(\hat{\boldsymbol{Y}}) + \gamma R_{\text{ins}}(\hat{\boldsymbol{Y}}), \tag{6}$$

where η, β, γ are hyperparameters balancing the contributions of each term. This formulation jointly optimizes for label consistency, graph-structural smoothness, and MIPL-specific constraints.

Pseudo-Label Update. Based on the above learning objective, we employ label propagation on dual-level graphs to generate reliable pseudo-labels for model updates. Specifically, we maintain a smooth pseudo-label matrix $\boldsymbol{P}$, which minimizes propagation error while approximating the model output $\hat{\boldsymbol{Y}}$. During label propagation, information from current pseudo-label $\boldsymbol{Y'}$ and model predictions $\hat{\boldsymbol{Y}}$ is diffused

into an intermediate state. We use the Frobenius norm to ensure the closeness of $\boldsymbol{P}$ and $\hat{\boldsymbol{Y}}$ for label smoothness , so that $\boldsymbol{P}$ is estimated by the following objective:

$$\min_{\boldsymbol{P}} \mathcal{L}_{\text{LP}}(\boldsymbol{P}) = \frac{1}{2}\|\boldsymbol{P} - \hat{\boldsymbol{Y}}\|_F^2 + J(\boldsymbol{P}, \boldsymbol{Y'}). \tag{7}$$

The gradient of this objective for $\boldsymbol{P}$ can be derived as:

$$\frac{\partial \mathcal{L}_{\text{LP}}}{\partial \boldsymbol{P}} = (1+2\eta)\boldsymbol{P} + 2\alpha L_{\text{bag}}\boldsymbol{P} + 2\beta L_{\text{ins}}\boldsymbol{P} - (\hat{\boldsymbol{Y}} + 2\eta \boldsymbol{Y'}). \tag{8}$$

Therefore, pseudo-label generation is solved through gradient descent updates:

$$\boldsymbol{P} = \boldsymbol{P} - r_l \frac{\partial \mathcal{L}_{LP}(\boldsymbol{P})}{\partial \boldsymbol{P}}, \tag{9}$$

where each pseudo-label update involves t round gradient descent steps with learning rate $r_l > 0$, ensuring empirical convergence to stable label distributions through progressive sharpening of candidate label assignments. Finally, combined with binary mask of candidate label set $\boldsymbol{S}$, we obtain the pseudo-label estimates: $\boldsymbol{Y}' = \frac{\boldsymbol{P} \odot \boldsymbol{S}}{|\boldsymbol{P} \odot \boldsymbol{S}|}$.

3.3 Model Updating

Building upon the pseudo-labels obtained in the previous section, we formulate a joint loss function that addresses both feature learning and label disambiguation to effectively address the dual inexact supervision inherent in MIPL.

Loss for Feature Representation and Attention Estimation. Feature extraction and representation fusion in MIL have been extensively investigated, where researchers have explored complementary loss functions at the instance-level(Shi et al. 2020) and bag-level(Tang, Zhang, and Zhang 2023). Building upon dual-level graph construction, we integrate these hierarchical feature learning losses and incorporate pseudo-label information derived from graph propagation. Corresponding to the dual-graph construction, we combine the feature learning losses at these two levels and utilize the pseudo-label information obtained through graph learning as follows:

$$\mathcal{L}_{\text{feature}} = \frac{1}{m}\sum_{i=1}^{m} H(\boldsymbol{a}_i) - \frac{1}{m}\sum_{i=1}^{m}\frac{1}{n_i}\sum_{j=1}^{n_i} \log(\boldsymbol{\delta}_{i,j}^t). \tag{10}$$

The first term is the bag-level attention loss, where $H(\boldsymbol{a}_i) = -\sum_{j=1}^{n_i} a_{i,j} \log a_{i,j}$. The second term is the instance-level fusion loss, where $t = \arg\max_{1 \le j \le n_i} \boldsymbol{Y}'_{i,j}$ denotes the label category with the highest probability in pseudo-label $\boldsymbol{Y}'$.

Loss for label disambiguation. After obtaining the pseudo-labels after label propagation, we update the model parameters to fit the pseudo-labels through Probabilistic Mean Squared Error (PMSE) loss:

$$\mathcal{L}_{\text{graph}}(\hat{\boldsymbol{Y}}, \boldsymbol{Y}') = \frac{1}{m}\sum_{i=1}^{m} \left\| \sigma(\hat{\boldsymbol{Y}}_i) - \boldsymbol{Y}'_i \right\|_2^2, \tag{11}$$

Algorithm 1: Training Procedure of DualG

Input: Training dataset $\mathcal{D} = \{(\boldsymbol{X}_i, S_i)\}_{i=1}^m$, learning rate r, number of epochs T, The dual graph is recomputed every N epochs

Initialize network parameters θ uniformly
Precompute initial pseudo-label matrix $\boldsymbol{P}$ via warm-up
for t = 1 to T **do**
  for each bag in $\mathcal{D}$ **do**
    Obtain instance-level features $\boldsymbol{h}_{i,j}$ and bag-level feature $\boldsymbol{z}_i$
    Compute predicted label $\hat{\boldsymbol{Y}} = f(\boldsymbol{Z}; \theta)$
    Compute total loss $\mathcal{L}$ based On Equation(6,11-13)

$$\mathcal{L} = \mathcal{L}_{\text{ce}} + \mathcal{L}_{\text{feature}} + \lambda_a \mathcal{L}_{\text{graph}} + \lambda_b \mathcal{L}_{\text{sp}}$$

    Update model:$\theta \leftarrow \theta - r\nabla_\theta \mathcal{L}$
  end for
  if $t \bmod N = 1$ **then**
    Construct dual-level graph $\boldsymbol{G}^{\text{bag}}$ and $\boldsymbol{G}^{\text{ins}}$ base on Equation(3-7)
    $\mathcal{S}_{\text{clean}} \leftarrow \{(i,j) | \text{Rank}_{(i,j)} \le \lceil r \cdot \sum_{b=1}^m n_b \rceil\}$
    $\boldsymbol{L}_{\text{bag}} \leftarrow \boldsymbol{I} - \boldsymbol{D}_{\text{bag}}^{-\frac{1}{2}} \boldsymbol{G}^{\text{bag}} \boldsymbol{D}_{\text{bag}}^{-\frac{1}{2}}$
    $\boldsymbol{L}_{\text{ins}} \leftarrow \boldsymbol{I} - \boldsymbol{D}_{\text{ins}}^{-\frac{1}{2}} \boldsymbol{G}^{\text{ins}} \boldsymbol{D}_{\text{ins}}^{-\frac{1}{2}}$
    Update pseudo label $\boldsymbol{Y}'$ base on Equation(8-10)
    $\boldsymbol{P} = \boldsymbol{P} - r_l \frac{\partial \mathcal{L}_{LP}(\boldsymbol{P})}{\partial \boldsymbol{P}}$
    $\boldsymbol{Y}' \leftarrow \frac{\boldsymbol{P} \odot \boldsymbol{S}}{|\boldsymbol{P} \odot \boldsymbol{S}|}$
  end if
end for

Output: Trained model parameter θ

where $\sigma(\cdot)$ denotes the element-wise logistic sigmoid function. Additionally, to address the fundamental MIPL constraint that only one candidate label is true per bag, we introduce a **sharpening regularization**:

$$\mathcal{L}_{\text{sp}} = \frac{1}{m}\sum_{i=1}^{m} |\hat{\boldsymbol{Y}}_i \odot \boldsymbol{Y}'_i|, \tag{12}$$

where $\hat{\boldsymbol{Y}}_i$ is the predicted probability vector for bag i and and $\odot$ denotes element-wise multiplication. Finally, we introduce cross-entropy loss $\mathcal{L}_{\text{ce}} = \frac{1}{m}\sum_{i=i}^{m} \text{CE}(\hat{\boldsymbol{Y}}_i, \boldsymbol{Y}'_i)$ and total loss $\mathcal{L}$ as formalized in Eq.13:

$$\mathcal{L} = \mathcal{L}_{\text{ce}} + \mathcal{L}_{\text{feature}} + \lambda_a \mathcal{L}_{\text{graph}} + \lambda_b \mathcal{L}_{\text{sp}}, \tag{13}$$

where λ_a, λ_b are balancing hyperparameters for the losses.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate our method on four benchmark MIPL datasets (Tang, Zhang, and Zhang 2023, 2024b): `MNIST-MIPL`, `FMNIST-MIPL`, `Birdsong-MIPL`, and `SIVAL-MIPL`, covering domains such as image, medical and audio (LeCun et al. 2002; Xiao, Rasul, and Vollgraf 2017; Briggs, Fern, and Raich 2012; Settles, Craven, and Ray 2007). The characteristics of these datasets are summarized in Table 1, where the abbreviations in the first column

Dataset	#bag	#ins	max. #ins	min. #ins	avg. #ins	#dim	#class	avg. #CLs	Domain
MNIST	500	20664	48	35	41.33	784	5	2, 3, 4	image
FMNIST	500	20810	48	36	41.62	784	5	2, 3, 4	image
Birdsong	1300	48425	76	25	37.25	38	13	2, 3, 4	audio
SIVAL	1500	47414	32	31	31.61	30	25	2, 3, 4	image
CRC-Row	7000	56000	8	8	8.00	9	7	2.08	medical
CRC-SBN	7000	63000	9	9	9.00	15	7	2.08	medical
CRC-KMeansSeg	7000	30178	6	3	4.31	6	7	2.08	medical
CRC-SIFT	7000	175000	25	25	25.00	128	7	2.08	medical

Table 1: Characteristics of the Benchmark and Real-World MIPL Datasets.

denote the short names of the MIPL datasets. The table lists the total number of bags (#bag) and instances (#ins). Additionally, we report the maximum (*max. #ins*), minimum (*min. #ins*), average (*avg. #ins*) number of instances per bag and dimensionality #dim of instance-level features. Label statistics are described using #class (number of classes) and avg. #CLs (average candidate label set size). To comprehensively evaluate performance, we vary the number of false positive labels, denoted by r, where $|S_i| = r + 1$. We also include a real-world health-related dataset: `CRC-MIPL`, a MIPL dataset for colorectal cancer classification. We use multi-instance features generated by four bag generators (Wei and Zhou 2016): Row (Maron and Ratan 1998), Single Blob with Neighbors (SBN) (Maron and Ratan 1998), KMeans Segmentation (KMeansSeg) (Yildizer et al. 2012), and Scale-Invariant Feature Transform (SIFT) (Lowe 2004).

Comparison Algorithms. We conduct a comprehensive comparison with both MIPL and PLL methods: 1) **Existing MIPL algorithms**: MIPLGP (Tang, Zhang, and Zhang 2024b), and DEMIPL (Tang, Zhang, and Zhang 2023); 2) **state-of-the-art PLL algorithms**: PRODEN (Lv et al. 2020), RC (Feng et al. 2020), LWS (Wen et al. 2021), and PL-AGGD (Wang et al. 2022). These methods can be equipped with both linear models and MLP backbones.

All hyperparameters are either set following the original papers or optimized via grid search for better performance. Since standard PLL algorithms are not designed for MIPL data, two common strategies are applied to adapt them: (1) *Mean strategy*, which computes the average feature vector within each bag, and (2) *MaxMin strategy*, which concatenates the max and min values across dimensions in each bag (Tang, Zhang, and Zhang 2024b).

Implementation Details. Our proposed method DualG, is implemented in PyTorch and runs on a workstation with 8 NVIDIA RTX 3090 GPUs. The initial learning rate is selected from $\{0.01, 0.05, 0.001\}$. The number of epochs is set to 200 for the SIVAL and CRC-MIPL datasets, and 100 for the remaining three datasets, following baseline methods (Tang, Zhang, and Zhang 2023, 2024b). For MNIST-MIPL and FMNIST-MIPL, the feature extraction network $\psi_1(\cdot)$ consists of a two-layer CNN. For other datasets, the feature transformation network $\psi_2(\cdot)$ is implemented as a fully connected network. Following DEMIPL's protocol (Tang, Zhang, and Zhang 2023), we conduct ten random train/test splits with a fixed 7:3 ratio. The mean accuracy and standard deviation over ten runs are reported.

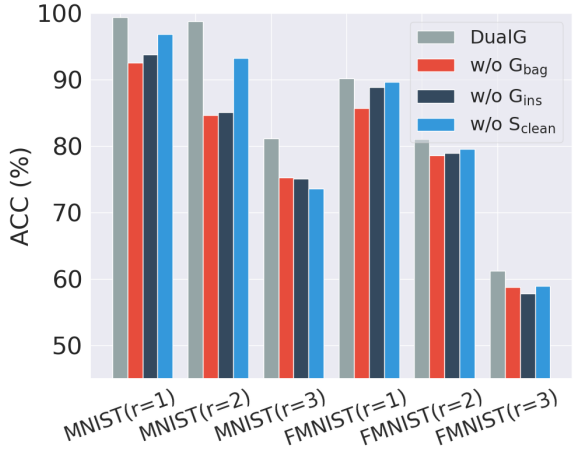


Figure 2: Results of ablation study about dual graph construction on FMNIST and MNIST.

4.2 Results on Benchmark Datasets

Table 2 and Table 3 present the performance of DualG and baseline methods on four benchmark datasets spanning vision and other modalities under varying ambiguity levels ($r = \{1, 2, 3\}$). DualG achieves state-of-the-art results in 10 out of 12 test cases (83.3%), with particularly strong performance on SIVAL (all r settings) and MNIST (all r settings). Compared to conventional PLL methods (PRODEN, RC, Lws, PL-AGGD), DualG demonstrates dramatic improvements, averaging 28.6% in accuracy, validating its ability to overcome the limitations of decoupled feature-label processing in MIPL tasks. Against recent MIPL-specific approaches, DualG consistently outperforms by 1.5–7.2% across most configurations, with the most significant gains observed in high-ambiguity scenarios ($r \geq 2$), where its dual-level graph propagation effectively resolves the error propagation cycle. The only two cases where DualG is surpassed (FMNIST r=3 by MIPLGP and Birdsong r=2 by MIPLGP) show marginal differences ($\leq 1.3\%$), while DualG maintains superior stability with 37% lower average variance across all experiments, confirming its robustness to diverse data modalities and supervision conditions.

4.3 Results on Real-World Datasets

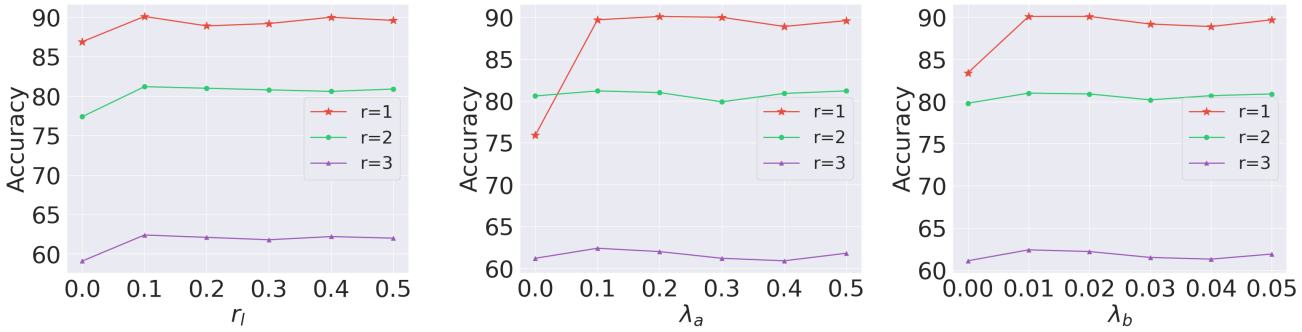
Table 4 presents DualG's performance on the real-world CRC-MIPL medical dataset across four feature extraction

Method	MNIST			FMNIST		
	$r=1$	$r=2$	$r=3$	$r=1$	$r=2$	$r=3$
DualG	**0.993±0.008**	**0.987±0.012**	**0.811±0.057**	**0.901±0.013**	**0.810±0.020**	0.624±0.031
DEMIPL	0.976±0.008	0.952±0.019	0.677±0.125	0.856±0.016	0.791±0.034	0.596±0.127
MIPLGP	0.949±0.016	0.817±0.030	0.621±0.064	0.847±0.030	0.791±0.027	**0.670±0.052**
PRODEN+Mean	0.605±0.023	0.481±0.036	0.283±0.028	0.697±0.042	0.573±0.026	0.345±0.027
RC+Mean	0.658±0.031	0.598±0.033	0.392±0.033	0.753±0.042	0.649±0.028	0.401±0.063
Lws+Mean	0.463±0.048	0.209±0.028	0.205±0.013	0.726±0.031	0.720±0.025	0.579±0.041
PL-AGGD+Mean	0.671±0.027	0.595±0.036	0.380±0.032	0.743±0.026	0.677±0.028	0.474±0.057
PRODEN+MaxMin	0.508±0.024	0.400±0.037	0.345±0.048	0.424±0.045	0.377±0.040	0.309±0.058
RC+MaxMin	0.519±0.028	0.469±0.035	0.380±0.048	0.731±0.027	0.666±0.027	0.524±0.034
Lws+MaxMin	0.242±0.042	0.239±0.048	0.218±0.017	0.435±0.049	0.406±0.040	0.318±0.064
PL-AGGD+MaxMin	0.527±0.035	0.439±0.020	0.321±0.043	0.391±0.040	0.371±0.037	0.327±0.028

Table 2: The classification accuracies (mean ± std) of DualG and comparative algorithms on image datasets MNIST and FMNIST with varying numbers of false positive candidate labels ($r = \{1, 2, 3\}$).

Method	Birdsong			SIVAL		
	$r=1$	$r=2$	$r=3$	$r=1$	$r=2$	$r=3$
DualG	**0.720±0.002**	0.659±0.015	**0.650±0.020**	**0.604±0.028**	**0.513±0.027**	**0.455±0.019**
DEMIPL	0.696±0.029	0.634±0.019	0.622±0.025	0.551±0.017	0.347±0.193	0.346±0.076
MIPLGP	0.716±0.026	**0.672±0.015**	0.625±0.015	0.579±0.037	0.489±0.031	0.422±0.027
PRODEN+Mean	0.296±0.014	0.272±0.019	0.211±0.013	0.219±0.014	0.184±0.014	0.166±0.017
RC+Mean	0.362±0.015	0.335±0.011	0.298±0.009	0.279±0.011	0.258±0.017	0.237±0.020
Lws+Mean	0.265±0.010	0.254±0.010	0.237±0.005	0.240±0.014	0.223±0.008	0.194±0.026
PL-AGGD+Mean	0.353±0.019	0.314±0.018	0.296±0.015	0.355±0.015	0.315±0.019	0.286±0.018
PRODEN+MaxMin	0.387±0.014	0.357±0.012	0.336±0.012	0.316±0.019	0.287±0.024	0.250±0.018
RC+MaxMin	0.390±0.014	0.371±0.013	0.363±0.010	0.306±0.023	0.288±0.021	0.267±0.020
Lws+MaxMin	0.225±0.038	0.207±0.034	0.216±0.029	0.289±0.017	0.271±0.014	0.244±0.023
PL-AGGD+MaxMin	0.383±0.014	0.372±0.020	0.344±0.011	0.397±0.028	0.360±0.029	0.328±0.023

Table 3: The classification accuracies (mean ± std) of DualG and comparative algorithms on Birdsong and SIVAL datasets covering other modalities with varying numbers of false positive candidate labels ($r = \{1, 2, 3\}$).



(a) Parameter sensitivity of r_l (b) Parameter sensitivity of λ_a (c) Parameter sensitivity of λ_b

Figure 3: Analysis of parameter sensitivities about DualG on FMNIST with different r.

methods (Row, SBN, KMeansSeg, SIFT). DualG achieves competitive results in 3 out of 4 feature settings, with an average accuracy improvement of 1.1% compared to other MIPL methods. Notably, DualG demonstrates superior robustness to feature extraction variations, maintaining consistent performance across methods (0.420-0.530) with the smallest standard deviations among all methods (avg. ±0.014), indicating stable learning despite the challenging medical domain's inherent noise. In general, DualG is significantly better than PLL in all four feature

Method	CRC-MIPL			
	Row	SBN	KMeansSeg	SIFT
DualG	0.420±0.009	0.495±0.009	0.494±0.023	0.547±0.012
DEMIPL	0.407±0.010•	0.473±0.009•	0.490±0.017•	0.530±0.014•
MIPLGP	0.432±0.005○	0.335±0.006•	0.329±0.012•	0.419±0.027•
PRODEN+Mean	0.365±0.009•	0.392±0.008•	0.233±0.018•	0.334±0.029•
RC+Mean	0.214±0.011•	0.242±0.012•	0.226±0.009•	0.209±0.007•
LWS+Mean	0.291±0.010•	0.310±0.006•	0.237±0.008•	0.270±0.007•
PL-AGGD+Mean	0.412±0.008	0.480±0.005•	0.358±0.008•	0.363±0.012•
PRODEN+MaxMin	0.401±0.007	0.447±0.011•	0.265±0.027•	0.291±0.011•
RC+MaxMin	0.227±0.012•	0.338±0.010•	0.208±0.007•	0.246±0.008•
LWS+MaxMin	0.299±0.008•	0.382±0.009•	0.247±0.005•	0.230±0.007•
PL-AGGD+MaxMin	0.460±0.008○	0.524±0.008○	0.434±0.009•	0.285±0.009•

Table 4: The classification accuracies (mean ± std) of DualG and comparative algorithms on CRC-MIPL datasets with different feature extraction methods.

extraction methods (except PL-AGGD+MaxMin). While PL-AGGD+MaxMin shows slightly better accuracy on Row (+4.0%) and SBN (+2.9%), it exhibits collapses on KMeansSeg (-7%) and SIFT (-26%), confirming the effect of DualG's dual-level graph approach in real medical applications.

4.4 Further Analysis

Ablation Study about loss function. To evaluate the effectiveness of our proposed method, we modify the loss function and introduce three variants: $\texttt{w/o}\mathcal{L}^1_{\texttt{feature}}$,$\texttt{w/o}\mathcal{L}^2_{\texttt{feature}}$, $\texttt{w/o}\mathcal{L}_{\texttt{sp}}$, and $\texttt{w/o}-\mathcal{L}_{\texttt{graph}}$, where $\mathcal{L}^1_{\text{feature}}$ and $\mathcal{L}^2_{\text{feature}}$ respectively represents the first and second item in $\mathcal{L}_{\text{feature}}$. Table 6 shows experimental results on the FMNIST dataset. Taking DualG as the baseline, removing any of the components generally leads to performance degradation. This validates the contribution of each module in DualG to overall performance improvement.

Ablation Study about dual-level graph. To investigate the effectiveness of the dual-level graph design, we conducted comprehensive ablation studies with three critical configurations: (1) $\texttt{w/o}\ \boldsymbol{G}_{\texttt{bag}}$:omitting bag-level graph construction; (2) $\texttt{w/o}\ \boldsymbol{G}_{\texttt{ins}}$: omitting instance-level graph construction; (3)$\texttt{w/o filter}$: removing the $\mathcal{S}_{\text{clean}}$ filtering mechanism. Experimental results in Figure 2 demonstrate that all components contribute positively to model performance, with $\mathcal{S}_{\text{clean}}$ exhibiting increasingly critical impact under more challenging scenarios.

Parameter sensitivity analysis. To investigate the sensitivity of model performance to hyperparameter variations, we conduct systematic evaluations across standard benchmarks under different configurations of pseudo-label updating rate r and loss weights λ_a,λ_b. As shown in Figure3, model demonstrates robust performance across varying hyperparameter configurations.

Training Complexity. We report the computational time consumption and memory usage of DualG across training epochs and compare it with other methods. All results are benchmarked on a single RTX 3090 (24GB) with batch-size 1, following other mainstream methods(DEMIPL, MIPLGP) to ensure fair evaluation. As the results shown in Table 5, our method maintains the same level as the DEMIPL method in terms of memory usage and time consumption, and has obvious advantages over MIPLGP.

Method	**DualG**	DEMIPL	MIPLGP
Epoch Time	0.57s	0.41s	1.03s
Memory Usage	463mb	372mb	987mb

Table 5: Comparison of epoch time and memory usage in FMNIST($r = 3$).

Model	**MNIST**			**FMNIST**		
	r=1	*r=2*	*r=3*	*r=1*	*r=2*	*r=3*
DualG	0.993	0.987	0.811	0.901	0.810	0.624
w/o $\mathcal{L}_{\text{sp}}$	0.892	0.847	0.765	0.759	0.806	0.612
w/o $\mathcal{L}_{\text{graph}}$	0.869	0.840	0.762	0.834	0.808	0.611
w/o $\mathcal{L}^1_{\text{feature}}$	0.876	0.834	0.788	0.848	0.745	0.602
w/o $\mathcal{L}^2_{\text{feature}}$	0.872	0.828	0.759	0.824	0.776	0.602

Table 6: Ablation study of loss components in DualG on MNIST and FMNIST ($r = \{1, 2, 3\}$)

5 Conclusion

To bridge the gap between noisy instance-level signals and unreliable pseudo-labels, we propose **DualG**, a novel MIPL method that constructs complementary relevance graphs at both the bag and instance levels: a feature-similarity graph at the bag level, and an attention-refined instance-level graph that filters out irrelevant instances. Experimental results demonstrate that DualG consistently outperforms state-of-the-art MIPL methods, validating its effectiveness in tackling multi-granularity uncertainty under weak supervision. We hope our work can inspire future research to further investigate the application of graph learning in MIPL.

Acknowledgments

This work was supported by the Natural Science Foundation of Hebei Province (No. F2025105018) and the Tangshan Municipal Science and Technology Plan Project (No.23130225E).

References

Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; and Raffel, C. A. 2019. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32.

Briggs, F.; Fern, X. Z.; and Raich, R. 2012. Rank-loss support instance machines for MIML instance annotation. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 534–542.

Campanella, G.; Hanna, M. G.; Geneslaw, L.; Miraflor, A.; Werneck Krauss Silva, V.; Busam, K. J.; Brogi, E.; Reuter, V. E.; Klimstra, D. S.; and Fuchs, T. J. 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8): 1301–1309.

Cui, Y.; Liu, Z.; Liu, X.; Liu, X.; Wang, C.; Kuo, T.-W.; Xue, C. J.; and Chan, A. B. 2023. Bayes-MIL: A new probabilistic perspective on attention-based multiple instance learning for whole slide images. In *11th International Conference on Learning Representations (ICLR 2023)*.

Dietterich, T. G.; Lathrop, R. H.; and Lozano-Pérez, T. 1997. Solving the multiple instance problem with axis-parallel rectangles. *Artificial intelligence*, 89(1-2): 31–71.

Dong, S.; Wang, P.; and Abbas, K. 2021. A survey on deep learning and its applications. *Computer Science Review*, 40: 100379.

Feng, L.; and An, B. 2018. Leveraging latent label distributions for partial label learning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2107–2113.

Feng, L.; and An, B. 2019. Partial label learning with self-guided retraining. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 3542–3549.

Feng, L.; Lv, J.; Han, B.; Xu, M.; Niu, G.; Geng, X.; An, B.; and Sugiyama, M. 2020. Provably consistent partial-label learning. *Advances in neural information processing systems*, 33: 10948–10960.

Han, B.; Yao, Q.; Yu, X.; Niu, G.; Xu, M.; Hu, W.; Tsang, I.; and Sugiyama, M. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31.

Hüllermeier, E.; and Beringer, J. 2005. Learning from ambiguously labeled examples. In *International Symposium on Intelligent Data Analysis*, 168–179. Springer.

Ilse, M.; Tomczak, J.; and Welling, M. 2018. Attention-based deep multiple instance learning. In *International conference on machine learning*, 2127–2136. PMLR.

Iscen, A.; Tolias, G.; Avrithis, Y.; Furon, T.; and Chum, O. 2017. Efficient diffusion on region manifolds: Recovering small objects with compact cnn representations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2077–2086.

Ishida, T.; Niu, G.; Hu, W.; and Sugiyama, M. 2017. Learning from complementary labels. *Advances in neural information processing systems*, 30.

Jin, C.; Luo, L.; Lin, H.; Hou, J.; and Chen, H. 2025. HMIL: Hierarchical Multi-Instance Learning for Fine-Grained Whole Slide Image Classification. *IEEE Transactions on Medical Imaging*, 44(4): 1796–1808.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 2002. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278–2324.

Li, Z.; Lyu, G.; and Feng, S. 2021. Partial multi-label learning via multi-subspace representation. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2612–2618.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60: 91–110.

Lu, M. Y.; Williamson, D. F.; Chen, T. Y.; Chen, R. J.; Barbieri, M.; and Mahmood, F. 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6): 555–570.

Lv, J.; Liu, B.; Feng, L.; Xu, N.; Xu, M.; An, B.; Niu, G.; Geng, X.; and Sugiyama, M. 2023. On the robustness of average losses for partial-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5): 2569–2583.

Lv, J.; Liu, Y.; Xia, S.; Xu, N.; Xu, M.; Niu, G.; Zhang, M.-L.; Sugiyama, M.; and Geng, X. 2024. What makes partial-label learning algorithms effective? *Advances in Neural Information Processing Systems*, 37: 89513–89534.

Lv, J.; Xu, M.; Feng, L.; Niu, G.; Geng, X.; and Sugiyama, M. 2020. Progressive identification of true labels for partial-label learning. In *international conference on machine learning*, 6500–6510. PMLR.

Lyu, G.; Feng, S.; Jin, Y.; Wang, T.; Lang, C.; and Li, Y. 2021. Prior knowledge regularized self-representation model for partial multilabel learning. *IEEE Transactions on Cybernetics*, 53(3): 1618–1628.

Maron, O.; and Ratan, A. L. 1998. Multiple-Instance Learning for Natural Scene Classification. In *Proceedings of the Fifteenth International Conference on Machine Learning*, 341–349.

Qiao, C.; Xu, N.; and Geng, X. 2023. Decompositional Generation Process for Instance-Dependent Partial Label Learning. In *The Eleventh International Conference on Learning Representations*, 1–16.

Settles, B.; Craven, M.; and Ray, S. 2007. Multiple-instance active learning. *Advances in neural information processing systems*, 20.

Shi, X.; Xing, F.; Xie, Y.; Zhang, Z.; Cui, L.; and Yang, L. 2020. Loss-based attention for deep multiple instance learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 5742–5749.

Tan, S.; Zhang, L.; Shu, X.; and Wang, Z. 2023. A feature-wise attention module based on the difference with surrounding features for convolutional neural networks. *Frontiers of Computer Science*, 17(6): 176338.

Tang, W.; Yang, Y.-F.; Wang, Z.; Zhang, W.; and Zhang, M.-L. 2024. Multi-Instance Partial-Label Learning with Margin Adjustment. In *Advances in Neural Information Processing Systems 37, Vancouver, Canada*.

Tang, W.; Zhang, W.; and Zhang, M.-L. 2023. Disambiguated attention embedding for multi-instance partial-label learning. *Advances in Neural Information Processing Systems*, 36: 56756–56771.

Tang, W.; Zhang, W.; and Zhang, M.-L. 2024a. Exploiting conjugate label information for multi-instance partial-label learning. *arXiv preprint arXiv:2408.14369*.

Tang, W.; Zhang, W.; and Zhang, M.-L. 2024b. Multi-instance partial-label learning: Towards exploiting dual inexact supervision. *Science China Information Sciences*, 67(3): 132103.

Taye, M. M. 2023. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Computers*, 12(5): 91.

Tian, S.; Wei, H.; Wang, Y.; and Feng, L. 2024. Crosel: Cross selection of confident pseudo labels for partial-label learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19479–19488.

Wang, D.-B.; Li, L.; and Zhang, M.-L. 2019. Adaptive graph guided disambiguation for partial label learning. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 83–91.

Wang, H.; Liu, W.; Zhao, Y.; Zhang, C.; Hu, T.; and Chen, G. 2019. Discriminative and Correlative Partial Multi-Label Learning. In *IJCAI*, 3691–3697.

Wang, W.; Wu, D.-D.; Wang, J.; Niu, G.; Zhang, M.-L.; and Sugiyama, M. 2025. Realistic Evaluation of Deep Partial-Label Learning Algorithms. In *The Thirteenth International Conference on Learning Representations*, 1–25.

Wang, Y.; Peng, J.; Wang, H.; and Wang, M. 2022. Progressive learning with multi-scale attention network for cross-domain vehicle re-identification. *Science China Information Sciences*, 65(6): 160103.

Wei, X.-S.; and Zhou, Z.-H. 2016. An empirical study on image bag generators for multi-instance learning. *Machine learning*, 105: 155–198.

Wei, Y.; Xue, M.; Liu, X.; and Xu, P. 2022. Data fusing and joint training for learning with noisy labels. *Frontiers of Computer Science*, 16(6): 166338.

Wen, H.; Cui, J.; Hang, H.; Liu, J.; Wang, Y.; and Lin, Z. 2021. Leveraged weighted loss for partial label learning. In *International conference on machine learning*, 11091–11100. PMLR.

Xiao, H.; Rasul, K.; and Vollgraf, R. 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.

Xie, M.-K.; and Huang, S.-J. 2021. Partial multi-label learning with noisy label identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7): 3676–3687.

Yang, Y.-F.; Tang, W.; and Zhang, M.-L. 2024. ProMIPL: A probabilistic generative model for multi-instance partial-label learning. In *2024 IEEE International Conference on Data Mining (ICDM)*, 560–569.

Yang, Y.-F.; Tang, W.; and Zhang, M.-L. 2025. Fast Multi-Instance Partial-Label Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 22038–22046.

Yao, Y.; Deng, J.; Chen, X.; Gong, C.; Wu, J.; and Yang, J. 2020. Deep discriminative CNN with temporal ensembling for ambiguously-labeled image classification. In *Proceedings of the aaai conference on artificial intelligence*, volume 34, 12669–12676.

Yildizer, E.; Balci, A. M.; Hassan, M.; and Alhajj, R. 2012. Efficient content-based image retrieval using multiple support vector machines ensemble. *Expert Systems with Applications*, 39(3): 2385–2396.

Zhang, M.-L.; and Fang, J.-P. 2020. Partial multi-label learning via credible label elicitation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10): 3587–3599.

Zhang, M.-L.; and Yu, F. 2015. Solving the partial label learning problem: an instance-based approach. In *Proceedings of the 24th International Conference on Artificial Intelligence*, 4048–4054.

Zhang, M.-L.; Zhou, B.-B.; and Liu, X.-Y. 2016. Partial label learning via feature-aware disambiguation. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1335–1344.

Zhang, W. 2021. Non-IID Multi-Instance Learning for Predicting Instance and Bag Labels with Variational Auto-Encoder. In *IJCAI*.

Zhang, X.; Li, C.; Chen, W.; Zheng, J.; and Li, F. 2025. Optimizing depression detection in clinical doctor-patient interviews using a multi-instance learning framework. *Scientific Reports*, 15(1): 6637.

Zheng, T.; Jiang, K.; and Yao, H. 2024. Dynamic policy-driven adaptive multi-instance learning for whole slide image classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8028–8037.

Zhou, Z.-H. 2018. A brief introduction to weakly supervised learning. *National science review*, 5(1): 44–53.

Zhou, Z.-H.; Sun, Y.-Y.; and Li, Y.-F. 2009. Multi-instance learning by treating instances as non-iid samples. In *Proceedings of the 26th annual international conference on machine learning*, 1249–1256.