Mutual Information Guided Reinforcement Learning for Ambiguous Label Disambiguation

Jinfu Fan[1], Jiangnan Li[1], Xiaohui Zhong[1], Kangrui Ren[2*], Linqing Huang[3 *]

[1] College of Computer Science and Technology, Qingdao University
[2] College of Electronics and Information Engineering, Tongji University
[3] School of Computer Science, Shanghai Jiao Tong University

{fan_jinfu,ljn99008}@163.com,{zhongxiaohui,2310213,huanglinqing}@{qdu,tongji,sjtu}.edu.cn

Abstract

Partial-label learning (PLL) addresses challenging scenarios where each instance is associated with a set of candidate labels and only one is the truth. Most existing PLL methods rely on static disambiguation heuristics, which are prone to error propagation when the ambiguity labels are high. To address this issue, we propose a novel *mutual information guided reinforcement learning framework for partial label disambiguation* (MR-PLL). In this framework, label disambiguation is formulated as a sequential Markov decision process, where an agent dynamically discriminates whether to retain, modify, or abstain from correcting ambiguous labels. To ensure reliable decision making, we introduce a mutual information guided gating mechanism that adaptively adjusts the confidence of soft label propagation according to the dependency between feature representations and labels. The abstention mechanism allows the model to postpone uncertain decisions, resulting in more reliable disambiguation. Furthermore, we design an information weighted reward term during the actor-critic process to gradually improve the label disambiguation ability of the policy, and provide theoretical analysis on convergence and bias reduction. Experiments on benchmark and real datasets verify the effectiveness of the proposed algorithm.

1 Introduction

Learning with ambiguous labels has become an increasingly important problem in modern machine learning, arising naturally in scenarios where precise annotations are difficult, costly, or even impossible to obtain. Therefore, *partial label learning* (PLL) [Cour *et al.*, 2011; Fan *et al.*, 2026a] has attracted much attention to address the impact of ambiguous labels. PLL is a weakly supervised learning framework, where each instance is associated with a candidate label set that contains the unknown ground-truth label but also includes multiple irrelevant labels. Although PLL can significantly reduce annotation cost, accurately identifying ground-truth labels from candidate label sets to reduce the impact of ambiguous labels poses a fundamental challenge to model training.

Intuitively, the basic method for accomplishing the PLL task is to perform disambiguation. To achieve this purpose, label confidence in PLL has been investigated according to two general strategies, i.e., *non-deep disambiguation strategy* (NDS) [Berg *et al.*, 2004; Luo and Orabona, 2010; Chai *et al.*, 2019; Fuchs *et al.*, 2024] and *deep disambiguation strategy* (DDS) [Yao *et al.*, 2020; Fan *et al.*, 2026b; Fan *et al.*, 2021; Li *et al.*, 2025]. For NDS, attempt to progressively identify the true label by iteratively updating disambiguation labels. DDS incorporates some training objectives compatible with stochastic optimization to distinguish the true label from candidate labels, such as adding regularization terms, data augmentation, and adjusting loss weight. Although existing PLL algorithms have achieved certain success, they still have limitations in relying on static or heuristic update rules, implicitly assuming that label correction is a deterministic process driven only by local loss minimization. *Therefore, these methods often lack the ability to actively explore different possibilities and are prone to accumulating errors, especially in training iterations that reinforce the impact of early error disambiguation decisions.*

To tackle this issue, a novel *mutual information guided reinforcement learning framework for partial label disambiguation* (MR-PLL) is proposed to address PLL from a decision-theoretic perspective. Unlike existing PLL approaches that treat label correction as a deterministic optimization problem, we formulate the disambiguation process as a reinforcement learning problem in a *Markov Decision Process* (MDP) [Sutton *et al.*, 1998]. Recently, *Reinforcement Learning* (RL) [Kaelbling *et al.*, 1996] has provided a principled framework for modeling sequential decision problems, where agents interact with the environment and learn optimal policies by maximizing long-term cumulative rewards. Therefore, RL can dynamically adjust its ambiguous labels correction strategy based on its behavioral feedback and optimize performance through a series of decisions. However, its potential for resolving ambiguous labels in PLL remains largely underexplored. A core challenge lies in designing meaningful reward signals that can effectively guide the label disambiguation process without access to ground-truth labels.

*The corresponding author.

To address the above limitations, this article designs an information weighted reward term for MR-PLL as the core component of RL. The reward integrates classification feedback, label consistency, and the mutual information between feature representations. *Mutual Information* (MI) provides a measure of similarity between representations of different classes, encouraging agents to promote feature alignment between the same labeled classes and avoid unreliable corrections. Meanwhile, MI estimation can be used to estimate the uncertainty between instances, which is used to adjust the value updates of RL agents. Specifically, the higher the uncertainty of the estimation, the smaller the corresponding update, which avoids ambiguous labeling leading to erroneous updates (values or policy functions) during the learning process, thereby making the agent unstable and forgetting all valuable states previously seen (catastrophic forgetting). Meanwhile, our method utilizes an instance selection-based discriminator to dynamically discriminate whether to retain, modify, or abstain from correcting ambiguous labels. The abstention mechanism is introduced the agent to temporarily withhold highly ambiguous instances from classifier updates, thus mitigating bias accumulation from ambiguous labels. To further enhance robustness, we propose a MI guided soft label fusion strategy. It adaptively blends original candidate labels with propagated pseudo-labels based on each instance reliability estimates. This design lowers gradient variance and stabilizes training, especially in early stages when model predictions are uncertain. The resulting framework unifies RL, information theory, and PLL into an end-to-end optimization pipeline. Furthermore, we provide a theoretical analysis of convergence and bias reduction. The contributions of this work are summarized as follows:

- We formulate PLL as a reinforcement learning problem, providing a novel decision-theoretic perspective on ambiguous supervision.
- We develop an information weighted reward term that encourages effective label disambiguation by aligning classification feedback, label consistency, and mutual information.
- A MI guided soft label fusion strategy and an abstention mechanism are proposed to adaptively balance uncertainty and bias control in label correction.

2 Related Work

2.1 Partial Label Learning

To alleviate the impact of ambiguous labels on classifier learning, two general strategies based on machine learning have been proposed for PLL, i.e., *non-deep disambiguation strategy* (NDS) and *deep disambiguation strategy* (DDS). The NDS operates on the principle that similar instances tend to share the same label. It resolves label ambiguity by leveraging the relationships among the candidate labels of neighboring instances[Chai *et al.*, 2019]. A wide range of machine learning techniques have been employed to tackle PLL challenges, including *k-Nearest Neighbors* (KNN)[Luo and Orabona, 2010; Zhang *et al.*, 2016], *Expectation Maximization* (EM) [Liu and Dietterich, 2012], *Low-Rank Representation* (LRR) [Fan *et al.*, 2022; Chen *et al.*, 2013], and *Support Vector Machine* (SVM) [Chai *et al.*, 2019; Fan *et al.*, 2023]. The progress in deep neural networks has motivated significant efforts to develop deep learning-based frameworks for PLL. DDS incorporates some training objectives compatible with stochastic optimization to distinguish the true label from candidate labels, such as adding regularization terms [Wang *et al.*, 2023; Feng *et al.*, 2025; Fan *et al.*, 2026c], data augmentation [Chai *et al.*, 2019], and the weight of the loss function [Fan *et al.*, 2026a; Wen *et al.*, 2021; Fan *et al.*, 2021], have been proposed to deal with ambiguous labels. Instance-dependent PLL offers a more realistic framework for ambiguous supervision. To tackle this challenge, methods such as contrastive learning, Bayesian models, and variational inference have been developed [Xu *et al.*, 2021; Xu *et al.*, 2024; Wu *et al.*, 2024].

The drawback of PLL is that it relies on static disambiguation heuristics, which are prone to error propagation when the ambiguity labels are high. Label disambiguation is inherently a dynamic process, early decisions directly affect the feature representations, which in turn constrain future label predictions. This creates a vicious cycle where an initial labeling error can be reinforced by subsequent updates, consequently causing severe and irreversible performance decline.

2.2 Reinforcement Learning

Reinforcement learning (RL) is a learning paradigm in which an autonomous agent acquires decision-making strategies through continuous interaction with an environment, optimizing its behavior over a sequence of actions to achieve long-term objectives [Sutton *et al.*, 1998]. A defining strength of RL lies in its capacity for effective exploration [Amin *et al.*, 2021] and its ability to optimize policies with respect to delayed and cumulative rewards [Xu *et al.*, 2023]. Owing to these advantages, RL has been widely adopted in real-world environments, especially in robotics-relevant applications [Kober *et al.*, 2013]. *Preference-based reinforcement learning* (PbRL) further extends this paradigm by enabling agents to learn directly from human preference feedback over pairs of trajectory segments, thereby alleviating the need for explicitly designed reward functions [Christiano *et al.*, 2017]. More recently, PbRL has demonstrated remarkable success in the fine-tuning of *Large Language Models* (LLMs) [Ouyang *et al.*, 2022], where human preferences provide a scalable and flexible alternative to manually specified reward signals. Unlike supervised learning, RL is a key artificial intelligence paradigm that enables an agent to learn optimal behavior through trial and error, guided by rewards. Therefore, the advantage of RL is that it is more suitable for labeling environments where data is scarce or unavailable.

3 MR-PLL Method

We propose a novel MI guided RL for ambiguous label disambiguation. The PLL is formulated as the action execution in the MDP, where an agent dynamically decides whether to retain, modify, or abstain from correcting ambiguous labels. Designing a suitable reward function plays a crucial role in building reinforcement learning models for real-world applications in the face of ambiguous labels in PLL. *It is difficult*

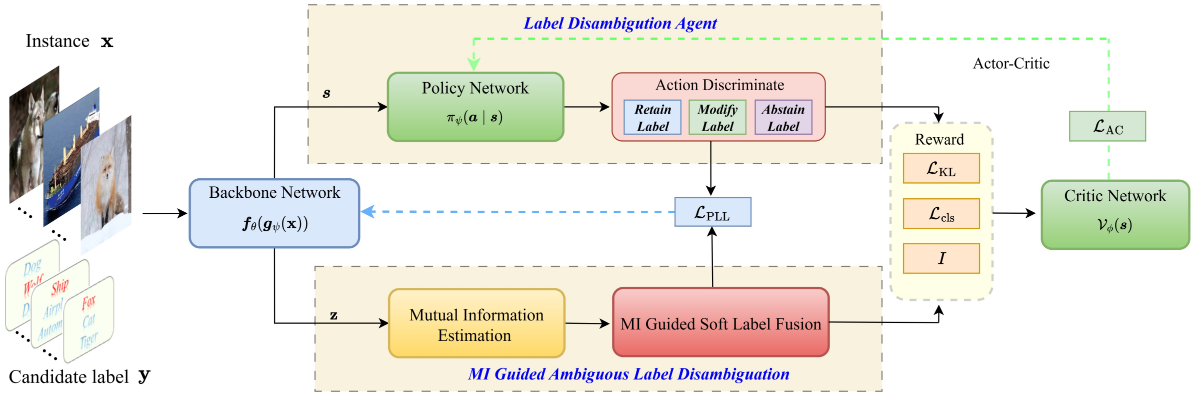


Figure 1: The framework of MR-PLL.

to design a reward function that produces credible rewards in the presence of ambiguous labels. This is because when the output of the reward function receives ambiguous feedback from candidate labels, it can lead to biased rewards for the classification model. To address this issue, we design an information weighted reward for actions to guide the action decision process. Meanwhile, the MI guided soft label propagation mechanism is adopted to train the disambiguation policy. The overall framework is illustrated in Figure 1, and the detailed components of the proposed method are introduced in the following subsections.

3.1 Problem Formulation

Let $\mathcal{X} = \mathbb{R}^d$ denote a d-dimensional feature space and $\mathcal{Y} = \{y_c : c = 1, \dots, Q\}$ denote the output space with Q class labels. The PLL training dataset $\mathcal{D} = \{(\mathbf{x}_i, S_i) | 1 \leq i \leq N\}$ of N instances contains feature vectors $\mathbf{x}_i \in \mathcal{X}$ and a set of candidate labels $S_i \subseteq \mathcal{Y}$. $|S_i|$ represents the number of ambiguous labels for instance $\mathbf{x}_i$. Particularly, instance $\mathbf{x}_i$ is annotated by a label vector $\mathbf{y}_i = \{y_c\}_{c=1}^{Q}$, where $y_{ic} \in \{0, 1\}$ denotes whether the label y_c is present in the candidate labels ('1') or the non-candidate labels ('0'). The objective is to remove the influence of false-positive labels and learn an effective prediction model $\boldsymbol{f}_\theta \circ \boldsymbol{g}_\psi : \mathcal{X} \mapsto \mathcal{Y}$ that can make correct predictions on unseen inputs, where $\boldsymbol{g}_\psi$ denotes the feature extractor with parameters ψ, and $\boldsymbol{f}_\theta$ represents the classifier with parameters θ.

3.2 RL for Ambiguous Label Disambiguation

RL problem is often modeled as a MDP, characterized by the tuple $(\mathcal{S}, \mathcal{A}, \pi, \mathcal{P}, \mathcal{R})$, where $\mathcal{S}$ denotes the state space, $\mathcal{A}$ represents the action space, $\mathcal{P}(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{a})$ defines the transition dynamics for $\boldsymbol{s}$, $\boldsymbol{s}' \in \mathcal{S}$ and $\boldsymbol{a} \in \mathcal{A}$, and $\mathcal{R}$ is a reward function. The goal of an RL agent is to learn a policy π that maximizes the expected cumulative reward.

State Representation

For each instance $\mathbf{x}_i$, the state $\boldsymbol{s}_i$ includes the potential information of an instance to be labeled, including both representation-level, prediction-level, and uncertainty-aware information. Specifically, we define the state as,

$$\boldsymbol{s}_i = \big[\mathbf{z}_i,\ \mathbf{p}_i,\ H(\mathbf{p}_i),\ \max(\mathbf{p}_i),\ \Delta(\mathbf{p}_i)\big] \tag{1}$$

where $\mathbf{z}_i = \boldsymbol{g}_\psi(\mathbf{x}_i)$ is the feature representation extracted by the backbone network, $\mathbf{p}_i = \text{softmax}(\boldsymbol{f}_\theta(\mathbf{z}_i))$ is the predicted label probability, $H(\mathbf{p}_i) = -\sum_{c=1}^{Q} p_{ic} \log p_{ic}$ is the predictive entropy, $\max(\mathbf{p}_i)$ is the maximum confidence score, and $\Delta(\mathbf{p}_i)$ denotes the margin between the top-2 predicted probabilities. This design provides a sufficient statistic capturing discriminability, uncertainty, and learning dynamics, which is critical for reliable policy learning to select action.

Action Space

Due to the PLL setting, our agent is required to determine the label of each class for instance. To avoid the influence of ambiguous labels, we define the action space $\mathcal{A}$ corresponding to three possible disambiguation decisions, such that for each action is defined as $\boldsymbol{a}_i = \{0, 1, 2\}$. $\boldsymbol{a}_i = 0$ represents keep the original label distribution, $\boldsymbol{a}_i = 1$ indicates the current label needs to be corrected, and $\boldsymbol{a}_i = 2$ indicates that uncertain instance need to be postponed for disambiguation to avoid updating the classifier with this instance. It allows the policy to explicitly model uncertainty and prevents premature and potentially erroneous label corrections.

MI Guided Ambiguous Label Disambiguation

1) KNN Attention for Label Refinement: To mitigate the noisy labels influence, we through KNN attention to identify potential ground truth labels by utilizing complementary information between instances and candidate labels, making the model more resistant to ambiguous labels. The underlying assumption is that a label consistent with those of its k-nearest neighbors is more likely to have the same label. To this end, we aggregate the labels of its k-nearest neighbors in the current state $\boldsymbol{s}$ using an attention mechanism to generate a new label prediction $\hat{\mathbf{y}}_i$, which can be computed as follows,

$$\hat{\mathbf{y}}_i = \sum_{j \in \mathcal{N}_k(\mathbf{x}_i)} \alpha_{ij} \mathbf{y}_j \tag{2}$$

where $\mathcal{N}_k(\mathbf{x}_i)$ denotes the KNN set of instance $\mathbf{x}_i$ based on the Euclidean distances calculated in the embedding space, calculated by the backbone network $\boldsymbol{g}_\psi$, and the attention weights α_{ij} is computed using similarities of instance pairs in the embedding space,

$$\alpha_{ij} = \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{l \in \mathcal{N}_k(\mathbf{x}_i)} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_l)/\tau)} \tag{3}$$

where τ is a temperature hyperparameter, and $\text{sim}(\cdot, \cdot)$ denotes the cosine similarity.

2) Mutual Information Estimation: A critical challenge in ambiguous label disambiguation is determining when such propagation is reliable. To this end, we introduce MI between positive and negative feature representations as a principled confidence measure. Then, we use variational f-divergence estimation [Poole *et al.*, 2019] to maximize MI, which can be calculated by,

$$I_i = 1 + \mathbb{E}_{\mathbb{P}}\big[T_\omega(\mathbf{z}_i, \mathbf{z}_i^+)\big] - \mathbb{E}_{\mathbb{Q}}\big[\exp\big(T_\omega(\mathbf{z}_i, \mathbf{z}_i^-)\big)\big] \tag{4}$$

where $T_\omega(\cdot)$ is the density estimate, $\mathbb{P}$ denotes the joint probability distribution, and $\mathbb{Q}$ denotes the marginal probability distribution. $\mathbf{z}_i^+$ and $\mathbf{z}_i^-$ are the positive and negative instance representations of $\mathbf{z}_i$, where $\mathbf{z}_i^+$ selects from the k-nearest neighbors in $\mathbf{z}_i$, and $\mathbf{z}_i^-$ selects based on the intersection of the candidate labels sets between the instances being empty. While maximizing MI, by optimizing the relationship between the joint distribution and the marginal distribution, we significantly enhance the similarity between instances and their positive counterparts while reducing the similarity between instances and negative counterparts. A good representation can effectively improve the disambiguation ability of the model. This design aims to achieve tight aggregation of intra-class instances and effective separation of inter-class instances, thereby enhancing the discriminative and generalization performance of the model.

3) MI Guided Soft Label Fusion: A low mutual information between representations indicates weak semantic alignment and unreliable supervisory signals, suggesting that the label is more likely to be noisy and in need of correction. Therefore, we propose an MI guided soft label fusion,

$$\check{\mathbf{y}}_i = (1 - \beta_i)\hat{\mathbf{y}}_i + \beta_i \boldsymbol{f}_\theta(\boldsymbol{g}_\psi(\mathbf{x}_i)) \tag{5}$$

$$\bar{\mathbf{y}}_i = \text{Normalize}(\eta \check{\mathbf{y}}_i + (1 - \eta)\mathbf{y}_i) \tag{6}$$

where $\text{Normalize}(\cdot)$ is normalization function and $\eta \in (0, 1)$ is the balance parameter. β_i is defined as $\beta_i = \sigma(I_i - \mathbb{E}[I])$, where $\sigma(\cdot)$ is the sigmoid function. Detailed derivations of MI guided soft label fusion can be found in **Appendix A.1**.

4) Deterministic Transition with Stochastic Policy: Given state $\boldsymbol{s}_i$, we determine the action $\boldsymbol{a}_i$ using the stochastic policy function to enhance the exploratory behavior of the learning process, the label distribution evolves as,

$$\tilde{\mathbf{y}}_i = \begin{cases} \mathbf{y}_i, & \boldsymbol{a}_i = 0, \\ \bar{\mathbf{y}}_i, & \boldsymbol{a}_i = 1, \\ \bar{\mathbf{y}}_i, & \boldsymbol{a}_i = 2, \end{cases} \tag{7}$$

where $\tilde{\mathbf{y}}_i$ denotes the propagated soft label. Notably, for the abstention action ($\boldsymbol{a}_i = 2$), the classifier parameters are not updated using this instance, ensuring that uncertain supervision does not affect representation learning. This mechanism maintains label uncertainty through soft label distributions and updates them via probabilistic corrections guided by noise likelihood. The introduced stochasticity promotes broader exploration of the state–action space, helping the model escape suboptimal solutions. Meanwhile, gradual label updates reduce instability caused by abrupt label changes, allowing the model to adapt smoothly to evolving learning dynamics and ultimately improving robustness under noisy supervision.

Actor-Critic Framework

MR-PLL adopts an actor-critic framework to learn the policy function by maximizing the expected cumulative reward. In this framework, the policy function π_ψ is considered as an "actor", where policy network is parameterized by ψ and outputs the probability $\pi_\psi(\boldsymbol{a} \mid \boldsymbol{s})$ for each action condition on the current state. The value function $\mathcal{V}_\phi(\boldsymbol{s})$ and is parameterized by ϕ is introduced as a "critic" to directly estimate the Q-value.

1) Reward Function: In RL, the reward function assesses the quality of an action $\boldsymbol{a}$ taken in a given state $\boldsymbol{s}$ and directs the learning process toward the desired objective. To achieve the goal of eliminating ambiguity labels for the proposed MR-PLL, we design the reward function to evaluate the quality of an action through the labels produced from taking the action. The reward is designed to reflect both classification performance, label consistency, and information reliability.

$$\mathcal{R} = -\mathcal{L}_{\text{cls}} - \gamma \mathcal{L}_{\text{KL}} + \lambda I \tag{8}$$

where γ and λ are balance parameter. $\mathcal{L}_{\text{cls}}$ is the partial label classification loss, which can be calculated as,

$$\mathcal{L}_{\text{cls}} = \mathbb{E}[(1 - \mathbb{I}[\boldsymbol{a}_i = 2]) \cdot \ell\left(\boldsymbol{f}_\theta\left(\boldsymbol{g}_\psi(\mathbf{x}_i)\right), \tilde{\mathbf{y}}_i\right)] \tag{9}$$

where $\mathbb{I}(\cdot)$ represents the indicator function and $\ell(\cdot)$ is cross-entropy loss. This item learns through a progressive guidance model with soft labels after disambiguation, first learning simple instances with high credibility, and then learning difficult instances. Furthermore, we introduce a label consistency reward $\mathcal{L}_{\text{KL}} = D_{\text{KL}}(\tilde{\mathbf{y}}_i \| \hat{\mathbf{y}}_i)$, which can encourage retaining confident predictions and revising uncertain ones. Meanwhile, the role of MI loss is to constrain the reward function to encourage actions that reduce classification loss, while also benefiting instances with high MI.

2) Advantage Estimation: The advantage function is defined as $A_i^{dv} = \mathcal{R}_i - \mathcal{V}_\phi(\boldsymbol{s}_i)$, which is normalized to stabilize training.

Therefore, updating the learning objectives of actors and critics through minimization,

$$\mathcal{L}_{\text{AC}} = -\mathbb{E}\left[\log \pi_\psi\left(\boldsymbol{a}_i \mid \boldsymbol{s}_i\right) A_i^{dv}\right] + \frac{1}{2}\mathbb{E}\left[\left(\mathcal{R}_i - \mathcal{V}_\phi(\boldsymbol{s}_i)\right)^2\right] \tag{10}$$

The final classifier objective is,

$$\mathcal{L}_{\text{PLL}} = \mathcal{L}_{\text{cls}} - \lambda_{MI} I \tag{11}$$

where λ_{MI} is the weighting coefficient. The final prediction model $\boldsymbol{f}_\theta \circ \boldsymbol{g}_\psi$ is used to make correct predictions on unseen inputs. The learning process of the proposed MR-PLL method is summarized in Algorithm 1. Meanwhile, the theoretical analysis of objective function is elaborated in **Appendix A.2**.

Algorithm 1 MR-PLL Framework

Input: Dataset $\mathcal{D} = (\mathbf{x}_i, \mathbf{y}_i)$, classifier $\boldsymbol{f}_\theta \circ \boldsymbol{g}_\psi$, the policy function π_ψ, and the value function $\mathcal{V}_\phi(\boldsymbol{s})$
Output: $\mathbf{y}^* = \arg\max(\boldsymbol{f}_\theta(\boldsymbol{g}_\psi(\mathbf{x})))$

Initialization:
while $t < maxIter$ **do**
  Compute state representation according to Eq.(1);
  MI guided soft label fusion computed according to Eq.(2), Eq.(4) and Eq.(6);
  Policy action: Sample action $\pi_\psi\left(\boldsymbol{a} \mid \boldsymbol{s}\right)$;
  Label transition according to Eq.(7);
  Reward computation according to Eq.(8);
  Update actor-critic according to Eq.(10);
  Update classifier according to Eq.(11);
end while

4 Experiments

4.1 Experimental Details

Datasets To assess the effectiveness of the proposed MR-PLL algorithm, we conduct experiments on four widely used benchmark datasets, including **MNIST** [LeCun *et al.*, 1998], **Fashion-MNIST** [Xiao *et al.*, 2017], **Kuzushiji-MNIST** [Clanuwat *et al.*, 2018], and **CIFAR-10** [Krizhevsky *et al.*, 2009]. For these datasets, partial label annotations are synthesized following the standard procedure introduced in [Yao *et al.*, 2021]. Specifically, the candidate label set for each instance is generated through $Q-1$ independent Bernoulli trials, where the parameter q controls the probability of adding an incorrect label [Fan *et al.*, 2024]. A larger value of q results in candidate sets with higher label ambiguity, thereby increasing the difficulty of the disambiguation task. In our experiments, we consider three noise levels, $q \in \{0.1, 0.3, 0.5\}$, to evaluate the robustness of MR-PLL under varying degrees of label noise across all four benchmark datasets. In addition to synthetic partial label settings, we further validate our method on five real-world partially labeled datasets, including **Lost** [Cour *et al.*, 2011], **BirdSong** [Briggs *et al.*, 2012], **MSRCv2** [Liu and Dietterich, 2012], **Soccer Player** [Zeng *et al.*, 2013], and **Yahoo!News** [Guillaumin *et al.*, 2010]. Detailed dataset statistics are provided in **Appendix B.1**.

Comparison Methods. To better verify the superiority of the MR-PLL algorithm, we compare the proposed method with the state-of-the-art PLL methods including CC [Feng *et al.*, 2020], PRODEN [Yao *et al.*, 2021], LW [Lv *et al.*, 2020], CAVL [Zhang *et al.*, 2021], DIRK [Wu *et al.*, 2024], VALEN [Xu *et al.*, 2024] and LS-PLL [Gong *et al.*, 2024]. Each comparison method sets parameters according to suggested settings in the original papers, and the details are shown in **Appendix B.2**.

Implementation Details. The implementation is based on PyTorch [Paszke *et al.*, 2019] training, and all experiments are conducted with an NVIDIA GeForce RTX 4090 GPU. The backbone network uses an MLP and ConvNet to apply to real-world partially labeled datasets and benchmark datasets, which outputs both class logits and feature representations. The RL component is implemented using an Actor–Critic architecture, where both the actor and critic are parameterized by two-layer multilayer perceptrons with $ReLU$ activations. We train models for 200 epochs, 128 batch size and use the SGD optimizer with a momentum of 0.9. Label refinement is performed using a k-nearest neighbor attention mechanism in the normalized feature space. The number of neighbors is fixed to $k = 8$, and the temperature parameter in the attention softmax is set to $\tau = 0.5$. The balance parameters η, γ, λ and λ_{MI} are set to 0.1, 0.2, 0.6 and 0.2, respectively.

Method	Datasets				
	Lost	**MSRCv2**	**Birdsong**	**Soccer Player**	**Yahoo!News**
CC	$69.29 \pm 0.31\%$	$48.57 \pm 0.29\%$	$56.31 \pm 0.81\%$	$49.12 \pm 0.34\%$	$52.47 \pm 0.34\%$
PRODEN	$65.17 \pm 0.72\%$	$52.00 \pm 0.46\%$	$63.38 \pm 0.52\%$	$51.45 \pm 0.19\%$	$59.06 \pm 0.38\%$
LW	$67.11 \pm 0.78\%$	$47.15 \pm 0.75\%$	$66.50 \pm 0.73\%$	$49.15 \pm 0.38\%$	$47.03 \pm 0.37\%$
CAVL	$63.11 \pm 0.72\%$	$52.84 \pm 0.16\%$	$70.50 \pm 0.92\%$	$54.27 \pm 0.37\%$	$63.86 \pm 0.58\%$
DIRK	$74.26 \pm 0.58\%$	$44.61 \pm 0.32\%$	$71.28 \pm 0.21\%$	$53.37 \pm 0.48\%$	$61.38 \pm 0.22\%$
VALEN	$71.56 \pm 0.80\%$	$45.89 \pm 0.55\%$	$72.30 \pm 0.49\%$	$53.91 \pm 0.12\%$	$\mathbf{67.73} \pm \mathbf{0.32}\%$
LS-PLL	$66.67 \pm 0.13\%$	$51.42 \pm 0.38\%$	$56.80 \pm 0.63\%$	$54.65 \pm 0.48\%$	$62.56 \pm 0.35\%$
MR-PLL	$\mathbf{75.85} \pm \mathbf{0.42}\%$	$\mathbf{58.85} \pm \mathbf{0.28}\%$	$\mathbf{75.35} \pm \mathbf{0.34}\%$	$\mathbf{56.84} \pm \mathbf{0.28}\%$	$64.98 \pm 0.37\%$

Table 1: Accuracy comparisons on real-world partially labeled datasets.

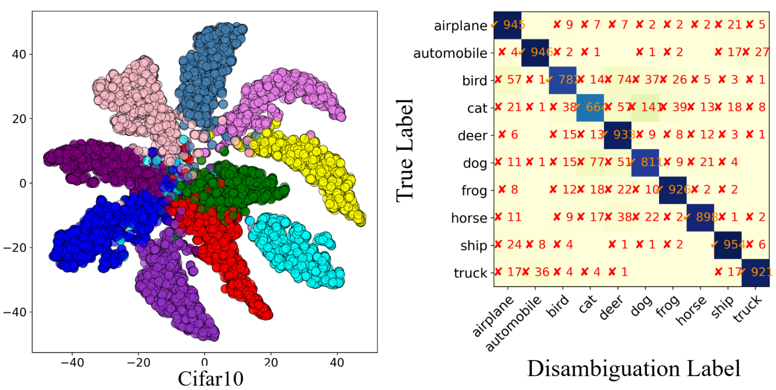
4.2 Experimental Results

Comparisons with State-of-the-arts methods. Across 35 experiments (5 real-world datasets $\times$ 7 comparison algorithms) in Table 1, MR-PLL significantly outperforms all other methods in 97.1% of the cases. Especially on datasets **BirdSong**, **MSRCv2**, and **Lost**, the MR-PLL achieved the best performance. Specifically, the mean classification accuracy of MR-PLL is higher than that of the best comparative method by 3.05%, 6.01%, 1.59%, 2.19%. As shown in Table 2, compared with the PLL algorithm designed based on the loss function, MR-PLL consistently performs better than PRODEN, LW, CC, and CAVL in the benchmark dataset. Compared with VALEN, DIRK, and LS-PLL, which have more complex structures, MR-PLL still shows better disambiguation performance and is superior to other methods on the **MNIST**, **F-MNIST**, and **K-MNIST** datasets. We attribute this satisfactory result to the design of the MR-PLL introduction of mutual information guided reinforcement learning disambiguation, which adaptively adjusts the confidence of soft label propagation based on the dependency between feature representations, and explores potential disambiguation strategies through reinforcement learning to gradually improve the label disambiguation ability of the model.

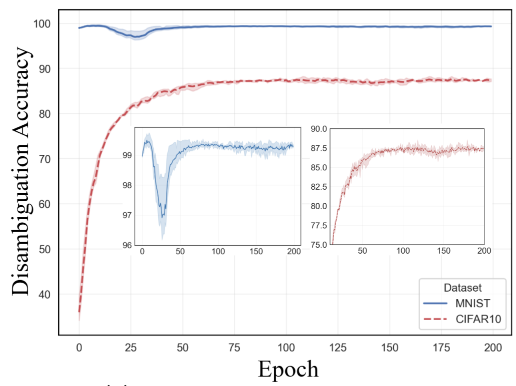
Analysis of ablation study. To evaluate the effectiveness of each component in MR-PLL, we conduct ablation studies on **CIFAR-10** ($q = 0.3$) and **BirdSong**. Specifically, ablation experiments were performed on four specific parts, i.e., MR-PLL-w/o-I removes the MI term from the reward and loss function, MR-PLL-w/o-$\mathcal{L}_{\mathrm{KL}}$ drops the label consistency reward, MR-PLL-w/o-Abs removes the abstain action and force the policy to choose between retain and modify, and MR-PLL-w/o-SLF drops MI guided soft label fusion. As reported in Table 3, all variants perform worse than the full MR-PLL model. Removing the MI term leads to a noticeable accuracy drop, demonstrating the importance of MI in exploiting latent semantic information for label disambiguation. Excluding $\mathcal{L}_{\mathrm{KL}}$ weakens label consistency during policy exploration, reducing robustness. Eliminating the abstain ac-

Dataset	q	Method							
		MR-PLL	PRODEN	LW	VALEN	CC	CAVL	DIRK	LS-PLL
CIFAR-10	0.1	90.49 ± 0.14%	84.12 ± 0.36%	81.43 ± 0.13%	83.21 ± 0.10%	85.35 ± 0.19%	83.37 ± 0.07%	**92.41 ± 0.23%**	84.31 ± 0.45%
	0.3	88.82 ± 0.27%	83.36 ± 0.46%	80.95 ± 0.17%	82.87 ± 0.26%	82.50 ± 0.24%	77.53 ± 0.34%	**92.14 ± 0.12%**	79.05 ± 0.52%
	0.5	84.72 ± 0.23%	77.52 ± 0.18%	78.72 ± 0.17%	81.88 ± 0.20%	79.33 ± 0.41%	73.25 ± 0.17%	**91.34 ± 0.14%**	77.82 ± 0.36%
MNIST	0.1	**99.62 ± 0.32%**	98.91 ± 0.03%	98.76 ± 0.03%	98.75 ± 0.02%	98.57 ± 0.38%	98.95 ± 0.04%	98.92 ± 0.25%	98.83 ± 0.24%
	0.3	**99.39 ± 0.12%**	98.71 ± 0.02%	98.64 ± 0.02%	98.66 ± 0.03%	98.28 ± 0.23%	98.90 ± 0.15%	98.54 ± 0.29%	98.52 ± 0.36%
	0.5	**99.25 ± 0.13%**	98.67 ± 0.03%	98.41 ± 0.03%	98.32 ± 0.02%	97.86 ± 0.19%	98.71 ± 0.04%	98.31 ± 0.13%	98.39 ± 0.51%
F-MNIST	0.1	**94.41 ± 0.32%**	88.27 ± 0.03%	88.18 ± 0.05%	90.72 ± 0.06%	90.83 ± 0.17%	90.32 ± 0.08%	93.71 ± 0.10%	90.49 ± 0.29%
	0.3	**93.45 ± 0.41%**	87.83 ± 0.04%	88.02 ± 0.04%	90.16 ± 0.06%	89.92 ± 0.41%	89.77 ± 0.04%	92.10 ± 0.19%	89.78 ± 0.25%
	0.5	**92.81 ± 0.27%**	87.01 ± 0.03%	87.66 ± 0.06%	89.51 ± 0.04%	88.76 ± 0.16%	88.92 ± 0.11%	91.23 ± 0.24%	84.09 ± 0.56%
K-MNIST	0.1	**97.58 ± 0.34%**	91.11 ± 0.03%	92.78 ± 0.04%	92.67 ± 0.06%	92.52 ± 0.26%	93.73 ± 0.16%	97.13 ± 0.32%	92.52 ± 0.37%
	0.3	**96.87 ± 0.23%**	90.67 ± 0.04%	91.28 ± 0.03%	91.75 ± 0.22%	91.17 ± 0.04%	93.57 ± 0.11%	96.49 ± 0.47%	86.38 ± 0.34%
	0.5	**96.36 ± 0.41%**	89.00 ± 0.05%	86.55 ± 0.06%	90.40 ± 0.03%	89.44 ± 0.48%	91.57 ± 0.22%	96.13 ± 0.51%	78.85 ± 0.54%

Table 2: Comparative disambiguation accuracy (mean ± std) across datasets.



(a) T-SNE Visualization of the **CIFAR-10** dataset.



(b) Convergence results of MR-PLL.

Figure 2: The effectiveness experimental results of MR-PLL.

tion degrades performance, indicating that postponing decisions on hard instances helps prevent error accumulation and improves stability. In addition, removing soft label fusion results in lower accuracy, verifying its effectiveness in providing reliable supervision signals for disambiguation. Overall, the superior performance of MR-PLL confirms that each component contributes positively to the disambiguation process.

Visualization of MI estimation in MR-PLL. To further illustrate the effectiveness of MR-PLL, we visualize the learned feature representations on **CIFAR-10** using T-SNE [Maaten and Hinton, 2008]. As shown in Figure 2(a), MR-PLL learns more discriminative features with clearer class boundaries and better clustering structures. By maximizing MI, samples from the same class become more compact, while features from different classes are better separated. This enhanced representation facilitates similarity estimation among instances, thereby improving ambiguous label detection and disambiguation. Moreover, MI estimation provides a reliable signal for assessing the credibility of candidate labels and dynamically refining the disambiguation process. The confusion matrix further confirms the effectiveness of MR-PLL in achieving accurate label disambiguation.

Uncertainty Quantification of MR-PLL. Existing machine learning algorithms not only require high accuracy, but also must reflect the effectiveness and reliability of the model [Guo *et al.*, 2017]. As shown in Figure 3, we analyzed the output uncertainty of various PLL methods. The results demonstrate that MR-PLL not only significantly mitigates the impact of label ambiguity but also enhances model robustness and reliability. This leads to more credible model outputs and yields competitive *expected calibration error* (ECE) scores. The results indicate that MR-PLL can enable agents to learn the best disambiguation behavior through trial and error under the guidance of rewards through RL. Meanwhile, the uncertainty estimation of MI is used to guide the reward function and adjust the update of candidate labels. This not only effectively reduces the impact of label ambiguity, but also makes the model more robust and reliable, thereby improving the credibility of the model output and achieving a competitive ECE score.

Convergence analysis. To prove the convergence of MR-PLL more intuitively, this section analyzes the convergence curves of MR-PLL on the **CIFAR-10** and **MNIST** datasets with $q = 0.3$. From Figure 2(b), MR-PLL achieves a good disambiguation effect. With the increase of learning iterations, MR-PLL approaches the steady-state value, so the convergence of the model can be proved through experiments. Meanwhile, from the comprehensive experimental test results, it can be seen that MR-PLL not only has excellent dis-

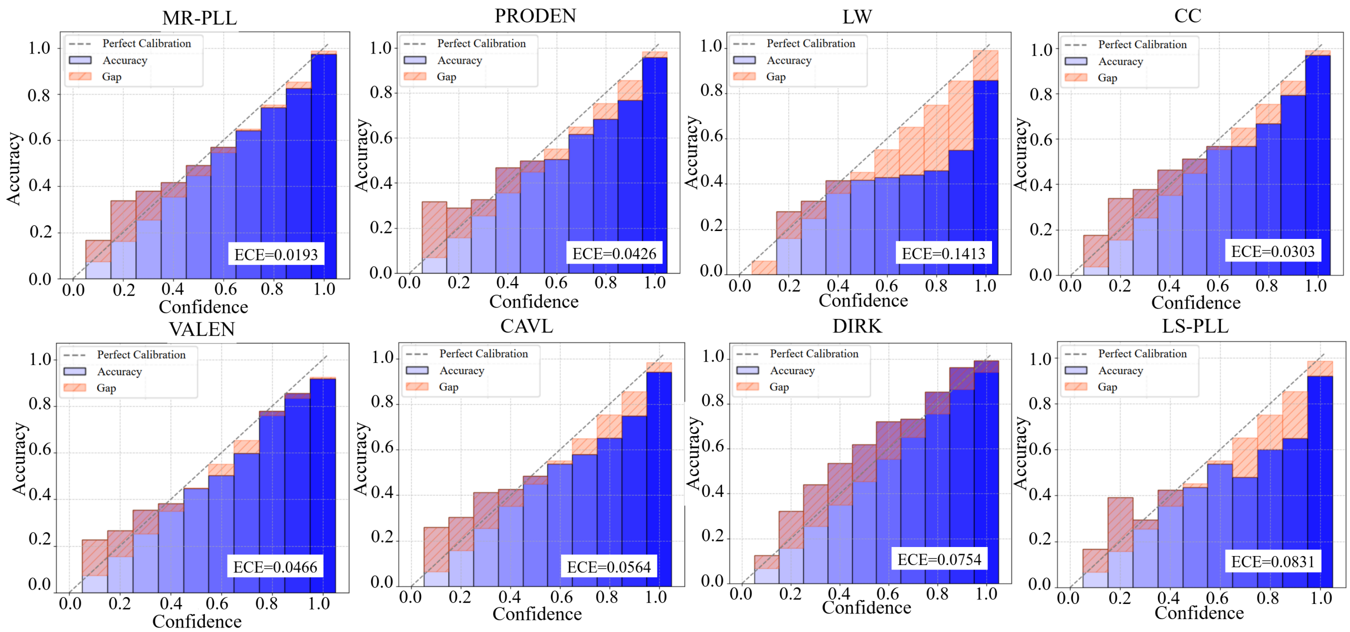


Figure 3: Reliability diagram and expected calibration error on **CIFAR-10**.

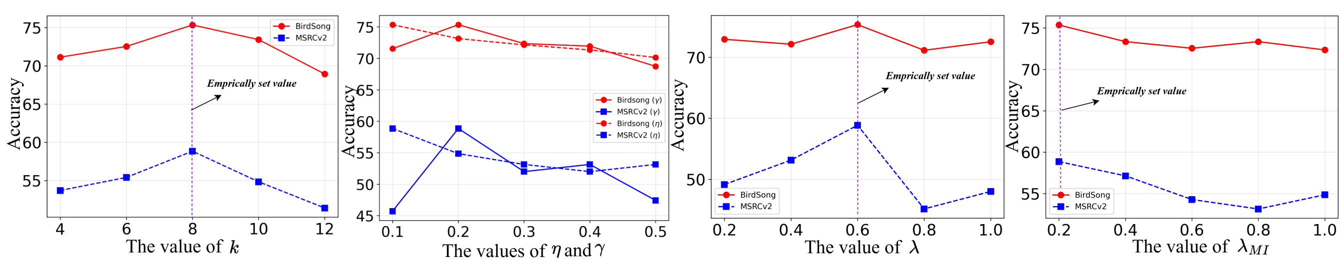


Figure 4: Parameter sensitivity analysis of MR-PLL.

Method	CIFAR-10	BirdSong
MR-PLL-w/o-I	$85.51 \pm 0.42\%$	$71.34 \pm 0.38\%$
MR-PLL-w/o-$\mathcal{L}_{\text{KL}}$	$86.14 \pm 0.34\%$	$73.54 \pm 0.27\%$
MR-PLL-w/o-Abs	$85.18 \pm 0.22\%$	$73.34 \pm 0.34\%$
MR-PLL-w/o-SLF	$83.59 \pm 0.39\%$	$70.9 \pm 0.25\%$
MR-PLL	$\mathbf{88.18 \pm 0.27\%}$	$\mathbf{75.35 \pm 0.34\%}$

Table 3: Ablation study of MR-PLL.

ambiguation performance but also has good stability.

Parameter sensitivity analysis. The MR-PLL involves main parameters. i.e. k, η, γ, λ and λ_{MI}. Figure 4 illustrates the performance of MR-PLL under various parameter configurations. For clarity, results are shown only on the **BirdSong** and **MSRCv2** datasets. In these tests, one parameter is changed at a time, while the other parameters are fixed at their default values. As shown in Figure 4, MR-PLL achieves optimal performance at $k = 8$, $\eta = 0.1$, $\gamma = 0.2$, $\lambda = 0.6$ and $\lambda_{MI} = 0.2$. These results further demonstrate the effectiveness of the MR-PLL algorithm among its various components. Meanwhile, the per-epoch computational complexity of the MR-PLL algorithm can reach $\mathcal{O}(N \cdot (d + Q + dlogN))$. The detailed calculation process is shown in **Appendix B.4**.

5 Conclusion

In this paper, we proposed a novel *mutual information guided reinforcement learning framework for partial label disambiguation* (MR-PLL). The highlight of MR-PLL is that it formulated partial label disambiguation as a sequential Markov decision process to enable dynamic correction of ambiguous labels through reinforcement learning. The integration of mutual information guided rewards and an abstention mechanism allows the model to perform reliable disambiguation while effectively mitigating bias accumulation caused by error disambiguation decisions. The method is superior to the state-of-the-art methods. Experiments on both benchmark and real-world datasets demonstrate its effectiveness. This work highlights the potential of combining reinforcement learning and information-theoretic guidance for addressing PLL and opens new directions for decision-theoretic weakly supervised learning.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Nos. 62401309, 62376130, 62401362), the China Postdoctoral Science Foundation (No. 2025M771692), the Natural Science Foundation of Shandong Province and Qingdao (No. 2R2024QF119, No. 24-4-4-zrjj-89-jch), and open project (No. 2024PY030).

References

[Amin *et al.*, 2021] Susan Amin, Maziar Gomrokchi, Harsh Satija, Herke Van Hoof, and Doina Precup. A survey of exploration methods in reinforcement learning. *arXiv preprint arXiv:2109.00157*, 2021.

[Berg *et al.*, 2004] Tamara Berg, Alexander Berg, Jaety Edwards, and David Forsyth. Who's in the picture. *Advances in neural information processing systems*, 17, 2004.

[Briggs *et al.*, 2012] Forrest Briggs, Xiaoli Z Fern, and Raviv Raich. Rank-loss support instance machines for miml instance annotation. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 534–542, 2012.

[Chai *et al.*, 2019] Jing Chai, Ivor W Tsang, and Weijie Chen. Large margin partial label machine. *IEEE Transactions on Neural Networks and Learning Systems*, 31(7):2594–2608, 2019.

[Chen *et al.*, 2013] Yi-Chen Chen, Vishal M Patel, Jaishanker K Pillai, Rama Chellappa, and P Jonathon Phillips. Dictionary learning from ambiguously labeled data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 353–360, Portland Oregon, USA, 2013.

[Christiano *et al.*, 2017] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.

[Clanuwat *et al.*, 2018] Tarin Clanuwat, Mikel Bober-Irizar, Asanobu Kitamoto, Alex Lamb, Kazuaki Yamamoto, and David Ha. Deep learning for classical japanese literature. *arXiv preprint arXiv:1812.01718*, 2018.

[Cour *et al.*, 2011] Timothee Cour, Ben Sapp, and Ben Taskar. Learning from partial labels. *The Journal of Machine Learning Research*, 12:1501–1536, 2011.

[Fan *et al.*, 2021] Jinfu Fan, Yang Yu, Zhongjie Wang, and Jinyi Gu. Partial label learning based on disambiguation correction net with graph representation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[Fan *et al.*, 2022] Jinfu Fan, Yang Yu, and Zhongjie Wang. Addressing label ambiguity imbalance in candidate labels: Measures and disambiguation algorithm. *Information Sciences*, 612:1–19, 2022.

[Fan *et al.*, 2023] Jinfu Fan, Zhencun Jiang, Yuanqing Xian, and Zhongjie Wang. A multi-class partial hinge loss for partial label learning. *Applied Intelligence*, 53(23):28333–28348, 2023.

[Fan *et al.*, 2024] Jinfu Fan, Linqing Huang, Chaoyu Gong, Yang You, Min Gan, and Zhongjie Wang. Kmt-pll: K-means cross-attention transformer for partial label learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[Fan *et al.*, 2026a] Jinfu Fan, Jiangnan Li, Linqing Huang, Min Gan, and CL Philip Chen. Partial label zero-shot learning with semantic mining and instance–label alignment. *Engineering Applications of Artificial Intelligence*, 177:114854, 2026.

[Fan *et al.*, 2026b] Jinfu Fan, Jiangnan Li, Xiaohui Zhong, Kangrui Ren, Zhencun Jiang, Min Gan, Tianhao Gu, and Linqing Huang. Evidential deep partial label learning to quantify disambiguation uncertainty. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24770–24779, 2026.

[Fan *et al.*, 2026c] JinFu Fan, Xiaohui Zhong, Kangrui Ren, JiangNan Li, Linqing Huang, Min Gan, and CL Philip Chen. Diffusion disambiguation models for partial label learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2026.

[Feng *et al.*, 2020] Lei Feng, Jiaqi Lv, Bo Han, Miao Xu, Gang Niu, Xin Geng, Bo An, and Masashi Sugiyama. Provably consistent partial-label learning. *Advances in neural information processing systems*, 33:10948–10960, 2020.

[Feng *et al.*, 2025] Jian Feng, Linqing Huang, JiangNan Li, Kangrui Ren, JinFu Fan, QingKai Bu, Min Gan, and CL Philip Chen. Partial label learning via mutual information representation learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

[Fuchs *et al.*, 2024] Tobias Fuchs, Florian Kalinke, and Klemens Böhm. Partial-label learning with a reject option. *arXiv preprint arXiv:2402.00592*, 2024.

[Gong *et al.*, 2024] Xiuwen Gong, Nitin Bisht, and Guandong Xu. Does label smoothing help deep partial label learning? In *Forty-first International Conference on Machine Learning*, 2024.

[Guillaumin *et al.*, 2010] Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid. Multiple instance metric learning from automatically labeled bags of faces. In *European conference on computer vision*, pages 634–647. Springer, 2010.

[Guo *et al.*, 2017] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.

[Kaelbling *et al.*, 1996] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

[Kober *et al.*, 2013] Jens Kober, J Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238–1274, 2013.

[Krizhevsky *et al.*, 2009] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[LeCun *et al.*, 1998] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[Li *et al.*, 2025] Zhiyong Li, Linqing Huang, Tianhao Gu, Qingkai Bu, Fuyu Qi, and Jinfu Fan. Partial multi-label learning via k-means graph transformer. *Knowledge-Based Systems*, 325:114017, 2025.

[Liu and Dietterich, 2012] Liping Liu and Thomas Dietterich. A conditional multinomial mixture model for superset label learning. *Advances in neural information processing systems*, 25, 2012.

[Luo and Orabona, 2010] Jie Luo and Francesco Orabona. Learning from candidate labeling sets. *Advances in Neural Information Processing Systems*, 23:1504–1512, 2010.

[Lv *et al.*, 2020] Jiaqi Lv, Miao Xu, Lei Feng, Gang Niu, Xin Geng, and Masashi Sugiyama. Progressive identification of true labels for partial-label learning. In *international conference on machine learning*, pages 6500–6510. PMLR, 2020.

[Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[Ouyang *et al.*, 2022] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

[Paszke *et al.*, 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[Poole *et al.*, 2019] Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In *International conference on machine learning*, pages 5171–5180. PMLR, 2019.

[Sutton *et al.*, 1998] Richard S Sutton, Andrew G Barto, et al. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

[Wang *et al.*, 2023] Haobo Wang, Ruixuan Xiao, Yixuan Li, Lei Feng, Gang Niu, Gang Chen, and Junbo Zhao. Pico+: Contrastive label disambiguation for robust partial label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Wen *et al.*, 2021] Hongwei Wen, Jingyi Cui, Hanyuan Hang, Jiabin Liu, Yisen Wang, and Zhouchen Lin. Leveraged weighted loss for partial label learning. In *International conference on machine learning*, pages 11091–11100. PMLR, 2021.

[Wu *et al.*, 2024] Dong-Dong Wu, Deng-Bao Wang, and Min-Ling Zhang. Distilling reliable knowledge for instance-dependent partial label learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 15888–15896, 2024.

[Xiao *et al.*, 2017] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

[Xu *et al.*, 2021] Ning Xu, Congyu Qiao, Xin Geng, and Min-Ling Zhang. Instance-dependent partial label learning. *Advances in Neural Information Processing Systems*, 34:27119–27130, 2021.

[Xu *et al.*, 2023] Ruiyang Xu, Jalaj Bhandari, Dmytro Korenkevych, Fan Liu, Yuchen He, Alex Nikulkov, and Zheqing Zhu. Optimizing long-term value for auction-based recommender systems via on-policy reinforcement learning. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 955–962, 2023.

[Xu *et al.*, 2024] Ning Xu, Congyu Qiao, Yuchen Zhao, Xin Geng, and Min-Ling Zhang. Variational label enhancement for instance-dependent partial label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[Yao *et al.*, 2020] Yao Yao, Jiehui Deng, Xiuhua Chen, Chen Gong, Jianxin Wu, and Jian Yang. Deep discriminative cnn with temporal ensembling for ambiguously-labeled image classification. In *Proceedings of the aaai conference on artificial intelligence*, volume 34, pages 12669–12676, 2020.

[Yao *et al.*, 2021] Yao Yao, Chen Gong, Jiehui Deng, and Jian Yang. Network cooperation with progressive disambiguation for partial label learning. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part II*, pages 471–488. Springer, 2021.

[Zeng *et al.*, 2013] Zinan Zeng, Shijie Xiao, Kui Jia, Tsung-Han Chan, Shenghua Gao, Dong Xu, and Yi Ma. Learning by associating ambiguously labeled images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 708–715, 2013.

[Zhang *et al.*, 2016] Min-Ling Zhang, Bin-Bin Zhou, and Xu-Ying Liu. Partial label learning via feature-aware disambiguation. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1335–1344, San Francisco, USA, 2016.

[Zhang *et al.*, 2021] Fei Zhang, Lei Feng, Bo Han, Tongliang Liu, Gang Niu, Tao Qin, and Masashi Sugiyama. Exploiting class activation value for partial-label learning. In *International conference on learning representations*, 2021.