CA2C: A Prior-Knowledge-Free Approach for Robust Label Noise Learning via Asymmetric Co-learning and Co-training

Mengmeng Sheng[1], Zeren Sun[1]*, Tianfei Zhou[2], Xiangbo Shu[1], Jinshan Pan[1], Yazhou Yao[1,3]*
[1]Nanjing University of Science and Technology, Nanjing, China
[2]Beijing Institute of Technology, Beijing, China
[3]State Key Laboratory of Intelligent Manufacturing of Advanced Construction Machinery, Nanjing, China

Abstract

Label noise learning (LNL), a practical challenge in real-world applications, has recently attracted significant attention. While demonstrating promising effectiveness, existing LNL approaches typically rely on various forms of prior knowledge, such as noise rates or thresholds, to sustain performance. This dependence limits their adaptability and practicality in real-world scenarios where such priors are usually unavailable. To this end, we propose a novel LNL approach, termed ***CA2C*** *(****C****ombined* ***A****symmetric* ***C****o-learning and* ***C****o-training), which alleviates the reliance on prior knowledge through an integration of complementary learning paradigms. Specifically, we first introduce an asymmetric co-learning strategy with paradigm deconstruction. This strategy trains two models simultaneously under distinct learning paradigms, harnessing their complementary strengths to enhance robustness against noisy labels. Then, we propose an asymmetric co-training strategy with cross-guidance label generation, wherein knowledge exchange is facilitated between the twin models to mitigate error accumulation. Moreover, we design a confidence-based re-weighting approach for label disambiguation, enhancing robustness against potential disambiguation failures. Extensive experiments on synthetic and real-world noisy datasets demonstrate the effectiveness and superiority of CA2C. Our source code has been made available at* https://github.com/NUST-Machine-Intelligence-Laboratory/CA2C.

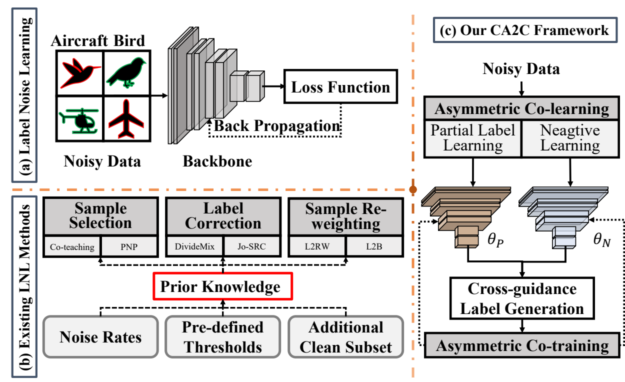

Figure 1. (a) Label noise learning primarily aims to mitigate the negative effects of noisy labels on model training, a challenge that is inevitable in real-world scenarios. (b) Existing LNL methods (*e.g.*, sample selection, label correction, and sample re-weighting) often face challenges in real-world applications due to their strong dependence on prior knowledge (*e.g.*, noise rates, pre-defined thresholds, or additional clean subsets) to sustain performance. (c) Contrarily, our CA2C deconstructs the LNL problem using a combined asymmetric co-learning and co-training framework, eliminating the demand for strong prior knowledge.

1. Introduction

In the past decade, deep neural networks (DNNs) have garnered substantial attention due to their remarkable performance in various applications, including image recognition [17, 20, 33, 35, 68], semantic segmentation [5, 6, 51, 67], and object detection [4, 28, 37, 39]. Among all factors contributing to these successes, the availability of large-scale, high-quality human-labeled datasets (*e.g.*, ImageNet [9]) serves as one of the most pivotal catalysts for these advancements. However, in real-world applications, acquiring accurately annotated data is time-consuming and expensive [59]. Accordingly, methods such as web crawlers, questionnaires, and crowd-sourcing have become widely adopted strategies for collecting large volumes of labeled data [57]. Nevertheless, these strategies often introduce varying degrees of mislabeled data (*i.e.*, label noise), which pose significant challenges for training deep learning models [70]. Label noise can severely degrade model performance, as deep neural networks (DNNs) possess remarkable capacity that empowers them to memorize and fit even incorrectly labeled samples [58], as shown in Fig. 1(a). Consequently,

*Corresponding Author.

developing robust methods for learning with noisy labels has become crucial in enhancing model generalization and ensuring reliable performance in real-world applications.

Driven by the *memorization effect* [2] (*i.e.*, DNNs tend to first fit clean samples and progressively memorize noisy ones), existing LNL approaches have evolved in three main categories: sample selection, label correction, and sample re-weighting. Sample selection methods aim to establish effective criteria for selecting a more precise "clean" subset to train the model while discarding "noisy" samples (*e.g.*, Co-teaching [14] and JoCoR [55]). Label correction approaches generally seek to rectify sample labels using noise transition matrices [12] or model predictions (*e.g.*, Jo-SRC [63] and PENCIL [66]). Sample re-weighting methods involve assigning larger weights to samples with higher confidence of being clean and smaller weights to those with lower confidence (*e.g.*, L2RW [38] and L2B [76]). However, existing approaches tend to demand various forms of prior knowledge, such as pre-defined drop rates, thresholds, or access to a clean validation subset. Consequently, these methods struggle to effectively handle real-world noisy scenarios where such prior knowledge is typically unavailable, as illustrated in Fig. 1(b).

Recently, researchers have explored various novel training and learning strategies [13, 40–42, 48]. For example, the symmetric co-training (SCT) strategy is proposed to enhance clean sample selection [14, 22, 31, 49, 55, 69]. Typical SCT methods train two models simultaneously with identical architectures but different initializations, enabling them to extract diverse knowledge from the training data and provide mutual guidance. However, these methods still require prior knowledge. Furthermore, the additional information gained from SCT is limited, as differences in the capabilities of the twin models primarily stem from their distinct initializations. The asymmetric co-training (ACT) [42] strategy is another attempt aimed at enhancing clean sample selection through asymmetrically training twin models with different strategies. However, ACT [42] is still limited in extracting diverse learning information between models, as the distinctions between the twin models depend solely on differences in their training data volumes.

To address the aforementioned issues, we propose a prior-knowledge-free LNL approach termed **CA2C**. Our CA2C deconstructs the traditional LNL paradigm with a **C**ombined **A**symmetric **C**o-learning and **C**o-training framework, as shown in Fig. 1(c). Firstly, we propose an asymmetric co-learning strategy with paradigm deconstruction, where twin models (*i.e.*, *P-model* and *N-model*) with identical architectures are trained simultaneously in different manners. The *P-model* follows a partial label learning paradigm, while the *N-model* is trained using negative learning. As such, our method enables better exploration of different learning capabilities between twin models while mitigating direct exposure to noisy labels during training. Then, to alleviate error accumulations, we introduce an asymmetric co-training strategy with cross-guidance label generation to facilitate mutual reinforcement between twin models through their distinct learning capabilities. Furthermore, we design a confidence-based re-weighting strategy for label disambiguation to improve *P-model*'s tolerance to disambiguation failures. Comprehensive experimental results and extensive ablation studies validate the effectiveness and superiority of our CA2C on both synthetical and real-world benchmarks. The main contributions are summarized as follows:

- We propose a novel approach for learning with noisy labels, termed CA2C. This method introduces a combined asymmetric co-learning and co-training framework that eliminates demands for strong prior knowledge. Our CA2C trains twin models simultaneously using distinct learning paradigms, leveraging their complementary strengths to improve robustness against noisy labels.
- We design an asymmetric co-training strategy that incorporates cross-guidance label generation. This strategy promotes knowledge exchange between twin models with distinct learning capabilities while reducing error accumulations within each network. Moreover, we introduce a confidence-based re-weighting strategy to enhance the effectiveness of partial label learning.
- Comprehensive experiments and ablation studies on both synthetically corrupted and real-world datasets demonstrate that CA2C exhibits remarkable robustness across various noise settings.

2. Related Work

2.1. Label Noise Learning

Existing methods for learning with noisy labels can be primarily categorized into three groups: sample selection, label correction, and sample re-weighting [7, 8, 11, 16, 23–25, 29, 32, 38, 53, 56, 62, 71, 73–75].

Sample Selection. Sample selection is a primary direction in label noise learning, aimed at designing robust criteria to identify clean samples for model training while discarding noisy ones [3, 27, 42, 60, 61]. Inspired by the *memorization effect* of DNNs, early researchers propose the small-loss criterion based on the observation that correctly labeled samples typically exhibit lower loss values. For example, Co-teaching [14] trains two networks simultaneously, guiding each other in selecting small-loss samples for training. DivideMix [22] utilizes Gaussian Mixture Models to fit the loss distributions of clean and noisy samples, thereby partitioning the dataset. However, these methods usually rely on prior knowledge (*e.g.*, a pre-defined drop rate or threshold) to assist in determining the proportion of clean samples, which limits their practicality in real-world scenarios.

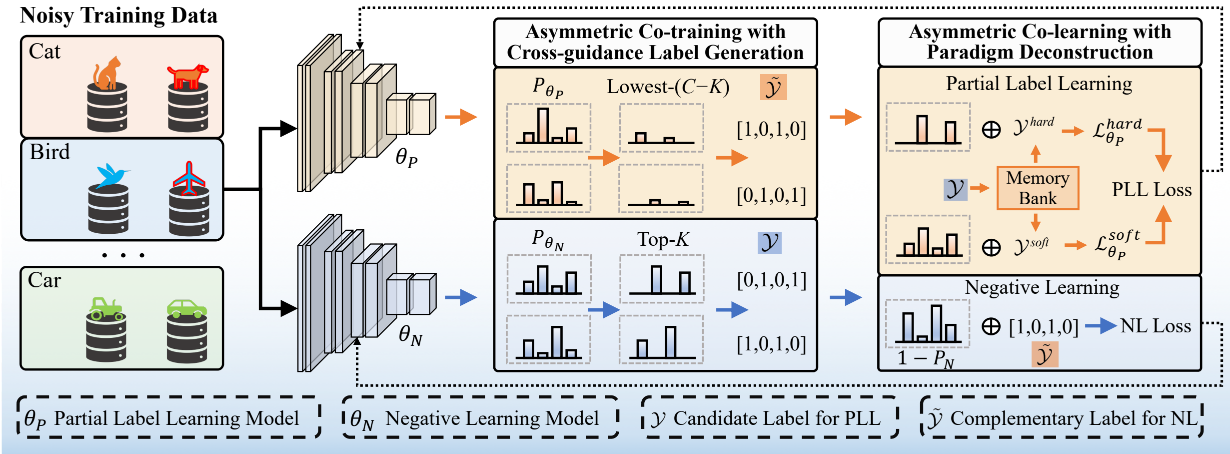

Figure 2. The overall framework of our proposed CA2C. In our CA2C, twin models (*i.e.*, θ_P and θ_N) with identical architectures are trained simultaneously but employ distinct learning paradigms: partial label learning and negative learning. To promote knowledge exchange between θ_P and θ_N, we exploit the paradigm independence inherent in our asymmetric co-learning strategy by using each model's predictions to cross-generate label spaces. For the *P-model*, we implement a memory bank to track the frequency of $\mathcal{Y}$ and design a confidence-based re-weighting strategy for label disambiguation, enhancing θ_P's robustness against disambiguation failures. For the *N-model*, we use the complementary labels $\tilde{\mathcal{Y}}$ generated from the *P-model* for negative learning.

Label Correction. Another effective approach for learning with noisy labels is to correct them before feeding them into networks [1, 50, 63, 65, 66]. Existing works often seek to correct corrupted labels by estimating noise transition matrices [12, 34] or using model predictions to generate pseudo-labels [48, 63]. For instance, Goldberger *et al.* [12] proposes to estimate the noise transition matrix with an additional layer. Jo-SRC [63] attempts to generate pseudo labels for noisy samples using a temporally averaged model. However, estimating noise transition matrices is challenging in real-world tasks, while unreliable pseudo-label estimation can lead to harmful error accumulations.

Sample Re-weighting. Some researchers have explored re-weighting training instances to mitigate the negative impact of noisy labels [10, 38, 43, 52, 54]. The idea behind sample re-weighting is straightforward: assign higher weights to clean samples and lower weights to noisy ones, thereby mitigating the model's tendency to overfit noisy data. For instance, L2RW [38] introduces a method for assigning sample weights using meta-learning. Similarly, L2B [76] dynamically adjusts importance weights between observed and generated labels. However, they rely heavily on dataset-specific prior knowledge, such as a clean validation subset. This limits their practicality in real-world scenarios.

2.2. Symmetric and Asymmetric Co-training

The symmetric co-training strategy has been widely used in existing sample selection methods [14, 22, 31, 55, 63, 69]. These approaches train two models simultaneously to extract additional information by enabling them to guide each other during the learning process, thereby enhancing the accuracy of clean sample selection. Unfortunately, as training progresses, models using SCT strategies tend to converge, ultimately causing the additional learning benefits to vanish, because their differences arise solely from random initialization. In contrast, ACT [42] introduces an asymmetric co-training strategy, where one model is trained on a selected clean subset, while the other is trained on the entire noisy dataset. However, ACT struggles to effectively capture diverse learning representations between models, as its design primarily hinges on varying the training data volumes assigned to each model. Moreover, it cannot guarantee that all samples in the selected "clean" subset are genuinely clean during training. In this paper, we propose to address the LNL problem by employing a combined asymmetric co-learning and co-training framework, which is more suitable for challenging real-world noisy scenarios.

3. Methods

3.1. Problem Statement

Considering a multi-class classification problem, let the feature space be denoted as $\mathcal{X} \in \mathbb{R}^d$ with d dimensions and $Y = \{1, 2, ..., C\}$ represents the label space with C distinct classes. In label noise learning, we are given a training set $D = \{(x_n, y_n)\}_{n=1}^N$ with N instances, where each sample $x_n \in \mathcal{X}$ and its one-hot noisy label $y_n \in \{0,1\}^C$. We assume that y_n^* represents the true label of x_n. $D_{test} = \{(x_n, y_n^*) | n = 1, ..., M\}$ is defined as the test set with M clean samples. Our goal is to train a robust classification model $\mathcal{F}(\cdot, \theta)$ on the noisy training set D to achieve higher prediction accuracy on the clean test set D_{test}, where θ denotes model parameters. However, the presence of noisy labels will misguide the optimization of the cross-entropy loss used in conventional classification problems:

$$\mathcal{L}_{ce} = -\frac{1}{N}\sum_{n=1}^{N} y_n \log(p(x_n, \theta)), \tag{1}$$

where $p(x_n, \theta)$ denotes the prediction of the classifier for the n-th training sample x_n.

3.2. Asymmetric Co-learning with Paradigm Deconstruction

We propose an asymmetric co-learning strategy with paradigm deconstruction. This strategy eliminates the requirement for strong prior knowledge, such as noise rates or selection thresholds, thereby enhancing its applicability in real-world scenarios. Specifically, two models with identical architectures (*i.e.*, *P-model* θ_P and *N-model* θ_N) are trained simultaneously but employ distinct training strategies. Unlike ACT [42], which differentiates the twin models solely by varying the volume of training data provided to each model, our approach employs different learning paradigms to fully exploit their unique learning capabilities. As demonstrated in NPN [40], the LNL problem can be decomposed into a combination of partial label learning (PLL) and negative learning (NL). Inspired by this insight, we apply partial label learning to the *P-model*, using candidate labels $\mathcal{Y}$ to guide its training. Contrarily, we employ complementary labels $\tilde{\mathcal{Y}}$ to train the *N-model* through negative learning. Losses used for optimizing θ_P and θ_N are:

$$\mathcal{L}_{\theta_P} = -\frac{1}{N}\sum_{n=1}^{N} \mathcal{Y}_n \log(p(x_n, \theta_P)), \tag{2}$$

$$\mathcal{L}_{\theta_N} = -\frac{1}{N}\sum_{n=1}^{N} \tilde{\mathcal{Y}}_n \log(1 - p(x_n, \theta_N)). \tag{3}$$

Notably, to achieve paradigm deconstruction and enable the learning of the twin models via Eqs. (2) and (3), we first decompose the given (noisy) label space Y into candidate labels $\mathcal{Y}$ and complementary labels $\tilde{\mathcal{Y}}$. Specifically, the Top-K categories with the highest confidence are selected to construct candidate labels $\mathcal{Y}$ for PLL, while the remaining reliable non-candidate labels are utilized to generate the complementary labels $\tilde{\mathcal{Y}}$ for NL. Accordingly, our method avoids direct exposure to noisy labels, which is an inherent issue in typical LNL approaches. The candidate labels $\mathcal{Y}_n$ used in Eq. (2) is defined as follows:

$$\begin{aligned} &\mathcal{Y}_n = \{\hat{y}_n^1, ..., \hat{y}_n^C\}, \\ &\hat{y}_n^c = \mathbb{1}_{c \in \{\text{Top-}K(\{p^1(x_n,\theta), ..., p^C(x_n,\theta)\})\}}. \end{aligned} \tag{4}$$

$\mathbb{1}_{(\cdot)}$ denotes the indicator function. Conversely, the complementary labels $\tilde{\mathcal{Y}}_n$ used in Eq. (3) is obtained as follows:

$$\tilde{\mathcal{Y}}_n = \neg \mathcal{Y}_n = \mathcal{I} - \mathcal{Y}_n. \tag{5}$$

$\mathcal{I}$ represents the entire label space.

Discussion: Although our method and NPN [40] both decompose the LNL problem into a combination of PLL and NL, they differ fundamentally. NPN is prone to error accumulations as it relies on a single model optimized using a joint loss. Moreover, NPN faces limitations in achieving performance improvements during the later stages of training. In contrast, our method introduces an asymmetric co-learning strategy, training two models using PLL and NL separately. This design mitigates the knowledge convergence issue that inherently arises from the mixed-paradigm learning in the single-model framework of NPN.

3.3. Asymmetric Co-training with Cross-guidance Label Generation

To enhance the performance of the *P-model* and *N-model*, we further propose an asymmetric co-training strategy with cross-guidance label generation. Each model facilitates the generation of supervision for its peer model based on the peer model's training mechanism. This cross-guidance-based label generation process facilitates knowledge exchange between the twin models trained under distinct paradigms, thereby mitigating error accumulation and improving overall model performance. Specifically, for the *N-model*, we construct its required complementary labels $\tilde{\mathcal{Y}}^*$ by selecting the categories with the Lowest-(C-K) predicted probabilities according to the prediction of the *P-model*. These categories are considered more confidently incorrect, thereby benefiting the reliability of the negative learning process. Accordingly, $\tilde{\mathcal{Y}}^*$ and $\mathcal{L}^*_{\theta_N}$ become:

$$\begin{aligned} &\tilde{\mathcal{Y}}^*_n = \{\hat{y}_n^1, ..., \hat{y}_n^C\}, \\ &\hat{y}_n^c = \mathbb{1}_{c \notin \{\text{Top-}K(\{p^1(x_n,\theta_P), ..., p^C(x_n,\theta_P)\})\}}, \end{aligned} \tag{6}$$

and

$$\mathcal{L}^*_{\theta_N} = -\frac{1}{N}\sum_{n=1}^{N} \tilde{\mathcal{Y}}^*_n \log(1 - p(x_n, \theta_N)). \tag{7}$$

Similarly, *N-model* generates candidate labels for the *P-model* to facilitate its partial label learning.

To improve the robustness of the *P-model* against disambiguation failures, we propose a confidence-based reweighting approach for label disambiguation. Specifically, we maintain a memory bank ($\mathcal{M}$) to track the frequency of candidate labels generated from *N-model* for each sample. This memory bank $\mathcal{M}$ not only aids in the selection of reliable candidate labels for label disambiguation but also serves as a proxy for confidence measurement, thereby mitigating the adverse effects of incorrect disambiguation. In the t-th epoch, $\mathcal{M}$ is updated as follows:

$$\mathcal{M}^t(x_n) = \begin{cases} 0, & if\ t = 0 \\ \mathcal{M}^{t-1}(x_n) + \mathcal{Y}_n^*, & if\ t \geq 1 \end{cases}, \tag{8}$$

in which,

$$\mathcal{Y}_n^* = \{\hat{y}_n^1, ..., \hat{y}_n^C\}, \\ \hat{y}_n^c = \mathbb{1}_{c \in \{\text{Top-}K(\{p^1(x_n,\theta_N),...,p^C(x_n,\theta_N)\})\}}. \tag{9}$$

To further leverage the memory bank to ensure the reliability of label disambiguation for the *P-model*, we calculate the confidence coefficient $\mathcal{W}(n)$ for the n-th sample's candidate labels as follows:

$$\mathcal{W}(n) = \frac{1}{\max(\mathcal{M}^t(x_n))}(\mathcal{M}^t(x_n)). \tag{10}$$

Hard disambiguation (*i.e.*, $\mathcal{Y}^{hard} = \arg\max\{\mathcal{M}^t(x)\}$) often encounters stability issues, whereas soft disambiguation (*i.e.*, $\mathcal{Y}^{soft} = \frac{\mathcal{M}^t(x)}{\text{sum}(\mathcal{M}^t(x))}$) is susceptible to underfitting. To the end, we propose a joint loss, which integrates hard disambiguation and soft disambiguation, for the training of the *P-model*. The loss function is as follows:

$$\begin{aligned} \mathcal{L}_{\theta_P}^* &= \lambda \mathcal{L}_{\theta_P}^{hard} + (1-\lambda)\mathcal{L}_{\theta_P}^{soft} \\ &= -\frac{\lambda}{N}\sum_{n=1}^{N} \mathcal{W}(n)\mathcal{Y}_n^{hard}\log(p(x_n,\theta_P)) \\ &- \frac{1-\lambda}{N}\sum_{n=1}^{N} \mathcal{W}(n)\mathcal{Y}_n^{soft}\log(p(x_n,\theta_P)). \end{aligned} \tag{11}$$

λ denotes a hyper-parameter that balances the weights of the hard and soft loss terms. By optimizing the *P-model* with this joint loss, we achieve complementary advantages from both hard and soft disambiguation, effectively mitigating their inherent limitations when applied independently.

Discussion: Our method and ACT [42] share a fundamental idea, wherein twin models undergo distinct training processes to develop asymmetric learning capabilities. However, ACT [42] achieves this differentiation by solely varying the amount of training data fed to each model, which imposes an inherent limitation on model performance. In contrast, our method introduces a novel approach to cultivate twin models with divergent capabilities by independently employing partial label learning and negative learning. This strategy effectively maximizes the utilization of all available training data, ultimately enhancing model performance. Resorting to Eqs. (6) and (9) for generating (candidate / complementary) labels under cross-guidance, our *P-model* and *N-model* exchange knowledge characterized by significant disparities arising from their divergent training processes, effectively guiding them toward optimal optimization.

Algorithm 1 The proposed CA2C

Input: Training dataset D, test dataset D_{test}, twin networks θ_P and θ_M, the number of warm-up epochs T_w and whole epochs T_{total}, the batch size bs.

for $epoch = 1, 2, \ldots, T_{total}$ **do**
  # Warmup Training.
  if $epoch \leq T_w$ **then**
    for $iteration = 1, 2, \ldots$ **do**
      Fetch $B = \{(x_n, y_n)\}_1^{bs}$ from D ;
      Calculate $\mathcal{L}_{\theta_P} = -\sum_{n=1}^{bs} y_n \text{log} p(x_n, \theta_P)$;
      Calculate $\mathcal{L}_{\theta_N} = -\sum_{n=1}^{bs} y_n \text{log} p(x_n, \theta_N)$.
      Update θ_P, θ_N by optimizing $\mathcal{L}_{\theta_P}$, $\mathcal{L}_{\theta_N}$.
    end for
  end if
  # Robust Training.
  if $T_w < epoch \leq T_{total}$ **then**
    for $iteration = 1, 2, \ldots$ **do**
      Fetch $B = \{(x_n, y_n)\}_1^{bs}$ from D
      # Cross-guidance Label Generation.
      Generate $\tilde{\mathcal{Y}}^*$ for NL based on Eq. (6);
      Obtain $\mathcal{M}$ for PLL based on Eq. (8);
      Generate $\mathcal{Y}^*$ for PLL based on Eq. (9);
      # Confidence-based Re-weighting Strategy.
      Obtain $\mathcal{W}$ for PLL based on Eq. (10);
      # Asymmetric Co-learning and Co-training.
      Calculate $\mathcal{L}_{\theta_N}^*$ based on Eq. (7);
      Calculate $\mathcal{L}_{\theta_P}^*$ based on Eq. (11);
      Update θ_P, θ_N by optimizing $\mathcal{L}_{\theta_P}^*$, $\mathcal{L}_{\theta_N}^*$.
    end for
  end if
end for

Output: The updated robust networks θ_P and θ_N.

3.4. The Overall Framework

In summary, we propose CA2C, a novel framework for learning with noisy labels, which integrates a combined asymmetric co-learning and co-training process. Unlike many existing LNL approaches, CA2C eliminates reliance on strong prior knowledge, such as predefined noise rates or selection thresholds, making it significantly more adaptable for real-world noisy scenarios where such prior information is often unavailable. First, we introduce the asymmetric co-learning strategy with paradigm deconstruction, where twin models are trained simultaneously under distinct learning paradigms. This design maximizes their divergence while exploiting their complementary strengths, thereby enhancing robustness against noisy labels. Then, we propose the asymmetric co-training strategy with cross-guidance label generation, which facilitates mutual guidance between twin models to mitigate error accumulations. Furthermore, we

Methods	Publication	CIFAR100N			CIFAR80N			Average
		Sym-20%	Sym-80%	Asym-40%	Sym-20%	Sym-80%	Asym-40%	
Standard	-	35.14	4.41	27.29	29.37	4.20	22.25	20.44
Decoupling	NeurIPS 2017	33.10	3.89	26.11	43.49	10.1	33.74	25.07
Co-teaching	NeurIPS 2018	43.73	15.15	28.35	60.38	16.59	42.42	34.44
Co-teaching+	ICML 2019	49.27	13.44	33.62	53.97	12.29	43.01	34.27
JoCoR	CVPR 2020	53.01	15.49	32.70	59.99	12.85	39.37	35.57
DivideMix	ICLR 2020	57.76	28.98	43.75	57.47	21.18	37.47	41.10
Jo-SRC	CVPR 2021	58.15	23.80	38.52	65.83	29.76	53.03	44.85
Co-LDL	TMM 2022	59.73	25.12	52.28	58.81	24.22	50.69	45.14
UNICON	CVPR 2022	55.10	31.49	49.90	54.50	36.75	51.50	46.54
SOP	ICML 2022	58.63	34.23	49.87	60.17	34.05	53.34	48.38
AGCE	TPAMI 2023	59.38	27.41	43.04	60.24	25.39	44.06	43.25
DISC	CVPR 2023	60.28	33.90	50.56	50.33	38.23	47.63	46.82
ANL	NeurIPS 2023	60.20	23.39	44.15	61.35	20.74	47.31	42.86
NPN	AAAI 2024	62.76	31.69	57.11	63.78	25.25	58.50	49.85
ACT	MM 2024	65.51	40.74	63.48	67.09	38.58	64.40	56.63
SED	ECCV 2024	66.50	38.15	58.29	69.10	**42.57**	60.87	55.91
Ours	-	**68.64**	**40.97**	**65.59**	**70.06**	40.47	**65.71**	**58.57**

Table 1. Average test accuracy (%) on CIFAR100N and CIFAR80N over the last ten epochs. Experiments are conducted under various noise conditions ("Sym" and "Asym" denote the symmetric and asymmetric label noise, respectively).

design a confidence-based re-weighting strategy for PLL, enhancing the robustness of the *P-model* against disambiguation failures. The overall architecture of our method is depicted in Algorithm 1 and Fig. 2, providing a comprehensive illustration of its entire pipeline.

4. Experiments

In this section, we validate the effectiveness of our proposed method through experiments conducted on synthetic and real-world noisy datasets. Moreover, we extensively perform ablation studies to investigate the contribution of each component and to analyze the impact of hyper-parameters.

4.1. Experiment Setup

Synthetic Noisy Datasets: Following [40], we evaluate our CA2C approach using two synthetic noisy datasets (*i.e.*, CIFAR100N and CIFAR80N), both of which are derived from the CIFAR100 dataset [19]. They are specifically designed to simulate closed-set and open-set noisy scenarios, respectively. Adhering to [40], we designate the last 20 categories from CIFAR100 as out-of-distribution ones in CIFAR80N. We generate noisy labels by randomly corrupting the sample labels, changing them from their ground-truth categories to other categories based on a predefined noise rate η, using two types of synthetic label noise: symmetric (Sym.) and asymmetric (Asym.). Symmetric noise is generated by randomly replacing the original labels with all other possible classes. Asymmetric noise is a more realistic setting where labels are replaced by similar classes.
Real-World Noisy Datasets: We also evaluate the performance of our CA2C on three widely used real-world noisy benchmarks: Web-Aircraft, Web-Bird, and Web-Car [46], whose images are sourced from web image search engines. These three web datasets are subsets of the web-image-based fine-grained image dataset WebFG-496 [46]. Food-101N [21] is another real-world benchmark dataset and its estimated noise rate is about 20%. In contrast to the controllable noise settings in synthetic noisy datasets, real-world noisy datasets present greater challenges due to their complex and unknown noise patterns.
Experiment Settings: Following previous works [40–42, 48], we employ a seven-layer CNN as the backbone when evaluating on synthetic datasets. The models are trained using SGD with a momentum of 0.9, and the learning rates decay in a cosine annealing schedule. The batch size and the learning rate are set to 128 and 0.01. For benchmarks on real-world datasets, we leverage ResNet50 [15] pre-trained on ImageNet-1K as our backbone. More details about the settings are in our supplementary material.

Baselines: We compare our method with the following state-of-the-art (SOTA) LNL algorithms on synthetic noisy datasets: Decoupling [31], Co-teaching [14], Co-teaching+ [69], JoCoR [55], DivideMix [22], Jo-SRC [63], Co-LDL [47], UNICON [18], SOP [30], AGCE [75], DISC [26], ANL [64], NPN [40], ACT [42], and SED [41]. ("Standard" refers to the conventional training on the entire noisy dataset.) When evaluating on real-world noisy datasets, we additionally include the following methods into comparison: PENCIL [66], AFM [36], PLC [72], WarPI [44], CoDis [61], and VRI [45]. The results of these competing SOTAs in Tables 1, 2, and 3 are primarily obtained from SED and VRI. We report the mean for each case based on five independent runs with different random seeds. The best performances in these tables are bolded.

Methods	Publication	Backbone	Performances(%)			
			Web-Aircraft	Web-Bird	Web-Car	Average
Standard	-	ResNet50	60.80	64.40	60.60	61.93
Decoupling	NeurIPS 2017	ResNet50	75.91	71.61	79.41	75.64
Co-teaching	NeurIPS 2018	ResNet50	79.54	76.68	84.95	80.39
Co-teaching+	ICML 2019	ResNet50	74.80	70.12	76.77	73.90
PENCIL	CVPR 2019	ResNet50	78.82	75.09	81.68	78.53
JoCoR	CVPR 2020	ResNet50	80.11	79.19	85.10	81.47
AFM	ECCV 2020	ResNet50	81.04	76.35	83.48	80.29
DivideMix	ICLR 2020	ResNet50	82.48	74.40	84.27	80.38
Jo-SRC	CVPR 2021	ResNet50	82.73	81.22	88.13	84.03
Co-LDL	TMM 2022	ResNet50	81.97	80.11	86.95	83.01
UNICON	CVPR 2022	ResNet50	85.18	81.20	88.15	84.84
SOP	ICML 2022	ResNet50	84.06	79.40	85.71	83.06
AGCE	TPAMI 2023	ResNet50	84.22	75.60	85.16	81.66
DISC	CVPR 2023	ResNet50	85.27	81.08	88.31	84.89
ANL	NeurIPS 2023	ResNet50	81.78	79.46	86.47	82.57
NPN	AAAI 2024	ResNet50	83.65	79.36	85.46	82.82
ACT	MM 2024	ResNet50	86.56	81.43	88.75	85.58
SED	ECCV 2024	ResNet50	86.62	82.00	88.88	85.83
Ours	-	ResNet50	**87.70**	**82.48**	**89.11**	**86.43**

Table 2. The comparison with SOTA approaches in test accuracy (%) on real-world noisy datasets: Web-Aircraft, Web-Bird, Web-Car.

Methods	Publication	Backbone	Acc (%)
Standard	-	ResNet50	84.50
Decoupling	NeurIPS 2017	ResNet50	85.53
Co-teaching	NeurIPS 2018	ResNet50	61.91
Co-teaching+	ICML 2019	ResNet50	81.61
JoCoR	CVPR 2020	ResNet50	77.94
DivideMix	ICLR 2020	ResNet50	85.88
PLC	ICML 2021	ResNet50	85.28
WarPI	PR 2022	ResNet50	85.91
CoDis	ICCV 2023	ResNet50	86.13
VRI	IJCV 2024	ResNet50	86.24
Ours	-	ResNet50	**86.83**

Table 3. The comparison with SOTA approaches in test accuracy(%) on the large-scale, real-world noisy dataset Food101N.

4.2. Results on Synthetic Noisy Datasets

We compare our proposed CA2C with existing SOTA methods on synthetic noisy datasets under various noise types (*i.e.*, symmetric and asymmetric) and noise rates (*i.e.*, 20%, 40%, and 80%). Table 1 provides the detailed results of comparison, including the average test accuracy achieved by each method across all noise settings. As shown in Table 1, our CA2C achieves state-of-the-art (or at least competitive) performance across various noise settings, particularly in challenging scenarios such as Asym-40%. Specifically, our CA2C achieves an accuracy of 65.59% on CIFAR100N (Asym-40%) and 65.71% on CIFAR80N (Asym-40%), demonstrating its remarkable robustness against label noise. Although CA2C achieves the second-best performance on CIFAR80N (Sym-80%), it surpasses all counterparts in terms of average accuracy. Notably, our CA2C establishes a new benchmark, achieving improvements of 8.72%, 1.94%, and 2.66% over the latest competing methods, namely NPN, ACT, and SED, respectively. These results highlight that our proposed CA2C is remarkably effective in addressing both closed-set and open-set noisy labels.

4.3. Results on Real-World Noisy Datasets

Tables 2 and 3 report the comparison between our CA2C and existing SOTA methods on real-world noisy datasets (*i.e.*, Web-Aircraft, Web-Bird, Web-Car, and Food101N). These four datasets are tailored for fine-grained image classification tasks, posing more challenges when handling noisy labels. As shown in Table 2, while both our CA2C and the competing methods exhibit robustness against real-world label noise, CA2C consistently outperforms its counterparts across datasets, achieving an average accuracy improvement of 0.60% over the second-best performer (*i.e.*, SED). Specifically, our CA2C achieves 87.70%, 82.48%, and 88.93% test accuracy on Web-Aircraft, Web-Bird, and Web-Car, respectively. Moreover, as shown in Table 3, the test accuracy of our CA2C exceeds that of the second-best performance (*i.e.*, VRI) by 0.59% on Food101N. Notably, our method consistently achieves optimal results without relying on any prior knowledge across these real-world benchmarks, underscoring its exceptional practicality and effectiveness in addressing real-world applications.

4.4. Ablation Studies

This section investigates the effectiveness of our key components: Asymmetric Co-learning with Paradigm Deconstruction (ACPD), Asymmetric Co-training with Cross-guidance Label Generation (ACLG), and Confidence-based Re-weighting for label disambiguation (CBRW) The detailed results are presented in Table 4.

ACPD	✗	✓
Sym-20%	35.14	65.22
Sym-80%	4.41	28.53
Asym-40%	27.29	62.02

ACLG	✗	✓
Sym-20%	65.22	67.37
Sym-80%	28.53	35.90
Asym-40%	62.02	65.06

CBRW	✗	✓
Sym-20%	67.37	68.64
Sym-80%	35.90	40.97
Asym-40%	65.06	65.59

ACPD	✗	✓
Sym-20%	29.37	64.81
Sym-80%	4.20	28.01
Asym-40%	22.25	60.39

ACLG	✗	✓
Sym-20%	64.81	68.46
Sym-80%	28.01	36.46
Asym-40%	60.39	64.63

CBRW	✗	✓
Sym-20%	68.46	70.06
Sym-80%	36.46	40.47
Asym-40%	64.63	65.71

Table 4. Effect of key components (*i.e.*, ACPD, ACLG and CBRW) in our CA2C on CIFAR100N (top) and CIFAR80N (bottom). Test accuracy (%) of our CA2C with (✓) and without (✗) the different components is compared under different settings.

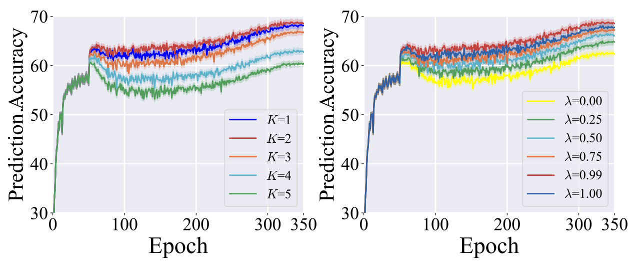

Figure 3. Sensitivity of Hyper-parameters: K (left) and λ (right). Experiments are conducted on CIFAR100N with Sym-20%.

Effects of the Asymmetric Co-learning Strategy with Paradigm Deconstruction: As described in Section 3.2, our CA2C deconstructs the LNL problem into a combined asymmetric co-learning and co-training framework, eliminating the demand for any strong prior knowledge. To achieve this, we propose the asymmetric co-learning strategy with paradigm deconstruction (ACPD), where twin models with identical architectures are trained using different learning paradigms to maximize their distinct learning capacities. The results in Table 4 (the left column) indicate that ACPD remarkably boosts model performance in various noise settings, firmly demonstrating its effectiveness.

Effects of Asymmetric Co-training Strategy with Cross-guidance Label Generation: To mitigate the error accumulation, we propose the asymmetric co-training strategy with cross-guidance label generation (ACLG) in Section 3.3. Specifically, without compromising the distinct learning capabilities, ACLG enables each model to leverage its peer to enhance its learning, facilitating knowledge exchange between models with diverse capabilities and thus preventing error accumulation. The results in Table 4 (the middle column) demonstrate that ACLG brings significant benefits to our model performance.

Effects of the Confidence-based Re-weighting: To enhance the tolerance of the *P-model* to inevitable disambiguation failures, we propose a confidence-based reweighting strategy. By maintaining a memory bank that tracks the frequencies of candidate labels, CBRW generates re-weighting coefficients to enhance the reliability of label disambiguation. As shown in Table 3 (the right column), the incorporation of CBRW enables our CA2C to achieve additional performance improvements.

Sensitivity of Hyper-parameters: We study the sensitivity of two key hyper-parameters: K and λ. K is used to control the number of labels in the candidate label set. As shown in Fig. 3 (left), the highest performance is achieved when $K = 2$, which minimizes the difficulty of label disambiguation in partial label learning while avoiding direct estimation of the ground-truth labels for noisy samples. λ is the hyper-parameter that controls the weights of the hard and soft disambiguation losses in Eq. (11). Fig. 3 (right) shows the results of employing different values of λ. When λ is set to 0.99, our CA2C attains the highest test accuracy throughout the robust training process.

5. Conclusion

In this paper, we proposed a novel noisy label learning (LNL) approach, termed CA2C. The method integrated asymmetric co-learning and co-training strategies to mitigate performance degradation in real-world noisy scenarios. Notably, our CA2C method eliminated requirements for strong prior knowledge, making it more adaptable in various real-world cases. Specifically, we trained two models simultaneously but employed distinct learning paradigms, aiming to maximize the twin models' disparities in learning capabilities and leverage their complementary strengths to improve robustness against noisy labels. Next, we designed an asymmetric co-training strategy with cross-guidance label generation to facilitate knowledge exchange between these twin models with distinct capabilities, effectively mitigating error accumulations. Furthermore, we introduced a confidence-based re-weighting strategy to improve the robustness of the *P-model* against disambiguation failures. Comprehensive experimental results on synthetic and real-world noisy datasets demonstrated the effectiveness and robustness of our CA2C in handling noisy labels.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 62202227, 62472222) and the Natural Science Foundation of Jiangsu Province (No. BK20220938, BK20240080).

References

[1] Eric Arazo, Diego Ortego, Paul Albert, Noel E. O'Connor, and Kevin McGuinness. Unsupervised label noise modeling and loss correction. In *Int. Conf. Mach. Learn.*, pages 312–321, 2019. 3

[2] Devansh Arpit, Stanislaw Jastrzebski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S. Kanwal, Tegan Maharaj, Asja Fischer, Aaron C. Courville, Yoshua Bengio, and Simon Lacoste-Julien. A closer look at memorization in deep networks. In *Int. Conf. Mach. Learn.*, pages 233–242, 2017. 2

[3] Yingbin Bai, Erkun Yang, Bo Han, Yanhua Yang, Jiatong Li, Yinian Mao, Gang Niu, and Tongliang Liu. Understanding and improving early stopping for learning with noisy labels. In *Adv. Neural Inform. Process. Syst.*, pages 24392–24403, 2021. 2

[4] Xinhao Cai, Qiuxia Lai, Yuwei Wang, Wenguan Wang, Zeren Sun, and Yazhou Yao. Poly kernel inception network for remote sensing detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 1

[5] Tao Chen, Yazhou Yao, and Jinhui Tang. Multi-granularity denoising and bidirectional alignment for weakly supervised semantic segmentation. *IEEE Trans. Multimedia*, 32:2960–2971, 2023. 1

[6] Tao Chen, Yazhou Yao, Lei Zhang, Qiong Wang, Guo-Sen Xie, and Fumin Shen. Saliency guided inter- and intra-class relation constraints for weakly supervised semantic segmentation. *IEEE Trans. Multimedia*, 25:1727–1737, 2023. 1

[7] De Cheng, Tongliang Liu, Yixiong Ning, Nannan Wang, Bo Han, Gang Niu, Xinbo Gao, and Masashi Sugiyama. Instance-dependent label-noise learning with manifold-regularized transition matrix estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 16609–16618, 2022. 2

[8] Hao Cheng, Zhaowei Zhu, Xing Sun, and Yang Liu. Mitigating memorization of noisy labels via regularization between representations. In *Int. Conf. Learn. Represent.*, 2023. 2

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 248–255, 2009. 1

[10] Tongtong Fang, Nan Lu, Gang Niu, and Masashi Sugiyama. Rethinking importance weighting for deep learning under distribution shift. In *Adv. Neural Inform. Process. Syst.*, 2020. 3

[11] Aritra Ghosh, Himanshu Kumar, and P. S. Sastry. Robust loss functions under label noise for deep neural networks. In *AAAI*, pages 1919–1925, 2017. 2

[12] Jacob Goldberger and Ehud Ben-Reuven. Training deep neural-networks using a noise adaptation layer. In *Int. Conf. Learn. Represent.*, 2017. 2, 3

[13] Shilin Gu, Chao Xu, Dewen Hu, and Chenping Hou. Adaptive learning for dynamic features and noisy labels. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(2):1219–1237, 2025. 2

[14] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor W. Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *Adv. Neural Inform. Process. Syst.*, pages 8536–8546, 2018. 2, 3, 6

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 770–778, 2016. 6

[16] Shenwang Jiang, Jianan Li, Ying Wang, Bo Huang, Zhang Zhang, and Tingfa Xu. Delving into sample loss curve to embrace noisy and imbalanced data. In *AAAI*, pages 7024–7032, 2022. 2

[17] Xiruo Jiang, Sheng Liu, Xili Dai, Guosheng Hu, Xingguo Huang, Yazhou Yao, Guo-Sen Xie, and Ling Shao. Deep metric learning based on meta-mining strategy with semiglobal information. *IEEE Trans. Neural Networks Learn. Syst.*, 35(4):5103–5116, 2024. 1

[18] Nazmul Karim, Mamshad Nayeem Rizve, Nazanin Rahnavard, Ajmal Mian, and Mubarak Shah. UNICON: combating label noise through uniform selection and contrastive learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 9666–9676, 2022. 6

[19] Alex Krizhevsky. Learning multiple layers of features from tiny images, 2009. 6

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Adv. Neural Inform. Process. Syst.*, pages 1106–1114, 2012. 1

[21] Kuang-Huei Lee, Xiaodong He, Lei Zhang, and Linjun Yang. Cleannet: Transfer learning for scalable image classifier training with label noise. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5447–5456, 2018. 6

[22] Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. In *Int. Conf. Learn. Represent.*, 2020. 2, 3, 6

[23] Jichang Li, Guanbin Li, Feng Liu, and Yizhou Yu. Neighborhood collective estimation for noisy label identification and correction. In *Eur. Conf. Comput. Vis.*, pages 128–145, 2022. 2

[24] Jichang Li, Guanbin Li, Hui Cheng, Zicheng Liao, and Yizhou Yu. Feddiv: Collaborative noise filtering for federated learning with noisy labels. In *AAAI*, pages 3118–3126, 2024.

[25] Yuncheng Li, Jianchao Yang, Yale Song, Liangliang Cao, Jiebo Luo, and Li-Jia Li. Learning from noisy labels with distillation. In *Int. Conf. Comput. Vis.*, pages 1928–1936, 2017. 2

[26] Yifan Li, Hu Han, Shiguang Shan, and Xilin Chen. DISC: learning from noisy labels via dynamic instance-specific selection and correction. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 24070–24079, 2023. 6

[27] Huafeng Liu, Mengmeng Sheng, Zeren Sun, Yazhou Yao, Xian-Sheng Hua, and Heng-Tao Shen. Learning with im-

balanced noisy data by preventing bias in sample selection. *IEEE Trans. Multimedia*, pages 1–12, 2024. 2

[28] Qipeng Liu, Luojun Lin, Zhifeng Shen, and Zhifeng Yang. Periodically exchange teacher-student for source-free object detection. In *Int. Conf. Comput. Vis.*, pages 6391–6401, 2023. 1

[29] Sheng Liu, Jonathan Niles-Weed, Narges Razavian, and Carlos Fernandez-Granda. Early-learning regularization prevents memorization of noisy labels. In *Adv. Neural Inform. Process. Syst.*, 2020. 2

[30] Sheng Liu, Zhihui Zhu, Qing Qu, and Chong You. Robust training under label noise by over-parameterization. In *Int. Conf. Mach. Learn.*, pages 14153–14172, 2022. 6

[31] Eran Malach and Shai Shalev-Shwartz. Decoupling "when to update" from "how to update". In *Adv. Neural Inform. Process. Syst.*, pages 960–970, 2017. 2, 3, 6

[32] Devraj Mandal, Shrisha Bharadwaj, and Soma Biswas. A novel self-supervised re-labeling approach for training with noisy labels. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1370–1379, 2020. 2

[33] Junzhu Mao, Yazhou Yao, Zeren Sun, Xingguo Huang, Fumin Shen, and Heng Tao Shen. Attention map guided transformer pruning for occluded person re-identification on edge device. *IEEE Trans. Multimedia*, 25:1592–1599, 2023. 1

[34] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1944–1952, 2017. 3

[35] Gensheng Pei, Tao Chen, Yujia Wang, Xinhao Cai, Xiangbo Shu, Tianfei Zhou, and Yazhou Yao. Seeing what matters: Empowering CLIP with patch generation-to-selection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 24862–24872, 2025. 1

[36] Xiaojiang Peng, Kai Wang, Zhaoyang Zeng, Qing Li, Jianfei Yang, and Yu Qiao. Suppressing mislabeled data via grouping and self-attention. In *Eur. Conf. Comput. Vis.*, pages 786–802, 2020. 6

[37] Joseph Redmon and Ali Farhadi. YOLO9000: better, faster, stronger. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6517–6525, 2017. 1

[38] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *Int. Conf. Mach. Learn.*, pages 4331–4340, 2018. 2, 3

[39] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. In *Adv. Neural Inform. Process. Syst.*, pages 91–99, 2017. 1

[40] Mengmeng Sheng, Zeren Sun, Zhenhuang Cai, Tao Chen, Yichao Zhou, and Yazhou Yao. Adaptive integration of partial label learning and negative learning for enhanced noisy label learning. In *AAAI*, pages 4820–4828, 2024. 2, 4, 6

[41] Mengmeng Sheng, Zeren Sun, Tao Chen, Shuchao Pang, Yucheng Wang, and Yazhou Yao. Foster adaptivity and balance in learning with noisy labels. In *Eur. Conf. Comput. Vis.*, pages 217–235, 2024. 6

[42] Mengmeng Sheng, Zeren Sun, Gensheng Pei, Tao Chen, Haonan Luo, and Yazhou Yao. Enhancing robustness in learning with noisy labels: An asymmetric co-training approach. In *ACM Int. Conf. Multimedia*, pages 4406–4415, 2024. 2, 3, 4, 5, 6

[43] Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. Meta-weight-net: Learning an explicit mapping for sample weighting. In *Adv. Neural Inform. Process. Syst.*, pages 1917–1928, 2019. 3

[44] Haoliang Sun, Chenhui Guo, Qi Wei, Zhongyi Han, and Yilong Yin. Learning to rectify for robust learning with noisy labels. *Pattern Recognition*, 124:108467, 2022. 6

[45] Haoliang Sun, Qi Wei, Lei Feng, Yupeng Hu, Fan Liu, Hehe Fan, and Yilong Yin. Variational rectification inference for learning with noisy labels. *Int. J. Comput. Vis.*, 133(2):652–671, 2025. 6

[46] Zeren Sun, Xian-Sheng Hua, Yazhou Yao, Xiu-Shen Wei, Guosheng Hu, and Jian Zhang. CRSSC: salvage reusable samples from noisy data for robust learning. In *ACM Int. Conf. Multimedia*, pages 92–101, 2020. 6

[47] Zeren Sun, Huafeng Liu, Qiong Wang, Tianfei Zhou, Qi Wu, and Zhenmin Tang. Co-ldl: A co-training-based label distribution learning method for tackling label noise. *IEEE Trans. Multimedia*, pages 1093–1104, 2022. 6

[48] Zeren Sun, Fumin Shen, Dan Huang, Qiong Wang, Xiangbo Shu, Yazhou Yao, and Jinhui Tang. Pnp: Robust learning from noisy labels by probabilistic noise prediction. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5311–5320, 2022. 2, 3, 6

[49] Cheng Tan, Jun Xia, Lirong Wu, and Stan Z. Li. Co-learning: Learning from noisy labels with self-supervision. In *ACM Int. Conf. Multimedia*, pages 1405–1413, 2021. 2

[50] Daiki Tanaka, Daiki Ikami, Toshihiko Yamasaki, and Kiyoharu Aizawa. Joint optimization framework for learning with noisy labels. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5552–5560, 2018. 3

[51] Quan Tang, Bowen Zhang, Jiajun Liu, Fagui Liu, and Yifan Liu. Dynamic token pruning in plain vision transformers for semantic segmentation. In *Int. Conf. Comput. Vis.*, pages 777–786, 2023. 1

[52] Yuanpeng Tu, Boshen Zhang, Yuxi Li, Liang Liu, Jian Li, Yabiao Wang, Chengjie Wang, and Cairong Zhao. Learning from noisy labels with decoupled meta label purifier. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 19934–19943, 2023. 3

[53] Yuanpeng Tu, Boshen Zhang, Yuxi Li, Liang Liu, Jian Li, Jiangning Zhang, Yabiao Wang, Chengjie Wang, and Cairong Zhao. Learning with noisy labels via self-supervised adversarial noisy masking. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 16186–16195, 2023. 2

[54] Yixin Wang, Alp Kucukelbir, and David M. Blei. Robust probabilistic modeling with bayesian data reweighting. In *Int. Conf. Mach. Learn.*, pages 3646–3655, 2020. 3

[55] Hongxin Wei, Lei Feng, Xiangyu Chen, and Bo An. Combating noisy labels by agreement: A joint training method with co-regularization. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 13723–13732, 2020. 2, 3, 6

[56] Jiaheng Wei, Hangyu Liu, Tongliang Liu, Gang Niu, Masashi Sugiyama, and Yang Liu. To smooth or not? when label smoothing meets noisy labels. In *Int. Conf. Mach. Learn.*, pages 23589–23614, 2022. 2

[57] Peter Welinder, Steve Branson, Serge J. Belongie, and Pietro Perona. The multidimensional wisdom of crowds. In *Adv. Neural Inform. Process. Syst.*, pages 2424–2432, 2010. 1

[58] Songhua Wu, Xiaobo Xia, Tongliang Liu, Bo Han, Mingming Gong, Nannan Wang, Haifeng Liu, and Gang Niu. Class2simi: A noise reduction perspective on learning with noisy labels. In *Int. Conf. Mach. Learn.*, pages 11285–11295, 2021. 1

[59] Xiaobo Xia, Tongliang Liu, Nannan Wang, Bo Han, Chen Gong, Gang Niu, and Masashi Sugiyama. Are anchor points really indispensable in label-noise learning? In *Adv. Neural Inform. Process. Syst.*, pages 6835–6846, 2019. 1

[60] Xiaobo Xia, Bo Han, Nannan Wang, Jiankang Deng, Jiatong Li, Yinian Mao, and Tongliang Liu. Extended \$t\$t: Learning with mixed closed-set and open-set noisy labels. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 3047–3058, 2023. 2

[61] Xiaobo Xia, Bo Han, Yibing Zhan, Jun Yu, Mingming Gong, Chen Gong, and Tongliang Liu. Combating noisy labels with sample selection by mining high-discrepancy examples. In *Int. Conf. Comput. Vis.*, pages 1833–1843, 2023. 2, 6

[62] Wenjie Xuan, Shanshan Zhao, Yu Yao, Juhua Liu, Tongliang Liu, Yixin Chen, Bo Du, and Dacheng Tao. Pnt-edge: Towards robust edge detection with noisy labels by learning pixel-level noise transitions. In *ACM Int. Conf. Multimedia*, pages 1924–1932, 2023. 2

[63] Yazhou Yao, Zeren Sun, Chuanyi Zhang, Fumin Shen, Qi Wu, Jian Zhang, and Zhenmin Tang. Jo-src: A contrastive approach for combating noisy labels. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5192–5201, 2021. 2, 3, 6

[64] Xichen Ye, Xiaoqiang Li, Songmin Dai, Tong Liu, Yan Sun, and Weiqin Tong. Active negative loss functions for learning with noisy labels. In *Adv. Neural Inform. Process. Syst.*, 2023. 6

[65] Kun Yi and Jianxin Wu. Probabilistic end-to-end noise correction for learning with noisy labels. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7017–7025, 2019. 3

[66] Kun Yi and Jianxin Wu. Probabilistic end-to-end noise correction for learning with noisy labels. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7017–7025, 2019. 2, 3, 6

[67] Hui Ying, Zhaojin Huang, Shu Liu, Tianjia Shao, and Kun Zhou. Embedmask: Embedding coupling for instance segmentation. In *IJCAI*, pages 1266–1273, 2021. 1

[68] Masakazu Yoshimura, Junji Otsuka, Atsushi Irie, and Takeshi Ohashi. Dynamicisp: Dynamically controlled image signal processor for image recognition. In *Int. Conf. Comput. Vis.*, pages 12820–12830, 2023. 1

[69] Xingrui Yu, Bo Han, Jiangchao Yao, Gang Niu, Ivor W. Tsang, and Masashi Sugiyama. How does disagreement help generalization against label corruption? In *Int. Conf. Mach. Learn.*, pages 7164–7173, 2019. 2, 3, 6

[70] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. In *Int. Conf. Learn. Represent.*, 2017. 1

[71] Shuo Zhang, Yuwen Li, Zhongyu Wang, Jianqing Li, and Chengyu Liu. Learning with noisy labels using hyperspherical margin weighting. In *AAAI*, pages 16848–16856, 2024. 2

[72] Yikai Zhang, Songzhu Zheng, Pengxiang Wu, Mayank Goswami, and Chao Chen. Learning with feature-dependent label noise: A progressive approach. In *Int. Conf. Learn. Represent.*, 2021. 6

[73] Zhilu Zhang and Mert R. Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In *Adv. Neural Inform. Process. Syst.*, pages 8792–8802, 2018. 2

[74] Xiong Zhou, Xianming Liu, Chenyang Wang, Deming Zhai, Junjun Jiang, and Xiangyang Ji. Learning with noisy labels via sparse regularization. In *Int. Conf. Comput. Vis.*, pages 72–81, 2021.

[75] Xiong Zhou, Xianming Liu, Deming Zhai, Junjun Jiang, and Xiangyang Ji. Asymmetric loss functions for noise-tolerant learning: Theory and applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(7):8094–8109, 2023. 2, 6

[76] Yuyin Zhou, Xianhang Li, Fengze Liu, Qingyue Wei, Xuxi Chen, Lequan Yu, Cihang Xie, Matthew P. Lungren, and Lei Xing. L2b: Learning to bootstrap robust models for combating label noise. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 2, 3