

Adaptive Integration of Partial Label Learning and Negative Learning for Enhanced Noisy Label Learning

Mengmeng Sheng¹, Zeren Sun^{1*}, Zhenhuang Cai¹, Tao Chen¹, Yichao Zhou¹, Yazhou Yao^{1*}

¹Nanjing University of Science and Technology, Nanjing, China

{shengmengmeng, zereens, czh, taochen, yczhou}@njust.edu.cn, yazhou.yao@outlook.com

Abstract

There has been significant attention devoted to the effectiveness of various domains, such as semi-supervised learning, contrastive learning, and meta-learning, in enhancing the performance of methods for noisy label learning (NLL) tasks. However, most existing methods still depend on prior assumptions regarding clean samples amidst different sources of noise (e.g., a pre-defined drop rate or a small subset of clean samples). In this paper, we propose a simple yet powerful idea called **NPN**, which revolutionizes Noisy label learning by integrating Partial label learning (PLL) and Negative learning (NL). Toward this goal, we initially decompose the given label space adaptively into the candidate and complementary labels, thereby establishing the conditions for PLL and NL. We propose two adaptive data-driven paradigms of label disambiguation for PLL: hard disambiguation and soft disambiguation. Furthermore, we generate reliable complementary labels using all non-candidate labels for NL to enhance model robustness through indirect supervision. To maintain label reliability during the later stage of model training, we introduce a consistency regularization term that encourages agreement between the outputs of multiple augmentations. Experiments conducted on both synthetically corrupted and real-world noisy datasets demonstrate the superiority of NPN compared to other state-of-the-art (SOTA) methods. The source code has been made available at <https://github.com/NUST-Machine-Intelligence-Laboratory/NPN>.

Introduction

Over the past decades, the revolution in supervised learning has been driven by large-scale, well-annotated datasets such as ImageNet (Deng et al. 2009). However, collecting these datasets has become a bottleneck due to its expense and time-consuming nature, hindering the scalability of models. The acquisition of large-scale datasets furnished with annotations of high quality for supervised learning algorithms is a formidable challenge. To address this issue, weakly supervised learning (Yan and Guo 2023; Chen, Yao, and Tang 2023; Chen et al. 2023b) has gained considerable attention, which includes, but is not limited to, noisy label learning (Li, Socher, and Hoi 2020; Han et al. 2018; Sun et al. 2022b), multi-label learning (Gong, Yuan, and Bao

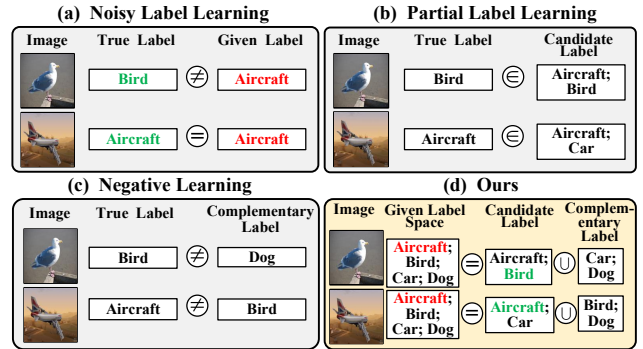


Figure 1: (a-c) The difference between Noisy Label Learning, Partial Label Learning, and Negative Learning. (d) Our decomposition of the given label space to candidate label and complementary label to establish the necessary conditions for Partial Label Learning and Negative Learning.

2022; Yan and Guo 2021; Xu, Liu, and Geng 2020), partial label learning (Wang et al. 2022; Xu et al. 2021), and semi-supervised learning (Chen et al. 2023a; Wang et al. 2023). In this paper, our primary focus is on one specific weakly supervised learning problem known as noisy label learning. NLL involves handling low-quality samples with noisy labels (as shown in Fig. 1 (a)), which mainly stem from crowd-sourcing platforms (Welinder et al. 2010) or web image search engines (Fergus et al. 2010) employed in dataset construction. The powerful learning capability of deep neural networks allows them to fit any noisy labels, leading to inferior performance in image classification tasks. Consequently, developing robust methods to mitigate the impact of noisy labels assumes paramount importance.

The existing NLL methods primarily focus on combating label noise by incorporating strategies such as semi-supervised learning, contrastive learning, and meta-learning (Liu et al. 2022a,b). Semi-supervised learning (SSL) often constitutes an appropriate choice to leverage the selected noisy samples through sample selection rather than abandoning them. For example, DivideMix (Li, Socher, and Hoi 2020) employs a mixture model to characterize the per-sample loss distribution, dynamically partitioning the training data into a labeled subset comprising clean sam-

*Corresponding Author

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

ples and an unlabeled subset comprising noisy samples. The model is subsequently trained on both labeled and unlabeled data, adopting a semi-supervised approach. UNICON (Karim et al. 2022) and NCE (Li et al. 2022a) propose refined sample selection strategies while incorporating semi-supervised learning. However, these methods tend to neglect the reliability of the pseudo-labels generated by the semi-supervised approach, which may lead to correction bias.

Another line of NLL methods boosts performance by integrating contrastive learning. This approach facilitates feature extraction that remains uninfluenced by label noise, thereby demonstrating promising results. Jo-SRC (Yao et al. 2021) takes predictions from two distinct views of each sample to estimate its “likelihood” of being clean or out-of-distribution in a contrastive learning manner. Sel-CL (Li et al. 2022b) proposes selective-supervised contrastive learning combined with a joint loss to enhance model generalization performance by introducing consistency regularization. Nevertheless, these methods are sensitive to hyperparameters. Recently, Meta-learning has also been employed to mitigate label noise. L2RW (Ren et al. 2018) proposes a meta-learning instance re-weighting approach to assign weights to noisy instances. MLC (Zheng, Awadallah, and Dumais 2021) proposes to adopt a label correction network as a meta-model to produce corrected labels for noisy samples. However, these meta-learning methods tend to require prior knowledge (*e.g.*, a small subset of clean samples).

To tackle the above challenges, we propose an innovative approach named **NPN** for noisy label learning. To be specific, NPN synergistically incorporates two distinct learning paradigms: partial label learning and negative learning, to combat label noise effectively. PLL enables each training example to be labeled with a coarse candidate set, which is well-suited for real-world data annotation scenarios with label ambiguity (as shown in Fig. 1 (b)). NL provides indirect supervision information using a complementary label that teaches the network “the input image does not belong to this complementary label” (as shown in Fig. 1 (c)). Thus, as a preliminary step, we decompose the given label space to the candidate and complementary labels to establish the conditions for PLL and NL (as shown in Fig. 1 (d)). For PLL, we propose two paradigms of label disambiguation in an adaptive data-driver manner: hard disambiguation and soft disambiguation. Additionally, we suggest generating reliable complementary labels using all non-candidate labels for NL to enhance model robustness through indirect supervision. Finally, a consistency regularization term is applied to improve both feature extraction and model prediction. We provide comprehensive experimental results and extensive ablation studies to validate the effectiveness and superiority of NPN on synthetically corrupted and real-world datasets. Our contributions can be summarized as follows:

(1) We present a simple yet powerful method, named NPN, to alleviate negative impacts induced by noisy labels. NPN introduces a paradigm shift in conventional noisy label learning by amalgamating partial label learning and negative learning. This involves partitioning the given label space into candidate labels for PLL and complementary labels for NL.

(2) We introduce two paradigms of label disambiguation

for PLL: hard disambiguation (NPN-hard) and soft disambiguation (NPN-soft). To enhance the effectiveness of NL, we advocate using all reliable non-candidate labels as the complementary label, as opposed to randomly selecting a single unreliable non-given label.

(3) Comprehensive evaluations on one synthetically corrupted and three real-world noisy datasets show that NPN outperforms SOTA methods in addressing noisy labels. Extensive ablation studies are conducted to further verify the effectiveness of the proposed method.

Related Work

Noisy Label Learning. Recently, numerous methods have been proposed for learning with noisy labels (Yao et al. 2021; Xia et al. 2023; Gong et al. 2023). They focus on combating noisy labels through sample selection, label correction, or noise regularization. Previous literature about sample selection attempts to detect noisy labels by exacerbating the natural resistance of neural networks to noise. BARE (Patel and Sastry 2023) proposes an adaptive sample selection strategy that relies only on batch statistics to provide robustness against label noise. Another line of research focuses on label correction, which typically attempts to rectify sample labels using the model predictions. Jo-SRC (Yao et al. 2021) uses the temporally averaged model (*i.e.*, mean-teacher model) to generate reliable pseudo-label distributions for training. Besides, PLS (Albert et al. 2023) proposes considering the confidence of corrected labels. Certain studies in the literature emphasize focus on noise regularization, such as mixup (Zhang et al. 2018), a dedicated loss term (Liu et al. 2020), or contrastive learning. Unsupervised regularization (Sun et al. 2022a; Li, Xiong, and Hoi 2021) has also been shown effective in improving the classification accuracy of neural networks trained on noisy datasets.

Partial Label Learning. Partial Label Learning allows each training example to be annotated with a candidate label set, in which the ground-truth label is guaranteed to be included. Two typical strategies have been proposed for label disambiguation in PLL: the average-based and identification-based disambiguation strategies. The average-based disambiguation strategy treats each candidate label equally during model learning (Zhang and Yu 2015). The identification-based disambiguation strategy considers the ground-truth label as a latent variable. It identifies the ground-truth label by deriving confidence scores for all candidate labels (Wang, Zhang, and Li 2022; Feng and An 2019). These label disambiguation strategies can serve as inspiration for noisy label learning, which can be regarded as a variation of PLL where candidate labels encompass all categories in the dataset.

Negative Learning. Negative Learning is an indirect learning method where CNNs are trained by utilizing a complementary label that indicates “input image does not belong to this complementary label.” NLNL (Kim et al. 2019) introduces a noise-robust learning approach that utilizes negative learning to mitigate the risk of learning incorrect information. JNPL (Kim et al. 2021) improves (Kim et al. 2019) and proposes a novel approach combining negative and positive learning. UPS (Rizve et al. 2021) employs negative learning on noisy and low-confidence labels to generalize negative

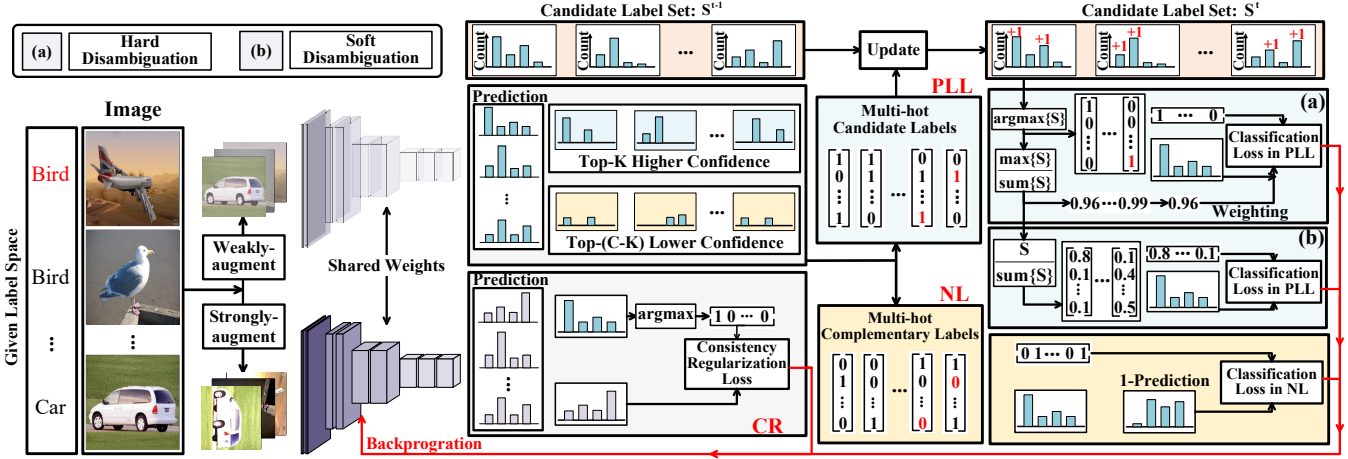


Figure 2: The overall framework of NPN. We first decompose the given label space into the candidate and complementary labels based on model predictions. For PLL, we introduce two adaptive, data-driven paradigms of label disambiguation: hard disambiguation and soft disambiguation. To enable adaptive, data-driven disambiguation, we maintain a candidate label set \mathcal{S} that stores the candidate labels for each sample during the training process. For NL, we incorporate reliable complementary labels, which encompass all non-candidate labels, to enhance the model’s robustness through indirect supervision. Finally, NPN introduces a consistency regularization term that promotes prediction agreement between sample augmentations.

pseudo-labels, which can be utilized for improving single-label classification. In this work, we propose to exploit the indirect supervision information by negative learning.

Methods

Problem Statement

Throughout this paper, we mainly focus on noisy label learning for image classification. Considering the problem of a C -class classification task, we denote the noisy training set $\mathcal{D}_{train} = \{(x_n, y_n) | n = 1, \dots, N\}$ as the input with N training samples. x_n denotes the n -th image, and $y_n \in \{0, 1\}^C$ denotes the corresponding one-hot label of C classes. The label y_n may not be equivalent to the ground-truth label denoted by y_n^* , which is not observable during training. Suppose a classification neural network $\mathcal{F}(\cdot, \theta)$ maps the input space to the C -dimensional score space $\mathcal{F} : \mathcal{X} \rightarrow \mathbb{R}^C$, where θ denotes network parameters. We further denote $p^c(x_n, \theta)$ as the predicted softmax probability of training sample x_n over its c -th class. The goal is to train $\mathcal{F}(\cdot, \theta)$ on the noisy training set \mathcal{D}_{train} to perform accurate prediction on \mathcal{D}_{test} , where $\mathcal{D}_{test} = \{(x_m, y_m^*) | m = 1, \dots, M\}$ is the test set with accurate labels. The loss function in the conventional training scheme is the cross-entropy loss as follows:

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

Existing literature (Wei et al. 2022; Xia et al. 2022; Yang et al. 2022) highlights that optimizing networks using the cross-entropy loss on noisy datasets often results in inferior performance. In this work, we aim to enhance noisy label learning by integrating partial label learning and negative learning. PLL allows each training example x_n to be labeled

with a coarse candidate label set \mathcal{Y}_n . NL is an indirect training method for CNNs, where the complementary label $\tilde{\mathcal{Y}}_n$ is employed. It is worth noting that y_n is encoded as a single label using one-hot encoding, whereas \mathcal{Y}_n and $\tilde{\mathcal{Y}}_n$ are encoded using multi-hot encoding in our NPN.

Label Space Decomposition

In the literature of NLL, more efforts have been devoted to semi-supervised learning, contrastive learning, and meta-learning. However, semi-supervised learning methods tend to overlook the reliability of the generated pseudo-labels, which may lead to correction bias. Methods employing contrastive learning methods are considerably sensitive to hyper-parameters. Meta-learning methods struggle in requiring prior knowledge (*i.e.*, a small subset of clean data).

To this end, we propose to revolutionize conventional noisy label learning by integrating partial label learning and negative learning. As an initial step, we decompose the given label space into a combination of the candidate label set in PLL and the complementary label in NL. This can prevent the model from overfitting noisy labels through the direct and indirect supervision information learned from PLL and NL, respectively. We construct the candidate label set in PLL by selecting the given label y_n and the category with the highest prediction confidence $\hat{y}_n \in \{0, 1\}^C$ as follows:

$$\mathcal{Y}_n = y_n + \hat{y}_n, \quad \hat{y}_n^k = \mathbb{1}_{k = \arg \max_{c \in \{1, \dots, C\}} p^c(x_n, \theta)}. \quad (2)$$

$\mathbb{1}_A$ is an indicator function. We generate complementary labels for NL using the remaining non-candidate labels in the given label space. Unlike NLNL (Kim et al. 2019) and JNPL (Kim et al. 2021), which randomly select a single label from non-given labels as the complementary label, we use all non-candidate labels as the complementary label set,

$$\tilde{\mathcal{Y}}_n = \neg \mathcal{Y}_n = \mathcal{I} - \mathcal{Y}_n, \quad (3)$$

where \mathcal{I} denotes the entire label space.

Discussion. Notably, the candidate labels within PLL should encompass the ground-truth label, while the complementary label within NL is the exact opposite. Consequently, upholding the reliability of both candidate and supplementary labels becomes paramount. We compare the probability of true labels among the top-k predicted categories on CI-FAR100N, as shown in Fig.3. Fig.3 (a) shows that the true label frequently falls between the given label and the label obtained the highest prediction confidence. In this case, our NPN achieves the highest test accuracy, as shown in Fig.3 (b). The candidate labels with higher confidence further reinforce the reliability of complementary labels used for NL, owing to the relationship $\tilde{\mathcal{Y}}_n + \mathcal{Y}_n = \mathcal{I}$.

Label Disambiguation in Partial Label Learning

The main challenge of PLL arises from the fact that the ground-truth label is concealed within the candidate label set and is not directly accessible to the learning algorithms. Thus, the primary focus of PLL is to tackle the task of candidate label disambiguation. Existing methods predominantly concentrate on two strategies: average-based disambiguation and identification-based disambiguation. The average-based disambiguation strategy treats each candidate label equally during the training phase. It computes the average of all the modeling outputs associated with each candidate label during the testing phase (Zhang and Yu 2015). On the other hand, the identification-based disambiguation strategy treats the ground-truth label as a latent variable (Wang, Zhang, and Li 2022). It aims to identify the true label by assigning different confidence scores to the candidate labels.

We propose two adaptive, data-driven paradigms of label disambiguation for PLL: hard disambiguation (*i.e.*, identification-based disambiguation) and soft disambiguation (*i.e.*, average-based disambiguation). Specifically, we first design a candidate label set \mathcal{S} to store the candidate labels for each sample during the training process as:

$$\mathcal{S}_n^t = \begin{cases} y_n, & t = 0 \\ \mathcal{S}_n^{t-1} + \mathcal{Y}_n^t, & t > 0 \end{cases} \quad (4)$$

We denote \mathcal{S}_n^t as the frequency distribution of the candidate labels of the n -th sample among C classes, which has been observed in the previous t epochs. \mathcal{Y}_n^t represents the aforementioned multi-label candidate label of the n -th sample in the t -th epoch. The sample with the highest frequency of occurrence in the candidate label set \mathcal{S}_n^t is more likely to be the ground-truth, allowing for the disambiguation of candidate labels. Following hard disambiguation, we construct \tilde{y} as the disambiguation labels and compute \mathcal{L}_{PLL} as follows:

$$\tilde{y}_n = \arg \max_{j=1, \dots, C} \mathcal{S}_n^t, \quad (5)$$

$$\mathcal{L}_{PLL}^{hard} = -\frac{1}{N} \sum_{n=1}^N \frac{\max\{\mathcal{S}_n^t\}}{\sum\{\mathcal{S}_n^t\}} \tilde{y}_n \log(p(x_n, \theta)). \quad (6)$$

We utilize $\frac{\max\{\mathcal{S}_n^t\}}{\sum\{\mathcal{S}_n^t\}}$ to measure the reliability of the disambiguated labels, aiming to mitigate the impact of inevitable errors from partially disambiguated samples.

Furthermore, we propose a soft disambiguation strategy:

$$\tilde{y}_n = \frac{\mathcal{S}_n^t}{\sum\{\mathcal{S}_n^t\}}, \quad (7)$$

$$\mathcal{L}_{PLL}^{soft} = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C \tilde{y}_n^c \log(p^c(x_n, \theta)). \quad (8)$$

The soft label disambiguation revolves around individually considering each label within the candidate label set and constructing soft labels based on their respective frequencies, thereby accomplishing effective label disambiguation.

Discussion. The effectiveness of PLL is directly influenced by the purity of the labels achieved through disambiguation. To further validate the impact of our label disambiguation strategies, we present the precision of label disambiguation *vs.* epochs in Fig.3 (c). The results demonstrate that selecting more labels into candidate labels increases the difficulty of label disambiguation, resulting in a decline in test accuracy. This rationale elucidates our decision to choose the given label and category with the highest prediction confidence as candidate labels \mathcal{Y} during each epoch. Thus, we ensure that the candidate labels encompass the ground-truth labels to the fullest extent possible while concurrently alleviating the complexities associated with label disambiguation.

Indirect Supervision in Negative Learning

Negative learning is generally carried out to assist the model in acquiring indirect information by randomly selecting a non-given label as the complementary label, as stated in NLNL and JNPL. Moreover, the complementary label should differ from the ground-truth label. However, the ground-truth label is present among the non-given labels when the given label is noisy. They encounter the challenge of potentially mis-selecting the ground-truth label as the complementary label. This becomes particularly pronounced in high-noise scenarios.

As mentioned above, the labels more likely to be ground-truth labels are often assigned to the candidate labels. The remaining labels, less likely to be the ground-truth labels, are thus assigned to the complementary labels. In addition, the complementary labels $\tilde{\mathcal{Y}}$ that we propose encompass a set of labels, as opposed to a randomly selected non-given label. It offers the model a richer source of supervision information, concurrently mitigating the impact of occasional mis-selection of true labels. Finally, we calculate the negative learning loss using the constructed complementary labels as follows:

$$\mathcal{L}_{NL} = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C \tilde{\mathcal{Y}}_n^c \log(1 - p^c(x_n, \theta)). \quad (9)$$

Eq.(9) facilitates the optimization of the probability value associated with the complementary label to approach zero. This adjustment leads to an increase in the probability values of other classes, aligning with the objectives of NL.

Discussion. To address the two drawbacks of NLNL and JNPL, we primarily focus on two aspects: (1) Our NL framework aims to discover more reliable (*i.e.*, lower prediction

Algorithm 1: Our proposed NPN algorithm

Input: The training set D_{train} , network θ , warm-up epochs E_w , total epochs E_{total} , batch size bs .

```

1: for  $epoch = 1, 2, \dots, E_{total}$  do
2:   if  $epoch \leq E_w$  then
3:     for  $iteration = 1, 2, \dots$  do
4:       Fetch  $B = \{(x_i, y_i)\}^{bs}$  from  $D_{train}$ 
5:       Calculate  $\mathcal{L}_{ce} = -\sum_{i=1}^{bs} y_i \log p(x_i, \theta)$ 
6:       Update  $\theta$  by optimizing  $\mathcal{L}_{ce}$ 
7:       Obtain candidate label  $\mathcal{Y}$  using Eq. (2)
8:       Update candidate label set  $\mathcal{S}$  using Eq. (4)
9:     end for
10:   end if
11:   if  $E_w < epoch \leq E_{total}$  then
12:     for  $iteration = 1, 2, \dots$  do
13:       Obtain candidate label  $\mathcal{Y}$  using Eq. (2)
14:       Obtain complementary label  $\tilde{\mathcal{Y}}$  using Eq. (3)
15:       Update candidate label set  $\mathcal{S}$  using Eq. (4)
16:       Calculate  $\mathcal{L}_{PLL}$  using Eq. (6) or Eq. (7)
17:       Calculate  $\mathcal{L}_{NL}$  using Eq. (9)
18:       Calculate  $\mathcal{L}_{REG}$  using Eq. (11)
19:       Calculate  $\mathcal{L} = \mathcal{L}_{PLL} + \alpha\mathcal{L}_{NL} + \beta\mathcal{L}_{REG}$ 
20:       Update  $\theta$  by optimizing  $\mathcal{L}$ 
21:     end for
22:   end if
23: end for

```

Output: Updated network θ .

probability) complementary labels. (2) By utilizing all non-candidate labels as the complementary label set rather than a single one, we can boost the performance of NL. As shown in Fig.3 (d), NPN attains superior performance when compared to previous approaches (*i.e.*, NLNL and JNPL).

Overall Framework

In summary, we merge partial label learning and negative learning to cope with noisy samples without demanding any strong prior assumptions. This approach is implemented in an adaptive, data-driven manner, capitalizing on both the direct supervision in PLL and the indirect supervision in NL to fortify the robustness of the model. The overall learning procedure of our NPN is illustrated in Fig. 2 and Algorithm 1. The objective loss function in our NPN is as follows:

$$\mathcal{L} = \mathcal{L}_{PLL} + \alpha\mathcal{L}_{NL} + \beta\mathcal{L}_{REG}, \quad (10)$$

where α and β are weighting factors.

\mathcal{L}_{REG} denotes the consistency regularization (CR) loss, which encourages prediction consistency between weakly- (A_w) and strongly-augmented (A_s) views of samples. This loss term implicitly enhances the robustness of the model by regularizing it as follows:

$$\mathcal{L}_{REG} = -\frac{1}{N} \sum_{n=1}^N \hat{y}_n \log(p(A_s(x_n), \theta)), \quad (11)$$

in which $\hat{y}_n = \arg \max p^j(A_w(x_n), \theta)$.

Experiments

In this section, we evaluate our NPN by comparing it with SOTA methods. We also conduct ablation studies to validate the effectiveness of each design choice in NPN. The experiments are conducted on a synthetically corrupted dataset, namely CIFAR100N (Krizhevsky 2009), as well as three real-world datasets, namely Web-Aircraft, Web-Car, and Web-Bird (Sun et al. 2021). Experiments follow the setup employed in (Sun et al. 2022a).

Experiment Setup

Synthetically Corrupted Datasets. We conduct experiments on CIFAR100N, which is derived from CIFAR100. CIFAR100 contains 100 classes, consisting of 50,000 training images and 10,000 test images. We follow (Sun et al. 2022a) and employ two approaches to generate noisy labels with a noise rate n : Symmetric noise and Asymmetric noise. **Real-World Datasets.** We also conduct experiments on real-world datasets (*i.e.*, Web-Aircraft, Web-Car, and Web-Bird). These datasets consist of training images obtained through web image search engines, resulting in inevitable label noise. Compared to the synthetically corrupted dataset, real-world datasets exhibit a more complex and realistic challenge in practical scenarios. They contain various types of label noise (*i.e.*, symmetric noise, asymmetric noise, and open-set noise).

Experiment Details. Following (Sun et al. 2022a), we use a seven-layer CNN network as the backbone for CIFAR100N. The network is trained using an SGD optimizer with a momentum of 0.9 for 300 epochs. During the first 100 epochs, a fixed learning rate (*i.e.*, 0.005) is used for warm-up training, followed by 200 epochs of training using a cosine-decay learning rate (starting from 0.005) for robust training. The batch size is 128. For real-world datasets, we leverage ResNet50 (He et al. 2016) pre-trained on ImageNet as our backbone. The network is also trained using an SGD optimizer with a momentum of 0.9 for 100 epochs. The batch size is 32, and the learning rate is 0.001. The learning rate decays in a cosine annealing manner.

Baselines. We compare NPN with the following SOTA methods for CIFAR100N: Decoupling (Malach and Shalev-Shwartz 2017), Co-teaching (Han et al. 2018), Co-teaching+ (Yu et al. 2019), JoCoR, Co-LDL, SPRL (Shi et al. 2023) and AGCE (Zhou et al. 2023). Besides, we compare NPN with the following SOTA methods for real-world datasets: PENCIL (Yi and Wu 2019), DivideMix (Li, Socher, and Hoi 2020) AFM (Peng et al. 2020), and Self-adaptive (Huang, Zhang, and Zhang 2020)). [†] means that we re-implement the method using its open-sourced code and default hyperparameters. Additionally, we provide results of using noisy training directly for training (denoted as Standard).

Evaluation on Synthetically Corrupted Datasets

The performance comparison on CIFAR100N is summarized in Table 1, clearly demonstrating the consistent superiority of both NPN-hard and NPN-soft approaches. Results of existing methods shown in Table 1 are mainly obtained from Co-LDL (Sun et al. 2022a). Our proposed NPN-hard outperforms all the competing methods across all noise types and

Methods	Symmetric				Asymmetric			
	10%	20%	40%	80%	10%	20%	40%	50%
Standard	43.27	35.50	21.02	3.84	45.89	40.86	28.43	23.40
Decoupling (Malach et al. 2017)	50.39	43.84	33.08	9.33	51.59	46.36	33.58	27.30
Co-teaching (Han et al. 2018)	57.69	56.21	52.09	22.83	57.31	52.78	37.26	25.99
Co-teaching+ (Yu et al. 2019)	54.99	52.87	45.64	18.55	54.26	51.25	38.78	27.46
JoCoR (Wei et al. 2020)	60.75	58.69	52.16	14.18	59.38	56.10	38.24	22.41
Co-LDL (Sun et al. 2022a)	61.55	59.73	53.56	25.12	61.05	60.40	52.28	30.50
SPRL [†] (Shi et al. 2023)	56.11	53.64	48.84	22.30	55.74	56.07	49.48	26.61
AGCE [†] (Zhou et al. 2023)	60.25	59.38	54.47	27.41	60.34	59.07	43.04	24.78
NPN-soft	64.75	62.76	58.80	31.69	64.52	63.55	57.11	33.23
NPN-hard	66.79	65.27	61.35	36.88	67.19	66.36	60.11	33.19

Table 1: Average test accuracy (%) on CIFAR100N over the last ten epochs. Experiments are conducted under various noise conditions (“Symmetric” and “Asymmetric” denote the symmetric and asymmetric label noise, respectively).

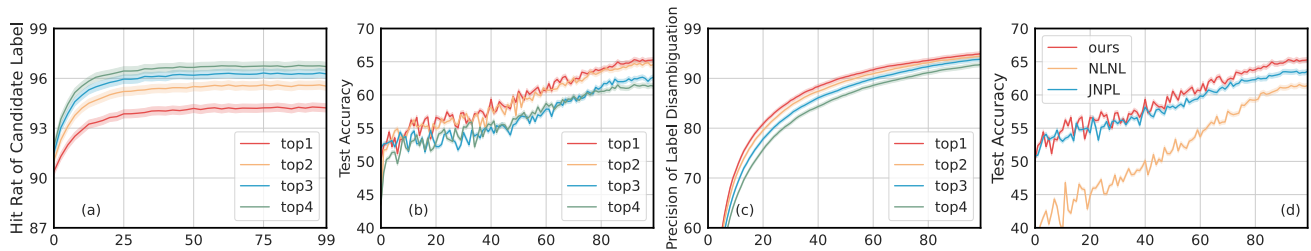


Figure 3: Comparison of label disambiguation using different top-k predicted labels (a-c) and comparison of different negative learning methods (d). (a) The overall hit rate of ground-truth labels within candidate labels (%) vs. epochs. (b) The test accuracy (%) vs. epochs. (c) The overall precision of label disambiguation (%) vs. epochs. (d) The test accuracy (%) of using different negative learning methods (%) vs. epochs.

noise rates. Similarly, NPN-soft also achieves a performance boost over existing methods, although it yields lower performance than NPN-hard in most cases. In the most challenging scenario (*i.e.*, Symmetric-80%), NPN-hard still outperforms other methods, achieving an accuracy of 36.88% (an 11.76% improvement). These experiments on CIFAR100N demonstrate the robustness of NPN in handling both symmetric and asymmetric label noise across various noise rates.

Evaluation on Real-World Datasets

In addition to the experimental results on synthetic datasets, Table 2 compares NPN with SOTA methods on three real-world datasets. Results of SOTA methods shown in Table 2 are also drawn from Co-LDL (Sun et al. 2022a). As shown in Table 2, our NPN-hard achieves accuracies of 86.02%, 80.91%, and 88.26% on Web-Aircraft, Web-Bird, and Web-Car, respectively, surpassing Co-LDL (Sun et al. 2022a) by 4.05%, 0.80%, and 1.31%. The results demonstrate the efficacy of our NPN in more challenging real-world scenarios. It showcases its adaptive, data-driven nature that operates effectively without additional prior knowledge, such as a pre-defined drop rate or threshold.

Ablation Studies

In this subsection, we conduct experiments to investigate the effectiveness of each component in NPN. Unless otherwise

stated, ablation experiments are performed based on our NPN-hard on CIFAR100N (Sym-20%). Table 3 and Fig. 4 show the results of the ablation experiments and provide insights into the contribution of each component in NPN.

Effects of Partial Label Learning. As stated above, we focus on combating label noise by transforming noisy label learning into partial label learning and negative learning. For PLL, we design a candidate label set \mathcal{S} to store the candidate labels for each sample during the training process. We evaluate two paradigms of label disambiguation based on the candidate label set: hard disambiguation and soft disambiguation. The results in Table 1 demonstrate that, in most cases, hard disambiguation outperforms soft disambiguation, except for a few extreme scenarios (*i.e.*, Asymmetric-50%). Furthermore, as shown in Table 3, incorporating PLL leads to a 5.92% performance gain.

To further validate the effect of label disambiguation, we present the comparison of label disambiguation among the top-k predicted labels. Fig.3 (c) demonstrates that selecting more labels into candidate labels increases the difficulty of label disambiguation. Consequently, it results in a decline in test accuracy, as shown in Fig.3 (a) and Fig.3 (b). This rationale elucidates our design to choose the given label and category with the highest prediction confidence as candidate labels during each epoch.

Effects of Negative Learning. We generate the complemen-

Methods	Backbone	Performances(%)			
		Web-Aircraft	Web-Bird	Web-Car	Average
Standard	ResNet50	60.80	64.40	60.60	61.93
Decoupling (Malach et al. 2017)	ResNet50	75.91	71.61	79.41	75.64
Co-teaching (Han et al. 2018)	ResNet50	79.54	76.68	84.95	80.39
Co-teaching+ (Yu et al. 2019)	ResNet50	74.80	70.12	76.77	73.90
PENCIL (Yi and Wu 2019)	ResNet50	78.82	75.09	81.68	78.53
JoCoR (Wei et al. 2020)	ResNet50	80.11	79.19	85.10	81.47
DivideMix (Li et al. 2020)	ResNet50	82.48	74.40	84.27	80.38
AFM (Peng et al. 2020)	ResNet50	81.04	76.35	83.48	80.29
Self-adaptive (Huang et al. 2020)	ResNet50	77.92	78.49	78.19	78.20
Co-LDL (Sun et al. 2022a)	ResNet50	81.97	80.11	86.95	83.01
SPRL [†] (Shi et al. 2023)	ResNet50	80.77	80.41	86.17	82.45
AGCE [†] (Zhou et al. 2023)	ResNet50	84.22	75.60	85.16	81.66
NPN-soft	ResNet50	83.65	79.36	85.46	82.82
NPN-hard	ResNet50	86.02	80.91	88.26	85.06

Table 2: Comparison with SOTA approaches in test accuracy (%) on real-world noisy datasets: Web-Aircraft, Web-Bird, Web-Car. Results of existing methods are mainly drawn from (Sun et al. 2022a).

Standard	NL	PLL	CR	Test Accuracy
✓				35.50
✓	✓			54.70
✓	✓	✓		60.62
✓	✓	✓	✓	65.27

Table 3: Effects of different ingredients in test accuracy (%) on CIFAR100N (noise type and noise rate are “symmetric” and 20%, respectively).

tary label for NL and boost the model robustness by the indirect supervision of NL. Different from the existing methods, we focus on mining more reliable complementary labels and utilizing all complementary labels to enhance the performance of NL. Table 3 illustrates that adopting NL boosts model performance by 19.2% compared to the baseline method (*i.e.*, Standard). We also conduct experiments by replacing our complementary label generation with typical generation methods that randomly select a non-given label (*i.e.*, NLNL and JNPL) in Fig.3 (d). It is evident that our NPN achieves the highest test accuracy.

Effects of Consistency Regularization. During the later stage of model training, it is inevitable that the model will overfit some noisy samples, which results in the inability to select reliable candidate labels and complementary labels. Accordingly, we propose to impose additional consistency regularization to achieve enhancement in both feature extraction and model prediction. As shown in Table 3, the results demonstrate that consistency regularization successfully boosts the model performance by 4.65%.

Effects of Hyper-parameters. Our NPN combines partial label learning and negative learning to effectively learn with noisy labels without strong prior assumptions. In our approach, we only use hyper-parameters (*i.e.*, α and β) to control the weights of different loss terms in Eq. (10). Fig. 4

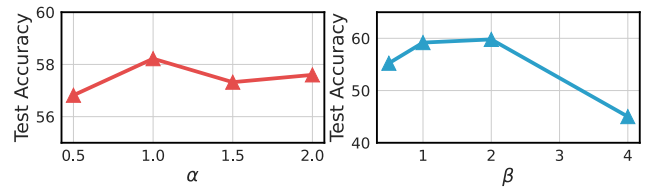


Figure 4: The hyper-parameter sensitivities of α and β .

provides an ablation study on the impact of different α and β settings. When α is 1.0, and β is 2.0, NPN achieves the highest performance.

Conclusion

This paper introduced a simple yet effective framework, termed NPN, designed to learn with noisy labels. It integrated two other distinct learning paradigms, namely partial label learning and negative learning, to effectively combat label noise. Different from traditional noisy label learning methods, NPN was implemented in an adaptive and data-driven manner. We started by decomposing the given label space into a combination of candidate labels for PLL and complementary labels for NL. For PLL, we proposed two adaptive, data-driven paradigms of label disambiguation: hard disambiguation and soft disambiguation. Additionally, we suggested generating reliable complementary labels using all non-candidate labels for NL to enhance model robustness through indirect supervision. Finally, we introduced a consistency regularization term that encouraged prediction agreement between different sample augmentations to improve both feature extraction and model prediction. Extensive experiments and ablation analysis confirmed the effectiveness and superiority of our proposed method.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62102182, 62202227, 62302217), Natural Science Foundation of Jiangsu Province (No. BK20220938, BK20220934, BK20220936), China Postdoctoral Science Foundation (No. 2022M721628, 2022M711635, 2022M721626), Jiangsu Funding Program for Excellent Postdoctoral Talent (No. 2022ZB271, 2022ZB267), and Fundamental Research Funds for the Central Universities (No.30923010303).

References

- Albert, P.; Arazo, E.; Krishna, T.; O'Connor, N. E.; and McGuinness, K. 2023. Is your noise correction noisy? PLS: Robustness to label noise with two stage detection. In *IEEE Winter Conference on Applications of Computer Vision*.
- Chen, H.; Tao, R.; Fan, Y.; Wang, Y.; Wang, J.; Schiele, B.; Xie, X.; Raj, B.; and Savvides, M. 2023a. SoftMatch: Addressing the Quantity-Quality Trade-off in Semi-supervised Learning. In *International Conference on Learning Representations*.
- Chen, T.; Yao, Y.; and Tang, J. 2023. Multi-Granularity Denoising and Bidirectional Alignment for Weakly Supervised Semantic Segmentation. *IEEE Transactions on Multimedia*, 32: 2960–2971.
- Chen, T.; Yao, Y.; Zhang, L.; Wang, Q.; Xie, G.; and Shen, F. 2023b. Saliency Guided Inter- and Intra-Class Relation Constraints for Weakly Supervised Semantic Segmentation. *IEEE Transactions on Multimedia*, 25: 1727–1737.
- Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.
- Feng, L.; and An, B. 2019. Partial Label Learning with Self-Guided Retraining. In *AAAI Conference on Artificial Intelligence*, 3542–3549.
- Fergus, R.; Fei-Fei, L.; Perona, P.; and Zisserman, A. 2010. Learning Object Categories From Internet Image Searches. *Proc. IEEE*, 1453–1466.
- Gong, C.; Ding, Y.; Han, B.; Niu, G.; Yang, J.; You, J.; Tao, D.; and Sugiyama, M. 2023. Class-Wise Denoising for Robust Learning Under Label Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2835–2848.
- Gong, X.; Yuan, D.; and Bao, W. 2022. Partial Multi-Label Learning via Large Margin Nearest Neighbour Embeddings. In *AAAI Conference on Artificial Intelligence*, 6729–6736.
- Han, B.; Yao, Q.; Yu, X.; Niu, G.; Xu, M.; Hu, W.; Tsang, I. W.; and Sugiyama, M. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *Advances in Neural Information Processing Systems*, 8536–8546.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
- Huang, L.; Zhang, C.; and Zhang, H. 2020. Self-Adaptive Training: beyond Empirical Risk Minimization. In *Advances in Neural Information Processing Systems*.
- Karim, N.; Rizve, M. N.; Rahnavard, N.; Mian, A.; and Shah, M. 2022. UNICON: Combating Label Noise Through Uniform Selection and Contrastive Learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 9666–9676.
- Kim, Y.; Yim, J.; Yun, J.; and Kim, J. 2019. NLNL: Negative Learning for Noisy Labels. In *IEEE International Conference on Computer Vision*, 101–110.
- Kim, Y.; Yun, J.; Shon, H.; and Kim, J. 2021. Joint Negative and Positive Learning for Noisy Labels. In *IEEE Conference on Computer Vision and Pattern Recognition*, 9442–9451.
- Krizhevsky, A. 2009. Learning Multiple Layers of Features from Tiny Images.
- Li, J.; Li, G.; Liu, F.; and Yu, Y. 2022a. Neighborhood Collective Estimation for Noisy Label Identification and Correction. In *European Conference on Computer Vision*, 128–145.
- Li, J.; Socher, R.; and Hoi, S. C. 2020. DivideMix: Learning with Noisy Labels as Semi-supervised Learning. In *International Conference on Learning Representations*.
- Li, J.; Xiong, C.; and Hoi, S. C. H. 2021. Learning from Noisy Data with Robust Representation Learning. In *IEEE International Conference on Computer Vision*, 9465–9474.
- Li, S.; Xia, X.; Ge, S.; and Liu, T. 2022b. Selective-Supervised Contrastive Learning with Noisy Labels. In *IEEE Conference on Computer Vision and Pattern Recognition*, 316–325.
- Liu, H.; Zhang, C.; Yao, Y.; Wei, X.; Shen, F.; Tang, Z.; and Zhang, J. 2022a. Exploiting Web Images for Fine-Grained Visual Recognition by Eliminating Open-Set Noise and Utilizing Hard Examples. *IEEE Transactions on Multimedia*, 24: 546–557.
- Liu, H.; Zhang, H.; Lu, J.; and Tang, Z. 2022b. Exploiting Web Images for Fine-Grained Visual Recognition via Dynamic Loss Correction and Global Sample Selection. *IEEE Transactions on Multimedia*, 24: 1105–1115.
- Liu, S.; Niles-Weed, J.; Razavian, N.; and Fernandez-Granda, C. 2020. Early-Learning Regularization Prevents Memorization of Noisy Labels. In *Advances in Neural Information Processing Systems*.
- Malach, E.; and Shalev-Shwartz, S. 2017. Decoupling “when to update” from “how to update”. In *Advances in Neural Information Processing Systems*, 960–970.
- Patel, D.; and Sastry, P. S. 2023. Adaptive Sample Selection for Robust Learning under Label Noise. In *IEEE Winter Conference on Applications of Computer Vision*.
- Peng, X.; Wang, K.; Zeng, Z.; Li, Q.; Yang, J.; and Qiao, Y. 2020. Suppressing Mislabeled Data via Grouping and Self-attention. In *European Conference on Computer Vision*, 786–802.
- Ren, M.; Zeng, W.; Yang, B.; and Urtasun, R. 2018. Learning to Reweight Examples for Robust Deep Learning. In *International Conference on Machine Learning*, 4331–4340.
- Rizve, M. N.; Duarte, K.; Rawat, Y. S.; and Shah, M. 2021. In Defense of Pseudo-Labeling: An Uncertainty-Aware Pseudo-label Selection Framework for Semi-Supervised

- Learning. In *International Conference on Learning Representations*.
- Shi, X.; Guo, Z.; Li, K.; Liang, Y.; and Zhu, X. 2023. Self-paced resistance learning against overfitting on noisy labels. *Pattern Recognition*, 134: 109080.
- Sun, Z.; Liu, H.; Wang, Q.; Zhou, T.; Wu, Q.; and Tang, Z. 2022a. Co-LDL: A Co-Training-Based Label Distribution Learning Method for Tackling Label Noise. *IEEE Transactions on Multimedia*, 1093–1104.
- Sun, Z.; Shen, F.; Huang, D.; Wang, Q.; Shu, X.; Yao, Y.; and Tang, J. 2022b. PNP: Robust Learning From Noisy Labels by Probabilistic Noise Prediction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 5311–5320.
- Sun, Z.; Yao, Y.; Wei, X.-S.; Zhang, Y.; Shen, F.; Wu, J.; Zhang, J.; and Shen, H.-T. 2021. Webly Supervised Fine-Grained Recognition: Benchmark Datasets and An Approach. In *IEEE International Conference on Computer Vision*, 10602–10611.
- Wang, D.; Zhang, M.; and Li, L. 2022. Adaptive Graph Guided Disambiguation for Partial Label Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 8796–8811.
- Wang, H.; Xia, M.; Li, Y.; Mao, Y.; Feng, L.; Chen, G.; and Zhao, J. 2022. SoLar: Sinkhorn Label Refinery for Imbalanced Partial-Label Learning. In *Advances in Neural Information Processing Systems*.
- Wang, Y.; Chen, H.; Heng, Q.; Hou, W.; Fan, Y.; Wu, Z.; Wang, J.; Savvides, M.; Shinozaki, T.; Raj, B.; Schiele, B.; and Xie, X. 2023. FreeMatch: Self-adaptive Thresholding for Semi-supervised Learning. In *International Conference on Learning Representations*.
- Wei, H.; Feng, L.; Chen, X.; and An, B. 2020. Combating Noisy Labels by Agreement: A Joint Training Method with Co-Regularization. In *IEEE Conference on Computer Vision and Pattern Recognition*, 13723–13732.
- Wei, J.; Liu, H.; Liu, T.; Niu, G.; Sugiyama, M.; and Liu, Y. 2022. To Smooth or Not? When Label Smoothing Meets Noisy Labels. In *International Conference on Machine Learning*, 23589–23614.
- Welinder, P.; Branson, S.; Belongie, S. J.; and Perona, P. 2010. The Multidimensional Wisdom of Crowds. In *Advances in Neural Information Processing Systems*, 2424–2432.
- Xia, X.; Han, B.; Wang, N.; Deng, J.; Li, J.; Mao, Y.; and Liu, T. 2023. Extended ST \bar{T} : Learning With Mixed Closed-Set and Open-Set Noisy Labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3047–3058.
- Xia, X.; Liu, T.; Han, B.; Gong, M.; Yu, J.; Niu, G.; and Sugiyama, M. 2022. Sample Selection with Uncertainty of Losses for Learning with Noisy Labels. In *International Conference on Learning Representations*.
- Xu, N.; Liu, Y.; and Geng, X. 2020. Partial Multi-Label Learning with Label Distribution. In *AAAI Conference on Artificial Intelligence*, 6510–6517.
- Xu, N.; Qiao, C.; Geng, X.; and Zhang, M. 2021. Instance-Dependent Partial Label Learning. In *Advances in Neural Information Processing Systems*, 27119–27130.
- Yan, Y.; and Guo, Y. 2021. Adversarial Partial Multi-Label Learning with Label Disambiguation. In *AAAI Conference on Artificial Intelligence*, 10568–10576.
- Yan, Y.; and Guo, Y. 2023. Mutual Partial Label Learning with Competitive Label Noise. In *International Conference on Learning Representations*.
- Yang, E.; Yao, D.; Liu, T.; and Deng, C. 2022. Mutual Quantization for Cross-Modal Search with Noisy Labels. In *IEEE Conference on Computer Vision and Pattern Recognition*, 7541–7550.
- Yao, Y.; Sun, Z.; Zhang, C.; Shen, F.; Wu, Q.; Zhang, J.; and Tang, Z. 2021. Jo-SRC: A Contrastive Approach for Combating Noisy Labels. In *IEEE Conference on Computer Vision and Pattern Recognition*, 5192–5201.
- Yi, K.; and Wu, J. 2019. Probabilistic End-To-End Noise Correction for Learning With Noisy Labels. In *IEEE Conference on Computer Vision and Pattern Recognition*, 7017–7025.
- Yu, X.; Han, B.; Yao, J.; Niu, G.; Tsang, I. W.; and Sugiyama, M. 2019. How does Disagreement Help Generalization against Label Corruption? In *International Conference on Machine Learning*, 7164–7173.
- Zhang, H.; Cissé, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2018. mixup: Beyond Empirical Risk Minimization. In *International Conference on Learning Representations*.
- Zhang, M.; and Yu, F. 2015. Solving the Partial Label Learning Problem: An Instance-Based Approach. In *International Joint Conference on Artificial Intelligence*, 4048–4054.
- Zheng, G.; Awadallah, A. H.; and Dumais, S. T. 2021. Meta Label Correction for Noisy Label Learning. In *AAAI Conference on Artificial Intelligence*, 11053–11061.
- Zhou, X.; Liu, X.; Zhai, D.; Jiang, J.; and Ji, X. 2023. Asymmetric Loss Functions for Noise-Tolerant Learning: Theory and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7): 8094–8109.