# Foster Adaptivity and Balance in Learning with Noisy Labels

# 在带噪声标签的学习中培养适应性和平衡性

Mengmeng Sheng [1] , Zeren Sun [1(∞)] , Tao Chen [1] , Shuchao Pang [1] , Yucheng Wang [2] , and Yazhou Yao [1(⋈)]
,

孟孟盛 [1] , 孙泽仁 [1(∞)] , 陈涛 [1] , 庞树超 [1] , 王宇成 [2] , 姚亚州 [1(⋈)] ,

[1] Nanjing University of Science and Technology, Nanjing, China

[1] 南京理工大学，南京，中国

{shengmengmemg, zerens, taochen, pangshuchao, yazhou.yao}@njust.edu.cn

{shengmengmemg, zerens, taochen, pangshuchao, yazhou.yao}@njust.edu.cn

[2] Horizon Robotics, Beijing, China

[2] Horizon Robotics，北京，中国

yucheng. wang@horizon.cc

yucheng.wang@horizon.cc

Abstract. Label noise is ubiquitous in real-world scenarios, posing a practical challenge to supervised models due to its effect in hurting the generalization performance of deep neural networks. Existing methods primarily employ the sample selection paradigm and usually rely on dataset-dependent prior knowledge (e.g., a pre-defined threshold) to cope with label noise, inevitably degrading the adaptivity. Moreover, existing methods tend to neglect the class balance in selecting samples, leading to biased model performance. To this end, we propose a simple yet effective approach named SED to deal with label noise in a Self-adaptivE and class-balanceD manner. Specifically, we first design a novel sample selection strategy to empower self-adaptivity and class balance when identifying clean and noisy data. A mean-teacher model is then employed to correct labels of noisy samples. Subsequently, we propose a self-adaptive and class-balanced sample re-weighting mechanism to assign different weights to detected noisy samples. Finally, we additionally employ consistency regularization on selected clean samples to improve model generalization performance. Extensive experimental results on synthetic and real-world datasets demonstrate the effectiveness and superiority of our proposed method. The source code has been made anonymously available at https://github.com/NUST-Machine-Intelligence-Laboratory/SED

摘要。标签噪声在实际场景中普遍存在，给监督模型带来实际挑战，因为它会损害深度神经网络 (Deep Neural Networks, DNNs) 的泛化性能。现有方法主要采用样本选择策略，通常依赖于数据集相关的先验知识 (例如预定义的阈值) 来应对标签噪声，不可避免地降低了适应性。此外，现有方法在样本选择时往往忽略类别平衡，导致模型性能偏差。为此，我们提出一种简单而有效的方法，命名为 SED，以自适应和类别平衡的方式处理标签噪声。具体而言，我们首先设计一种新颖的样本选择策略，以增强在识别干净和噪声数据时的自适应性和类别平衡。然后，采用均值教师模型 (mean-teacher model) 对噪声样本的标签进行校正。随后，我们提出一种自适应且类别平衡的样本重加权机制，为检测到的噪声样本分配不同的权重。最后，我们在选中的干净样本上额外采用一致性正则化，以提升模型的泛化能力。在合成数据和真实数据集上的大量实验结果证明了我们所提出方法的有效性和优越性。源代码已在 https://github.com/NUST-Machine-Intelligence-Laboratory/SED 匿名公开。

Keywords: Noisy labels - Self-adaptive - Class-balanced - Sample selection and re-weighting

关键词: 噪声标签 - 自适应 - 类别平衡 - 样本选择与重加权

## 1 Introduction

## 1 引言

Deep neural networks (DNNs) have witnessed remarkable achievements in many computer vision tasks, such as image classification [24, 39, object detection [42, 44], face recognition [5], and instance segmentation [8, 9]. The superior performance of DNNs is highly attributed to supervised training with large-scale and high-quality human-labeled training datasets (e.g., ImageNet [12]). However, collecting large-scale datasets with accurate annotations is expensive and time-consuming, especially for tasks requiring expert annotation knowledge (e.g., medical images [61]). To alleviate this problem, researchers start to resort to alternative methods, such as crowd-sourcing platforms [60] or web image search engines [14], for obtaining cheaper label annotations. Unfortunately, these methods usually result in unavoidable noisy labels, which tend to cause inferior model performance due to the strong learning ability of DNNs [70]. Consequently, developing robust models for learning with noisy labels is of significant importance.

深度神经网络 (DNNs) 在许多计算机视觉任务中取得了显著成就，如图像分类 [24, 39]、目标检测 [42, 44]、人脸识别 [5] 和实例分割 [8, 9]。深度学习的优异表现很大程度上归功于使用大规模高质量的人工标注训练数据集 (如 ImageNet[12]) 进行监督训练。然而，收集具有准确标注的大规模数据集成本高昂且耗时，尤其是需要专家标注知识的任务 (如医学图像 [61])。为缓解这一问题，研究人员开始采用替代方法，例如众包平台 [60] 或网络图像搜索引擎 [14]，以获取更便宜的标签标注。不幸的是，这些方法通常会引入不可避免的噪声标签，由于深度神经网络强大的学习能力，这些噪声标签往往导致模型性能下降 [70]。因此，开发在带噪声标签条件下具有鲁棒性的学习模型具有重要意义。

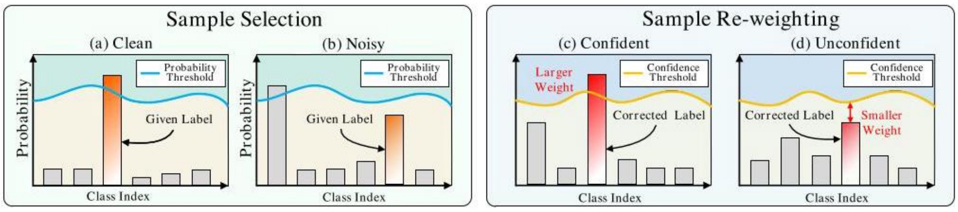


Fig. 1: (a-b) Self-adaptive and class-balanced sample selection based on predicted probability w.r.t. given labels. The blue curve indicates the class-specific selection thresholds. (c-d) Self-adaptive and class-balanced sample re-weighting based on correction confidence. The orange curve represents the class-specific confidence threshold.

图 1:(a-b) 基于预测概率相对于给定标签的自适应和类别平衡样本选择。蓝色曲线表示类别特定的选择阈值。(c-d) 基于校正置信度的自适应和类别平衡样本重加权。橙色曲线代表类别特定的置信度阈值。

Recently, a growing number of methods have been proposed for addressing the label noise problem [2, 4, 6, 17, 30, 54, 59, 62, 65]. Label correction and sample selection/re-weighting are two major strategies for tackling noisy labels. Label correction methods typically attempt to rectify labels using the noise transition matrix [15] or model predictions [29]. For example, 40 proposes to correct corrupted labels by estimating the noise transition matrix. Jo-SRC [66] uses the temporally averaged model (i.e., mean-teacher model) to generate reliable pseudo-label distributions for providing supervision. However, on the one hand, the noise transition matrix is hard to estimate in real-world scenarios. On the other hand, networks tend to have better recognition capability on simple categories than hard ones. This recognition bias usually results in imbalanced label corrections (i.e., samples are more likely to be corrected into simple categories) in prediction-based label correction methods, hurting the final model performance.

近年来，针对标签噪声问题，提出的方法不断增加 [2, 4, 6, 17, 30, 54, 59, 62, 65]。标签校正和样本选择/重加权是应对噪声标签的两大主要策略。标签校正方法通常试图利用噪声转移矩阵 [15] 或模型预测 [29] 来修正标签。例如，文献 [40] 提出通过估计噪声转移矩阵来校正被污染的标签。Jo-SRC[66] 使用时间平均模型 (即均值教师模型) 生成可靠的伪标签分布以提供监督。然而，一方面，噪声转移矩阵在实际场景中难以估计；另一方面，网络在识别简单类别时的能力通常优于困难类别。这种识别偏差常导致基于预测的标签校正方法出现类别不平衡 (即样本更可能被校正为简单类别)，从而影响最终模型性能。

Another line of research focuses on the sample selection/re-weighting [19, 23, 32, 45, 49, 52, 66, 67]. Sample selection methods primarily seek to split samples into two subsets: a noisy subset and a clean subset [18, 19,66]. Prior methods tend to regard samples with small losses as clean ones [19,58]. For example, JoCoR [58] exploits a joint loss to select small-loss samples to encourage agreement between models. However, these methods often require proper prior knowledge (e.g., a pre-defined drop rate or threshold) to achieve effective sample selection. Moreover, previous literature usually neglects class balance during sample selection, leading to biased model performance. Sample re-weighting can be deemed as a variant of sample selection, smoothing its $0/1$ weighting scheme to a softer one. Samples with higher confidence are assigned larger weights, while those with lower confidence are assigned smaller weights. For example, L2RW [43] proposes to assign different sample weights based on meta-learning. However, existing sample re-weighting methods also tend to require prior knowledge (e.g., a small subset of clean samples).

另一项研究集中在样本选择/重加权 [19, 23, 32, 45, 49, 52, 66, 67]。样本选择方法主要旨在将样本划分为两个子集: 噪声子集和干净子集 [18, 19,66]。以往的方法倾向于将损失较小的样本视为干净样本 [19,58]。例如，JoCoR[58] 利用联合损失选择损失较小的样本，以促进模型之间的一致性。然而，这些方法通常需要事先的先验知识 (例如预定义的丢弃率或阈值) 以实现有效的样本选择。此外，之前的文献通常忽略了样本选择过程中的类别平衡，导致模型性能偏差。样本重加权可以看作是样本选择的变体，它通过平滑其 0/1 加权方案，使之更柔和。置信度较高的样本被赋予更大的权重，而置信度较低的样本则被赋予较小的权重。例如，L2RW[43] 提出根据元学习 (meta-learning) 为不同样本分配不同的权重。然而，现有的样本重加权方法也往往需要先验知识 (例如一小部分干净样本)。

To alleviate the aforementioned issues, we propose a simple yet effective method, named SED, to learn with noisy labels in a Self-adaptivE and class-balanceD manner. Our SED integrates sample selection, label correction, and sample re-weighting. Specifically, we propose to identify clean samples based on the predicted probability w.r.t. the given labels of input samples. To promote self-adaptivity and class balance in sample selection, we propose to integrate global and local thresholds for each category when distinguishing between clean and noisy data (as shown in Fig. 1 (a) and (b)). The global and local thresholds are dynamically updated during training. Once the clean and noisy subsets are obtained, we employ a mean-teacher model to correct labels for identified noisy samples. Subsequently, we propose to re-weight label-corrected noisy samples in a self-adaptive and class-balanced fashion to alleviate the confirmation bias caused by imbalanced label correction. We impose larger/smaller weights on noisy samples with higher/lower correction confidence according to an estimated truncated normal distribution (as shown in Fig. 1 (c) and (d)). Finally, we employ an additional regularization loss term on identified clean samples to further enhance the performance and robustness of the model. Comprehensive experimental results have been provided to verify the effectiveness and superiority of our proposed SED on synthetically corrupted datasets and real-world datasets. Our contributions are summarized as follows:

为了解决上述问题, 我们提出了一种简单而有效的方法, 命名为 SED(Self-adaptivE and class-balanceD learning)，以自适应和类别平衡的方式学习带有噪声标签的数据。我们的 SED 结合了样本选择、标签校正和样本重加权。具体而言，我们提出根据输入样本的预测概率与给定标签的关系，识别干净样本。为了促进样本选择的自适应性和类别平衡，我们建议在区分干净与噪声数据时，为每个类别整合全局阈值和局部阈值 (如图 1(a) 和 (b) 所示)。全局和局部阈值在训练过程中动态更新。一旦获得干净和噪声子集，我们采用均值教师模型 (mean-teacher model) 对识别出的噪声样本进行标签校正。随后，我们提出以自适应和类别平衡的方式对标签校正后的噪声样本进行重加权，以缓解由不平衡标签校正引起的确认偏差。我们根据估算的截断正态分布 (如图 1(c) 和 (d) 所示) 对置信度较高/较低的噪声样本赋予较大/较小的权重。最后，我们在识别出的干净样本上加入额外的正则化损失，以进一步提升模型的性能和鲁棒性。大量实验结果验证了我们提出的 SED 在合成噪声数据集和真实世界数据集上的有效性和优越性。我们的贡献总结如下:

(1) We propose a simple yet effective method, named SED, to combat noisy labels. SED selects and re-weights samples in a self-adaptive and class-balanced manner, alleviating the demand for dataset-dependent prior knowledge and the negative effect caused by class imbalance.

(1) 我们提出了一种简单而有效的方法，命名为 SED，用于应对带有噪声的标签。SED 以自适应和类别平衡的方式选择和重加权样本，减轻了对数据集依赖的先验知识需求以及类别不平衡带来的负面影响。

(2) Our proposed SED selects samples according to class-specific thresholds that are estimated in a data-driven

manner, encouraging self-adaptivity and class balance in sample selection. In addition, we propose to re-weight samples based on a truncated normal distribution that is updated periodically, mitigating performance downgrade due to imbalanced label corrections.

(2) 我们提出的 SED 根据类别特定的阈值进行样本选择，这些阈值以数据驱动的方式估算，促进样本选择的自适应性和类别平衡。此外，我们建议基于定期更新的截断正态分布对样本进行重加权，从而减轻不平衡标签校正带来的性能下降问题。

(3) We provide comprehensive experimental results on synthetic and real-world datasets to illustrate the superiority of our proposed SED. Extensive ablation studies are conducted to further verify the effectiveness of our method.

(3) 我们在合成和真实数据集上提供了全面的实验结果，展示了我们提出的 SED 的优越性。还进行了大量消融实验，以进一步验证我们方法的有效性。

## 2 Related Work

## 2 相关工作

Label Correction. The intuitive idea for handling noisy labels is to correct corrupted labels before feeding them into networks [11, 15, 16, 33, 34, 40, 57, 64, 67]. Early works propose to correct the training labels by estimating the noise transition matrix. [67] introduces an intermediate class to avoid directly estimating the noisy class posterior and then factorizes the transition matrix into the product of two sub-matrices. However, the transition matrix is hard to estimate accurately in real-world scenarios. Some other methods propose to model label noise by using predictions of DNNs [27, 55, 56, 68]. [68] proposes to directly learn label distributions for corrupted samples in an end-to-end manner. Nevertheless, since DNNs tend to learn better on simple categories than hard ones, pseudo-labels are more likely to fall into the simple class set, leading to imbalanced label correction. In this work, we resort to the re-weighting strategy to alleviate the issue caused by imbalanced label correction.

标签校正。处理带噪声标签的直观思路是先对受污染的标签进行校正，然后再输入网络 [11, 15, 16, 33, 34, 40, 57, 64, 67]。早期的工作提出通过估算噪声转移矩阵 (noise transition matrix) 来校正训练标签。[67] 引入了中间类别 (intermediate class)，以避免直接估算噪声类别的后验概率，然后将转移矩阵分解为两个子矩阵的乘积。然而，在实际场景中准确估算转移矩阵较为困难。其他一些方法则试图利用深度神经网络 (DNNs) 的预测结果对标签噪声进行建模 [27, 55, 56, 68]。[68] 提出端到端地直接学习受污染样本的标签分布。然而，由于 DNNs 倾向于在简单类别上学习得更好，伪标签 (pseudo-labels) 更容易落入简单类别集，导致标签校正不平衡。在本工作中，我们采用重加权策略以缓解由不平衡标签校正引起的问题。

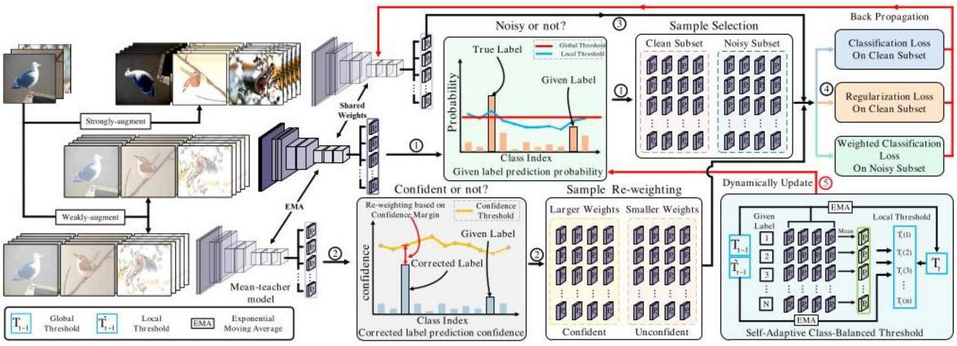


Fig. 2: The overall framework of our SED. We first divide the training set into a clean subset and a noisy subset based on global and local thresholds that are dynamically updated. Our threshold design enables self-adaptivity and class balance in sample selection. We then employ a mean-teacher model to correct labels for noisy samples. Based on the correction confidence, SED adaptively assigns different weights to label-corrected noisy samples and uses them for training. Finally, SED further boosts the model performance by imposing an additional consistency regularization loss on selected clean samples. The final objective loss integrates the classification losses on clean and noisy samples and the regularization loss on clean samples.

图 2: 我们的 SED(噪声检测与纠正) 整体框架。我们首先根据动态更新的全局和局部阈值将训练集划分为干净子集和噪声子集。我们的阈值设计实现了样本选择的自适应性和类别平衡。然后，我们采用均值教师模型对噪声样本的标签进行纠正。基于纠正的置信度，SED 自适应地为标签纠正的噪声样本分配不同的权重，并将其用于训练。最后，SED 通过在选定的干净样本上施加额外的一致性正则化损失，进一步提升模型性能。最终的目标损失融合了干净样本和噪声样本的分类损失以及干净样本的正则化损失。

Sample Selection. Another type of classical method to deal with label noise is sample selection, which divides the training set into a clean subset and a noisy subset [19, 51, 58, 66] . Previous sample selection methods primarily employ the cross-entropy loss as the selection criterion, regarding samples with small losses as clean ones. For example, Co-teaching [19] proposes to cross-update two networks using small-loss samples selected by peer networks. Some recent methods propose new selection criteria for finding clean samples [29,66]. Jo-SRC [66] proposes to employ Jensen-Shannon Divergence for selecting clean samples globally. DISC [31] proposes to select reliable instances based on the insight of memorization strength. However, these methods usually demand pre-defined drop rates or thresholds. Furthermore, previous methods neglect the class imbalance issue in the selection process, leading to inferior and biased model performance. In this work, we employ predicted probability as the selection criterion and propose a novel threshold mechanism to enable self-adaptive and class-balanced selection.

样本选择。另一类经典处理标签噪声的方法是样本选择，它将训练集划分为干净子集和噪声子集 [19, 51, 58, 66] 。以往的样本选择方法主要采用交叉熵损失作为选择标准，将损失较小的样本视为干净样本。例如，Co-teaching [19] 提出通过同行网络选择的小损失样本进行交叉更新两个网络。一些最新方法提出了新的选择标准以寻找干净样本 [29,66]。Jo-SRC [66] 建议采用 Jensen-Shannon 散度 (Jensen-Shannon Divergence) 在全局范围内选择干净样本。DISC [31] 基于记忆强度的洞察，提出选择可靠的实例。然而，这些方法通常需要预定义的丢弃率或阈值。此外，之前的方法在选择过程中忽略了类别不平衡问题，导致模型性能较差且偏向某些类别。在本工作中，我们采用预测概率作为选择标准，并提出一种新颖的阈值机制，实现自适应和类别平衡的样本选择。

Sample Re-weighting. Recently, some researchers have been devoted to re-weighting training samples to cope with noisy labels [13, 47, 53, 63]. These methods usually assign larger weights to samples that are more likely to be clean while smaller weights to others, minimizing the misleading impact of noisy samples. For example, L2RW [43] proposes a meta-learning algorithm that learns to assign weights to training examples based on their gradient directions. However, existing methods tend to require considerable prior knowledge (e.g., a small subset of clean samples), posing a limit to their practicability. In this work, we design a novel re-weighting scheme to empower self-adaptivity and class balance when leveraging label-corrected noisy samples.

样本重加权。近年来，一些研究者致力于通过重加权训练样本来应对噪声标签 [13, 47, 53, 63]。这些方法通常为更可能干净的样本赋予较大的权重，而对其他样本赋予较小的权重，以减少噪声样本的误导影响。例如，L2RW [43] 提出一种元学习算法，根据梯度方向学习为训练样本分配权重。然而，现有方法往往需要大量先验知识 (例如一小部分干净样本)，限制了其实用性。在本工作中，我们设计了一种新颖的重加权方案，以在利用标签纠正的噪声样本时实现自适应和类别平衡。

# 3 Method

# 3 方法

## 3.1 Problem Statement

## 3.1 问题描述

Formally, considering a $C$ -class classification problem, we denote $D_{\text{train}} = \{(x_i, y_i) \mid i = 1, \ldots, N\}$ as the training set with label noise, in which $x_i$ denotes the $i$ -th training sample and $y_i \in \{0,1\}^C$ is its associated label (potentially "incorrect"). We use $y_i^*$ to represent the ground-truth label of $x_i$ and denote $D_{\text{test}} = \{(x_i, y_i^*) \mid i = 1, \ldots, M\}$ as the test set with accurate labels. $N$ and $M$ represent the total number of samples in the training set $D_{\text{train}}$ and test set $D_{\text{test}}$ , respectively. The goal is to train a robust classification neural network $\mathcal{F}(\cdot, \theta)$ ($\theta$ denotes network parameters) on the noisy training set $D_{\text{train}}$ to perform accurate prediction on the test set $D_{\text{test}}$ . The conventional classification task usually hypothesizes that given labels of training samples are accurate (i.e., $y_i = y_i^*$ ), thus using the following cross-entropy loss to optimize the network.

正式地，考虑一个 $C$ -类别分类问题，我们用 $D_{\text{train}} = \{(x_i, y_i) \mid i = 1, \ldots, N\}$ 表示带有标签噪声的训练集，其中 $x_i$ 表示第 $i$ 个训练样本，$y_i \in \{0,1\}^C$ 是其相关标签 (可能“错误”)。我们用 $y_i^*$ 表示 $x_i$ 的真实标签，用 $D_{\text{test}} = \{(x_i, y_i^*) \mid i = 1, \ldots, M\}$ 表示带有准确标签的测试集。$N$ 和 $M$ 分别代表训练集 $D_{\text{train}}$ 和测试集 $D_{\text{test}}$ 中的样本总数。目标是在带有噪声的训练集 $D_{\text{train}}$ 上训练一个鲁棒的分类神经网络 $\mathcal{F}(\cdot, \theta)$ ($\theta$ (网络参数) 以对测试集 $D_{\text{test}}$ 进行准确预测。传统的分类任务通常假设训练样本的标签是准确的 (即 $y_i = y_i^*$ )，因此使用以下交叉熵损失来优化网络。

$$\mathcal{L}_{ce} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_i^c \log \left(p^c\left(x_i, \theta\right)\right), \tag{1}$$

in which $p^c\left(x_i, \theta\right)$ denotes the predicted softmax probability of the $i$ -th training sample $x_i$ over its $c$ -th class.

其中 $p^c(x_i, \theta)$ 表示第 $i$ 个训练样本 $x_i$ 在其第 $c$ 个类别上的预测 softmax 概率。

Due to the memorization effect [1] (i.e., models tend to fit clean and simple samples first and then gradually memorize noisy ones), the network optimization based on the above loss usually leads to an ill-suited solution. One potentially useful remedy is to integrate sample selection, label correction, and sample re-weighting. In this work, we follow this paradigm to combat noisy labels by encouraging self-adaptivity and class balance.

由于记忆效应 [1](即模型倾向于先拟合干净且简单的样本，然后逐渐记忆噪声样本)，基于上述损失的网络优化通常会导致不适合的解决方案。一种潜在的有效方法是结合样本选择、标签校正和样本重加权。在本工作中，我们遵循这一范式，通过鼓励自适应性和类别平衡来应对噪声标签的问题。

## 3.2 Adaptive and Balanced Sample Selection

## 3.2 自适应与平衡的样本选择

Previous studies [19, 58, 66] usually select small-loss samples as clean ones based on pre-defined drop rates or thresholds. It should be noted that drop rates can be easily converted to thresholds during selection, thus we only discuss thresholds hereafter. The selection thresholds are usually dataset-dependent, making it challenging to adapt them to different real-world datasets. Although existing methods employ scheduling strategies (e.g., a gradually increasing schedule [19]) to adjust thresholds during training for fully exploiting the model capability, these scheduling designs are rather heuristic and still require pre-defined initial and final threshold values. Moreover, few works consider the different difficulties in learning various categories, leading to biased selection results and inferior model performance.

以往研究 [19, 58, 66] 通常根据预定义的丢弃率或阈值选择小损失样本作为干净样本。需要注意的是，丢弃率可以在选择过程中方便地转换为阈值，因此我们在此仅讨论阈值。选择阈值通常依赖于数据集，难以适应不同的实际数据集。虽然现有方法采用调度策略 (例如逐步增加的调度 [19]) 在训练过程中调整阈值以充分利用模型能力，但这些调度设计较为启发式，仍需预定义初始和最终阈值。此外，很少有工作考虑不同类别学习难度的差异，导致选择结果偏差，模型性能较差。

To this end, we propose a self-adaptive and class-balanced sample selection (SCS) strategy to address the above problems. SCS adaptively adjusts the threshold in an epoch-wise and class-wise manner to enable effective clean sample identification. Specifically, we employ global and local thresholds, which are both self-adaptive, to distinguish between clean and noisy samples in each category. Since the cross-entropy loss is unbounded, we propose to rely on predicted probability w.r.t. the given labels $p^{y_i}(x_i, \theta)$ to determine whether the samples are clean. Samples with higher $p^{y_i}(x_i, \theta)$ are more likely to have correct labels.

为此，我们提出了一种自适应且类别平衡的样本选择 (SCS) 策略，以解决上述问题。SCS 在每个 epoch 和每个类别中自适应调整阈值，以实现有效的干净样本识别。具体而言，我们采用全局阈值和局部阈值，这两者都是自适应的，用于区分每个类别中的干净样本和噪声样本。由于交叉熵损失 (cross-entropy loss) 没有界限，我们建议依赖于与给定标签相关的预测概率 <b0> 来判断样本是否干净。预测概率越高，样本的标签越可能正确。

We estimate the global threshold based on the averaged predicted probability w.r.t. given labels over all training samples to reflect the overall learning state of the network. This design makes the global threshold data-driven, thus eliminating the demand for pre-defined thresholds. Moreover, we employ the exponential moving average (EMA) to further refine the global threshold, alleviating unstable training caused by large perturbation of the averaged predicted probability. By adopting an initial value of $T_0 = \frac{1}{C}$ , our final global threshold at the $t$ -th epoch is defined as:

我们基于所有训练样本的平均预测概率 <b0>，估算全局阈值，以反映网络的整体学习状态。这一设计使得全局阈值具有数据驱动特性，免去了预定义阈值的需求。此外，我们采用指数移动平均 (EMA) 进一步优化全局阈值，缓解由平均预测概率大幅波动引起的不稳定训练。通过设定初始值 <b0>，我们在第 <b1> 轮训练中的最终全局阈值定义为:

$$T_t = \begin{cases} \frac{1}{C}, & t = 0 \\ mT_{t-1} + (1-m)\frac{1}{N}\sum_{i=1}^{N} p^{y_i}(x_i, \theta), & t > 0 \end{cases} \tag{2}$$

Our design of the global threshold scheduling implicitly complies with the memorization effect [1]. As the training progresses, the predicted probability w.r.t. the given label gradually increases, leading to the monotonic increase of $T_t$ . Consequently, the network can learn from more samples in the early stage but fewer samples in the later stage.

我们的全局阈值调度设计隐含符合记忆效应 [1]。随着训练的进行，预测概率相对于给定标签逐渐增加，导致 <b0> 单调上升。因此，网络在早期可以学习更多样本，而在后期则学习较少样本。

As stated above, using only a global threshold to divide the training set neglects the difference among various categories and will result in imbalanced sample selection (i.e., fewer samples of complicated categories will be selected as clean data). Samples of easy categories tend to be better learned and have higher $p^{y_i}(x_i, \theta)$ , thus requiring larger thresholds to distinguish between clean and noisy data. Therefore, we additionally propose a local threshold scheme to further adjust the global threshold. We first estimate the expectation of the model's predictions $\widetilde{E}_t(c)$ on each class $c$ at the $t$ -th epoch to reveal the class-specific learning status.

如上所述，仅使用全局阈值划分训练集忽略了不同类别之间的差异，可能导致样本选择不平衡 (即，复杂类别的样本作为干净数据的比例较少)。容易类别的样本更易学习，且具有更高的 <b0>，因此需要更大的阈值来区分干净与噪声数据。因此，我们还提出一种局部阈值方案，进一步调整全局阈值。我们首先在第 <b3> 轮估算模型对每个类别 <b2> 的预测期望 <b1>，以揭示类别特定的学习状态。

$$\widetilde{E}_t(c) = \begin{cases} \frac{1}{C}, & t = 0 \\ m\widetilde{E}_{t-1}(c) + (1-m)\frac{1}{N}\sum_{i=1}^{N} p^{c}(x_i, \theta), & t > 0 \end{cases} \tag{3}$$

Accordingly, we obtain local threshold $\widetilde{T}_t(c)$ for each class $c$ by normalizing $\widetilde{E}_t(c)$ and integrating it with global threshold $T_t$ as:

据此，我们通过归一化 <b2> 并结合全局阈值 <b3>，为每个类别 <b1> 获得局部阈值 <b0>，其表达式为:

$$\widetilde{T}_t(c)=\frac{\widetilde{E}_t(c)}{\max\left\{\widetilde{E}_t(c:c\in[C]\right\}}T_t. \tag{4}$$

On the one hand, the design of our global threshold ensures that sufficient clean samples are identified and learned by the network. On the other hand, the design of our local threshold ensures that selected clean samples are class-balanced. Finally, by unifying our proposed global and local thresholds, we divide the training set $D_{\text{train}}$ into a clean subset $D_c$ and a noisy subset $D_n$ in each epoch according to Eq. 5).

一方面，我们的全局阈值设计确保网络能够识别并学习足够的干净样本；另一方面，我们的局部阈值设计确保所选的干净样本具有类别平衡性。最终，通过统一我们提出的全局和局部阈值，我们在每个 epoch 中根据式 (5) 将训练集 <b0> 划分为干净子集 <b1> 和噪声子集 <b2>。

$$\begin{cases} D_c=\left\{(x_i,y_i)\mid(x_i,y_i)\in D_{\text{train}},p^{y_i}(x_i,\theta)>\widetilde{T}_t(y_i)\right\} \\ D_n=\{(x_i,y_i)\mid(x_i,y_i)\in D_{\text{train}},(x_i,y_i)\notin D_c\}\end{cases}. \tag{5}$$

## 3.3 Adaptive and Balanced Re-weighting

## 3.3 自适应与平衡的重加权

Recent researches propose to cope with noisy samples in a semi-supervised-learning-like (SSL-like) manner by integrating sample selection and label correction [29, 66]. Identified clean samples are used conventionally for model training, while detected noisy samples are assigned pseudo labels to correct their supervision before being used for training. However, existing methods tend to treat label-corrected noisy samples equally, neglecting their difference in reliability. Moreover, due to different learning difficulties in various categories, label correction results may be imbalanced (noisy samples are more likely to be assigned labels of simple classes), resulting in biased label correction and sub-optimal model performance.

近期研究提出通过结合样本选择和标签校正 [29, 66]，以半监督学习 (SSL-like) 方式应对噪声样本。已识别的干净样本通常用于模型训练，而检测到的噪声样本则被赋予伪标签以校正其监督信息，然后再用于训练。然而，现有方法倾向于将标签校正后的噪声样本一视同仁，忽视了它们在可靠性上的差异。此外，由于不同类别的学习难度不同，标签校正结果可能存在不平衡 (噪声样本更可能被赋予简单类别的标签)，导致偏差的标签校正和次优的模型性能。

To mitigate the above issue, we propose a self-adaptive and class-balanced re-weighting (SCR) mechanism to adaptively assign different weights to samples according to their confidence. Specifically, we use a temporally averaged model (i.e., mean-teacher model $\theta^*$ ) to generate reliable pseudo labels for detected noisy samples. By introducing the historical models, we obtain corrected labels $y^{\text{corr}}$ using $\theta^*$ to promote the reliability of label correction and alleviate error-propagation issues. The mean-teacher model $\theta^*$ is not updated in the gradient back-propagation. $\theta^*$ is updated in each training step $t'$ as follows:

为解决上述问题，我们提出了一种自适应和类别平衡的重加权 (SCR) 机制，根据样本的置信度自适应分配不同的权重。具体而言，我们使用一个时间平均模型 (即均值教师模型 $\theta^*$ ) 为检测到的噪声样本生成可靠的伪标签。通过引入历史模型，我们利用 $\theta^*$ 获得校正标签 $y^{\text{corr}}$ ，以提升标签校正的可靠性并缓解误差传播问题。均值教师模型 $\theta^*$ 在梯度反向传播中不更新。$\theta^*$ 在每个训练步骤 $t'$ 中按如下方式更新:

$$\theta_{t'}^* = \alpha\theta_{t'-1}^* + (1-\alpha)\,\theta_{t'} \tag{6}$$

in which $\theta_0^*$ is initialized using the initial model parameters of $\theta$ . Accordingly, noisy samples are assigned pseudo labels as follows:

其中 $\theta_0^*$ 使用 $\theta$ 的初始模型参数初始化。因此，噪声样本的伪标签赋值如下:

$$y_i^{\text{corr}} = \underset{j=1,\ldots,C}{\arg\max} p^j\left(x_i, \theta^*\right). \tag{7}$$

As mentioned above, the label correction results could be imbalanced due to the biased capability of the network. Consequently, we propose a re-weighting method to adaptively assign larger weights to (noisy) samples with higher correction confidence. We employ the prediction probability w.r.t. the corrected label to reveal the correction confidence. Inspired by the semi-supervised learning methods3,7,10,48, we propose to fit the underlying sample weights to a dynamic truncated normal distribution, whose mean and variance values at the $t$ -th epoch are $\mu_t$ and $\sigma_t$ . The sample weights are therefore derived in a self-adaptive fashion as:

如上所述，由于网络能力的偏差，标签校正结果可能存在不平衡。因此，我们提出一种重加权方法，对具有较高校正置信度 (噪声) 样本赋予更大的权重。我们利用相对于校正标签的预测概率来反映校正置信度。受到半监督学习方法 [3,7,10,48] 的启发，我们建议将样本权重拟合为一个动态截断正态分布，其在第 $t$ 轮的均值和方差分别为 $\mu_t$ 和 $\sigma_t$ 。因此，样本权重以自适应方式导出如下:

$$\lambda\left(x_i\right) = \begin{cases} \lambda_m \exp\left(\frac{\left(p^{y_i^{corr}}(x_i,\theta)-\mu_t\right)^2}{-2\sigma_t^2}\right), p^{y_i^{corr}}\left(x_i, \theta\right) < \mu_t \\ \lambda_m, \end{cases}, \tag{8}$$

in which $\lambda_m$ is the upper bound of sample weights. Assuming sample weights to follow the dynamic truncated normal distribution is equivalent to treating the deviation of correction confidence from $\mu_t$ as a proxy measure of the correctness of the label correction. Samples with higher confidence are less prone to be erroneously label-corrected than those with lower confidence, thus being assigned larger weights.

其中 $\lambda_m$ 是样本权重的上限。假设样本权重遵循动态截断正态分布，将校正置信度偏离 $\mu_t$ 视为标签校正正确性的代理指标。置信度较高的样本比置信度较低的样本更不易被错误校正，因此被赋予更大的权重。

Moreover, to enable class-balanced re-weighting and promote training stability, we propose to estimate $\mu_t\left(c\right)$ and $\sigma_t^2\left(c\right)$ for each class $c$ based on their historical estimations using EMA:

此外，为实现类别平衡的重加权并促进训练稳定性，我们提出基于指数移动平均 (EMA) 的方法，估算每个类别 $c$ 的 $\mu_t\left(c\right)$ 和 $\sigma_t^2\left(c\right)$ ，基于它们的历史估计:

$$\mu_t(c) = \begin{cases} \frac{1}{C}, & t = 0 \\ m\mu_{t-1}(c) + (1-m)\widetilde{\mu}(c), & t > 0 \end{cases} \tag{9}$$

$$\sigma_t^2(c) = \begin{cases} 1.0, & t = 0 \\ m\sigma_{t-1}^2(c) + (1-m)\widetilde{\sigma}^2(c), & t > 0 \end{cases} \tag{10}$$

in which,

其中,

$$\widetilde{\mu}(c) = \frac{1}{|D_n|}\sum_{i=1}^{|D_n|} p^{y_i^{corr}}(x_i, \theta), \text{ if } y_i^{corr} = c, \tag{11}$$

$$\widetilde{\sigma}^2(c) = \frac{1}{|D_n|}\sum_{i=1}^{|D_n|} \left(p^{y_i^{corr}}(x_i, \theta) - \widetilde{\mu}(c)\right)^2, \text{ if } y_i^{corr} = c. \tag{12}$$

$\mu_t$ and $\sigma_t$ of the dynamic truncated normal distribution can be adaptively estimated from the correction confidence distribution based on Eqs. (9) and [10]). As the model performance improves during training, $\mu_t$ gradually increases and $\sigma_t$ decreases. Since the tail of the normal distribution grows exponentially tighter, the samples with lower correction confidence are given lower weights. Besides, we estimate class-specific $\mu_t$ and $\sigma_t$ . This effectively alleviates the class imbalance in the label correction process caused by the biased model ability.

$\mu_t$ 和 $\sigma_t$ 的动态截断正态分布参数可以根据校正置信度分布，利用公式 (9) 和 [10]) 自适应估算。随着训练的进行，模型性能提升，$\mu_t$ 逐渐增大，$\sigma_t$ 逐渐减小。由于正态分布尾部指数级收紧，校正置信度较低的样本被赋予较低的权重。此外，我们还估算类别特定的 $\mu_t$ 和 $\sigma_t$ ，有效缓解由偏差模型能力引起的标签校正过程中的类别不平衡问题。

## 3.4 Overall Framework

## 3.4 整体框架

In summary, our proposed SED follows the paradigm that integrates sample selection, label correction, and sample re-weighting for addressing noisy labels. Details of our SED are shown in Fig. 2 and Algorithm 1.

总之，我们提出的 SED(样本选择、标签校正与重加权结合的框架) 旨在应对带噪声标签的问题。我们的 SED 的详细内容见图 2 和算法 1。

Firstly, SED divides $D_{\text{train}}$ into a clean subset $D_c$ and a noisy subset $D_n$ in a self-adaptive and class-balanced manner. For samples in the clean subset $D_c$ , we take their given labels to calculate the classification loss $\mathcal{L}_{D_c}$ as follow:

首先，SED 以自适应和类别平衡的方式将 $D_{\text{train}}$ 划分为干净子集 $D_c$ 和噪声子集 $D_n$ 。对于干净子集 $D_c$ 中的样本，我们使用其给定的标签计算分类损失 $\mathcal{L}_{D_c}$ ，如下:

$$\mathcal{L}_{D_c} = -\frac{1}{|D_c|}\sum_{(x,y)\in D_c} y \log p(x, \theta). \tag{13}$$

For samples in the noisy subset $D_n$ , we discard their given labels and perform label correction based on a mean-teacher model using Eq. (7). Then, we calculate the loss of the noisy subset $\mathcal{L}_{D_n}$ as

对于噪声子集 $D_n$ 中的样本，我们舍弃其给定标签，基于均值教师模型使用公式 (7) 进行标签校正。然后，计算噪声子集 $\mathcal{L}_{D_n}$ 的损失如下:

$$\mathcal{L}_{D_n} = -\frac{1}{|D_n|} \sum_{(x,y)\in D_n} \lambda(x)\, y^{\text{corr}} \log p(\widehat{x}, \theta), \tag{14}$$

Algorithm 1 Our proposed SED algorithm

算法 1 我们提出的 SED 算法

---

Input: The training set $D_{\text{train}}$ , network $\theta$ , mean-teacher network $\theta^*$ , total epochs

输入: 训练集 $D_{\text{train}}$ ， 网络 $\theta$ ， 教师网络 (mean-teacher network) $\theta^*$ ， 总训练轮数

$E_{\text{total}}$ , batch size bs.

$E_{\text{total}}$ ， 批次大小 bs。

for epoch $= 1, 2, \ldots, E_{\text{total}}$ do

对于轮次 $= 1, 2, \ldots, E_{\text{total}}$ 执行

Obtain $T_t$ and $\widetilde{T}_t$ by Eqs. (2),(3) and (4)

通过方程 (2)、(3) 和 (4) 获得 $T_t$ 和 $\widetilde{T}_t$

Obtain $D_c$ and $D_n$ based on Eq. (5).

基于方程 (5) 获得 $D_c$ 和 $D_n$

Obtain $y^{\text{corr}}$ , $\widetilde{\mu}$ , and $\widetilde{\sigma}^2$ by Eqs. (7),(9) and (10) .

通过方程 (7)、(9) 和 (10) 获得 $y^{\text{corr}}$ , $\widetilde{\mu}$ 和 $\widetilde{\sigma}^2$

Obtain $\lambda(x)$ based on Eq. 8).

基于方程 (8) 获得 $\lambda(x)$

for iteration $= 1, 2, \ldots$ do

对于迭代 $= 1, 2, \ldots$ 执行

Fetch $B = \{(x_i, y_i)\}^{bs}$ from $D_{\text{train}}$

从 $D_{\text{train}}$ 获取 $B=\{(x_i, y_i)\}^{bs}$

Obtain $B_{\text{clean}} \subseteq D_c$ and $B_{\text{noise}} \subseteq D_n$

获得 $B_{\text{clean}} \subseteq D_c$ 和 $B_{\text{noise}} \subseteq D_n$

Calculate $\mathcal{L}=\mathcal{L}_{D_c}+\mathcal{L}_{D_n}+\mathcal{L}_{\text{reg}}$

计算 $\mathcal{L}=\mathcal{L}_{D_c}+\mathcal{L}_{D_n}+\mathcal{L}_{\text{reg}}$

Update $\theta$ by optimizing $\mathcal{L}$

通过优化 $\mathcal{L}$ 更新 $\theta$

Update $\theta^*$ by Eq. 6

通过方程 6 更新 $\theta^*$

end for

结束循环

end for

结束轮次

utput: Updated network $\theta$ .

输出: 更新后的网络 $\theta$ 。

---

in which $\widehat{x}$ denotes the strongly-augmented view of the sample $x.\lambda(x)$ represents the sample weight computed by Eq. (8). Finally, we incorporate an additional weighted classification loss on clean samples w.r.t. corrected labels (similar to $\mathcal{L}_{D_n}$ ) to further enhance the robustness of the model. This loss term implicitly encourages prediction consistency between weakly- and strongly-augmented views of samples from the clean subset, regularizing the model to achieve better performance. Thus, we term this loss as the consistency regularization loss and compute it as follows:

其中 $\widehat{x}$ 表示样本的强增强视图，$x.\lambda(x)$ 代表根据公式 (8) 计算的样本权重。最后，我们在干净样本的校正标签上加入额外的加权分类损失 (类似于 $\mathcal{L}_{D_n}$ )， 以进一步增强模型的鲁棒性。该损失项隐式地鼓励干净子集样本的弱增强视图与强增强视图之间的预测一致性, 从而正则化模型以获得更好的性能。因此，我们将此损失称为一致性正则化损失，并按如下方式计算:

$$\mathcal{L}_{\text{reg}}=-\frac{1}{|D_c|} \sum_{(x,y)\in D_c} \lambda(x)\, y^{\text{corr}} \log p(\widehat{x}, \theta), \tag{15}$$

where $\lambda(x)$ is also computed based on Eq. (8). Accordingly, the final objective loss function in our SED is:

其中 $\lambda(x)$ 也基于公式 (8) 计算。因此，我们的 SED 中的最终目标损失函数为:

$$\mathcal{L} = \mathcal{L}_{D_c} + \mathcal{L}_{D_n} + \mathcal{L}_{\text{reg}}\,. \tag{16}$$

# 4 Experiments

# 4 实验

In this section, we conduct experiments on two synthetically corrupted datasets (i.e., CIFAR100N and CIFAR80N [66]) and three real-world datasets (i.e., Web-Aircraft, Web-Car, and Web-Bird [49]). We demonstrate the superiority of our method in coping with noisy labels by comparing SED with various state-of-the-art (SOTA) methods. Moreover, we conduct extensive ablation studies to evaluate the effectiveness of each component in our SED.

在本节中, 我们在两个合成噪声数据集 (即 CIFAR100N 和 CIFAR80N [66]) 以及三个真实世界数据集 (即 Web-Aircraft、Web-Car 和 Web-Bird [49]) 上进行了实验。我们通过将 SED 与多种最先进 (SOTA) 方法进行比较，展示了我们方法在应对噪声标签方面的优越性。此外，我们还进行了大量消融实验，以评估 SED 中各个组件的有效性。

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). Results of existing methods are mainly drawn from [50]. $^\dagger$ means that we re-implement the method using its open-sourced code and default hyper-parameters.

表 1: 在 CIFAR100N 和 CIFAR80N 上最后十个 epoch 的平均测试准确率 (%)。实验在不同噪声条件下进行 (“Sym”表示对称噪声，“Asym”表示非对称噪声)。现有方法的结果主要来自 [50]。$^\dagger$ 表示我们使用其开源代码和默认超参数重新实现该方法。

| Methods | Publication | CIFAR100N | | | CIFAR80N | | |
|---|---|---|---|---|---|---|---|
| | | 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 |
| Decoupling | NeurIPS 2017 | 33.10 | 3.89 | 26.11 | 43.49 | 10.1 | 33.74 |
| Co-teaching | NeurIPS 2018 | 43.73 | 15.15 | 28.35 | 60.38 | 16.59 | 42.42 |
| Co-teaching+69 | ICML 2019 | 49.27 | 13.44 | 33.62 | 53.97 | 12.29 | 43.01 |
| JoCoR58 | CVPR 2020 | 53.01 | 15.49 | 32.70 | 59.99 | 12.85 | 39.37 |
| Jo-SRC66 | CVPR 2021 | 58.15 | 23.80 | 38.52 | 65.83 | 29.76 | 53.03 |
| SELC36 | IJCAI 2022 | 55.44 | 23.54 | 45.19 | 57.51 | 22.79 | 47.50 |
| DivideMix29 | ICLR 2020 | 57.76 | 28.98 | 43.75 | 57.47 | 21.18 | 37.47 |
| Co-LDL50 | TMM 2022 | 59.73 | 25.12 | 52.28 | 58.81 | 24.22 | 50.69 |
| UNICON†25 | CVPR 2022 | 55.10 | 31.49 | 49.90 | 54.50 | 36.75 | 51.50 |
| NCE† 28 | ECCV 2022 | 54.58 | 35.23 | 49.90 | 58.53 | 39.34 | 56.40 |
| SOP† 35 | ICML 2022 | 58.63 | 34.23 | 49.87 | 60.17 | 34.05 | 53.34 |
| SPRL † 46 | PR 2023 | 57.04 | 28.61 | 49.38 | 47.90 | 22.25 | 40.86 |
| AGCE† 71 | TPAMI 2023 | 59.38 | 27.41 | 43.04 | 60.24 | 25.39 | 44.06 |
| DISC† 31 | CVPR 2023 | 60.28 | 33.90 | 50.56 | 50.33 | 38.23 | 47.63 |
| Ours | - | 66.50 | 38.15 | 58.29 | 69.10 | 42.57 | 60.87 |

| 方法 | 出版 | CIFAR100N | | | CIFAR80N | | |
|---|---|---|---|---|---|---|---|
| | | 对称-20% | 对称-80% | 非对称-40% | 对称-20% | 对称-80% | 非对称-40% |
| 标准 | - | 35.14 | 4.41 | 27.29 | 29.37 | 4.20 | 22.25 |
| 解耦 | NeurIPS 2017 | 33.10 | 3.89 | 26.11 | 43.49 | 10.1 | 33.74 |
| 共同教学 | NeurIPS 2018 | 43.73 | 15.15 | 28.35 | 60.38 | 16.59 | 42.42 |
| 共同教学 +69 | ICML 2019 | 49.27 | 13.44 | 33.62 | 53.97 | 12.29 | 43.01 |
| JoCoR58 | CVPR 2020 | 53.01 | 15.49 | 32.70 | 59.99 | 12.85 | 39.37 |
| Jo-SRC66 | CVPR 2021 | 58.15 | 23.80 | 38.52 | 65.83 | 29.76 | 53.03 |
| SELC36 | IJCAI 2022 | 55.44 | 23.54 | 45.19 | 57.51 | 22.79 | 47.50 |
| DivideMix29 | ICLR 2020 | 57.76 | 28.98 | 43.75 | 57.47 | 21.18 | 37.47 |
| Co-LDL50 | TMM 2022 | 59.73 | 25.12 | 52.28 | 58.81 | 24.22 | 50.69 |
| UNICON†25 | CVPR 2022 | 55.10 | 31.49 | 49.90 | 54.50 | 36.75 | 51.50 |
| NCE† 28 | ECCV 2022 | 54.58 | 35.23 | 49.90 | 58.53 | 39.34 | 56.40 |
| SOP† 35 | ICML 2022 | 58.63 | 34.23 | 49.87 | 60.17 | 34.05 | 53.34 |
| SPRL † 46 | PR 2023 | 57.04 | 28.61 | 49.38 | 47.90 | 22.25 | 40.86 |
| AGCE† 71 | TPAMI 2023 | 59.38 | 27.41 | 43.04 | 60.24 | 25.39 | 44.06 |
| DISC† 31 | CVPR 2023 | 60.28 | 33.90 | 50.56 | 50.33 | 38.23 | 47.63 |
| 我们的方法 | - | 66.50 | 38.15 | 58.29 | 69.10 | 42.57 | 60.87 |

## 4.1 Experiment Setup

## 4.1 实验设置

Synthetically Corrupted Datasets. CIFAR100N and CIFAR80N are mainly derived from CIFAR100 [26]. CIFAR100 consists of 60,000 RGB images (50,000 for training and 10,000 for testing). We follow [66] to create the closed-set noisy dataset CIFAR100N and the open-set noisy dataset CIFAR80N. In particular, to construct the open-set noisy dataset CIFAR80N, we regard the last 20 categories in CIFAR100 as out-of-distribution samples. We adopt two classical noise structures: symmetric and asymmetric, with a noise ratio $n \in (0, 1)$ .

合成噪声数据集。CIFAR100N 和 CIFAR80N 主要源自 CIFAR100 [26]。CIFAR100 包含 60,000 张 RGB 图像 (训练用 50,000 张，测试用 10,000 张)。我们遵循 [66] 创建了封闭集噪声数据集 CIFAR100N 和开放集噪声数据集 CIFAR80N。特别地，为了构建开放集噪声数据集 CIFAR80N，我们将 CIFAR100 的最后 20 个类别视为超出分布样本。我们采用两种经典的噪声结构: 对称噪声和非对称噪声，噪声比例为 $n \in (0,1)$ 。

Real-World Datasets. To further verify the effectiveness of our SED in practical scenarios, we conduct experiments on the three real-world noisy datasets (i.e., Web-Aircraft, Web-Car, and Web-Bird [49]), whose training images are crawled from web image search engines. The noise rates and structures of real-world datasets are all unknown. No label verification information is provided.

真实世界数据集。为了进一步验证我们的方法 (SED) 在实际场景中的有效性，我们在三个真实世界噪声数据集 (即 Web-Aircraft、Web-Car 和 Web-Bird [49]) 上进行了实验，这些训练图像均来自网络图像搜索引擎。真实数据集的噪声率和结构均未知，也未提供标签验证信息。

Implementation Details. On synthetically corrupted datasets, we follow [66] to conduct experiments with a seven-layer CNN network as the backbone. The network is trained using SGD with a momentum of 0.9 for 100 epochs (including 20 warm-up epochs). The batch size is 128, and the initial learning rate is 0.05 . For real-world datasets, we follow [50] and leverage ResNet50 [20] pre-trained on ImageNet as our backbone. We use the SGD optimizer with a momentum of 0.9 to train the network for 110 epochs. The batch size, the initial learning rate, and the weight decay are32,0.005, and 0.0005 . The learning rate decays in a cosine annealing manner. We train the network for 110 epochs, in which the first 10 epochs are warm-up. The EMA coefficients $m$ and $\alpha$ are set to 0.99 and 0.95 . $\lambda_m$ is set to 1.0 for all datasets.

实现细节。在合成噪声数据集上，我们遵循 [66] 使用七层卷积神经网络 (CNN) 作为骨干网络进行实验。网络采用带动量的 SGD(动量为 0.9) 训练 100 个周期 (包括 20 个预热周期)。批量大小为 128 , 初始学习率为 0.05 。对于真实世界数据集，我们遵循 [50]，使用在 ImageNet 上预训练的 ResNet50 [20] 作为骨干网络。采用带动量的 SGD(动量为 0.9) 训练网络 110 个周期。批量大小为 32 , 初始学习率为 0.005 , 权重衰减为 0.0005 。学习率采用余弦退火方式衰减。训练总共 110 个周期，前 10 个为预热。EMA(指数移动平均) 系数 $m$ 和 $\alpha$ 分别设为 0.99 和 0.95 。$\lambda_m$ 在所有数据集上均设为 1.0。

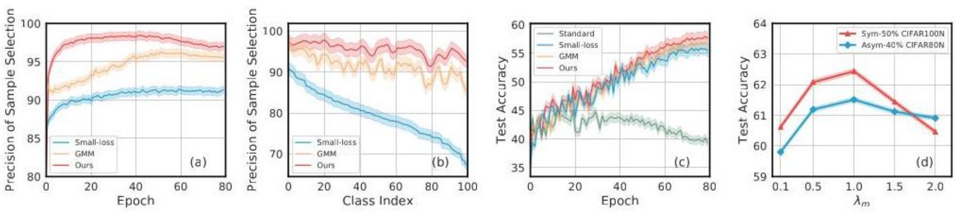


Fig. 3: Comparison of different sample selection methods and the ablation results of the parameter $\lambda_m$ . (a) The overall precision of sample selection (%) vs. epochs. (b) The class-wise precision of sample selection (%) vs. classes. (c) The test accuracy (%) $vs$ . epochs. (d) The test accuracy (%) of using different $\lambda_m$ .

图 3: 不同样本选择方法的比较及参数 $\lambda_m$ 的消融结果。(a) 样本选择的整体精度 (%) 与训练轮数的关系。(b) 按类别划分的样本选择精度 (%) 与类别的关系。(c) 测试准确率 (%) $vs$ 。训练轮数。(d) 使用不同 $\lambda_m$ 时的测试准确率 (%)。

Baselines. For CIFAR100N and CIFAR80N, we compare our method with the following SOTA methods: Decoupling [37], Co-teaching [19], Co-teaching- [69], JoCoR [58], Jo-SRC [66], SELC [36], Co-LDL [50], UNICON [25], SOP [35], AGCE [71, and DISC [31]. For Web-Aircraft, Web-Bird, and Web-Car, besides the above methods, we additionally compare SED with other SOTA methods (e.g., PENCIL [68], Hendrycks et al. [21], mCT-S2R [38], AFM [41], and Self-adaptive [22]). Moreover, we perform conventional training using the entire noisy dataset. The result is provided as a baseline (denoted as Standard). Results in Tables 1 and 2 are mainly obtained from [66] and [50].

基线方法。对于 CIFAR100N 和 CIFAR80N，我们将我们的方法与以下最新的最先进 (SOTA) 方法进行比较:Decoupling [37]、Co-teaching [19]、Co-teaching- [69]、JoCoR [58]、Jo-SRC [66]、SELC [36]、Co-LDL [50]、UNICON [25]、SOP [35]、AGCE [71] 和 DISC [31]。对于 Web-Aircraft、Web-Bird 和 Web-Car，除了上述方法外，我们还将 SED 与其他 SOTA 方法 (如 PENCIL [68]、Hendrycks 等 [21]、mCT-S2R [38]、AFM [41] 和自适应方法 [22]) 进行比较。此外，我们还对整个噪声数据集进行常规训练，结果作为基线 (标记为 Standard)。表 1 和表 2 的结果主要来自 [66] 和 [50]。

## 4.2 Evaluation on Synthetic Datasets

## 4.2 在合成数据集上的评估

We show the comparison results between our SED and existing SOTA methods on the synthetic datasets (i.e., CIFAR100N and CIFAR80N) in Table 1.

我们在表 1 中展示了我们的方法 (SED) 与现有 SOTA 方法在合成数据集 (即 CIFAR100N 和 CIFAR80N) 上的对比结果。

Results on CIFAR100N. Table 1 shows that SED consistently achieves the best performance compared to SOTA methods on CIFAR100N. In particular, it should be noted that SED can better adapt to severely noisy situations (i.e., Sym-80%), while most SOTA approaches almost fail in the most inferior case. It should be emphasized that the asymmetric noise case is often more challenging than the symmetric one. Our SED shows a significant improvement (i.e., $\geq$ 6.01% ) on Asym-40%. Experiments on CIFAR100N show that SED can effectively deal with closed-set label noise in different noise situations.

CIFAR100N 的结果。表 1 显示，SED 在 CIFAR100N 上始终优于其他 SOTA 方法，表现出最佳性能。特别值得注意的是, SED 能更好地适应严重噪声 (即 Sym-80%) 的情况, 而大多数 SOTA 方法在最差情况下几乎失效。需要强调的是, 非对称噪声 (Asymmetric Noise) 通常比对称噪声 (Symmetric Noise) 更具挑战性。我们的 SED 在非对称 40% 噪声情况下表现出显著提升 (即 $\geq$ 6.01% )。在 CIFAR100N 上的实验表明，SED 能有效应对不同噪声情况下的封闭集标签噪声 (closed-set label noise)。

Results on CIFAR80N. To simulate real-world scenarios, CIFAR80N contains both closed-set and open-set noisy labels, making it undoubtedly more challenging. Results shown in Table 1 illustrate: (1) in the case of Sym-20%, our SED can achieve a 3.27% performance improvement. (2) in the case of Sym-80%, while most SOTA approaches fail to tackle the massive noisy labels, SED achieves the best result. (3) when the noise scenario becomes harder (i.e., Asym-40%), our SED consistently obtains the best performance, outperforming the second-best result by 7.53% . Table 1 proves that SED performs consistently better than existing methods when coping with open-set noisy datasets.

在 CIFAR80N 上的结果。为了模拟真实世界场景，CIFAR80N 包含封闭集和开放集的噪声标签，使其无疑更具挑战性。表 1 中的结果显示:(1) 在 Sym-20% 的情况下，我们的 SED(噪声检测模型) 可以实现 3.27% 的性能提升。(2) 在 Sym-80% 的情况下，尽管大多数 SOTA(最先进技术) 方法未能应对大量噪声标签，SED 仍取得了最佳结果。(3) 当噪声场景变得更困难 (即 Asym-40%) 时，我们的 SED 始终获得最佳性能，优于第二名结果 7.53% 。表 1 证明了在处理开放集噪声数据集时，SED 的表现始终优于现有方法。

Table 2: The 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 50. † means that we re-implement the method using its open-sourced code and default hyper-parameters.

表 2: 在真实世界噪声数据集 Web-Aircraft、Web-Bird、Web-Car 上的测试准确率 (%) 与 SOTA(最先进技术) 方法的比较。现有方法的结果主要来自 50。† 表示我们使用其开源代码和默认超参数重新实现了该方法。

| Methods | Publication | Backbone | Performances(%) | | | |
|---|---|---|---|---|---|---|
| | | | Web-Aircraft | Web-Bird | Web-Car | Average |
| Standard | - | ResNet50 | 60.80 | 64.40 | 60.60 | 61.93 |
| Decoupling 37 | NeurIPS 2017 | ResNet50 | 75.91 | 71.61 | 79.41 | 75.64 |
| Co-teaching [19 | NeurIPS 2018 | ResNet50 | 79.54 | 76.68 | 84.95 | 80.39 |
| Co-teaching+ 69 | ICML 2019 | ResNet50 | 74.80 | 70.12 | 76.77 | 73.90 |
| PENCIL 68 | CVPR 2019 | ResNet50 | 78.82 | 75.09 | 81.68 | 78.53 |
| Hendrycks et al. 21 | NeurIPS 2019 | ResNet50 | 73.24 | 70.03 | 73.81 | 72.36 |
| mCT-S2R 38 | WACV 2020 | ResNet50 | 79.33 | 77.67 | 82.92 | 79.97 |
| JoCoR 58 | CVPR 2020 | ResNet50 | 80.11 | 79.19 | 85.10 | 81.47 |
| AFM 41 | ECCV 2020 | ResNet50 | 81.04 | 76.35 | 83.48 | 80.29 |
| DivideMix 29 | ICLR 2020 | ResNet50 | 82.48 | 74.40 | 84.27 | 80.38 |
| Self-adaptive 22 | NeurIPS 2020 | ResNet50 | 77.92 | 78.49 | 78.19 | 78.20 |
| Co-LDL 50 | TMM 2022 | ResNet50 | 81.97 | 80.11 | 86.95 | 83.01 |
| UNICON†25 | CVPR 2022 | ResNet50 | 85.18 | 81.20 | 88.15 | 84.84 |
| NCE†28 | ECCV 2022 | ResNet50 | 84.94 | 80.22 | 86.38 | 83.85 |
| SOP†35 | ICML 2022 | ResNet50 | 84.06 | 79.40 | 85.71 | 83.06 |
| SPRL†46 | PR 2023 | ResNet50 | 84.40 | 76.36 | 86.84 | 82.53 |
| AGCE†71 | TPAMI 2023 | ResNet50 | 84.22 | 75.60 | 85.16 | 81.66 |
| DISC†31 | CVPR 2023 | ResNet50 | 85.27 | 81.08 | 88.31 | 84.89 |
| Ours | - | ResNet50 | 86.62 | 82.00 | 88.88 | 85.83 |

| 方法 | 出版 | 骨干网络 | 性能 (%) | | | |
|---|---|---|---|---|---|---|
| | | | Web- 飞机 | Web-鸟 | Web-汽车 | 平均值 |
| 标准 | - | ResNet50(残差网络 50 层) | 60.80 | 64.40 | 60.60 | 61.93 |
| 解耦 37 | NeurIPS 2017 | ResNet50(残差网络 50 层) | 75.91 | 71.61 | 79.41 | 75.64 |
| 协同教学 [19 | NeurIPS 2018 | ResNet50(残差网络 50 层) | 79.54 | 76.68 | 84.95 | 80.39 |
| 协同教学 +69 | ICML 2019 | ResNet50(残差网络 50 层) | 74.80 | 70.12 | 76.77 | 73.90 |
| PENCIL 68 | CVPR 2019 | ResNet50(残差网络 50 层) | 78.82 | 75.09 | 81.68 | 78.53 |
| Hendrycks 等人 21 | NeurIPS 2019 | ResNet50(残差网络 50 层) | 73.24 | 70.03 | 73.81 | 72.36 |
| mCT-S2R 38 | WACV 2020 | ResNet50(残差网络 50 层) | 79.33 | 77.67 | 82.92 | 79.97 |
| JoCoR 58 | CVPR 2020 | ResNet50(残差网络 50 层) | 80.11 | 79.19 | 85.10 | 81.47 |
| AFM 41 | ECCV 2020 | ResNet50(残差网络 50 层) | 81.04 | 76.35 | 83.48 | 80.29 |
| DivideMix 29 | ICLR 2020 | ResNet50(残差网络 50 层) | 82.48 | 74.40 | 84.27 | 80.38 |
| 自适应学习 22 | NeurIPS 2020 | ResNet50(残差网络 50 层) | 77.92 | 78.49 | 78.19 | 78.20 |
| Co-LDL 50 | TMM 2022 | ResNet50(残差网络 50 层) | 81.97 | 80.11 | 86.95 | 83.01 |
| UNICON$^{\dagger}$25 | CVPR 2022 | ResNet50(残差网络 50 层) | 85.18 | 81.20 | 88.15 | 84.84 |
| NCE$^{\dagger}$28 | ECCV 2022 | ResNet50(残差网络 50 层) | 84.94 | 80.22 | 86.38 | 83.85 |
| SOP$^{\dagger}$35 | ICML 2022 | ResNet50(残差网络 50 层) | 84.06 | 79.40 | 85.71 | 83.06 |
| SPRL$^{\dagger}$46 | PR 2023 | ResNet50(残差网络 50 层) | 84.40 | 76.36 | 86.84 | 82.53 |
| AGCE$^{\dagger}$71 | TPAMI 2023 | ResNet50(残差网络 50 层) | 84.22 | 75.60 | 85.16 | 81.66 |
| DISC$^{\dagger}$31 | CVPR 2023 | ResNet50(残差网络 50 层) | 85.27 | 81.08 | 88.31 | 84.89 |
| 我们的方法 | - | ResNet50(残差网络 50 层) | 86.62 | 82.00 | 88.88 | 85.83 |

## 4.3 Evaluation on Real-world Datasets

## 4.3 在真实世界数据集上的评估

Table 2 shows the experimental results of existing methods and SED on Web-Aircraft, Web-Bird, and Web-Car, which contain open-set and closed-set noise simultaneously. From this table, we can find that SED can achieve better (or comparable) performance against SOTA approaches in different datasets. SED achieves performances of 86.62%, 82.00% , and 88.88% on test sets of Web-Aircraft, Web-Bird, and Web-Car, respectively. The average test accuracy outperforms existing SOTA methods by 0.94%. It should be noted that the second and third-best methods (i.e., DISC and UNICON) involve the Mixup training trick and two simultaneously trained networks respectively, while SED trains only one network without Mixup. Compared to existing methods, our SED eliminates the demand for dataset-dependent prior knowledge (e.g., pre-defined drop rate/threshold), making it easier to adapt to different datasets.

表 2 显示了在 Web-Aircraft(网络飞机)、Web-Bird(网络鸟类) 和 Web-Car(网络汽车) 上现有方法与 SED 的实验结果，这些数据集同时包含开放集和封闭集噪声。从表中可以看出，SED 在不同数据集上都能实现优于 (或与之相当的) 最先进 (SOTA) 方法的性能。SED 在 Web-Aircraft、Web-Bird 和 Web-Car 的测试集上分别达到了 86.62%, 82.00% 、88.88% 的性能。平均测试准确率比现有的 SOTA 方法高出 0.94%。值得注意的是，第二和第三名的方法 (即 DISC 和 UNICON) 分别采用了 Mixup 训练技巧和同时训练的两个网络，而 SED 仅训练一个网络且未使用 Mixup。与现有方法相比，我们的 SED 无需依赖于数据集特定的先验知识 (例如预定义的丢弃率/阈值)，因此更易于适应不同的数据集。

## 4.4 Ablation Studies

## 4.4 消融实验

In this section, we demonstrate the effectiveness of each component in our SED (i.e., SCS, SCR, and CR). Besides, we investigate the effect of the hyper-parameter $\lambda_m$ in Eq. (8). Unless otherwise stated, ablation experiments are conducted on CIFAR100N (Sym-50%). Table 3 and Table 4 show the impact of each component.

在本节中，我们展示了 SED 中各个组件 (即 SCS、SCR 和 CR) 的有效性。此外，我们还研究了公式 (8) 中超参数 $\lambda_m$ 的影响。除非另有说明，消融实验均在 CIFAR100N(Sym-50%) 上进行。表 3 和表 4 展示了各个组件的影响。

Table 3: Effect of each component in the test accuracy (%) on CIFAR100N.

表 3: 在 CIFAR100N 上各组件对测试准确率 (%) 的影响。

| Model | Test Accuracy |
|---|---|
| Standard | 34.10 |
| Standard+SCS w/o local threshold | 53.36 |
| Standard+SCS w/o global threshold | 55.64 |
| Standard+SCS w/o EMA | 54.72 |
| Standard+SCS | 58.21 |
| Standard+SCS+SCR w/o re-weighting | 59.75 |
| Standard+SCS+SCR w/o EMA | 60.08 |
| Standard+SCS+SCR | 60.43 |
| Standard+SCS+SCR+CR | 62.65 |

| 模型 | 测试准确率 |
|---|---|
| 标准 | 34.10 |
| 标准 +SCS(自适应协同筛选) 不使用局部阈值 | 53.36 |
| 标准 +SCS(自适应协同筛选) 不使用全局阈值 | 55.64 |
| 标准 +SCS(自适应协同筛选) 不使用指数移动平均 (EMA) | 54.72 |
| 标准 +SCS(自适应协同筛选) | 58.21 |
| 标准 +SCS+SCR(自适应协同筛选 + 自适应筛选) 不进行重加权 | 59.75 |
| 标准 +SCS+SCR(自适应协同筛选 + 自适应筛选) 不使用指数移动平均 (EMA) | 60.08 |
| 标准 +SCS+SCR(自适应协同筛选 + 自适应筛选) | 60.43 |
| 标准 +SCS+SCR+CR(自适应协同筛选 + 自适应筛选 + 类别重采样) | 62.65 |

Effects of Self-adaptive and Class-balanced Sample Selection. As analyzed above, existing sample selection methods tend to struggle with the demand for dataset-dependent prior knowledge, such as pre-defined drop rate/threshold. However, these hyper-parameters are usually unknown and hard to estimate in real-world datasets. The proposed SCS strategy in our method allows adaptive sample selection in a class-balanced manner, making our SED have better generalization performance on different datasets. As shown in Table 3, employing SCS achieves a 24.11% performance gain compared to the baseline Standard. We also provide the result of using SCS without local thresholds and global thresholds. This proves that our threshold design is crucial for improving the robustness of the model.

自适应样本选择与类别平衡样本选择的效果。如上所述，现有的样本选择方法在依赖于数据集先验知识 (如预定义的丢弃率/阈值) 方面存在困难。然而，这些超参数在实际数据集中通常未知且难以估算。我们方法中提出的 SCS 策略允许以类别平衡的方式进行自适应样本选择，使我们的 SED 在不同数据集上具有更好的泛化性能。如表 3 所示，采用 SCS 相比基线标准方法 (Standard) 性能提升了 24.11%。我们还提供了在没有局部阈值和全局阈值条件下使用 SCS 的结果。这证明了我们的阈值设计对于提升模型鲁棒性至关重要。

To further demonstrate the superiority of our SCS over previous sample selection strategies, we compare our SCS with two commonly-used methods (i.e., small-loss [19], and GMM [29]) in Fig. 3. As shown in Fig. 3 (a), our SCS is shown to be more effective in selecting clean samples accurately compared with the other two strategies. Additionally, we compare the sample selection accuracy for each category in the selected clean subset and present the comparison in Fig. 3 (b). It illustrates that the selection results of SCS are more balanced. The curves of test accuracy are shown in Fig. 3 (c), revealing the leading performance of our SCS compared with the other two methods and the baseline.

为了进一步展示我们的 SCS 优于以往的样本选择策略，我们在图 3 中将 SCS 与两种常用方法 (即小损失 [19] 和 GMM[29]) 进行了比较。如图 3(a) 所示，我们的 SCS 在准确选择干净样本方面比其他两种策略更有效。此外，我们还比较了在所选干净子集中的每个类别的样本选择准确率，并在图 3(b) 中展示了对比结果。结果显示，SCS 的选择结果更为平衡。测试准确率的曲线如图 3(c) 所示，显示出我们的 SCS 在性能上优于其他两种方法和基线方法。

Effects of Self-adaptive and Class-balanced Sample Re-weighting. Our SED follows an SSL-like paradigm. Selected clean samples are learned conventionally, while detected noisy samples are also fed into the network for training after label correction. However, the biased model capability tends to result in imbalanced label correction, hurting the model performance. We accordingly propose SCR to re-weight detected noisy samples in a self-adaptive and class-balanced manner when using their corrected labels for training. Table 3 shows a performance gain of 2.19% by employing our proposed SCR. The only involved hyper-parameter in the SCR is the $\lambda_m$ in Eq. (8). Fig. 3 (d) exhibits the influence of different $\lambda_m$ values on the test accuracy when experimenting with CIFAR100N (Sym-50%) and CIFAR80N (Asym-40%). It can be observed that the best performance is achieved when $\lambda_m = 1.0$ on CIFAR100N (Sym-50%) and CIFAR80N (Asym-40%).

自适应样本重加权与类别平衡样本重加权的效果。我们的 SED 遵循类似 SSL(半监督学习) 范式。选中的干净样本进行常规学习，而检测到的噪声样本在标签校正后也会被输入网络进行训练。然而，偏置模型能力倾向于导致标签校正不平衡，影响模型性能。因此，我们提出了 SCR，在使用校正标签进行训练时，以自适应且类别平衡的方式对检测到的噪声样本进行重加权。表 3 显示，采用我们提出的 SCR 后，性能提升了 2.19%。SCR 中唯一涉及的超参数是公式 (8) 中的 $\lambda_m$ 。图 3(d) 展示了不同 $\lambda_m$ 值对在 CIFAR100N(对称 50%) 和 CIFAR80N(非对称 40%) 上测试准确率的影响。可以观察到，在 CIFAR100N(对称 50%) 和 CIFAR80N(非对称 40%) 上，最佳性能出现在 $\lambda_m = 1.0$ 值时。

Table 4: Effect of promoting class balance on CIFAR100N (left) and CIFAR80N (right). Test accuracy (%) of SED with and without the class-balanced design is compared under different settings.

表 4: 在 CIFAR100N(左) 和 CIFAR80N(右) 上促进类别平衡的效果。比较在不同设置下，采用和不采用类别平衡设计的 SED 的测试准确率 (%)。

| Class-balanced? | ✘ | ✓ |
|---|---|---|
| Sym-20% | 64.16 | 66.59 |
| Sym-80% | 38.08 | 39.32 |
| Asym-40% | 52.78 | 58.80 |

| 类别平衡吗? | ✘ | ✓ |
|---|---|---|
| 对称-20% | 64.16 | 66.59 |
| 对称-80% | 38.08 | 39.32 |
| 非对称-40% | 52.78 | 58.80 |

| Class-balanced? | ✘ | ✓ |
|---|---|---|
| Sym-20% | 67.20 | 68.75 |
| Sym-80% | 39.74 | 42.90 |
| Asym-40% | 57.00 | 61.51 |

| 类别平衡吗? | ✘ | ✓ |
|---|---|---|
| 对称-20% | 67.20 | 68.75 |
| 对称-80% | 39.74 | 42.90 |
| 非对称-40% | 57.00 | 61.51 |

Effects of Consistency Regularization. Although clean samples selected by SED are more accurate and balanced than previous methods, it is inevitable that some noisy data will be mistakenly selected into the clean subset. Therefore, we impose an additional CR on the selected clean samples to enhance the model's robustness. Table 3 shows that CR successfully boosts model performance by 2.02% , revealing the benefits that CR brings to our model.

一致性正则化的效果。虽然由 SED(自适应样本选择) 筛选出的干净样本比以往方法更准确且更平衡，但不可避免地会有一些噪声数据被误选为干净子集。因此，我们在所选的干净样本上施加了额外的 CR(一致性正则化)，以增强模型的鲁棒性。表 3 显示，CR 成功提升了模型性能 2.02% ，揭示了 CR 为我们的模型带来的益处。

Effects of Promoting Class Balance. As stated in SCS and SCR, our SED favors the class-balanced design. Specifically, SCS estimates local thresholds on each class to avoid imbalanced sample selection, while SCR also estimates $\mu_t$ and $\sigma_t^2$ of the dynamic truncated normal distribution for each class to encourage balanced re-weighting. As shown in Table 4, we investigate the effect of the class-balanced design in SED. We can find that our method consistently achieves better performance when incorporated with the class-balanced design, especially in harder scenarios. Table 4 effectively demonstrates that the class-balanced design in our SED is beneficial for model performance.

促进类别平衡的效果。如在 SCS(局部阈值估计) 和 SCR(动态截断正态分布) 中所述，我们的 SED 偏向于类别平衡设计。具体而言，SCS 在每个类别上估算局部阈值以避免样本不平衡的选择，而 SCR 还估算每个类别的 $\mu_t$ 和 $\sigma_t^2$ ，以鼓励平衡的重加权。如表 4 所示，我们研究了类别平衡设计在 SED 中的效果。可以发现，结合类别平衡设计后，我们的方法在更困难的场景中表现更优。表 4 有效地证明了我们 SED 中的类别平衡设计对模型性能的有益作用。

## 5 Conclusion

## 5 结论

In this paper, we proposed a simple yet effective approach named SED to address the inferior model performance caused by noisy labels. We designed a self-adaptive and class-balanced sample selection strategy to distinguish between clean and noisy samples. Clean samples were learned conventionally. A mean-teacher model was employed to correct the labels of detected noisy samples. Subsequently, SED re-weighted noisy samples in a self-adaptive and class-balanced fashion based on the correction confidence when leveraging them for model training. Finally, we additionally imposed consistency regularization on the clean subset to further improve model performance. Comprehensive experiments and ablation analysis on synthetic and real-world noisy datasets validated the superiority of our SED.

本文提出了一种简单而有效的方法，命名为 SED(自适应样本选择)，旨在解决由噪声标签引起的模型性能下降问题。我们设计了一种自适应且类别平衡的样本选择策略，用于区分干净样本和噪声样本。干净样本采用传统学习方式。采用均值教师 (mean-teacher) 模型对检测到的噪声样本的标签进行校正。随后，SED 根据校正置信度，以自适应且类别平衡的方式对噪声样本进行重加权，在模型训练中利用它们。最后，我们在干净子集上额外施加了一致性正则化，以进一步提升模型性能。在合成和真实噪声数据集上的全面实验和消融分析验证了我们 SED 的优越性。

## References

## 参考文献

1. Arpit, D., Jastrzebski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M.S., Maharaj, T., Fischer, A., Courville, A.C., Bengio, Y., Lacoste-Julien, S.: A closer look at memorization in deep networks. In: Int. Conf. Mach. Learn. pp. 233-242 (2017)

1. Arpit, D., Jastrzebski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M.S., Maharaj, T., Fischer, A., Courville, A.C., Bengio, Y., Lacoste-Julien, S.: 深度网络中的记忆机制分析。发表于: 国际机器学习会议 (2017) 第 233-242 页

2. Bai, Y., Yang, E., Han, B., Yang, Y., Li, J., Mao, Y., Niu, G., Liu, T.: Understanding and improving early stopping for learning with noisy labels. In: Adv. Neural Inform. Process. Syst. pp. 24392-24403 (2021)

2. Bai, Y., Yang, E., Han, B., Yang, Y., Li, J., Mao, Y., Niu, G., Liu, T.: 理解与改进带噪声标签学习中的早停策略。发表于: 神经信息处理系统进展 (2021) 第 24392-24403 页

3. Berthelot, D., Carlini, N., Goodfellow, I.J., Papernot, N., Oliver, A., Raffel, C.: Mixmatch: A holistic approach to semi-supervised learning. In: Adv. Neural Inform. Process. Syst. pp. 5050-5060 (2019)

3. Berthelot, D., Carlini, N., Goodfellow, I.J., Papernot, N., Oliver, A., Raffel, C.: Mixmatch: 一种半监督学习的整体方法。发表于: 神经信息处理系统进展 (2019) 第 5050-5060 页

4. Berthon, A., Han, B., Niu, G., Liu, T., Sugiyama, M.: Confidence scores make instance-dependent label-noise learning possible. In: Int. Conf. Mach. Learn. pp. 825-836 (2021)

4. Berthon, A., Han, B., Niu, G., Liu, T., Sugiyama, M.: 置信度评分使得实例依赖的标签噪声学习成为可能。发表于: 国际机器学习会议 (2021) 第 825-836 页

5. Boutros, F., Damer, N., Kirchbuchner, F., Kuijper, A.: Elasticface: Elastic margin loss for deep face recognition. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 1577-1586 (2022)

5. Boutros, F., Damer, N., Kirchbuchner, F., Kuijper, A.: Elasticface: 用于深度人脸识别的弹性边缘损失。发表于:IEEE 计算机视觉与模式识别会议 (2022) 第 1577-1586 页

6. Bucarelli, M.S., Cassano, L., Siciliano, F., Mantrach, A., Silvestri, F.: Leveraging inter-rater agreement for classification in the presence of noisy labels. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 3439-3448 (2023)

6. Bucarelli, M.S., Cassano, L., Siciliano, F., Mantrach, A., Silvestri, F.: 利用评审者间一致性进行带噪声标签的分类。发表于:IEEE 计算机视觉与模式识别会议 (2023) 第 3439-3448 页

7. Chen, H., Tao, R., Fan, Y., Wang, Y., Wang, J., Schiele, B., Xie, X., Raj, B., Sav-vides, M.: Softmatch: Addressing the quantity-quality trade-off in semi-supervised learning. In: Int. Conf. Learn. Represent. (2023)

7. Chen, H., Tao, R., Fan, Y., Wang, Y., Wang, J., Schiele, B., Xie, X., Raj, B., Savvides, M.: Softmatch: 解决半监督学习中的数量与质量权衡问题。发表于: 学习表征国际会议 (2023)

8. Chen, T., Yao, Y., Tang, J.: Multi-granularity denoising and bidirectional alignment for weakly supervised semantic segmentation. IEEE Trans. Multimedia 32, 2960-2971 (2023)

8. Chen, T., Yao, Y., Tang, J.: 多粒度去噪与双向对齐用于弱监督语义分割。IEEE 多媒体杂志 32, 2960-2971 (2023)

9. Chen, T., Yao, Y., Zhang, L., Wang, Q., Xie, G., Shen, F.: Saliency guided inter-and intra-class relation constraints for weakly supervised semantic segmentation. IEEE Trans. Multimedia 25, 1727-1737 (2023)

9. Chen, T., Yao, Y., Zhang, L., Wang, Q., Xie, G., Shen, F.: 基于显著性引导的类间与类内关系约束用于弱监督语义分割。IEEE 多媒体杂志 25, 1727-1737 (2023)

10. Chen, Y., Tan, X., Zhao, B., Chen, Z., Song, R., Liang, J., Lu, X.: Boosting semi-supervised learning by exploiting all unlabeled data. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 7548-7557. IEEE (2023)

10. Chen, Y., Tan, X., Zhao, B., Chen, Z., Song, R., Liang, J., Lu, X.: 通过利用所有未标记数据提升半监督学习。发表于:IEEE 计算机视觉与模式识别会议 (IEEE Conf. Comput. Vis. Pattern Recog.) 第 7548-7557 页 (2023)

11. Cheng, D., Liu, T., Ning, Y., Wang, N., Han, B., Niu, G., Gao, X., Sugiyama, M.: Instance-dependent label-noise learning with manifold-regularized transition matrix estimation. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 16609-16618 (2022)

11. Cheng, D., Liu, T., Ning, Y., Wang, N., Han, B., Niu, G., Gao, X., Sugiyama, M.: 利用流形正则化转移矩阵估计进行实例依赖标签噪声学习。发表于:IEEE 计算机视觉与模式识别会议 (IEEE Conf. Comput. Vis. Pattern Recog.) 第 16609-16618 页 (2022)

12. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 248- 255 (2009)

12. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, L.: ImageNet: 一个大规模层次化图像数据库。发表于:IEEE 计算机视觉与模式识别会议 (IEEE Conf. Comput. Vis. Pattern Recog.) 第 248-255 页 (2009)

13. Fang, T., Lu, N., Niu, G., Sugiyama, M.: Rethinking importance weighting for deep learning under distribution shift. In: Adv. Neural Inform. Process. Syst. (2020)

13. Fang, T., Lu, N., Niu, G., Sugiyama, M.: 重新思考在分布偏移 (distribution shift) 下深度学习的重要性加权。发表于:Advances in Neural Information Processing Systems(2020)

14. Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A.: Learning object categories from internet image searches. Proc. IEEE pp. 1453-1466 (2010)

14. Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A.: 从互联网图像搜索中学习对象类别。会议论文:IEEE(2010)

15. Goldberger, J., Ben-Reuven, E.: Training deep neural-networks using a noise adaptation layer. In: Int. Conf. Learn. Represent. (2017)

15. Goldberger, J., Ben-Reuven, E.: 使用噪声适应层训练深度神经网络。发表于: 国际学习表征会议 (Int. Conf. Learn. Represent.)(2017)

16. Goldberger, J., Ben-Reuven, E.: Training deep neural-networks using a noise adaptation layer. In: Int. Conf. Learn. Represent. (2017)

16. Goldberger, J., Ben-Reuven, E.: 使用噪声适应层训练深度神经网络。发表于: 国际学习表征会议 (Int. Conf. Learn. Represent.)(2017)

17. Gong, C., Ding, Y., Han, B., Niu, G., Yang, J., You, J., Tao, D., Sugiyama, M.: Class-wise denoising for robust learning under label noise. IEEE Trans. Pattern Anal. Mach. Intell. pp. 2835-2848 (2023)

17. Gong, C., Ding, Y., Han, B., Niu, G., Yang, J., You, J., Tao, D., Sugiyama, M.: 针对标签噪声的类别级去噪以实现稳健学习。IEEE 模式分析与机器智能学报 (IEEE Trans. Pattern Anal. Mach. Intell.) 第 2835-2848 页 (2023)

18. Gui, X., Wang, W., Tian, Z.: Towards understanding deep learning from noisy labels with small-loss criterion. In: IJCAI. pp. 2469-2475 (2021)

18. Gui, X., Wang, W., Tian, Z.: 通过小损失准则理解带噪声标签的深度学习。发表于: 国际人工智能联合会议 (IJCAI)(2021)

19. Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I.W., Sugiyama, M.: Co-teaching: Robust training of deep neural networks with extremely noisy labels. In: Adv. Neural Inform. Process. Syst. pp. 8536-8546 (2018)

19. Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I.W., Sugiyama, M.: 协同训练 (Co-teaching): 应对极端噪声标签的深度神经网络稳健训练。发表于:Advances in Neural Information Processing Systems(2018)

20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 770-778 (2016)

20. He, K., Zhang, X., Ren, S., Sun, J.: 深度残差学习 (Deep Residual Learning) 用于图像识别。发表于:IEEE 计算机视觉与模式识别会议 (CVPR)(2016)

21. Hendrycks, D., Mazeika, M., Kadavath, S., Song, D.: Using self-supervised learning can improve model robustness and uncertainty. In: Adv. Neural Inform. Process. Syst. pp. 15637-15648 (2019)

21. Hendrycks, D., Mazeika, M., Kadavath, S., Song, D.: 使用自监督学习 (self-supervised learning) 可以提升模型的鲁棒性和不确定性。发表于:Advances in Neural Information Processing Systems(2019)

22. Huang, L., Zhang, C., Zhang, H.: Self-adaptive training: beyond empirical risk minimization. In: Adv. Neural Inform. Process. Syst. (2020)

22. Huang, L., Zhang, C., Zhang, H.: 自适应训练 (Self-adaptive training): 超越经验风险最小化。发表于:Advances in Neural Information Processing Systems(2020)

23. Jiang, L., Zhou, Z., Leung, T., Li, L., Fei-Fei, L.: Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In: Int. Conf. Mach. Learn. pp. 2309-2318 (2018)

23. Jiang, L., Zhou, Z., Leung, T., Li, L., Fei-Fei, L.: Mentornet: 为极深神经网络在受污染标签 (corrupted labels) 上学习数据驱动的课程。发表于: 国际机器学习会议 (ICML)(2018)

24. Jiang, X., Liu, S., Dai, X., Hu, G., Huang, X., Yao, Y., Xie, G.S., Shao, L.: Deep metric learning based on meta-mining strategy with semiglobal information. IEEE Trans. Neural. Netw. Learn. Syst. **35** (4) , 5103 – 5116 (2024)

24. Jiang, X., Liu, S., Dai, X., Hu, G., Huang, X., Yao, Y., Xie, G.S., Shao, L.: 基于元挖掘策略 (meta-mining strategy) 和半全局信息的深度度量学习。IEEE 神经网络与学习系统学报 (IEEE Trans. Neural. Netw. Learn. Syst.) **35** (4) , 5103 – 5116 (2024)

25. Karim, N., Rizve, M.N., Rahnavard, N., Mian, A., Shah, M.: UNICON: combating label noise through uniform selection and contrastive learning. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 9666-9676 (2022)

25. Karim, N., Rizve, M.N., Rahnavard, N., Mian, A., Shah, M.: UNICON: 通过均匀选择 (uniform selection) 和对比学习 (contrastive learning) 应对标签噪声。发表于:IEEE 计算机视觉与模式识别会议 (IEEE Conf. Comput. Vis. Pattern Recog.) 第 9666-9676 页 (2022)

26. Krizhevsky, A.: Learning multiple layers of features from tiny images (2009)

26. Krizhevsky, A.: 从微小图像中学习多层特征 (2009)

27. Lee, K.H., He, X., Zhang, L., Yang, L.: Cleannet: Transfer learning for scalable image classifier training with label noise. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 5447-5456 (2018)

27. Lee, K.H., He, X., Zhang, L., Yang, L.: Cleannet: 用于带标签噪声的可扩展图像分类器训练的迁移学习。发表于:IEEE 计算机视觉与模式识别会议，第 5447-5456 页 (2018)

28. Li, J., Li, G., Liu, F., Yu, Y.: Neighborhood collective estimation for noisy label identification and correction. In: Eur. Conf. Comput. Vis. pp. 128-145 (2022)

28. Li, J., Li, G., Liu, F., Yu, Y.: 邻域集体估计用于噪声标签识别与校正。发表于: 欧洲计算机视觉会议，第 128-145 页 (2022)

29. Li, J., Socher, R., Hoi, S.C.: Dividemix: Learning with noisy labels as semi-supervised learning. In: Int. Conf. Learn. Represent. (2020)

29. Li, J., Socher, R., Hoi, S.C.: Dividemix: 将带噪声标签的学习视为半监督学习。发表于: 国际学习表示会议 (2020)

30. Li, S., Xia, X., Ge, S., Liu, T.: Selective-supervised contrastive learning with noisy labels. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 316-325 (2022)

30. Li, S., Xia, X., Ge, S., Liu, T.: 选择性监督对比学习与噪声标签。发表于:IEEE 计算机视觉与模式识别会议，第 316-325 页 (2022)

31. Li, Y., Han, H., Shan, S., Chen, X.: DISC: learning from noisy labels via dynamic instance-specific selection and correction. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 24070-24079 (2023)

31. Li, Y., Han, H., Shan, S., Chen, X.: DISC: 通过动态实例特定选择与校正实现噪声标签学习。发表于:IEEE 计算机视觉与模式识别会议，第 24070-24079 页 (2023)

32. Liu, H., Sheng, M., Sun, Z., Yao, Y., Hua, X.S., Shen, H.T.: Learning with imbal-anced noisy data by preventing bias in sample selection. IEEE Trans. Multimedia 26, 7426-7437 (2024)

32. Liu, H., Sheng, M., Sun, Z., Yao, Y., Hua, X.S., Shen, H.T.: 通过防止样本选择偏差应对不平衡噪声数据。IEEE 多媒体杂志 26，7426-7437(2024)

33. Liu, H., Zhang, C., Yao, Y., Wei, X., Shen, F., Tang, Z., Zhang, J.: Exploiting web images for fine-grained visual recognition by eliminating open-set noise and utilizing hard examples. IEEE Trans. Multimedia 24, 546-557

(2022)

33. Liu, H., Zhang, C., Yao, Y., Wei, X., Shen, F., Tang, Z., Zhang, J.: 利用网络图像进行细粒度视觉识别，消除开放集噪声并利用困难样本。IEEE 多媒体杂志 24，546-557(2022)

34. Liu, H., Zhang, H., Lu, J., Tang, Z.: Exploiting web images for fine-grained visual recognition via dynamic loss correction and global sample selection. IEEE Trans. Multimedia 24, 1105-1115 (2022)

34. Liu, H., Zhang, H., Lu, J., Tang, Z.: 通过动态损失校正和全局样本选择利用网络图像进行细粒度视觉识别。IEEE 多媒体杂志 24，1105-1115(2022)

35. Liu, S., Zhu, Z., Qu, Q., You, C.: Robust training under label noise by over-parameterization. In: Int. Conf. Mach. Learn. pp. 14153-14172 (2022)

35. Liu, S., Zhu, Z., Qu, Q., You, C.: 通过过参数化实现对标签噪声的鲁棒训练。发表于: 国际机器学习会议 (2022)

36. Lu, Y., He, W.: SELC: self-ensemble label correction improves learning with noisy labels. In: IJCAI. pp. 3278-3284 (2022)

36. Lu, Y., He, W.: SELC: 自集成标签校正提升带噪声标签的学习效果。发表于: 国际人工智能联合会议 (2022)

37. Malach, E., Shalev-Shwartz, S.: Decoupling ”when to update” from ”how to update”. In: Adv. Neural Inform. Process. Syst. pp. 960-970 (2017)

37. Malach, E., Shalev-Shwartz, S.: 将“何时更新”与“如何更新”解耦。发表于: 先进神经信息处理系统 (2017)

38. Mandal, D., Bharadwaj, S., Biswas, S.: A novel self-supervised re-labeling approach for training with noisy labels. In: IEEE Winter Conference on Applications of Computer Vision. pp. 1370-1379 (2020)

38. Mandal, D., Bharadwaj, S., Biswas, S.: 一种新颖的自监督重标注方法用于带噪声标签的训练。发表于:IEEE 计算机视觉应用冬季会议 (2020)

39. Mao, J., Yao, Y., Sun, Z., Huang, X., Shen, F., Shen, H.T.: Attention map guided transformer pruning for occluded person re-identification on edge device. IEEE Trans. Multimedia 25, 1592-1599 (2023)

39. Mao, J., Yao, Y., Sun, Z., Huang, X., Shen, F., Shen, H.T.: 基于注意力图引导的变换器剪枝，用于边缘设备上的遮挡行人再识别。IEEE 多媒体杂志 25，1592-1599(2023)

40. Patrini, G., Rozza, A., Krishna Menon, A., Nock, R., Qu, L.: Making deep neural networks robust to label noise: A loss correction approach. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 1944-1952 (2017)

40. Patrini, G., Rozza, A., Krishna Menon, A., Nock, R., Qu, L.: 使深度神经网络对标签噪声具有鲁棒性的方法: 一种损失校正策略。发表于:IEEE 计算机视觉与模式识别会议，第 1944-1952 页 (2017)

41. Peng, X., Wang, K., Zeng, Z., Li, Q., Yang, J., Qiao, Y.: Suppressing mislabeled data via grouping and self-attention. In: Eur. Conf. Comput. Vis. pp. 786-802 (2020)

41. Peng, X., Wang, K., Zeng, Z., Li, Q., Yang, J., Qiao, Y.: 通过分组和自注意力抑制误标数据。发表于: 欧洲计算机视觉会议，第 786-802 页 (2020)

42. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: IEEE Conf. Com-put. Vis. Pattern Recog. pp. 6517-6525 (2017)

42. Redmon, J., Farhadi, A.: YOLO9000: 更好、更快、更强大。在:IEEE 计算机视觉与模式识别会议，第 6517-6525 页 (2017)

43. Ren, M., Zeng, W., Yang, B., Urtasun, R.: Learning to reweight examples for robust deep learning. In: Int. Conf. Mach. Learn. pp. 4331-4340 (2018)

43. Ren, M., Zeng, W., Yang, B., Urtasun, R.: 学习重新加权样本以实现鲁棒深度学习。在: 国际机器学习会议，第 4331-4340 页 (2018)

44. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Adv. Neural Inform. Process. Syst. pp. 91-99 (2017)

44. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster R-CNN: 面向实时目标检测的区域建议网络。在: 先进神经信息处理系统，第 91-99 页 (2017)

45. Sheng, M., Sun, Z., Cai, Z., Chen, T., Zhou, Y., Yao, Y.: Adaptive integration of partial label learning and negative learning for enhanced noisy label learning. In: AAAI. pp. 4820-4828 (2024)

45. Sheng, M., Sun, Z., Cai, Z., Chen, T., Zhou, Y., Yao, Y.: 自适应整合部分标签学习与负样本学习以增强噪声标签学习。在:AAAI 会议，第 4820-4828 页 (2024)

46. Shi, X., Guo, Z., Li, K., Liang, Y., Zhu, X.: Self-paced resistance learning against overfitting on noisy labels. Pattern Recognition 134, 109080 (2023)

46. Shi, X., Guo, Z., Li, K., Liang, Y., Zhu, X.: 针对噪声标签过拟合的自我调节抗拒学习。模式识别，134， 109080(2023)

47. Shu, J., Xie, Q., Yi, L., Zhao, Q., Zhou, S., Xu, Z., Meng, D.: Meta-weight-net: Learning an explicit mapping for sample weighting. In: Adv. Neural Inform. Process. Syst. pp. 1917-1928 (2019)

47. Shu, J., Xie, Q., Yi, L., Zhao, Q., Zhou, S., Xu, Z., Meng, D.: 元加权网络: 学习样本加权的显式映射。在: 先进神经信息处理系统，第 1917-1928 页 (2019)

48. Sosea, T., Caragea, C.: Marginmatch: Improving semi-supervised learning with pseudo-margins. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 15773-15782 (2023)

48. Sosea, T., Caragea, C.: Marginmatch: 利用伪边缘改善半监督学习。在:IEEE 计算机视觉与模式识别会议，第 15773-15782 页 (2023)

49. Sun, Z., Hua, X.S., Yao, Y., Wei, X.S., Hu, G., Zhang, J.: Crssc: salvage reusable samples from noisy data for robust learning. In: ACM Int. Conf. Multimedia. pp. 92-101 (2020)

49. Sun, Z., Hua, X.S., Yao, Y., Wei, X.S., Hu, G., Zhang, J.: Crssc: 从噪声数据中回收可重用样本以实现鲁棒学习。在:ACM 多媒体国际会议，第 92-101 页 (2020)

50. Sun, Z., Liu, H., Wang, Q., Zhou, T., Wu, Q., Tang, Z.: Co-Idl: A co-training-based label distribution learning method for tackling label noise. IEEE Trans. Multimedia pp. 1093-1104 (2022)

50. Sun, Z., Liu, H., Wang, Q., Zhou, T., Wu, Q., Tang, Z.: Co-Idl: 一种基于协同训练的标签分布学习方法，用于应对标签噪声。IEEE 多媒体学报，第 1093-1104 页 (2022)

51. Sun, Z., Shen, F., Huang, D., Wang, Q., Shu, X., Yao, Y., Tang, J.: Pnp: Robust learning from noisy labels by probabilistic noise prediction. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 5311-5320 (2022)

51. Sun, Z., Shen, F., Huang, D., Wang, Q., Shu, X., Yao, Y., Tang, J.: Pnp: 通过概率噪声预测实现对噪声标签的鲁棒学习。在:IEEE 计算机视觉与模式识别会议，第 5311-5320 页 (2022)

52. Sun, Z., Yao, Y., Wei, X.S., Zhang, Y., Shen, F., Wu, J., Zhang, J., Shen, H.T.: Webly supervised fine-grained recognition: Benchmark datasets and an approach. In: Int. Conf. Comput. Vis. pp. 10602-10611 (2021)

52. Sun, Z., Yao, Y., Wei, X.S., Zhang, Y., Shen, F., Wu, J., Zhang, J., Shen, H.T.: Webly 监督的细粒度识别: 基准数据集与方法。在: 国际计算机视觉会议，第 10602-10611 页 (2021)

53. Tu, Y., Zhang, B., Li, Y., Liu, L., Li, J., Wang, Y., Wang, C., Zhao, C.: Learning from noisy labels with decoupled meta label purifier. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 19934-19943 (2023)

53. Tu, Y., Zhang, B., Li, Y., Liu, L., Li, J., Wang, Y., Wang, C., Zhao, C.: 利用解耦元标签净化器从噪声标签中学习。在:IEEE 计算机视觉与模式识别会议，第 19934-19943 页 (2023)

54. Tu, Y., Zhang, B., Li, Y., Liu, L., Li, J., Zhang, J., Wang, Y., Wang, C., Zhao, C.: Learning with noisy labels via self-supervised adversarial noisy masking. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 16186-16195 (2023)

54. Tu, Y., Zhang, B., Li, Y., Liu, L., Li, J., Zhang, J., Wang, Y., Wang, C., Zhao, C.: 通过自监督对抗噪声掩码实现噪声标签学习。在:IEEE 计算机视觉与模式识别会议，第 16186-16195 页 (2023)

55. Vahdat, A.: Toward robustness against label noise in training deep discriminative neural networks. In: Adv. Neural Inform. Process. Syst. pp. 5596-5605 (2017)

55. Vahdat, A.: 面向深判别神经网络训练中标签噪声鲁棒性的研究。在: 先进神经信息处理系统，第 5596-5605 页 (2017)

56. Veit, A., Alldrin, N., Chechik, G., Krasin, I., Gupta, A., Belongie, S.: Learning from noisy large-scale datasets with minimal supervision. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 6575-6583 (2017)

56. Veit, A., Alldrin, N., Chechik, G., Krasin, I., Gupta, A., Belongie, S.: 在最少监督下从大规模噪声数据集中学习。在:IEEE 计算机视觉与模式识别会议，第 6575-6583 页 (2017)

57. Wang, X., Hua, Y., Kodirov, E., Clifton, D.A., Robertson, N.M.: Proselflc: Progressive self label correction for training robust deep neural networks. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 752-761 (2021)

57. Wang, X., Hua, Y., Kodirov, E., Clifton, D.A., Robertson, N.M.: Proselflc: 用于训练鲁棒深度神经网络的渐进式自我标签校正。在:IEEE 计算机视觉与模式识别会议，第 752-761 页 (2021)

58. Wei, H., Feng, L., Chen, X., An, B.: Combating noisy labels by agreement: A joint training method with co-regularization. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 13723-13732 (2020)

58. Wei, H., Feng, L., Chen, X., An, B.: 通过一致性应对噪声标签: 一种联合训练方法与协正则化。在:IEEE 计算机视觉与模式识别会议，第 13723-13732 页 (2020)

59. Wei, J., Liu, H., Liu, T., Niu, G., Sugiyama, M., Liu, Y.: To smooth or not? when label smoothing meets noisy labels. In: Int. Conf. Mach. Learn. pp. 23589-23614 (2022)

59. Wei, J., Liu, H., Liu, T., Niu, G., Sugiyama, M., Liu, Y.: 是否平滑？当标签平滑遇到噪声标签。在: 国际机器学习会议，第 23589-23614 页 (2022)

60. Welinder, P., Branson, S., Belongie, S.J., Perona, P.: The multidimensional wisdom of crowds. In: Adv. Neural Inform. Process. Syst. pp. 2424-2432 (2010)

60. Welinder, P., Branson, S., Belongie, S.J., Perona, P.: 群体的多维智慧。在: 先进神经信息处理系统，第 2424-2432 页 (2010)

61. Wu, T., Dai, B., Chen, S., Qu, Y., Xie, Y.: Meta segmentation network for ultra-resolution medical images. In: IJCAI. pp. 544-550 (2020)

61. Wu, T., Dai, B., Chen, S., Qu, Y., Xie, Y.: 超分辨率医学图像的元分割网络。在: 国际人工智能会议，第 544-550 页 (2020)

62. Xia, X., Han, B., Wang, N., Deng, J., Li, J., Mao, Y., Liu, T.: Extended $t$t: Learning with mixed closed-set and open-set noisy labels. IEEE Trans. Pattern Anal. Mach. Intell. pp. 3047-3058 (2023)

62. Xia, X., Han, B., Wang, N., Deng, J., Li, J., Mao, Y., Liu, T.: 扩展的 t-t: 混合封闭集和开放集噪声标签的学习。在:IEEE 模式分析与机器智能杂志，第 3047-3058 页 (2023)

63. Xia, X., Liu, T., Han, B., Gong, M., Yu, J., Niu, G., Sugiyama, M.: Sample selection with uncertainty of losses for learning with noisy labels. In: Int. Conf. Learn. Represent. (2022)

63. Xia, X., Liu, T., Han, B., Gong, M., Yu, J., Niu, G., Sugiyama, M.: 利用损失不确定性进行样本选择以应对噪声标签的学习。在: 国际学习表征会议 (2022)

64. Xia, X., Liu, T., Wang, N., Han, B., Gong, C., Niu, G., Sugiyama, M.: Are anchor points really indispensable in label-noise learning? In: Adv. Neural Inform. Process. Syst. pp. 6835-6846 (2019)

64. Xia, X., Liu, T., Wang, N., Han, B., Gong, C., Niu, G., Sugiyama, M.: 在带噪声标签学习中，锚点真的不可或缺吗？在: 先进神经信息处理系统，第 6835-6846 页 (2019)

65. Yang, E., Yao, D., Liu, T., Deng, C.: Mutual quantization for cross-modal search with noisy labels. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 7541-7550 (2022)

65. Yang, E., Yao, D., Liu, T., Deng, C.: 用于带噪声标签的跨模态搜索的互量化。在:IEEE 计算机视觉与模式识别会议，第 7541-7550 页 (2022)

66. Yao, Y., Sun, Z., Zhang, C., Shen, F., Wu, Q., Zhang, J., Tang, Z.: Jo-src: A contrastive approach for combating noisy labels. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 5192-5201 (2021)

66. Yao, Y., Sun, Z., Zhang, C., Shen, F., Wu, Q., Zhang, J., Tang, Z.: Jo-src: 一种对抗噪声标签的对比方法。在:IEEE 计算机视觉与模式识别会议，第 5192-5201 页 (2021)

67. Yao, Y., Liu, T., Han, B., Gong, M., Deng, J., Niu, G., Sugiyama, M.: Dual t: Reducing estimation error for transition matrix in label-noise learning. In: Adv. Neural Inform. Process. Syst. (2021)

67. Yao, Y., Liu, T., Han, B., Gong, M., Deng, J., Niu, G., Sugiyama, M.: 双 t: 在带噪声标签学习中减少转移矩阵的估计误差。在: 先进神经信息处理系统 (2021)

68. Yi, K., Wu, J.: Probabilistic end-to-end noise correction for learning with noisy labels. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 7017-7025 (2019)

68. Yi, K., Wu, J.: 用于带噪声标签学习的概率端到端噪声校正。在:IEEE 计算机视觉与模式识别会议，第 7017-7025 页 (2019)

69. Yu, X., Han, B., Yao, J., Niu, G., Tsang, I.W., Sugiyama, M.: How does disagreement help generalization against label corruption? In: Int. Conf. Mach. Learn. pp. 7164-7173 (2019)

69. Yu, X., Han, B., Yao, J., Niu, G., Tsang, I.W., Sugiyama, M.: 争议如何帮助模型在标签污染中实现泛化？在: 国际机器学习会议，第 7164-7173 页 (2019)

70. Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O.: Understanding deep learning requires rethinking generalization. In: Int. Conf. Learn. Represent. (2017)

70. Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O.: 理解深度学习需要重新思考泛化能力。在: 国际表征学习会议 (2017)

71. Zhou, X., Liu, X., Zhai, D., Jiang, J., Ji, X.: Asymmetric loss functions for noise-tolerant learning: Theory and applications. IEEE Trans. Pattern Anal. Mach. In-tell. **45** (7) , 8094 − 8109 (2023)

71. Zhou, X., Liu, X., Zhai, D., Jiang, J., Ji, X.: 用于噪声容忍学习的非对称损失函数: 理论与应用。在:IEEE 模式分析与机器智能杂志 **45** (7) , 8094 − 8109 (2023)