DOI:10.16356/j.1005-2615.2025.05.005

开放世界下带有分布内和分布外噪声的长尾学习

郑金鹏[1], 李绍园[1,2], 朱晓林[3], 黄圣君[1], 陈松灿[1], 王康侃[4]

(1.南京航空航天大学计算机科学与技术学院模式分析与机器智能工业和信息化部重点实验室,南京 211106;
2.南京大学计算机软件新技术全国重点实验室,南京 210093; 3.江苏省产品质量监督检验研究院,南京 210001;
4.南京理工大学计算机科学与工程学院,南京 210094)

摘要: 在训练深度神经网络时,实际应用数据常存在长尾类别分布、分布内噪声和分布外噪声等偏差。现有方法多单独解决类别不平衡或含噪声标签问题,很少同时考虑两者,尤其是两种噪声并存时。本文提出不平衡噪声标签校准(Imbalanced noisy label calibration, INLC)方法,用模型一致性预测筛选分布外样本并赋予均匀标签,增强模型对其检测能力;对分布内样本,利用 Jensen-Shannon 散度区分噪声,减少干净样本误分类,尤其针对尾部类别;引入额外语义分类器,缓解伪标签对多数类的偏向性以应对类别不平衡;采用基于强数据增强的一致性正则化方法提升模型泛化性能。在模拟和真实数据集上的实验表明,INLC 显著减轻了标签噪声和类别不平衡的影响,分类准确率较优异基线方法提高 2% 以上。

关键词: 长尾学习;开放世界;分布内和分布外噪声;伪标签;不平衡

中图分类号: TP311.5 **文献标志码:** A **文章编号:** 1005-2615(2025)05-0842-10

Long-Tailed Learning with In- and Out-of-Distribution Noisy Labels in the Open World

ZHENG Jinpeng[1], *LI Shaoyuan*[1,2], *ZHU Xiaolin*[3], *HUANG Shengjun*[1],
CHEN Songcan[1], *WANG Kangkan*[4]

(1. MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China; 2. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China; 3. Jiangsu Product Quality Testing and Inspection Institute, Nanjing 210001, China; 4. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: When training deep neural networks in practical application scenarios, the data used often have various biases, such as long-tailed category distributions, in-distribution noise, and out-of-distribution noise. Most existing methods focus on solving the problem of category imbalance or dealing with noisy labels, but rarely consider both aspects simultaneously, especially when in- and out-of-distribution noises exist at the same time. We propose an imbalanced noisy labels calibration (INLC) method to address this challenge. To handle out-of-distribution samples, we use the model's consistent predictions to filter them out and assign uniform labels, thereby enhancing the model's ability to detect out-of-distribution samples. For in-distribution samples, we use the Jensen-Shannon divergence to distinguish noise and reduce misclassification of clean samples, especially in tail categories. To address the problem of category imbalance, we introduce an

基金项目: 中央高校基本科研业务费专项资金(NS2024059);国家自然科学基金(62376126)。
收稿日期: 2024-01-30;**修订日期:** 2024-04-25

通信作者: 李绍园,女,副教授,E-mail:lisy@nuaa.edu.cn。

引用格式: 郑金鹏,李绍园,朱晓林,等. 开放世界下带有分布内和分布外噪声的长尾学习[J]. 南京航空航天大学学报(自然科学版),2025,57(5):842-851. ZHENG Jinpeng, LI Shaoyuan, ZHU Xiaolin, et al. Long-tailed learning with in- and out-of-distribution noisy labels in the open-world[J]. Journal of Nanjing University of Aeronautics & Astronautics(Natural Science Edition),2025, 57(5):842-851.

additional semantic classifier to mitigate the bias of pseudo-labels towards majority categories. Finally, we adopt a consistency regularization method based on strong data augmentation to further improve the model's generalization performance. We conducted extensive experiments on simulated and real-world datasets, covering different levels of category imbalance from low to high and different proportions of label noise. Experimental results show that INLC significantly alleviates the impact of label noise and category imbalance, and improves the classification accuracy by more than 2% compared with the previous state-of-the-art baseline methods.

Key words: long-tailed learing; open world; in- and out-of-distribution noisy; pseudo-labels; imbalance

深度神经网络(Deep neural networks，DNNs)在图像分类任务中取得了巨大成功[1]。然而，训练所需的大规模标注数据在许多场景难以获取，一种常见的替代方法是通过互联网收集样本[2]，但这往往会导致所构建的数据集中存在偏置。图 1 展示了相关的数据偏置。在不平衡的类别分布中，“猫”是一个头部类别，而“响尾蛇”是一个尾部类别。每个类别都可能包含：分布内(In-distribution，ID)噪声：样本来自已知类，但标签标注错误；分布外(Out-of-distribution，OOD)噪声：样本属于未知类，但被标注为已知类。在现实世界的数据收集中，通过网络爬虫获取的图像数据集往往存在类别分布不均衡的长尾现象，而基于查询搜索得到的样本则常因标签与实际类别不符而携带噪声。值得关注的是，OOD 噪声在实际场景中更为普遍。以网络爬虫构建的 mini-WebVision 数据集为例，其中 25% 的数据属于分布外噪声，而分布内噪声仅占 5%[3]。具体来看网络图像数据：搜索引擎返回的结果不仅呈现显著的类别失衡(如“宠物猫”图片数量远超过“稀有鸟类”)，且由于自动抓取机制或用户上传过程中的误差，标签错误率居高不下，尤其是 OOD 样本的占比已达到不可忽视的程度。在这种类别分布偏差与标签噪声交织的复杂环境中，传统单一的处理策略已难以有效应对。在长尾数据集中，特征空间在很大程度上受到多数(头部)类的影响，这使得在这些分布上训练的分类器倾向于关注多数类别，而忽视少数(尾部)类[4]。此外，深度神经网络容易过度拟合噪声标签，从而导致泛化性能较差[5]。因此，现实面临着一个多种数据偏置共存的复杂场景：长尾类别不平衡、分布内噪声以及 OOD 噪声。

图 1　问题说明

Fig.1　Illustration of the problem setting

在相关文献中，长尾学习方法主要侧重帮助模型对少数类的学习[4]。然而，当样本包含噪声标签时，重平衡(Re-balance)方法处理的数据总会包含少数类的噪声样本，从而加剧了过参数化 DNNs 中的过拟合问题[4,6]。对于处理噪声标签的学习方法而言，由于少数类样本的数据稀缺使得准确恢复被错误标注的标签变得极具挑战[7]。尽管已经有一些研究工作针对长尾数据和噪声标签的结合问题提出了解决方案[8-10]，但它们主要处理 ID 噪声，而忽略了 OOD 噪声。由于 OOD 样本的真实标签位于数据分布之外，无法进行推断。因此，模型不仅要从长尾数据中学习，还必须有效地缓解对混合噪声样本的过拟合问题。

解决噪声样本问题的一种常见方法是利用 DNNs 往往会先学习简单样本，之后才会过拟合噪声的特性，在半监督学习框架内先筛选噪声样本[7]。为了更好地利用数据，被筛选出来的样本通常不会被丢弃，而是被当作无标签样本处理。对于 ID 噪声，基于伪标签的自举法(Bootstrapping)常被用于恢复噪声样本的标签[11-12]。在处理 OOD 噪声方面，近期的研究工作[13-15]表明，引入一个包含 OOD 样本辅助数据集，能帮助模型提高 ID 样本分类的鲁棒性，从而生成更具判别性的表征。一种常见的做法是给 OOD 样本重新赋予均匀标签，并训练模型使其拟合均匀分布。

然而，选择合适的指标来筛选噪声样本仍然具有挑战性，因为许多方法依赖于交叉熵损失[3]。在类别不平衡的数据集中，少数类样本由于数量稀少从而训练不充分，其损失结果往往与噪声样本相似。此外，由于分类器往往倾向于对头部类，造成了自举生成的伪标签带有偏置，从而加剧类别不平衡[3,8]。此外，在实际场景中，OOD 数据很少作为单独的辅助数据集提供；相反，它通常会随

机混入训练集中。这就要求在训练过程中,对ID噪声样本和OOD噪声样本进行区分,并分别进行处理。

本文提出了一种不平衡噪声标签校准(Imbalanced noisy label calibration,INLC)框架,用于处理不平衡数据集中的两种类型噪声。INLC框架如图2所示,图中变量含义见后文。首先,针对OOD样本分布外的特性,利用模型的一致性预测来分离出OOD样本,并为它们重新分配均匀标签,从而增强模型对OOD样本的检测能力。受文献[16]启发,本文采用詹森-香农(Jensen-Shannon,JS)散度作为ID样本的噪声分离指标。本文实验表明,JS散度在识别干净样本方面具有更高的召回率和精确率,尤其是在尾部类别中。为了减轻模型对头部类别的固有偏置,本文借鉴了文献[17]中的技术,引入了一个额外的分类器,它通过比较类原型相似度生成语义伪标签,与通过自举生成的伪标签结合,以缓解生成不平衡伪标签带来的偏置。最后,本文引入了采用强数据增强策略的一致性正则化方法,以进一步提升模型的泛化性能。

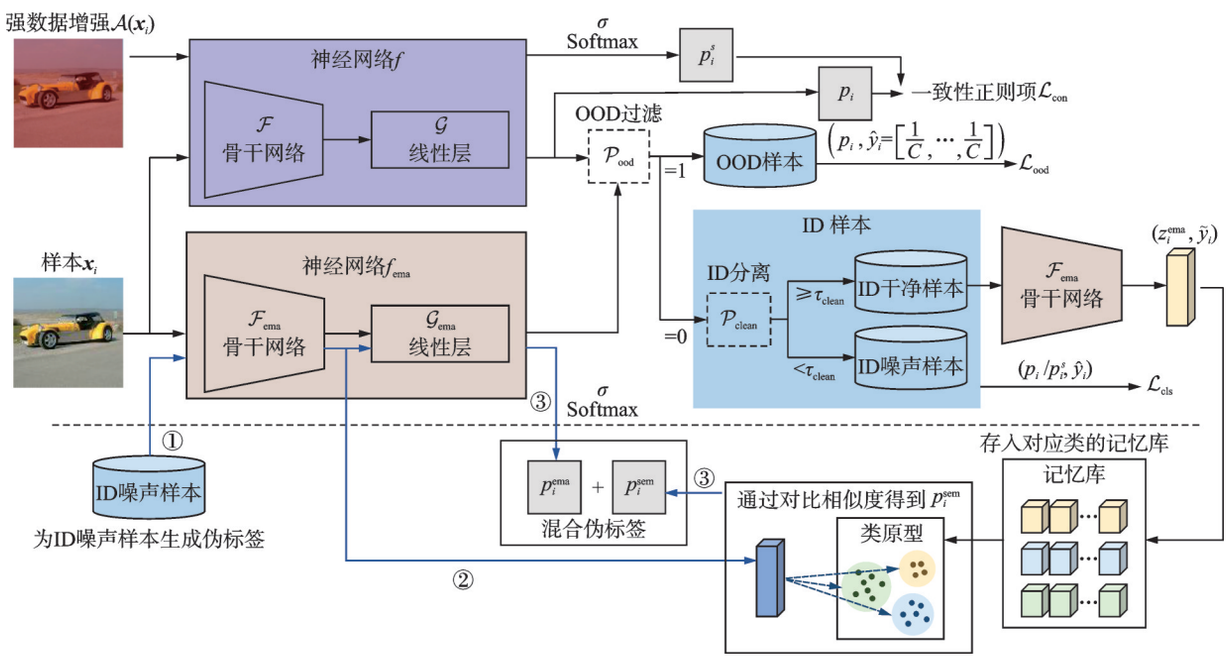

图2 INLC框架图

Fig.2 INLC framework

1 相关工作

1.1 OOD噪声标签学习

关于OOD噪声标签的学习,最近已有许多相关研究方法被提出,它们旨在研究数据集中同时存在ID噪声与OOD噪声的场景[18]。文献[16]通过JS散度来选择干净样本,并利用一致性预测区分ID样本和OOD样本。文献[3]调整了交叉熵损失的计算,通过使用Beta混合模型拟合ID样本和OOD样本损失实现分离,并动态调整OOD样本的标签。文献[19]通过训练两个网络来同时预测类别标签和噪声类型。文献[9]基于文献[8]通过对比学习和K近邻(K-nearest neighbor,KNN)以检测并排除OOD噪声。本文额外考虑了类别不平衡的情况。

1.2 长尾数据下的噪声标签学习

长尾数据下的噪声标签学习同时考虑了数据呈长尾分布和存在ID噪声的场景[20-21]。文献[10]设计了正则化方法,用于统一处理噪声样本和少数类样本。文献[22]利用了噪声数据以及尾部类样本训练时损失变化趋势计算合适的样本权重。文献[23]引入了额外的度量标准,用于区分噪声样本和少数类别。文献[24]利用无监督对比学习表征,并采用知识迁移的方式帮助分类器对尾部类的学习。文献[11]将样本划分为干净样本子集和含噪声样本子集,并过滤掉置信度较低的样本。本文额外考虑了存在OOD噪声的场景。

2 本文方法

2.1 预备知识

考虑一个共有C类的多分类任务,训练集$\mathcal{D}=\{(x_i,\tilde{y}_i)\}_{i=1}^{N}\subset\mathcal{X}\times\mathcal{Y}$共有$N$个样本,其中$\mathcal{X}$表示输入空间,$\mathcal{Y}$表示标签空间,并且$\mathcal{Y}=\{1,2,\cdots,C\}$。由于噪声标签的存在,观察到的标签$\tilde{y}_i$可能不同于真实标签(Ground-truth),使用one-hot变化处理

$\tilde{y}_i$ 并引入标签平滑因子 ϵ，得到长度为 C 向量 $\hat{\boldsymbol{y}}_i$。对于ID噪声样本与OOD噪声样本的比例分别用 r_{id} 与 r_{ood} 表示。第 c 类的样本数表示为 n_c，且 $N=\sum_{c=1}^{C} n_c$。假设类别标签的是按样本数递减的顺序标注，即 $n_1>n_2>\cdots>n_C$，则不平衡率 $r_{\text{imb}}=n_1/n_C$。本文的目标是训练一个基于DNNs的映射函数 $f=\mathcal{G}\circ\mathcal{F}$，其中，$\mathcal{F}:\mathcal{X}\to\mathcal{Z}$ 表示特征提取的骨干网络，$\mathcal{Z}$ 表示特征空间，$\mathcal{G}:\mathcal{Z}\to\mathcal{Y}$ 表示由线性层实现的分类头，模型参数表示为 $\boldsymbol{\theta}$。模型预测概率分布 $\boldsymbol{p}_i=\sigma(f(\boldsymbol{x}_i))$，其中 σ 表示 softmax 激活函数，$f(\boldsymbol{x}_i)$ 表示 $\boldsymbol{x}_i$ 的 logits。交叉熵损失函数表示为 l。

2.2　框架概述

在每个训练轮次（epoch）的每次批次迭代中，考虑到OOD样本的真实标签在 $\mathcal{Y}$ 之外，本文首先使用去偏模型的一致性预测过滤OOD样本，并为OOD样本分配均匀标签，增强模型区分ID/OOD样本的能力。对于ID噪声，本文发现与基于损失的方法相比，使用JS散度对尾部样本能实现更高的召回率和精确率，有效地防止了损失较大的尾部样本被错误归类为ID噪声。为了进一步解决不平衡数据集中伪标签偏向多数类的问题本文额外引入了一个语义分类器。具体来说，本文根据给定的标签 $\hat{\boldsymbol{y}}_i$，将干净样本的表征存储在记忆库中，然后从记忆库中构建每个类的类原型，并分别与ID噪声样本的表征比较，以生成语义伪标签。语义标签会与线性层预测得到线性伪标签结合，实现对ID噪声样本的标签校准。为了进一步提高模型泛化能力，本文应用了数据增强和一致性正则化。框架整体流程如图2所示。

2.3　样本分离和标签校准

2.3.1　OOD噪声校准

由于OOD样本的真实标签在 $\mathcal{Y}$ 之外，DNNs在对其进行预测时往往会表现出不确定性[13]。基于以上观察，通过比较协同猜测（Co-guessing）的一致性来过滤OOD噪声。为此，引入 $f_{\text{ema}}=\mathcal{G}_{\text{ema}}\circ\mathcal{F}_{\text{ema}}$，其架构与 f 相同。然而，它的参数 $\boldsymbol{\theta}_{\text{ema}}$ 是使用 $\boldsymbol{\theta}$ 的动量系数为 m 的指数移动平均（Exponential moving average，EMA）更新，即

$$\boldsymbol{\theta}_{\text{ema}}\leftarrow m\boldsymbol{\theta}_{\text{ema}}+(1-m)\boldsymbol{\theta} \tag{1}$$

考虑到类别不平衡，模型往往会偏向于多数类。因此，对logits $f(\boldsymbol{x}_i)$ 进行后调整，有

$$f^{*}(\boldsymbol{x}_i)=f(\boldsymbol{x}_i)-\log\boldsymbol{\pi} \tag{2}$$

式中 $\boldsymbol{\pi}$ 表示训练集的类别先验概率，可提前计算得到。然后比较它们的预测结果，有

$$\mathcal{P}_{\text{ood}}(\boldsymbol{x}_i)=\min(1,|\text{argmax}f^{*}(\boldsymbol{x}_i)-\text{argmax}f^{*}_{\text{ema}}(\boldsymbol{x}_i)|) \tag{3}$$

式中 $f^{*}_{\text{ema}}(\boldsymbol{x}_i)$ 表示对 $f_{\text{ema}}(\boldsymbol{x}_i)$ 进行后调整后 logits。显然，$\mathcal{P}_{\text{ood}}(\boldsymbol{x}_i)\in\{0,1\}$，使用二值变量 $\mathcal{P}_{\text{ood}}(\boldsymbol{x}_i)$ 表示 $\boldsymbol{x}_i$ 属于OOD噪声样本概率：$\mathcal{P}_{\text{ood}}(\boldsymbol{x}_i)=1$ 的样本视为OOD噪声，而其余样本则被认定为ID样本。

文献[13-14]研究表明，合理利用OOD样本可以增强所学习到的表征鲁棒性。文献[13]指出，OOD噪声可能会引导模型趋向于更平坦的极小值点，从而使模型对OOD样本做出更保守的预测。对于被过滤的OOD样本，将 $\hat{\boldsymbol{y}}_i$ 的每个分量都设置为 $1/C$（即 $\hat{y}_{i,c}=1/C$），以此鼓励模型的预测 $\boldsymbol{p}_i$ 接近均匀分布。关于OOD样本的损失计算表示为

$$\mathcal{L}_{\text{ood}}=\mathbb{E}_{\mathcal{P}_{\text{ood}}(\boldsymbol{x}_i)=1}[l(\boldsymbol{p}_i,\hat{\boldsymbol{y}}_i)] \tag{4}$$

2.3.2　ID噪声校准

对于剩余的ID样本，本文采用半监督学习范式来区分含噪声样本，并恢复它们的真实标签。在含噪声标签场景下，使用交叉熵损失训练的DNNs往往会倾向于优先拟合简单样本[7]。这一现象催生了各种不同的方法[3,9]，这些方法都强调选择损失较小样本作为干净样本。然而，尾部类样本数量较少，模型难以充分学习，并表现出较高的损失[10]。为此，使用JS散度计算干净样本的概率，即

$$D_{\text{JS}}(\boldsymbol{p}_i\|\hat{\boldsymbol{y}}_i)=\frac{D_{\text{KL}}\left(\boldsymbol{p}_i\middle\|\frac{\boldsymbol{p}_i+\hat{\boldsymbol{y}}_i}{2}\right)+D_{\text{KL}}\left(\hat{\boldsymbol{y}}_i\middle\|\frac{\boldsymbol{p}_i+\hat{\boldsymbol{y}}_i}{2}\right)}{2} \tag{5}$$

JS散度用于量化两个概率分布之间的差异，与无界的交叉熵损失不同，它的值域在[0,1]之间。干净样本的概率表示为

$$\mathcal{P}_{\text{clean}}(\boldsymbol{x}_i)=1-D_{\text{JS}}(\boldsymbol{p}_i\|\hat{\boldsymbol{y}}_i)\in[0,1] \tag{6}$$

本文将 $\mathcal{P}_{\text{clean}}(\boldsymbol{x}_i)\geqslant\tau_{\text{clean}}$ 的样本认定为干净样本，记作 $\mathbb{S}_{\text{clean}}$。图3(a)展示了 $r_{\text{id}}=0.1$，$r_{\text{ood}}=0.1$，$r_{\text{imb}}=100$ 的CIFAR-10下的尾部类召回率与精确率变化，不同的颜色代表了使用JS散度和交叉熵损失作为过滤干净样本标准的结果。在整个训练过程中，JS散度表现出更高的召回率，并且随着训练的进行，其精确率逐渐超过交叉熵损失的精确率。如图3(b)所示，剩余的ID样本，不仅是干净样本还是噪声样本，都表现出类别不平衡现象。本文将其视为一个类别不平衡的半监督学习任务。对于ID噪声，通过线性分类器预测自举恢复噪声标签是一种常见方法[3,8]。然而，标准的线性分类器在长尾分布中往往倾向于多数类，并且使用不准确的估计可能会导致性能退化。

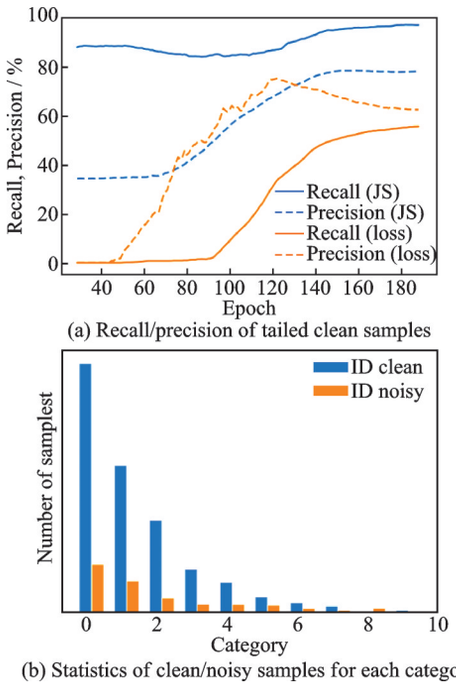

图 3 尾部类样本召回率、精确率变化和不同类别的 ID 干净/噪声样本统计

Fig.3 Recall, precision of tailed clean samples and statistics of clean/noisy samples for each category

本文筛选出满足 $\mathcal{P}_{\text{clean}}(\boldsymbol{x}_i)<\tau_{\text{clean}}$ 与 $\max(\boldsymbol{p}_i)\geqslant\tau_{\text{id}}$ 的样本,记为高置信度 ID 噪声样本 $\mathbb{S}_{\text{id}}$,并为这些样本分配伪标签。文献[17]指出语义伪标签会偏向于尾部类。基于此引入了一个语义分类器。具体来说,为每个类别建立一个记忆库,以构建类原型:$\mathcal{Q}=\{Q_c\}_{c=1}^{C}$,每个记忆库都有固定的容量 $|Q_c|$。对于干净样本,使用 EMA 更新的骨干网络提取表征:$\boldsymbol{z}_i^{\text{ema}}=\mathcal{F}_{\text{ema}}(\boldsymbol{x}_i)$,并将其存储在与给定标签 $\tilde{y}_i$ 对应的记忆库中。如果对应记忆库未达到容量上限,就将 $\boldsymbol{z}_i^{\text{ema}}$ 存入。否则,在添加新表征前,会移除最早存入的项。然后,根据每个类的记忆库内容计算相应的类原型,即

$$\boldsymbol{o}_c=\frac{1}{|Q_c|}\sum_{j=1}^{|Q_c|}\boldsymbol{z}_{c,j} \tag{7}$$

式中 $\boldsymbol{z}_{c,j}$ 表示 Q_c 中的第 j 个元素 $\boldsymbol{z}_j^{\text{ema}}$。类原型集合表示为 $\mathcal{O}=\{\boldsymbol{o}_c\}_{c=1}^{C}$。语义伪标签由基于相似度分类器得到的,该分类器计算样本表征 $\boldsymbol{z}_i$ 与 $\mathcal{O}$ 之间的相似度,具体为

$$\boldsymbol{p}_i^{\text{sem}}=\sigma(\text{sim}(\boldsymbol{z}_i,\mathcal{O})/t_{\text{proto}}) \tag{8}$$

式中:$\text{sim}(\cdot,\cdot)$ 表示余弦相似度,t_{proto} 为分类器的一个温度超参数。将语义伪标签 $\boldsymbol{p}_i^{\text{sem}}$ 与线性层预测的线性伪标签 $\boldsymbol{p}_i^{\text{ema}}=\sigma(f_{\text{ema}}(\boldsymbol{x}_i))$ 融合得到 ID 噪声样本的伪标签,有

$$\hat{\boldsymbol{y}}_i=w_{\text{id}}\cdot\boldsymbol{p}_i^{\text{ema}}+(1-w_{\text{id}})\cdot\boldsymbol{p}_i^{\text{sem}} \tag{9}$$

式中 w_{id} 为一个超参数,用于控制 $\boldsymbol{p}_i^{\text{ema}}$ 和 $\boldsymbol{p}_i^{\text{sem}}$ 间的权衡。为了增强模型的泛化性,本文使用非对称数据增强。对于 ID 样本训练目标定义为

$$\mathcal{L}_{\text{cls}}=\mathbb{E}_{\boldsymbol{x}_i\in\mathbb{S}_{\text{clean}}\cup\mathbb{S}_{\text{id}}}[l(\boldsymbol{p}_i,\hat{\boldsymbol{y}}_i)+l(\boldsymbol{p}_i^s,\hat{\boldsymbol{y}}_i)] \tag{10}$$

式中 $\boldsymbol{p}_i^s=\sigma(f(\mathcal{A}(\boldsymbol{x}_i)))$,$\mathcal{A}(\boldsymbol{x}_i)$ 表示强数据增强[25]。

为了充分利用有限的数据,提升模型检测 OOD 样本的能力,并强化表征学习,在两种不同程度的数据增强间引入一致性正则项,有

$$\mathcal{L}_{\text{con}}=w\cdot\mathbb{E}_{\boldsymbol{x}_i\in\mathcal{D}}[D_{\text{KL}}(\boldsymbol{p}_i\|\boldsymbol{p}_i^s)+D_{\text{KL}}(\boldsymbol{p}_i^s\|\boldsymbol{p}_i)] \tag{11}$$

式中:如果 $\boldsymbol{x}_i$ 属于 ID 样本,$w=1$;否则,$w=-1$。一方面,式(11)通过自监督的方式,促使模型对分布内样本和干净样本在不同视图下做出一致性预测。另一方面,它增加了模型对 OOD 样本预测的不确定性。

最终的训练目标损失函数定义为

$$L=\mathcal{L}_{\text{cls}}+\alpha\mathcal{L}_{\text{ood}}+\beta\mathcal{L}_{\text{con}} \tag{12}$$

式中 α 和 β 为控制损失项的超参数。

3 实验过程与结果

3.1 数据集与实现细节

(1)数据集 CIFAR-10/100[26]。在 CIFAR-10/1000 上模拟了类别不平衡且包含 ID/OOD 噪声数据的场景:

① 首先,设置 c 类的样本数 $n_c=n_0\cdot\left(\frac{1}{r_{\text{imb}}}\right)^{(c-1)/(n-1)}$,其中 n_0 表示 CIFAR-10/100 中 c 类的原始样本数,$r_{\text{imb}}\in\{10,100\}$。

② 随机挑选 $r_{\text{ood}}\cdot N$ 样本,替换为 ImageNet32 数据集[27]中的图像作为 OOD 样本,$r_{\text{ood}}\in\{0.1,0.2\}$。

③考虑两种 ID 噪声场景。对称噪声:噪声标签翻转为其他任意一种标签的概率相同。非对称噪声:噪声的翻转仅发生在语义相似的类。对于剩余干净样本,每个样本有 r_{id} 概率发生标签翻转,$r_{\text{id}}\in\{0.1,0.2\}$。

(2)数据集 WebVision[28]。参照文献[10]的做法,采用了一个简化版本的训练集(Mini-WebVision)。然后,在 WebVision 和 ImageNet ILSVRC12[1]验证集上对训练好的模型进行评估。

3.2 基线模型

本文工作是首次在不平衡数据集中同时处理 ID/OOD 噪声问题。为了进行公平的比较,保持了各个基线模型的原始设置,并在一致的问题设置下汇总可复现方法的结果。(1)经验风险最小化(Empirical risk minimization, ERM),仅使用标准交叉熵损失。(2)长尾学习的方法,包括 Focal[29]、Logit Adjustment[30]、LDAM[4]、MiSLAS[31]

和CSA[32]。(3)处理ID噪声的方法，包括Mixup[33]、ELR+[34]。(4)处理ID与OOD噪声的方法，包括Jo-SRC[16]、PNP[19]、DSOS[3]和MDivideMix[9]。(5)在长尾分布中处理ID噪声的方法，包括HAR[10]和CBS[11]。在每个epoch的训练阶段结束后计算测试数据上的Top-1准确率，并在实验记录中汇总最后5次epoch评估的平均准确率。

本文统一设置动量系数$m=0.999$，记忆库容量$|Q_c|=256$，语义伪标签温度参数$t_{\text{proto}}=0.5$，ID样本伪标签权重系数$w_{\text{id}}=0.5$，损失项$\alpha=0.1$，以及$\beta=0.1$。对于CIFAR-10/100，骨干网络使用PreAct ResNet18[35]，标签平滑系数$\epsilon=0.6$，高置信度ID噪声样本阈值$\tau_{\text{id}}=0.4$。对于WebVision，骨干网络使用InceptionResNet-v2[36]，$\epsilon=0.7$，$\tau_{\text{id}}=0.1$。

3.3　模拟数据集的表现

表1展示了使用CIFAR-10/100模拟场景下的平均测试分类准确率，本文使用*符号表示ID数据中的非对称噪声，并对最优结果加粗。在CIFAR-10数据集上，由于其类别数量相对较少，大多方法表现出色。在高不平衡和高噪声比例的场景中，长尾学习的方法(例如MiSLAS和CBS)取得了更好的结果。在CIFAR-100数据集上，由于其类别数量更多，处理噪声的方法与处理类别不平

表1　在具有不同不平衡和噪声率的模拟CIFAR-10/100上测试准确率

Table 1　Test accuracy on simulated CIFAR-10/100 with varying imbalance and noise rates　%

数据集	对比模型	$r_{\text{imb}}=10$				$r_{\text{imb}}=100$				均值
		$r_{\text{ood}},r_{\text{id}}=(0.1,0.1)$	$r_{\text{ood}},r_{\text{id}}=(0.2,0.1)$	$r_{\text{ood}},r_{\text{id}}=(0.2,0.2)$	$r_{\text{ood}},r_{\text{id}}=(0.2,0.2^*)$	$r_{\text{ood}},r_{\text{id}}=(0.1,0.1)$	$r_{\text{ood}},r_{\text{id}}=(0.2,0.1)$	$r_{\text{ood}},r_{\text{id}}=(0.2,0.2)$	$r_{\text{ood}},r_{\text{id}}=(0.2,0.2^*)$	
CIFAR-10	ERM	71.40	68.43	57.58	69.26	47.55	49.31	42.87	52.07	57.31
	Focal	68.66	66.75	55.26	70.41	45.32	42.67	38.85	50.21	54.77
	Logit Adj.	70.71	69.15	57.37	71.13	51.98	48.58	43.44	60.51	59.11
	LDAM	83.87	82.73	79.40	80.20	68.46	66.55	58.99	64.09	73.04
	MiSLAS	85.91	84.67	78.97	82.70	68.59	65.24	55.98	**67.82**	73.73
	CSA	80.11	72.76	68.58	76.89	59.94	56.85	50.88	59.28	65.66
	Mixup	75.27	72.17	65.90	70.65	54.70	54.79	46.62	53.56	61.71
	ELR+	82.07	85.35	83.87	**82.79**	59.91	57.29	53.46	60.88	70.70
	Jo-SRC	74.91	74.39	69.22	69.28	50.01	47.61	38.96	41.61	58.25
	PNP	73.62	69.54	61.13	66.87	51.69	47.33	42.92	47.38	57.56
	DSOS	82.84	82.64	76.77	79.11	64.98	57.42	50.58	62.48	69.60
	MDivideMix	70.29	60.88	57.90	61.31	33.49	28.24	26.36	29.86	46.06
	HAR	75.71	74.72	68.43	73.38	58.06	54.21	50.34	56.86	63.69
	CBS	78.91	79.63	78.65	79.47	63.13	63.52	57.62	65.37	70.79
	Ours	87.85	86.80	**84.02**	81.58	**73.16**	**71.58**	62.98	64.13	**76.51**
	Ours (Focal)	87.50	85.51	78.22	80.66	72.27	68.92	**63.21**	64.60	75.11
	Ours (S2)	**88.38**	**86.99**	81.45	80.38	70.37	68.51	62.03	64.48	75.73
CIFAR-100	ERM	34.18	28.32	23.93	29.95	22.16	17.79	17.16	15.98	23.68
	Focal	35.78	31.12	24.45	29.06	20.82	16.48	14.10	16.90	23.59
	Logit Adj.	36.56	33.12	23.02	30.67	24.09	20.02	14.57	19.56	25.20
	LDAM	48.75	47.70	43.01	45.75	32.51	29.88	24.12	25.28	37.12
	MiSLAS	53.17	51.31	44.40	48.22	32.28	28.34	22.33	28.86	38.61
	CSA	47.84	45.66	37.94	43.80	30.28	27.89	22.32	27.31	35.38
	Mixup	43.81	42.32	33.21	37.79	25.78	26.65	21.01	22.16	31.59
	ELR+	58.85	55.83	49.61	52.39	38.51	36.83	31.60	33.14	44.59
	Jo-SRC	42.08	39.44	36.47	33.39	24.22	21.38	19.75	22.27	29.88
	PNP	34.92	30.19	26.75	28.41	21.26	19.92	16.60	18.17	24.53
	DSOS	48.84	45.77	41.12	43.41	31.04	26.88	21.58	27.42	35.76
	MDivideMix	36.59	37.69	28.36	31.77	26.73	28.13	25.62	22.79	29.71
	HAR	43.75	41.39	36.28	38.79	28.44	24.74	21.87	24.45	32.46
	CBS	46.34	45.44	43.64	45.51	29.58	28.11	25.16	28.60	36.55
	Ours	61.92	58.78	54.49	55.01	**40.44**	**38.06**	33.09	33.39	46.90
	Ours (Focal)	61.41	57.12	54.79	52.23	40.29	33.99	**33.28**	**33.92**	45.88
	Ours (S2)	**62.25**	**60.26**	**56.26**	**57.40**	39.16	37.06	33.08	33.34	**47.35**

注：Focal表示使用Focal损失函数，S2表示改变噪声分离顺序。

衡的方法之间的性能差距变得更加明显。MiSLAS将Mixup技术融入长尾学习中,即使在高噪声和不平衡条件下也表现鲁棒的性能。ELR+引入正则化项,展现出对噪声样本的抵抗力。然而,与其他噪声学习方法类似,随着不平衡率增加,其性能显著下降。在CIFAR-100场景中,类别数量增加且每个类可用样本减少,大多数方法的性能显著下降。尽管如此,本文提出的INLC框架,旨在处理不同的数据偏差,在这种设置下仍然保持了强大的性能。与ERM相比,本文方法在CIFAR-10上平均准确率提高了19%,在CIFAR-100上提高了23%。

为了评估不同方法在样本量不同的类别上的性能,将数据集划分为多样本类、中样本类和少样本类。表2展示了$r_{\text{ood}}=0.2$,$r_{\text{id}}=0.2$以及$r_{\text{imb}}=100$的模拟CIFAR-10结果。MiSLAS有效地校准了对少样本类的预测,并减少了过度自信,显著提高了尾部类的性能。然而,它在多样本类和中样本类中未能保持正常的置信水平,导致性能下降。相比之下,ELR+忽略了数据不平衡问题,主要在数据丰富的多样本类中表现出色。而且,CBS在少样本类中的表现优于多样本类。本文方法通过考虑整体数据分布,在各类别之间实现了更为均衡的性能表现。具体而言,语义分类器通过衡量未标记样本与各类原型(即每一类样本特征的平均表示)在特征空间中的相似度,来生成语义伪标签。在面对类别不平衡问题时,由于少数类样本数量较少,其类原型可能无法准确地反映该类的真实特征分布。然而,这种基于相似度的分类机制倾向于将与少数类原型具有一定相似性的样本归为少数类,即使这些样本实际上可能属于多数类。这一特性虽然可能导致部分误分类,但也使得更多真正的少数类样本被识别出来,从而有效提升了少数类的召回率。

表2 不同样本量下的类的测试准确率

Table 2 Test accuracy for classes under different sample sizes %

模型	多样本	中样本	少样本	整体
ERM	71.03	43.90	12.43	42.87
Focal	73.10	35.35	8.13	38.85
Logit Adj.	56.03	46.35	27.30	43.44
MiSLAS	62.46	58.25	**46.83**	55.98
Mixup	80.83	42.55	16.53	46.62
ELR+	90.33	57.05	11.90	53.46
CBS	42.70	66.92	57.83	57.62
Ours	**91.66**	**68.07**	28.70	62.98
Ours(Focal)	89.90	63.65	38.63	**63.21**
Ours(S2)	88.76	59.65	38.26	62.03

在本文方法中,仅在标签校准阶段考虑了类别不平衡。本文尝试使用Focal损失函数(表中用"Focal"标注)计算$\mathcal{L}_{\text{cls}}$。然而,结果显示,使用Focal损失并不总是比交叉熵损失更好。与大多数重平衡方法类似,这种方法对头部类别产生了负面影响,未能兼顾其他类别。

此外,本文还展示了改变噪声分离顺序的结果(表中用"S2"标注)。具体而言,在S2中,首先应用式(5)和式(6)来过滤干净样本,然后使用式(3)从剩余数据中分离出OOD样本,最后对OOD和ID样本的标签进行校正。观察表2可知,与原始策略相比,改变过滤顺序对整体性能的影响微乎其微。然而,S2显著提高了尾部类别的性能。S2在尾部类别中对干净样本实现了更高的召回率,但精度较低,这导致头部类性能下降,更侧重于尾部类,这与许多处理长尾数据的方法类似。

3.4 真实数据集表现

表3展示了在真实数据集中的性能结果,包括mini-WebVision和ImageNet ILSVRC12验证集。其他方法的结果均引自文献[3,10]。从表3可以看出,本文提出的INLC框架表现出色,在两个基准测试中均取得了优异的结果。

表3 mini-WebVision测试准确率

Table 3 Test accuracy on mini-WebVision %

模型	mini-WebVision		ILSVRC12	
	Top-1	Top-5	Top-1	Top-5
ERM	62.50	80.80	58.50	81.80
Mixup	75.44	90.12	71.44	89.40
Co-Teaching	63.58	85.20	61.48	84.70
DivideMix	77.32	91.64	**75.20**	90.84
ELR+	77.78	91.68	70.29	89.76
DSOS	77.76	92.04	74.36	90.80
HAR	75.50	90.70	70.30	90.00
Ours	**78.12**	**93.36**	72.36	**92.20**

3.5　消融实验

为了评估本文提出方法中每个组件的贡献，本文在CIFAR-10数据上分别对每个组件进行了消融实验，结果如表4所示。在表4中，f^*表示式(2)中的去偏操作；D_{JS}表示使用JS散度代替交叉熵损失来过滤干净样本；$\mathcal{A}$表示应用强数据增强。结果表明，这些组件都对整体性能有积极贡献。特别是在高不平衡和高噪声率的情况下，使用交叉熵损失过滤样本会导致尾部类别的召回率和精度较低。这使得少样本类别和一些中等样本类别被错误分类，最终导致性能极差。相比之下，使用JS散度作为优化目标能够更好地平衡头部与尾部类别的学习过程，其对称性和分布匹配机制有助于提升模型对尾部类别的识别能力，同时增强对噪声样本的鲁棒性。JS散度通过鼓励输出更平滑的概率分布，避免模型过度自信于某些类别，从而在不牺牲整体性能的前提下，显著改善尾部类别的召回表现。

表4　各个组件影响下的测试准确率

Table 4　Test accuracy under influence of each component　%

f^*	D_{JS}	$\mathcal{A}$	$r_{\mathrm{imb}}=10$		$r_{\mathrm{imb}}=100$	
			$r_{\mathrm{ood}}, r_{\mathrm{id}}=$ (0.2,0.2)	$r_{\mathrm{ood}}, r_{\mathrm{id}}=$ (0.1,0.1)	$r_{\mathrm{ood}}, r_{\mathrm{id}}=$ (0.2,0.1)	$r_{\mathrm{ood}}, r_{\mathrm{id}}=$ (0.2,0.2)
√	√	√	84.02	73.16	71.58	62.98
	√	√	83.93	72.51	69.05	60.84
√		√	83.99	63.37	60.86	35.52
√	√		79.02	64.45	60.32	56.71

元组$(r_{\mathrm{imb}}, r_{\mathrm{ood}}, r_{\mathrm{id}})$表示CIFAR-10模拟设置，图4展示了超参数变化对性能的影响。在图4(a)中，设置$w_{\mathrm{id}}=1$或$w_{\mathrm{id}}=0$分别对应仅使用线性伪标签p_{ema}或语义伪标签p_{sem}。观察到语义伪标签在高噪声水平下鲁棒性不足，仅依赖其会导致训练失败。当w_{id}过大或过小时，性能均会下降。综合考虑，在其他实验中设置$w_{\mathrm{id}}=0.5$。关于损失项权重α和β，图4(b)和(c)的结果验证了$\mathcal{L}_{\mathrm{ood}}$和$\mathcal{L}_{\mathrm{con}}$的有效性。基于这些结果，在后续实验中设置$\alpha=0.1$和$\beta=1$。在图4(d)中，为避免梯度爆炸，使用$\epsilon=0.01$模拟标签平滑消融实验，并展示不同平滑程度下的性能。当平滑程度较低时，JS散度无法有效分离干净样本，导致召回率和精确率下降。根据实验结果，在其他实验中设置$\epsilon=0.6$。

表5展示了在$(r_{\mathrm{imb}}, r_{\mathrm{ood}}, r_{\mathrm{id}})=(100, 0.2, 0.2)$条件下的CIFAR-10数据集上，记忆库容量对模型性能的影响。实验数据表明，本文方法对记忆库容量$|Q_c|$的设置具有较低敏感性。综合性能表现，将$|Q_c|$的最优值设定为256。当使用FP32精度、特征维度512、$|Q_c|=256$时，每个类将额外消耗约

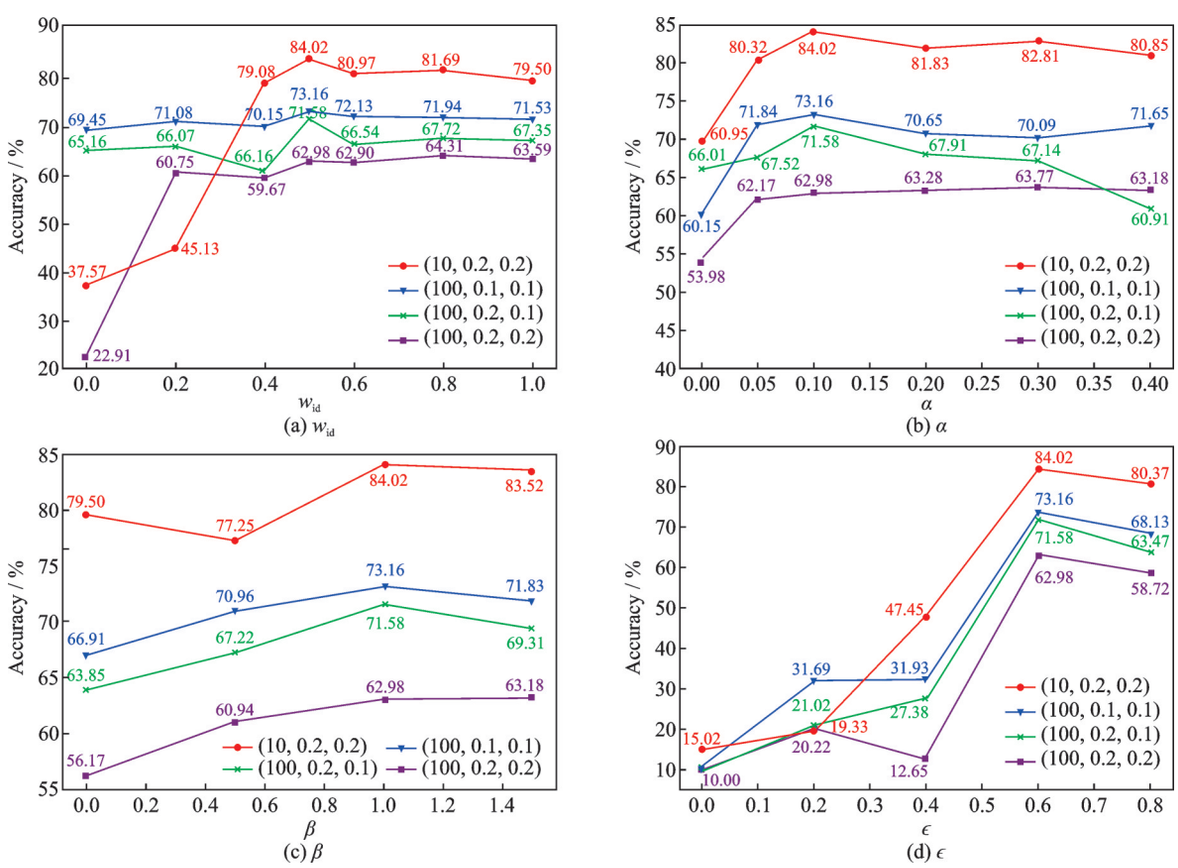

图4　超参数敏感性分析

Fig.4　Hyperparameter sensitivity analysis

512×256×4 Bytes的显存(约0.5 MB),该开销主要来自特征向量的存储。

表5 $|Q_c|$对测试准确率的影响

Table 5 Test accuracy under influence of $|Q_c|$ %

$|Q_c|$	32	64	128	256	512
Accuracy	62.11	62.07	62.70	62.98	62.03

4 结 论

本文提出了INLC框架,以解决在长尾数据上同时处理ID噪声和OOD噪声问题。首先,在考虑类别不平衡的基础上,基于预测一致性分离OOD噪声,并为其分配均匀标签以增强模型的判别能力。本文利用预测分布与对应标签之间的JS散度来过滤ID噪声。对于ID噪声,引入了一个额外的语义分类器来综合生成类别平衡的伪标签。此外,本文结合一致性正则化,以鼓励模型在不同数据增强下产生一致的预测,从而提高其泛化能力。综合实验验证了本文方法的有效性。当面对极端OOD噪声比例(如OOD噪声率超过30%)时,INLC仍存在潜在的不足。此时模型可能因大量OOD样本的干扰,导致对ID样本的特征学习出现偏差,尤其是在OOD样本与ID样本特征分布高度相似的情况下,可能会误将部分ID样本识别为OOD样本,进而影响整体分类性能。未来,本文计划从引入类别层次的软标签机制入手进行优化。考虑构建基于类别语义层次的树状结构,将OOD样本的特征通过对比学习或迁移学习的方式映射到ID样本的类别层次空间中,而非直接赋予均匀硬标签。例如,通过度量OOD样本与ID类别原型的特征相似度,动态生成介于各ID类别之间的软标签分布,使模型能够在特征迁移过程中捕捉OOD样本与ID类别间的潜在语义关联。

参考文献:

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[2] YAO Yazhou, ZHANG Jian, SHEN Fumin, et al. Web images for dataset construction: A domain robust approach[J]. IEEE Transactions on Multimedia, 2017, 19(8): 1771-1784.

[3] ALBERT P, ORTEGO D, ARAZO E, et al. Addressing out-of-distribution label noise in webly-labelled data[C]//Proceedings of the 22nd IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway, NJ: IEEE, 2022: 2393-2402.

[4] CAO K D, WEI C, GAIDON A, et al. Learning imbalanced datasets with labeldistribution-aware margin loss[C]//Proceedings of the 33rd Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2019: 1565-1576.

[5] ZHANG C Y, BENGIO S, HARDT M, et al. Understanding deep learning (still) requires rethinking generalization[J]. Communications of the ACM, 2021, 64(3): 107-115.

[6] KANG B Y, XIE S N, ROHRBACH M, et al. Decoupling representation and classifier for long-tailed recognition[C]//Proceedings of the 8th International Conference on Learning Representations. San Diego, CA: OpenReview.net, 2020.

[7] ARPIT D, JASTRZEBSKI S, BALLAS N, et al. A closer look at memorization in deep networks[C]//Proceedings of the 34th International Conference on Machine Learning. New York: ACM, 2017: 233-242.

[8] LI J, SOCHER R, HOI S C. Dividemix: Learning with noisy labels as semisupervised learning[C]//Proceedings of the 8th International Conference on Learning Representations. San Diego, CA: OpenReview.net, 2020.

[9] FOOLADGAR F, TO M N N, MOUSAVI P, et al. Manifold dividemix: A semisupervised contrastive learning framework for severe label noise[C]//Proceedings of the 41st IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2024: 4012-4021.

[10] CAO K D, CHEN Y N, LU J W, et al. Heteroskedastic and imbalanced deep learning with adaptive regularization[C]//Proceedings of the 9th International Conference on Learning Representations. San Diego, CA: OpenReview.net, 2021.

[11] LIU Huafeng, SHENG Mengmeng, SUN Zeren, et al. Learning with imbalanced noisy data by preventing bias in sample selection[J]. IEEE Transactions on Multimedia, 2024, 26: 7426-7437.

[12] 刘雅芬,郑艺峰,江铃燚,等.深度半监督学习中伪标签方法综述[J].计算机科学与探索,2022,16(6):1279-1290.
LIU Yafen, ZHENG Yifeng, JIANG Lingyi, et al. Survey on pseudo-labeling methods in deep semi-supervised learning[J]. Journal of Frontiers of Computer Science & Technology, 2022, 16(6): 1279-1290.

[13] WEI H X, TAO L, XIE R C Z, et al. Open-set label noise can improve robustness against inherent label noise[C]//Proceedings of the 35th Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2021: 7978-7992.

[14] WEI H X, TAO L, XIE R C Z, et al. Open-sampling: Exploring out-of-distribution data for re-balancing long-tailed datasets[C]//Proceedings of the 39th

International Conference on Machine Learning. New York: ACM, 2022: 23615-23630.

[15] YANG Yang, JIANG Nan, XU Yi, et al. Robust semi-supervised learning by wisely leveraging open-set data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 8334-8347.

[16] YAO Y Z, SUN Z R, ZHANG C Y, et al. Jo-SRC: A contrastive approach for combating noisy labels [C]//Proceedings of the 38th IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 5192-5201.

[17] OH Y, KIM D J, KWEON I S. Daso: Distribution-aware semantics-oriented pseudo-label for imbalanced semi-supervised learning[C]//Proceedings of the 39th IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 9786-9796.

[18] 聂晖,王瑞平,陈熙霖.开放世界物体识别与检测系统:现状、挑战与展望[J].计算机研究与发展,2024, 61(9): 2128-2141.
NIE Hui, WANG Ruiping, CHEN Xilin. Open world object recognition and detection systems: Landscapes, challenges and prospects[J]. Journal of Computer Research and Development, 2024, 61(9): 2128-2141.

[19] SUN Z R, SHEN F M, HUANG D, et al. PNP: Robust learning from noisy labels by probabilistic noise prediction[C]//Proceedings of the 39th IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 5301-5310.

[20] ZHOU B Y, CUI Q, WEI X S, et al. BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition[C]//Proceedings of the 37th IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 97199728.

[21] WEI Tong, SHI Jiangxin, TU Weiwei, et al. Robust long-tailed learning under label noise[EB/OL]. (2021-08-26).https://arxiv.org/abs/2108.11569.

[22] JIANG S W, LI J N, WANG Y, et al. Delving into sample loss curve to embrace noisy and imbalanced data[C]//Proceedings of the 36th AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI, 2022: 7024-7032.

[23] LU Y, ZHANG Y L, HAN B, et al. Label-noise learning with intrinsically long tailed data[C]//Proceedings of the 19th International Conference on Computer Vision. Piscataway, NJ: IEEE, 2023: 1369-1378.

[24] ZHANG M Y, ZHAO X Y, YAO J, et al. When noisy labels meet long tail dilemmas: A representation calibration method[C]//Proceedings of the 19th International Conference on Computer Vision. Piscataway, NJ: IEEE, 2023: 15890-15900.

[25] CUBUK E D, ZOPH B, MANE D, et al. Auto augment: Learning augmentation policies from data[EB/OL].(2019-04-11).https://arxiv.org/abs/1805.09501.

[26] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[M]. Toronto: University of Toronto, 2009.

[27] CHRABASZCZ P, LOSHCHILOV I, HUTTER F. A downsampled variant of imagenet as an alternative to the cifar datasets[EB/OL]. (2017-08-23).https://arxiv.org/abs/1707.08819.

[28] LI Wen, WANG Limin, LI Wei, et al. Webvision database: Visual learning and understanding from web data[EB/OL]. (2017-08-09). https://arxiv.org/abs/1708.02862.

[29] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the 16th International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 2980-2988.

[30] MENON A K, JAYASUMANA S, RAWAT A S, et al. Long-tail learning via logit adjustment[C]//Proceedings of the 9th International Conference on Learning Representations. San Diego, CA: OpenReview.net, 2021.

[31] ZHONG Z S, CUI J Q, LIU S, et al. Improving calibration for long-tailed recognition[C]//Proceedings of the 38th IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 16489-16498.

[32] SHI J X, WEI T, XIANG Y K, et al. How re-sampling helps for long-tail learning?[C]//Proceedings of the 37th Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2023: 75669-75687.

[33] ZHANG Hongyi. Mixup: Beyond empirical risk minimization[EB/OL]. (2018-04-27). https://arxiv.org/abs/1710.09412.

[34] LIU S, NILES-WEED J, RAZAVIAN N, et al. Early-learning regularization prevents memorization of noisy labels[C]//Proceedings of the 34th Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2020: 20331-20342.

[35] HE K M, ZHANG X Y, REN S Q, et al. Identity mappings in deep residual networks[C]//Proceedings of the 14th European Conference on Computer Vision. Berlin: Springer, 2016: 630-645.

[36] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI, 2017: 4278-4284.

(编辑:刘彦东)