

# 基于鲁棒表示学习的噪声数据处理

## 摘要

大规模数据集在收集和处理过程中产生的误差和不准确性称为噪声。模型直接拟合这些包含噪声的数据会学习其潜在的误差，导致泛化性能下降。因此，处理噪声数据是一个关键的实际应用问题。本研究选取一个利用鲁棒对比学习框架来获取数据样本的鲁棒表征的 RRL 模型来有效处理噪声数据。该框架通过对比实例级和类别级的表征来缓解噪声的影响。此外，RRL 利用样本表征的局部信息来平滑矫正标签噪声，进一步抑制噪声数据的不良影响。在复现 RRL 模型后，本文深入分析了实验结果，并引入动态类别相关阈值的策略，进一步提升模型的性能。实验结果表明，RRL 模型在多个含噪声数据集上表现出显著的鲁棒性，证明了其在处理噪声数据方面的有效性。

**关键词：**噪声学习；鲁棒表示学习；深度学习

## 1 引言

深度学习模型已经成功应用在许多领域中，例如模式识别、数据挖掘、医疗诊断等等。该过程成功的关键是大规模的数据。然而，人工收集大规模数据显然是费时费力的。作为可行的方案之一，一般采用众包或者网络爬虫等自动化方式收集大量数据。此外，由于缺乏专家知识，这些大规模的数据往往存在噪声，而这主要包括特征噪声和标签噪声。特征噪声主要指图像数据中的干扰，例如雾霾、噪点或模糊等；而标签噪声通常指数据集中的错误标注或标签缺失等问题。若深度学习网络拟合这两种噪声，模型性能都会出现下降趋势。因此，如何有效处理噪声数据称为了一个重要的实际应用问题。本选题采用的 RRL 模型通过鲁棒表示的方法来处理噪声数据，从而显著提高模型的鲁棒性。

## 2 相关工作

本节概述了标签噪声学习和对比学习领域的相关工作，并对这些研究进行分类和描述。

### 2.1 噪声学习

基于深度学习的标签噪声学习可分为两类：鲁棒损失函数和样本选择。鲁棒损失函数通过设计特定的损失函数来增强模型对噪声数据的鲁棒性。基于贝叶斯风险最小化定理，可以推导出一系列鲁棒的损失函数。文献 [15] 表明 01 损失是一种鲁棒损失函数，它将预测的标签与真实标签进行比较，如果不一致则损失为 1，否则损失为 0。相比于类别交叉熵 CCE，MAE [4] 损失对噪声更加鲁棒的。而在面对复杂数据时，MAE 的泛化性能会下降，因此 GCE

损失 [23] 被提出以扩展 MAE 损失。进一步地, Amid 等人 [1] 引入基于 Tsallis 发散的两个温度来扩展 GCE 损失。课程损失 CL [13] 和主动与被动损失 APL [14] 也被用于鲁棒的标签噪声学习中。尽管鲁棒损失函数在一些情况下表现良好, 特别是学习过程较为容易或者数据集类别数量较少时。然而, 引入鲁棒的损失函数一般会增加模型的复杂度, 从而导致学习过程更加困难, 增加训练时间。

另一方面, 样本选择方法通过筛选干净样本进行训练, 从而提高模型的鲁棒性。协同训练广泛用于多网络训练, 它一般要求包含教师网络和学生网络。MentorNet [8] 使用预训练的教师网络以协同训练的方式指导学生网络的训练。类似地, Co-teaching [6] 和 Co-teaching+ [21] 也需要两个深度学习网络, 并且每个深度学习网络选择一定数量的干净样本, 用于给另一个深度学习网络进行进一步学习。此外, 多轮迭代学习也应用于选择干净样本。NGC [20] 利用样本的低维嵌入表示来迭代构建最近邻图结构, 并进一步聚合邻域信息来选择干净样本。TopoFilter [19] 利用样本表示的空间拓扑结构来选择干净样本, 而不依赖于含噪声信号的分类器预测。此外, 结合半监督学习, 样本选择策略能够进一步提升标签噪声学习模型的性能。DivideMix [10] 使用一维双分量高斯混合模型选择噪声数据和干净数据, 分别将它们作为无标记数据和标记数据, 进而以半监督学习方式进行优化。然而, 样本选择通常基于一些强假设, 如噪声数据和干净数据分布的差异性, 这在高噪声环境下可能不再适用, 因此在噪声率较高的情况下, 样本选择方法的性能可能会有所下降。

## 2.2 对比学习

对比学习主要分为实例级对比学习和类别级对比学习两大类。实例级对比学习着重于从同一样本的不同视图提取鲁棒的表征, 而类别级对比学习则侧重于通过比较不同样本的视图来学习鲁棒的类别表征。这两种方法都旨在提升模型的特征提取和泛化能力。

实例级对比学习是一种学习样本内部结构和变化的方法。在这种方法中, 同一个实例通过不同的数据增强技术生成多个视图, 例如旋转、裁剪、颜色调整等。这些变化后的视图在本质上仍代表同一个实例, 但在表现形式上却有所不同。模型的任务是识别这些不同视图之间的内在一致性, 从而学习到对变化不敏感的鲁棒表征。这种方法常用于自监督学习领域, 因为它不依赖于外部的标注信息, 而是通过样本自身的信息来进行学习。SimCLR [3] 是一种简单而高效的对比学习框架, 它通过不同的数据增强方法生成样本视图, 并利用这些视图之间的相似性来学习鲁棒的特征表示。MoCo [7] 使用动量更新机制来维护一个动态的样本表示字典, 以增强学习过程的效率。BYOL [5] 是一种无需负样本对的自监督学习策略, 其通过预测同一样本不同视图之间的关系来学习鲁棒的表征。SwAV [2] 通过交换不同视图的类簇分配来进行有效的自监督学习, 强调了特征聚类在对比学习中的作用。实例级对比学习的关键在于如何设计有效的正负样本对。正样本对通常是同一实例的不同视图, 而负样本对则是来自不同实例的视图。通过区分这些正负样本对, 模型能够更好地理解实例的内在特性, 并在特征空间中有效地区分不同的实例。

与实例级对比学习不同, 类别级对比学习更侧重于学习不同类别之间的区别。在这种方法中, 模型不仅要识别同一类别内样本的相似性, 而且还要区分不同类别样本之间的差异。这通常通过构建类内正样本对和类间负样本对来实现, 即同一类别的样本视图形成正样本对, 不同类别的样本视图形成负样本对。类别级对比学习在监督学习和半监督学习场景中尤为重要, 因为它利用了类别标签信息来引导学习过程。Khosla 等人 [9] 研究了一种结合监督学习的对

比学习方法，该方法通过聚集相同类别样本的特征表示，同时分离不同类别样本，提高了类别区分度。PCL [12] 引入了原型对比学习，结合原型表示的有监督对比学习方法，用于提高不同类别间的区分能力。Tian 等人 [17] 提出了一种从不同视角编码信息的对比学习方法，该方法通过利用视觉和语音等多种模态的信息来增强模型的特征提取能力。

总体而言，无论是实例级还是类别级对比学习，核心目标都是通过对比和区分不同视图来学习鲁棒、区分性强的特征表示，从而提升模型在各种任务上的性能。

### 3 本文方法

此部分对本文复现的工作进行概述，包括整体模型，以及不同的模块。

#### 3.1 本文方法概述

RRL 模型 [11] 的整体框架图如图1所示：

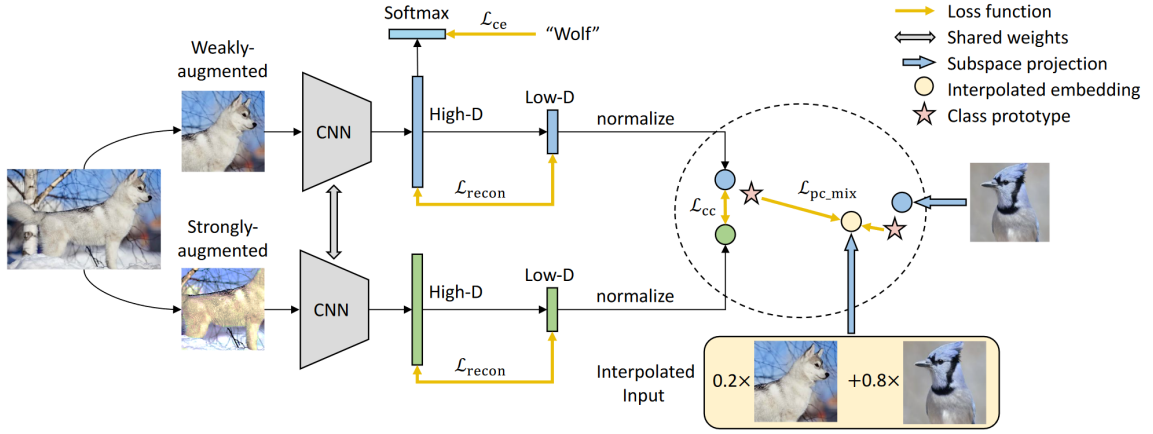


图 1. RRL 方法示意图

该模型接受含噪声数据作为输入，旨在获得良好分类性能。首先，它使用一个深度学习神经网络将输入数据转化为一个高维向量。其次，RRL 模型将该高维向量输入到一个标准 softmax 分类器中，并利用以及一个全连接线性投影层获得低维嵌入向量，该低维嵌入向量作为样本的抽象表征。最后，RRL 模型对低维嵌入构建重构损失、一致性损失以及混合原型损失，从而进一步优化模型。

#### 3.2 噪声鲁棒对比学习

首先，构建实例级无监督对比学习。为了提供不含标签噪声信号来优化表征学习过程，RRL 模型引入实例级无监督对比学习。它认为同一个图像弱增强和强增强是对比学习的正对，而负对则由该样本的弱增强与其它样本组成。通过构造这样的对比学习损失，RRL 构造正对，能够令模型对低层次图像扰动鲁棒。

其次，构建弱监督的对比原型学习。为了给对比学习的优化提供类别嵌入知识，RRL 模型通过引入弱监督的对比原型学习，学习鲁棒类别表征。具体而言，模型在每个 epoch 对同类样本的低维表示取均值，获得类别原型。RRL 模型认为样本与相应的类别原型是正对，与不同类别原型是负对，从而进行原型对比学习。为了缓解样本标签噪声的影响，使用 MixUp

对输入样本进行插值，并利用 MixUp 的标签算原型损失。RRL 模型使用 MixUp 缓解由标签噪声产生的错误嵌入知识，更好地学习类别相关的表征。

最后，构建重构损失。为了进一步提高鲁棒性，RRL 模型通过构建重构损失，提供不含标签噪声的监督信号来优化表征本身。具体而言，对模型输出的高维向量添加一个编码器（线性层），进而输出低维特征，并为该特征添加一个解码器（线性层），获得重构特征，两者计算 L2 损失进行优化。该过程类似于应用主成分分析方法，令原始特征的低维表示对输入噪声更加鲁棒。此外，SDAE [18] 模型表明最小化重构损失等价于最大化两者的互信息，因此从上述对比学习（都是利用低维特征/表示计算相似度）获得的知识能够最大化保留到高维表示中，进而辅助正则化分类器。

### 3.3 噪声标签矫正

RRL 模型利用低维子空间的结构信息生成矫正噪声的伪标签，而不直接在自身的 softmax 输出上处理，从而降低确认偏差。具体而言，RRL 模型获得样本低维子空间表示的  $k$  近邻，并根据距离生成权重。进一步地，该样本的伪标签由  $k$  近邻的模型预测加权结果与样本的模型预测结果做均值获得。此外，RRL 模型利用两个阈值挑选干净样本和伪标签样本。对于干净样本，要求模型预测中对应自身标签的概率大于阈值  $\eta_0$  (RRL 取 0.1)；对于伪标签样本，模型预测的最大概率大于阈值  $\eta_1$  (RRL 取 0.9)。RRL 模型的噪声标签矫正策略利用  $k$  近邻信息能够缓解确认偏差问题，而生成的伪标签同时利用局部平滑和模型的预测结果信息，降低用于模型优化的数据包含噪声的比例。

## 4 复现细节

### 4.1 与已有开源代码对比

本文实验代码参考了开源代码仓库 [RRL](#)，并在此基础上进行创新。

### 4.2 实验环境搭建

本文实验均在 Ubuntu 20.04.1 LTS 系统下进行，使用 Python 3.8 作为编程语言，使用 Pytorch 1.12.0 作为深度学习框架。关于硬件环境，本文使用 Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz 和 NVIDIA GeForce GTX 3090 Ti GPU 进行实验。

### 4.3 创新点

对于复现的 RRL 模型，本文对其进行了创新，主要包括以下四点：

- 引入动态类别相关阈值；
- 完美过滤开放集噪声；
- 添加 OVA-classifier [16]，并利用困难负类分类器采样策略优化；
- 结合开放熵最小化策略进一步优化 OVA-classifier；



## 5 实验结果分析

本部分对实验结果进行展示和分析，包括实验的复现结果以及添加创新模块的结果。

### 5.1 复现结果

RRL 模型在不同类型的噪声数据上进行实验，本文主要关注封闭集噪声以及混合噪声情况下的实验结果。表 1 展示了 RRL 模型在封闭集噪声数据上的实验结果。其中，该结果包含了两种类型的封闭集噪声，分别为对称噪声和非对称噪声，记为 Sym 和 Asym。CIFAR-10 数据集的对称噪声率分别为 20% 和 50%，非对称噪声率为 40%；CIFAR-100 数据集的对称噪声率分别为 20% 和 50%。表 1 中的 Paper 表示原论文中的实验结果，Ours 表示本文复现的实验结果。从表中可以看出，本文复现的实验结果与原论文中的实验结果基本一致，证明了本文复现的有效性。

表 1. 在封闭集噪声数据上的复现结果

| Dataset    |     | CIFAR-10 |         |          | CIFAR-100 |         |
|------------|-----|----------|---------|----------|-----------|---------|
| Noise type |     | Sym 20%  | Sym 50% | Asym 40% | Sym 20%   | Sym 50% |
| Paper      | cls | 95.8     | 94.3    | 91.9     | 79.1      | 74.8    |
|            | knn | 95.9     | 94.5    | 92.4     | 79.4      | 75.0    |
| Ours       | cls | 96.0     | 94.8    | 91.0     | 78.9      | 74.8    |
|            | knn | 96.0     | 94.7    | 91.2     | 78.7      | 75.2    |

进一步地，本文还对混合噪声情况进行了复现，也即包含封闭集噪声和开放集噪声，结果展示在表 2 中。两种测试情况的 CIFAR-10 数据集都包含 50% 的对称噪声。CIFAR-10+CIFAR-100 20k 表示 CIFAR-10 数据集和 CIFAR-100 数据集的混合噪声数据，从 CIFAR-100 中随机选取与 CIFAR-10 类别不相交的 20K 张图片作为开放集噪声数据。类似地，CIFAR-10+SVHN 20K 表示从 SVHN 数据集中随机选取与 CIFAR-10 类别不相交的 20K 张图片作为开放集噪声数据。从表中可以看出，本文复现的实验结果大部分与原论文中的实验结果基本一致，证明了本文复现的有效性。针对 CIFAR-10+SVHN 20k 使用分类器的情况出现性能下降，将在后续部分进行分析。该结果表明，由于 CIFAR-10 与 CIFAR-100 的图像来源相似，也即数据域相似，导致模型难以区分封闭集噪声和开放集噪声，从而使得性能下降。而 SVHN 数据集的图像来源与 CIFAR-10 不同，也即数据域不同，因此模型能够更好地区分封闭集噪声和开放集噪声，从而使得性能提升。

表 2. 在混合噪声数据上的复现结果

| Input noise |     | CIFAR-10+CIFAR-100 20k | CIAFR10+SVHN 20k |
|-------------|-----|------------------------|------------------|
| Paper       | cls | 91.5                   | 93.3             |
|             | knn | 93.1                   | 93.9             |
| Ours        | cls | 91.3                   | 80.8             |
|             | knn | 92.8                   | 93.4             |

此外，本文还使用 t-SNE 降维技术，对 RRL 模型拟合过程的数据低维嵌入表征分布进行可视化展示，结果分别由图 2、图 3 以及图 4 展示。这些结果表明 RRL 能够在迭代过程中学习分布良好的样本特征表示，但是一些噪声仍然没有处于真实类别的簇中，这可能是由于噪声数据的分布与真实数据的分布相差较大，导致模型难以学习到噪声数据的真实表征。

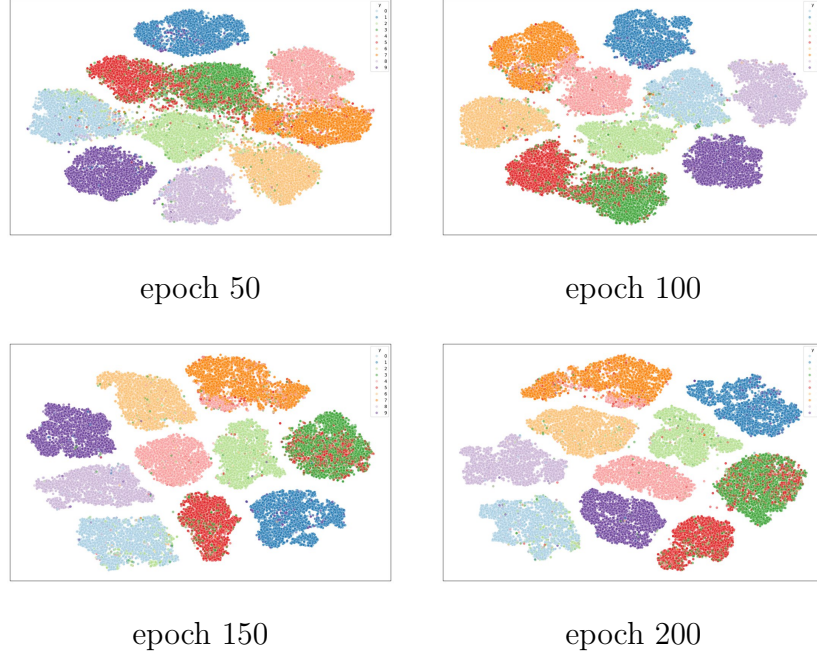


图 2. 在 CIFAR-10 数据集带有 40% 非对称噪声时，t-SNE 可视化的 RRL 模型在不同迭代步时样本低维表征分布结果。

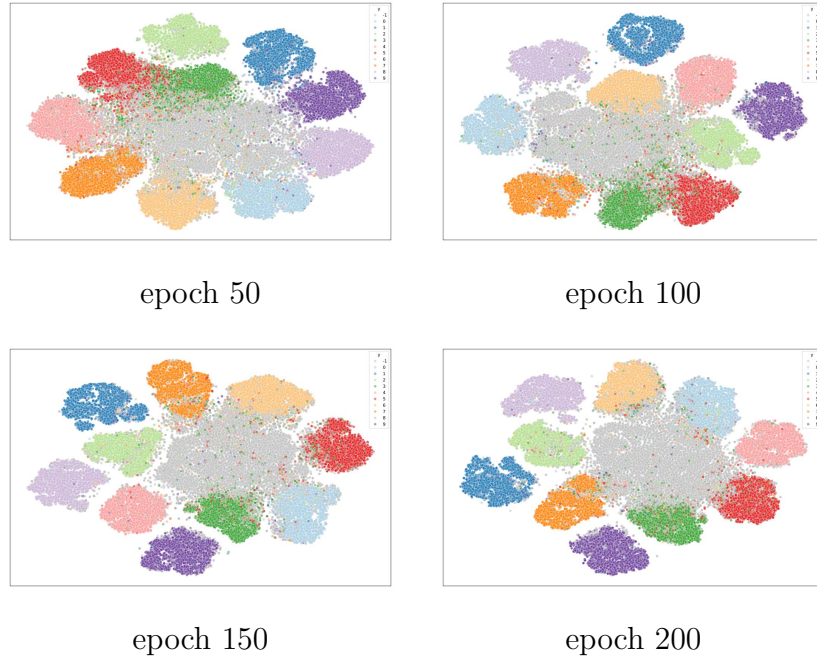


图 3. 在 CIFAR-10 数据集带有 50% 对称噪声且添加 20k 来自 CIFAR-100 的开放集噪声样本时，t-SNE 可视化的 RRL 模型在不同迭代步时样本低维表征分布结果。

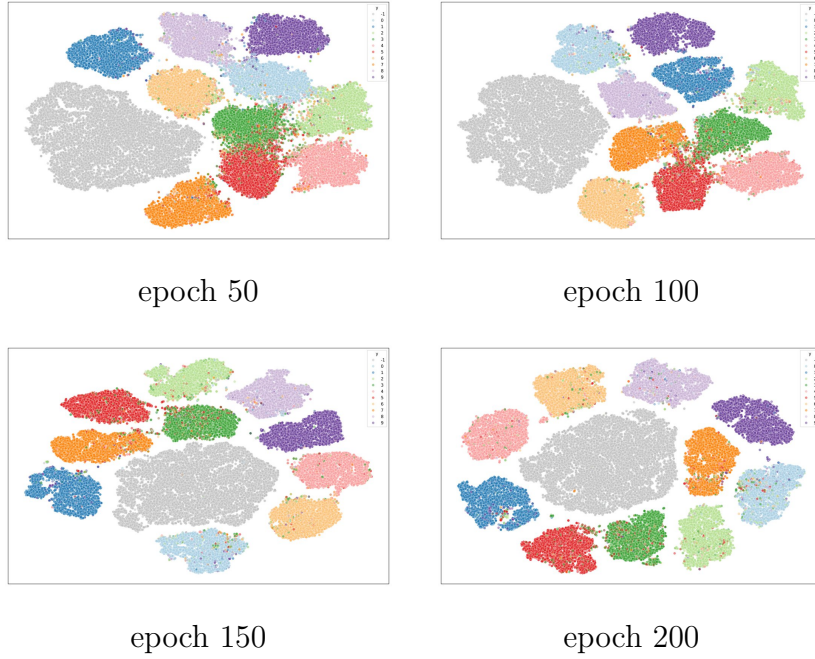


图 4. 在 CIFAR-10 数据集带有 50% 对称噪声且添加 20k 来自 SVHN 的开放集噪声样本时，t-SNE 可视化的 RRL 模型在不同迭代步时样本低维表征分布结果。

## 5.2 创新结果

**引入动态类别相关阈值。**RRL 模型在选择干净标签时利用固定的全局阈值，而模型对于不同类别的学习程度是不相同的，这可能会导致无法学习困难样本。从直觉上而言，对学习得好的类别，设定较高阈值，尽可能学习对模型性能有提升有用的；对于学习的不好的类，设置较低的阈值，使得模型有机会去学习该类别样本。因此根据 FlexMatch [22] 的思想，利用模型对类别预测的样本数量反映其对该类别的学习程度好坏。具体而言，统计模型预测相应的类别的训练样本数量，并除以最大的样本数量进行归一化。进而利用公式  $a(c) = (N_{norm}(c))/(2 - N_{norm}(c))$ ，获得模型对类别  $c$  的学习效果。最终，乘以基础全局阈值  $\tau$ ，获得类别相关的阈值  $\tau(c) = a(c) \cdot \tau$  进行伪标签过滤。在封闭集和开放集噪声上的结果如表 3 所示。表 3 的结果表明，虽然引入动态类别相关阈值策略在封闭集噪声上的性能没有得到提升，而在混合噪声上的 knn 分类性能得到一定的提升。这说明动态类别相关阈值策略能够有效地提升模型对开放集噪声的鲁棒性。

表 3. 动态类别相关阈值策略的实验结果

| Input noise |     | CIFAR-100 (Sym 50%) | +CIFAR-100 20k | +SVHN 20k |
|-------------|-----|---------------------|----------------|-----------|
| Ours        | cls | 74.8                | 91.3           | 80.8      |
|             | knn | 75.2                | 92.8           | 93.4      |
| Ours+flex   | cls | 73.0                | 91.4           | 93.7      |
|             | knn | 73.1                | 93.1           | 94.0      |

**完美过滤开放集噪声。**进一步地，为了探究开放集噪声在混合噪声情况下对模型的影响。在学习过程中挑选干净伪标记样本时，完全过滤开放集噪声，实验结果如表 4 所示。该结果

表明，在伪标记样本集中过滤开放集噪声对性能有一定的提升，而由于没有在对比学习中约束开放集噪声的影响，因此相比于完全封闭集噪声情况下的性能有所下降。

表 4. 过滤开放集噪声策略的实验结果

| Input noise   |     | CIFAR-10 (Sym50%) | +CIFAR-100 20k |
|---------------|-----|-------------------|----------------|
| Ours          | cls | 94.8              | 91.3           |
|               | knn | 94.7              | 92.8           |
| No open noise | cls | -                 | 93.5           |
|               | knn | -                 | 94.1           |

**添加 OVA-classifier。**由于 OVA-classifier 能够判别分布内 (ID) 和分布外 (OOD) 样本，因此使用 OVA-classifier 来近似上述过滤开放集噪声的策略。然而，由于相关损失，也即困难负类分类器采样策略和开放熵最小化策略，直接用于噪声学习会引起无法选择特定类别样本，进而导致原型计算出现 NaN 值，使得模型优化过程终止。

### 5.3 结果分析

**引入动态类别相关阈值。**针对封闭集噪声，该策略几乎不起作用，反而降低性能。图 5 是 CIFAR-100 带有 50% 对称噪声的准确率、选择的伪标记样本数量以及包含的噪声比例随着训练过程变化的结果。该结果表明动态阈值策略学习的伪标记样本含有较多噪声样本，导致性能降低。

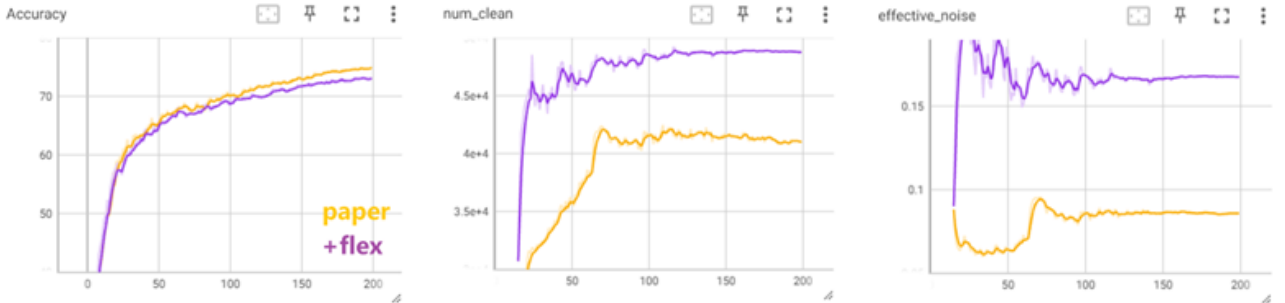


图 5. 在 CIFAR-100 包含 50% 对称噪声时，模型添加动态类别相关阈值后的准确率、选择的伪标记样本数量以及其包含的噪声比例随训练迭代变化的结果。

针对开放集噪声，动态类别相关阈值能够加速模型收敛速度，并能缓解模型过拟合问题。图 6 是 CIFAR-10 带有 50% 对称噪声并且添加 20k 来自 CIFAR-100 的开放集噪声样本的准确率、选择的伪标记样本数量以及包含的噪声比例随着训练过程变化的结果。这主要是因为添加该策略的模型能够较早地利用足够多的样本来优化监督损失，然而又因为伪标记样本噪声率较高，导致模型性能的提升被限制。

另一方面，动态类别相关阈值能够缓解模型过拟合。图 7 是 CIFAR-10 带有 50% 对称噪声并且添加 20k 来自 SVHN 的开放集噪声样本的准确率、选择的伪标记样本数量以及包含的噪声比例随着训练过程变化的结果。这主要是因为伪标签样本集的噪声率在逐步降低。

**完美过滤开放集噪声。**该策略对于混合噪声情况是有效的，但是依旧难以媲美仅在封闭集的性能，可能的原因是无监督损失（也即对比学习和重构损失）引入开放集噪声后会降低





图 6. 在 CIFAR-10 包含 50% 对称噪声并且添加 20k 来自 CIFAR-100 的开放集噪声样本时，模型添加动态类别相关阈值后的准确率、选择的伪标记样本数量以及其包含的噪声比例随训练迭代变化的结果。

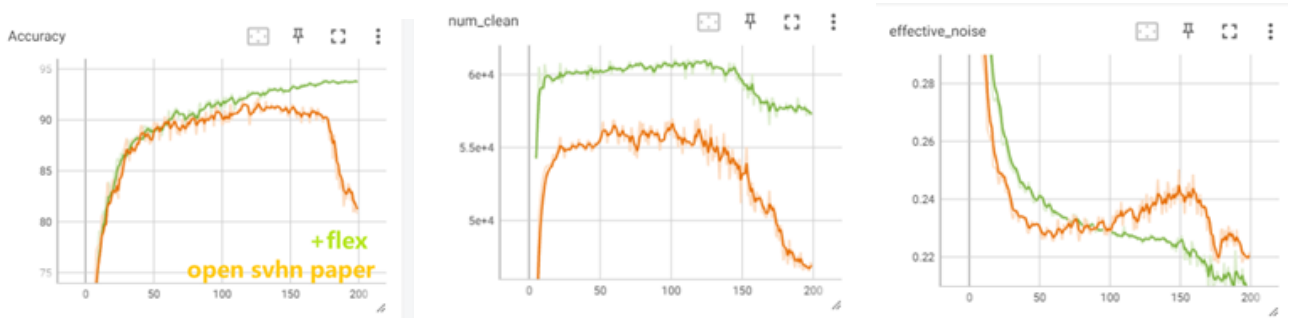


图 7. 在 CIFAR-10 包含 50% 对称噪声并且添加 20k 来自 SVHN 的开放集噪声样本时，模型添加动态类别相关阈值后的准确率、选择的伪标记样本数量以及其包含的噪声比例随训练迭代变化的结果。

模型性能。图 8 展示了 CIFAR-10 带有 50% 对称噪声且添加其它策略模型的准确率、选择的伪标记样本数量以及包含的噪声比例随着训练过程变化的结果。该结果表明，模型学习后期阶段会倾向于选择较少的干净样本。

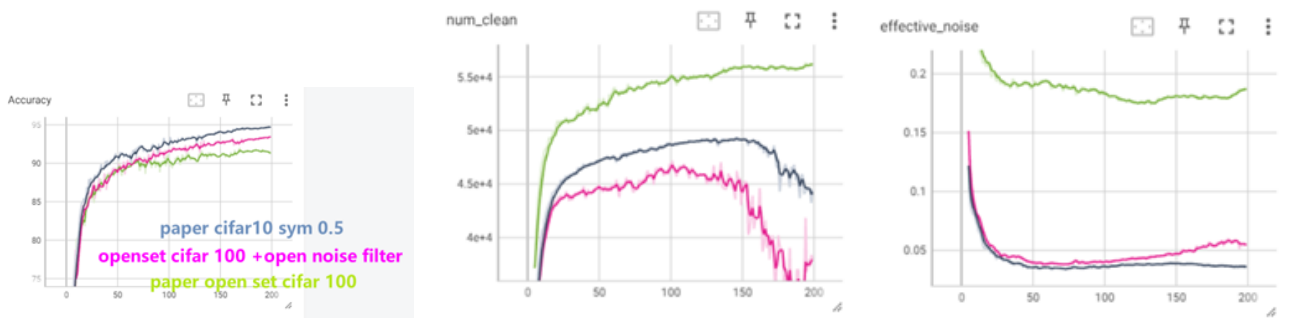


图 8. 在 CIFAR-10 带有 50% 对称噪声且添加其它策略时，模型的准确率、选择的伪标记样本数量以及其包含的噪声比例随训练迭代变化的结果。

阈值选择策略在高噪声情况，或者选择伪标签的阈值/outlier 概率不可信时，会导致无法选择某一类的伪标记样本，从而无法计算对应类原型。添加 OVA-classifier，以及原始 RRL 模型在 CIFAR100 带有对称噪声 80% 时，都会因为某一类原型计算为 NaN，导致模型学习过程终止。这一定程度说明基于局部邻域的 KNN 生成伪标签的策略在高噪声下难以起作用。

在 CIFAR10 数据集中，图 9 表明模型的学习后期阶段会倾向于选择较少的干净样本的

数量，但是选择的样本集合的噪声率没有下降，且性能还在逐步提高。这可能是因为引入一些困难学习样本，从而提高模型的泛化能力。

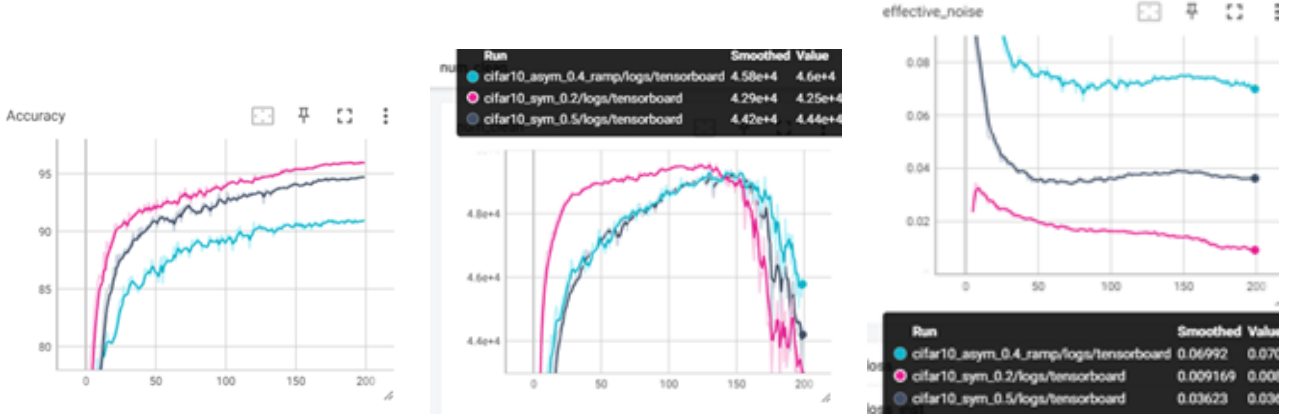


图 9. 在 CIFAR-10 带有不同的封闭集噪声时，模型的准确率、选择的伪标记样本数量以及其包含的噪声比例随训练迭代变化的结果。

图 10 表明在引入开放集噪声后，模型的训练后期阶段重构损失会逐渐上升。这说明模型后期阶段会拟合和重构开放集噪声样本，从而导致模型性能下降。



图 10. 在 CIFAR-10 带有不同的开放集噪声时，模型的重构损失随训练迭代变化的结果。

## 6 总结与展望

在数据集含有噪声的情况下，本文选择的 RRL 模型通过鲁棒表示学习，能够获得较好的模型性能。本文对 RRL 模型进行了复现，并在此基础上进行了创新，主要包括引入动态类别相关阈值、完美过滤开放集噪声、添加 OVA-classifier 以及结合开放熵最小化策略进一步优化的 OVA-classifier。实验结果表明，引入动态类别相关阈值能够缓解模型过拟合问题，提高模型对开放集噪声的鲁棒性；完美过滤开放集噪声能够提高模型对开放集噪声的鲁棒性，但是难以媲美仅在封闭集噪声的性能；添加 OVA-classifier 则在高噪声情况下难以优化，这可能是因为基于局部邻域的 KNN 生成伪标签的策略难以起作用。未来，将结合局部信息和全局信息，利用模型不同层输出的特征构造一个样本的多层次原型描述，从而缓解高噪声带来的情况，并且考虑进一步利用开放集噪声的信息，提高模型的鲁棒性。

## 参考文献

- [1] Ehsan Amid, Manfred K Warmuth, and Sriram Srinivasan. Two-temperature logistic regression based on the tsallis divergence. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2388–2396. PMLR, 2019.
- [2] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33:9912–9924, 2020.
- [3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.
- [4] Aritra Ghosh, Himanshu Kumar, and P Shanti Sastry. Robust loss functions under label noise for deep neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.
- [5] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.
- [6] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31, 2018.
- [7] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.
- [8] Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *International conference on machine learning*, pages 2304–2313. PMLR, 2018.
- [9] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.
- [10] Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv preprint arXiv:2002.07394*, 2020.
- [11] Junnan Li, Caiming Xiong, and Steven CH Hoi. Learning from noisy data with robust representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9485–9494, 2021.

- [12] Junnan Li, Pan Zhou, Caiming Xiong, and Steven Hoi. Prototypical contrastive learning of unsupervised representations. In *International Conference on Learning Representations*, 2020.
- [13] Yueming Lyu and Ivor W Tsang. Curriculum loss: Robust learning and generalization against label corruption. *arXiv preprint arXiv:1905.10045*, 2019.
- [14] Xingjun Ma, Hanxun Huang, Yisen Wang, Simone Romano, Sarah Erfani, and James Bailey. Normalized loss functions for deep learning with noisy labels. In *International conference on machine learning*, pages 6543–6553. PMLR, 2020.
- [15] Naresh Manwani and PS Sastry. Noise tolerance under risk minimization. *IEEE transactions on cybernetics*, 43(3):1146–1151, 2013.
- [16] Kuniaki Saito and Kate Saenko. Ovanet: One-vs-all network for universal domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9000–9009, 2021.
- [17] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 776–794. Springer, 2020.
- [18] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, and Léon Bottou. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12), 2010.
- [19] Pengxiang Wu, Songzhu Zheng, Mayank Goswami, Dimitris Metaxas, and Chao Chen. A topological filter for learning with label noise. *Advances in neural information processing systems*, 33:21382–21393, 2020.
- [20] Zhi-Fan Wu, Tong Wei, Jianwen Jiang, Chaojie Mao, Mingqian Tang, and Yu-Feng Li. Ngc: A unified framework for learning with open-world noisy data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 62–71, 2021.
- [21] Xingrui Yu, Bo Han, Jiangchao Yao, Gang Niu, Ivor Tsang, and Masashi Sugiyama. How does disagreement help generalization against label corruption? In *International Conference on Machine Learning*, pages 7164–7173. PMLR, 2019.
- [22] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34:18408–18419, 2021.
- [23] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31, 2018.