

# 深度长尾学习：综述

Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, *Fellow, IEEE* and Jiashi Feng

**Abstract**—深度长尾学习是视觉识别中最具挑战性的问题之一，旨在从大量遵循长尾分布的图像中训练性能良好的深度模型。在过去的十年中，深度学习已经成为学习高质量图像表示的强大识别模型，并在通用视觉识别上取得了显著突破。然而，长尾类别不平衡是实际视觉识别任务中的一个常见问题，它常常限制了基于深度网络的识别模型在实际应用中的实用性，因为它们很容易偏向于多数类，而在尾部类上的性能很差。为了解决这个问题，近年来进行了大量的研究，在深度长尾学习领域取得了可喜的进展。考虑到这一领域的快速发展，本文旨在综述深度长尾学习的最新进展。具体而言，本文将现有的深度长尾学习研究分为三大类（即类别重平衡、信息增强和模块改进），并按照这种分类法详细回顾这些方法。然后，本文对几种最先进的方法进行了实证分析，通过一个新提出的评估指标（即相对准确度）来评估这些方法在多大程度上解决了类别不平衡的问题。最后，本文强调了深度长尾学习的重要应用，并确定了未来研究的几个有希望的方向。

**Index Terms**—长尾学习, 深度学习, 不平衡学习, 卷积神经网络.

## 1 介绍

深度学习允许由多个处理层组成的计算模型学习具有多个抽象层次的数据表示 [1], [2]，在计算机视觉方面取得了惊人的进步 [3], [4], [5], [6], [7], [8]。深度学习崛起的关键因素是大规模的公开数据集、GPU 的出现以及深度网络结构的进步 [9]。由于从高质量数据中学习到的强大的表示能力，深度神经网络已成功应用于许多视觉识别任务，包括图像分类 [6], [10]、目标检测 [7], [11] 和语义分割 [8], [12]。

在实际应用中，训练样本通常呈现长尾类别分布，即少部分类别具有大量样本，而大部分其他类别仅有少量样本 [13], [14], [15], [16]。这种训练样本数量的类别不平衡给基于深度网络的识别模型的训练带来了巨大挑战。如图1所示，经过训练的模型很容易偏向于具有大量训练数据的头部类别，从而导致在数据有限的尾部类别上的性能较差 [17], [18], [19]。因此，由经验风险最小化 [20] 的通常做法训练的深层模型无法处理实际应用中的长尾类别不平衡问题，例如人脸识别 [21], [22]、物种分类 [23], [24]、医学图像诊断 [25]、城市场景理解 [26] 和无人机检测 [27]。

为了解决长尾类别不平衡问题，近年来进行了大量的深度长尾学习研究 [15], [16], [28], [29], [30]。尽管这一领域发展迅速，但目前还没有系统的研究来回顾和讨论现有的进展。为了填补这一空白，本文旨在为2021年年中之前进行的最新长尾学习研究提供全面综述。

如图2所示，我们根据现有方法的主要技术贡献将其分为三大类，即类别重平衡、信息增强和模块改进；这些类别可进一步分为九个子类别：重采样、代价敏感学习、logit调整、迁移学习、数据增强、表征学习、分类器设计、解耦训练和集成学习。根据这种分类法，本文对现有的方法进行了全面的回顾，并通过使用一种新的评估指标（即相对准确度）来评估几种最先进的方法处理类不平衡的能力，对其进行了实证分析。本文通过介绍几个深度长尾学习的实际应用场景并确定这一领域未来可以探索的几个有希望的研究方向来作为结束。

这篇综述的核心贡献可以总结如下。

- Y. Zhang, B. Hooi, 来自新加坡国立大学计算机学院, 新加坡. 邮箱: yifan.zhang@u.nus.edu, dcsbhk@nus.edu.sg.
- B. Kang, S. Yan, J. Feng, 来自SEA AI 实验室, 新加坡. 邮箱: bingyikang@gmail.com, yansc@sea.com, fengjs@sea.com.
- 翻译作者: 焦文沛, 来自南开大学人工智能学院, 天津, 中国. 邮箱: jwp@mail.nankai.edu.cn.

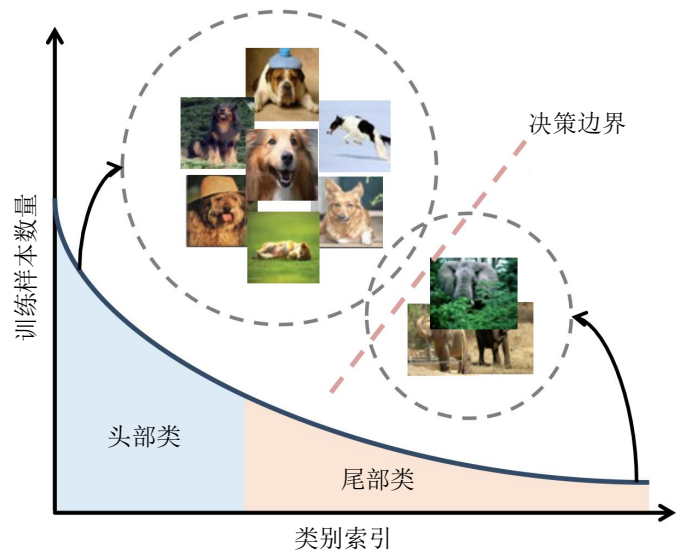


Fig. 1: 长尾数据集的标签分布（例如，iNaturalist物种数据集 [23]，有8000多个类）。在这些样本上学习到的头部类别的特征空间通常大于尾部类别，而决策边界通常偏向于优势类。

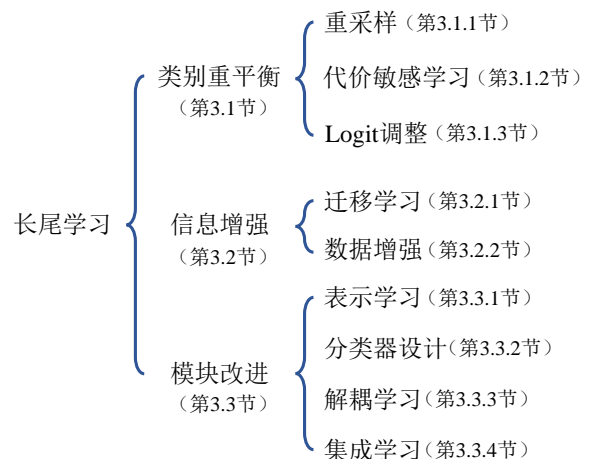


Fig. 2: 现有的深度长尾学习方法分类。

- 据我们所知，这是第一次对深度长尾学习进行全面综述，这将使研究人员和社区对深度神经网络的长尾视觉学习有更好的理解。
- 本文对高级长尾学习研究进行了深入的回顾，并通过一个新的相对准确度指标评估了最先进的方法在多大程度上解决了长尾类别不平衡。
- 本文确定了四个潜在的研究方向以及八个新的深度长尾学习任务设置以供未来研究。

本文的其余部分将组织如下：第2节介绍了问题定义，并介绍了广泛使用的数据集、指标和骨干网络。第3节全面回顾了先进的长尾学习方法，第4节根据新的评估指标实证分析了几种最先进的方法。第5节介绍了深度长尾学习的应用场景，第6节确定了未来的研究方向。在第7节中对全文进行了总结。

## 2 问题定义和基本概念

### 2.1 问题定义

深度长尾学习寻求从具有长尾类别分布的训练数据集中学习深度神经网络模型，其中一小部分类别具有大量样本，其余类别仅有少量样本（参考图1）。设有长尾训练集 $\{x_i, y_i\}_{i=1}^n$ ，其中，每个样本 $x_i$ 都有相应的类别标签 $y_i$ 。训练集中 $K$ 个类别的样本总数为 $n = \sum_{k=1}^K n_k$ ，其中 $n_k$ 表示类别 $k$ 的样本数量；令 $\pi$ 表示标签频率向量，其中， $\pi_k = n_k/n$ 表示类别 $k$ 的标签频率。在不丧失一般性的情况下，长尾学习中的一个常见假设 [31], [32]是，类别是按照样本数量降序排序的（即：如果 $i_1 < i_2$ ，则 $n_{i_1} \geq n_{i_2}$ ，且 $n_1 \gg n_K$ ），则不平衡率定义为 $n_1/n_K$ 。

因为两个困难的存在使得这项任务具有挑战性：（1）类别间的数据数量不平衡使得深度模型偏向于头部类别，而在尾部类别上表现不佳；（2）尾部类别样本的缺乏使得训练模型进行尾部类别分类更具挑战性。这种基础任务可能发生在各种视觉识别任务中，例如图像分类 [15], [32]、检测 [19], [33]和分割 [26], [34], [35]。

TABLE 1: 长尾数据集统计。“Cls.”表示图像分类；“Det.”表示目标检测；“Seg.”表示实例分割。

任务	数据集	# 类别	# 训练数据	# 测试数据
Cls.	ImageNet-LT [15]	1,000	115,846	50,000
	CIFAR100-LT [18]	100	50,000	10,000
	Places-LT [15]	365	62,500	36,500
	iNaturalist 2018 [23]	8,142	437,513	24,426
Det./Seg.	LVIS v0.5 [36]	1,230	57,000	20,000
	LVIS v1 [36]	1,203	100,000	19,800
多标签Cls.	VOC-LT [37]	20	1,142	4,952
	COCO-LT [37]	80	1,909	5,000
视频Cls.	VideoLT [38]	1,004	179,352	51,244

### 2.2 数据集

近年来发布了各种用于长尾学习的视觉数据集，这些数据集在任务、类别数量和样本数量上有所不同。在表1中，我们总结了在深度长尾学习社区中广泛使用的九个视觉数据集。

在长尾图像分类中，有四个基准数据集：ImageNet-LT [15]、CIFAR100-LT [18]、Places-LT [15]和iNaturalist 2018 [23]。前三个分别遵循帕累托分布从ImageNet [39]、CIFAR100 [40]和Places365 [41]采样得到，而iNaturalist是一个真实的长尾数据集。ImageNet-LT、Places-LT和iNaturalist的不平衡率分别为256、996和500；CIFAR100-LT有三种不同的不平衡率{10, 50, 100}的版本。

在长尾目标检测和实例分割中，LVIS [36]是广泛使用的基准，它提供了精确边界框和掩码标注。在多标签图像分类中，基准是VOC-LT [37]和COCO-LT [37]，分别是PASCAL VOC 2012 [42]和COCO [43]中采样得到的。最近发布了一个大规模的“未裁剪”视频数据集VideoLT [38]，用于长尾视频识别。

### 2.3 评估指标

在长尾学习中，通常会报告所有类别的总体表现以及头部、中部、尾部类别的表现。所使用的评估指标在不同的任务中有所不同。例如，Top-1精度（或错误率）是长尾图像分类中广泛使用的指标，而平均精度均值（mAP）[44]用于长尾目标检测和实例分割。此外，mAP还被用作长尾多标签图像分类的一个指标，而视频识别同时应用Top-1精度和mAP进行评估。

### 2.4 主流骨干网络

现有的长尾学习方法是基于通用骨干网络开发的，这些骨干网络在不同的数据集上有所不同。ImageNet-LT上的常见做法是不同深度的ResNet [10]和ResNeXt [45]，其中ResNet-50和ResNeXt-50是最常见的。此外，ResNet-32通常用于CIFAR100-LT；ResNet-50通常用于iNaturalist 2018；ResNet-152通常在ImageNet上预训练后再用于Places-LT。对于LVIS数据集，广泛使用的网络结构是基于ResNet-50的带特征金字塔网络（FPN）[46]的Mask R-CNN [47]或Faster R-CNN [7]。在多标签分类中，预训练的ResNet-50是VOC-LT和COCO-LT的通常选择，在视频识别中，预训练的ResNet-50和ResNet-101都被应用于VideoLT。除了这些通用的骨干网络之外，最近的方法还探索了多网络分支（即：multi-expert），以改进骨干网络 [30], [48]。

### 2.5 长尾学习挑战

长尾学习中最受欢迎的挑战赛事包括iNat [23]和LVIS [36]。

**iNat挑战：**iNaturalist (iNat) 挑战是CVPR的大规模细粒度物种分类竞赛。这项挑战旨在推动具有大量类别（包括植物和动物在内）的真实图像的自动图像分类技术的发展。与其他分类挑战（如ImageNet大规模视觉识别挑战）相比，该挑战中的iNaturalist数据集 [23] 呈现长尾类别分布，这促进了图像分类的进步。

**LVIS挑战：**大规模细粒度词汇级标记数据集（LVIS）[36]是一个具有超过1000个目标类别的高质量实例分割数据集。由于类别是长尾分布的，LVIS在ICCV/ECCV上提出了一个新的实例分割和对象检测挑战，这与著名的COCO挑战不同。

### 2.6 与其他任务的关系

本文随后简要讨论了长尾学习与类别不平衡学习、少样本学习和域外泛化的区别。这些关系在深度学习和非深度学习之间是一致的。

类别不平衡学习 [5], [49]旨在从类别不平衡样本中训练模型。总的来说，长尾学习可以被视为类别不平衡学习中更具体、更具挑战性的子任务。相比之下，在类别不平衡学习中，类别的数量可能非常少（例如2个），劣势类别的数据量不一定很小；而在长尾学习中有非常多类别，且尾部类样本往往非常稀少。

少样本学习 [50], [51], [52], [53] 旨在从有限数量（例如：1或5）的样本中训练模型。相比之下，由于长尾学习



中尾部类的样本数量通常非常少，因此少样本学习可以被视为长尾学习的一个子任务。

域外泛化 [54], [55]是指训练分布与未知测试分布不一致的一类任务。这种不一致性包括不一致的数据边界分布（例如，域自适应 [56], [57], [58], [59], [60], [61]和域泛化 [62], [63]）、不一致的类别分布（例如，长尾学习 [15], [28], [32]、开放集学习 [64], [65]），以及前两种情况的组合。从这个角度来看，长尾学习可以被视为域外泛化中的一项特定任务。

### 3 经典方法

如图2所示，我们将现有的深度长尾学习方法分为三大类，包括类别重平衡、信息增强和模块改进。更具体地说，类别重平衡包括三个子类别：重采样、代价敏感学习（CSL）和logit调整（LA）。信息增强包括迁移学习（TL）和数据增强（Aug）。模块改进包括表征学习（RL）、分类器设计（CD）、解耦训练（DT）和集成学习（Ensemble）。根据这种分类法，我们在表2中对现有的深度长尾学习方法进行了分类，并将对其进行如下详细审查。

#### 3.1 类别重平衡

类别重平衡是长尾学习的主流范式，旨在平衡模型训练中不同类别的训练样本数。我们从基于重采样的方法开始，然后是代价敏感学习和logit调整。

##### 3.1.1 重采样

在过去几十年中，重采样是解决类别不平衡问题最广泛使用的方法之一 [32], [34], [66], [67], [68], [69], [70]。重采样的常见做法是随机过采样（ROS）和随机欠抽样（RUS）。为了使类别重新平衡，ROS随机重复尾部类的样本，而RUS随机丢弃头部类的样本。然而，当类别间数量差别极为严重时，ROS倾向于对尾部类过拟合，而RUS倾向于降低头部类的模型性能。最近的长尾学习研究没有使用随机重采样，而是发展了各种各样的采样方法，包括类别平衡重采样和方案导向的采样。

类别平衡重采样。

我们从解耦（Decoupling） [32]开始，该研究对长尾识别中表征学习的各种采样策略进行了实证评估。具体来说，采样策略包括实例平衡采样、类别平衡采样、平方根采样和渐进平衡采样。在实例平衡采样中，每个样本被采样的概率相等，而在类别平衡采样中，每个类被选中的概率相等。此外，平方根采样 [71]是实例平衡采样的一种变体，其中每个类别的采样概率与相应类别中样本大小的平方根相关。渐进平衡采样 [32]在实例和类平衡采样之间进行渐进插值。

简单校准（SimCal） [34]提出了一种新的双层类别平衡采样策略来处理长尾实例分割。具体而言，双层采样策略将图像级重采样和实例级重采样相结合，以缓解实例分割中的类不平衡。

动态课程学习（DCL） [72]开发了一种新的课程策略，通过动态采样数据来使得类别重新平衡。具体来说，随着训练的进行，从一个类中采样的实例越多，该类随后被采样的概率越低。遵循这一思想，DCL首先进行随机采样以学习一般表示，然后根据课程策略采样更多的尾部类实例来处理长尾类别不平衡问题。

平衡的meta-softmax [73]开发了一种基于元学习的采样方法，用于估计长尾学习不同类别的最佳采样率。具体而言，所提出的元学习方法是一种双层优化策略，通过在一个平衡元验证集上优化模型分类性能来学习最佳样本分布参数。

特征增强和采样自适应（FASA） [74]建议使用平衡元验证集（作为度量）上的模型分类损失来调整不同类别的特征采样率，以便可以对表示不足的尾部类别进行更多采样。

具有分类均衡的长尾目标检测器（LOCE） [33]建议使用平均分类预测分数（即：运行预测概率）来监控不同类别上的模型训练，并指导记忆增强特征采样以提高尾部类别的性能。

VideoLT [38]旨在解决长尾视频识别问题，引入了一种新的帧堆栈（FrameStack）方法，该方法进行帧级采样以重新平衡类别分布。具体来说，FrameStack在训练期间根据运行模型的性能动态调整不同类的采样率，以便可以从尾部类（通常运行性能较低）中采样更多的视频帧，从头部类中采样更少的帧。

方案导向的采样旨在为长尾学习提供一些特定的学习方案，如度量学习和集成学习。例如，大边距局部嵌入（LMLE） [75]开发了一种新的五元组采样方案用于度量学习，以学习保持簇间和类间边距的高质量特征。与三元组损失（triplet loss） [76]不同的是，LMLE提出了一个五元组采样器来采样四个对比对，包括一个阳性对和三个阴性对。阳性对是距离最远的簇内样本，而阴性对包括来自同一类的两个簇间样本（一个是同一簇内距离最近的，一个是最远的）和最接近的类间样本。随后，LMLE引入了五元组损失，以鼓励采样的五元组遵循特定的距离顺序。这样，学习到的表示不仅类内间距较小，而且类间间距较大。此外，五元组损失中的每个数据批次包含来自不同类别的相同数量的样本，用于类别重平衡。

分区水库采样（PRS） [77]提出了一种基于重放的采样方法来处理连续长尾学习。一个关键的挑战是重放内存无法考虑类不平衡的问题，因为没有关于未来输入的信息可用。为了解决这个问题，PRS开发了一种在线内存维护算法，该算法动态维护来自不同类的样本的运行统计信息。PRS可以根据运行统计信息动态调整不同类别的内存大小和样本输入/输出操作方案。

双边分支网络（BBN） [48]开发了两个网络分支（即传统学习分支和重平衡分支），基于一个新的双边采样策略来处理类别不平衡。具体而言，在传统分支，BBN应用均匀采样来模拟原始的长尾训练分布；同时，在重平衡分支，BBN应用反向采样器以采样更多的尾部类样本，从而提高尾部类性能。最终预测是两个网络分支的加权和。之后，长尾多标签视觉识别（LTML） [78]扩展了双边分支网络，以解决长尾多标签分类问题。几何结构迁移（GIST） [79]也探讨了这种从头部类到尾部类知识迁移的双边采样策略。

除了双边分支的抽样外，平衡组softmax（BAGS） [80]建议根据每个类中的样本数量将类划分为几个平衡组，其中每个组的类具有相似数量的训练数据。然后，BAGS使用不同的样本组来训练不同的分类头，以便它们在具有相似数量训练数据的类上执行softmax操作，从而避免由于不平衡而产生严重偏差的分类器。之后，学习分割尾部（LST） [81]还将训练样本划分为几个平衡的子集，并基于类别增量学习对每个子集进行处理。为了解决类别增量学习中的灾难性遗忘问题，LST开发了一种类别平衡的数据应答/采样策略，该策略为知识蒸馏保留了一个相对平衡的样本集。

联合互补专家（ACE） [82]没有将样本划分为几个平衡组，而是将样本划分为几个技能多样的子集，其中一个子集包含所有类，一个子集包含中间类和尾部类，另一个子集仅包含尾部类。基于这些子集，ACE培训不同的专家，使其具备集成学习的特定和互补技能。

##### 3.1.2 代价敏感学习

代价敏感学习旨在通过调整训练期间不同类别的损失值来对类别进行重新平衡 [119], [120], [121], [122], [123], [124],

TABLE 2: 2021年年中之前在顶级会议上发布的现有深度长尾学习方法总结。主要有三类：类别重平衡、信息增强和模块改进。在本表中，“CSL”表示代价敏感学习；“LA”表示logit调整；“TL”代表迁移学习；“Aug”表示数据增强；“RL”表示表征学习；“CD”表示分类器设计，旨在为长尾识别设计新的分类器或预测方案；“DT”表示解耦训练，其中特征提取器和分类器分别进行训练；“Ensemble”表示基于集成学习的方法。我们还开源了我们收集的长尾学习资源：<https://github.com/Vanint/Awesome-LongTailed-Learning>

Method	Publication	Year	Class Re-balancing			Augmentation		Module Improvement			
			Re-sampling	CSL	LA	TL	Aug	RL	CD	DT	Ensemble
LMLE [75]	CVPR	2016	✓					✓			
HFL [83]	CVPR	2016						✓			
Focal loss [84]	ICCV	2017		✓							
Range loss [21]	ICCV	2017						✓			
CRL [85]	ICCV	2017						✓			
MetaModelNet [86]	NeurIPS	2017				✓					
DSTL [87]	CVPR	2018				✓					
CB [16]	CVPR	2019		✓							
Bayesian estimate [88]	CVPR	2019		✓							
FTL [89]	CVPR	2019				✓	✓				
Unequal-training [90]	CVPR	2019						✓			
OLTR [15]	CVPR	2019						✓			
DCL [72]	ICCV	2019	✓								
Meta-Weight-Net [91]	NeurIPS	2019		✓							
LDAM [18]	NeurIPS	2019		✓							
Decoupling [32]	ICLR	2020	✓	✓				✓	✓	✓	
LST [81]	CVPR	2020	✓			✓					
BBN [48]	CVPR	2020	✓								✓
BAGS [80]	CVPR	2020	✓								✓
Domain adaptation [28]	CVPR	2020		✓							
Equalization loss (ESQL) [19]	CVPR	2020		✓							
DBM [22]	CVPR	2020		✓							
M2m [92]	CVPR	2020				✓	✓				
LEAP [93]	CVPR	2020				✓	✓	✓			
IEM [94]	CVPR	2020						✓			
SimCal [34]	ECCV	2020	✓							✓	✓
PRS [77]	ECCV	2020	✓								
Distribution-balanced loss [37]	ECCV	2020		✓							
OFA [95]	ECCV	2020				✓	✓			✓	
LFME [96]	ECCV	2020				✓					✓
Deep-RTC [97]	ECCV	2020							✓		
Balanced Meta-Softmax [73]	NeurIPS	2020	✓	✓							
UNO-IC [98]	NeurIPS	2020			✓						
De-confound-TDE [99]	NeurIPS	2020			✓				✓		
SSP [100]	NeurIPS	2020				✓		✓			
Logit adjustment [14]	ICLR	2021			✓						
RIDE [17]	ICLR	2021				✓					✓
KCL [13]	ICLR	2021						✓		✓	
LTML [78]	CVPR	2021	✓								✓
Equalization loss v2 [101]	CVPR	2021		✓							
Seesaw loss [102]	CVPR	2021		✓							
ACSL [103]	CVPR	2021		✓							
PML [104]	CVPR	2021		✓							
LADE [31]	CVPR	2021		✓	✓						
RoBal [105]	CVPR	2021		✓	✓				✓		
DisAlign [29]	CVPR	2021		✓	✓					✓	
MiSLAS [106]	CVPR	2021		✓			✓			✓	
CReST [107]	CVPR	2021				✓					
Conceptual 12M [108]	CVPR	2021				✓					
RSG [109]	CVPR	2021				✓					
MetaSAug [110]	CVPR	2021					✓				
Hybrid [111]	CVPR	2021						✓			
Unsupervised discovery [35]	CVPR	2021						✓			
VideoLT [38]	ICCV	2021	✓								
LOCE [33]	ICCV	2021	✓	✓							
GIST [79]	ICCV	2021	✓			✓			✓		
FASA [74]	ICCV	2021	✓				✓				
ACE [82]	ICCV	2021	✓								✓
IB [112]	ICCV	2021		✓							
DARS [26]	ICCV	2021				✓					
SSD [113]	ICCV	2021				✓					
DiVE [114]	ICCV	2021				✓					
MosaicOS [115]	ICCV	2021				✓					
PaCo [116]	ICCV	2021						✓			
DRO-LT [117]	ICCV	2021						✓			
DT2 [118]	ICCV	2021								✓	

TABLE 3: 损失函数总结。在该表中,  $z$ 和 $p$ 表示样本 $x$ 的预测logits和softmax概率, 其中 $z_y$ 和 $p_y$ 对应于类别 $y$ 。此外,  $n$ 表示训练数据的总数, 其中 $n_y$ 是类别 $y$ 的样本数。此外,  $\pi$ 表示采样频率的向量, 其中 $\pi_y = n_y/n$ 表示类别 $y$ 的标签频率。若没有给出更具体的值, 类别权重由 $w$ 表示, 类别边距由 $\Delta$ 表示。损失相关参数包括 $\gamma$ 。

损失	表达式
Softmax 损失	$\mathcal{L}_{ce} = -\log(p_y)$
加权Softmax 损失	$\mathcal{L}_{wce} = -\frac{1}{\pi_y} \log(p_y)$
焦点损失 [84]	$\mathcal{L}_{fl} = -(1 - p_y)^\gamma \log(p_y)$
类别平衡损失 [16]	$\mathcal{L}_{cb} = -\frac{1-\gamma}{1-\gamma} \log(p_y)$
平衡Softmax 损失 [73]	$\mathcal{L}_{bs} = -\log\left(\frac{\pi_y \exp(z_y)}{\sum_j \pi_j \exp(z_j)}\right)$
均衡损失 [19]	$\mathcal{L}_{eq} = -\log\left(\frac{\exp(z_y)}{\sum_j w_j \exp(z_j)}\right)$
LDAM 损失 [18]	$\mathcal{L}_{ldam} = -\log\left(\frac{\exp(z_y - \Delta_y)}{\sum_j \exp(z_j - \Delta_j)}\right)$

[125]。最近的研究开发了各种对代价敏感的长尾学习方法来处理类别不平衡, 包括类别级重加权和类别级重边距调整 (re-margining)。

类别级重加权。最直观的方法是直接使用训练样本的标签频率进行损失重新加权, 即加权softmax损失 (参考表3)。基于样本影响 [112]或模型预测和平衡参考分布之间的分布对齐 [29], 通过调整标签频率对损失权重的影响, 可以进一步改善这种损失。除了损失值重新加权外, 平衡softmax [73]还建议在训练期间使用标签频率来调整模型预测, 以便通过先验知识来缓解类别不平衡的偏差。随后, LADE [31]引入了标签分布分离损失, 将学习模型从长尾训练分布中分离出来, 然后在测试标签频率可用的情况下, 模型可以适应任意测试类分布。

类别平衡损失 (CB) [16]没有使用标签频率, 而是引入了有效样本数的新概念, 以近似不同类别的预期样本数。这里, 有效样本数是训练样本数的指数函数。根据这一概念, CB损失将类别平衡加权参数设置为与各类别的有效样本数成反比, 以解决类别不平衡问题 (参考表3)。

焦点损失 (focal loss) [84]探索了重加权的类别预测困难。具体而言, 焦点损失的灵感来自于这样一个观察结果, 即类别不平衡通常会增加尾部类的预测难度, 使得尾部类的预测概率将低于头部类的预测概率。基于这一观察, 焦点损失使用预测概率来对类别进行反向重加权 (参考表3), 因此, 它可以为较难的尾部类分配较高的权重, 而为较容易的头部类分配较低的权重。

除了使用预定义的权重函数外, 还可以从数据中学习类别权重。在元权重网络 [91]中, 在平衡验证集的指导下, 由一层MLP来近似权重函数, 通过更新这个MLP来拟合长尾分布, 从而可以学习到一个在均匀测试集上性能良好的模型。此外, 分布对齐 (DisAlign) [29]开发了一种自适应校准函数来校准模型分类器。这个校准函数通过最小化调整后的预测分布和给定的平衡参考分布之间的KL散度来进行自适应学习。

长尾学习中的另一个问题是负梯度过度抑制 [19], [126]。也就是说, 在softmax或sigmoid交叉熵中, 一个类的每个正样本可以被视为其他类的负样本, 导致尾部类收到了更多的抑制梯度。为了解决这个问题, 分布平衡损失 [37]通过一种新的负容忍正则化来缓解梯度过度抑制。同时, 它还评估了每个类别的预期采样频率和实际采样频率之间的差距, 然后使用这两个频率的比值来对不同类别的损失值进行重加权。

均衡损失 [19]在尾部类样本作为大量头部类样本的负样本时, 直接降低尾部类样本的损失值权重。均衡损失v2 [101]进一步扩展了均衡损失 [19], 将多类别检测问题建模为一组独立的子任务, 其中每个子任务专注于一类。更具体地说, 均衡损失v2引入了一种新的梯度引导重新加权机制, 用

于动态增加正梯度的权重, 并减少负梯度的权重, 以便对每个子任务进行模型训练。

跷跷板损失 [102]通过两个重新加权因子, 即缓解因子和补偿因子, 来重新平衡每个类别的正梯度和负梯度。为了解决梯度过度抑制问题, 缓解因子基于不同类别之间累积样本数的动态比率, 在训练期间减轻对尾部类的惩罚。同时, 如果观察到一个假阳性样本, 补偿因子会将惩罚加权到相应的类别, 以改进模型判别。

自适应类别抑制损失 (ACSL) [103]使用输出置信度来决定是否抑制负标签的梯度。具体而言, 如果负标签的预测概率大于预设的阈值, 则将该类的权重设置为1 来提高模型的判别力。否则, 将权重设置为0 以避免负梯度过度抑制。

类别级重边距调整旨在通过调整不同类别学习到的特征和模型分类器之间的最小边距 (即距离) 来处理类别不平衡。例如, 标签分布感知边距调整 (LDAM) [18]通过执行基于标签频率的类别依赖的边距, 并鼓励尾部类别具有更大的边距, 扩展了现有的边距调整损失 [127], [128]。然而, 仅仅使用LDAM损失在经验上不足以处理类别不平衡。因此, LDAM进一步引入了延迟重平衡优化调度, 该调度在学习LDAM损失一段时间后, 通过类平衡的方式重加权LDAM损失来重新平衡类别。

贝叶斯估计 [88]发现, 类别预测的不确定性与训练标签频率成反比, 即尾部类更不确定。受此启发, 贝叶斯估计 [88]建议使用估计的类别级不确定性来对损失进行重边距调整, 以使具有更高类别不确定性的尾部类得到更高的损失值, 从而在特征和分类器之间具有更大的边距。

域平衡 [22]研究了一个长尾域问题, 其中少量域 (包含多个类) 经常出现, 而其他域较少出现。为了解决这一问题, 这篇工作引入了一种新的基于特征的类型紧性的域频率指示器, 并使用该指示器对尾部域的特征空间进行重边距调整。

LOCE [33]使用平均分类预测分数来监控不同类别的学习状态, 并将其应用于指导类别级边距调整, 以提高尾部类的性能。

渐进边距损失 (PML) [104]通过两个边距项 (顺序边距和变化边距) 调整长尾学习的类别边距。顺序边距旨在提取识别特征并保持时间顺序关系。在长尾训练样本中, 变化边距试图逐步抑制头部类以处理类别不平衡。

RoBal [105]认为, 现有的重边距调整方法鼓励尾部类获得更大边距, 但可能会降低头部类的特征学习。因此, RoBal执行了一个额外的边距项, 以扩大头部类的特征边距。

### 3.1.3 Logit调整

Logit调整, 即基于标签频率对模型的logit进行后期移动, 是在类别不平衡问题中获得较大的类间相对边距的经典思路 [14], [129]。最近, 一项研究 [14]全面分析了长尾识别中的logit调整方法, 并从理论上表明logit调整是费雪一致<sup>1</sup>的, 以最小化每个类别的平均错误。根据这一想法, RoBal [105]应用后处理策略, 根据训练标签频率调整余弦分类边界。

LADE [31]建议使用测试数据的标签频率 (如果可用的话) 来后调整模型输出, 而不是使用训练数据的标签频率, 以便可以针对任意的测试类别分布来校准训练模型。UNO-IC [98]建议使用一个在平衡的元验证集上调整过的超参数来校准模型分类器, 以处理类别不平衡, 从而在统一的测试集上获得更好的性能。De-Confund [99]引入了因果分类器 (参考第3.3.2节), 该分类器通过计算训练期间特征的指数移动平均值来记录偏差信息, 然后通过减去推理期间的偏差信息

1. 译者注: 费雪一致 (Fisher consistent), 即当损失函数降到最低时, 识别错误率也要达到最小。



来消除不良因果效应。DisAlign [29]为logit调整应用了自适应校准函数，其中校准函数是通过将校准预测分布与相对平衡的类别分布相匹配来学习的。

### 3.1.4 讨论

与其他长尾学习范式相比，类别重平衡方法相对简单，但可以取得相当甚至更好的性能。其中一些方法，尤其是代价敏感学习方法，在理论上受到启发或者说保证能够处理长尾问题 [16], [18], [31]。这些优点使得类别重平衡方法成为实际应用的一个很好的候选者。

然而，这种方法的一个缺点是，大多数类别重平衡方法都以头部类性能为代价来改善尾部类性能，这就像玩性能跷跷板一样。虽然总体性能得到了改善，但它无法从根本上解决信息缺少的问题，尤其在数据量很有限的尾部类上。为了解决这一限制，一个可行的解决方案是对所有类进行信息增强，如下所示。

## 3.2 信息增强

基于信息增强的方法寻求在模型训练中引入额外的信息，以便在长尾学习中提高模型性能。这种类型的方法有两种：迁移学习和数据增强。

### 3.2.1 迁移学习

迁移学习 [86], [95], [109], [130], [131] 试图从源域（例如数据集、任务或类）迁移知识，以增强目标域上的模型训练。在深度长尾学习中，主要有四种迁移学习方案，即从头部类到尾部类的知识迁移、模型预训练、知识蒸馏和自训练（self-training）。

从头部到尾部的知识迁移旨在迁移头部类的知识，以增强尾部类上的模型性能。例如，特征转移学习（FTL） [89]发现，尾部类样本的类内方差比头部类样本小得多，导致特征空间和决策边界有偏差。为了解决这个问题，FTL利用头部类的类内方差知识来指导尾部类样本的特征增强，从而使尾部类特征具有更高的类内方差，从而获得更好的尾部类性能。随后，LEAP [93]为每个类构建了“特征云”，并试图通过在特征空间中增加具有一定干扰的尾部类样本，来转移头部类特征云的知识，以增强尾部类特征云的类内变化。结果，不同类间的类内特征方差的失真得到了缓解。

在线特征增强（OFA） [95]建议使用类激活映射 [132]将样本特征解耦为类确定的和类不可知的特征。然后，OFA通过将尾部类样本的类确定特征与头部类样本的类不可知特征相结合来扩充尾部类。之后，所有增强的和原始的特征将用于使用重平衡采样器对模型分类器进行微调，从而获得更好的长尾学习性能。

稀有类别样本生成器（RSG） [109]还观察到，在长尾问题中，尾部类的特征空间远小于头部类的特征空间。为了解决这个问题，RSG建议生成新的尾部类样本，以扩大尾部类的特征空间并“推开”决策边界。为此，RSG动态估计每个类的一组特征中心，并使用头部类样本特征和最近的类内特征中心之间的特征位移来增强每个尾部类样本特征。为了进一步最大化特征位移距离并增加生成的尾部类样本特征的多样性，RSG引入了最大向量损失，以强制特征位移的方向和样本特征的方向为“共线”<sup>2</sup>。

除了特征级的头部到尾部的迁移外，主次迁移（M2m） [92]还提出通过基于扰动的优化将头部类样本转换为尾部类样本来扩充尾部类，这本质上类似于对抗性攻击。转换后的尾部类样本将用于构建更平衡的模型训练集。此外，GIST

[79]建议在分类器级别进行头部到尾部的迁移。通过利用头部类相对较大的分类器几何信息来增强尾部类的分类器权重，GIST能够获得更好的尾部类性能。

MetaModelNet [86]提出学习一种元网络，将少样本模型参数映射到多样本模型参数，其中少样本模型在少量样本上训练，而多样本模型在大量样本上训练。具体来说，元网络是在头部类上训练的，其中多样本模型直接在头部类训练集上训练，而少样本模型是在这些类别的样本子集上训练的。然后，在头部类上学习的元网络被用于映射在尾部类上训练的少样本模型，以获得更好的尾部类性能。

模型预训练是一种流行的深度模型训练方案 [133], [134], [135], [136], [137]。特定域迁移学习（DSTL） [87]首先使用所有长尾样本对模型进行预训练，以进行表征学习，然后在更平衡的训练子集上微调模型。通过这种方式，DSTL将学习到的特征慢慢地迁移到尾部类，从而在所有类别之间获得更平衡的性能。此外，自监督预训练（SSP） [100]建议首先使用自监督学习（例如对比学习 [138]或旋转预测 [139]）进行模型预训练，然后对长尾数据进行标准训练。实证结果表明了SSP的有效性，其中尾部类表现出更大的性能增益。还探索了一种方案来处理带有噪声标签的长尾数据 [140]。最近，一个新的视觉和语言预训练数据集（Conceptual 12M [108]）被提出，并被证明对下游长尾识别是有效的。

知识蒸馏旨在根据训练有素的教师模型的输出来训练学生模型 [141], [142]。最近的几项研究探索了长尾学习中的知识蒸馏。LST [81]开发了一种类别增量学习策略来处理长尾实例分割，其中知识蒸馏用于克服增量学习过程中的灾难性遗忘。从多个专家学习（LFME） [96]将整个长尾数据集划分为几个类别不平衡程度较小的子集，并使用不同的样本子集训练多个专家。基于这些专家，LFME以从易到难的课程实例选择方式，使用自适应知识蒸馏来训练统一的学生模型。遵循多专家框架，路由多样化分布感知专家（RIDE） [17]引入了一种知识蒸馏方法，通过使用更少的专家模型来训练学生网络，从而减少多专家模型的参数。

最近，自监督蒸馏（SSD） [113]开发了一种新的自蒸馏方案，以增强解耦训练。具体而言，SSD首先通过解耦训练方案（参考第3.3.3节）基于监督和自监督信息训练一个校准模型，然后使用校准模型为所有样本生成软标签。然后，使用生成的软标签和原始的长尾硬标签来蒸馏新的学生模型，接着在下一个阶段微调一个新的分类器。此外，蒸馏实际样例（DiVe） [114]表明，在长尾学习的知识蒸馏中，使用类别平衡模型作为老师是有效的。

自训练旨在从少量标记样本和大量未标记样本中学习性能良好的模型 [143], [144], [145]。具体来说，它首先使用标记样本来训练监督模型，然后使用该模型为未标记数据生成伪标签。之后，使用标记样本和伪标记样本重新训练模型。通过这种方式，自训练可以利用大量未标记样本中的知识来提高长尾学习性能。

然而，这种范式不能直接用于处理长尾问题，因为标记和未标记的数据集都可能遵循长尾类别分布。在这种情况下，在标记数据上训练的模型可能偏向于头部类，并且倾向于为未标记样本生成更多的头部类伪标签，这将导致更大程度的类别不平衡。到目前为止，如何提高自训练以解决长尾半监督学习问题仍然是一个有待探索的重要问题。

类别重平衡自训练（CReST） [107]探索了长尾分类中的自训练，发现监督模型在尾部类别上的精度<sup>3</sup>出奇地高。基于这一发现，CReST提出在每次迭代中选择更多的尾部类样本进行在线伪标记，以便重新训练的模型能够在尾部类上获得更好的性能。

2. 译者注：共线（co-linear），即特征位移的方向和样本特征的方向在同一条线上。

3. 译者注：精度（precision），注意这里的精度不是准确度（accuracy），而是查准率。

分布对齐和随机采样 (DARS) [26] 增强了处理长尾半监督语义分割的自训练。为了解决标记样本和未标记样本之间类别不平衡的潜在不一致性, DARS 将标记训练数据的标签频率视为真实的类别分布, 并强制生成的伪标签的标签频率与标记的标签频率一致。同时, DARS 采用采样策略严格控制每个类别中伪标签的数量。这样, 生成的伪标签将与真正的标签更加一致, 从而提高模型性能。

MosaicOS [115] 采用了额外的以对象为中心 (用于图像分类) 的图像来增强长尾目标检测。具体来说, 它首先使用原始检测数据集中标记的以场景为中心的图像对模型进行预训练, 然后使用预训练的模型为以对象为中心的图像生成伪边界框, 例如 ImageNet-1K [39]。然后, MosaicOS 分两个阶段对预训练好的模型进行微调, 即首先使用伪标记的以对象为中心的图像进行微调, 然后使用原始标记的以场景为中心的图像进行微调。通过这种方式, MosaicOS 减轻了数据差异的负面影响, 并有效地提高了长尾学习性能。

### 3.2.2 数据增强

数据增强旨在打包一组增强技术, 以增强模型训练数据集的大小和质量 [146], [147]。在长尾学习中, 已经探索了两种数据增强方法, 包括基于迁移的增强 (请参阅第 3.2.1 节中的从头部到尾部知识迁移) 和常规 (非迁移) 增强。

非迁移增强旨在改进或设计传统的数据增强方法, 以解决长尾问题。MiSLAS [106] 研究了长尾学习中的数据混合 (mixup), 发现 (1) 数据混合有助于纠正模型的过度自信; (2) 在解耦训练方案中, 数据混合对表征学习有积极影响, 但对分类器学习有消极或可忽略的影响 [32]。根据这些观察, MiSLAS 建议在解耦方案中使用数据混合来增强表征学习。此外, Remix [148] 还利用数据混合进行长尾学习, 并引入了一种重平衡的混合方法来专门增强尾部类。

FASA [74] 提出, 根据先前观察到的样本估计一个高斯先验的均值和方差, 然后基于这一高斯先验来生成类特征。此外, FASA 利用平衡验证集上的模型分类损失来调整不同类别特征的采样率, 从而可以进一步增强表示不足的尾部类别。

元语义扩充 (MetaSAug) [110] 提出用隐式语义数据增强 (ISDA) 的一个变体来增强尾部类 [149]。具体地说, ISDA 估计类别条件统计量 (即样本特征的协方差矩阵) 以获得语义方向, 并通过转换样本特征和不同语义有意义的方向来生成不同的增强样本。然而, 尾部类样本的不足使得估计尾部类的协方差矩阵无效。为了解决这一问题, MetaSAug 探索了元学习, 以指导每个类别的协方差矩阵的学习, 并使其具有类平衡损失 [16]。这样, 可以更准确地估计尾部类的协方差矩阵, 从而使生成的尾部类特征信息更加丰富。

### 3.2.3 讨论

由于引入了额外的知识, 基于迁移学习的方法在不牺牲头部类性能的情况下提高了尾部类性能。考虑到缺乏足够的尾部类样本是长尾学习的关键问题之一, 这种方法值得进一步探索。

数据增强是一种相对基础的技术, 可以用于各种长尾问题, 这使得这种方法在现实世界的应用中比其他方法更实用。然而, 简单地使用现有的类别不可知增强技术来改进长尾学习是不利的, 因为考虑到头部类有更多的样本并且将被扩充更多, 它们可能会进一步增加不平衡。如何更好地为长尾学习进行数据增强仍是一个悬而未决的问题。

## 3.3 模块改进

除了类别重平衡和信息增强外, 研究人员还探索了在长尾学习中改进网络模块的方法。这些方法可分为四类: (1) 表

征学习改进了特征提取器; (2) 分类器设计增强了模型分类器; (3) 解耦训练增强了特征提取器和分类器的学习;

(4) 集成学习改进了整个体系结构。

### 3.3.1 表征学习

现有的长尾学习表征学习方法基于四种主要范式, 即度量学习、顺序训练、原型学习和迁移学习。

度量学习旨在设计特定于任务的距离度量, 以确定目标之间的相似性或相异性; 在长尾学习中, 基于度量学习的方法寻求探索基于距离的损失, 以学习更具区分性的特征空间。LMLE [75] 就是一个例子, 它引入了一种五元组损失学习表示法, 可以保持簇间和类间的边距 (margin)。此外, 幅度损失 [21] 通过使用一个小批次内所有样本对之间的总距离, 创新了表征学习。换言之, 幅度损失使用整个批次的统计信息, 而不是实例级别的统计信息, 从而缓解了总体类别数据数量不平衡的偏差。更具体地说, 幅度损失通过最大化小批量中任意两个类中心的距离来扩大类间距离, 并通过最小化类内样本之间的最大距离来减少类内变化。通过这种方式, 幅度损失可以获得具有更好的识别能力和更少的不平衡偏差的特征。

类别校正损失 (CRL) [85] 旨在增强尾部类样本表示, 使其具有更大程度的类内紧凑性和类间距离。为此, CRL 为尾部类构造了大量的难样本对 (hard-pair) 三元组, 并将类别校正损失 (类似于三元组损失 [76]) 作为类别平衡约束。这样, 学习模型克服了类别不平衡对表征学习的负面影响。

最近的研究也探讨了长尾问题的对比学习。KCL [13] 提出了一种  $k$ -正对比损失学习均衡特征空间的方法, 有助于缓解类间的不平衡, 提高模型的泛化能力。随后, Hybrid [111] 引入了一种典型的对比学习策略来增强长尾学习。参数对比学习 (PaCo) [116] 进一步创新了监督对比学习, 增加了一组参数可学习类中心, 如果将类中心作为分类器权重, 则这些类中心起到与分类器相同的作用。DRO-LT [117] 将原型对比学习与分布鲁棒优化进行了扩展 [150], 使学习模型对数据分布偏移更加鲁棒。

顺序训练。分层特征学习 (HFL) [83] 的灵感来自于每个类别在有区分力的视觉表征中都有各自独特的特征。因此, HFL 分层地将目标聚类到视觉上相似的类别组中, 形成分层聚类树。在该聚类树中, 原始节点中的模型在 ImageNet-1K 上进行预训练; 每个子节点中的模型继承其父节点的模型参数, 然后用簇节点中的样本进行微调。通过这种方式, 从拥有大量类别的组中获得的知识逐渐转移到拥有较少类别的组中。

不平等训练 [90] 提出将数据集划分为头部类和尾部类子集, 并在训练过程中对它们进行不同的处理。首先, 不平等训练使用头部类样本, 通过一个来新的抗噪损失, 来训练相对具有识别力和抗噪能力的特征。然后, 利用尾部类样本, 通过难特征挖掘和一种新的中心分散损失来增强表征的类间区分。这里, 中心分散损失基于每个类别中的归一化特征。

基于原型学习的方法旨在学习特定于类别的特征原型, 以提高长尾学习性能。开放式长尾识别 (OLTR) [15] 创新性地探索了在开放世界中处理长尾识别的特征原型的想法, 其中测试集包括头部、尾部和开放类。这里, 开放类表示训练集中不存在的测试类。为了完成这项任务, OLTR 维护了一个包含区分性特征原型的视觉元内存, 并使用从视觉内存中提取的特征来扩充原始特征。这样, 学习到的特征空间将更具区分性, 来自新类别的样本特征将远离内存, 更接近原点。这样的特征空间使 OLTR 能够区分闭集类以及检测新类。此外, OLTR 还探索了一种自我注意机制来增强特征学习。

随后, 膨胀片段记忆 (IEM) [94] 通过一种动态更新方案进一步创新了元嵌入记忆, 其中每个类别都有独立且可区分的记忆块, 而每个记忆块记录了对应类别最具辨别力的特征



原型。由于动态内存库只包含最具辨别力的特征原型，因此它们不受类别数量不平衡问题的影响。此外，IEM还探索了一种区域自注意力机制，以进一步加强表征学习。

TABLE 4: 分类器总结，其中 $w, f, b, \phi$ 和 $p$ 分别表示模型分类器、样本特征、偏差项、softmax函数和预测概率。此外， $\hat{d}$ 是指数移动平均特征的单位向量。温度因子用 $\tau$ 表示，其他与分类器相关的超参数包括 $\gamma$ 和 $\alpha$ 。

分类器	表达式
线性分类器	$p = \phi(w^\top f + b)$
余弦分类器	$p = \phi(\tau \frac{w^\top f}{\ w\  \ f\ } + b)$
$\tau$ -归一化分类器	$p = \phi(\frac{w^\top f}{\ w\ _2} + b)$
因果分类器	$p = \phi(\tau \frac{(w)^\top f}{(\ w\  + \gamma)\ f\ } - \alpha \frac{\cos(x, \hat{d})(w)^\top \hat{d}}{\ w\  + \gamma})$

迁移学习。在第3.2.1节中，我们介绍了一些基于迁移学习的长尾方法，这些方法改进了表征学习，包括SSP [100]和LEAP [93]。除此之外，无监督发现 (UD) [35]建议使用无监督学习来帮助从具有长尾目标的图像中发现新的、更细粒度的目标。具体来说，UD首先使用预先训练好的类别不可知的掩码建议网络为所有可能的目标生成目标边界框和分割掩码。然后，UD基于边界框和语义掩码的特征，应用三种新的自监督三元组损失来学习双曲特征空间。UD最后根据学习到的特征进行无监督聚类 and 专有标签分配，以发现新的、更细粒度的目标。

除上述学习方案外，decoupling [32]还创新了长尾表征学习，采用了不同的采样策略，包括实例平衡、类平衡、平方根和渐进平衡采样。在实例平衡抽样之后，MiSLAS [106]实证研究了数据混合 (mixup) 对长尾表征学习的影响。

### 3.3.2 分类器设计

除了表征学习，研究人员还设计了不同类型的分类器来解决长尾问题。在一般视觉问题 [10], [138]中，深度学习的常见实践是使用线性分类器 $p = \phi(w^\top f + b)$ ，其中 $\phi$ 表示softmax函数，偏差项 $b$ 可以丢掉。

然而，长尾类别不平衡通常会导致头部类的分类器权重范数大于尾部类的分类器权重范数 [89]，这使得线性分类器容易偏向优势类。为了解决这一问题，一些研究 [93], [105]建议使用尺度不变的余弦分类器 $p = \phi((\frac{w^\top f}{\|w\| \|f\|})/\tau + b)$ ，其中分类器权重和样本特征均被归一化。这里，应合理选择温度 $\tau$  [125]，否则会对分类器性能产生负面影响。

$\tau$ -归一化分类器 [32]通过 $\tau$ -归一化过程调整分类器权重范数，从而纠正决策边界的不平衡。形式上，设 $\tilde{w} = \frac{w}{\|w\|_2}$ ，其中 $\tau$ 是归一化的温度系数。当 $\tau = 1$ 时， $\tau$ -归一化退化为 $L_2$ 归一化，而当 $\tau = 0$ 时，不施加尺度变化。注意，超参数 $\tau$ 也可以通过类别平衡采样进行训练，得到的分类器被命名为可学习的权重缩放分类器 [32]。此外，最近类别均值分类器 [32]计算训练集中每个类的 $L_2$ -标准化均值特征，然后基于最近邻算法 [151] 进行预测，或者使用余弦相似性或欧氏距离。

基于现实分类学的分类器 (RTC) [97]提出用分层分类来解决类别不平衡问题。具体而言，RTC将图像映射到一个分类学的树状结构中，这个树的层次结构由一组分类节点和节点关系定义。不同的样本在不同的层次上进行自适应分类，其中进行预测的层次取决于样本分类的难度和分类器的置信度。这样的设计有利于中间层级节点的正确决策，而不是叶子节点的错误决策。

因果分类器 [99]借助因果推理来在长尾学习中保持好的并消除坏的动量因果效应。好的因果效应指示了稳定梯度和加速训练的有利因素，而坏的因果效应表明累积的长尾偏差导致了较差的尾部性能。为了更好地

逼近偏差信息，因果分类器采用多头策略将模型权重和数据特征的通道（或维度）平均划分为 $K$ 组。形式上，因果分类器通过 $p = \phi(\frac{\tau}{K} \sum_{k=1}^K \frac{(w^k)^\top f^k}{(\|w^k\| + \gamma)\|f^k\|})$ 计算原始logits，其中 $\tau$ 是温度因子， $\gamma$ 是超参数。当 $\gamma = 0$ 时，该分类器本质上是余弦分类器。在推理过程中，因果分类器通过在输入为空时减去预测来消除不良因果效应，即 $p = \phi(\frac{\tau}{K} \sum_{k=1}^K \frac{(w^k)^\top f^k}{(\|w^k\| + \gamma)\|f^k\|} - \alpha \frac{\cos(x, \hat{d})(w^k)^\top \hat{d}^k}{\|w^k\| + \gamma})$ ，其中 $\hat{d}$ 是指数移动平均特征的单位向量， $\alpha$ 是控制直接和间接影响的权衡参数。更直观地说，分类器在训练过程中通过计算指数移动平均特征来记录偏差，然后在推理过程中通过从预测logits中减去偏差来消除不良的因果效应。

GIST分类器 [79]试图将头部类的几何结构转换为尾部类。具体而言，GIST分类器由一个特定于类的权重中心（用于编码类别位置）和一组位移（用于编码类别几何形状）组成。通过利用头部类相对较大的位移来增强尾部类地权重中心，GIST分类器能够在尾部类上获得更好的性能。

### 3.3.3 解耦训练

解耦训练将学习过程解耦为表征学习和分类器训练。Decoupling [32]是引入两阶段训练方案的开创性工作。它在第一阶段（参考第3.1.1节）对表征学习的不同采样策略进行了实证评估，然后在第二阶段通过固定经过训练的特征提取器来评估不同的分类器训练方案。在分类器学习阶段，也有四种方法，包括使用类别平衡采样重新训练分类器、最近类别均值分类器、 $\tau$ -归一化分类器和可学习的权重缩放方案。主要观察结果有两方面：在长尾识别中，（1）实例平衡采样是表征学习的最佳策略；（2）分类中的魔鬼：重新调整分类器可以带来显著地性能提升。

随后，KCL [13]根据经验观察到，平衡的特征空间有利于长尾学习。因此，它创新了解耦的训练方案，通过开发一个 $k$ -正对比损失来学习一个类别更平衡的、类别之间区分度更高的特征空间，从而获得更好的长尾学习性能。

MiSLAS [106]根据经验观察到，在两阶段解耦训练方案下，数据混合 (mixup) 有利于特征学习，但对分类器训练有负面/可忽略的影响。因此，MiSLAS建议在第一阶段通过数据混合来增强表示学习。在第二阶段，MiSLAS采用标签感知平滑策略来更好地进行模型泛化，并通过不断更新批次归一化层 (batch normalization) 中的运行均值和方差，进一步解决两个训练阶段之间的分布偏移问题。

最近的一些研究通过增强分类器的训练阶段，创新了解耦训练方案。OFA [95]通过尾部类特征增强创新了分类器的重新训练，具体做法是从尾部类中提取到的特定特征与从头部类中提取到的通用特征相结合。SimCal [34]提出通过使用一种新的用于长尾实例分割的双层类别平衡采样策略来校准分类头，从而增强分类器的训练阶段。DisAlign [29]利用一种新的自适应校准策略创新了分类器训练。具体而言，通过最小化校准预测分布和平衡参考分布之间的KL-散度来学习新的自适应校准函数，用于调整原始分类器的输出logits。总之，DisAlign本质上是使用了一个额外的分类器层来校准原始分类器，通过将校准的预测分布与相对平衡的类别分布相匹配来实现这一目的。

最近，DT2 [118]将解耦训练方案应用于场景图生成任务，这证明了解耦训练在处理视觉关系学习中的长尾类别不平衡方面的有效性。

### 3.3.4 集成学习

基于集成学习的方法战略性地生成和组合多个网络模块（即多个专家），以解决长尾视觉学习问题。图3总结了现有基于集成方法的主要方案，具体如下。



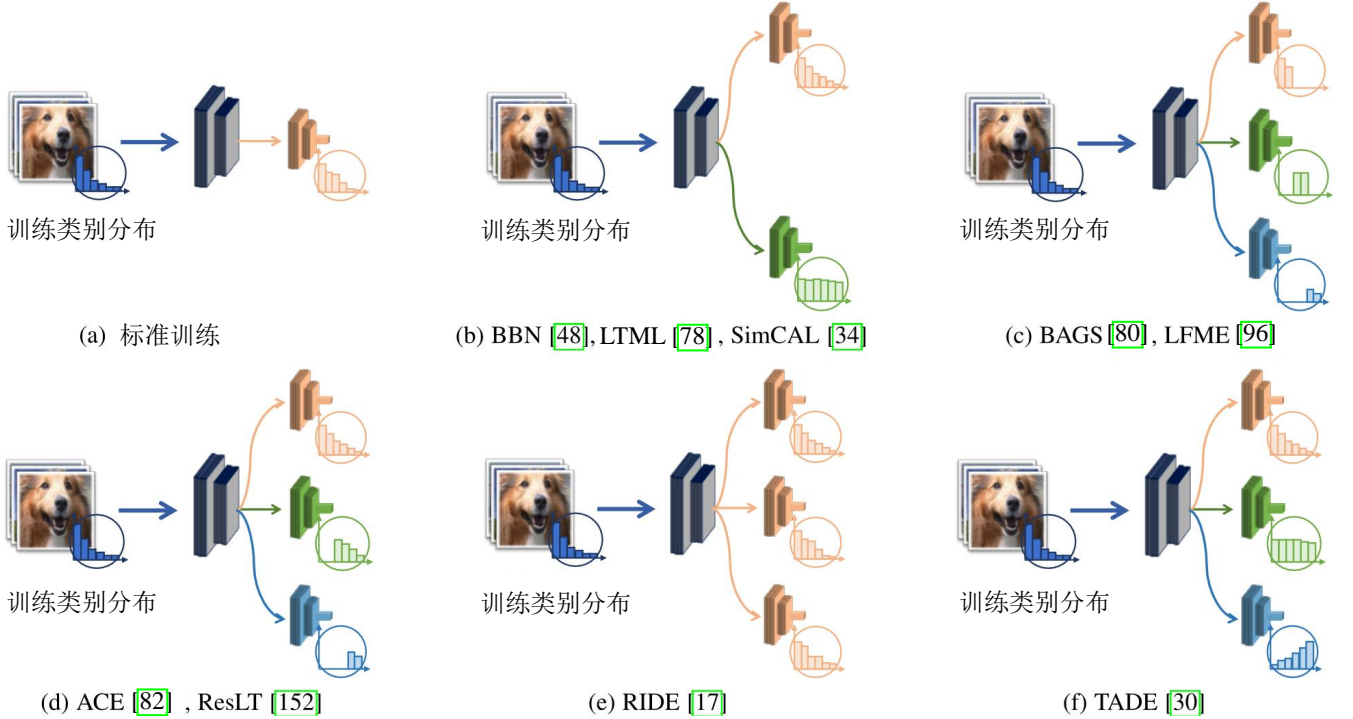


Fig. 3: 现有基于集成的长尾学习方法的示例。训练过的专家可能具有不同的技能，例如，熟练掌握不同的类别分布或不同的类别子集。

BBN [48]建议使用两个网络分支，即传统学习分支和重平衡分支来处理长尾识别。具体而言，传统学习分支采用均匀采样来模拟原始的长尾训练分布；重平衡分支使用反向采样器，在每个小批次中采样更多的尾部类样本，以提高尾部类性能。在训练过程中，两个分支的预测结果被动态组合，因此BBN的学习重点逐渐从头部类转向尾部类（通过重平衡分支）。

继BBN之后，LTML [78]探索了双边分支网络方案，以解决长尾多标签分类问题。具体来说，LTML使用sigmoid交叉熵损失对每个分支进行多标签分类，并强制执行logit一致性损失以提高两个分支的一致性。此外，LTML采用标签平滑和logit补偿来提高模型的泛化能力。

与BBN类似，SimCal [34]探索了一种双重分类头方案来解决长尾实例分割问题。具体而言，SimCal维护了两个分类头：原始分类头和校准分类头。基于一种新的双层抽样策略，校准后的分类头能够改善尾部分类的性能，而原始分类头旨在保持头部分类的性能。

不同于双边分支，BAGS [80]探索了一种多头方案来解决长尾目标检测问题。具体来说，BAGS的灵感来自一项观察，即用较少的样本学习更均匀的分布有时比用更多的样本学习长尾分布更容易。因此，BAGS首先将类别划分为若干子组，其中每个子组中的类别具有相似数量的训练数据。然后，BAGS在共享的特征提取器上使用多个分类头进行预测，其中不同的分类头在不同的数据子组上进行训练。这样，每个分类头对具有相似数量训练数据的类执行softmax操作，从而避免了类别不平衡的负面影响。此外，BAGS还将“其他”类引入每组，以缓解不同分类头之间的不一致。

与BAGS类似，LFME [96]将长尾数据集划分为几个类别不平衡程度较小的子集，并使用不同的样本子集培训多个专家。然后，基于这些专家，LFME使用来自多个教师专家的自适应知识蒸馏来学习统一的学生模型。

与将数据集划分为几个平衡的子集不同，ACE [82]将类别划分为几个技能不同的子集：一个子集包含所有类别；

一个包含中间类和尾部类；另一个只有尾部类。然后ACE训练具有不同类别子集的多个专家，以便不同的专家具有特定和互补的技能。此外，考虑到不同的子集具有不同的样本数，ACE还使用分布式自适应优化器来调整不同专家的学习率。ResLT [152]也探讨了ACE的类似概念。

RIDE [17]没有进行数据划分，而是根据所有训练样本对每个专家进行独立的softmax损失训练，并强制执行基于KL散度的损失，以提高不同专家的多样性。之后，RIDE应用专家分配模块来提高计算效率。注意，使用softmax损失对每位专家进行独立训练可以大大提高长尾学习的整体性能。

测试阶段集成多元专家（TADE）[30]探索了处理测试分布不可知的长尾识别的多专家方案，其中测试类别分布可以是均匀的，也可以是长尾的。具体而言，TADE开发了一个新的多专家框架，并通过引入促进多样性的专业知识引导损失来创新了专家训练方案，以训练不同的专家处理不同的类别分布。通过这种方式，学习之后的专家更加多样化，从而获得更好的集成性能，使得整体的模型能够适应不同的测试类别分布。基于这一特性，TADE进一步引入了一种新的测试阶段自监督学习方法，即预测稳定性最大化，以仅基于未标记的测试数据自适应集成专家，更好地处理未知的测试类别分布。

### 3.3.5 讨论

表征学习和分类器设计是深度长尾学习的基本问题，值得进一步探索。近年来，解耦训练越来越受到人们的关注；在该方案中，类平衡分类器学习的第二阶段不会引入太多的计算开销，但会带来显著的性能提升。一种批评 [82]是，累积的训练阶段使解耦训练变得不太实用，无法与其他长尾问题（例如，目标检测和实例分割）中现有的公式化方法集成。尽管如此，解耦训练的思想在概念上很简单，因此可以很容易地用于设计解决各种长尾学习问题的新方法。

与其他类型的长尾学习方法相比，基于集成的方法通常在头部类和尾部类上都能获得更好的性能。这些方法的一个

问题是，由于使用多个专家，它们通常会导致更高的计算成本。然而，通过使用共享特征提取器可以缓解这种担忧。此外，面向效率的专家分配和知识蒸馏策略 [17], [96]也可以降低计算复杂性。

## 4 实证研究

本节实证分析了现有的长尾学习方法。首先，我们引入一个新的评估指标。

### 4.1 新的评估指标

长尾学习的关键目标是处理类别不平衡，以获得更好的模型性能。因此，通用评估协议 [13], [22] 直接使用top-1测试精度（由 $A_t$ 表示）来判断长尾方法的性能，以及哪种方法处理类别不平衡更好。然而，这种度量不能准确反映处理类别不平衡时不同方法之间的相对优势，因为top-1精度还受类别不平衡以外的其他因素的影响。例如，集成学习（或数据增强）等长尾方法也可以提高在平衡训练集上训练的模型的性能。在这种情况下，很难判断性能的提高是来自于类别不平衡的缓解，还是来自于更好的网络架构（或更多的数据信息）。

为了更好地评估长尾方法处理类别不平衡的有效性，我们提出了一个新的度量，即相对准确度 $A_r$ ，以减轻长尾学习中不必要因素的影响。为此，我们首先计算一个经验上较高的参考精度 $A_u = \max(A_v, A_b)$ ，其中一般精度 $A_v$ 是在平衡训练集上使用交叉熵损失训练的相应的骨干网络得到的，平衡精度 $A_b$ 是在平衡训练集上使用相应的长尾方法得到的模型结果，参考精度 $A_u$ 是这两者中的最大值。这里，平衡训练集是长尾训练集的一个变体，具有相似的总样本数，但每个类别具有的样本数是相同的。从平衡训练集中获得的这一较高的参考精度被用于缓解类别不平衡的影响；然后，相对精度被定义为 $A_r = \frac{A_t}{A_u}$ 。在我们的实验中，top-1测试精度 $A_t$ 、较高参考精度 $A_u$ 和相对准确度 $A_r$ 都将用于评估。

### 4.2 实验设置

我们接下来介绍实验设置。

数据集。考虑到广泛使用的ImageNet-LT [15]具有很大的类别数目（1000）和大的不平衡率（256），我们采用了ImageNet-LT作为实证研究的基准长尾数据集。与ImageNet-LT对应的平衡训练集是基于 [13]进行采样的。ImageNet-LT的总样本数见表1。除了在所有类别上的总体表现外，我们还报告了在ImageNet-LT的三个类别子集：头部类（超过100张图像）、中间类（20张图像100张图像）和尾部类（少于20张图像）上的表现。

基线。我们通过以下标准来选择长尾学习方法：

(1) 官方源代码公开或易于复现的；(2) 在相应的论文中，在ImageNet-LT数据集上对这些方法进行了评估的。据此，本文对20多种方法进行了实证评估，包括基线（Softmax），代价敏感学习（加权Softmax、焦点损失 [84]、LDAM [18]、ESQL [19]、平衡Softmax [73]、LADE [31]），logit调整（UNO-IC [98]），迁移学习（SSP [100]），数据增强（RSG [109]），表征学习（OLTR [15]，PaCo [116]），分类器设计（De-confound [99]），解耦训练（Decouple-IB-CRT [32]，CB-CRT [32]，SR-CRT [32]，PB-CRT [32]，MiSLAS [106]），集成学习（BBN [48]，LFME [96]，RIDE [17]，ResLT [152]，TADE [30]）。有关这些方法的更多详细信息，请参见第3节。

实现细节。我们在PyTorch中实现了所有实验。延续 [17], [31], [32]，我们使用ResNeXt-50作为所有方法的骨干网络。我们使用SGD优化器，批次大小设置为256，动量设

置为0.9，权重衰减系数为0.0005，学习率为0.1（线性LR衰减）。对于与方法相关的超参数，我们直接按照原始论文设置，如果默认值表现不佳，则手动进行调整。此外，我们对所有方法使用相同的基本数据增强（即，随机调整大小并裁剪为224，随机水平翻转，颜色抖动和归一化），而基于数据增强的长尾方法中提出的其他增强技术将用于这些基本增强操作之上。

### 4.3 所有类别上的结果

表5和图4报告了所有类别上的平均性能。从这些结果中，我们对总体方法进展和不同方法类型进行了一些观察。

所有方法上的观察。如表5所示，几乎所有的长尾方法在准确性方面都优于Softmax基线，这证明了长尾学习的有效性。即便如此，有两种方法的性能比Softmax稍差，即Decouple-CB-CRT [32]和BBN [48]。我们推测，Decouple-CB-CRT表现不佳的原因是在解耦训练的第一阶段，通过类别平衡采样导致了不佳的表征学习（更多实证观察请参考 [32]）。BBN（基于官方代码）的不良结果可能来自累积学习策略，该策略逐渐将学习重点从头部类调整到尾部类；然而，在训练结束时，它可能会把太多的注意力放在尾部类。因此，尽管尾部类性能更好了，但头部类的模型精度显著下降（参考表6），导致平均性能更差。

除了普通的准确度之外，我们还评估了基于较高的参考精度（UA）和相对准确度（RA）的长尾方法。表5显示，大多数方法的UA与基线模型相同，但仍有一些方法的UA较高，例如SSP、MiSLAS、TADE。对于这些方法，性能的提高不仅来自于类别不平衡的缓解，还来自其他因素，如数据增强或更好的网络架构。因此，仅仅使用准确度进行评估是不够准确的，而我们提出的RA度量提供了一个很好的补充，因为它减轻了类别不平衡之外的因素的影响。例如，基于数据混合的MiSLAS在90个训练周期下比平衡Softmax具有更高的精度，但它也具有更高的UA。因此，MiSLAS的RA低于平衡Softmax，这意味着在90个训练周期下，平衡Softmax比MiSLAS更好地缓解了类别不平衡。当训练周期增加到200时，MiSLAS的RA高于平衡Softmax。也就是说，尽管有其他因素可以提高性能，但经过充分训练的MiSLAS也比平衡的Softmax具有更好的处理类别不平衡的能力。200个训练周期下的更多示例可在图4（a,c）中找到。

尽管最近一些高精度方法的RA较低，但长尾学习的总体发展趋势仍然是积极的，如图4所示。这种性能趋势表明，最近对长尾学习的研究取得了真正的进展。此外，最先进的TADE的RA为93.0，这意味着未来仍有提升的空间。

此外，我们还在表5中评估了不同训练周期（即90和200）对模型训练的影响。总的来说，使用200个周期的训练可以提高大多数长尾方法的性能，因为充分的训练可以使深层模型更好地拟合数据并学习更好的视觉表示。然而，也有一些方法在只训练90个周期时表现更好，例如，De-confound和Decouple-CB-CRT。我们推测，对于这些方法，90个周期足以很好地训练模型，而训练更多周期不会带来额外的好处，但会增加训练难度，因为这也会影响学习率衰减方案。

对不同方法类型的观察。我们还从不同方法类型的角度讨论了表5中的结果。首先，与基线模型相比，几乎所有的类别重平衡（CR）方法都有利于长尾学习性能。具体而言，LADE、平衡Softmax和LDAM在这种方法类型中达到了最先进的水平。此外，还提出了焦距损失来处理不平衡的目标检测 [84]；然而，当处理大量的长尾类别时（例如，ImageNet-LT中的1000个类别），焦距损失表现得不好，只能带来微小的提升。在LDAM中，除了LDAM损失外，还有一个延迟重平衡优化方案；注意，在没有延迟方案的情况下，仅使用LDAM丢失进行学习可能无法获得令



TABLE 5: 在ImageNet-LT上训练90或200个周期的准确度 (Acc)、较高的参考精度 (UA)、相对准确度 (RA) 结果。在该表中, CR、IA和MI分别表示类别重平衡、信息增强和模块改进。

类型	方法	90 epochs			200 epochs		
		Acc	UA	RA	Acc	UA	RA
基线	Softmax	45.5	57.3	79.4	46.8	57.8	81.0
CR	加权Softmax	47.9	57.3	83.6	49.1	57.8	84.9
	焦点损失 [84]	45.8	57.3	79.9	47.2	57.8	81.7
	LDAM [18]	51.1	57.3	89.2	51.1	57.8	88.4
	ESQL [19]	47.3	57.3	82.5	48.0	57.8	83.0
	UNO-IC [98]	45.7	57.3	81.4	46.8	58.6	79.9
	平衡Softmax [73]	50.8	57.3	88.7	51.2	57.8	88.6
IA	LADE [31]	51.5	57.8	89.1	51.6	57.8	89.3
	SSP [100]	53.1	59.6	89.1	53.3	59.9	89.0
	RSG [109]	49.6	57.3	86.7	52.9	57.8	91.5
MI	OLTR [15]	46.7	57.3	81.5	48.0	58.4	82.2
	PaCo [116]	52.7	58.7	89.9	54.4	59.6	91.3
	De-confound [99]	51.8	57.7	89.8	51.3	57.8	88.8
	Decouple-IB-CRT [32]	49.9	57.3	87.1	50.3	58.1	86.6
	Decouple-CB-CRT [32]	44.9	57.3	78.4	43.0	57.8	74.4
	Decouple-SR-CRT [32]	49.3	57.3	86.0	48.5	57.8	83.9
	Decouple-PB-CRT [32]	48.4	57.3	84.5	48.1	57.8	83.2
	MisLAS [106]	51.4	58.3	88.2	53.4	59.7	89.4
	BBN [48]	41.2	57.3	71.9	44.7	57.8	77.3
	LFME [96]	47.0	57.3	82.0	48.0	57.8	83.0
	ResLT [152]	51.6	57.3	90.1	53.2	58.1	91.6
	RIDE [17]	55.5	60.2	92.2	56.1	60.9	92.1
	TADE [30]	57.3	61.9	92.6	58.8	63.2	93.0

人满意的性能。此外, 如表5所示, 大多数代价敏感学习方法的UA上限相同, 因此RA与普通准确度呈正相关 (参考图4 (b))。因此, 这种方法类型中准确度的提高可以准确反映类别不平衡的缓解。

在信息增强 (IA) 类方法中, 迁移学习 (SSP) 和数据增强 (RSG) 都有助于处理长尾类别不平衡。虽然SSP也提高了UA, 但RA的提高更为显著, 这意味着性能的提高主要来自类别不平衡的处理。考虑到缺乏足够的尾部类样本是关键挑战之一, 而IA在训练中引入了更多信息, 因此这类方法值得进一步探索。

在模块改进 (MI) 类方法中, 所有细分方法都有助于处理类不平衡。到目前为止, 在准确度和RA方面最先进的方法是集成学习的长尾方法, 即TADE [30]和RIDE [17]。虽然集成学习也提高了UA, 但处理不平衡问题带来的性能增益更为显著, 即RA更高 (参考图4 (d))。

#### 4.4 类别子集上的结果

本节给出了不同类别子集上各方法的性能。如表6所示, 几乎所有的方法都以牺牲头部类性能为代价来提高尾部类和中间类的性能。然而, 头部类在长尾学习中也很重要, 因此有必要在不牺牲头部类性能的情况下提高长尾性能。潜在的解决方案包括信息增强和集成学习, 例如SSP、RIDE和TADE (需要足够的训练时间, 例如200个周期)。

通过比较表5和表6, 我们可以发现, 总体性能的提高在很大程度上取决于中间类和尾部类的提高; 因此, 如何提高它们的表现仍然是长尾学习未来最重要的目标。

到目前为止, TADE [30]在精确度和RA方面取得了最佳的整体性能 (参考表5), 但TADE并没有在所有类别子集上表现出最好的性能 (参考表6)。例如, 当训练200个周期时, TADE的头部类性能比RIDE差, 而尾部类性能比BBN差。总而言之, TADE的平均性能较高, 这意味着获得更好的长尾性能的关键是在所有类别之间进行更好的权衡。

总之, 当前深度长尾学习的最佳实践是同时使用集成学习和类别重平衡。请注意, 除了基于数据增强的方法之外, 所有这些方法都只使用基本的增强操作。如果使用更强的数据

TABLE 6: 在ImageNet-LT上训练90或200个周期得到的头部类、中间类、尾部类的准确度 (Acc) 结果。在该表中, WS表示加权softmax, BS表示平衡softmax。方法的类型与表5相同。

方法	90 epochs			200 epochs		
	头部	中间	尾部	头部	中间	尾部
Softmax	66.5	39.0	8.6	66.9	40.4	12.6
WS	66.3	42.2	15.6	57.9	46.2	34.0
焦点损失 [84]	66.9	39.2	9.2	67.0	41.0	13.1
LDAM [18]	62.3	47.4	32.5	60.0	49.2	31.9
ESQL [19]	62.5	44.0	15.7	63.1	44.6	17.2
UNO-IC [98]	66.3	38.7	9.3	67.0	40.3	12.7
BS [73]	61.7	48.0	29.9	62.4	47.7	32.1
LADE [31]	62.2	48.6	31.8	63.1	47.7	32.7
SSP [100]	65.6	49.6	30.3	67.3	49.1	28.3
RSG [109]	68.7	43.7	16.2	65.0	49.4	31.1
OLTR [15]	58.2	45.5	19.5	62.9	44.6	18.8
PaCo [116]	59.7	51.7	36.6	63.2	51.6	39.2
De-confound [99]	63.0	48.5	31.4	64.9	46.9	28.1
IB-CRT [32]	62.6	46.2	26.7	64.2	46.1	26.0
CB-CRT [32]	62.4	39.3	14.9	60.9	36.9	13.5
SR-CRT [32]	64.1	43.9	19.5	66.0	42.3	18.0
PB-CRT [32]	63.9	45.0	23.2	64.9	43.1	20.6
MisLAS [106]	62.1	48.9	32.6	65.3	50.6	33.0
BBN [48]	40.0	43.3	40.8	43.3	45.9	43.7
LFME [96]	60.6	43.5	22.0	64.1	42.3	22.8
ResLT [152]	57.8	50.4	40.0	61.6	51.4	38.8
RIDE [17]	66.9	52.3	34.5	67.9	52.3	36.0
TADE [30]	65.3	55.2	42.0	67.2	55.3	40.0

增强操作, 例如RandAugment [153] 或Cutmix [154], 模型性能可以进一步提高。

#### 4.5 关于代价敏感损失的更多讨论

在本节中, 我们进一步评估了基于解耦训练方案 [32] 的不同代价敏感学习损失的性能, 该方案将表征和分类器学习解耦为两个阶段。在第一阶段, 我们使用不同的代价敏感学习损失来训练模型主干以进行表征学习; 在第二阶段, 我们使用四种不同的分类器训练策略 [32], 即无需再训练的联合训练、最近类别均值分类器 (NCM)、类别平衡分类器再训练 (CRT) 和可学习的权重缩放 (LWS)。

如表7所示, 与联合训练相比, 除平衡softmax (BS) 外, 解耦训练可以进一步提高大多数代价敏感方法的总体性能。这些方法在解耦训练下可以获得与BS相当的性能, BS在联合训练下表现最好。这些结果特别有趣, 因为它们意味着, 尽管这些代价敏感损失在联合训练下表现不同, 但它们基本上学习到了具有相似质量的特征表示。

TABLE 7: 在ImageNet-LT上进行解耦训练200个周期得到的各种代价敏感损失的性能结果。这里, “Joint”表示单阶段的端到端联合训练; “NCM”是最近类别均值分类器 [32]; “CRT”表示类别平衡分类器再训练 [32]; “LWS”是指可学习权重缩放 [32]。此外, BS表示平衡softmax方法 [73]。

Test Dist.	所有类准确度				头部类准确度			
	Joint	NCM	CRT	LWS	Joint	NCM	CRT	LWS
Softmax	46.8	50.2	50.2	50.8	66.9	63.5	65.0	64.6
焦点损失 [84]	47.2	50.7	50.7	51.5	67.0	62.6	64.5	64.3
ESQL [19]	48.0	49.8	50.6	50.5	63.1	60.2	64.0	63.3
BS [73]	51.2	50.4	50.6	51.1	62.4	62.4	64.9	64.3
Test Dist.	中间类准确度				尾部类准确度			
	Joint	NCM	CRT	LWS	Joint	NCM	CRT	LWS
Softmax	40.4	45.8	45.3	46.1	12.6	28.1	25.5	28.2
焦点损失 [84]	41.0	47.0	46.4	47.3	13.1	30.1	26.9	30.2
ESQL [19]	44.6	46.6	46.5	46.1	17.2	31.1	27.1	29.5
BS [73]	47.7	46.8	46.1	46.7	32.1	29.1	26.2	29.4

与联合训练相比, 解耦训练下的BS整体表现较差, 这可能意味着BS已经很好地进行了类别重平衡; 进一步使用分类器再训练进行重平衡不但不会带来额外的好处, 甚至会降低

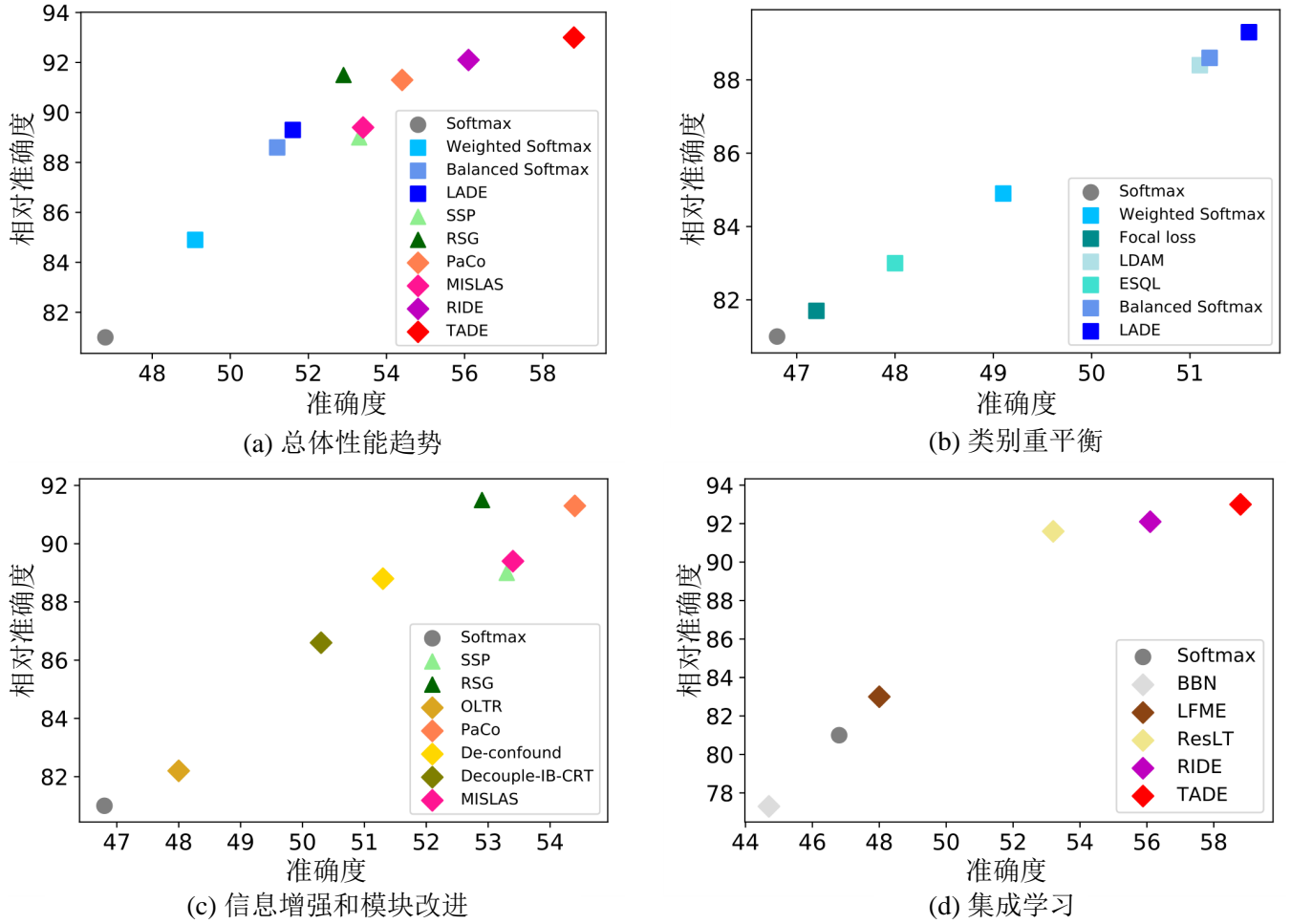


Fig. 4: 在200个训练周期下，长尾学习方法在准确度和相对准确度方面的性能趋势。图（a）包括来自多种方法类型的方法；图（b）指类别重平衡；图（c）包括信息增强和部分模块改进方法；图（d）指集成学习。这里○表示softmax基线；□表示类别重平衡；△和◇分别代表信息增强和模块改进方法。不同的颜色，在所有子图形中一致，表示不同的方法。

通过端到端联合训练得到的网络参数的一致性。类似的观察也可以在头部类、中间类和尾部类的表现中找到。

## 5 应用

本节讨论了深度长尾学习的主要视觉应用，包括图像分类、图像检测和分割以及视觉关系学习。

### 5.1 图像分类

长尾学习最常见的应用是多类别分类 [15], [30], [32], [48], [99]。如第2.2节所述，有很多从广泛使用的图像分类数据集（即ImageNet、CIFAR和Places）中人工采样得到的长尾数据集。基于这些数据集，提出了各种长尾学习方法，如第3节所示。除了这些人工任务外，长尾学习还应用于现实世界的图像分类任务，包括物种分类 [23], [24], [155]、人脸识别 [21], [22], [90], [93]、年龄分类 [104]、标识检测 [156]、钢轨表面缺陷检测 [157]和医学图像诊断 [25], [158]。

除了多类别分类外，长尾学习还应用于基于人工任务 [37], [78]（即VOC-LT和COCO-LT）和真实任务的多标签分类，包括web图像分类 [77]、人脸属性分类 [85]和布料属性分类 [85]。

### 5.2 图像检测和分割

目标检测和实例分割在长尾学习社区中引起了越来越多的关注 [19], [35], [74], [80], [84], [126], [159], [160]，其中大多数现有研究都是基于LVIS和COCO进行的。除了这些广泛使用的基准之外，还探索了许多其他应用，包括城市场景理解 [26], [161]、无人机检测 [27]、点云分割 [162], [163]。

### 5.3 视觉关系学习

视觉关系学习对于图像理解非常重要，在长尾学习社区中受到越来越多的关注。一个重要的应用是长尾场景图生成 [118], [164]。在未来，长尾视觉问答和图像字幕是值得探索的 [165], [166]。

## 6 未来方向

在本节中，我们从方法创新和任务创新两个角度确定了深度长尾学习未来的几个研究方向。

### 6.1 新方法

我们首先讨论了深度长尾学习方法的几个潜在的创新方向。

不使用标签频率的类别重平衡。一些现实世界中的长尾任务，例如多标签分类或目标检测，除了类别不平衡之外，还



可能遇到另一个问题，即标签同现。具体而言，标签同现表明了头部类标签经常与尾部类标签同时出现的情况，这可能会使模型训练过程中的不平衡程度发生偏差，并且难以获得准确的标签频率。考虑到这个问题，现有的基于标签频率的类别重平衡方法往往会失败。这个问题在长尾学习中仍是一个悬而未决的问题。

使用未标记数据进行迁移学习。长尾学习的一个关键挑战是缺乏足够的尾部类样本。从其他未标记样本中迁移知识是一种可行的解决方案，例如自监督学习、知识蒸馏和自训练。然而，现有的迁移方法可能无法很好地处理长尾学习。例如，CReST [107]发现监督训练模型在长尾图像分类中通常对尾部类具有较高的精度（precision），因此建议选择更多的尾部类数据进行伪标记和模型训练。然而，这样的发现可能不适用于长尾目标检测或多标签分类。因此，如何更好地利用未标记数据进行长尾学习值得进一步探索。

多任务数据增强。现有的长尾方法通常是针对特定任务设计的，例如图像分类或图像检测。然而，由于不同任务之间的差异，针对特定任务的现有方法可能无法处理其他任务，导致方法泛化能力较差。考虑到数据增强是所有可视化任务的基础，因此，设计更好的基于数据增强的能够同时解决多个长尾任务的长尾方法是很有价值的。

提高所有类别的集成学习。大多数现有的长尾方法以头部类性能为代价来提高尾部类性能。一种解决方案是集成学习，它利用不同专家的不同专业知识，在头部和尾部类别之间取得更好的权衡，从而在长尾学习方面取得最先进的性能 [30]。由于所有类别潜在的性能改进，集成学习将是未来研究的一个有希望的方向。

## 6.2 新任务设置

除了方法创新之外，还有一些新的长尾学习任务设置有待解决。

测试集不可知的长尾学习。现有的长尾学习方法通常假设测试类别分布均衡。然而，实际的测试分布往往违反了这一假设（例如，长尾甚至反向长尾），这可能导致现有方法在实际应用中失效。为了克服这一限制，LADE [31]放松了这一假设，假设测试类别分布可以任意倾斜，但测试分布的先验信息可用。之后，TADE [30]进一步创新了这个任务，其中测试类别分布不仅是任意倾斜的，而且是未知的。除了类别不平衡之外，这项任务还带来了另一个挑战，即在训练样本和测试样本之间的未确定类别分布移动。

开放集长尾学习。现实世界的样本通常具有长尾和开放的类别分布。受此启发，开放集长尾学习（open-set long-tailed learning）[15], [94]寻求从长尾数据中学习，并在包括头部类、尾部类和开放类的平衡测试集上优化分类精度。有两个主要挑战：（1）如何在头部类和尾部类之间共享视觉知识；（2）如何减少尾部类和开放类之间的混淆。

联邦长尾学习。现有的长尾学习研究通常假设所有的训练样本都可以在模型训练期间被访问。然而，在实际应用中，长尾训练数据可能分布在许多移动设备或物联网上 [167]，这需要对深度模型进行分散训练。这种任务设置被称为联邦长尾学习，它有两个主要挑战：（1）长尾类别不平衡；（2）不同客户端的局部数据之间的未知类别分布移动。

类别增量长尾学习。在实际应用中，长尾数据可能以连续和类别增量的方式出现 [77], [168]。为了应对这种情况，类别增量长尾学习旨在从类别增量长尾数据中学习深层模型，这一任务面临两个关键挑战：（1）当不同的类别依次出现时，如何处理长尾类别不平衡，并且模型没有关于类别以及标签频率的未来输入信息；（2）如何克服在学习新类别时对先前类别知识的灾难性遗忘。这种任务设置也可以称为连续长尾学习。

多领域长尾学习。当前的长尾方法通常假设所有长尾样本来自相同的数据边缘分布。然而，在实践中，长尾数据也可能来自具有不同数据分布的不同领域 [28], [169]，如DomainNet数据集 [170]。基于此，多领域长尾学习旨在同时处理类别不平衡和领域分布转移。一个更具挑战性的问题可能是不同领域之间类别不平衡的不一致性。换句话说，不同的领域可能具有不同的类别分布，这进一步扩大了多领域长尾学习中的领域转移。

鲁棒长尾学习。真实世界的长尾样本也可能遭受图像噪声 [105], [171] 或标签噪声 [140], [145]。然而，大多数长尾方法都假设所有图像和标签都是干净的，这导致实际应用中的模型鲁棒性较差。这个问题对于尾部类来说尤其严重，因为它们的训练样本非常有限。受此启发，鲁棒长尾学习寻求在处理类别不平衡的同时提高模型的鲁棒性。

长尾回归。现有的长尾视觉学习研究大多集中在分类、检测和分割上，这些任务具有带类别索引的离散标签。然而，许多任务涉及连续标签，其中类别之间不存在硬分类边界。基于此，长尾回归 [172]旨在处理具有连续标签空间的长尾学习。在这一任务中，如何同时解决长尾类别不平衡和处理某些标签潜在的数据缺失仍是一个悬而未决的问题。

长尾视频学习。现有的深度长尾学习研究大多集中在图像层面，而忽视了视频领域也存在长尾类别不平衡的问题。考虑到视频数据中额外的时间维度，长尾视频学习应该比长尾图像学习更困难。得益于最近发布的VideoLT数据集 [38]，在不久的将来可以进行长尾视频学习的探索。

## 7 结论

在这篇综述中，我们根据类别重平衡、信息增强和模块改进的分类方式，广泛回顾了2021年中之前提出的经典的深度长尾学习方法。我们对几种最先进的长尾方法进行了实证分析，根据一个新提出的相对准确度指标，评估了它们在多大程度上解决了类别不平衡的问题。随后，我们讨论了长尾学习的主要应用场景，并确定了在方法和任务设置两个方面的潜在创新方向。我们希望，这一及时的综述不仅能为研究人员和社区提供对长尾学习的更好理解，而且还能促进未来的研究。

## REFERENCES

- [1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, may 2015. [Online]. Available: <https://doi.org/10.1038/2Fnature14539>
- [2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [3] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–13, 2018. [Online]. Available: <https://doi.org/10.1155/2F2018/2F7068349>
- [4] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, feb 2016. [Online]. Available: <https://doi.org/10.1109/2Ftpami.2015.2439281>
- [5] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3365–3387, oct 2021. [Online]. Available: <https://doi.org/10.1109/2Ftpami.2020.2982166>
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, may 2017. [Online]. Available: <https://doi.org/10.1145/2F3065386>

- [7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, jun 2017. [Online]. Available: <https://doi.org/10.1109/2Ftpami.2016.2577031>
- [8] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, apr 2017. [Online]. Available: <https://doi.org/10.1109/2Ftpami.2016.2572683>
- [9] Y. Bengio, Y. Lecun, and G. Hinton, "Deep learning for AI," *Communications of the ACM*, vol. 64, no. 7, pp. 58–65, jul 2021. [Online]. Available: <https://doi.org/10.1145/2F3448250>
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2016. [Online]. Available: <https://doi.org/10.1109/2Fcvpr.2016.90>
- [11] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," *Advances in neural information processing systems*, vol. 26, 2013.
- [12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2014. [Online]. Available: <https://doi.org/10.1109/2Fcvpr.2014.81>
- [13] B. Kang, Y. Li, S. Xie, Z. Yuan, and J. Feng, "Exploring balanced feature spaces for representation learning," in *International Conference on Learning Representations*, 2020.
- [14] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit, and S. Kumar, "Long-tail learning via logit adjustment," *arXiv preprint arXiv:2007.07314*, 2020.
- [15] Z. Liu, Z. Miao, X. Zhan, J. Wang, B. Gong, and S. X. Yu, "Large-scale long-tailed recognition in an open world," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2019. [Online]. Available: <https://doi.org/10.1109/2Fcvpr.2019.00264>
- [16] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2019. [Online]. Available: <https://doi.org/10.1109/2Fcvpr.2019.00949>
- [17] X. Wang, L. Lian, Z. Miao, Z. Liu, and S. X. Yu, "Long-tailed recognition by routing diverse distribution-aware experts," *arXiv preprint arXiv:2010.01809*, 2020.
- [18] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," *Advances in neural information processing systems*, vol. 32, 2019.
- [19] J. Tan, C. Wang, B. Li, Q. Li, W. Ouyang, C. Yin, and J. Yan, "Equalization loss for long-tailed object recognition," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109/2Fcvpr42600.2020.01168>
- [20] V. Vapnik, "Principles of risk minimization for learning theory," *Advances in neural information processing systems*, vol. 4, 1991.
- [21] X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, "Range loss for deep face recognition with long-tailed training data," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, oct 2017. [Online]. Available: <https://doi.org/10.1109/2Ficcv.2017.578>
- [22] D. Cao, X. Zhu, X. Huang, J. Guo, and Z. Lei, "Domain balancing: Face recognition on long-tailed domains," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109/2Fcvpr42600.2020.00571>
- [23] G. V. Horn, O. M. Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, and S. Belongie, "The iNaturalist species classification and detection dataset," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2018. [Online]. Available: <https://doi.org/10.1109/2Fcvpr.2018.00914>
- [24] Z. Miao, Z. Liu, K. M. Gaynor, M. S. Palmer, S. X. Yu, and W. M. Getz, "Iterative human and automated identification of wildlife images," *Nature Machine Intelligence*, vol. 3, no. 10, pp. 885–895, oct 2021. [Online]. Available: <https://doi.org/10.1038/2Fs42256-021-00393-0>
- [25] L. Ju, X. Wang, L. Wang, T. Liu, X. Zhao, T. Drummond, D. Mahapatra, and Z. Ge, "Relational subsets knowledge distillation for long-tailed retinal diseases recognition," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Springer International Publishing, 2021, pp. 3–12. [Online]. Available: [https://doi.org/10.1007/2F978-3-030-87237-3\\_1](https://doi.org/10.1007/2F978-3-030-87237-3_1)
- [26] R. He, J. Yang, and X. Qi, "Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109/2Ficcv48922.2021.00685>
- [27] W. Yu, T. Yang, and C. Chen, "Towards resolving the challenge of long-tail distribution in UAV images for object detection," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, jan 2021. [Online]. Available: <https://doi.org/10.1109/2Fwacv48630.2021.00330>
- [28] M. A. Jamal, M. Brown, M.-H. Yang, L. Wang, and B. Gong, "Rethinking class-balanced methods for long-tailed visual recognition from a domain adaptation perspective," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109/2Fcvpr42600.2020.00763>
- [29] S. Zhang, Z. Li, S. Yan, X. He, and J. Sun, "Distribution alignment: A unified framework for long-tail visual recognition," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109/2Fcvpr46437.2021.00239>
- [30] Y. Zhang, B. Hooi, L. Hong, and J. Feng, "Test-agnostic long-tailed recognition by test-time aggregating diverse experts with self-supervision," *arXiv preprint arXiv:2107.09249*, 2021.
- [31] Y. Hong, S. Han, K. Choi, S. Seo, B. Kim, and B. Chang, "Disentangling label distribution for long-tailed visual recognition," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109/2Fcvpr46437.2021.00656>
- [32] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, "Decoupling representation and classifier for long-tailed recognition," *arXiv preprint arXiv:1910.09217*, 2019.
- [33] C. Feng, Y. Zhong, and W. Huang, "Exploring classification equilibrium in long-tailed object detection," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109/2Ficcv48922.2021.00340>
- [34] T. Wang, Y. Li, B. Kang, J. Li, J. Liew, S. Tang, S. Hoi, and J. Feng, "The devil is in classification: A simple framework for long-tail instance segmentation," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 728–744. [Online]. Available: [https://doi.org/10.1007/2F978-3-030-58568-6\\_43](https://doi.org/10.1007/2F978-3-030-58568-6_43)
- [35] Z. Weng, M. G. Ogut, S. Limonchik, and S. Yeung, "Unsupervised discovery of the long-tail in instance segmentation using hierarchical self-supervision," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109/2Fcvpr46437.2021.00263>
- [36] A. Gupta, P. Dollar, and R. Girshick, "LVIS: A dataset for large vocabulary instance segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2019. [Online]. Available: <https://doi.org/10.1109/2Fcvpr.2019.00550>
- [37] T. Wu, Q. Huang, Z. Liu, Y. Wang, and D. Lin, "Distribution-balanced loss for multi-label classification in long-tailed datasets," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 162–178. [Online]. Available: [https://doi.org/10.1007/2F978-3-030-58548-8\\_10](https://doi.org/10.1007/2F978-3-030-58548-8_10)
- [38] X. Zhang, Z. Wu, Z. Weng, H. Fu, J. Chen, Y.-G. Jiang, and L. Davis, "VideoLT: Large-scale long-tailed video recognition," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109/2Ficcv48922.2021.00786>
- [39] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2009. [Online]. Available: <https://doi.org/10.1109/2Fcvpr.2009.5206848>
- [40] A. Krizhevsky, G. Hinton et al., "Learning multiple layers of features from tiny images," 2009.
- [41] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," *Advances in neural information processing systems*, vol. 27, 2014.



- [42] M. Everingham, S. M. A. Eslami, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, jun 2014. [Online]. Available: <https://doi.org/10.1007%2F11263-014-0733-5>
- [43] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Computer Vision – ECCV 2014*. Springer International Publishing, 2014, pp. 740–755. [Online]. Available: [https://doi.org/10.1007%2F978-3-319-10602-1\\_48](https://doi.org/10.1007%2F978-3-319-10602-1_48)
- [44] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2019. [Online]. Available: <https://doi.org/10.1109%2FCvpr.2019.00075>
- [45] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jul 2017. [Online]. Available: <https://doi.org/10.1109%2FCvpr.2017.634>
- [46] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jul 2017. [Online]. Available: <https://doi.org/10.1109%2FCvpr.2017.106>
- [47] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, feb 2020. [Online]. Available: <https://doi.org/10.1109%2FTpami.2018.2844175>
- [48] B. Zhou, Q. Cui, X.-S. Wei, and Z.-M. Chen, "BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109%2FCvpr42600.2020.00974>
- [49] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, sep 2009. [Online]. Available: <https://doi.org/10.1109%2FTkde.2008.239>
- [50] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Advances in neural information processing systems*, vol. 30, 2017.
- [51] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2018. [Online]. Available: <https://doi.org/10.1109%2FCvpr.2018.00131>
- [52] Q. Sun, Y. Liu, T.-S. Chua, and B. Schiele, "Meta-transfer learning for few-shot learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2019. [Online]. Available: <https://doi.org/10.1109%2FCvpr.2019.00049>
- [53] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM computing surveys (csur)*, vol. 53, no. 3, pp. 1–34, 2020.
- [54] D. Krueger, E. Caballero, J.-H. Jacobsen, A. Zhang, J. Binas, D. Zhang, R. Le Priol, and A. Courville, "Out-of-distribution generalization via risk extrapolation (rex)," in *International Conference on Machine Learning*. PMLR, 2021, pp. 5815–5826.
- [55] Z. Shen, J. Liu, Y. He, X. Zhang, R. Xu, H. Yu, and P. Cui, "Towards out-of-distribution generalization: A survey," *arXiv preprint arXiv:2108.13624*, 2021.
- [56] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, feb 2011. [Online]. Available: <https://doi.org/10.1109%2FTnn.2010.2091281>
- [57] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jul 2017. [Online]. Available: <https://doi.org/10.1109%2FCvpr.2017.316>
- [58] Y. Zhang, H. Chen, Y. Wei, P. Zhao, J. Cao, X. Fan, X. Lou, H. Liu, J. Hou, X. Han, J. Yao, Q. Wu, M. Tan, and J. Huang, "From whole slide imaging to microscopy: Deep microscopy adaptation network for histopathology cancer image classification," in *Lecture Notes in Computer Science*. Springer International Publishing, 2019, pp. 360–368. [Online]. Available: [https://doi.org/10.1007%2F978-3-030-32239-7\\_40](https://doi.org/10.1007%2F978-3-030-32239-7_40)
- [59] Y. Zhang, Y. Wei, Q. Wu, P. Zhao, S. Niu, J. Huang, and M. Tan, "Collaborative unsupervised domain adaptation for medical image diagnosis," *IEEE Transactions on Image Processing*, vol. 29, pp. 7834–7844, 2020. [Online]. Available: <https://doi.org/10.1109%2FTip.2020.3006377>
- [60] Z. Qiu, Y. Zhang, H. Lin, S. Niu, Y. Liu, Q. Du, and M. Tan, "Source-free domain adaptation via avatar prototype generation and adaptation," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, aug 2021. [Online]. Available: <https://doi.org/10.24963%2Fijcai.2021%2F402>
- [61] H. Wu, H. Zhu, Y. Yan, J. Wu, Y. Zhang, and M. K. Ng, "Heterogeneous domain adaptation by information capturing and distribution matching," *IEEE Transactions on Image Processing*, vol. 30, pp. 6364–6376, 2021. [Online]. Available: <https://doi.org/10.1109%2FTip.2021.3094137>
- [62] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, "Deeper, broader and artier domain generalization," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, oct 2017. [Online]. Available: <https://doi.org/10.1109%2Ficcv.2017.591>
- [63] H. Li, S. J. Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2018. [Online]. Available: <https://doi.org/10.1109%2FCvpr.2018.00566>
- [64] L. Neal, M. Olson, X. Fern, W.-K. Wong, and F. Li, "Open set learning with counterfactual images," in *Computer Vision – ECCV 2018*. Springer International Publishing, 2018, pp. 620–635. [Online]. Available: [https://doi.org/10.1007%2F978-3-030-01231-1\\_38](https://doi.org/10.1007%2F978-3-030-01231-1_38)
- [65] Y. Fu, X. Wang, H. Dong, Y.-G. Jiang, M. Wang, X. Xue, and L. Sigal, "Vocabulary-informed zero-shot and open-set learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 12, pp. 3136–3152, dec 2020. [Online]. Available: <https://doi.org/10.1109%2FTpami.2019.2922175>
- [66] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, jun 2002. [Online]. Available: <https://doi.org/10.1613%2Fjair.953>
- [67] A. Estabrooks, T. Jo, and N. Japkowicz, "A multiple resampling method for learning from imbalanced data sets," *Computational Intelligence*, vol. 20, no. 1, pp. 18–36, feb 2004. [Online]. Available: <https://doi.org/10.1111%2Fj.0824-7935.2004.t01-1-00228.x>
- [68] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2005, pp. 878–887. [Online]. Available: [https://doi.org/10.1007%2F11538059\\_91](https://doi.org/10.1007%2F11538059_91)
- [69] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 539–550, apr 2009. [Online]. Available: <https://doi.org/10.1109%2FTsmcb.2008.2007853>
- [70] Z. Zhang and T. Pfister, "Learning fast sample re-weighting without reward data," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.00076>
- [71] D. Mahajan, R. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li, A. Bharambe, and L. van der Maaten, "Exploring the limits of weakly supervised pretraining," in *Computer Vision – ECCV 2018*. Springer International Publishing, 2018, pp. 185–201. [Online]. Available: [https://doi.org/10.1007%2F978-3-030-01216-8\\_12](https://doi.org/10.1007%2F978-3-030-01216-8_12)
- [72] Y. Wang, W. Gan, J. Yang, W. Wu, and J. Yan, "Dynamic curriculum learning for imbalanced data classification," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2019. [Online]. Available: <https://doi.org/10.1109%2Ficcv.2019.00512>
- [73] J. Ren, C. Yu, X. Ma, H. Zhao, S. Yi et al., "Balanced meta-softmax for long-tailed visual recognition," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4175–4186, 2020.
- [74] Y. Zang, C. Huang, and C. C. Loy, "FASA: Feature augmentation and sampling adaptation for long-tailed instance segmentation," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.00344>
- [75] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *2016 IEEE*

- Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, jun 2016. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2016.580>
- [76] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," *arXiv preprint arXiv:1703.07737*, 2017.
- [77] C. D. Kim, J. Jeong, and G. Kim, "Imbalanced continual learning with partitioning reservoir sampling," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 411–428. [Online]. Available: [https://doi.org/10.1007%2F978-3-030-58601-0\\_25](https://doi.org/10.1007%2F978-3-030-58601-0_25)
- [78] H. Guo and S. Wang, "Long-tailed multi-label visual recognition by collaborative training on uniform and re-balanced samplings," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.01484>
- [79] B. Liu, H. Li, H. Kang, G. Hua, and N. Vasconcelos, "GistNet: a geometric structure transfer network for long-tailed recognition," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.00810>
- [80] Y. Li, T. Wang, B. Kang, S. Tang, C. Wang, J. Li, and J. Feng, "Overcoming classifier imbalance for long-tail object detection with balanced group softmax," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109%2Fcvpr42600.2020.01100>
- [81] X. Hu, Y. Jiang, K. Tang, J. Chen, C. Miao, and H. Zhang, "Learning to segment the tail," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109%2Fcvpr42600.2020.01406>
- [82] J. Cai, Y. Wang, and J.-N. Hwang, "ACE: Ally complementary experts for solving long-tailed recognition in one-shot," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.00018>
- [83] W. Ouyang, X. Wang, C. Zhang, and X. Yang, "Factors in finetuning deep model for object detection with long-tail distribution," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2016. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2016.100>
- [84] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, oct 2017. [Online]. Available: <https://doi.org/10.1109%2Ficcv.2017.324>
- [85] Q. Dong, S. Gong, and X. Zhu, "Class rectification hard mining for imbalanced deep learning," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, oct 2017. [Online]. Available: <https://doi.org/10.1109%2Ficcv.2017.205>
- [86] Y.-X. Wang, D. Ramanan, and M. Hebert, "Learning to model the tail," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [87] Y. Cui, Y. Song, C. Sun, A. Howard, and S. Belongie, "Large scale fine-grained categorization and domain-specific transfer learning," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2018. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2018.00432>
- [88] S. Khan, M. Hayat, S. W. Zamir, J. Shen, and L. Shao, "Striking the right balance with uncertainty," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2019. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2019.00019>
- [89] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, "Feature transfer learning for face recognition with under-represented data," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2019. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2019.00585>
- [90] Y. Zhong, W. Deng, M. Wang, J. Hu, J. Peng, X. Tao, and Y. Huang, "Unequal-training for deep face recognition with long-tailed noisy data," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2019. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2019.00800>
- [91] J. Shu, Q. Xie, L. Yi, Q. Zhao, S. Zhou, Z. Xu, and D. Meng, "Meta-weight-net: Learning an explicit mapping for sample weighting," *Advances in neural information processing systems*, vol. 32, 2019.
- [92] J. Kim, J. Jeong, and J. Shin, "M2m: Imbalanced classification via major-to-minor translation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109%2Fcvpr42600.2020.01391>
- [93] J. Liu, Y. Sun, C. Han, Z. Dou, and W. Li, "Deep representation learning on long-tailed data: A learnable embedding augmentation perspective," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109%2Fcvpr42600.2020.00304>
- [94] L. Zhu and Y. Yang, "Inflated episodic memory with region self-attention for long-tailed visual recognition," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109%2Fcvpr42600.2020.00440>
- [95] P. Chu, X. Bian, S. Liu, and H. Ling, "Feature space augmentation for long-tailed data," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 694–710. [Online]. Available: [https://doi.org/10.1007%2F978-3-030-58526-6\\_41](https://doi.org/10.1007%2F978-3-030-58526-6_41)
- [96] L. Xiang, G. Ding, and J. Han, "Learning from multiple experts: Self-paced knowledge distillation for long-tailed classification," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 247–263. [Online]. Available: [https://doi.org/10.1007%2F978-3-030-58558-7\\_15](https://doi.org/10.1007%2F978-3-030-58558-7_15)
- [97] T.-Y. Wu, P. Morgado, P. Wang, C.-H. Ho, and N. Vasconcelos, "Solving long-tailed recognition with deep realistic taxonomic classifier," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 171–189. [Online]. Available: [https://doi.org/10.1007%2F978-3-030-58598-3\\_11](https://doi.org/10.1007%2F978-3-030-58598-3_11)
- [98] J. Tian, Y.-C. Liu, N. Glaser, Y.-C. Hsu, and Z. Kira, "Posterior re-calibration for imbalanced datasets," *Advances in Neural Information Processing Systems*, vol. 33, pp. 8101–8113, 2020.
- [99] K. Tang, J. Huang, and H. Zhang, "Long-tailed classification by keeping the good and removing the bad momentum causal effect," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1513–1524, 2020.
- [100] Y. Yang and Z. Xu, "Rethinking the value of labels for improving class-imbalanced learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 19 290–19 301, 2020.
- [101] J. Tan, X. Lu, G. Zhang, C. Yin, and Q. Li, "Equalization loss v2: A new gradient balance approach for long-tailed object detection," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.00173>
- [102] J. Wang, W. Zhang, Y. Zang, Y. Cao, J. Pang, T. Gong, K. Chen, Z. Liu, C. C. Loy, and D. Lin, "Seesaw loss for long-tailed instance segmentation," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.00957>
- [103] T. Wang, Y. Zhu, C. Zhao, W. Zeng, J. Wang, and M. Tang, "Adaptive class suppression loss for long-tail object detection," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.00312>
- [104] Z. Deng, H. Liu, Y. Wang, C. Wang, Z. Yu, and X. Sun, "PML: Progressive margin loss for long-tailed age classification," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.01036>
- [105] T. Wu, Z. Liu, Q. Huang, Y. Wang, and D. Lin, "Adversarial robustness under long-tailed distribution," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.00855>
- [106] Z. Zhong, J. Cui, S. Liu, and J. Jia, "Improving calibration for long-tailed recognition," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.01622>
- [107] C. Wei, K. Sohn, C. Mellina, A. Yuille, and F. Yang, "CReST: A class-rebalancing self-training framework for imbalanced semi-supervised learning," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.01071>
- [108] S. Changpinyo, P. Sharma, N. Ding, and R. Soricut, "Conceptual 12m: Pushing web-scale image-text pre-training to recognize long-tail visual concepts," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021.



- [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.00356>
- [109] J. Wang, T. Lukasiewicz, X. Hu, J. Cai, and Z. Xu, "RSG: A simple but effective module for learning imbalanced datasets," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.00378>
- [110] S. Li, K. Gong, C. H. Liu, Y. Wang, F. Qiao, and X. Cheng, "MetaSAug: Meta semantic augmentation for long-tailed visual recognition," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.00517>
- [111] P. Wang, K. Han, X.-S. Wei, L. Zhang, and L. Wang, "Contrastive learning based hybrid networks for long-tailed image classification," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvpr46437.2021.00100>
- [112] S. Park, J. Lim, Y. Jeon, and J. Y. Choi, "Influence-balanced loss for imbalanced visual classification," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.00077>
- [113] T. Li, L. Wang, and G. Wu, "Self supervision to distillation for long-tailed visual recognition," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.00067>
- [114] Y.-Y. He, J. Wu, and X.-S. Wei, "Distilling virtual examples for long-tailed recognition," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.00030>
- [115] C. Zhang, T.-Y. Pan, Y. Li, H. Hu, D. Xuan, S. Changpinyo, B. Gong, and W.-L. Chao, "MosaicOS: A simple and effective use of object-centric images for long-tailed object detection," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.00047>
- [116] J. Cui, Z. Zhong, S. Liu, B. Yu, and J. Jia, "Parametric contrastive learning," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.00075>
- [117] D. Samuel and G. Chechik, "Distributional robustness loss for long-tail learning," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.00936>
- [118] A. Desai, T.-Y. Wu, S. Tripathi, and N. Vasconcelos, "Learning of visual relations: The devil is in the tails," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.01512>
- [119] C. Elkan, "The foundations of cost-sensitive learning," in *International joint conference on artificial intelligence*, vol. 17, no. 1. Lawrence Erlbaum Associates Ltd, 2001, pp. 973–978.
- [120] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, jan 2006. [Online]. Available: <https://doi.org/10.1109%2Ftkde.2006.17>
- [121] Y. Sun, M. S. Kamel, A. K. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognition*, vol. 40, no. 12, pp. 3358–3378, dec 2007. [Online]. Available: <https://doi.org/10.1016%2Fj.patcog.2007.04.009>
- [122] P. Zhao, Y. Zhang, M. Wu, S. C. H. Hoi, M. Tan, and J. Huang, "Adaptive cost-sensitive online classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 2, pp. 214–228, feb 2019. [Online]. Available: <https://doi.org/10.1109%2Ftkde.2018.2826011>
- [123] Y. Zhang, P. Zhao, J. Cao, W. Ma, J. Huang, Q. Wu, and M. Tan, "Online adaptive asymmetric active learning for budgeted imbalanced data," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 2768–2777.
- [124] Y. Zhang, P. Zhao, S. Niu, Q. Wu, J. Cao, J. Huang, and M. Tan, "Online adaptive asymmetric active learning with limited budgets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp. 2680–2692, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Ftkde.2019.2955078>
- [125] H.-J. Ye, H.-Y. Chen, D.-C. Zhan, and W.-L. Chao, "Identifying and compensating for feature deviation in imbalanced deep learning," *arXiv preprint arXiv:2001.01385*, 2020.
- [126] T.-I. Hsieh, E. Robb, H.-T. Chen, and J.-B. Huang, "Droploss for long-tail instance segmentation," in *AAAI*, vol. 3, no. 6, 2021, p. 15.
- [127] F. Wang, J. Cheng, W. Liu, and H. Liu, "Additive margin softmax for face verification," *IEEE Signal Processing Letters*, vol. 25, no. 7, pp. 926–930, jul 2018. [Online]. Available: <https://doi.org/10.1109%2Fslsp.2018.2822810>
- [128] V. Koltchinskii and D. Panchenko, "Empirical margin distributions and bounding the generalization error of combined classifiers," *The Annals of Statistics*, vol. 30, no. 1, feb 2002. [Online]. Available: <https://doi.org/10.1214%2Faoas%2F015362183>
- [129] F. Provost, "Machine learning from imbalanced data sets 101," in *Proceedings of the AAAI'2000 workshop on imbalanced data sets*, vol. 68, no. 2000. AAAI Press, 2000, pp. 1–3.
- [130] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, oct 2010. [Online]. Available: <https://doi.org/10.1109%2Ftkde.2009.191>
- [131] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Artificial Neural Networks and Machine Learning – ICANN 2018*. Springer International Publishing, 2018, pp. 270–279. [Online]. Available: [https://doi.org/10.1007%2F978-3-030-01424-7\\_27](https://doi.org/10.1007%2F978-3-030-01424-7_27)
- [132] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2016. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2016.319>
- [133] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, "Why does unsupervised pre-training help deep learning?" in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 201–208.
- [134] K. He, R. Girshick, and P. Dollar, "Rethinking ImageNet pre-training," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2019. [Online]. Available: <https://doi.org/10.1109%2Ficcv.2019.00502>
- [135] D. Hendrycks, K. Lee, and M. Mazeika, "Using pre-training can improve model robustness and uncertainty," in *International Conference on Machine Learning*. PMLR, 2019, pp. 2712–2721.
- [136] B. Zoph, G. Ghiasi, T.-Y. Lin, Y. Cui, H. Liu, E. D. Cubuk, and Q. Le, "Rethinking pre-training and self-training," *Advances in neural information processing systems*, vol. 33, pp. 3833–3845, 2020.
- [137] Y. Zhang, B. Hooi, D. Hu, J. Liang, and J. Feng, "Unleashing the power of contrastive self-supervised visual models via contrast-regularized fine-tuning," *Advances in Neural Information Processing Systems*, vol. 34, 2021.
- [138] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109%2Fcvpr42600.2020.00975>
- [139] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," *arXiv preprint arXiv:1803.07728*, 2018.
- [140] S. Karthik, J. Revaud, and B. Chidlovskii, "Learning from long-tailed data with noisy labels," *arXiv preprint arXiv:2108.11096*, 2021.
- [141] G. Hinton, O. Vinyals, J. Dean *et al.*, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, vol. 2, no. 7, 2015.
- [142] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *International Journal of Computer Vision*, vol. 129, no. 6, pp. 1789–1819, mar 2021. [Online]. Available: <https://doi.org/10.1007%2F0151263-021-01453-z>
- [143] X. J. Zhu, "Semi-supervised learning literature survey," 2005.
- [144] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised self-training of object detection models," in *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05) - Volume 1*. IEEE, jan 2005. [Online]. Available: <https://doi.org/10.1109%2Ffacvmot.2005.107>
- [145] T. Wei, J.-X. Shi, W.-W. Tu, and Y.-F. Li, "Robust long-tailed learning under label noise," *arXiv preprint arXiv:2108.11569*, 2021.

- [146] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017.
- [147] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, jul 2019. [Online]. Available: <https://doi.org/10.1186%2Fs40537-019-0197-0>
- [148] H.-P. Chou, S.-C. Chang, J.-Y. Pan, W. Wei, and D.-C. Juan, "Remix: Rebalanced mixup," in *Computer Vision – ECCV 2020 Workshops*. Springer International Publishing, 2020, pp. 95–110. [Online]. Available: [https://doi.org/10.1007%2F978-3-030-65414-6\\_9](https://doi.org/10.1007%2F978-3-030-65414-6_9)
- [149] Y. Wang, X. Pan, S. Song, H. Zhang, G. Huang, and C. Wu, "Implicit semantic data augmentation for deep networks," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [150] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Operations Research*, vol. 58, no. 4-part-1, pp. 902–917, aug 2010. [Online]. Available: <https://doi.org/10.1287%2Fopre.1090.0795>
- [151] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, jan 1967. [Online]. Available: <https://doi.org/10.1109%2FTit.1967.1053964>
- [152] J. Cui, S. Liu, Z. Tian, Z. Zhong, and J. Jia, "ResLT: Residual learning for long-tailed recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2022. [Online]. Available: <https://doi.org/10.1109%2FTpami.2022.3174892>
- [153] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "RandAugment: Practical automated data augmentation with a reduced search space," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109%2Fcvprw50498.2020.00359>
- [154] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2019. [Online]. Available: <https://doi.org/10.1109%2Ficcv.2019.00612>
- [155] M. R. Keaton, R. J. Zaveri, M. Kovur, C. Henderson, D. A. Adjero, and G. Doretto, "Fine-grained visual classification of plant species in the wild: Object detection as a reinforced means of attention," *arXiv preprint arXiv:2106.02141*, 2021.
- [156] X. Jia, H. Yan, Y. Wu, X. Wei, X. Cao, and Y. Zhang, "An effective and robust detector for logo detection," *arXiv preprint arXiv:2108.00422*, 2021.
- [157] Z. Zhang, S. Yu, S. Yang, Y. Zhou, and B. Zhao, "Rail-5k: a real-world dataset for rail surface defects detection," *arXiv preprint arXiv:2106.14366*, 2021.
- [158] A. Galdran, G. Carneiro, and M. A. G. Ballester, "Balanced-MixUp for highly imbalanced medical image classification," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Springer International Publishing, 2021, pp. 323–333. [Online]. Available: [https://doi.org/10.1007%2F978-3-030-87240-3\\_31](https://doi.org/10.1007%2F978-3-030-87240-3_31)
- [159] T. Weyand, A. Araujo, B. Cao, and J. Sim, "Google landmarks dataset v2 – a large-scale benchmark for instance-level recognition and retrieval," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109%2Fcvpr42600.2020.00265>
- [160] J. Wu, L. Song, T. Wang, Q. Zhang, and J. Yuan, "Forest r-CNN: Large-vocabulary long-tailed object detection and instance segmentation," in *Proceedings of the 28th ACM International Conference on Multimedia*. ACM, oct 2020. [Online]. Available: <https://doi.org/10.1145%2F3394171.3413970>
- [161] J. Mao, M. Niu, C. Jiang, H. Liang, J. Chen, X. Liang, Y. Li, C. Ye, W. Zhang, Z. Li *et al.*, "One million scenes for autonomous driving: Once dataset," *arXiv preprint arXiv:2106.11037*, 2021.
- [162] Y. Zhang, Z. Zhou, P. David, X. Yue, Z. Xi, B. Gong, and H. Foroosh, "PolarNet: An improved grid representation for online LiDAR point clouds semantic segmentation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. [Online]. Available: <https://doi.org/10.1109%2Fcvpr42600.2020.00962>
- [163] X. Chen, C. Zhang, G. Lin, and J. Han, "Compositional prototype network with multi-view comparison for few-shot point cloud semantic segmentation," *arXiv preprint arXiv:2012.14255*, 2020.
- [164] N. Dhingra, F. Ritter, and A. Kunz, "BGT-net: Bidirectional GRU transformer network for scene graph generation," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, jun 2021. [Online]. Available: <https://doi.org/10.1109%2Fcvprw53098.2021.00244>
- [165] J. Chen, A. Agarwal, S. Abdelkarim, D. Zhu, and M. Elhoseiny, "Reltransformer: Balancing the visual relationship detection from local context, scene and memory," *arXiv preprint arXiv:2104.11934*, 2021.
- [166] Z. Li, E. Stengel-Eskin, Y. Zhang, C. Xie, Q. Tran, B. V. Durme, and A. Yuille, "Calibrating concepts and operations: Towards symbolic reasoning on real images," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2021. [Online]. Available: <https://doi.org/10.1109%2Ficcv48922.2021.01464>
- [167] M. Luo, F. Chen, D. Hu, Y. Zhang, J. Liang, and J. Feng, "No fear of heterogeneity: Classifier calibration for federated learning with non-iid data," *Advances in Neural Information Processing Systems*, vol. 34, 2021.
- [168] S. Niu, J. Wu, G. Xu, Y. Zhang, Y. Guo, P. Zhao, P. Wang, and M. Tan, "Adaxpert: Adapting neural architecture for growing data," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8184–8194.
- [169] Y. Zhang, S. Niu, Z. Qiu, Y. Wei, P. Zhao, J. Yao, J. Huang, Q. Wu, and M. Tan, "Covid-da: deep domain adaptation from typical pneumonia to covid-19," *arXiv preprint arXiv:2005.01577*, 2020.
- [170] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2019. [Online]. Available: <https://doi.org/10.1109%2Ficcv.2019.00149>
- [171] K. Cao, Y. Chen, J. Lu, N. Arechiga, A. Gaidon, and T. Ma, "Heteroskedastic and imbalanced deep learning with adaptive regularization," *arXiv preprint arXiv:2006.15766*, 2020.
- [172] Y. Yang, K. Zha, Y. Chen, H. Wang, and D. Katabi, "Delving into deep imbalanced regression," in *International Conference on Machine Learning*. PMLR, 2021, pp. 11 842–11 851.