

华中科技大学

人工智能与自动化学院

数字图像处理 - 课程设计

深度长尾学习研究综述

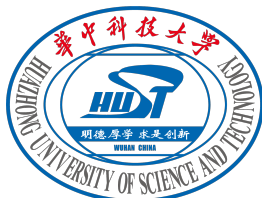
孟繁鹏 (U201914689)

自动化 1905 班

fanpengmeng@hust.edu.cn

指导老师：谭山

2022 年 7 月 10 日



课程设计说明

本课程设计中。笔者选取了**任务五**，并选择了深度长尾学习（Deep Long-tailed Learning）这一课题进行了调研、综述和探索。本课程设计中，笔者主要完成了以下工作：

- 阐述了深度长尾学习的概念，并详细介绍了深度学习中遇到的长尾分布问题。
- 对现有的（截至 2022 年 ECCV 会议）深度长尾学习方法进行了分类、归纳和总结；对其中代表性的方法进行了实验，验证其性能。
- 提出了一种参数自适应的多专家网络，实验验证其达到 SOTA 行列。此外，其强大的泛化能力为解决更广泛的不平衡分布和未知分布问题提供了可能。
(鉴于笔者将于年底投稿此方法，暂时未提供代码，如老师感兴趣可以持续关注和交流。)
- 归纳总结了深度长尾学习目前遇到的问题和未来的发展方向。

笔者已经附上了实验代码，大部分算法都参考了作者发布的官方代码。由于算力限制，笔者只在 CIFAR10-LT 数据集上进行了实验，如有算力条件可能会进行更多实验来验证算法的性能。实验代码已经开源在github上，为了展示实验结果，笔者一并上传了训练记录。

课程设计报告正文从下页起。

摘要

深度长尾学习是视觉识别中最具挑战性的问题之一，它旨在从大量遵循长尾类分布的图像中训练出表现良好的深度模型。在过去的十年中，深度学习已经成为解决通用视觉问题的一大主流模型。然而，具有长尾分布的类不平衡是实际视觉任务中常见的问题，往往限制了深度网络识别模型在现实应用中的实用性，因为它们容易偏向主导类，在尾部类上表现不佳。为了解决这一问题，近年来进行了大量的研究，在深度长尾学习领域取得了巨大进展。考虑到这一领域的快速发展，本文旨在全面介绍深度长尾学习的最新进展。具体来说，笔者将现有的深度长尾学习研究分为三个主要类别（即类重新平衡、信息增强和模块改进），并按照这个分类方法详细回顾这些方法。之后，笔者实证分析了几种最先进的方法，通过评估它们在多大程度上解决了长尾分布问题。在最后，笔者提出了一种新方法，采用自适应的多专家网络来解决长尾分布问题，甚至是更广泛的不平衡分布问题。鉴于作者将在年底发表该方法，不对该方法的细节进行过多阐述。最后，笔者调研了深度长尾学习的重要应用，并确定了几个有可能的研究方向。

目录

1 概述	3
2 主流框架	4
2.1 多专家重采样	4
2.2 集成学习	5
3 代表性方法介绍	6
3.1 BBN：双边分支网络	6
3.2 TLC：可解释的长尾分类	7
3.3 TADE：测试不可知的长尾识别	7
3.4 VL-LTR：类的视觉-语言表示	8
3.5 SAMEN：自适应的多专家网络	9
4 实验设置及结果	9
4.1 实验设置	9
4.2 实验参数	9
4.3 实验结果	9
4.4 实验构想	10
5 应用和未来	10
5.1 主要应用	10
5.1.1 图像分类长尾学习	10
5.1.2 图像检测和分割	10
5.1.3 视觉联系学习	10
5.2 未来方向	10
5.2.1 新方法	10
5.2.2 新任务	11
6 结论及感想	11
参考文献	13

1 概述

近十年来，深度学习在计算机视觉方面取得了令人难以置信的进展 [1]。深度神经网络具有学习高质量数据表示的强大能力，已成功地应用于许多视觉识别任务，包括图像分类、目标检测和语义分割等。然而在现实应用中，训练样本通常表现出长尾类分布，即其中一小部分类占据了大量的样本点，但其他类只与少数样本相关 [2] [3]。然而，这种训练样本数的类不平衡，使得基于深度网络的视觉模型的训练非常具有挑战性。如图1所示，训练后的模型很容易偏向于具有大量训练数据的头部类，导致数据有限的尾类的模型性能较差 [4]。因此，采用经验风险最小化训练的深度模型不能处理长尾类不平衡的真实应用。

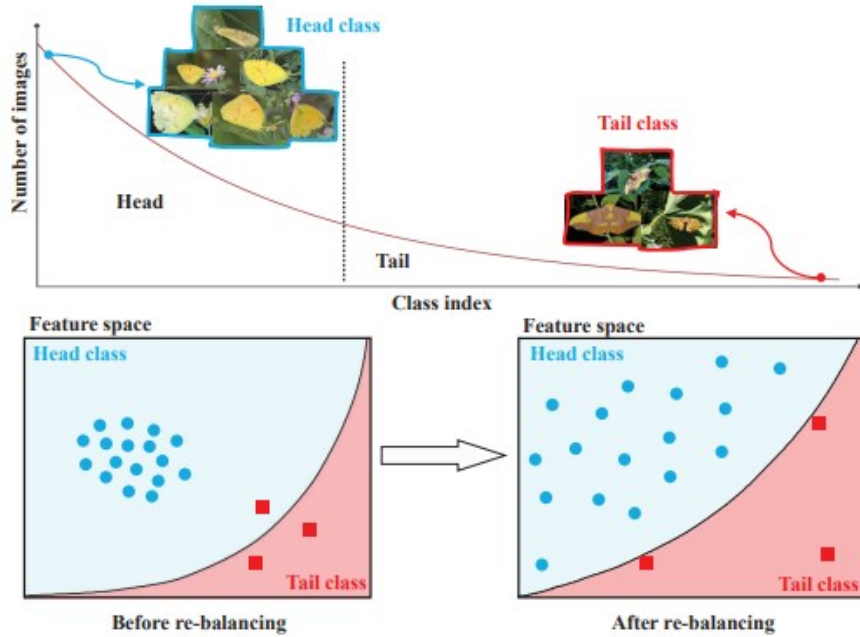


图 1: 视觉问题中的长尾分布问题

为了解决长尾阶级失衡问题，近年来开展了大规模的深度长尾学习研究 [2] [5] [6] [7] [8]。在工作 [6] 中，根据其主要采用的技术手段将其分为三大类，即类重平衡、信息增强和模块改进；这些类别可以进一步分为 9 个子类：重采样、代价敏感学习、逻辑调整、迁移学习、数据增强、表示学习、分类器设计、解耦训练和集成学习，如图2所示。根据这一分类法，笔者对现有的方法进行了全面的回顾，总结归纳了当前的主流算法框架。基于归纳的框架调研了如下几种算法，他们分别是不同时期下的 SOTA (state-of-art) 算法，并且有一定的革命性创新：

- BBN: 双边分支网络 [9]。富有创新性地构造了反采样分支，第一次尝试结合解耦训练和集成学习（融合学习）方法，发表于 CVPR2020。原文链接
- TLC: 可解释的长尾分类 [10]。通过专家和专业知识的设置，可解释地建立了多个子网络，并通过联合学习整合专家知识。发表于 CVPR2022。原文链接
- TADE: 测试不可知的长尾识别 [11]。通过在测试时自监督地聚合不同专家信息，实现对于不可知的测试集的精准预测。目前收录于 Arxiv。原文链接
- VL-LTR: 类的视觉-语言表示 [12]。创造性地引入了文本模态信息进行数据增强，目前收录于 Arxiv。原文链接

由于时间和篇幅限制，笔者只选取了重采样 + 集成学习算法详解，原因在于：（1）这些算法比较具有代表性，近些年取得了不错的效果；（2）学界认为这种思路将会是未来深度长尾学习的主流解决方法 [6]。

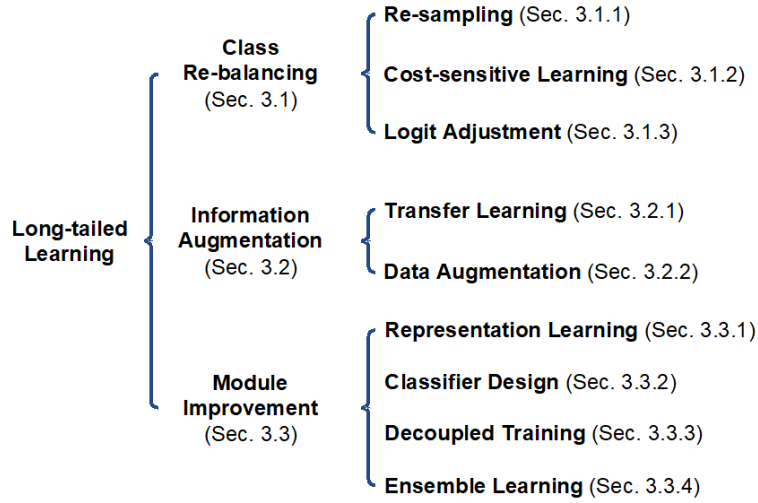


图 2: 长尾学习方法分类 [6]

2 主流框架

2.1 多专家重采样

类重平衡是深度长尾学习的主流范式，旨在在模型训练过程中平衡不同类的训练样本数量。而构造多个专家（分支）重采样是解决近年来最广泛使用的方法之一 [13] [14] [15] [16] [17] [18]。重新抽样的常见做法是随机过采样 (random over-sampling, ROS) 和随机欠采样 (random under-sampling, RUS)。为了重新平衡类，ROS 随机重复尾部类中的样本，而 RUS 随机丢弃头部类中的样本。然而，当类极度倾斜时，ROS 倾向于过拟合尾部类，而 RUS 倾向于降低头部类的模型性能。最近的长尾学习研究发展了各种抽样方法，而不是使用随机重抽样，包括类平衡重抽样和方案导向抽样。

类平衡重新采样。从解耦 [13] 开始，它通过经验评估了在长尾识别上的表示学习的各种抽样策略。具体来说，抽样策略包括实例平衡抽样、类平衡抽样、平方根抽样和逐步平衡抽样。在实例平衡抽样中，每个样本被抽样的概率相等，而在类平衡抽样中，每个类被选择的概率相等。此外，平方根抽样 [19] 是实例平衡抽样的一种变体，其中每个类的采样概率与相应类中样本大小的平方根有关。逐步平衡抽样 [13] 在实例平衡抽样和类平衡抽样之间逐步进行插值。

动态内容学习 (DCL) [20] 创造了一种新的内容策略来动态抽样数据以进行类重平衡。具体地说，随着训练的进行，从一个类中采样的实例越多，之后这个类的采样概率就会越低。根据这一想法，DCL 首先进行随机抽样来学习一般表示，然后根据类策略抽样更多的尾类实例，以处理长尾类不平衡。

平衡元 Softmax [21] 开发了一种基于元学习的抽样方法来估计不同类别的长尾学习的最优抽样率。具体来说，所提出的元学习方法是一种双层优化策略，通过在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) [22] 提出利用平衡元验证集上的模型分类损失（作为度量）来调整不同类的特征采样率，以便对代表不足的尾类进行更多的采样。

面向方案的抽样。使用一些特定的学习方案，如度量学习、集成学习等，来提高长尾学习的表现。例如，大边际局部嵌入 (LMLE) [23] 开发了一种新的五联体抽样方案，以学习保持类间边缘的高质量特征。

LMLE 提供了一个五胞胎采样器对四对对比对采样, 包括一个正对和三对负对。正对是最遥远的簇内样本, 而负对包括来自同一类的两个簇间样本 (一个是最近的, 一个是同一簇内最遥远的) 和最最近的类间样本。随后, LMLE 引入了一个五胞胎损失, 以鼓励采样的五胞胎遵循一个特定的距离顺序。通过这种方式, 学习到的表示不仅保留了类内集群之间的局部性, 而且还保留了类之间的区别。此外, 五联体丢失中的每个数据批都包含来自不同类的相同数量的样本, 用于类的重新平衡。

划分储层采样 (PRS) [24] 提出了一种基于重播的采样方法来处理连续的长尾学习。一个关键的挑战是, 重播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, 它可以动态地维护来自不同类的样本的运行统计信息。基于运行统计, PRS 可以动态调整不同类的内存大小和采样输入输出操作方案。

双边分支网络 (BBN) [9] 基于一种新的双边采样策略, 开发了两个网络分支 (即一个传统的学习分支和一个再平衡分支) 来处理类不平衡。具体来说, BBN 对传统分支采用均匀采样来模拟原始的长尾训练分布; 同时, BBN 对再平衡分支采用反向采样器对更多的尾类样本进行采样, 以提高尾类性能。最终的预测是两个网络分支的加权和。随后, 长尾多标签视觉识别 (LTML) [23] 扩展了双边分支网络来解决长尾多标签分类问题。几何结构转移 (GIST) [25] 也探索了这种从头到尾类的知识转移的双边抽样策略。

除了对双边分支进行抽样外, 平衡组 Softmax (BAGS) [26] 还提出根据每个类中的样本数量将类划分为几个平衡组, 其中每个组具有具有相似训练数据数量的类。随后, BAGS 使用不同的样本组来训练不同的分类头, 以便对具有相似训练数据的类执行 Softmax 操作, 从而避免由于不平衡而产生的严重偏差分类器。然后, 学习分割尾部 (LST) [27] 也将训练样本划分为几个平衡子集, 并基于类增量学习处理每个子集。为了解决类增量学习过程中的灾难性遗忘问题, LST 开发了一种类平衡的数据回复/采样策略, 该策略保持了一个相对平衡的样本集用于知识蒸馏。

盟友互补专家 (ACE) [28] 没有分成几个平衡的组, 而是将样本分成几个技能多样化的子集, 其中一个子集包含所有类, 一个包含中间类和尾部类, 另一个只包含尾部类。基于这些子集, ACE 训练不同的专家, 具有特定和互补的集成学习技能。

与其他长尾学习范式相比, 重采样方法相对简单, 并且可以轻易地实现类的重平衡, 但可以达到相当甚至更好的性能。这些优点使类重采样成为现实世界应用程序的一个很好的候选对象。然而, 这种方法的一个缺点是, 大多数重采样法以牺牲头类性能为代价来提高尾类性能, 这就像在性能跷跷板上玩一样。

2.2 集成学习

基于集成学习的方法有策略地生成和组合多个网络模块 (即多个专家) 来解决长尾视觉学习问题。我们在图 3 中总结了现有的基于集成的方法的主要方案, 具体介绍如下。

BBN [9] 提出使用传统的学习分支和再平衡分支这两个网络分支来处理长尾识别。具体来说, 传统的学习分支采用均匀抽样来模拟原始的长尾训练分布; 再平衡分支采用反向采样器在每个小批中采样更多的尾类样本, 以提高尾类性能。在训练过程中, 对两个分支的预测进行动态组合, 使 BBN 的学习焦点逐渐从头类转变为尾类 (通过重新平衡分支)。在 BBN 之后, LTML [23] 探索了双边分支网络方案来解决长尾多标签分类问题。具体来说, LTML 使用交叉熵损失对每个分支进行多标签分类, 并强制执行逻辑一致性损失, 以提高两个分支的一致性。此外, LTML 应用标签平滑和 logit 补偿来改进模型的泛化。

LFME [29] 将长尾数据集划分为几个具有较小的“类长尾度”的子集, 并使用不同的样本子集训练多个专家。在这些专家的基础上, LFME 利用多个教师专家的自适应知识蒸馏来学习一个统一的学生模型。ACE [28] 没有将类划分为几个平衡的子组, 而是将类划分为几个技能多样化的子集: 一个子集包含所有类; 一个包含中产阶级和尾类; 另一个只有尾类。然后, ACE 用不同的类子集培训多个专家, 从而使不同的专家拥有特定的和互补的技能。此外, 考虑到不同的子集有不同的样本数, ACE 还应用了分布式自适应优化器来调整不同专家的学习率。在 ResLT [30] 中也探讨了类似的 ACE 概念。

在没有数据划分的情况下，RIDE [31] 基于所有训练样本独立地使用灵活最大损失训练每个专家，并实施基于 kl 散度的损失，以提高不同专家的多样性。然后，RIDE 应用了一个专家分配模块来提高计算效率。需要注意的是，用 Softmax 损失独立地训练每个专家可以大大提高长尾学习的集成性能。

测试时间聚合不同的专家 (TADE) [11] 探索了多专家方案来处理测试分布不可知的长尾识别，其中测试类分布可以是均匀的，也可以是长尾的。具体来说，TADE 开发了一种新的跨多专家框架，并通过引入促进多样性的专业知识损失来创新专家培训方案，培训不同的专家处理不同的类分布。通过这种方式，学习过的专家更加多样化，从而导致更好的集成性能，并综合跨越了广泛的可能的类分布。在此基础上，TADE 进一步引入了一种新的测试时间自监督学习方法，即预测稳定性最大化，以自适应地聚合专家，仅基于未标记的测试数据，更好地处理未知的测试类分布。

3 代表性方法介绍

3.1 BBN: 双边分支网络

本文 [9] 富有创新性地构造了反采样分支，尝试通过反采样分支对尾部类进行特别关注。第一次尝试结合多分支（专家）和集成学习（融合学习）方法解决深度长尾学习问题，如图3所示。发表于 CVPR2020。

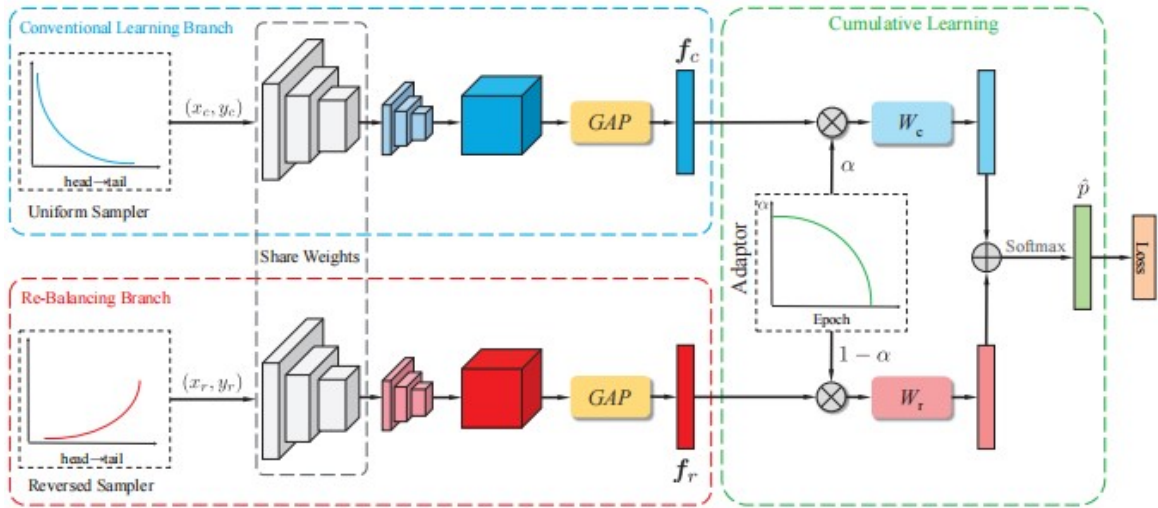


图 3: BBN 双边分支网络 [9]

双边分支网络。传统学习分支和再平衡分支均执行了自己的职责（即表示学习和分类器学习）。对于这两个分支，作者设计了如下细节。（1）**数据采样器。**传统学习分支的输入数据来自一个统一的采样器，保留了原始分布的特征，因此有利于表示学习。而重平衡分支的目的是缓解极端的不平衡，特别是提高尾部类的分类精度 [32]，其输入数据来自一个反向采样器。对于反向采样器，每个类的采样可能性与其样本大小的倒数成正比，即一个类中的样本越多，该类的采样可能性就越小。（2）**权值共享。**我们出于以下两个目的保留了共享权值。一方面，传统学习分支学习良好的表示可以有利于重新平衡分支的学习。另一方面，共享权值将极大地降低推理阶段的计算复杂度。

融合学习提出了累积学习策略，通过控制两个分支产生的特征的权重和分类损失，将学习焦点转移到双边分支之间，首先学习通用模式，然后逐步关注尾部数据。在训练阶段，常规学习分支的特征 f_c 乘以 α ，再平衡分支的特征 f_r 乘以 $1-\alpha$ ，其中 α 根据训练时期自动生成。具体地说，总训练期表示为 T_{max} ，当前时期为 T 。 α 计算如下：

$$\alpha = 1 - \left(\frac{T}{T_{max}}\right)^2$$

3.2 TLC: 可解释的长尾分类

本文 [10] 通过专家和专业知识的设置, 可解释地建立了多个子网络, 并通过联合学习整合专家知识, 如图4所示。发表于 CVPR2022。

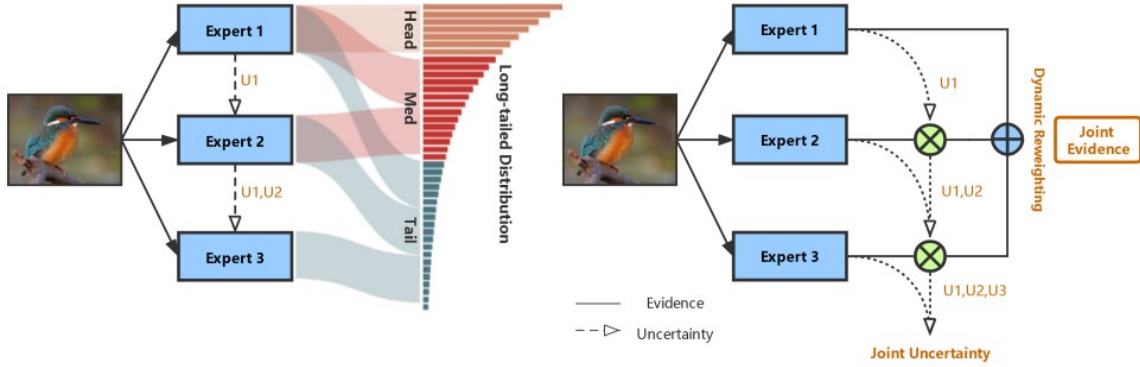


图 4: TLC 可解释性的长尾分类 [10]

在我们的多专家框架中, 每个专家都可以从输入中获取证据, 从而诱导一个分类意见 [33]。TLC 动态地分配给尾类的样本平均比头类更多的专家。这种分配是通过识别具有不确定性的硬样本来自动实现的。同时我们训练联合学习的动态参数, 得到完整的模型参数。

不确定性的计算。因此, 对于可信的长尾分类, 我们在登普斯特-谢弗证据理论 (DST) 下引入了基于循证的不确定性 (EvU), 以同时提高可信度和效率。DST 是对主观概率 [34] 的贝叶斯理论的推广。而基于 DST, 主观逻辑 (SL) 明确地将认知的不确定性和来源信任考虑到 [35]。形式上, 主观逻辑定义了狄利克雷分布 [36] 上的分配:

$$D(p|\alpha) = \begin{cases} \frac{1}{B(\alpha)} \prod_{k=1}^K p_k^{\alpha_k-1} & \text{for } p \in S_K \\ 0 & \text{otherwise} \end{cases}$$

3.3 TADE: 测试不可知的长尾识别

本文 [11] 通过在测试时自监督地聚合专家信息, 实现对不可知测试集的精准预测。目前收录于 Arxiv。

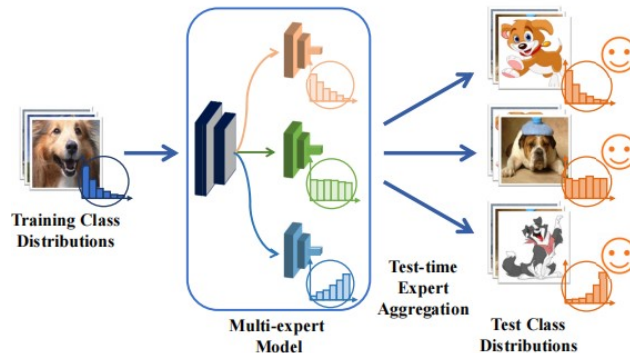


图 5: TADE 的差异化专家构建 [10]

差异化技能的多专家框架。不同的专家被训练为精通不同的类分布, 如图5所示, TADE 构建了一个由三位专家组成的模型, 它由两个组件组成: (1) 是一个专家共享的骨干 f ; (2) 独立的专家网络 E_1 、 E_2 和 E_3 。以 ResNet [37] 为例来说明多专家模型。由于浅层提取更多的一般特征, 更深的层提取更多的任务

特定特征 [38]，三专家模型使用 ResNet 的前两个阶段作为专家共享主干，而 ResNet 的后期和全连接层构成每个专家的独立组件。最终的预测是这些专家的预测日志的算术平均值，然后是一个 Softmax 函数。为了学习技能多样的专家。

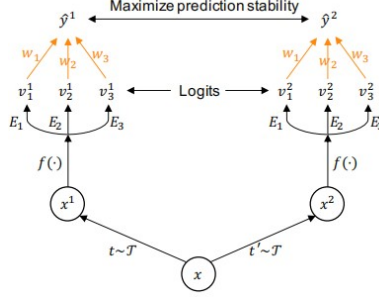


图 6: TADE 测试时自监督聚合 [10]

测试时的自监督聚合。专家聚集的一个基本原则是，专家应该在他们有专业知识的情况下发挥更大的作用，如图6。然而，如何检测各种（未知的）测试类分布的强大专家仍然是未知的。我们的关键见解是，强大的专家应该更稳定地预测其熟练阶层的样本，即使这些样本受到了干扰。为了验证这一点，我们通过比较专家对一个样本的两个增广视图的预测之间的余弦相似度来估计专家的预测稳定性。在这里，数据视图是由 MoCov2 [39] 中的数据增强技术生成的。专业知识与预测稳定性呈正相关关系，即较强的专家在其有利类别的样本的不同观点之间具有较高的预测相似性。根据这一发现，我们建议探索相对预测稳定性，以检测未知的测试类分布的强专家和权重专家。

3.4 VL-LTR：类的视觉-语言表示

本文 [12] 创造性地引入了文本信息来辅助深度长尾问题。目前收录于 Arxiv。

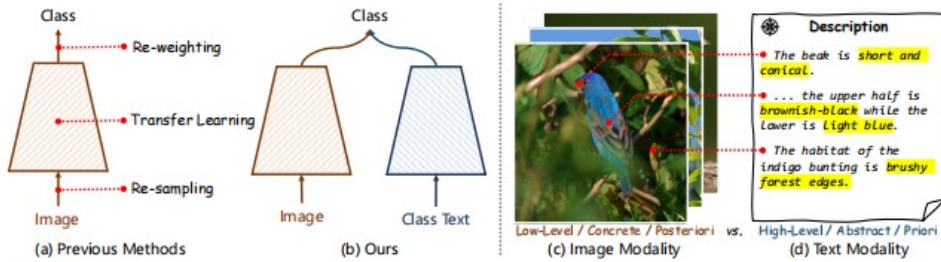


图 7: TADE 测试时自监督聚合 [10]

这篇文章一改往日的问题设定。虽然本文作者自己手动标记了图片中的描述，即此方法引入了额外信息，如图7，之后的 BALLAD [40] 则是基于本算法的启发，做到了目前的 SOTA 水平。

VL-LTR 的整体架构。整个模型有两个阶段。在第一阶段，课堂视觉语言预训练 (CVLP) 将每个课堂的图像和文本作为输入，学习通过课堂对比学习将两种模式的表征连接起来。在第二阶段，语言引导识别 (LGR) 头部使用学习到的视觉-语言表征来进行图像分类。

3.5 SAMEN: 自适应的多专家网络

本方法为笔者新提出的解决方案，构造了更加灵活，同时具有自适应的采样分支，结合多专家的知识进行分析。同时，该框架具有弱特征依赖性，可以推广到更广泛的不平衡学习问题，包括 OOD(Out-of-Distribution) 问题等。预计将发表在年底的 ICLR2023，故在此只做简单阐释，并不提交代码。

差异化采样器。对于每个专家，我们按照一个高斯分布要求进行采样，这个分布的均值和方差随着训练的过程不断迭代，采用 RS ((Random Search) 或 GS(Gradient Search) 寻找最优解。

自适应的参数调节我们设计了一个可以动态调整的集成学习参数网络，首先，我们保证：

$$\sum_{i=1}^n \omega_i = 1$$

然后利用一个 MLP 网络，输入一张图片，最后再通过 Softmax 方法得到参数。训练过程中以分类结果作为 GT 来调节网络参数。在测试时，对于某一个样本，这里面每个参数的大小意味着不同专家对于这个样本的学习熟练程度。

4 实验设置及结果

4.1 实验设置

本次实验中，笔者选取了诸多算法中较具有代表性的 6 个算法，进行实验，分别是：PaCo [41]（表示学习）、LADE [42]（代价敏感学习）、BBN [9]、ResLT [30]、TADE [11]、TLC [10]（集成学习）。

在深度长尾学习中，图像分类问题常见的数据集有 CIFAR-LT [4]、ImageNet-LT [2]、Places-LT [2]、iNaturalist [43] 等，由于算力限制，笔者本次实验只选用了 CIFAR 数据集的 CIFAR00-LT 进行实验。该数据集有 100 个类，通过调整其不平衡率改变其长尾分布的倾斜程度。

4.2 实验参数

CIFAR 数据集的不平衡率设置为 0.1，训练 200 个 epoch，50、100、200epoch 时分别记录准确率。对于多专家（分支）算法 [30] [11] [10]，其专家数均为 3。

4.3 实验结果

分类	算法	准确率%/epoch		
		50	100	200
N-EL	LADE [42]	31.8	30.0	43.9
	PaCo [41]	30.9	31.1	43.6
EL	BBN [9]	34.3	36.6	45.5
	ResLT [30]	32.1	33.2	46.6
	TADE [11]	30.9	33.6	48.3
	TLC [10]	31.1	34.2	49.1

表中 EL 表示集成学习 (Ensemble Learning)、N-EL 表示非集成学习。由于 CIFAR 数据集过小，实验结果方差可能很大，故学界一般也不取之为主要参考，但由于笔者算力限制无法在大数据集上运行，所以从仅做粗略分析。从实验结果，我们大致可以看出，集成学习的方法普遍表现较好，所以我们通常为该范式可能是未来解决深度长尾学习问题的主流方法。此外，BBN [9] 算法的收敛速度非常快，可能非常适合短、平、快地解决现实中的长尾分布问题。

4.4 实验构想

由于算力限制，笔者的实验做的并不完善，如果算力允许，笔者还将尝试进行下列实验。

- 在 CIFAR 数据集不同不平衡率下、以及不同数据集下进行实验，以尝试挖掘一些，和分布情况弱耦合的算法，这些算法可能对于普适性的不平衡学习问题更有帮助。
- 选取更多算法进行试验，获得更全面的对照结果。
- 对于多专家（分支）算法，增加专家个数。从理论上来分析，专家越多知识学得越精细。但是，是否会出现过拟合现象？

5 应用和未来

5.1 主要应用

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，以及视觉联系学习。

5.1.1 图像分类长尾学习

最常见的应用是多类分类 [9] [11] [10] [13]。有许多从广泛使用的图像分类数据集中进行人工采样的长尾数据集，即 ImageNet、CIFAR 和 Places 等。基于这些数据集，研究者们提出了各种长尾学习方法，如章节2所示。除了这些人工任务外，长尾学习还应用于真实世界的图像分类任务，包括物种分类 [43] [44] [45]，人脸识别 [46] [47] [48]，年龄分类 [49]、轨道表面缺陷检测 [50] 和医学图像诊断 [51] [52] 等。除多类分类外，长尾学习还应用于基于人工任务（即 VOC-LT 和 COCO-LT）和真实世界任务的多标签分类 [53] [54]，包括 web 图像分类 [24]、面部属性分类 [55] 和布匹属性分类 [55] 等。

5.1.2 图像检测和分割

对象检测和实例分割在长尾学习领域中引起了越来越多的关注 [56] [57] [22] [58] [59] [60]，大多基于 LVIS 和 COCO 开展研究。除了这些广泛使用的基准，许多其他应用也被探索，包括城市场景理解 [61] [62]，点云分割 [63] [64] 等。

5.1.3 视觉联系学习

视觉联系学习对图像理解非常重要，并且在长尾学习领域中吸引着越来越多的关注。一个重要的应用是长尾场景图的生成 [32] [65]。在未来，长尾视觉问题回答和图像字幕也是值得探索的方向 [66] [67]。

5.2 未来方向

在本节中，笔者从新方法和新任务两个角度确定了深度长尾学习的几个未来研究方向。

5.2.1 新方法

标签频率不可知的类重新平衡。在一些长尾任务，如多标签分类或目标检测，除了类不平衡之外，还可能出现标签共现。具体来说，标签共现意味着头类标签经常与尾类标签一起出现的情况，这可能会使得现有的依赖标签分布的模型在训练过程中难以获得准确的标签频率，无法得到准确的结果。所以现有的基于标签频率的算法在这一场景下十分受限，如何解决这一问题值得探索。

使用无标记的数据进行迁移学习。在很多长尾学习中尾类样本的数量是十分匮乏的，所以从未标记样本中转移知识是一种可行的解决方案，如自监督学习、知识蒸馏和自训练等等。然而，现有的迁移学习方法可能不能很好地处理长尾学习。例如，CReST [68] 发现，监督训练模型在长尾图像分类中对尾类往往具有较高的关注度，因此提出选择更多的尾类数据应用伪标签技术进行模型训练。然而，这种发现可能不适用于多标签分类或长尾目标检测。因此，如何更好地使用未标记数据进行长尾学习值得进一步探索。

任务泛化能力强的长尾模型。现有的长尾方法通常是针对特定的任务而设计的。然而，由于不同任务之间的差异，现有的针对特定任务的方法可能无法处理其他任务，导致方法泛化效果较差。相比于提出具体模型，针对这一问题，我们更应该提出一种解决方案，或者提出一种可替换主干网络的框架，这也是笔者在后续工作中尝试跟进的研究方向。

使用集成学习全面提高各个类表现。大多数现有的长尾方法以牺牲头类性能为代价来提高尾类性能。一种解决方案是集成学习，利用不同专家的不同知识，在头类和尾类之间获得更好的权衡，从而使得长尾学习的表现提升更加全面。由于所有类的潜在性能均能获得提高，多专家结合集成学习将成为今后深度长尾学习的主流解决方案，实验结果也佐证了这一点 [9] [10]。

引入多模态信息辅助。目前许多视觉问题都引入了文本信息作为辅助，在 CLIP [69] 等视觉-文本融合模型的加持下，大多数视觉任务的表现都得到了质的提升。VL-LTR [12] 首先尝试通过手动引入图片描述来提升尾部类的表现。而此后，逐渐有人开始关注通过图片提取文本，再利用文本增强图片的信息，以提高视觉任务的效果 [40]。其他模态的信息，如音频等也在考虑范围内。在如今主流视觉任务都在尝试引入多模态方法来提升表现的情况下，引入多模态信息来解决长尾问题也是好的尝试。

5.2.2 新任务

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集类分布保持平衡。然而，实际的测试分布经常违背这一假设（例如是长尾的，甚至是反长尾的），这将导致现有方法失效。为了克服这一限制，LADE [42] 假设测试类分布可以任意偏移，但可以用测试分布的先验来放宽这一假设。之后，TADE [11] 进一步更新了这个任务，其中测试类的分布是任意便宜且未知的。除了类不平衡，这个任务还提出了另一个挑战，即训练和测试样本之间的类分布转移。

开放式长尾学习。现实世界的样本通常有一个长尾且开放的类分布。受此启发，开放集长尾学习 [2] [70] 寻求从长尾数据中学习，并在包括头、尾和开放类的平衡测试集上优化分类精度。该问题有两个主要的挑战：（1）如何在头部和尾部类之间共享视觉知识；（2）如何减少尾部和开放类之间的混淆。

联邦长尾学习。现有的长尾学习研究一般假设所有的训练样本在模型训练过程中都是可访问的。然而，在现实应用中，长尾训练数据可以分布在许多移动设备或物联网上 [71]，这需要对深度模型的分散训练。这种任务设置被称为联邦长尾学习，它有两个主要挑战：（1）长尾类不平衡；（2）未知的类分布在不同客户端的本地数据之间转移。

多域长尾学习。目前的长尾方法一般假设所有的长尾样本都来自相同的数据域。然而，在实践中，长尾数据也可能来自数据分布不同的域 [7] [72]。在此的激励下，多域长尾学习寻求同时处理类的不平衡和域的分布的转移。一个更具有挑战性的问题可能是不同域之间的不平衡，即不同的域可能有不同的类分布，这进一步扩大了多域长尾学习中的域偏移。

6 结论及感想

在本综述中，笔者根据类再平衡、信息增强和模块改进的分类，广泛回顾了近年提出的经典的深度长尾学习方法。笔者分析并验证了几种最先进的长尾方法，并提出了一种基于多专家和集成学习的自适应深度长尾学习框架。最后，笔者讨论了长尾学习的主要应用场景，并为新方法和任务确定了潜在研究方向。

笔者之前虽然接触了一些计算机视觉与图像处理的实践和科研项目，但是一直没有系统性地学习图像处理知识。在谭老师的讲解下，笔者第一次完整地构建了图像处理的知识体系，从传统的图像处理到现在主流的深度学习方法，都有了更深刻全面的了解。在本课程结束后，笔者将持续在计算机视觉和深度学习领域深耕，并希望能做出一些有意义的工作。再次感谢谭老师在本课程中的讲解和指导，也欢迎老师持续指导和关注我的工作。

参考文献

- [1] VOULODIMOS A, DOULAMIS N, DOULAMIS A, et al. Deep learning for computer vision: A brief review[J]. Computational intelligence and neuroscience, 2018, 2018.
- [2] LIU Z, MIAO Z, ZHAN X, et al. Large-scale long-tailed recognition in an open world[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2537–2546.
- [3] KANG B, LI Y, XIE S, et al. Exploring balanced feature spaces for representation learning[C] // International Conference on Learning Representations. 2020.
- [4] CAO K, WEI C, GAIDON A, et al. Learning imbalanced datasets with label-distribution-aware margin loss[J]. Advances in neural information processing systems, 2019, 32.
- [5] CUI Y, JIA M, LIN T-Y, et al. Class-balanced loss based on effective number of samples[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9268–9277.
- [6] ZHANG Y, KANG B, HOOI B, et al. Deep long-tailed learning: A survey[J]. arXiv preprint arXiv:2110.04596, 2021.
- [7] JAMAL M A, BROWN M, YANG M-H, et al. Rethinking class-balanced methods for long-tailed visual recognition from a domain adaptation perspective[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 7610–7619.
- [8] YANG L, JIANG H, SONG Q, et al. A Survey on Long-Tailed Visual Recognition[J]. International Journal of Computer Vision, 2022: 1–36.
- [9] ZHOU B, CUI Q, WEI X-S, et al. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 9719–9728.
- [10] LI B, HAN Z, LI H, et al. Trustworthy Long-Tailed Classification[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 6970–6979.
- [11] ZHANG Y, HOOI B, HONG L, et al. Test-agnostic long-tailed recognition by test-time aggregating diverse experts with self-supervision[J]. arXiv preprint arXiv:2107.09249, 2021.
- [12] TIAN C, WANG W, ZHU X, et al. VL-LTR: Learning Class-wise Visual-Linguistic Representation for Long-Tailed Visual Recognition[J]. arXiv preprint arXiv:2111.13579, 2021.
- [13] KANG B, XIE S, ROHRBACH M, et al. Decoupling representation and classifier for long-tailed recognition[J]. arXiv preprint arXiv:1910.09217, 2019.
- [14] WANG T, LI Y, KANG B, et al. The devil is in classification: A simple framework for long-tail instance segmentation[C] // European conference on computer vision. 2020: 728–744.
- [15] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002, 16: 321–357.
- [16] ESTABROOKS A, JO T, JAPKOWICZ N. A multiple resampling method for learning from imbalanced data sets[J]. Computational intelligence, 2004, 20(1): 18–36.
- [17] HAN H, WANG W-Y, MAO B-H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning[C] // International conference on intelligent computing. 2005: 878–887.
- [18] LIU X-Y, WU J, ZHOU Z-H. Exploratory undersampling for class-imbalance learning[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2008, 39(2): 539–550.
- [19] MAHAJAN D, GIRSHICK R, RAMANATHAN V, et al. Exploring the limits of weakly supervised pretraining[C] // Proceedings of the European conference on computer vision (ECCV). 2018: 181–196.
- [20] WANG Y, GAN W, YANG J, et al. Dynamic curriculum learning for imbalanced data classification[C] // Proceedings of the IEEE/CVF international conference on computer vision. 2019: 5017–5026.
- [21] REN J, YU C, MA X, et al. Balanced meta-softmax for long-tailed visual recognition[J]. Advances in neural information processing systems, 2020, 33: 4175–4186.
- [22] ZANG Y, HUANG C, LOY C C. Fasa: Feature augmentation and sampling adaptation for long-tailed instance segmentation[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 3457–3466.
- [23] HUANG C, LI Y, LOY C C, et al. Learning deep representation for imbalanced classification[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 5375–5384.
- [24] KIM C D, JEONG J, KIM G. Imbalanced continual learning with partitioning reservoir sampling[C] // European Conference on Computer Vision. 2020: 411–428.
- [25] LIU B, LI H, KANG H, et al. Gistnet: a geometric structure transfer network for long-tailed recognition[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 8209–8218.

- [26] LI Y, WANG T, KANG B, et al. Overcoming classifier imbalance for long-tail object detection with balanced group softmax[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10991–11000.
- [27] HU X, JIANG Y, TANG K, et al. Learning to segment the tail[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 14045–14054.
- [28] CAI J, WANG Y, HWANG J-N. Ace: Ally complementary experts for solving long-tailed recognition in one-shot[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 112–121.
- [29] XIANG L, DING G, HAN J. Learning from multiple experts: Self-paced knowledge distillation for long-tailed classification[C] // European Conference on Computer Vision. 2020: 247–263.
- [30] CUI J, LIU S, TIAN Z, et al. Reslt: Residual learning for long-tailed recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
- [31] WANG X, LIAN L, MIAO Z, et al. Long-tailed recognition by routing diverse distribution-aware experts[J]. arXiv preprint arXiv:2010.01809, 2020.
- [32] DESAI A, WU T-Y, TRIPATHI S, et al. Learning of visual relations: The devil is in the tails[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 15404–15413.
- [33] KIELA D, GRAVE E, JOULIN A, et al. Efficient large-scale multi-modal classification[C] // Proceedings of the AAAI Conference on Artificial Intelligence: Vol 32. 2018.
- [34] DEMPSTER A P. A generalization of Bayesian inference[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1968, 30(2): 205–232.
- [35] JSANG A. Subjective Logic: A formalism for reasoning under uncertainty[M]. [S.l.]: Springer, 2018.
- [36] KOTZ S, BALAKRISHNAN N, JOHNSON N L. Continuous multivariate distributions, Volume 1: Models and applications: Vol 1[M]. [S.l.]: John Wiley & Sons, 2004.
- [37] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770–778.
- [38] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks?[J]. Advances in neural information processing systems, 2014, 27.
- [39] CHEN X, FAN H, GIRSHICK R, et al. Improved baselines with momentum contrastive learning[J]. arXiv preprint arXiv:2003.04297, 2020.
- [40] MA T, GENG S, WANG M, et al. A simple long-tailed recognition baseline via vision-language model[J]. arXiv preprint arXiv:2111.14745, 2021.
- [41] CUI J, ZHONG Z, LIU S, et al. Parametric contrastive learning[C] // Proceedings of the IEEE/CVF international conference on computer vision. 2021: 715–724.
- [42] HONG Y, HAN S, CHOI K, et al. Disentangling label distribution for long-tailed visual recognition[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 6626–6636.
- [43] VAN HORN G, MAC AODHA O, SONG Y, et al. The inaturalist species classification and detection dataset[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8769–8778.
- [44] MIAO Z, LIU Z, GAYNOR K M, et al. Iterative human and automated identification of wildlife images[J]. Nature Machine Intelligence, 2021, 3(10): 885–895.
- [45] KEATON M R, ZAVERI R J, KOVUR M, et al. Fine-grained visual classification of plant species in the wild: Object detection as a reinforced means of attention[J]. arXiv preprint arXiv:2106.02141, 2021.
- [46] ZHANG X, FANG Z, WEN Y, et al. Range loss for deep face recognition with long-tailed training data[C] // Proceedings of the IEEE International Conference on Computer Vision. 2017: 5409–5418.
- [47] CAO D, ZHU X, HUANG X, et al. Domain balancing: Face recognition on long-tailed domains[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 5671–5679.
- [48] ZHONG Y, DENG W, WANG M, et al. Unequal-training for deep face recognition with long-tailed noisy data[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7812–7821.
- [49] DENG Z, LIU H, WANG Y, et al. Pml: Progressive margin loss for long-tailed age classification[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 10503–10512.
- [50] ZHANG Z, YU S, YANG S, et al. Rail-5k: a Real-World Dataset for Rail Surface Defects Detection[J]. arXiv preprint arXiv:2106.14366, 2021.
- [51] JU L, WANG X, WANG L, et al. Relational subsets knowledge distillation for long-tailed retinal diseases recognition[C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. 2021: 3–12.

- [52] GALDRAN A, CARNEIRO G, GONZÁLEZ BALLESTER M A. Balanced-mixup for highly imbalanced medical image classification[C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. 2021 : 323–333.
- [53] WU T, HUANG Q, LIU Z, et al. Distribution-balanced loss for multi-label classification in long-tailed datasets[C] // European Conference on Computer Vision. 2020 : 162–178.
- [54] GUO H, WANG S. Long-tailed multi-label visual recognition by collaborative training on uniform and re-balanced samplings[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021 : 15089–15098.
- [55] DONG Q, GONG S, ZHU X. Class rectification hard mining for imbalanced deep learning[C] // Proceedings of the IEEE International Conference on Computer Vision. 2017 : 1851–1860.
- [56] TAN J, WANG C, LI B, et al. Equalization loss for long-tailed object recognition[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020 : 11662–11671.
- [57] WENG Z, OGUT M G, LIMONCHIK S, et al. Unsupervised discovery of the long-tail in instance segmentation using hierarchical self-supervision[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021 : 2603–2612.
- [58] HSIEH T-I, ROBB E, CHEN H-T, et al. Droploss for long-tail instance segmentation[C] // Proceedings of the AAAI conference on artificial intelligence : Vol 35. 2021 : 1549–1557.
- [59] WEYAND T, ARAUJO A, CAO B, et al. Google landmarks dataset v2-a large-scale benchmark for instance-level recognition and retrieval[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020 : 2575–2584.
- [60] WU J, SONG L, WANG T, et al. Forest r-cnn: Large-vocabulary long-tailed object detection and instance segmentation[C] // Proceedings of the 28th ACM International Conference on Multimedia. 2020 : 1570–1578.
- [61] HE R, YANG J, QI X. Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021 : 6930–6940.
- [62] MAO J, NIU M, JIANG C, et al. One million scenes for autonomous driving: Once dataset[J]. arXiv preprint arXiv:2106.11037, 2021.
- [63] ZHANG Y, ZHOU Z, DAVID P, et al. Polarnet: An improved grid representation for online lidar point clouds semantic segmentation[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020 : 9601–9610.
- [64] CHEN X, ZHANG C, LIN G, et al. Compositional prototype network with multi-view comparison for few-shot point cloud semantic segmentation[J]. arXiv preprint arXiv:2012.14255, 2020.
- [65] DHINGRA N, RITTER F, KUNZ A. BGT-Net: Bidirectional GRU transformer network for scene graph generation[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021 : 2150–2159.
- [66] CHEN J, AGARWAL A, ABDELKARIM S, et al. RelTransformer: Balancing the Visual Relationship Detection from Local Context, Scene and Memory[J]. arXiv preprint arXiv:2104.11934, 2021.
- [67] LI Z, STENGEL-ESKIN E, ZHANG Y, et al. Calibrating Concepts and Operations: Towards Symbolic Reasoning on Real Images[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021 : 14910–14919.
- [68] WEI C, SOHN K, MELLINA C, et al. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021 : 10857–10866.
- [69] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C] // International Conference on Machine Learning. 2021 : 8748–8763.
- [70] ZHU L, YANG Y. Inflated episodic memory with region self-attention for long-tailed visual recognition[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020 : 4344–4353.
- [71] LUO M, CHEN F, HU D, et al. No fear of heterogeneity: Classifier calibration for federated learning with non-iid data[J]. Advances in Neural Information Processing Systems, 2021, 34 : 5972–5984.
- [72] ZHANG Y, NIU S, QIU Z, et al. COVID-DA: Deep domain adaptation from typical pneumonia to COVID-19[J]. arXiv preprint arXiv:2005.01577, 2020.