



基于平衡对比学习策略的长尾识别方法

孔令权^{1,2}, 翟俊海^{1,2}

(1. 河北大学 数学与信息科学学院, 河北 保定 071002; 2. 河北省机器学习与计算智能重点实验室[河北大学], 河北 保定 071002)

摘要 长尾识别是计算机视觉领域最具挑战性的问题之一。在现实世界中长尾识别具有广泛的应用,研究长尾识别具有重要意义。对于长尾分布数据来说,由于类与类之间样本量不平衡,以及占比众多的尾部类缺少足够的训练样本,使其在训练过程中很难找到各类间的明确界限。为解决这一问题,将元预训练和监督对比学习结合起来,提出了基于平衡对比学习策略的长尾识别方法 MBCP-BB (meta balanced contrastive pre-training and batch balance)。MBCP-BB 采用解耦学习方式对模型进行训练:通过预训练获得具有优异特征表示能力的特征提取器,在微调阶段,固定特征提取器,重新训练分类器。该方法突出特征学习的重要性,设计了平衡对比学习策略指导特征学习过程,从而使监督对比学习技术能有效应用于长尾识别场景。进行特征学习时,首先适当减少头部类样本,并利用少样本图像生成技术为尾部类生成新样本;之后以每类的类原型作为补充样本用于训练。解耦学习训练模式下,充分挖掘了特征提取器与分类器的潜力,在增强模型特征学习能力的同时,大大简化了分类器的训练过程。在几个长尾基准数据集上进行了大量实验,并与 7 个代表性的算法从多个角度进行了实验比较,实验结果表明该方法优于比较的算法。

关键词 长尾识别;元学习;预训练;监督对比学习;批次平衡训练

中图分类号:TP391 DOI:10.16152/j.cnki.xdxbzr.2024-04-010

A long-tailed recognition method based on balanced contrastive learning strategy

KONG Lingquan^{1,2}, ZHAI Junhai^{1,2}

(1. College of Mathematics and Information Science, Hebei University, Baoding 071002, China;

2. Key Laboratory of Machine Learning and Computational Intelligence, Hebei University, Baoding 071002, China)

Abstract Long-tailed recognition is one of the most challenging problems in computer vision. Long-tailed recognition has a wide range of applications in the real world, and it is of great significance to study long-tailed recognition. For long-tailed distribution data, due to the unbalanced sample size between classes and the lack of sufficient training samples for the large tail classes, it is difficult to find a clear boundary between classes during the training process. To address this issue, we combine meta pre-training and supervised contrastive learning, and propose MBCP-BB (meta balanced contrastive pre-training and batch balance), a long-tailed

收稿日期:2023-10-20

基金项目:河北省科技计划重点研发项目(19210310D);河北省自然科学基金(F2017201026)。

第一作者:孔令权,男,从事深度学习研究,15530098103@163.com。

通信作者:翟俊海,男,教授,博士生导师,从事深度学习、小样本学习和非平衡学习研究,mczjh@126.com。

recognition method based on a balanced contrastive learning strategy. MBCP-BB adopts a decoupled learning method for model training: A feature extractor with excellent feature representation ability is obtained through pre-training, and in the fine-tuning stage, the feature extractor is fixed and the classifier is retrained. This method highlights the importance of feature learning, and designs a balanced contrastive learning strategy to guide the feature learning process, so that supervised contrastive learning techniques can be effectively applied to long-tailed recognition scenarios. When performing feature learning, first reduce the samples of the head classes appropriately, and use the few-shot image generation technology to generate new samples for the tail classes; then use the class prototype of each class as supplementary samples for training. In the decoupled learning training mode, the potential of the feature extractor and classifier is fully mined, and the training process of the classifier is greatly simplified while enhancing the feature learning ability of the model. A large number of experiments are carried out on several long-tailed benchmark datasets, and compared with seven representative algorithms from multiple perspectives, the experimental results show that the proposed method is superior to the compared algorithms.

Keywords long-tailed recognition; meta-learning; pre-training; supervised contrastive learning; batch balance training

近年来,随着深度学习技术的发展和计算能力的提升,大规模视觉数据集的建立和整理变得更加完善。这些进步为计算机视觉领域带来了显著突破,尤其是深度神经网络的应用,使得图像分类、图像分割以及目标检测等技术取得了巨大进展。然而,这些技术严重依赖大规模平衡训练数据,难以有效处理更贴近现实的长尾识别问题^[1-3]。与此同时,多数现实场景中的训练数据呈长尾分布,如图 1 所示。现实场景中需要对长尾数据进行有效处理和识别,例如医学影像分析、无人驾驶和视频监控等。

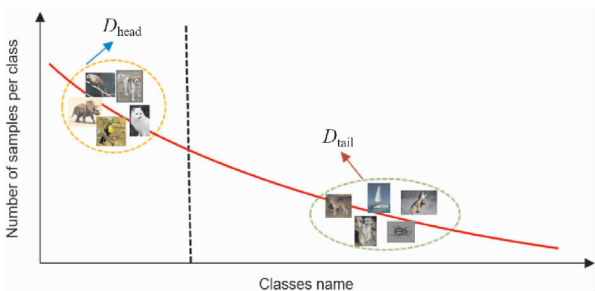


图 1 长尾分布示意图

Fig. 1 Illustration of long-tailed distribution

处理长尾识别问题的方法很多,传统的长尾识别方法包括重采样^[4-8]、重加权^[9-12]和数据增强^[13-18]3类。最近的研究^[7,19-22]探索了一些其他的解决方案,标签分布感知边际损失^[19](label-distribution-aware margin loss, LDAM)鼓励模型在每类边际间进行最佳权衡,同时合理使用再平衡策略,在基准数据集上获得了更佳的性能收益。文献^[7,22]将模型训练过程解耦为特征学习和

分类器训练两个阶段,进一步提升了模型性能。同时,由于长尾识别问题中造成模型性能低下的主要原因是尾部类数据的匮乏,这和少样本学习问题^[23]不谋而合。因此,可以将少样本学习看作长尾识别的一个子任务。处理少样本学习问题的元学习方法^[24-32]也可用于解决长尾识别问题。其中,Finn等提出的模型无关元学习算法(model agnostic meta-learning, MAML)从参数优化的角度出发,使训练出的模型易于微调且在新任务上具有良好的泛化能力^[30]。

除上述长尾识别方法外,特征学习的重要性逐渐被人们认知和发掘。从提高模型特征表示能力的角度出发,能进一步解决长尾识别问题。Wang等和Zhu等引入监督对比学习来辅助长尾识别中的特征学习,取得了十分出色的效果^[33-34]。Wang等人提出一种新的混合网络结构,以联合学习特征和分类器,利用改进的监督对比损失进行特征学习,基于交叉熵损失训练分类器,训练时从特征学习逐步过渡到分类器学习^[33]。Zhu等人提出了平衡对比学习(balanced contrastive learning, BCL),设计了一种新的平衡对比损失,与监督对比损失相比,进行了类平均和类补足两项改进,类平均有效平衡了负类的梯度贡献,类补足则使所有类别的样本都能出现在每个训练批次中^[34]。

受特征学习和解耦学习的启发,本文提出一种基于平衡对比学习策略的长尾识别方法 MB-CP-BB。采用的监督对比学习技术能在训练时将

同一类样本的嵌入表示聚在一起,并将不同类样本分开,从而大大强化了模型的特征表示能力,降低了训练过程中参数调整和优化的难度。元预训练算法和平衡对比学习策略的结合能有效增强 MBCP-BB 的泛用性,少样本图像生成技术的应用也为该方法提供了重要支撑。

本文的贡献包括3个方面:

1)突出特征学习的重要性,设计了平衡对比学习策略指导特征学习过程,从而使监督对比学习技术能有效应用于长尾识别场景;

2)采用解耦学习方式进行模型训练,充分发挥了特征提取器与分类器的潜力,在增强模型特征表示能力的同时,大大简化了分类器的训练过程;

3)通过研究不同采样策略对模型微调阶段的影响,证明本文提出的批次平衡训练策略更适用于长尾条件下的分类器训练。

1 相关工作

1.1 长尾识别

处理长尾识别问题的方法可分为重采样^[4-8]、重加权^[9-12]和数据增强^[13-18]3类。重采样通过平衡头部类每类包含的样本数来处理长尾识别问题。一般的重采样方法分为对尾部类随机过采样^[4]和对头部类随机欠采样^[4-5]两种。最近的研究应用了各种新式采样方法,具体包括样例均衡采样、类均衡采样、平方根采样^[6]、渐进均衡采样^[7]和动态采样策略^[8]。重加权又称代价敏感学习,即给训练中不同类别的损失分配不同的权重,以此达到损失平衡。类平衡损失^[9](class-balanced loss, CB)利用每个类别的有效样本数来重加权。焦点损失^[10](focal loss)用预测分数反向加权每类的损失。Ren 等提出一种元学习算法,其基于训练样本的梯度方向分配权重。数据增强的思路是在模型训练中引入额外信息^[11]。Mullick 等使用生成对抗模型为尾部类生成新样本^[13]。文献[14-17]从分布的角度出发,研究如何从头部类迁移知识到尾部类。Wang 等提出一个将少样本模型参数映射到多样本模型参数的元网络,用渐进的方式从头部类转移知识到尾部类^[18]。

此外,最近的研究^[7,19-22]也探索了一些其他解决方案。LDAM 鼓励模型在每类边际之间进行

最佳权衡,让尾部类别拥有更大的边际。Logits 调整^[20]基于标签频率在训练时或训练后调整模型的 Logits。一个类别的每个正样本都可被视为其他类别的负样本,这使得尾部类会接收更多的负面梯度,在此基础上,Tan 等提出一种均衡损失,通过简单地忽略尾部类的梯度来处理长尾不平衡问题^[21]。解耦学习将模型训练过程解耦为特征学习和分类器训练两个阶段,文献[7]是进行这种训练的开创性工作,其通过大量实验研究了不同因素对模型长尾识别能力的贡献,结果表明特征学习更适合使用随机采样,而类均衡采样是分类器学习的更好选择。双边分支网络^[22](bilateral-branch network, BBN)是在文献[7]的基础上提出的,该模型具有两路分支,其中一个分支使用随机采样来学习头部类数据,另一个分支则使用反向采样来增强尾部类数据。Wang 等和 Zhu 等研究了不平衡数据的特征学习,并且引入监督对比学习来辅助特征学习^[33-34]。

1.2 元学习

长尾识别和少样本学习联系紧密,处理少样本学习的元学习范式同样适用于长尾识别领域。元学习按学习角度的不同可分为基于度量的元学习^[25-27]、基于优化的元学习^[28-29]和基于模型的元学习^[30-32]。

基于度量的元学习通过学习合适的度量空间进行分类。匹配网络^[25]用不同的嵌入函数分别学习查询集和支撑集样本的嵌入。原型网络^[26]将支撑集每类样本嵌入的均值作为该类的类原型,并根据类原型和查询集样本嵌入之间的欧氏距离对查询集样本分类。关系网络^[27]则设计了一个关系模块作为可学习的度量。

基于优化的元学习旨在学习优化器本身。文献[28]以递归神经网络(recursive neural network, RNN)代替梯度下降法,学习通过 RNN 来执行梯度下降。文献[29]则用长短期记忆网络(long short term memory networks, LSTM)来代替梯度下降法。

基于模型的元学习从学习最优初始化参数入手,加快模型的收敛,从而实现快速学习。MAML 尝试从大量任务中训练具有良好泛化能力的最优初始化参数,在该参数的基础上利用新任务的少量训练样本并经过几次梯度更新便可使模型收敛。MAML + ^[31]对 MAML 进行了各种修改,这些修改不仅使训练过程更稳定,还可明显减少计

算开销。任务无关元学习^[32](task agnostic meta-learning, TAML)采用熵最大化策略防止任务偏差,进一步优化了模型的泛化能力。

1.3 对比学习

对比学习以构建正负样本对的方式训练模型,通过聚合语义上相似的样本,区分语义上不相似的样本,用于各种任务的特征学习^[35-36]。文献[37]是典型的自监督对比学习方法,监督对比学习^[38]则利用标签信息区分正负样本,以使图像分类性能达到更先进的水平。

然而,传统对比学习方法不能直接用于长尾识别,因为长尾条件下头部类会主导特征学习过程。于是针对长尾场景下的对比学习,涌现出一些新方法。文献[39]利用自监督和半监督对比学习来增强长尾识别性能。参数对比^[40](parametric contrastive, PaCo)通过引入一组逐类可学习中心来克服监督对比学习在长尾条件下的性能下降。文献[41]采用两阶段学习模式,并在每个批次的类别中均使用相同数量的正样本。Hybrid-SC^[33](Contrastive learning based hybrid networks)构建了新的双分支网络用于训练。BCL^[34]对监督对比损失进行了类平均和类补足两项改进,纠正了监督对比学习在长尾条件下的优化行为。

2 MBCP-BB 方法

对长尾分布数据来说,由于类与类之间样本量的不平衡,以及占比众多的尾部类缺少足够的训练样本,使其在训练过程中很难找到各类间的明确界限。为了解决这一问题,将元预训练和监督对比学习结合起来,提出了基于平衡对比学习策略的长尾识别方法 MBCP-BB。由于监督对比学习能在训练时将同一类样本的嵌入表示聚在一起,并将不同类样本分开,从而大大强化了模型的特征表示能力,降低了训练过程中参数调整和优化的难度,突破了原有的性能局限。此外,元预训练算法和平衡对比学习策略的结合能有效增强 MBCP-BB 方法的泛用性,少样本图像生成技术的应用也为该方法提供了重要支撑。

2.1 方法总体设计

MBCP-BB 包括两个阶段,如图 2 所示。第一阶段的预训练包括元预训练和基于平衡对比学习策略的特征学习两个步骤;第二阶段的微调基于

批次平衡训练策略进行。两个阶段的特征提取器均使用相同的残差网络,不同阶段模型的分层器设置不同(元预训练步骤使用标准线性分类器,特征学习步骤未设置分类器,微调阶段使用可缩放余弦度量分类器)。经实验验证,MBCP-BB 在所有长尾基准数据集上均取得了最优性能。具体来说,在 CIFAR-100-LT 上比其他先进方法^[17,19]提升 1.2% ~ 3.5% 的精度;在 CIFAR-10-LT 上相较 MPBB(训练时仅执行元预训练和批次平衡训练)有大幅提升,不平衡比率为 100 时提升最明显,精度提高了 3.62%;在 ImageNet-LT 上的精度比 MPBB 提升了 0.8%,比 LDAM 高 2.3%;可缩放余弦度量分类器的设置为模型提供了稳定的性能增益。

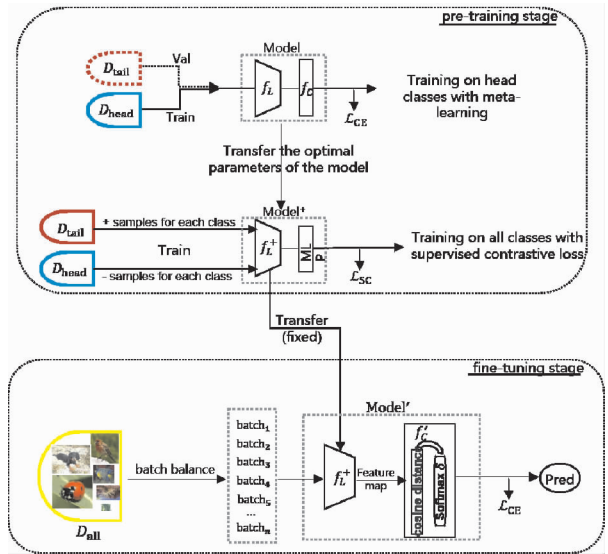


图 2 MBCP-BB 模型结构图

Fig. 2 Structure diagram of MBCP-BB model

第一阶段可细分为两个步骤,首先是在元预训练算法的指导下进行的元预训练;之后按照平衡对比学习策略进行特征学习。两个步骤的细节见 2.2 节和 2.3 节。这一阶段重视模型的特征学习能力,参考了对比学习的相关技术,这里列出自监督对比损失式(1)和监督对比损失式(2)。

$$\mathcal{L}_{\text{self}} = - \sum_{i \in I} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_{j(i)} / \tau)}{\sum_{\alpha \in A(i)} \exp(\mathbf{z}_i \cdot \mathbf{z}_{\alpha} / \tau)} \quad (1)$$

$$\mathcal{L}_{\text{SC}} = \sum_{i \in I} \frac{-1}{|P(i)|} \cdot \sum_{p \in P(i)} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p / \tau)}{\sum_{\alpha \in A(i)} \exp(\mathbf{z}_i \cdot \mathbf{z}_{\alpha} / \tau)} \quad (2)$$

第二阶段的主要任务是重新训练分类器。得益于上一阶段的训练成果,改善了模型的特征表示能力,简化了分类器的训练,从而加快了模型的收敛速度,测试精度也有大幅提升。此阶段采用批次平衡训练策略,在每个 batch 中使每类样本的数量保持平衡。与随机采样相比,采用此策略能稳定提高模型性能。

2.2 元预训练算法

受 MAML 启发,本文提出了一种长尾识别场景下的元预训练算法,具体如算法 1 所示。该算法从长尾识别的实际场景出发,利用 D_{head} 的丰富样本训练模型,并使用 D_{tail} 指导模型的训练。算法的目标是从 $P(T)$ 中提取对尾部类有益的信息,以增强尾部类,即寻找一组对任务变化敏感的最优参数 θ_{best} 。

算法 1 长尾识别场景下的元预训练算法

输入: 训练任务分布 $P(T)$, 验证任务分布 $Q(E)$, 最大迭代轮次 N , 模型参数 θ , 学习率 α, β 。
输出: 适用于 $P(T)$ 和 $Q(E)$ 的最优参数 θ_{best} 。

- ① 随机初始化参数 θ
- ② for $i \leftarrow 1$ to N do
- ③ $T_i \sim P(T)$ /* 从 $P(T)$ 中采样若干 task 形成一个 batch */
- ④ for all T_i do

- ⑤ 针对 T_i 支撑集样例计算 $\nabla_{\theta} \mathcal{L}_{T_i}(f_{\theta})$
- ⑥ $\theta'_i \leftarrow \theta - \alpha \nabla_{\theta} \mathcal{L}_{T_i}(f_{\theta})$ /* 用梯度下降法计算自适应参数 */
- ⑦ end for
- ⑧ Update $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{T_i \sim P(T)} \mathcal{L}_{T_i}(f_{\theta'_i})$
- ⑨ $E_j \sim Q(E)$ /* 从 $Q(E)$ 中采样若干 task 进行测试 */
- ⑩ for all E_i do
- ⑪ 在 θ 的基础上用 E_j 的支撑集进行微调,在查询集上验证精度 a_j
- ⑫ end for
- ⑬ $a_{\text{ave}} \leftarrow \frac{1}{n} \sum_{E_j \sim Q(E)} a_j$ /* 计算在任务集 E 上的平均验证精度 a_{ave} */
- ⑭ if a_{ave} is max do
- ⑮ $\theta_{\text{best}} \leftarrow \theta$
- ⑯ end if
- ⑰ end for

2.3 平衡对比学习策略

为寻求在长尾不平衡数据上进行特征学习的更佳方案,本节详细分析了用于长尾识别的基于交叉熵损失、标准监督对比损失及平衡对比学习策略的特征学习,如图 3 所示。

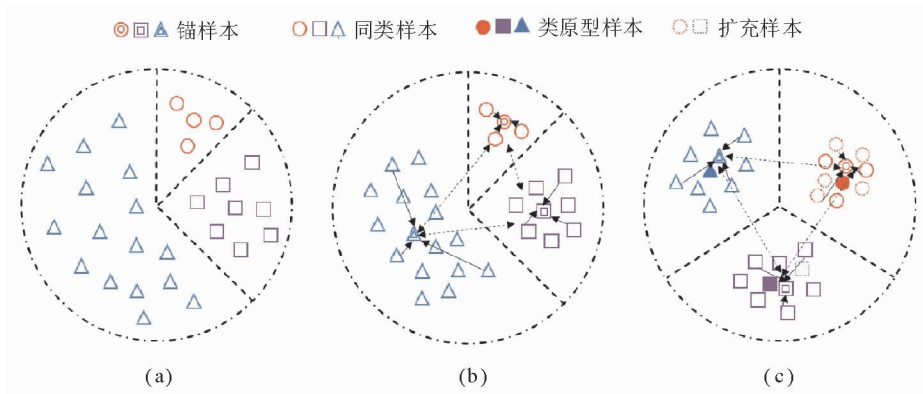


图 3 基于不同策略的特征学习示意图

Fig. 3 Illustration of feature learning based on different strategies

首先,构建了一个多分类(这里以 3 分类为例)的长尾识别场景,三角形代表长尾数据集的头部类,正方形代表样本较少的中间类,圆形代表样本匮乏的尾部类。图 3(a)描述了基于交叉熵损失的特征学习,显然,样本数量众多的头部类占据主导地位,特征学习更倾向于学习头部类特征,而中间类和尾部类的特征未得到足够重视。

因此,基于交叉熵损失进行特征学习并不是理想的选择。基于标准监督对比损失的特征学习如图 3(b)所示,在特征空间中,锚样本与同类正样本的距离被拉近,和异类负样本的距离被拉远,于是,模型可以学习到更集中的类内特征和间隔更清晰的类间特征。虽然基于监督对比损失能学到更好的特征表示,但是头部类依旧主导特征学习。

为适应长尾识别场景,本文提出了平衡对比学习策略来进行特征学习,如图 3(c) 所示。首先,适当减少头部类样本,利用少样本图像生成技术^[42]分别为中间类和尾部类生成不同数量的新样本,使长尾数据再次平衡。平衡后数据的每类样本数量更加接近中间类别的样本数量。之后,引入每类的类原型作为训练样本,有效避免了小

批次训练中的类别缺失问题,这一改进在数据集类别数多时效果更明显。应用平衡对比学习策略,能使监督对比损失适应长尾条件下的特征学习,避免特征学习被头部类主导,从而更稳定地学习长尾数据的类内特征和类间特征,具体的模型训练过程如图 4 所示。

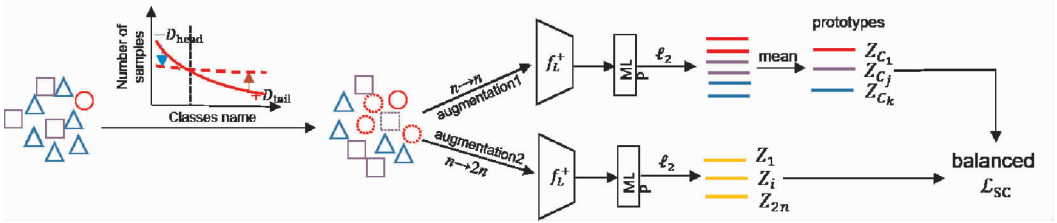


图 4 模型训练示意图

Fig. 4 Illustration of model training

实际训练中,特征学习步骤使用的模型由特征提取器 f_L^* 和一个多层感知器(MLP)组成。 f_L^* 实际上是一个残差网络模块,初始参数继承自元预训练步骤中训练好的最优模型参数 θ_{best} 。MLP是全连接网络,对特征向量进行非线性变换。模型的训练过程介绍如下。

1) 采用图 3(c) 所示的折中的数据再平衡方案,使长尾分布数据再次平衡,如图 4 所示。处理后的数据集每类样本数量处于原始数据集的平均水平。

2) 假设再平衡后数据集的样本总量为 n ,接下来数据的处理分两路进行:第一路经数据增强后总量不变;第二路经数据增强后样本总量变为 $2n$ 。

3) 之后,第一路的 n 个样本被用于计算每类的类原型,第二路的 $2n$ 个样本用于正常的对比学习训练,两路使用同一组网络模块($f_L^* + \text{MLP}$)。两路的样本(用 $\mathbf{x}$ 表示其中的任意样本)都要经过两次非线性变换,先经 f_L^* 变换后得到特征向量 $\mathbf{r}$,再经 MLP 变换后得到特征向量 $\mathbf{r}'$ 。对特征向量 $\mathbf{r}'$ 进行 $\mathcal{L}_2$ 归一化后得到特征向量 $\mathbf{z}$,向量 $\mathbf{z}$ 的计算方式如式(3)所示。

$$\mathbf{z} = \frac{\mathbf{r}'}{\sqrt{\sum_{i=1}^d \mathbf{r}'_i^2}} = \frac{\mathbf{r}'}{\sqrt{\mathbf{r}'^T \mathbf{r}'}} \quad (3)$$

4) 在第一路中,假设共有 k 个类别,类标签索引集 $C = \{c_1, \dots, c_j, \dots, c_k\}$,类原型集 $Z_C = \{\mathbf{z}_{c_1}, \dots, \mathbf{z}_{c_j}, \dots, \mathbf{z}_{c_k}\}$;第二路中,训练样本特征向量集合 $Z = \{\mathbf{z}_1, \dots, \mathbf{z}_j, \dots, \mathbf{z}_{2n}\}$, i 表示训练样本的索引编号($i \in I = \{1, \dots, 2n\}$)。以 Z_y 表示 Z 的子集,其

包含属于类别 y 的所有样本对应的特征向量, I_y 表示 Z_y 对应的索引集。未引入 Z_C 时,在一个训练批次中,每个锚样本对应的监督对比损失为

$$\mathcal{L}_i = \frac{-1}{|I_y| - 1} \cdot \sum_{p \in I_y \setminus (i)} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p / \tau)}{\sum_{\alpha \in \Lambda(i)} \exp(\mathbf{z}_i \cdot \mathbf{z}_\alpha / \tau)} \quad (4)$$

若 $I = I_{\text{head}} \cup I_{\text{tail}}$,则头尾部类及整个训练批次对应的监督对比损失分别为

$$\mathcal{L}_{\text{SC}}^{\text{head}} = \sum_{i \in I_{\text{head}}} \mathcal{L}_i \quad (5)$$

$$\mathcal{L}_{\text{SC}}^{\text{tail}} = \sum_{i \in I_{\text{tail}}} \mathcal{L}_i \quad (6)$$

$$\mathcal{L}_{\text{SC}} = \mathcal{L}_{\text{SC}}^{\text{head}} + \mathcal{L}_{\text{SC}}^{\text{tail}} = \sum_{i \in I} \mathcal{L}_i \quad (7)$$

$\mathcal{L}_{\text{SC}}$ 由 $\mathcal{L}_{\text{SC}}^{\text{head}}$ 和 $\mathcal{L}_{\text{SC}}^{\text{tail}}$ 共同决定,若不执行再平衡操作,则 $|I_{\text{head}}| \gg |I_{\text{tail}}|$, $\mathcal{L}_{\text{SC}}$ 的计算结果更多取决于头部类,于是头部类将主导模型的特征学习;执行提出的再平衡操作后, $|I_{\text{head}}|$ 不再远大于 $|I_{\text{tail}}|$,样本的类别分布比较均衡,这避免了模型学到过度倾斜的特征,使监督对比学习能有效应用于长尾识别场景。实际的训练是分批次进行的,由于随机采样和批次大小的影响,一个批次中可能缺少某些类的数据,计算损失时会出现较大误差,从而限制了模型性能。为避免小批次训练中的类别缺失,在每个批次的训练中引入类原型集 Z_C ,则每个锚样本对应的监督对比损失为

$$\mathcal{L}_i' = \frac{-1}{|I_y|} \cdot$$

$$\sum_{p \in |I_y \setminus \{i\}| \cup \{c_y\}|} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p / \tau)}{\sum_{\alpha \in |I \setminus \{i\}| \cup C} \exp(\mathbf{z}_i \cdot \mathbf{z}_\alpha / \tau)} \quad (8)$$

结合式(8),最终可得到基于平衡对比学习策略的监督对比损失,其表达式如式(9)所示。

$$\mathcal{L}_{SC}^{balance} = \sum_{i \in I} \mathcal{L}_i' = \sum_{i \in I} \frac{-1}{|I_y|} \cdot \sum_{p \in |I_y \setminus \{i\}| \cup \{c_y\}|} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p / \tau)}{\sum_{\alpha \in |I \setminus \{i\}| \cup C} \exp(\mathbf{z}_i \cdot \mathbf{z}_\alpha / \tau)} \quad (9)$$

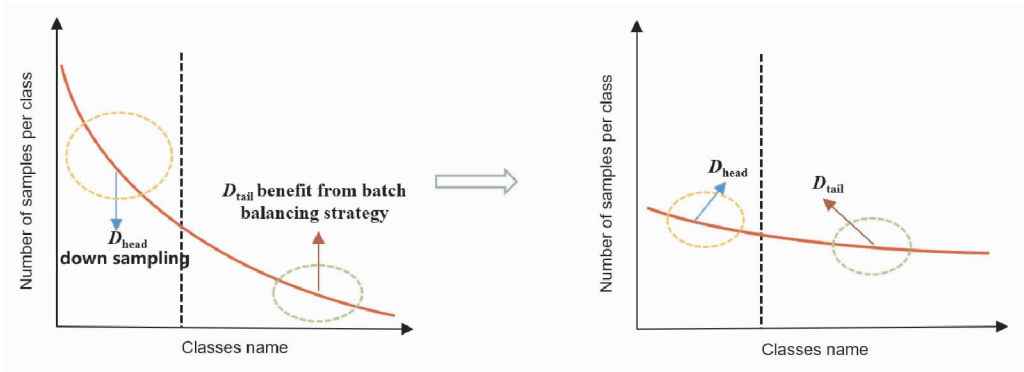


图5 批次平衡训练策略示意图

Fig. 5 Illustration of batch balance training strategy

2.5 少样本图像生成技术介绍

本文使用的少样本图像生成技术源自文献[42]提出的FAML,重新实现了FAML的代码,并利用该方法为尾部类生成新样本。

图6展示了在CIFAR-LT上的尾部类样本生成情况(对抗训练阶段经历了75 000 episodes,数据生成阶段微调了50 episodes),图7展示了在ImageNet-LT上的尾部类样本生成情况(对抗训练阶段经历了90 000 episodes,数据生成阶段微调了60 episodes),图6、7第一行为原始图像,后三行为生成图像。



图6 在CIFAR-LT上生成的尾部类图像

Fig. 6 Tail classes images generated on CIFAR-LT

2.4 批次平衡训练策略

提出的批次平衡训练策略如图5所示。其关键思想是在训练中调整每类样本去最大程度适应模型,从而让每个类对模型优化有一个近似的贡献。该策略首先对 D_{head} 执行适当的下采样操作,然后对数据集执行批次平衡训练策略,在训练过程中调整batch,使每个batch每类包含的样本数量保持平衡,这样的操作有益于增强 D_{tail} 对模型的影响,改善模型整体性能。

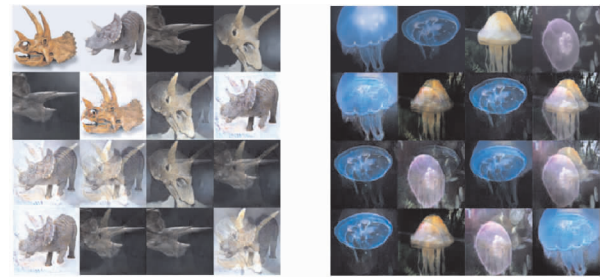


图7 在ImageNet-LT上生成的尾部类图像

Fig. 7 Tail classes images generated on ImageNet-LT

3 实验

在CIFAR-10-LT、CIFAR-100-LT和ImageNet-LT上与多种代表性的长尾识别方法进行了大量的对比实验,也与MPBB方法进行了多个角度的比较,实验以Top-1精度作为评价指标。此外,还进行了充分的消融实验验证MBCP-BB的有效性。

3.1 数据集

CIFAR-10-LT和CIFAR-100-LT:为方便与其他方法进行公平比较,本文使用与文献[19]相同的长尾版本,这个版本的训练集由原始版本抽样

得到,并且可以控制训练集的数据不平衡程度,使样本量在不同类别之间呈指数衰减。同时,本文定义了不平衡比率 γ 来描述数据不平衡程度, γ 是数据集样本最多的类与样本最少的类的训练样本数的比值,即 $\gamma = \frac{N_{\max}}{N_{\min}}$ 。本文在实验中用到的不平衡比率是 10、50 和 100。

ImageNet-LT:该版本由 Liu 等^[1]提出,它由原始 ImageNet 数据集按 Pareto 分布采样得到,是原始 ImageNet 数据集的一个子集。该数据集包含来自 1 000 个类别的 115.8K 张彩色图片,每个类别最多有 1 280 张图片,最少有 5 张图片。

3.2 实验细节

本节按数据集的不同对实验细节介绍如下。

1) CIFAR-LT。特征学习步骤中,以 ResNet-32 作为骨干网络,MLP 隐藏层大小设置为 512,输出层大小设置为 128,训练批次大小设置为 512,标量超参数 τ 设置为 0.1。在模型微调阶段,使用 SGD 优化器,动量设置为 0.9,初始学习率设为 0.1,余弦缩放参数 δ 的初始值为 10,训练进行了 100 轮。

2) ImageNet-LT。特征学习步骤中,以 ResNet-50 作为骨干网络,MLP 的单个线性层的大小为 1 024,训练批次大小设置为 512,标量超参数 τ 设置为 0.1。在微调阶段,初始学习率设置为 0.05,余弦缩放参数 δ 的初始值为 10,训练进行了 200 轮。

3.3 对比实验

首先,在 CIFAR-LT 上进行了对比实验,数据集的不平衡比率有 100、50 和 10 三种设置,实验结果列于表 1 中。对比项由 7 种代表性的长尾识别算法和本文设计的 MPBB 构成,对比项包括基于交叉熵损失训练、代价敏感学习^[9-10]、间隔修改^[19]、数据增强^[17]和两阶段训练等各种处理长尾识别问题的方案。表 1 的实验结果显示,MB-CP-BB 的 Top-1 精度明显高于各对比项,在 CIFAR-LT 上取得了最优性能。具体来说,在 CIFAR-100-LT 上比性能较高的对比项^[17,19]提升 1.2%~3.5%的精度,比 MPBB 提高了 1.17%~2.08%的精度,在不平衡比率为 10 时提升更明显;在 CIFAR-10-LT 上的实验精度相较 MPBB 有大幅提升,不平衡比率为 100 时提升最为明显,精度提高了 3.62%。

在 ImageNet-LT 上也进行了对比实验,进一步验证了 MBCP-BB 的普适性和有效性,结果列于表 2 中。

其中,OLTR 和 CE 等传统方法的实验效果都比较差;与 LDAM 相比,MBCP-BB 提升了 2.3%的精度;在 MPBB 的基础上,MBCP-BB 的精度提升了 0.8%,验证了改进方案的可行性;与采用解耦学习方案的文献[7]相比,MBCP-BB 优势明显,分析 MBCP-BB 的优势主要源于模型特征表示能力的加强。

表 1 在 CIFAR-LT 上同其他方法比较的实验结果
Tab. 1 Experimental results compared with other methods on CIFAR-LT

方法	CIFAR-10-LT			CIFAR-100-LT		
	$\gamma = 100$	$\gamma = 50$	$\gamma = 10$	$\gamma = 100$	$\gamma = 50$	$\gamma = 10$
CE	70.36	74.81	86.39	38.32	43.85	55.71
Focal loss ^[10]	70.38	76.72	86.66	38.41	44.32	55.78
CB-Focal ^[9]	74.57	79.27	87.10	39.60	45.17	57.99
CE-DRW ^[19]	76.34	79.97	87.56	41.51	45.29	58.12
CE-DRS ^[19]	75.61	79.81	87.38	41.61	45.48	58.11
LDAM-DRW ^[19]	77.03	81.03	88.16	42.04	46.62	58.71
M2m ^[17]	79.10	—	87.50	43.50	—	57.60
MPBB(ours)	75.68	79.93	87.62	43.66	46.93	58.97
MBCP-BB(ours)	79.30	82.16	88.56	44.83	48.36	61.05

表 2 在 ImageNet-LT 上同其他方法比较的实验结果
Tab. 2 Experimental results compared with other methods on ImageNet-LT

方法	ImageNet-LT
OLTR ^[1]	37.9
CE	44.3
LDAM-DRW ^[19]	48.8
τ -normalized ^[7]	46.5
cRT ^[7]	49.6
LWS ^[7]	49.8
MPBB(ours)	50.3
MBCP-BB(ours)	51.1

3.4 消融实验

本文进行了大量消融实验,以深入研究 MB-CP-BB 方法在各个方面的特性,并对实验结果进行了合理分析。首先,在 CIFAR-100-LT 上应用不同策略进行了消融研究,以验证平衡对比学习策略对模型的有效性。随后,针对模型的微调阶段设置了一组消融实验,旨在研究不同采样策略对分类器训练的影响。

3.4.1 应用不同策略的消融研究

为研究 MBCP-BB 各方面特性,设置了基于不同策略的 6 组消融实验:基于交叉熵损失训练(None)、基于标准监督对比损失进行特征学习并基于交叉熵损失进行分类器训练(SC-CE)、基于平衡对比学习策略进行特征学习并基于交叉熵损失进行分类器训练(BCP-CE)、基于元预训练和批次平衡训练策略训练模型(MPBB)、未应用可缩放余弦度量分类器设置的 MBCP-BB(No δ)和使用了余弦度量分类器的 MBCP-BB(δ)。注意,SC-CE 和 BCP-CE 中未使用元预训练算法,这两组实验仅包含特征学习和分类器训练两阶段的训练过程。消融实验结果列于表 3 中。

其中,基于交叉熵损失训练的实验精度是最低的,明显落后于其他方法;比较 MPBB 和 BCP-CE 两组实验,发现 BCP-CE 在大多数情况下的性能更优,这表明平衡对比学习策略为模型提供的性能增益要普遍高于元预训练算法,特征学习是

解决长尾识别问题的更有效手段;比较应用不同分类器设置的两组实验,发现使用余弦度量分类器能稳定提高模型性能,具体来说,这一设置为模型提供了 0.3% 左右的稳定性能增益;在 3 种不平衡比率设置下,BCP-CE 的精度均明显高于 SC-CE,这说明提出的平衡对比学习策略比标准监督对比学习更适用于长尾识别场景,同时,MBCP-BB(δ)的精度又明显高于 BCP-CE,这说明元预训练算法和平衡对比学习策略有很好的兼容性。

表 3 在 CIFAR-100-LT 上应用不同策略的消融实验结果
Tab. 3 Results of ablation experiments using different strategies on CIFAR-100-LT

策略	CIFAR-100-LT		
	$\gamma = 100$	$\gamma = 50$	$\gamma = 10$
None	38.53	43.96	55.87
SC-CE	42.35	45.89	59.18
BCP-CE	43.58	47.13	60.12
MPBB	43.66	46.93	58.97
MBCP-BB(No δ)	44.52	47.99	60.81
MBCP-BB(δ)	44.83	48.36	61.05

3.4.2 针对模型微调阶段设置的消融实验

本文设置实验研究了不同采样策略对分类器训练的影响,实验结果列于表 4 中。

在此实验中,分别采用随机数据采样和批次平衡训练策略来训练分类器。表 4 中的结果表明,使用批次平衡训练策略可以获得更好的实验效果,相比使用随机数据采样训练分类器,精度提高了 0.97% 至 2.98%。如图 8、9 所示,经进一步分析发现:不平衡比率越大,使用批次平衡训练策略获得的性能优势越为显著;同时,在 CIFAR-100-LT 上,两种策略的精度相差 1.89% ~ 2.98%,而在类别数量较少的 CIFAR-10-LT 上,精度差异仅为 0.97% ~ 2.28%,这说明批次平衡训练策略能更有效地处理高难度的长尾分类问题。综上所述,在固定特征提取器的情况下,批次平衡训练策略能进一步促进分类器的训练,从而为模型提供更多的性能增益。

表 4 在 CIFAR-LT 上应用不同采样策略进行分类器训练

Tab. 4 Applying different sampling strategies for classifier training on CIFAR-LT

策略	CIFAR-10-LT			CIFAR-100-LT		
	$\gamma = 100$	$\gamma = 50$	$\gamma = 10$	$\gamma = 100$	$\gamma = 50$	$\gamma = 10$
Random sampling	77.02	80.38	87.59	41.85	45.8	59.16
Batch balance	79.30	82.16	88.56	44.83	48.36	61.05

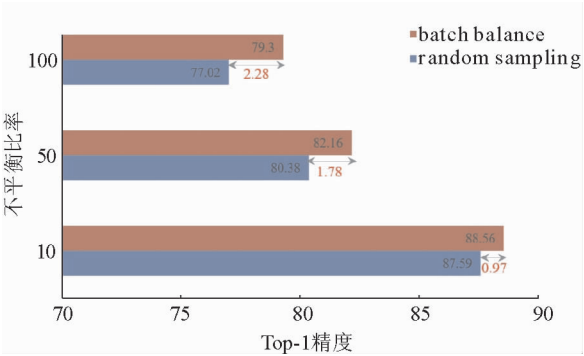


图 8 在 CIFAR-10-LT 上的 Top-1 精度对比

Fig.8 Top-1 accuracy comparison on CIFAR-10-LT

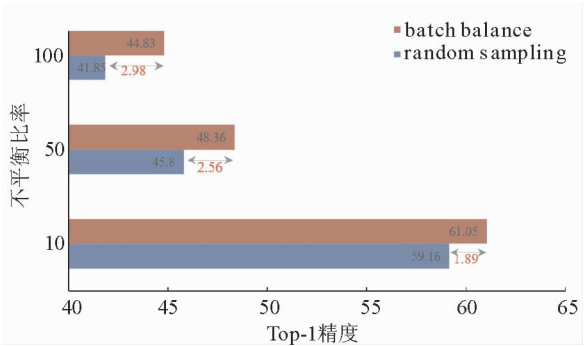


图 9 在 CIFAR-100-LT 上的 Top-1 精度对比

Fig.9 Top-1 accuracy comparison on CIFAR-100-LT

4 结论

本文提出了一种基于平衡对比学习策略的长尾识别方法 MBCP-BB, MBCP-BB 重视模型的特征表示能力,设计了平衡对比学习策略指导特征学习过程,从而使监督对比学习技术能有效应用于长尾识别场景。平衡对比学习策略中,折中的数据集再平衡方案兼顾了头部类和尾部类的性能,以每类类原型作为训练样本有效避免了小批次训练中的类别缺失。同时,采用的解耦训练方式,在增强模型特征学习能力的同时,简化了分类器的训练。通过大量的对比实验发现 MBCP-BB

方法在 CIFAR-LT、ImageNet-LT 等长尾数据集上都能取得更优的性能。此外,在对一系列消融实验进行详细分析后,得出元预训练算法和平衡对比学习策略有很好的兼容性,批次平衡训练策略能进一步促进分类器的训练。

在未来的研究中,希望探索将本文提出的方法应用于更具挑战性的任务上,例如开放环境下的长尾识别、长尾目标检测等。

参考文献

[1] LIU Z W, MIAO Z Q, ZHAN X H, et al. Large-scale long-tailed recognition in an open world [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 2532-2541.

[2] ZHANG Y F, KANG B Y, HOOI B, et al. Deep long-tailed learning: A survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45 (9): 10795-10816.

[3] 张明,翟俊海,许垒,等. 长尾识别研究进展[J]. 南京师范大学学报(工程技术版),2022,22(2):63-72.

ZHANG M, ZHAI J H, XU L, et al. Research advance in long-tailed recognition[J]. Journal of Nanjing Normal University (Engineering Technology Edition), 2022, 22(2):63-72.

[4] BUDA M, MAKI A, MAZUROWSKI M A. A systematic study of the class imbalance problem in convolutional neural networks [J]. Neural Networks, 2018, 106: 249-259.

[5] DRUMMOND C, HOLTE R C. C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling [C] // Workshop on Learning from Imbalanced Datasets II. Washington: PRML, 2003, 11: 1-8.

[6] 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). Munich: Springer, 2018: 185-201.
- [7] KANG B Y, XIE S N, ROHRBACH M, et al. Decoupling representation and classifier for long-tailed recognition [C]//Proceedings of the International Conference on Learning Representations (ICLR). Virtual Conference, 2020: 1-16.
- [8] WANG Y R, GAN W H, YANG J, et al. Dynamic curriculum learning for imbalanced data classification [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea: IEEE, 2019: 5016-5025.
- [9] CUI Y, JIA M L, 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. Long Beach: IEEE, 2019: 9260-9269.
- [10] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2999-3007.
- [11] REN M Y, ZENG W Y, YANG B, et al. Learning to reweight examples for robust deep learning [C]//International Conference on Machine Learning. Stockholm: PRML, 2018: 4334-4343.
- [12] HUANG C, LI Y N, LOY C C, et al. Learning deep representation for imbalanced classification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 5375-5384.
- [13] MULLICK S S, DATTA S, DAS S. Generative adversarial minority oversampling [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea: IEEE, 2019: 1695-1704.
- [14] YIN X, YU X, SOHN K, et al. Feature transfer learning for face recognition with under-represented data [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 5697-5706.
- [15] LIU J L, SUN Y F, HAN C C, et al. Deep representation learning on long-tailed data: A learnable embedding augmentation perspective [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 2967-2976.
- [16] WANG J F, LUKASIEWICZ T, HU X L, et al. RSG: A simple but effective module for learning imbalanced datasets [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE, 2021: 3783-3792.
- [17] KIM J, JEONG J, SHIN J. M2M: Imbalanced classification via major-to-minor translation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 13893-13902.
- [18] WANG Y X, RAMANAN D, HEBERT M. Learning to model the tail [C]//31st Conference on Neural Information Processing Systems (NIPS). Long Beach: IEEE, 2017: 1-11.
- [19] CAO K D, WEI C, GAIDON A, et al. Learning imbalanced datasets with label-distribution-aware margin loss [C]//33rd Conference on Neural Information Processing Systems (NIPS). Vancouver: IEEE, 2019: 1-12.
- [20] MENON A K, JAYASUMANA S, RAWAT A S, et al. Long-tail learning via logit adjustment [EB/OL]. (2021-07-09) [2023-03-10]. <https://arxiv.org/abs/2007.07314>.
- [21] TAN J R, WANG C B, LI B Y, et al. Equalization loss for long-tailed object recognition [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 11659-11668.
- [22] ZHOU B Y, CUI Q, WEI X S, et al. BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 9716-9725.
- [23] 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述 [J]. 软件学报, 2021, 32(2): 349-369.
- ZHAO K L, JIN X L, WANG Y Z. Survey on few-shot learning [J]. Journal of Software, 2021, 32(2): 349-369.
- [24] 李凡长, 刘洋, 吴鹏翔, 等. 元学习研究综述 [J]. 计算机学报, 2021, 44(2): 422-446.
- LI F Z, LIU Y, WU P X, et al. A survey on recent advances in meta-learning [J]. Journal of Computers, 2021, 44(2): 422-446.
- [25] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning [J]. Advances in Neural Information Processing Systems, 2016: 3637-3645.
- [26] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning [C]//31st Conference on

- Neural Information Processing Systems (NIPS). Long Beach: IEEE, 2017: 1-11.
- [27] SUNG F, YANG Y X, ZHANG L, et al. Learning to compare: Relation network for few-shot learning [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake: IEEE, 2018: 1199-1208.
- [28] ANDRYCHOWICZ M, DENIL M, GOMEZ S, et al. Learning to learn by gradient descent by gradient descent [C] // 30th Conference on Neural Information Processing Systems (NIPS). Barcelona: IEEE, 2016: 1-12.
- [29] RAVI S, LAROCHELLE H. Optimization as a model for few-shot learning [C] // 5th International Conference on Learning Representations. Toulon, France: IEEE, 2017: 1-11.
- [30] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [C] // International Conference on Machine Learning. Sydney: PMLR, 2017: 1126-1135.
- [31] ANTONIOU A, EDWARDS H, STORKEY A. How to train your MAML [C] // Proceedings of the International Conference on Learning Representations. New Orleans: ICLR, 2019: 1-11.
- [32] JAMAL M A, QI G J. Task agnostic meta-learning for few-shot learning [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 11711-11719.
- [33] WANG P, HAN K, WEI X S, et al. Contrastive learning based hybrid networks for long-tailed image classification [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE, 2021: 943-952.
- [34] ZHU J G, WANG Z, CHEN J J, et al. Balanced contrastive learning for long-tailed visual recognition [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 6898-6907.
- [35] WANG R, WU Z X, WENG Z J, et al. Cross-domain contrastive learning for unsupervised domain adaptation [J]. IEEE Transactions on Multimedia, 2022, 25: 1665-1673.
- [36] HAN T D, XIE W D, ZISSERMAN A. Self-supervised co-training for video representation learning [J]. Advances in Neural Information Processing Systems, 2020, 33: 5679-5690.
- [37] CHEN T, KORNBILTH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations [EB/OL]. (2020-02-13) [2023-03-10]. <https://arxiv.org/abs/2002.05709>.
- [38] KHOSLA P, TETERWAK P, WANG C, et al. Supervised contrastive learning [J]. Advances in Neural Information Processing Systems, 2020, 33: 18661-18673.
- [39] YANG Y Z, XU Z. Rethinking the value of labels for improving class-imbalanced learning [J]. Advances in Neural Information Processing Systems, 2020, 33: 19290-19301.
- [40] CUI J Q, ZHONG Z S, LIU S, et al. Parametric contrastive learning [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 695-704.
- [41] KANG B Y, LI Y, XIE S, et al. Exploring balanced feature spaces for representation learning [EB/OL]. (2021-05-03) [2023-03-10]. <https://api.semanticscholar.org/CorpusID:235613459>.
- [42] PHAPHUANGWITTAYAKUL A, GUO Y, YING F L. Fast adaptive meta-learning for few-shot image generation [J]. IEEE Transactions on Multimedia, 2022, 24: 2205-2217.

(编辑 张 欢)