

边缘侧领域自适应中长尾视觉识别技术研究

欧阳昱中, 韩锐*, 刘驰

(北京理工大学计算机学院, 北京 100081)

摘要: 将深度学习部署到边缘侧存在训练数据的域偏移、长尾分布、计算资源有限的问题, 因此需要应用领域自适应方法进行在线重训缓解域偏移, 重训时利用长尾削减技术缓解长尾分布问题, 且需要考虑计算开销。然而现有长尾削减技术大多计算开销较大或无法与领域自适应方法有效结合。为此, 提出一种边缘侧结合领域自适应和长尾分布削减技术的算法 EdgeTailor。EdgeTailor 将边缘合成少数类过采样技术和类平衡损失作为长尾削减优化策略, 通过对连续无监督自适应过程进行优化, 并引入缓冲区解决在线学习时尾部类数据量过少的问题, 使其能够在进行在线连续领域自适应的同时缓解在线学习时数据的长尾问题。实验结果表明, 在两组存在域偏移的长尾数据集, 以 5 种深度神经网络作为模型骨架构建的边缘侧领域自适应任务中, EdgeTailor 相比基线在目标域上平均 Top-1 准确率提升了约 8.10%; 在计算开销方面, EdgeTailor 的每秒浮点运算次数 (FLOPs) 和参数量均保持在相对较低的水平, FLOPs 相比基线中效果较好的数据合成方法减小了大约 29.84%。EdgeTailor 具有边缘侧高性能和低开销的优点, 有效缓解领域自适应中的长尾视觉识别问题。

关键词: 边缘智能; 深度学习; 长尾; 领域自适应; 在线学习

中图分类号: TP301

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0069287

Research on Long-Tail Visual Recognition Technology with Edge-Side Domain Adaptation

OUYANG Yuzhong, HAN Rui*, LIU Chi

(College of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

【Abstract】 The deployment of deep learning models at the edge is hindered by challenges such as domain offset, long-tail distribution, and limited computing resources in the training data. Therefore, domain adaptation methods must be applied for online retraining to alleviate the domain offset, and long-tail reduction techniques must be applied during retraining to alleviate the long-tail problem while considering computational costs. However, most existing long-tail reduction techniques have high computational costs or cannot be effectively combined with domain adaptation methods. To address these issues, this paper proposes EdgeTailor, a long-tail optimization method specifically designed for edge-side domain adaptation. EdgeTailor optimizes the continuous unsupervised adaptive process by using synthetic minority class oversampling techniques and class-balanced loss as strategies for tail truncation. Consequently, a buffer is introduced to address the issue of insufficient data for tail classes during online learning, allowing it to mitigate the long-tail problem while conducting online continuous domain adaptation. Experimental results demonstrate the effectiveness of EdgeTailor in edge domain adaptation tasks involving two long-tail datasets with domain shift. Using five deep neural networks as the model backbone, EdgeTailor improves average Top-1 accuracy by approximately 8.10% compared with the baseline in the target domain. In terms of computational cost, EdgeTailor maintains a low level of Floating Point Operations Per Second (FLOPs) and parameter count, reducing FLOPs by approximately 29.84% compared with the data synthesis method, with better performance than the baseline. Overall, EdgeTailor achieves high performance and low cost in addressing both domain adaptation and long-tail visual recognition challenges in edge deployment.

【Key words】 edge intelligence; deep learning; long-tail; domain adaptation; online learning

0 引言

随着边缘智能受到关注, 越来越多的深度神经网络被部署到边缘侧^[1-5]。边缘设备直接从现实场

景中获得数据, 并对深度神经网络进行在线训练。同时由于训练数据直接从现实场景中采集, 因此采集的数据往往具有连续域偏移和长尾分布的特点。

在边缘侧对深度神经网络进行训练时主要面临

收稿日期: 2024-01-23 修回日期: 2024-04-02

基金项目: 国家重点研发计划(N2021YFB3301503); 国家自然科学基金(62272046, 62132019, 61872337); 算力互联网与信息安全教育部重点实验室开放课题(2023PY002)。

通信作者 E-mail: * hanrui@bit.edu.cn

3 个挑战:1)连续域偏移^[6-8],源域训练集和目标域测试集的数据样本存在特征或标签空间偏移,源域上训练的模型不能很好地适用于目标域,域偏移会让模型的推理精度降低;2)长尾分布^[9-10],数据类别样本量不平衡,少数头部类有大量样本量,而其余尾部类的样本量很少,使用长尾数据训练模型会引起模型偏向头部类,同时在尾部类上表现不佳;3)计算资源有限,大多数边缘计算设备资源有限,无法像数据中心使用高算力设备进行长周期大规模的离线学习^[11],边缘设备上的在线学习需要考虑计算开销。

针对上述边缘侧深度学习面临的挑战,现有连续领域自适应^[12-13]和长尾分布削减技术^[14-19]可以分别缓解连续域偏移和长尾分布问题。但是在将现有长尾分布削减技术与现有连续领域自适应方法相结合,在缓解连续域偏移的同时解决重训练数据的长尾分布问题时效果却不尽如人意。其原因有多种:现有长尾分布削减技术大多用于离线深度学习,其中部分方法^[14-15]有着较高的资源开销,另一些方法^[20-24]对训练数据量有要求。同时许多现有长尾分布削减技术^[16-19]无法与现有连续领域自适应方法有效结合也是导致结合方法效果不佳的关键因素。

因此,本文提出使用长尾分布削减策略与连续领域自适应有效结合的高效方法 EdgeTailor,应用在边缘侧能高效缓解在线学习时数据的连续域偏移和长

尾分布问题。为了使长尾分布削减技术与领域自适应有效结合,EdgeTailor 使用边缘合成少数类过采样技术^[25]和类平衡损失^[17]作为长尾分布削减优化策略与连续无监督自适应(CUA)^[12]进行结合,并通过构建缓冲区存储少量在线学习的尾部数据来缓解尾类样本严重不足的问题,保证边缘合成少数类过采样技术具有一定的有效性。为控制计算开销,EdgeTailor 严格控制边缘合成少数类过采样技术的使用范围。

为验证 EdgeTailor 在边缘侧领域自适应中缓解长尾分布的有效性,将领域自适应和现有长尾分布削减技术结合作为基线,选用广泛应用的深度神经网络进行实验测试。实验结果表明,EdgeTailor 在边缘侧轻量深度神经网络上应对领域自适应中的长尾分布问题取得优势,相比基线目标域平均 Top-1 准确率提升 8.10%,同时计算开销相对基线处于较低水平。

1 相关工作

图 1 展示边缘侧视觉识别所面临的主要挑战(彩色效果见《计算机工程》官网 HTML 版,下同)。针对这些挑战,本节将对长尾分布削减技术和领域自适应方法的工作进行总结,选出具有代表性的长尾分布削减技术和符合边缘侧环境的领域自适应方法,并分析它们的局限性。

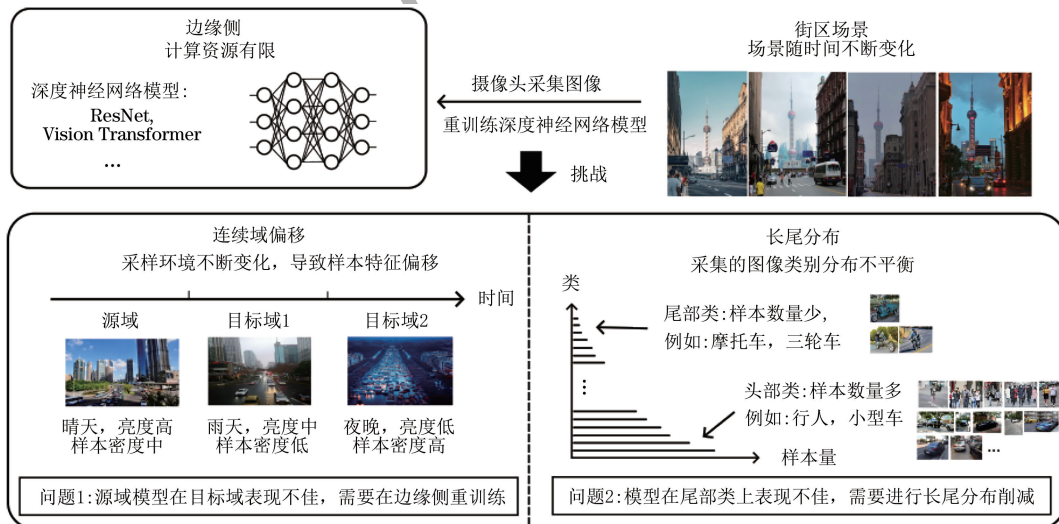


图 1 边缘侧视觉识别面临的挑战

Fig. 1 Challenges faced by edge side visual recognition

1.1 长尾分布削减技术

在现实视觉识别中,训练样本通常表现为长尾分布,即其中一小部分类具有大量样本点,而其他类仅有少数样本^[9]。训练样本数量的类不平衡导致模型偏向头部类,在尾部类上表现不佳。针对上述问题,研究人员提出长尾分布削减技术缓解长尾视觉

识别问题。长尾分布削减主要有 2 种思路:1)通过优化数据量分布削减长尾分布,包括重采样、数据合成等;2)通过在训练过程优化削减长尾分布,包括重加权、对数调整、迁移学习和模型改进等。

1) 数据优化^[14-15]。

重采样是应用最广泛的长尾分布削减技术,它

旨在平衡不同类的训练样本数量,可以分为对尾部类样本进行过采样,对头部类样本进行欠采样。数据合成指通过合成一些与尾部类样本相似的数据来平衡数据集。

2)训练过程优化^[16-24]。

重加权通过调整不同类别的权重来影响计算损失过程,从而实现类别平衡。对数调整通过对模型输出的对数概率进行调整来校正分类概率。迁移学习的基本思路是从 1 个大规模平衡域学习知识,并将这些知识迁移到目标域,该领域可以是长尾。模块改进指改进长尾学习中深度学习网络的模块,包括特征优化、分类器优化及解耦等。

长尾分布削减技术在应用中可能会用多种方式实现长尾削减。表 1 所示为近年来长尾分布削减技术及其实现方式。

表 1 现有长尾分布削减技术结合领域自适应方法对准确率的影响

方法	实现方式	准确率变化幅度
DeepSMOTE ^[14]	数据合成	-3.1%~10.5%
CFA ^[15]	数据合成	-2.4%~10.8%
IBloss ^[16]	重加权	-0.2%~2.6%
CBloss ^[17]	重加权	-0.2%~5.9%
GCLloss ^[18]	对数调整	0.4%~5.1%
LA ^[19]	对数调整	0~4.5%
Mosaicos ^[23]	迁移学习	—
RSG ^[24]	迁移学习,数据合成	—
WB ^[20]	重加权,模型优化	0~8.7%
MiSLA ^[21]	模型优化,对数调整	-0.1%~8.2%

1.2 领域自适应

在领域自适应中,训练集和测试集分别被称为源域和目标域。领域自适应通常从源域中的标注数据学习模型,该模型可以通过最小化领域分布之间的差异来推广到目标域^[7]。

边缘侧的领域自适应可能还需要面临连续域变换的问题,指目标域在连续变化,比如,随着时间的推移,目标域从阴天环境街景图像变换为夜晚环境街景图像等,这被称为连续领域自适应。

连续无监督自适应^[12]是一个可应用在域连续变化环境中的领域自适应方法,该方法利用域偏移的连续性,从 1 个源域迭代地适应连续移动的目标域。该方法结构简单,主要分为 2 个部分:通过计算域偏移距离来适应不断变化的目标域;同时使用先前样本的重放来保持对先前域的性能。CUA 模型简单轻量,通过在线学习的方式进行连续领域自适

应,适用于本文边缘侧的环境。

1.3 现有工作

将近年长尾分布削减技术与 CUA 相结合的方法可以初步缓解领域自适应中的长尾分布问题。但将结合方法应用在边缘侧进行实验测试效果不佳。为充分展示方法的效果,将长尾分布削减技术与领域自适应方法相结合,在线训练相同时间,将模型与不进行长尾分布削减模型的推理精度进行对比。表 1 展示了一些现有长尾分布削减技术对模型推理精度的影响。

基于数据合成^[14-15]的长尾分布削减技术与领域自适应相结合的方法对准确率有相对较高的影响,但数据合成会影响领域自适应过程的计算开销,有可能产生负面效果。基于重加权^[16-17]和对数调整^[18-19]的长尾分布削减技术与领域自适应相结合的方法对结果的影响微弱,主要原因是它们无法与领域自适应过程有效结合,重采样和对数调整无法优化偏移度量的计算过程。基于迁移学习^[23-24]的长尾分布削减方法与领域自适应相结合的方法需要在平衡的大型目标域数据集中进行训练,不适用于边缘侧环境,因此本文不做相关讨论。基于模型优化^[20-21]的长尾分布削减技术与领域自适应相结合的准确率可获得显著提升,但其准确率提升上限不如数据合成。

综上所述,在边缘侧将现有长尾分布削减技术与领域自适应相结合的方法会存在 2 个主要问题:1)计算开销过大,大部分模型优化和数据优化时间开销较大,在时间资源有限的边缘侧无法得到充分训练;2)长尾分布削减技术效果无法得到充分发挥,大部分长尾分布削减技术无法对领域自适应过程中的偏移度量计算过程进行优化,导致其效果不佳,有些甚至不适用与领域自适应结合,例如迁移学习的长尾分布削减技术^[23-24]。针对上述 2 个问题,本文提出 EdgeTailor,它使用边缘合成少数类过采样技术与类平衡损失作为长尾削减优化策略,对领域自适应过程进行优化,在保证领域自适应全过程都通过长尾分布削减技术进行优化的同时,控制高开销长尾分布削减方式的使用,以控制方法的整体开销,从而实现边缘侧将领域自适应与长尾分布削减技术相结合的学习方法。

2 EdgeTailor 设计

2.1 EdgeTailor 介绍

边缘侧领域自适应和长尾分布削减技术相结合会面临计算资源有限、长尾分布削减技术的优势无法充分发挥的问题,本文提出基于数据合成和重加权的长尾分布削减技术结合连续领域自适应的高效

优化方法 EdgeTailor。

图 2 展示了 EdgeTailor 的总体框架,它主要通过连续领域自适应器和长尾分布削减器 2 个模块来分别缓解边缘侧的连续域偏移和长尾分布问题。长尾分布削减器包括数据合成和重加权 2 个策略,其中数据合成处理用于域偏移度量的长尾目标域数据,但由于其计算开销较大,因此使用重加权处理样本重放过

程,在缓解重放样本长尾分布的同时减小计算开销。连续领域自适应器包括域偏移度量和样本重放 2 个策略,域偏移度量利用源域和目标域数据的特征偏差计算偏移量,从而适应连续变化的目标分布,样本重放通过重放过去的样本来保持模型之前的性能。最后使用偏移量和样本重放损失共同优化深度神经网络的特征。

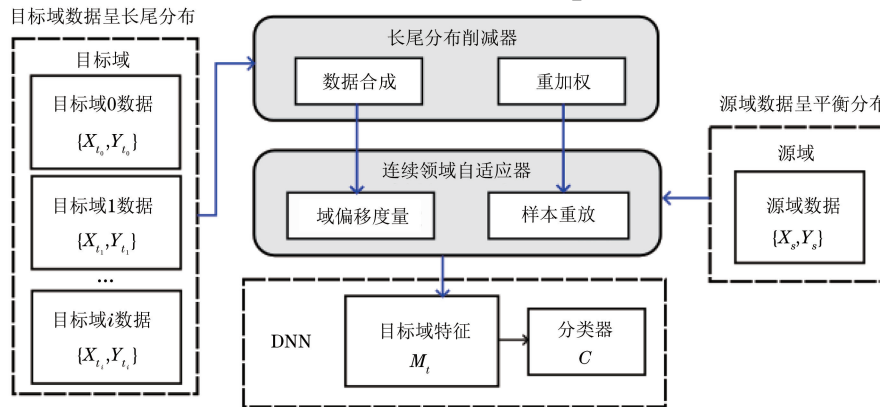


图 2 EdgeTailor 总体框架

Fig. 2 Overall framework of EdgeTailor

2.2 连续领域自适应器

一般的领域自适应通常通过学习源域 $\{X_s, Y_s\}$ 和目标域 $\{X_t, Y_t\}$ 特征表示 M 的映射来缓解域偏移的影响。然而在许多情况下,目标域数据 $\{X_{t_0}, Y_{t_0}\}, \{X_{t_1}, Y_{t_1}\}, \dots, \{X_{t_i}, Y_{t_i}\}$ 连续变化,偏移不能被 2 个静态域有效表示,这称为连续域偏移。

为使 EdgeTailor 在边缘侧高效运行,连续领域自适应器使用简单快速的连续无监督自适应作为主要架构。连续领域自适应器通过最小化源域和目标域之间域偏移的距离 d 来让特征 M_t 适应不断变化的目标域。同时为缓解特征忘记过去的知识,连续领域自适应器引入重放数据 $\{X_p, Y_p\}$ 并通过计算出的重放损失 L_{replay} 进行学习。

算法 1 展示了连续领域自适应器的运行步骤。在进行连续领域自适应前,先对源域数据进行下采样,生成重放数据 $\{X_p, Y_p\}$ (第 1 行),并使用源域特征 M_s 初始化目标域特征 M_t (第 2 行)。在连续领域自适应过程中,目标域特征通过最小化源域特征和目标域特征映射之间的距离 d ,以及重放数据 $\{X_p, Y_p\}$ 在目标域特征 M_t 与分类器 C 上计算出的重放损失 L_{replay} 来学习(第 3 和 4 行),其中偏移距离 d 选用最大平均差异度量计算得到, λ 为重放损失超参数。最后对目标域数据进行下采样并将其加入重放数据中(第 5 行)。

算法 1 连续领域自适应

输入 源域数据 $\{X_s, Y_s\}$; 目标域数据 $\{X_{t_0}, Y_{t_0}\}$,

$\{X_{t_1}, Y_{t_1}\}, \dots, \{X_{t_i}, Y_{t_i}\}$; 源域特征 M_s ; 分类器 C ; 目标域总数 N ; 重放权重 λ

1. $\{X_p, Y_p\} \leftarrow \text{subsample}\{X_s, Y_s\}$;
2. $M_t \leftarrow M_s$;
3. for $i \in \{1, \dots, N\}$ do
4. $M_t \leftarrow \underset{M_t}{\text{argmind}} (M_s(X_s), M_t(X_{t_i})) + \lambda \times L_{\text{replay}} (M_t(X_p), Y_p)$;
5. $\{X_p, Y_p\} \leftarrow \{X_p, Y_p\} \cup \text{subsample}\{X_{t_i}, Y_{t_i}\}$
6. End for

综合来看,使用算法 1 进行连续领域自适应能够让特征学习不断变化目标域之间的距离 d 以适应连续变化的目标域,同时重放损失 L_{replay} 会缓解特征遗忘问题。连续领域自适应器能够高效地进行连续领域自适应。

2.3 长尾分布削减器

2.3.1 长尾分布削减器设计

目标域数据呈现长尾分布将会影响连续领域自适应器的运行效果,因此在使用目标域数据进行连续领域自适应之前需要进行长尾分布削减处理。由于 EdgeTailor 运行环境为边缘侧,因此长尾分布削减器需要在考虑能够与领域自适应有效结合的同时尽量保持较低的计算开销。长尾分布削减器选用边缘合成少数类过采样技术(Borderline-SMOTE)处理目标域数据,平衡目标域数据头尾类,缓解长尾分布对偏移距离 d 的影响,为保证目标域尾部类有足够样本进行数据合成,引入缓冲区储存部分目标域尾部类;选

用类平衡损失(CBloss)处理重放损失 L_{replay} 的计算过程,缓解长尾分布对重放损失 L_{replay} 的影响。

2.3.2 Borderline-SMOTE 数据合成

考虑到边缘侧时间资源有限,EdgeTailor 使用简单的数据合成方法,Borderline-SMOTE 处理计算距离度量时的目标域样本。Borderline-SMOTE 的原理可以简述为:寻找边界样本及其临近样本作为基准样本 x_{base} 和参考样本 x_{ref} ,计算它们之间的差值 $d_{\text{diff}} = x_{\text{ref}} - x_{\text{base}}$,将差值乘以 1 个随机数再加到基准样本上,生成新的合成样本 $x_{\text{new}} = x_{\text{base}} + r \times d_{\text{diff}}$ 。EdgeTailor 可以对近邻数和合成样本数量进行设置,使 Borderline-SMOTE 的时间复杂度在可控范围之内。为保证在线学习过程中有足够目标域尾部类数据进行数据合成,长尾分布削减器使用缓冲区存储部分目标域尾部类样本,随着在线学习的过程不断更新。

2.3.3 CBloss 重加权

重加权对重放损失 L_{replay} 的计算过程进行优化,使用 CBloss 计算重放损失 L_{replay} 。CBloss 应用类别平衡因子在聚焦损失^[22]的基础上进行类别平衡。连续领域自适应器中的重放损失 L_{replay} 计算过程如下:

$$L_{\text{replay}}(C(M_i(X_p)), Y_p) = \frac{1-\beta}{1-\beta^\gamma} (1-C(M_i(X_p)))^\gamma \circ \ln C(M_i(X_p)) \circ Y_p \quad (1)$$

式中: \circ 表示矩阵内逐元素相乘; $\frac{1-\beta}{1-\beta^\gamma}$ ($\beta \in [0,1)$) 表示类别平衡因子; γ 表示焦点损失难易样本权重。重加权的不会时间复杂度产生影响,对计算资源的影响极低。

综合来看,长尾分布削减器同时使用 Borderline-SMOTE 数据合成和 CBloss 重加权对连续领域自适应过程进行长尾分布削减。长尾分布削减器可以和连续领域自适应器有效结合,使其在充分发挥效果的同时将计算开销控制在较小范围内。

3 实验评估

本节使用 3 组存在域偏移的长尾数据集和 5 种深度神经网络,将本文方法 EdgeTailor 与基准方法在准确率和计算开销 2 个方面进行实验对比,并对结果进行分析。

3.1 实验设置

3.1.1 实验环境

本文在边缘设备 NVIDIA AGX Orin 上进行实验,该设备采用 ARM 架构,具有 8 核处理器和

32 GB RAM。

3.1.2 数据集

本文一共选用 7 个数据集: CIFAR10, STL10, CIFAR10C, MNIST, SVHN, USPS 和 iNaturalist 2018。第 1 组为 CIFAR10, STL10 和 CIFAR10C,它们包括汽车、鸟、猫等 10 个物体类别,它们样本之间存在域偏移;第 2 组为 MNIST, SVHN 和 USPS,它们包括数字从 0~9 这 10 个类,且它们的样本之间存在域偏移;第 3 组为 iNaturalist2018,它是 1 个大规模的真实世界数据集,包含 437 513 个训练图像和 24 426 个测试图像,共有 8 142 个类,同时数据分布呈现长尾,其长尾失衡比为 500。

本文通过对原数据集进行简单随机采样,减少选定的尾部类样本数量来获取长尾数据。产生的长尾数据集通过在尾部加上“LT”来命名,得到长尾数据集 CIFAR10LT, STL10LT, CIFAR10CLT, MNISTLT, SVHNLT 和 USPSLT。本文实验通过失衡比 ρ 来调整类别分布的长尾程度。类别 i 的样本数按照 $N_i = N_0 \cdot \rho^{\frac{i}{n-1}}$ 递减,其中 n 为类别数, N_0 为平衡数据集下各个类的样本量。本文实验在边缘设备上,资源有限,因此使用 iNaturalist2018 的子集,并使用与 iNaturalist2018 存在抗性过滤偏移的数据集 iNaturalist2018-A。

本文设置 3 组数据集:第 1 组以 CIFAR10 为源域数据, STL10LT、CIFAR10CLT 为目标域数据;第 2 组以 MNIST 为源域数据, SVHNLT、USPSLT 为目标域数据;第 3 组以 iNaturalist 2018 为源域数据, iNaturalist2018-A 为目标域数据。这样设置满足源域和目标域间的连续域偏移性,且目标域数据为长尾分布。

3.1.3 评估指标

本文通过 2 个评价指标进行全面地评估不同方法的性能。

1) 平均 Top-1 准确率 $\text{Acc}_{\text{Top-1}}^{\text{avg}}$ 。

为有效评估 EdgeTailor 和基线训练出的模型在目标域上的整体效果,本文选用模型在各个目标域上的平均 Top-1 准确率来评估模型整体的准确率,即在完成对最后一批目标域数据的训练后,在所有的目标域 $n \in \{1, 2, \dots, N\}$ 上进行测试,得到各个域上 Top-1 准确率 $\text{Acc}_{\text{Top-1}}^n$ 的均值,即:

$$A_{\text{Acc}_{\text{Top-1}}^{\text{avg}}} = \frac{1}{N} \sum_{n=1}^N A_{\text{Acc}_{\text{Top-1}}^n} \quad (2)$$

式中: $A_{\text{Acc}_{\text{Top-1}}^{\text{avg}}}$ 表示平均 Top-1 准确率; $A_{\text{Acc}_{\text{Top-1}}^n}$ 表示各个域上 Top-1 准确率。

2)每秒浮点运算次数与参数量。

为评估 EdgeTailor 和基线的计算开销,本文记录它们在训练过程中的每秒浮点运算次数(FLOPs)和参数量。FLOPs 可以衡量计算设备在单位时间内执行浮点运算来评估模型的计算复杂度,参数量可以评估模型复杂度和规模。FLOPs 和参数量越大意味着模型的计算复杂度越高,运行消耗更多的计算资源。

3.1.4 基线

本文使用多种长尾分布削减技术分别与 CUA 结合生成的派生模型在边缘侧对长尾数据集进行领域自适应作为基线。其中选用 8 种长尾分布削减技术对在线连续领域自适应进行长尾分布削减。

1)数据优化包括数据合成 DeepSMOTE 和 CFA。它们通过合成目标域数据和重放数据尾部样本来优化样本分布,从而缓解连续领域自适应中的长尾分布。

2)训练过程优化包括重加权 IBloss、CBloss 和对数调整 GCLloss、LA。重加权 IBloss、CBloss 根据重放样本数量对类进行加权,影响重放损失的计算;对数调整 GCLloss、LA 通过对重放样本的输出对数概率进行调整来影响重放损失的计算。

3)复合方法中 WB 使用重加权和模型优化, MiSLA 使用模型优化、对数调整。WB 和 MiSLA 除了上述的优化方式以外,还会对深度神经网络的分类器进行训练。

3.1.5 神经网络及实验流程

本文选用的 5 个神经网络 ResNet 20^[26]、Vision Transformer^[27]、MobileNetV3^[28]、RegNet^[29] 和 EfficientNet^[30] 作为网络骨干。其中 ResNet 20 和 Vision Transformer 是经典的视觉识别深度神经网络, MobileNetV3、RegNet 和 EfficientNet 是应用广泛的轻量深度神经网络。为排除离线源域预训练对 EdgeTailor 和基线实验结果的影响,实验中 EdgeTailor 和基线使用相同的源域预训练深度神经网络模型。

实验流程为先使用 80 000 个源域样本进行离线预训练。在预训练模型的基础上分别在每个目标域上选用 5 000 个样本进行在线训练。训练过程中批大小设置为 32,选用 SGD 优化器,其中学习率设置为 0.000 1,动量设置为 0.9。EdgeTailor 的重放损失超参数 $\lambda=0.01$ 。

3.2 结果及分析

3.2.1 准确率结果及分析

不同方法在 ResNet 模型上的平均 Top-1 准确率对比如表 2 所示,加粗表示最优数据。

表 2 不同方法的平均 Top-1 准确率结果

Table 2 Average Top-1 accuracy results of different methods

方法	第 1 组数据集	第 2 组数据集	第 3 组数据集	%
DeepSMOTE ^[14]	74.35	63.51	42.64	
CFA ^[15]	73.41	63.84	41.42	
IBloss ^[16]	68.16	62.61	40.23	
CBloss ^[17]	67.20	61.37	40.22	
GCLloss ^[18]	69.11	62.24	40.36	
LA ^[19]	68.78	60.97	40.32	
WB ^[20]	71.87	63.14	41.23	
MiSLA ^[21]	70.71	63.21	40.32	
EdgeTailor (ours)	78.97	67.03	43.23	

1)现有方法的结果分析。

重加权 IBloss、CBloss 和对数调整 GCLloss、LA 长尾分布削减方法应对领域自适应中的长尾分布表现不佳,在第 1 组和第 2 组数据集中平均 Top-1 准确率分别没超过 70%和 63%,明显低于其他方法。它们的优化方式相似,都无法对偏移距离进行计算,只能优化重放损失,即无法与领域自适应有效结合,无法充分发挥作用,因此优化效果较弱。模型优化包括 WB、MiSLA,它们对长尾分布的优化效果优于重加权和和对数调整,能够在优化重放损失的同时优化分类器。此外,数据合成 DeepSMOTE 和 CFA 直接对源域数据和重放数据进行优化,平衡头尾类,能与领域自适应充分结合,因此相对上述方法效果较好。

2)本文方法的结果分析。

相比基线,EdgeTailor 的长尾分布削减技术和领域自适应方法高效结合。EdgeTailor 的长尾分布削减技术结合了数据合成和重加权,对源域数据进行数据合成,对重放损失的计算进行重加权。它在 3 组实验中平均 Top-1 准确率分别是 78.97%、67.03%和 43.23%,相比重加权和和对数调整 Top-1 准确率提升了 13.50%、7.81%和 6.82%,相比数据合成基线有 6.45%、5.01%和 2.78%的模型准确率提升。EdgeTailor 在应对连续领域自适应中的长尾分布问题时相比基线的准确率得到提升,平均提升为 8.10%。

为了进一步了解 EdgeTailor 和基线在整个连续领域自适应过程中各个目标域上的表现,本文还记录了模型在每个目标域上进行在线领域自适应后的准确率。图 3 所示为在第 1 组数据集和 ResNet 模型上,不同方法在各个目标域上的 Top-1 准确率结果。实验结果表明,EdgeTailor 在各个目标域上

的 Top-1 准确率都超过 75%，相比基线在各个目标域上的 Top-1 准确率均取得优势，说明 EdgeTailor 相比基线在各个目标域上的提升较为平缓稳定。换言之，EdgeTailor 相比基线在连续目标域上的准确率方面具有稳定优势。

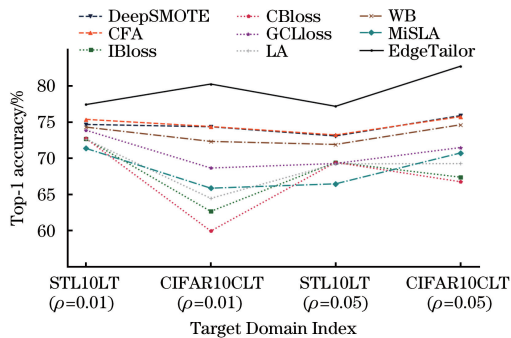


图 3 不同方法在各个目标域上的 Top-1 准确率结果
Fig. 3 Top-1 accuracy results of different methods on various target domains

表 3 不同方法在深度神经网络上平均 Top-1 准确率结果

Table 3 Average Top-1 accuracy results of different methods on deep neural networks

方法	ResNet	Vision Transformer	MobileNetV3	RegNet	EfficientNet	%
DeepSMOTE ^[14]	74.35	62.35	57.38	67.07	68.37	
CFA ^[15]	73.41	62.73	59.97	66.06	67.67	
IBloss ^[16]	68.16	61.54	58.72	65.99	67.50	
CBloss ^[17]	67.20	62.30	58.14	66.46	67.79	
GCLloss ^[18]	69.11	61.62	57.16	66.49	67.25	
LA ^[19]	68.78	61.03	57.91	65.09	67.70	
WB ^[20]	71.87	61.71	59.19	66.97	67.97	
MiSLA ^[21]	70.71	61.49	58.93	66.21	67.16	
EdgeTailor (ours)	78.97	63.06	63.84	69.28	71.19	

3.2.2 计算开销结果及分析

表 4 展示了在第 1 组实验数据和 ResNet 模型上，不同方法的计算开销对比结果。

表 4 不同方法的计算开销结果

Table 4 Computational cost results of different methods

方法	平均 Top-1 准确率/%	GFLOPs	参数量
DeepSMOTE ^[14]	74.35	4.53	272 474
CFA ^[15]	73.41	3.48	272 474
IBloss ^[16]	68.16	2.64	272 474
CBloss ^[17]	67.20	2.64	272 474
GCLloss ^[18]	69.11	2.64	272 474
LA ^[19]	68.78	2.68	272 474
WB ^[20]	71.87	2.64	272 474
MiSLA ^[21]	70.71	2.65	273 207
EdgeTailor (ours)	78.97	2.81	272 474

1) FLOPs 分析。

FLOPs 表示设备在单位时间内执行浮点运算次数。测试不同方法在 ResNet 模型上运行时

为了解 EdgeTailor 和基线在不同深度神经网络中的效果，本文使用第 1 组实验数据在 5 个深度神经网络模型上进行实验，并记录了它们在目标域上的平均 Top-1 准确率。表 3 所示为不同方法在深度神经网络上平均 Top-1 准确率。EdgeTailor 在轻量且结构简单的深度神经网络 ResNet, MobileNetV3, RegNet 和 EfficientNet 上相比基线有明显的优势，平均 Top-1 准确率提升分别是 10.79%, 8.48%, 4.31%, 5.29%，但在体量相对较大且结构复杂的深度神经网络 Vision Transformer 上提升不明显，仅为 1.92%。其原因为 Vision Transformer 和传统的深度神经网络结构有很大不同，使用子注意力机制及有着较高的网络深度和广度。上述结果说明 EdgeTailor 更适用于结构简单的轻量深度神经网络，在轻量深度神经网络上能取得优于基线的平均 Top-1 准确率。

FLOPs 可以直观展示在模型运行占用时的计算资源，用于对比不同方法的计算开销。基于数据优化的长尾分布削减技术会对实验数据进行处理，这种方式有很好的长尾分布削减效果，但数据合成过程需要较大的计算量，因此有较大的 FLOPs。基于训练过程优化的长尾分布削减技术只对重放损失的计算过程进行优化，或者额外优化分类器，这些方法无法与领域自适应方法高效结合，但不会产生大量计算需求，因此它们有着较低的准确率和 FLOPs。而 EdgeTailor 则结合数据优化和训练过程优化方法，在训练过程优化的重放训练部分使用重加权优化，只能数据优化的域偏移量计算部分使用数据合成进行优化。在保证准确率的同时，尽量降低 EdgeTailor 的 FLOPs。从表 4 可以看出，EdgeTailor 在准确率大于只使用数据合成的基线时，FLOPs 明显低于基线的 FLOPs，大约减少了 29.84%。

2) 参数量分析。

参数量指深度神经网络模型中的可训练参数数

量,以评估模型复杂度和规模。从表 4 可以看出,当不同方法在模型 ResNet 上运行时,包括 EdgeTailor 在内的大多数方法不改变参数量,只有 MiSLA 因为额外引入分类器导致参数量增加。EdgeTailor 相比数据合成和训练过程优化方法没有参数量的提升,说明 EdgeTailor 不会直接影响模型的复杂度和规模。

综上所述,EdgeTailor 通过将数据合成方法和重加权方法作为长尾分布削减策略结合连续领域自适应,有助于深度神经网络模型在具有连续域偏移和长尾分布的数据中进行学习,相比基线在目标域上提升了 8.10% 的平均 Top-1 准确率。同时 EdgeTailor 严格控制计算开销,FLOPs 和参数量均处于较低水平,其中 FLOPs 相比精确度较高的数据合成方法平均减少了 29.84%。EdgeTailor 在目标域上提高准确率的同时将计算开销控制在相对较低的水平。

4 结束语

深度学习在边缘侧面临连续域偏移、长尾分布和计算资源有限 3 个方面的挑战。为应对当前长尾分布削减技术无法在领域自适应方法中充分发挥作用的问题,本文提出结合长尾分布削减技术的领域自适应方法 EdgeTailor,将数据合成方法和重加权方法作为长尾分布削减方法与无监督连续域自适应方法有效结合,让深度神经网络在边缘侧数据具有连续域偏移和长尾分布的情况下进行在线学习。同时,EdgeTailor 通过缩小数据合成方法的使用范围,在保证模型性能的情况下,将计算开销控制在相对较低的水平。在不同数据集和不同深度神经网络上的实验结果表明,EdgeTailor 具有较优的性能。本文工作只专注于边缘侧视觉识别中的多分类任务,适用范围存在一定的局限性。后续将研究在目标检测、图像分割等视觉识别任务上进行领域自适应时存在的长尾分布问题解决方案。同时,将 EdgeTailor 应用于领域自适应时计算开销依然会有所增加,因此对计算开销的进一步优化也是本文重点的研究方向。

参考文献

- [1] ZHOU Z, CHEN X, LI E, et al. Edge intelligence: paving the last mile of artificial intelligence with edge computing[J]. *Proceedings of the IEEE*, 2019, 107(8): 1738-1762.
- [2] WANG X F, HAN Y W, LEUNG V C M, et al. Convergence of edge computing and deep learning: a comprehensive survey[J]. *IEEE Communications Surveys & Tutorials*, 2020, 22(2): 869-904.
- [3] 郭斌,刘思聪,刘琰,等.智能物联网:概念、体系架构与关键技术[J].*计算机学报*,2023,46(11):2259-2278.
GUO B, LIU S C, LIU Y, et al. AIoT: the concept, architecture and key techniques[J]. *Journal of Computer Science and Technology*, 2023, 46(11): 2259-2278. (in Chinese)
- [4] 张祥俊,伍卫国,张弛,等.面向移动边缘计算网络的高能效计算卸载算法[J].*软件学报*,2023,34(2):849-867.
ZHANG X J, WU W G, ZHANG C, et al. Energy-efficiency computing offloading algorithm for mobile edge computing networks[J]. *Journal of Software*, 2023, 34(2): 849-867. (in Chinese)
- [5] 李亚国,李冠良,张凯,等.基于人工智能与边缘代理的物联网框架设计[J].*计算机工程*,2023,49(10):313-320.
LI Y G, LI G L, ZHANG K, et al. Design of Internet of Things framework based on artificial intelligence and edge agents[J]. *Computer Engineering*, 2023, 49(10): 313-320. (in Chinese)
- [6] LIU X, YOO C, XING F, et al. Deep unsupervised domain adaptation: a review of recent advances and perspectives[J]. *APSIPA Transactions on Signal and Information Processing*, 2022, 11(1):92-96.
- [7] KOUW W M, LOOG M. A review of domain adaptation without target labels [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 43(3): 766-785.
- [8] WANG M, DENG W H. Deep visual domain adaptation: a survey[J]. *Neurocomputing*, 2018, 312: 135-153.
- [9] ZHANG Y F, KANG B Y, HOUI B, et al. Deep long-tailed learning: a survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(9): 10795-10816.
- [10] YANG L, JIANG H, SONG Q, et al. A survey on long-tailed visual recognition [J]. *International Journal of Computer Vision*, 2022, 130(7): 1837-1872.
- [11] 袁佳伟,宋庆增,王雪纯,等.边缘计算设备的性能功耗测量与分析[J].*计算机工程*,2021,47(2):233-238,245.
YUAN J W, SONG Q Z, WANG X C, et al. Performance and power consumption measurement and analysis of edge computing devices[J]. *Computer Engineering*, 2021, 47(2): 233-238,245. (in Chinese)
- [12] BOBU A, TZENG E, HOFFMAN J, et al. Adapting to continuously shifting domains [EB/OL]. [2023-12-18]. <https://openreview.net/pdf?id=BJsBjPJvf>.
- [13] WANG Q, FINK O, VAN GOOL L, et al. Continual test-time domain adaptation[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2022: 7201-7211.
- [14] DABLAIN D, KRAWCZYK B, CHAWLA N V. DeepSMOTE: fusing deep learning and SMOTE for imbalanced data[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 34(9): 6390-6404.
- [15] TEMRAZ M, KEANE M T. Solving the class imbalance problem using a counterfactual method for data augmentation [EB/OL]. [2023-12-18]. <https://arxiv.org/abs/2111.03516?context=cs.AI>.
- [16] PARK S, LIM J, JEON Y H, et al. Influence-balanced loss for imbalanced visual classification[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Washington D. C., USA: IEEE Press, 2021: 735-744.
- [17] 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*. Washington D. C., USA: IEEE Press, 2019: 9268-9277.
- [18] LI M K, CHEUNG Y M, LU Y. Long-tailed visual recognition via Gaussian clouded logit adjustment [C]//*Proceedings of the IEEE/CVF Conference on Computer*

- Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2022: 6929-6938.
- [19] MENON A K, JAYASUMANA S, RAWAT A S, et al. Long-tail learning via logit adjustment[EB/OL]. [2023-12-18]. <https://arxiv.org/abs/2007.07314>.
- [20] ALSHAMMARI S, WANG Y X, RAMANAN D, et al. Long-tailed recognition via weight balancing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2022: 6897-6907.
- [21] ZHONG Z S, CUI J Q, LIU S, et al. Improving calibration for long-tailed recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2021: 16489-16498.
- [22] 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. Washington D. C. , USA; IEEE Press, 2017: 2980-2988.
- [23] ZHANG C, PAN T Y, LI Y, et al. Mosaicos: a simple and effective use of object-centric images for long-tailed object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA; IEEE Press, 2021: 417-427.
- [24] 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. Washington D. C. , USA; IEEE Press, 2021: 3784-3793.
- [25] HAN H, WANG W Y, MAO B H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning [C] // Proceedings of International Conference on Intelligent Computing. Berlin, Germany: Springer, 2005: 878-887.
- [26] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2016: 770-778.
- [27] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words; Transformers for image recognition at scale[EB/OL]. [2023-12-18]. <https://arxiv.org/abs/2010.11929v2>.
- [28] HOWARD A, SANDLER M, CHU B, et al. Searching for MobileNetV3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA; IEEE Press, 2019: 1314-1324.
- [29] PESCHL M F, FUNDNEIDER T. Why space matters for collaborative innovation networks: on designing enabling spaces for collaborative knowledge creation[J]. International Journal of Organisational Design and Engineering, 2014, 3(3/4): 358-391.
- [30] TAN M X, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[EB/OL]. [2023-12-18]. <https://arxiv.org/abs/1905.11946?context=cs.LG>.

编辑 薛晋栋