

样本-特征协同的长尾识别算法^{*}

张恩豪¹, 李超华^{2,3,4}, 王志华¹, 陈松灿^{3,4}

¹(郑州大学 网络空间安全学院, 河南 郑州 450002)

²(西安邮电大学 计算机学院, 陕西 西安 710121)

³(南京航空航天大学 计算机科学与技术学院, 江苏 南京 211106)

⁴(模式分析与机器智能工业和信息化部重点实验室 (南京航空航天大学), 江苏 南京 211106)

通信作者: 陈松灿, E-mail: s.chen@nuaa.edu.cn



摘 要: 现实场景中的数据分布普遍呈现长尾模式, 导致所训练的深度学习模型常遭遇头部偏置困境, 即模型偏向于头部类而在尾部类上表现欠佳. 一种简单且有效的策略是通过增广尾部类样本来平衡数据分布. 尽管多数基于该策略的方法能在样本数量上实现重平衡, 但生成的样本仍存在语义漂移和多样性不足的风险, 导致特征分布松散和决策边界有偏, 限制了模型性能. 为此, 提出一种样本-特征协同的长尾识别算法, 旨在构建对数据分布不敏感的分类模型. 具体而言, 在样本层面, 基于傅里叶变换提出“幅值增广”策略, 借助幅值迁移来转换增广样本的风格, 能够在丰富样本多样性的同时保留原有语义信息; 在特征层面, 依据神经坍塌理论提出“特征坍塌”损失, 将类原型对齐至具有最大可分性的等角紧框架, 并促使特征收敛至对应的类原型, 实现最大的类间间隔. 该方法从样本和特征两个层面缓解头部偏置问题, 进而增强类内紧凑性并校准决策边界. 多个基准数据集上的实验结果表明该方法可显著提高长尾识别性能.

关键词: 长尾识别; 傅里叶变换; 神经坍塌; 数据增广

中图法分类号: TP18

中文引用格式: 张恩豪, 李超华, 王志华, 陈松灿. 样本-特征协同的长尾识别算法. 软件学报. <http://www.jos.org.cn/1000-9825/7629.htm>

英文引用格式: Zhang EH, Li CH, Wang ZH, Chen SC. Sample-feature Collaborative Long-tailed Recognition Algorithm. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/7629.htm>

Sample-feature Collaborative Long-tailed Recognition Algorithm

ZHANG En-Hao¹, LI Chao-Hua^{2,3,4}, WANG Zhi-Hua¹, CHEN Song-Can^{3,4}

¹(School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China)

²(School of Computer Science & Technology, Xi'an University of Posts & Telecommunications, Xi'an 710121, China)

³(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

⁴(MIT Key Laboratory of Pattern Analysis and Machine Intelligence (Nanjing University of Aeronautics and Astronautics), Nanjing 211106, China)

Abstract: Data distributions in real-world scenarios commonly exhibit long-tail patterns, causing deep models to suffer from head bias, where performance is biased toward head classes while tail classes are poorly recognized. A simple and effective strategy to alleviate this problem is to balance data distributions by augmenting tail-class samples. Although most methods following this strategy achieve a quantitative rebalancing, the generated samples often suffer from semantic shift and insufficient diversity, which leads to dispersed feature distributions and biased decision boundaries, thereby limiting overall model performance. To address these issues, this study proposes a sample-feature collaborative learning framework for long-tailed recognition, aiming to construct classification models that are insensitive to

* 基金项目: 国家自然科学基金 (62376126); 航空发动机及燃气轮机重大专项基础研究项目 (J2019-IV-0018-0086)

收稿时间: 2025-09-30; 修改时间: 2025-12-01; 采用时间: 2025-12-29; jos 在线出版时间: 2026-07-08

data distribution imbalance. At the sample level, a “magnitude augmentation” strategy based on the Fourier transform is introduced, in which amplitude shifting is employed to modify the style of augmented samples, while preserving their original semantic information. At the feature level, a “feature collapse” loss inspired by neural collapse theory is proposed to align class prototypes into an equiangular tight frame with maximum separability and to encourage features to converge toward their corresponding class prototypes, achieving maximum inter-class separation. By jointly addressing head bias from both the sample and feature perspectives, the proposed framework enhances intra-class compactness and calibrates decision boundaries. Experimental results across multiple benchmark datasets demonstrate that the proposed method significantly improves long-tailed recognition performance.

Key words: long-tailed recognition (LTR); Fourier transform; neural collapse (NC); data augmentation

深度神经网络 (deep neural network, DNN) 在图像分类^[1]、目标检测^[2,3]、语义分割^[4,5]等多个领域已取得突破性进展,这与可供使用的大规模标注数据密不可分。这些数据通常经过手工预处理使其近似服从均匀分布,典型的如 CIFAR^[6]和 ImageNet^[7]数据集。然而,受广泛存在的幂律分布影响,现实应用中的数据通常遵循长尾分布:少数类别拥有大部分样本(头部类),而剩余的大量类别仅包含较少的样本(尾部类)^[8-10]。如图 1(a) 所示,在现实场景中,由于猫的样本数量远超过熊猫,致使收集到的数据呈现出长尾模式。尾部类(熊猫)样本的匮乏使其真实分布难以被充分学习,因此其决策边界受到头部类(猫)的挤压,最终导致模型过度偏好于头部类而在尾部类上表现不佳^[11,12]。因此,长尾识别 (long-tailed recognition, LTR) 任务近年来受到广泛关注。

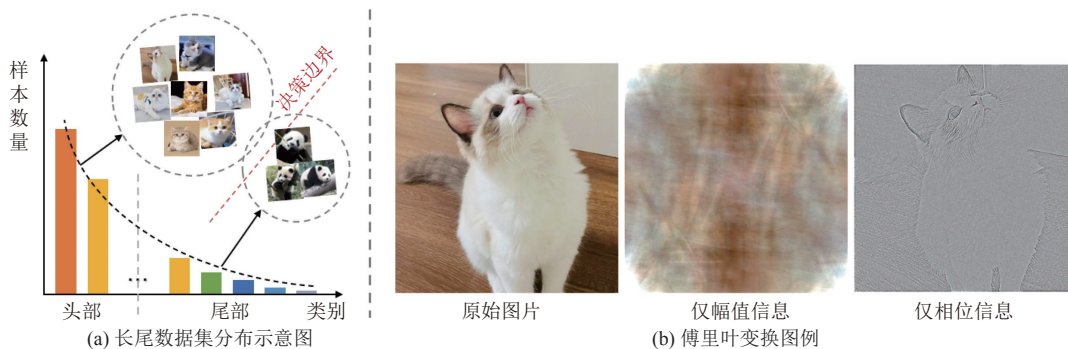


图 1 长尾数据集分布示意图及傅里叶变换图例

为应对 LTR 问题,研究者们已提出诸多策略。其中通过增广尾部类样本来重平衡数据分布是一种简单有效的策略,常见的数据增广方法包括随机翻转、缩放和添加噪声等。然而,由于尾部类样本匮乏,由此获得的增广样本常缺乏多样性信息,导致模型易遭遇过拟合风险。为缓解该问题,受 Mixup 及其变体的启发^[13],部分研究期望借助在数据/特征与对应标签进行线性插值构建混合样本,用于增广尾部类,以缓解样本不平衡问题。例如, Xu 等人^[14]提出 UniMix 算法,其采用偏向于少数类的混合权重和采样器随机生成更多的尾部类样本; Park 等人^[15]提出 CMO 算法,其通过混合背景和前景图像生成以尾部类为中心的样本,进而丰富尾部类的语义信息。虽然这些方法能在一定程度上缓解长尾问题,但是生成的样本具有随机性无法保证其真实性和合理性。为此, Gao 等人^[16]提出 OTmix 算法,其额外考虑前景图像和背景图像之间的语义距离,以混合生成更具确定性和真实性的尾部类样本。

尽管上述方法能在一定程度上提升尾部类的多样性,但是它们选择用于混合的样本/特征具有不同的自身语义信息,生成的图像可能存在语义漂移风险。如图 2(上)所示,基于 Mixup 生成的图像同时具备猫和狗的信息,这类模棱两可的样本可能会造成松散的特征分布和模糊的分类边界,难以提升模型性能。鉴于此,本文提出一种样本-特征协同的长尾识别 (a sample-feature collaborative learning framework for long-tailed recognition, ASFC) 算法,其既能丰富尾部类多样性、规避语义漂移风险,又能强化特征的类内紧凑性和类间可分性,进而习得对数据分布不敏感的识别模型。

ASFC 方法的核心在于通过样本与特征两个层面的协同创新来应对 LTR 问题。首先,在样本层面,本文提出“幅值增广 (amplitude augmentation, AmpAug)”机制。AmpAug 基于傅里叶变换的特性 (相位分量承载高级语义信

息, 而幅值分量主导低级背景统计信息^[17], 如图 1(b) 所示, 其中“仅幅值信息”表示将相位分量设为常数, 重构出具有幅值信息的图像; “仅相位信息”表示类似的操作), 通过计算类间语义信息, 为目标类 (主要是尾部类) 样本选择风格匹配的候补类样本, 并进行幅值迁移, 从而生成背景多样但语义一致的高质量样本, 如图 2(下) 所示. 该机制不仅增广了尾部类样本, 而且从根源上避免了语义漂移风险. 进一步, 在特征层面, 为确保增广样本被有效利用且习得高度可分的特征, 本文受“神经坍塌 (neural collapse, NC)”现象^[18]启发, 设计一种“特征坍塌 (feature collapse, FC)”损失. NC 现象揭示了在训练的最终阶段, 特征、类均值和分类器向量均收敛至一个单纯形等角紧框架 (equiangular tight frame, ETF) 的顶点, 达到最优的类间可分性, 但长尾分布会破坏该结构^[18,19]. 为此, FC 通过将类原型对齐至 ETF 顶点, 并约束特征向量收敛至其类原型, 以在长尾背景下获得最优的可分离特征. AmpAug 机制与 FC 损失分别从外部数据增广和内部表征结构两个层面相互协作, 促使 ASFC 方法极大缓解长尾问题所带来的偏差问题, 进而显著提升模型性能.

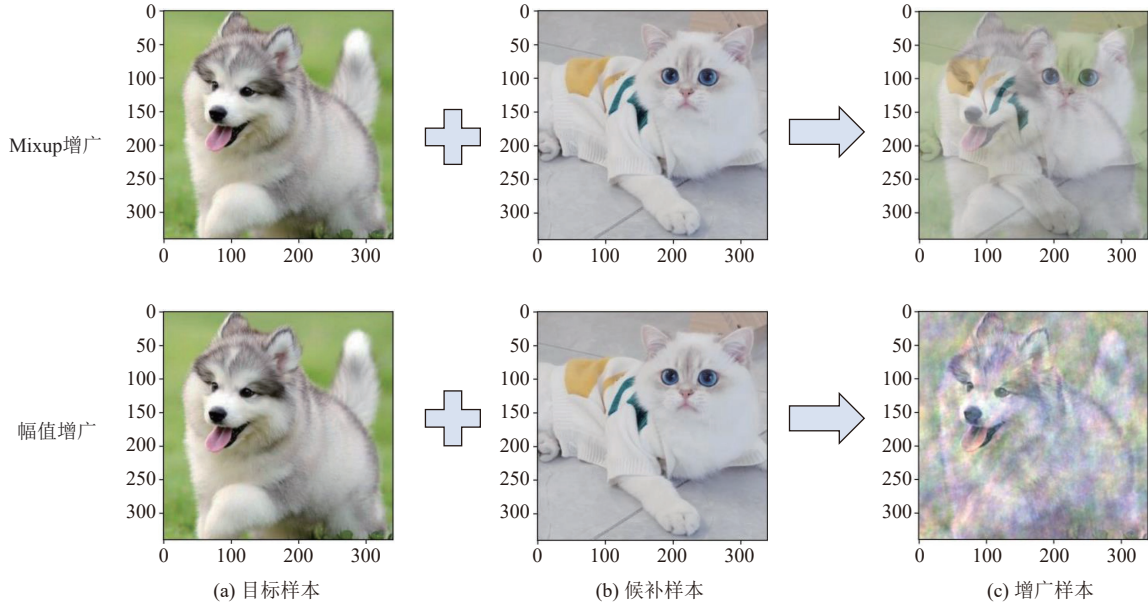


图 2 不同增广方式的可视化结果

本文的主要贡献可以概括如下.

(1) 提出一种样本-特征协同的无偏长尾识别算法, 其同时考虑样本和特征两个层面, 不仅可以避免过拟合和语义漂移风险, 而且可以有效缓解模型对头部类的偏好.

(2) 基于傅里叶变换设计一种“幅值增广”机制, 使得所生成的样本既能丰富目标类的多样性又能保持原有的语义信息. 进一步, 本文借助“神经坍塌”现象构造一种“特征坍塌”损失, 促使在不平衡背景下, 所学特征能与最优的分类结构对齐.

(3) 在大量的基准数据集上的实验表明, ASFC 的模型精度均高于所比较的基线方法. 此外, 进一步的实验表明 ASFC 能够获得近似的分类性能, 验证了其能显著缓解模型偏差问题.

本文第 1 节回顾与本文研究密切相关的工作. 第 2 节介绍本文方法所涉及的相关问题定义和预备知识. 第 3 节详细阐述 ASFC 方法的总体架构与实现机制. 第 4 节根据系统的实验验证本文方法的有效性. 最后在第 5 节总结全文并展望未来的研究方向.

1 相关工作

本节简要回顾与本文方法密切相关的几个研究策略, 包括: 重平衡学习、快速傅里叶变换和神经坍塌.

1.1 重平衡学习

现有 LTR 方法旨在通过对头、尾部类给予同等程度的关注,以重平衡模型训练^[20-22].当前广泛采用的策略主要有重采样和重加权两种.尽管这些方法简单有效,但易导致模型获得次优的性能^[23].例如,当采样不足时,可能会遗漏有价值的信息,进而不可避免地降低模型性能;而采样过多则往往会导致尾部类出现过拟合风险.后续研究期望通过丰富尾部类的语义信息,提升模型的泛化能力.例如,Zada 等人^[24]提出了 OPeN 框架,该框架采用来自相同高斯分布的纯噪声图像作为额外的样本来重平衡训练数据. Shi 等人^[25]系统地研究了重采样如何提高 LTR 模型性能.最近,一些研究受 Mixup 及其变体具有良好性能的启发,通过构建混合样本添加至尾部类中,以缓解样本不平衡问题.例如,Zhang 等人^[26]提出 CaUIRL 算法,通过设计高阶 Mixup 算法,构建具有类感知能力 Universum 样本,以从数量和质量两个方面来平衡样本分布.然而,现有基于 Mixup 的 LTR 算法通常将具有不同语义信息的样本进行混合,导致生成的样本易遭遇语义漂移的风险.为此,本文设计一种“幅值增广”机制,仅需迁移目标样本的背景信息即可构造额外的尾部类样本.所生成的样本不仅能丰富目标类的样本多样性而且能在一定程度上避免语义漂移风险.

1.2 快速傅里叶变换

快速傅里叶变换 (fast Fourier transform, FFT) 是一种用于计算离散傅里叶变换及其逆变换的数学算法^[27].在计算机视觉领域中,FFT 通过将原始域(如时间域或空间域)的复杂数据转换至频域并实现反向转换,使其能够从噪声或高维数据集中提取关键特征^[28,29].在 FFT 过程中,所获得的相位分量保留原始输入的高阶语义信息,而幅值分量则包含更多的背景信息,这使其具有很强的跨域泛化能力^[30,31].因此,当前大多数视觉领域算法将 FFT 应用至域自适应或者域泛化等任务.例如,Yang 等人^[32]通过交换源域和目标域的低频谱来缩小两者差异,进而实现域对齐. Xu 等人^[33]利用傅里叶相位信息蕴含高层次语义特征且不易受领域偏移影响的特性提出一种基于傅里叶变换的域泛化模型.尽管 FFT 具有显著的实用性和有效性,但在长尾识别任务中仍未被探索.为此,本文借助 FFT 提出一种“幅值增广”策略,用于生成尾部类样本,进而平衡数据分布.

1.3 神经坍塌

在平衡场景中,Papayan 等人^[18]通过实验观察发现在模型训练的最终阶段,DNN 往往会形成一种特定的几何结构,这种结构被定义为“神经坍塌(NC)”.NC 可以促使特征、类均值和分类器向量达到类间最佳分离状态^[34],详细定义如第 2.3 节所述.随后的工作探究了这一现象的物理机制,推导出 NC 解在某些约束条件下是全局最优的^[35].尽管大量研究从经验和理论上验证了 NC 的有效性,但一旦转向不平衡场景时,少数类对应的类均值就会受到挤压,不可避免地导致这种最佳结构被破坏^[36].因此,现有工作希望通过改进分类器来重构 NC 现象^[37,38].例如,Yang 等人^[39]采用固定的 ETF 分类器代替可学习分类器,试图恢复 NC.不同于此,本文从特征层面提出一种“特征坍塌”损失,将类原型对齐至 ETF 结构,并促使特征收敛至类原型,以达到最优的类间可分性.

2 预备知识

本节介绍长尾识别任务的相关问题设置,以及所涉及的傅里叶变换和神经坍塌的相关概念.

2.1 问题设置

对于给定的不平衡训练数据集 $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$,共包含 C 个类别,每个类由 N_i 个训练样本组成.令 $N = \sum_{i=1}^C N_i$ 表示 $\mathcal{D}$ 中的样本总数.不失一般性,本文假设 $N_1 \geq N_2 \geq \dots \geq N_C$,那么数据集的不平衡率可以表示为 $\beta = \frac{N_1}{N_C}$,而测试集是平衡的.定义 $\mathbf{z} = f(\mathbf{x})$ 为样本 $\mathbf{x}$ 的特征,其中 $f: \mathcal{X} \rightarrow \mathcal{Z}$ 将输入样本投影到特征空间 $\mathcal{Z} \subseteq \mathbb{R}^s$ 上.最终的预测结果是 $\hat{\mathbf{y}} = g(\mathbf{z})$, $g: \mathcal{Z} \rightarrow \mathcal{Y}$ 是分类器.将 $P_{\text{train}}(\mathbf{X}, \mathbf{Y})$ 和 $P_{\text{test}}(\mathbf{X}, \mathbf{Y})$ 分别定义为潜在的训练集分布和测试集分布.通常情况下,长尾识别任务假定测试数据与训练数据具有相同的类条件概率,即 $P_{\text{train}}(\mathbf{X} | \mathbf{Y}) = P_{\text{test}}(\mathbf{X} | \mathbf{Y})$,而它们的类先验概率是不同的,即 $P_{\text{train}}(\mathbf{Y}) \neq P_{\text{test}}(\mathbf{Y})$.由乘法公式可以得到: $P_{\text{train}}(\mathbf{X}, \mathbf{Y}) \neq P_{\text{test}}(\mathbf{X}, \mathbf{Y})$.因此,长尾识别的主要目的是利用不平衡的训练集训练 f 和 g ,从而在平衡的测试集上有更好的泛化效果.

2.2 傅里叶变换

对于一个单信道图片 $\mathbf{x}$, 其对应的傅里叶变换 $\mathcal{F}$ 可以表示为:

$$\mathcal{F}(\mathbf{x})(m, n) = \sum_{h, w} \mathbf{x}(h, w) e^{-j2\pi(\frac{h}{H}m + \frac{w}{W}n)}, j^2 = -1 \quad (1)$$

$\mathcal{F}^{-1}(\mathbf{x})$ 表示对应的反傅里叶变换. 令 $\mathcal{F}^A, \mathcal{F}^P: \mathbb{R}^{H \times W \times 3} \rightarrow \mathbb{R}^{H \times W \times 3}$ 是 $\mathcal{F}$ 所对应的幅值 (amplitude) 和相位 (phase) 部分, 根据 FFT 算法可以得到如下表示:

$$\begin{cases} \mathcal{F}^A(\mathbf{x})(m, n) = [R^2(\mathbf{x})(m, n) + I^2(\mathbf{x})(m, n)]^{1/2} \\ \mathcal{F}^P(\mathbf{x})(m, n) = \arctan \left[\frac{I(\mathbf{x})(m, n)}{R(\mathbf{x})(m, n)} \right] \end{cases} \quad (2)$$

其中, $R(\mathbf{x})$ 和 $I(\mathbf{x})$ 分别表示 $\mathcal{F}(\mathbf{x})$ 的实部和虚部. 对于 RGB 图像, 通过独立计算每个通道的傅里叶变换, 得到相应的幅值和相位信息.

2.3 神经坍塌

Papayan 等人^[18]通过实验发现了 NC 现象, 即在平衡数据集训练的最后阶段, 最后一层的特征收敛到相应的类内均值, 类内均值与分类器向量收敛至同一个单纯形 ETF 的顶点. 下面, 本节首先介绍单纯形 ETF 的相关定义.

定义 1. 单纯形 ETF. 对于一组向量 $\mathbf{v}_c \in \mathbb{R}^q$, $c = 1, 2, \dots, C$ 满足一个单纯形 ETF 结构, 如果有:

$$\mathbf{V} = \sqrt{\frac{C}{C-1}} \mathbf{U} \left(\mathbf{I}_C - \frac{1}{C} \mathbf{1}_C \mathbf{1}_C^T \right) \quad (3)$$

其中, $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_C] \in \mathbb{R}^{q \times C}$, $\mathbf{U} \in \mathbb{R}^{q \times C}$ 满足 $\mathbf{U}^T \mathbf{U} = \mathbf{I}_C$, $\mathbf{I}_C$ 是一个单位矩阵, $\mathbf{1}_C$ 是一个全 1 的向量. 因此, 对于 $\forall i, j \in [1, C]$ 满足:

$$\mathbf{v}_i^T \mathbf{v}_j = \begin{cases} 1, & i = j \\ -\frac{1}{C-1}, & i \neq j \end{cases} \quad (4)$$

那么, NC 现象可以形式化概括为以下 4 个方面.

- (NC1) 变量坍塌: 随着特征坍塌到其类均值, 对应的类内方差趋近于 0:

$$\Sigma_W \rightarrow 0 \quad (5)$$

其中, $\Sigma_W \triangleq \text{Ave}_{i,c} \{(\mathbf{z}_{i,c} - \boldsymbol{\mu}_c)(\mathbf{z}_{i,c} - \boldsymbol{\mu}_c)^T\}$, $\boldsymbol{\mu}_c \triangleq \text{Ave}_i \{\mathbf{z}_{i,c}\}$ ($c = 1, 2, \dots, C$) 是第 c 类的均值. $\mathbf{z}_{i,c}$ 表示第 c 类第 i 个样本对应的最后一层的特征.

- (NC2) 收敛到单纯形 ETF: 被全局中心化后的类均值会收敛到定义 1 中单纯形 ETF 结构的顶点, 即 $\{\tilde{\boldsymbol{\mu}}_c\}_{c=1}^C$ 满足公式 (4), 其中 $\tilde{\boldsymbol{\mu}}_c \triangleq (\boldsymbol{\mu}_c - \boldsymbol{\mu}_G) / \|(\boldsymbol{\mu}_c - \boldsymbol{\mu}_G)\|_2$, $\boldsymbol{\mu}_G \triangleq \text{Ave}_{i,c} \{\mathbf{z}_{i,c}\}$.

- (NC3) 收敛到自对偶: 类均值和分类器向量相互收敛, 直至满足:

$$\Delta = \left\| \frac{\mathbf{W}^T}{\|\mathbf{W}\|_F} - \frac{\tilde{\mathbf{M}}}{\|\tilde{\mathbf{M}}\|_F} \right\|_F \rightarrow 0 \quad (6)$$

其中, $\tilde{\mathbf{M}} = [\boldsymbol{\mu}_c - \boldsymbol{\mu}_G, c = 1, \dots, C]$, $\mathbf{W} = [\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_C]$ 表示分类器向量矩阵, 这确保两类之间的混淆最小化.

- (NC4) 简单的判定规则: 当 NC1–NC3 成立时, 模型预测可简化为近邻的类中心预测:

$$\arg \max_{c'} \langle \boldsymbol{\omega}_{c'}, \mathbf{h} \rangle + b_{c'} \rightarrow \arg \min_{c'} \|\mathbf{h} - \boldsymbol{\mu}_{c'}\|_2 \quad (7)$$

其中, b 表示偏差项, $\mathbf{h}$ 是最后一层进行预测的样本特征.

3 样本-特征协同的无偏长尾识别算法

本文提出一种新颖的样本-特征协同的长尾识别方法 ASFC, 主要包括“幅值增广 (AmpAug)”和“特征坍塌 (FC)”两部分, 二者相互作用, 以缓解不平衡数据所带来的模型偏差. 本节将详细阐述 ASFC 方法的实现细节. 首先,

在第 3.1 节提出 AmpAug 机制, 其在增加尾部类的类内多样性的同时, 有效规避语义漂移风险. 其次, 在第 3.2 节提出 FC 损失, 其通过同时优化特征表示和类均值分布, 引导模型获得最优的类间可分离结构; 最后, 本节集成各部分组件, 给出完整的训练损失函数. 本文方法的总框架如图 3(a) 所示.

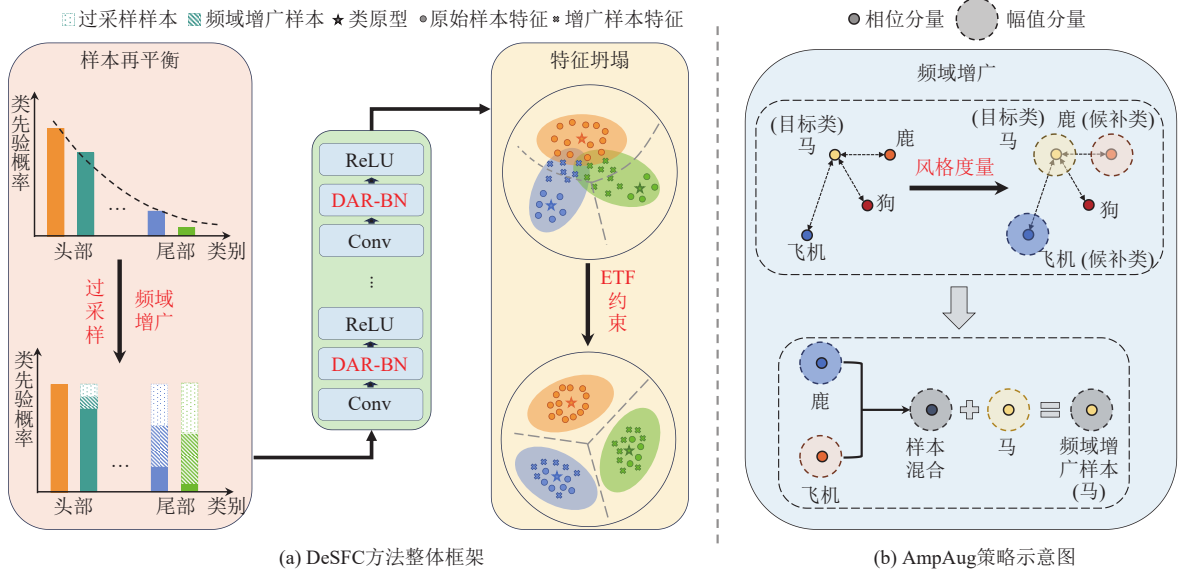


图 3 相关算法图例

3.1 幅值增广 (AmpAug) 策略

基于傅里叶谱中相位分量主导高阶语义信息的特性, 本节设计一种 AmpAug 机制, 其通过迁移幅值信息构建增广样本来重平衡数据分布. 该机制首先为给定的目标类样本, 依据语义风格相似度筛选合适的候补类别; 随后, 从候选类中随机选取样本并进行幅值谱间的线性插值, 进而生成既保留目标类语义, 又融入多样背景风格的增广样本, 如图 3(b) 所示, 其中, 实线的小圆表示样本的相位信息, 虚线的大圆表示幅值信息. 最终, 本文使用混合样本的幅值信息与目标样本的相位信息进行增广, 获得频域增广样本.

具体地, 对于给定的不平衡数据集 $\mathcal{D}$, 分别计算每类的类内均值 $\{\mu_i\}_{i=1}^C$, 其中 $\mu_c = \frac{1}{N_c} \sum_{i=1}^{N_c} \mathbf{z}_i$, 并将其作为对应的类原型. 根据类原型的余弦相似度分别计算第 i 类对应的最小语义类 $\phi_{\min}(i)$ 和最大语义类 $\phi_{\max}(i)$ 作为候补类, 可以形式化为:

$$\mu_{\phi_{\min}(i)} = \arg \min_{j=1, \dots, C, j \neq i} \cos(\mu_i, \mu_j), \mu_{\phi_{\max}(i)} = \arg \max_{j=1, \dots, C, j \neq i} \cos(\mu_i, \mu_j) \quad (8)$$

给定目标类 c 中的样本 $\mathbf{x}_i^c$, 分别从候补类中随机采样两个样本 $\mathbf{x}_j^{\phi_{\min}(c)}$ 和 $\mathbf{x}_k^{\phi_{\max}(c)}$, 并根据公式 (9) 计算候选幅值信息 $\mathcal{F}^A(\tilde{\mathbf{x}}_i^c)$:

$$\mathcal{F}^A(\tilde{\mathbf{x}}_i^c) = \alpha \mathcal{F}^A(\mathbf{x}_j^{\phi_{\min}(c)}) + (1 - \alpha) \mathcal{F}^A(\mathbf{x}_k^{\phi_{\max}(c)}) \quad (9)$$

其中, α 是可调节参数, 用于控制丰富背景信息的强度. 之后对样本 $\mathbf{x}_i^c$ 进行幅值增广, 可以表示为:

$$\tilde{\mathcal{F}}^A(\mathbf{x}_i^c) = \lambda \mathcal{F}^A(\mathbf{x}_i^c) + (1 - \lambda) \mathcal{F}^A(\tilde{\mathbf{x}}_i^c)$$

其中, $\lambda \sim U(0, \eta)$, η 控制增广强度. 将混合振幅与原始相位相结合, 从而形成一个新的傅里叶表示形式:

$$\mathcal{F}(\hat{\mathbf{x}}_i^c)(m, n) = \tilde{\mathcal{F}}^A(\mathbf{x}_i^c)(m, n) \times e^{-j \times \tilde{\mathcal{F}}^P(\mathbf{x}_i^c)(m, n)}$$

然后, 根据反傅里叶变换 $\mathcal{F}^{-1}$, 生成增广样本 $\hat{\mathbf{x}}_i^c$:

$$\hat{\mathbf{x}}_i^c = \mathcal{F}^{-1}[\mathcal{F}(\hat{\mathbf{x}}_i^c)(m, n)] \quad (10)$$

本文将这种基于傅里叶变化的增广策略记为“幅值增广”, $\hat{\mathbf{x}}_i^c$ 为对应的增广样本.

最终, 根据伯努利分布将一部分过采样得到的样本替换为“幅值增广”生成的样本, 即:

$$\text{Bernoulli}(p_c) = \begin{cases} 1, & \mathbf{x}_i^c \rightarrow \hat{\mathbf{x}}_i^c \\ 0, & \text{未替换} \end{cases},$$

其中, $p_c = 1 - N_c/N_1$ ($c = 1, 2, \dots, C$) 为类内表征率. 当类内样本量 N_c 越大时, 那么被增广样本替换的概率越小, 反之亦然. 这能够保证尾部类获得更多的增广样本, 以丰富类内样本多样性. 不难发现, 增广样本 $\hat{\mathbf{x}}$ 仅改变原始样本的幅值信息, 而相位信息保持不变, 以缓解语义漂移问题. 此外, 由于相位混合可能产生一定程度的域漂移问题, 本文使用 DAR-BN^[24] 替换网络中原始的 BN 层. DAR-BN 通过感知训练样本的分布执行分批正则化, 以处理来自不同分布/域的输入样本, 进而提高模型的泛化能力并抑制样本过拟合问题.

3.2 特征坍塌 (FC) 损失

本节借助“神经坍塌”理论设计一种新的“特征坍塌”损失, 旨在将类原型对齐至 ETF 结构并同时约束特征收敛至对应的类原型, 以获得具有最优类间可分性和类内紧性的特征簇 (如图 4 所示). 具体地, 根据公式 (4) 定义最优类间可分角 ρ_{ij} :

$$\rho_{ij} \triangleq \frac{C}{C-1} \delta_{ij} - \frac{1}{C-1} \quad (11)$$

其中, δ_{ij} 表示 Kronecker 符号. 令标准化后的类均值为 $\{\tilde{\boldsymbol{\mu}}_i\}_{i=1}^C$, 然后定义损失 $\mathcal{L}(\tilde{\boldsymbol{\mu}})$ 将其与 ETF 结构对齐:

$$\mathcal{L}(\tilde{\boldsymbol{\mu}}) = \frac{1}{C^2} \sum_{i,j=1}^C (\tilde{\boldsymbol{\mu}}_i^T \tilde{\boldsymbol{\mu}}_j - \rho_{ij})^2 \quad (12)$$

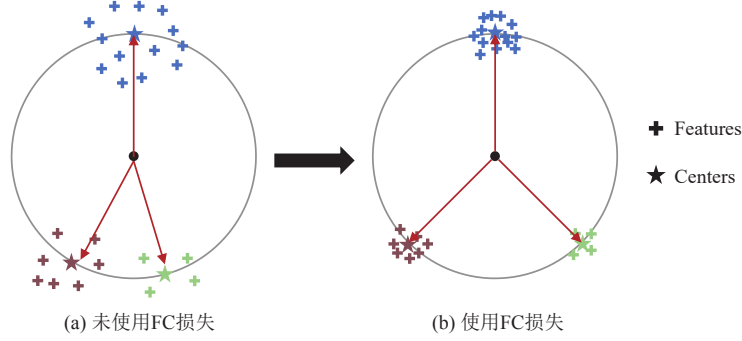


图 4 特征坍塌损失示意图

当损失 $\mathcal{L}(\tilde{\boldsymbol{\mu}}) = 0$ 时, 可得 $\tilde{\boldsymbol{\mu}}_i^T \tilde{\boldsymbol{\mu}}_j = \frac{C}{C-1} \delta_{ij} - \frac{1}{C-1}$, 实现类原型的最优可分性. 进一步, 为促使特征收敛至对应的类原型, 定义损失 $\mathcal{L}(\mathbf{z})$:

$$\mathcal{L}(\mathbf{z}) = -\cos(\mathbf{z}, \boldsymbol{\mu}) \quad (13)$$

当损失 $\mathcal{L}(\mathbf{z})$ 最小化时, 特征 $\mathbf{z}$ 与对应的类原型 $\boldsymbol{\mu}$ 的夹角为 0, 即特征坍塌至类均值对应的类内方差趋于 0, 实现最优的类内紧性. 因此, 特征坍塌损失定义为:

$$\mathcal{L}_t = \mathcal{L}(\tilde{\boldsymbol{\mu}}) + \mathcal{L}(\mathbf{z}) \quad (14)$$

最终, 定义交叉熵损失为 $\mathcal{L}_{\text{CE}}$, 那么本文的总损失可以表示为:

$$\mathcal{L}_{\text{all}} = \mathcal{L}_{\text{CE}} + \eta \mathcal{L}_t \quad (15)$$

其中, η 为可调节参数, 用于控制 FC 损失的调节强度.

4 实验分析

4.1 数据集和实验设置

(1) 数据集介绍. 本文主要使用 CIFAR10-LT、CIFAR100-LT、ImageNet-LT 和 iNaturalist2018 数据集进行实验分析.

● CIFAR10/100-LT 数据集^[6]: CIFAR10/100 包含 60 000 张图片, 其中 50 000 张用于训练, 10 000 张用于测试, 分别有 10/100 个类别. 为进行公平比较, 本文实验采用长尾版 CIFAR 数据集, 其设置与 Zada 等人^[24]的设置相同. 它通过对训练集中每类图像的数量进行指数衰减采样来建立长尾版本 CIFAR10/100-LT, 而其对应的测试集保持不变. 相应地, 实验选择的不平衡率分别为 $\beta=10$, $\beta=50$ 和 $\beta=100$.

● ImageNet-LT 数据集^[40]: ImageNet-LT 是原始 ImageNet-2012 数据集的长尾版本, 通过按照帕累托分布对幂值 $\alpha=6$ 的子集进行采样. 通常, 它有来自 1 000 个类别的 115.8k 张图片, 每个类别最多 1 280 张, 每个类别最少 5 张.

● iNaturalist2018 数据集^[41]: iNaturalist 2018 是一个从自然学家观察生物多样性和自然所拍摄收集的大规模分类数据集. 该数据集本质上是不平衡的, 并遵循长尾分布, 包括 8 142 个类别和 437 513 张图片. 通常, 数据集中使用的不平衡比率最高可达 500. 实验所使用的数据集的相关指标统计汇总在表 1.

表 1 实验所用数据集的相关统计信息

数据集	训练样本	测试样本	类别	最大类尺寸	最小类尺寸	不平衡因子 β	适用任务
CIFAR10	50 000	10 000	10	5 000	50–500	10–100	分类
CIFAR100	50 000	10 000	100	500	5–50	10–100	分类
ImageNet-LT	115 846	50 000	1 000	1 280	5	256	分类
iNaturalist2018	437 513	24 426	8 142	1 000	2	500	分类

(2) 实验设置. 在无特别说明的情况下, 本节中的所有实验都使用 DAR-BN 替换网络中原始的 BN 层, 以缓解域漂移问题. 所有模型都使用 SGD 优化器和交叉熵损失进行训练, 其中动量设置为 0.9 和权重衰减为 $2\text{E}-4$. 所有实验使用 PyTorch 1.8.0 在 Ubuntu 20.04 操作系统上使用 NVIDIA 3090 GPU 进行实验. 实验中的所有模型都是从头开始训练的. 在所有实验中, 损失 L_{all} 中的超参数 η 固定为 8.

对于 CIFAR10/100-LT 数据集, 遵循 Zada 等人^[24]和 Zhang 等人^[26]的实验设置, 使用 WideResNet-28-10 作为网络骨干. 本实验设置 mini-batch 尺寸为 128, 使用初始值为 0.1 的逐步衰减学习率, 然后在 epoch 为 160 和 180 时降低为原来的 0.01. 在前 5 个 epoch 选择了一种线性预热学习率技术, 以与一些基线方法兼容, 即使它对所提方法影响很小. 在最后 40 个 epoch 使用 ASFC.

对于 ImageNet-LT 和 iNaturalist2018 数据集, 实验使用 ResNet-50 作为骨干网和批尺寸设置为 256, 动量为 0.9 和权重衰减为 $5\text{E}-4$ 的 SGD 优化器训练模型, 初始学习率为 $5\text{E}-2$ 的逐步衰减学习率, 然后分别在 epoch 为 160 和 170 衰减 10 倍. ASFC 被推迟到最后 40 个 epoch.

4.2 模型性能比较

(1) 对比方法. 本文方法将与以下方法进行比较.

● 基线方法: 经验风险最小化 (ERM): 在不使用任何重平衡策略的情况下利用交叉熵损失进行训练; 过采样: 通过对少数类进行过采样重平衡数据分布. 这些基线方法与本文算法共享相同的训练参数和网络架构.

● 最近的方法. 比较两种经典的两阶段方法: LDAM-DRW^[34,42]、MisLAS^[34,42]和两种基于 Mixup 的方法: cRT+Mixup^[42]、LWS+Mixup^[42]. 此外, 该实验还与 4 个最近的基于数据增广的方法: M2m^[43]、OPeN^[24]、CSA^[25]、FFN^[44]和 CaUIRL^[26]进行比较.

(2) 在 CIFAR10/100-LT 数据集上的实验结果比较. 根据表 2 的实验结果可以发现, 本文提出的 ASFC 方法在不同数据集及不同不平衡率下均表现出优于其他方法的性能. 具体而言, 在 CIFAR10/100-LT 数据集上, 当不平衡率分别为 10、50 和 100 时, ASFC 相较于性能次优的方法分别提升了 0.4% (3.7%)、0.6% (3.9%) 和 0.9% (3.9%). 进一步分析表明, ASFC 在更困难的场景 (即不平衡率更高) 具有更显著的优势. 这得益于 ASFC 从不同样本和特征两个维度缓解不平衡问题, 从而增强了模型的泛化能力.

(3) 在大规模数据集 ImageNet-LT 和 iNaturalist2018 上的实验结果比较. 为全面评估所提方法的有效性与泛化能力, 本文在更具挑战性的大规模数据集 ImageNet-LT 和 iNaturalist2018 上执行了对比实验, 结果如表 2 所示. 可以观察到, ASFC 在所有对比方法中仍保持最优性能, 尤其在真实采样的 iNaturalist2018 数据集上, 其类别分布更

接近自然长尾特性. 该结果进一步说明, 本文所提出的样本-特征协同机制在不同规模与不同分布特性的数据集上均具备良好的适应能力, 能够有效缓解真实场景中因数据分布不平衡所带来的模型偏差问题.

表 2 各个数据集上 Top-1 准确率的比较结果 (%)

方法	CIFAR10-LT			CIFAR100-LT			ImageNet-LT	iNaturalist2018
	$\beta = 10$	$\beta = 50$	$\beta = 100$	$\beta = 10$	$\beta = 50$	$\beta = 100$		
ERM	—	83.6	79.6	—	52.4	45.1	44.6	61.2
过采样	—	82.2	75.1	—	48.0	42.5	45.1	64.7
LDAM-DRW	—	85.3	80.5	—	52.6	46.8	48.8	68.0
M2m	—	85.5	81.3	—	52.9	46.5	—	—
MiSLAS	91.3	86.9	82.6	67.9	55.8	52.1	—	—
cRT+Mixup	90.1	86.3	81.1	63.9	52.4	48.2	51.7	70.2
LWS+Mixup	91.2	85.7	80.0	63.7	53.0	46.9	52.0	70.9
OPeN	—	87.9	84.6	—	56.3	51.5	52.4	—
CMO	—	—	—	—	—	—	49.1	68.9
CSA	90.3	85.4	82.7	68.3	53.2	46.1	49.7	—
FFN	89.4	85.8	83.1	57.6	49.3	46.8	50.5	—
CaUIRL	92.5	89.1	86.5	72.1	57.4	52.6	53.9	71.0
ASFC	92.9	89.7	87.6	75.8	61.3	56.5	54.1	71.3

4.3 消融研究

(1) 不同组件的影响

本节在 CIFAR100-LT ($\beta = 100$) 数据集上对 ASFC 中的关键组件进行了消融研究, 其中“w/o AmpAug&FC”表示本文使用相同的设置复现的 OPeN 算法. 特别地, 本节进一步统计了在 3 种不同类别尺寸上的准确率, 包括头部类 (类内训练样本 >100)、中部类 (类内训练样本在 20–100 之间) 和 Few-shot 类 (类内训练样本 <20), 以更全面评估所提出的模型. 根据表 3 所示的实验结果可以发现, 本文提出的所有组件均能显著提高模型性能. 此外, 当使用“AmpAug”和“FC”时, 尾部类的分类准确率分别提升了 2.6% 和 4.1%, 这表明它们可以有效缓解尾部类被头部类挤压问题, 能大幅度改善模型偏好.

表 3 在 CIFAR100-LT 数据集上的消融结果 (%)

组件	头部类	中部类	尾部类	总体
ASFC	76.9	58.4	34.2	56.5
w/o AmpAug	77.4	57.6	32.1	55.7
w/o FC	76.2	57.9	33.6	55.9
w/o AmpAug&FC	78.1	52.0	29.5	53.2

(2) 特征可视化结果

为更直观地验证 ASFC 学习所得特征的质量, 本节基于 t-SNE (t-distributed stochastic neighbor embedding) 对 CIFAR10-LT 数据集在 $\beta = 100$ 的不平衡率设定下的特征分布进行了可视化分析. 如图 5(a) 所示, Oversampling 方法因多样性缺乏, 导致尾部类特征分布较为松散, 进而影响模型在尾部类别上的判别性能. 相比于过采样方法, 尽管 OPeN (图 5(b)) 和 CaUIRL (图 5(c)) 方法能在一定程度上提升尾部类特征的聚合程度, 但仍存在部分样本与头部类特征混淆. 而根据本文提出的 ASFC 方法 (图 5(d)) 对应的可视化结果可见, 头部类和尾部类均呈现出更紧凑的类内分布和更清晰的类间间隔. 这说明 ASFC 能有效促进尾部类特征的学习, 进而推动特征分布向理想的最优分类结构对齐.

(3) 分类性能的可视化结果

为深入验证 ASFC 方法在缓解模型头部偏好、提升尾部类别识别性能方面的作用, 本节在 CIFAR100-LT 数据集上对各单一类别的分类准确率进行了可视化分析. 图 6 结果显示, ASFC 方法在绝大多数类别上均能提高分

类准确率. 特别值得指出的是, 随着类别样本数量的减少, ASFC 带来的性能提升幅度更为显著, 表明该方法对样本稀缺的尾部类别具有突出的增强效果. 这一现象说明, ASFC 能够有效改善模型在长尾分布下的判别偏差, 增强对尾部类别的识别能力, 从而在类别层面上进一步证实其缓解头部偏好问题的有效性.

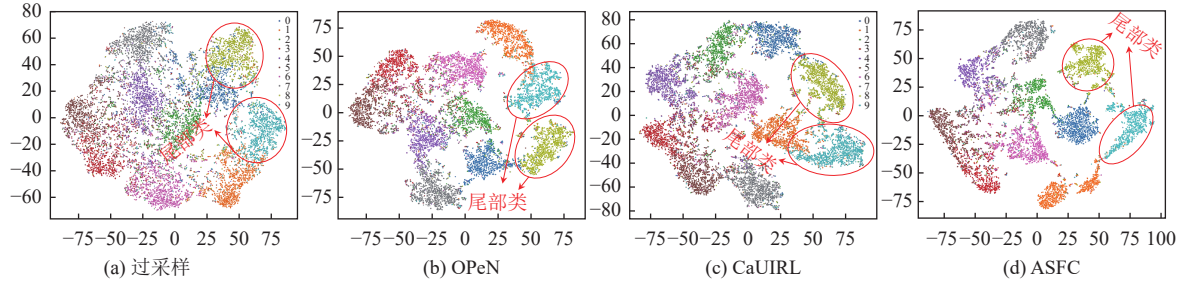


图5 在 CIFAR10-LT 数据集上, 使用 t-SNE 获得的不同算法在测试集上特征的可视化结果

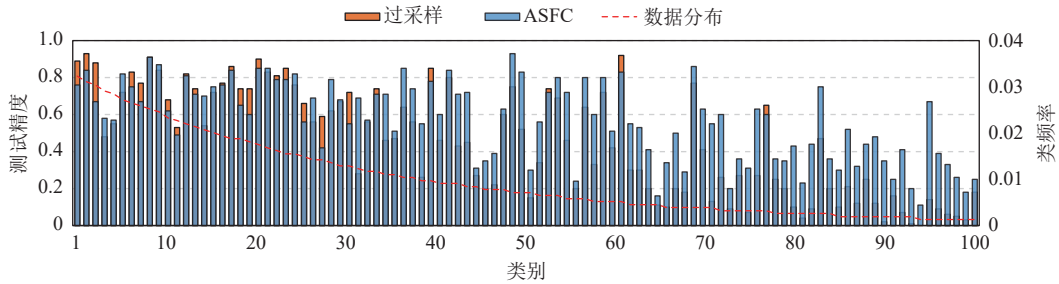


图6 在 CIFAR100-LT 数据集上, 不同方法对应的每个类别的 Top-1 分类准确率

(4) 幅值混合参数 α 对模型性能的影响

在公式 (9) 中, 可调节参数 α 用于控制幅值增广的强度: α 值越大, 表示所选取的语义背景与原始样本越相近, 增广强度越低; 反之, α 值越小, 则增广强度越高. 为了探究幅值增广强度对模型性能的影响, 本节通过调节 α 来探究模型在 CIFAR10/100-LT 数据集上的性能变化. 根据图 7 所示的性能变化曲线可以看出, 当 α 值较小时, 模型性能相对较低, 这是由于所选择的幅值信息比较相近, 导致增广样本的多样性不足, 难以有效提升模型泛化能力. 随着 α 值的逐渐增大, 模型性能稳步上升. 然而, 当 α 取值接近于 1 时, 模型性能呈下降趋势. 这归因于 α 值过大时, 所选择的幅值信息与原始样本差异过大, 导致增广生成的图片可能存在域漂移风险, 从而扰模型学习. 基于上述分析, 在无特殊说明的情况下, 本文所有实验均设定 $\alpha=0.7$, 以在多样性与语义一致性之间取得平衡.

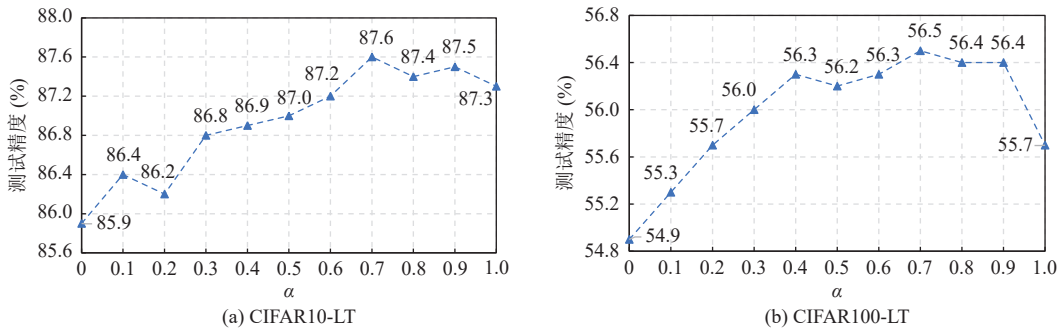


图7 在 CIFAR10/100-LT 数据集上, 不同超参数 α 对模型性能的影响

(5) 增广样本可视化结果

为了更直观地展示使用 FFT 的有效性, 本文对生成的增广样本进行可视化. 如图 8 所示, 图 8(a) 上表示目标

样本 (提供相位信息), 图 8(a) 下表示候选样本 (提供幅值信息); 而对于下面一行情况相似, 目标样本是图 8(a) 下, 候选样本是图 8(a) 上. γ 表示目标样本和候选样本的幅值混合比例, 其值越大表示候选样本的幅值信息占比越大. 不难发现, 随着 γ 的增大, 目标样本的背景越来越接近候选样本, 大幅增加了目标样本多样性. 这进一步说明了 FFT 对于少数类样本扩增的有效性.

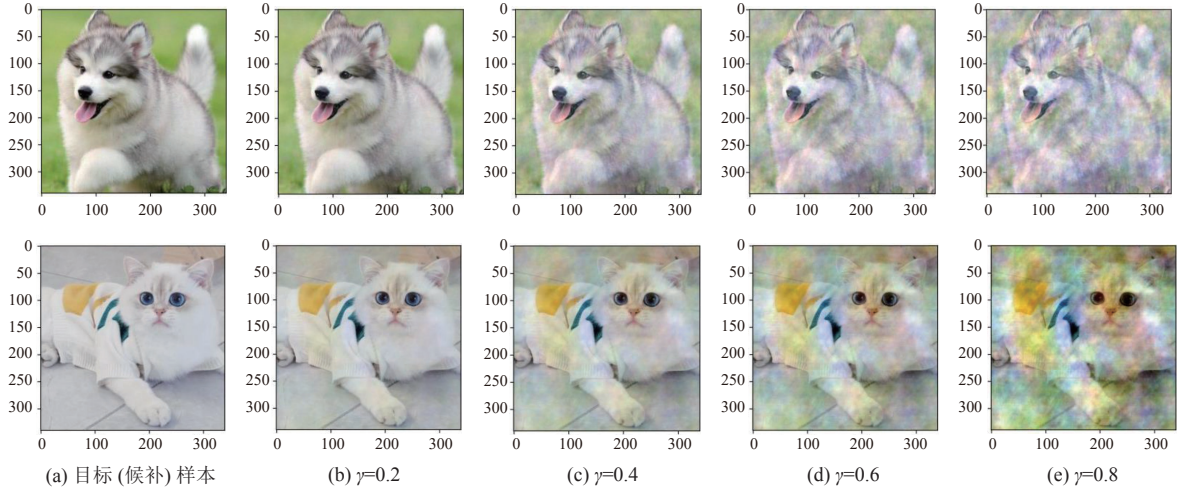


图 8 使用 FFT 得到的扩增样本可视化结果

(6) 模型运行效率探究

为了进一步验证所提模型 ASFC 的计算效率, 本文在大规模数据集 ImageNet-LT 和 iNaturalist2018 上记录在第 2 阶段执行一个批次 (batchsize=64) 不同方法所需的时间 (s). 表 4 统计了基于数据扩增的 CaUIRL 方法和本文方法所需的时间和对应的分类精度, 其中 CaUIRL 使用 Universum 数据扩增尾部类. 不难发现, 相比于 CaUIRL 方法, ASFC 的训练时间较少, 且能显著提高模型的识别性能.

表 4 不同方法运行一个批次的时间和 Top-1 分类精度

方法	ImageNet-LT		iNaturalist2018	
	时间 (s)	精度 (%)	时间 (s)	精度 (%)
CaUIRL	0.114	53.9	0.098	71.0
ASFC	0.109	54.1	0.093	71.3

5 总 结

本文从样本和特征两个层面构建分类模型, 提出 ASFC 方法以缓解长尾识别中的模型偏差问题. 具体而言, 在样本层面, 基于傅里叶变化提出一种“幅值增广”策略, 其在丰富尾部类样本多样性的同时有效避免语义漂移风险; 在特征层面, 受神经坍塌理论启发, 本文设计一种“特征坍塌”损失, 旨在促进特征向量和类原型对齐至理想的最大可分结构. 本文在 4 个基准数据集的不同不平衡率上进行了广泛的实验, 包括: CIFAR10/100-LT、ImageNet-LT 和 iNaturalist2018 数据集. 实验结果表明, ASFC 能显著提高长尾识别性能, 并有效缓解模型对头部类的偏好.

在限制与可拓展性方面, 尽管本文所提出的方法能显著提高模型性能, 但与基于数据扩增的方法类似, 其在大规模数据环境下的训练效率仍有提升空间. 在应用层面, 本文目前主要聚焦于监督长尾识别任务. 未来, 我们将致力于将方法拓展至更具挑战性的长尾半监督学习与开集长尾识别场景, 以验证其更广泛的适用性与鲁棒性. 此外, 本文所提出的“特征坍塌”损失自身具有较好的可迁移性, 可进一步应用于多模态学习和目标检测等方向. 例如, 在多模态学习中, 该损失可推动不同模态下同一语义的特征在隐空间中对齐与聚合, 从而增强跨模态理解的协调性与一致性, 具有重要的延伸研究价值.

References

- [1] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90)]
- [2] Guo JS, Gao X, Yan YX, Li GH, Pu J. Dark-ISP: Enhancing RAW image processing for low-light object detection. In: Proc. of the 2025 IEEE/CVF Int'l Conf. on Computer Vision. Honolulu: IEEE, 2025. 9583–9593. [doi: [10.1109/ICCV51701.2025.00894](https://doi.org/10.1109/ICCV51701.2025.00894)]
- [3] Hu ZX, Li DX, Ma XZ, Tang SX, Li XT, Yang WH, Duan LY. Adaptive dual uncertainty optimization: Boosting monocular 3D object detection under test-time shifts. In: Proc. of the 2025 IEEE/CVF Int'l Conf. on Computer Vision. Xi'an: IEEE, 2025. 7273–7283.
- [4] Li JY, Jiang JJ, Liang PW, Ma JY, Nie LQ. MaeFuse: Transferring omni features with pretrained masked autoencoders for infrared and visible image fusion via guided training. IEEE Trans. on Image Processing, 2025, 34: 1340–1353. [doi: [10.1109/TIP.2025.3541562](https://doi.org/10.1109/TIP.2025.3541562)]
- [5] Zhang P, Yang SR, Sun JS, Li WQ, Su ZY. Open-world point cloud semantic segmentation: A human-in-the-loop framework. IEEE Trans. on Circuits and Systems for Video Technology, 2025, 35(1): 832–845. [doi: [10.1109/TCSVT.2025.3596238](https://doi.org/10.1109/TCSVT.2025.3596238)]
- [6] Krizhevsky A. Learning multiple layers of features from tiny images. 2009. <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>
- [7] Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. ImageNet: A large-scale hierarchical image database. In: Proc. of the 2009 IEEE Conf. on Computer Vision and Pattern Recognition. Miami: IEEE, 2009. 248–255. [doi: [10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848)]
- [8] Zhang EH, Geng CX, Li CH, Chen SC. Dynamic learnable logit adjustment for long-tailed visual recognition. IEEE Trans. on Circuits and Systems for Video Technology, 2024, 34(9): 7986–7997. [doi: [10.1109/TCSVT.2024.3383962](https://doi.org/10.1109/TCSVT.2024.3383962)]
- [9] Du CQ, Wang YL, Song SJ, Huang G. Probabilistic contrastive learning for long-tailed visual recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2024, 46(9): 5890–5904. [doi: [10.1109/TPAMI.2024.3369102](https://doi.org/10.1109/TPAMI.2024.3369102)]
- [10] Wang YH, Wu GS, Wang LM. Long-tailed temporal action detection based on semi-supervised learning. Ruan Jian Xue Bao/Journal of Software, 2025, 36(2): 625–643 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/7154.htm> [doi: [10.13328/j.cnki.jos.007154](https://doi.org/10.13328/j.cnki.jos.007154)]
- [11] Xuan SY, Zhang SL. Decoupled contrastive learning for long-tailed recognition. In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Vancouver: AAAI Press, 2024. 6396–6403. [doi: [10.1609/aaai.v38i6.28459](https://doi.org/10.1609/aaai.v38i6.28459)]
- [12] Zhao QH, Dai YL, Lin S, Hu W, Zhang F, Liu J. LTRL: Boosting long-tail recognition via reflective learning. In: Proc. of the 18th European Conf. on Computer Vision. Milan: Springer, 2024. 1–18. [doi: [10.1007/978-3-031-72855-6_1](https://doi.org/10.1007/978-3-031-72855-6_1)]
- [13] Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. Mixup: Beyond empirical risk minimization. In: Proc. of the 2018 Int'l Conf. on Learning Representations. Vancouver, 2018.
- [14] Xu ZZ, Chai ZH, Yuan C. Towards calibrated model for long-tailed visual recognition from prior perspective. In: Proc. of the 35th Int'l Conf. on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2021. 546.
- [15] Park S, Hong Y, Heo B, Yun S, Chui JY. The majority can help the minority: Context-rich minority oversampling for long-tailed classification. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 6877–6886. [doi: [10.1109/CVPR52688.2022.00676](https://doi.org/10.1109/CVPR52688.2022.00676)]
- [16] Gao JT, Zhao H, Li Z, Guo DD. Enhancing minority classes by mixing: An adaptative optimal transport approach for long-tailed classification. In: Proc. of the 2023 Annual Conf. on Neural Information Processing Systems. New Orleans, 2023. 60329–60348.
- [17] Nussbaumer HJ. The fast Fourier transform. In: Nussbaumer HJ, ed. Fast Fourier Transform and Convolution Algorithms. Berlin: Springer, 1981. 80–111. [doi: [10.1007/978-3-662-00551-4_4](https://doi.org/10.1007/978-3-662-00551-4_4)]
- [18] Papayan V, Han XY, Donoho DL. Prevalence of neural collapse during the terminal phase of deep learning training. Proc. of the National Academy of Sciences of the United States of America, 2020, 117(40): 24652–24663. [doi: [10.1073/pnas.2015509117](https://doi.org/10.1073/pnas.2015509117)]
- [19] Zhang EH, Li CH, Geng CX, Chen SC. All-around neural collapse for imbalanced classification. IEEE Trans. on Knowledge and Data Engineering, 2025, 37(8): 4460–4470. [doi: [10.1109/TKDE.2025.3569563](https://doi.org/10.1109/TKDE.2025.3569563)]
- [20] Li MK, Liu Y, Lu Y, Zhang YQ, Cheung YM, Huang H. Improving visual prompt tuning by Gaussian neighborhood minimization for long-tailed visual recognition. In: Proc. of the 38th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2024. 3304.
- [21] Shi JX, Wei T, Li YF. LIFT+: Lightweight fine-tuning for long-tail learning. arXiv:2504.13282, 2025.
- [22] Zhou ZH, Fang SY, Zhou ZJ, Wei T, Wan YY, Zhang ML. Continuous contrastive learning for long-tailed semi-supervised recognition. In: Proc. of the 38th Int'l Conf. on Neural Information Processing System. Vancouver: Curran Associates Inc., 2024. 1627.
- [23] Li MK, Hu ZK, Lu Y, Lan WC, Cheung YM, Huang H. Feature fusion from head to tail for long-tailed visual recognition. In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Vancouver: AAAI Press, 2024. 13581–13589. [doi: [10.1609/aaai.v38i12.29262](https://doi.org/10.1609/aaai.v38i12.29262)]
- [24] Zada S, Benou I, Irani M. Pure noise to the rescue of insufficient data: Improving imbalanced classification by training on random noise

- images. In: Proc. of the 39th Int'l Conf. on Machine Learning. Baltimore: PMLR, 2022. 25817–25833.
- [25] Shi JX, Wei T, Xiang YK, Li YF. How re-sampling helps for long-tail learning? In: Proc. of the 37th Int'l Conf. on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2023. 75669–75687.
- [26] Zhang EH, Geng CX, Chen SC. Class-aware Universum inspired re-balance learning for long-tailed recognition. Pattern Recognition, 2025, 161: 111337. [doi: [10.1016/j.patcog.2024.111337](https://doi.org/10.1016/j.patcog.2024.111337)]
- [27] Almeida LB. The fractional Fourier transform and time-frequency representations. IEEE Trans. on Signal Processing, 1994, 42(11): 3084–3091. [doi: [10.1109/78.330368](https://doi.org/10.1109/78.330368)]
- [28] Chi L, Jiang BR, Mu YD. Fast Fourier convolution. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 376.
- [29] Zeng RJ, Han C, Wang QF, Wu CS, Geng T, Huang LF, Wu YN, Liu DF. Visual Fourier prompt tuning. In: Proc. of the 38th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2024. 180.
- [30] Buchholz TO, Jug F. Fourier image Transformer. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 1846–1854.
- [31] Nguyen TM, Pham M, Nguyen T, Nguyen K, Osher SJ, Ho N. FourierFormer: Transformer meets generalized Fourier integral theorem. In: Proc. of the 36th Int'l Conf. on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022. 2126.
- [32] Yang YC, Soatto S. FDA: Fourier domain adaptation for semantic segmentation. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 4085–4095. [doi: [10.1109/CVPR42600.2020.00414](https://doi.org/10.1109/CVPR42600.2020.00414)]
- [33] Xu QW, Zhang RP, Zhang Y, Wang YF, Tian Q. A Fourier-based framework for domain generalization. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 14378–14387. [doi: [10.1109/CVPR46437.2021.01415](https://doi.org/10.1109/CVPR46437.2021.01415)]
- [34] Markou E, Ajanthan T, Gould S. Guiding neural collapse: Optimising towards the nearest simplex equiangular tight frame. In: Proc. of the 38th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2024. 35544–35573.
- [35] Tirer T, Bruna J. Extended unconstrained features model for exploring deep neural collapse. In: Proc. of the 39th Int'l Conf. on Machine Learning. Baltimore: PMLR, 2022. 21478–21505.
- [36] Fang C, He HF, Long Q, Su WJ. Exploring deep neural networks via layer-peeled model: Minority collapse in imbalanced training. Proc. of the National Academy of Sciences of the United States of America, 2021, 118(43): e2103091118. [doi: [10.1073/pnas.2103091118](https://doi.org/10.1073/pnas.2103091118)]
- [37] Seo M, Koh H, Jeung W, Li M, Kim S, Lee H, Cho S, Choi S, Kim H, Choi J. Learning equi-angular representations for online continual learning. In: Proc. of the 2024 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 23933–23942. [doi: [10.1109/CVPR52733.2024.02259](https://doi.org/10.1109/CVPR52733.2024.02259)]
- [38] Saadabadi MSE, Dabouei A, Malakshan SR, Nasrabadi NM. Hyperspherical classification with dynamic label-to-prototype assignment. In: Proc. of the 2024 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 17333–17342. [doi: [10.1109/CVPR52733.2024.01641](https://doi.org/10.1109/CVPR52733.2024.01641)]
- [39] Yang YB, Chen SX, Li XT, Xie L, Lin ZC, Tao DC. Inducing neural collapse in imbalanced learning: Do we really need a learnable classifier at the end of deep neural network? In: Proc. of the 36th Int'l Conf. on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022. 2753.
- [40] Liu ZW, Miao ZQ, Zhan XH, Wang JY, Gong BQ, Yu SX. Large-scale long-tailed recognition in an open world. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 2537–2546. [doi: [10.1109/CVPR.2019.00264](https://doi.org/10.1109/CVPR.2019.00264)]
- [41] van Horn G, Aodha OM, Song Y, Cui Y, Sun C, Shepard A. The iNaturalist species classification and detection dataset. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8769–8778. [doi: [10.1109/CVPR.2018.00914](https://doi.org/10.1109/CVPR.2018.00914)]
- [42] Cao KD, Wei C, Gaidon A, Arechiga N, Ma T. Learning imbalanced datasets with label-distribution-aware margin loss. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 140.
- [43] Kim J, Jeong J, Shin J. M2m: Imbalanced classification via major-to-minor translation. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 13893–13902. [doi: [10.1109/CVPR42600.2020.01391](https://doi.org/10.1109/CVPR42600.2020.01391)]
- [44] Zhou XS, Zhai JH, Cao Y. Feature fusion network for long-tailed visual recognition. Pattern Recognition, 2023, 144: 109827. [doi: [10.1016/j.patcog.2023.109827](https://doi.org/10.1016/j.patcog.2023.109827)]

附中文参考文献

- [10] 王雨虹, 武港山, 王利民. 基于半监督学习的长尾时序动作检测. 软件学报, 2025, 36(2): 625–643. <http://www.jos.org.cn/1000-9825/7154.htm> [doi: [10.13328/j.cnki.jos.007154](https://doi.org/10.13328/j.cnki.jos.007154)]

作者简介

张恩豪, 博士, 讲师, 主要研究领域为计算机视觉, 模式识别.

李超华, 博士, 副研究员, 主要研究领域为机器学习, 模式识别.

王志华, 博士, 教授, CCF 高级会员, 主要研究领域为密码学及应用, 网络安全, 隐私计算.

陈松灿, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为机器学习, 模式识别, 神经计算.