面向长尾分布甲骨文识别的自适应难度感知解耦学习方法

孙君为 关素妍 陈信宇 王 堃 蔡院强*
(北京邮电大学计算机学院(国家示范性软件学院) 北京 100876)

摘 要：针对甲骨文识别场景中普遍存在的类别长尾分布问题，以及现有方法难以兼顾头部类别判别力与尾部稀缺样本识别性能的技术瓶颈，该文提出一种两阶段解耦学习方法。第1阶段结合混合数据增强与标签分布感知间隔损失，学习全局泛化的特征空间并构建鲁棒决策边界；第2阶段冻结骨干网络，提出基于类别识别难度的自适应采样策略，结合类别加权损失优化分类器，通过动态分配采样权重聚焦尾部与困难类别，实现头部与尾部类别识别性能的协同提升。在OBC306公开数据集上的实验结果表明，该文方法总体识别准确率达94.34%，平均识别准确率达89.89%，综合性能优于主流方法，可为低资源古文字识别提供技术支撑。

关键词：甲骨文识别；长尾分布；自适应难度采样；解耦学习

中图分类号：TN911.73; TP391.41; TP391.43; TP181 **文献标识码**：A **文章编号**：1009-5896(2026)12-0001-12

DOI: 10.11999/JEIT260327 **CSTR**: 32379.14.JEIT260327

1 引 言

甲骨文起源于商朝，是迄今发现的中国历史上最早的成熟文字系统。如图1所示，因其多刻写于龟甲与兽骨之上，故得名甲骨文。其内容广泛涵盖国家祭祀等重大事务，亦涉及农事生产、气象记录等日常社会生活场景。对甲骨文的系统研究，是探寻商代社会历史与厘清汉字起源脉络的重要途径。葛亮[1]的研究表明，目前已有约160 000件甲骨拓片被发现。然而，由于人工识别与释读甲骨文不仅效率偏低，且受限于专家资源的稀缺性，亟须发展高效的甲骨文智能识别与释读技术，为研究者提供精准的辅助工具。近年来，得益于计算机视觉领域的飞速发展，基于深度学习与神经网络的智能方法被广泛应用于甲骨文识别与释读领域，为古文字研究者的学术工作提供了重要支撑。

Guo等人[2]率先构建了一个包含261个类别的甲骨文数据集Oracle-20K，该数据集收录了约20 000个由专家手工摹写的甲骨文字符，并提出了一种新的多层级图像表示方法，并与经典卷积神经网络相结合，有效提升了甲骨文自动化识别的效果。为进一步扩大手写甲骨文数据的规模与多样性，Zhang等人[3]构建了涵盖2 583个类别的Oracle-AYNU，其中包含约40 000张手写字符图像。该研究采用卷积神经网络将字符图像映射至欧几里德空间，使得样本间距离能够反映其形态相似性，进而通过最近邻规则实现分类。在拓片字符识别领域，Huang等人[4]发布了迄今规模最大的OBC306。该数据集涵盖306个字形类别，包含约300 000张真实拓片字符图像，并对多种深度卷积神经网络进行了性能评估，为拓片字符识别提供了性能基准。为探索甲骨文字符的演变规律并辅助未知字符的释读，Guan等人[5]构建了6个历史阶段的古汉字演变数据集EVOBC，涵盖13 714个不同字符类别的229 170张图像，为通过字形演变轨迹进行甲骨文识别与甲骨文破译提供了极其丰富的数据支撑。

尽管上述数据集在规模与多样性上取得了显著进展，但甲骨文数字化识别研究仍面临数据集固有的长尾分布的挑战。高频字符因反复使用而遗存丰富，而生僻字、专名及异体字则因甲骨残损、存世稀少与释读困难，样本量严重不足。以OBC306[4]数据集为例，样本共划分为306个类别，其中70个类别的样本占总样本数量的83.82%，类别数仅占22.88%，少数类别占据了数据样本的主体，而多数类别则样本匮乏，严重制约了深度学习模型在跨类别识别中的泛化能力。图2展示了OBC306的训练集与测试集的样本分布，与通用场景下的长尾分布不同[6]，甲骨文数据集的长尾特性不仅存在于训练集，在测试集中同样显著呈现。

为解决数据集不平衡分布导致的尾部类别识别精度低的问题，Li等人[7]将Mix-up技术作为数据增强方式，通过在特征空间中显式增强少数类样本，结合3元组损失约束特征空间，有效缓解数据分布倾斜问题。关联字形变换与纹理迁移生成对抗网络(Associate Glyph-transformation and Texture-transfer GAN, AGTGAN)[8]聚焦数据生成以解决

收稿日期：2026-03-25；改回日期：2026-06-28
*通信作者： 蔡院强， caiyuanqiang@bupt.edu.cn
基金项目：国家自然科学基金(62272058)
Foundation Item: The National Natural Science Foundation of China (62272058)

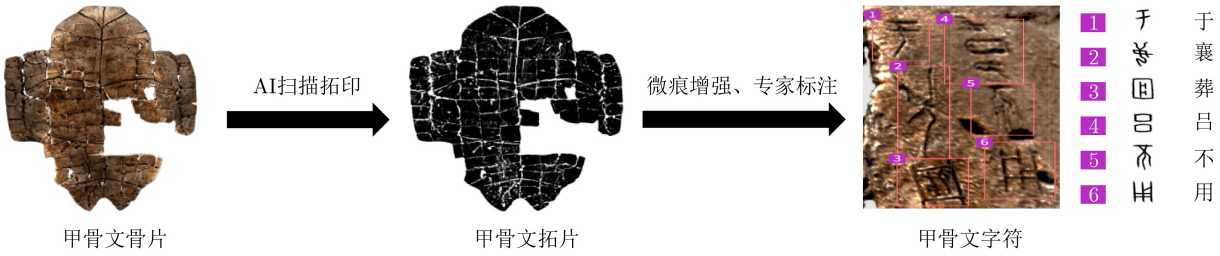
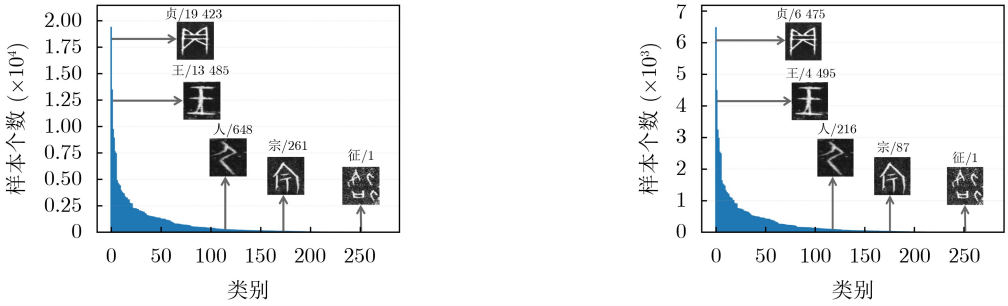


图 1 甲骨文由甲骨文骨片经多种算法技术数字化为图片形式样本



(a) 训练集分布 (b) 测试集分布

图 2 OBC306数据集的真实分布图

数据稀缺问题，该模型借助关联对抗训练机制，协同优化两个子生成对抗网络，生成符合数据集特征分布的甲骨文字符，将OBC306[4]数据集的平均识别准确率从80.16%提升至85.36%。Diff-Oracle[9]突破了单一拓片域的限制，借助于高质量的手写数据辅助新样本的生成，通过配对的内容图像与风格图像，生成过程更为稳定，在识别准确率上取得最优性能。除上述数据增强与扩充策略外，Li等人[10]使用了解耦结构和多阶段学习的方法，通过先训练骨干网络获取鲁棒特征、再优化分类器的策略，结合可学习权重缩放与知识蒸馏，得到泛化识别性能更好的无偏模型。

在上述方法中，由于甲骨文字本身具有的类内变种特性，传统数据增强技术虽能增强多数类别识别的鲁棒性，但可能生成并不具有甲骨文字语义的样本。生成式模型[8]通过学习样本特征生成补充数据，但拓片样本常因历史久远，不可避免地带有不同程度的噪声干扰，要生成语义有效、风格契合的拓片字符仍面临挑战。Li等人[10]引入知识蒸馏过程与KL散度损失，维持对头部类关注，却削弱了分类器对尾部类的学习优先级，制约了模型对长尾分布数据的适配能力。此外，目前的甲骨文长尾识别方法多以优化少数类识别为核心，但部分类别存在丰富的类内变种。这些类别既需学习更具代表性的判别特征，还需捕捉自身可变部分的特征，导致特征学习难度显著提升。本文的方法旨在平衡少数类与困难类的学习权重，通过自适应难度采样器，使困难类与少数类得到更充分的训练。同时，借助高效数据增强技术生成甲骨文样本，促使模型精准捕获类别特异性判别特征，实现模型鲁棒性表征学习。

综上，本文的主要贡献如下：(1)提出一种高效的数据增强策略，融合CutMix[11]与RandAug(随机增强)[12]方法，使模型学习更具判别力的类别特征，并显著提升其在轻微扰动下的识别鲁棒性。(2)设计一种基于类别难度的采样机制，动态平衡类别样本数量与学习复杂度，适度降低简单多样本类别的训练优先级，将模型学习重心导向少样本困难类别，在全局层面更充分地挖掘甲骨文的本质特征。(3)构建两阶段解耦学习框架，并设计类别加权标签分布感知边缘损失(Class-Balanced Label-Distribution-Aware Margin Loss, CBL)，根据各类别样本数量分配损失权重与类别间隔，引导模型学习蕴含更丰富语义信息的类别表征，强化类别判别特征的区分度与有效性。

2 相关工作

2.1 通用甲骨文识别

甲骨文识别是甲骨文研究的核心环节之一。传统依赖专家人工识别的方式耗时费力且成本高昂，随着计算机视觉技术发展，甲骨文识别从专家人工识别向自动化识别转型[13]，相关技术在该领域广泛应用。早期两级分类[14]将每个甲骨文视为一个无向图，提取其拓扑属性进行初步识别。图同构识别[15]将字符转换为无向标记图，并用准邻接矩阵编码以比较其结构。

深度学习为甲骨文识别带来突破性进展，其端到端特征学习能力有效弥补了传统方法的短板，大幅提升了识别性能。Guo等人[2]采用卷积神经网络提取特征，并结合其提出的特征表征方案，实现了优异的甲骨文识别性能。针对甲骨文数据集匮乏的问题，Zhang等人[3]于2019年构建了Oracle-AYNU数据集，并提出基于深度度量学习的识别方法。同期，Huang等人[4]构建了包含300 000样本、306个类别的拓片字符的OBC306数据集，并在该数据集上验证了Inception-v4[16]等主流神经网络的性能。OBC306数据集存在严重的长尾分布问题，传统深度学习方法未能对头部类别与尾部类别进行差异化处理，导致尾部类别识别性能受限。近年来，研究者通过优化网络架构提升模型对复杂拓片图像的特征提取能力。例如，Mai等人[17]针对甲骨文拓片噪声大、残缺的特点，提出一种融合Inception模块与残差连接的新型卷积神经网络，并行多尺度卷积与数据增强策略以取得优异的识别性能。然而，此类基于架构改进的方法大多在类别数量较少且分布高度均衡的数据集上进行验证的，面对真实甲骨文数据集中的极端长尾分布问题时，网络架构改进由于梯度被多数类主导而失效。近年来，小样本学习与零样本学习在甲骨文识别领域取得一定进展，毕等人[18]基于对比学习提升了极少样本的特征提取能力，刘等人[19]通过引入字模语义实现了对未见字符的跨模态视觉匹配。然而，小样本学习方法[18]主要针对人工构造的均衡任务，若应用于包含数百个类别的全局长尾分类场景中，头部类别将彻底主导模型的梯度更新与特征空间划分，导致尾部特征被严重挤压；而零样本匹配方法[19]则高度依赖高质量的专家字模作为辅助信息，对于缺乏标准参考字模、伴随严重噪声和类内变种的长尾数据集，语义提取与跨域匹配极易失效。这两种范式均未能根本解决由极度不平衡的数据分布引发的分类器决策边界严重偏移的问题。因此一系列面向长尾分布的优化方法被引入甲骨文识别领域，以针对性改善尾部类别识别效果。

2.2 长尾分布场景下的甲骨文识别

甲骨文识别领域存在显著的长尾分布问题[3,4]，Oracle-AYNU[3]与OBC306[4]数据集为该问题的研究提供了重要数据支撑。Li等人[7]提出一种融合Mix-up数据增强与3元组损失函数的识别方法，通过Remix变体在特征空间合成新样本，扩充少数类数据、平衡类别分布，使用3元组损失提升特征表示的类内紧凑性与类间区分度，但该方法合成样本的质量难以评估。针对该局限，生成对抗网络[8,20]实现样本扩充的方法所生成的样本具备直观可评估的视觉质量，精准地补充尾部类数据，缓解甲骨文数据集的长尾分布问题。AGTGAN[8]提出一种无监督生成模型，通过分离学习全局与局部字形特征，引入笔画敏感的纹理迁移方法，并结合关联性对抗学习策略，生成字形多样且纹理逼真的甲骨文字符，其在OBC306[4]数据集上针对少样本与零样本类的实验验证了其对尾部类的适配能力。循环一致对抗网络(Cycle-Consistent Adversarial Network, CycleGAN)[20]学习手写字形域与真实拓片域的映射关系大规模生成补充样本，扩充尾部类样本规模改善类别不平衡问题。Li等人[21]在对抗生成的基础上，设计尾部融合方法，引入其他混淆类的样本以解决尾部类别在对抗生成时由于类别样本不足而产生的过拟合的问题，但该方法可能导致尾部类别与混淆类别之间的决策边界趋于模糊。由于甲骨文字的类内变种多的特性，整体风格差异较大的类内变种样本会在对抗生成过程中被丢弃，导致无法学习到类别内部更鲁棒的判别特征。

3 本文方法

本文提出一种两阶段训练方法，实现特征提取与分类决策的解耦优化，如图3所示。第1阶段采用实例平衡采样，样本经CutMix[11]与RandAug[12]组合增强扩充鲁棒特征后，引入标签分布感知边缘(Label-Distribution-Aware Margin, LDAM)[22]损失对logit实施差异化缩放约束，提升头部类别的决策阈值，降低尾部类别的判别标准，缓解模型的类别偏置问题。针对一阶段采样权重均等、甲骨文字符类内变种多样问题，第2阶段引入类别难度自适应采样策略，量化类别难度系数分配差异化采样权重，提升困难类别的采样优先级。2阶段使用并冻结第1阶段中训练完成的骨干网络，继承其特征判别能力。2阶段训练的核心在于采用难度感知采样，集中优化分类头，以精细化其决策边界。本方法维护头部类别性能的同时，提升了尾部类别与困难类别的分类效果，实现了整体性能的均衡。

3.1 LDAM损失

LDAM损失[22]是为解决长尾识别问题所设计的适应数据集不平衡的损失函数，通过为不同类别的样本设置自适应分类边际，缓解类别不均衡带来的偏置，从而在理论上能够最小化基于边际的泛化误差上界。具体而言，该损失函数对样本量大的头部类别施加更大的分类边际，迫使模型以更高置信度对其进行分类，而对样本量少的尾部类别则设置较小的边际，以降低其分类难度。这种差异化策略的数学表达如式(1)所示

$$L_{\mathrm{LDAM}} = -\ln \frac{\mathrm{e}^{s\cdot(z_y-\Delta_y)}}{\mathrm{e}^{s\cdot(z_y-\Delta_y)}+\sum\limits_{j\neq y}\mathrm{e}^{s\cdot z_j}} \tag{1}$$

其中，s是用于缩放logit值的超参数，z_y和z_j分别表示神经网络模型对输入样本的真实类别y和非真实类别j输出的原始预测logits，Δ_y为真实类别y的自适应分类边际。对于$\Delta_y = C/n_y^{1/4}$，n_y表示类别y的样本个数，C是用于控制边际幅度的超参数。

LDAM损失在特征空间层面为尾部类别设置较小的分类边际，保留了其更大的决策区域，以此抑制头部类别的主导倾向，并增强模型对尾部特征的敏感性。在优化层面，该损失函数通过缩放logit，提高模型正确分类时的置信度，并对分类错误施以严厉惩罚，迫使模型学习更清晰的决策边界。最终确保模型在面对所有类别时，均能做出高置信度的准确预测。

3.2 难度采样

在长尾识别任务中，传统重采样方法通常依据类别样本数量，为头部与尾部类别分别赋予低与高的采样权重。然而，此类方法仅从宏观层面关注类别不均衡，而忽视了样本自身的学习难度所带来的挑战。受Yu等人[23]启发，本文提出基于实例难度的重采样策略，核心思想是依据样本的学习速度动态调整采样概率，从而在实例级别同时关注困难样本与少数类样本，以有效解决头部类中也存在难以学习的样本这一关键问题，提升模型整体识别性能。尽管如此，该策略需在实例级别记录所有样本的难度信息，这将带来显著的性能开销与训练效率瓶颈，使其难以适用于大数据集场景。鉴于此，本文提出一种更高效的采样策略。本文以类别的平均损失作为其整体难度的代理指标，并设计一种融合历史与当前周期难度信息的动态权重分配机制。本策略在保留难度感知核心思想的同时，将计算粒度从实例级提升至类别级，大幅降低了计算开销，使其能够直接应用于大规模长尾数据集。设当前epoch内的所有样本索引集合为S，类别c的本轮类别难度定义为D_c^{current}，如式(2)所示

$$D_c^{\mathrm{current}} = \frac{\sum\limits_{i\in S} L_i \cdot \boldsymbol{1}(y_i = c)}{\sum\limits_{i\in S} \boldsymbol{1}(y_i = c)} \tag{2}$$

其中，L_i为样本i的分类损失，y_i为样本i的标签。由于训练中使用了CutMix混合增强策略，每个样本可能包含两个类别的标签信息，在计算难度时，本文同步统计了这两个类别各自的难度与样本数。式(3)定义了第t个epoch的类别c难度评估更新规则$D_c^{(t)}$。该公式采用指数移动平均的方式融合了历史信息与当前观测值。其中，$\mathrm{clip}\left(D_c^{\mathrm{current}}, D_{\min}, D_{\max}\right)$将难度截断到$[D_{\min}, D_{\max}]$之内，保证数值稳定，避免某个类别权重过大或过小影响训练。

$$D_c^{(t)} = \alpha_c \cdot D_c^{(t-1)} + (1-\alpha_c)\cdot \mathrm{clip}\left(D_c^{\mathrm{current}}, D_{\min}, D_{\max}\right) \tag{3}$$

式(3)中，平滑系数α_c依据类别的样本数量进行自适应调整。参照长尾识别领域的通用标准[6]，将样本数大于100、介于20～100以及小于20的类别分别定义为头部类、中部类和尾部类。鉴于本实验采用3:1的比例划分训练集与测试集，本文对训练阶段的判定阈值进行了等比缩放，在训练集中，样本数大于75的为头部类，15～75的为中部类，小于15的为尾部类。因此，α_c的具体设置如式(4)，n_c表示训练集中类别c的样本个数

$$\alpha_c = \begin{cases} 0.9, & n_c > 75 \\ 0.5, & 15 < n_c \le 75 \\ 0.1, & n_c \le 15 \end{cases} \tag{4}$$

针对不同类别的样本分布特性，本文采用了差异化的平滑策略。在式(4)中，鉴于头部类别样本充足，训练过程相对收敛且稳定，本文将α_c设为0.9，使难度评估更侧重于历史信息的积累，有效平滑了由单个epoch引起的随机波动，确保了采样过程的稳定性。对于中等类别，本文采取折中策略，设置$\alpha_c = 0.5$，兼顾训练过程的数值稳定性与对难度变化的动态响应能力。考虑到尾部类别数据稀缺且往往存在较大的类内差异，导致其训练状态具有高度不稳定性，因此，本文设置较低的平滑系数$\alpha_c = 0.1$，以赋予当前观测值更大的权重，使模型能够快速捕捉并适应尾部类别的难度突变，从而及时优化采样权重，提升长尾识别性能。本文将上述计算出的难度$D_c^{(t)}$，作为类别c在当前轮次t的采样权重W_c。

与基于实例难度的重采样策略[23]不同，本文提出的策略在类别层级进行采样权重分配。具体而言，本文通过统计类别的平均损失来计算难度值，同一类别内的所有样本共享相同的采样权重。相较重采样策略[23]需要维护所有实例的难度信息，本文方法在大规模长尾数据集上显著降低了计算开销，简化设计逻辑的同时实现了效果相当的性能表现。该策略根据实际训练状态动态调整关注点：对于识别困难的尾部类，较高的损失值赋予其更大的采样权重，以缓解数据稀缺带来的欠拟合；而对于特征易于判别的简单尾部类，模型快速收敛至较低损

失，从而分配较小的采样频率，避免无效的过采样。针对头部与中部类别，尽管样本量充足，但甲骨文数据普遍存在严重的类内变异，不同变体风格差异巨大。针对模型尚未充分学习的类别，该策略会自动提升其采样权重，确保模型将训练重心聚焦于难以判别的特征空间，而非单纯依赖样本数量。

3.3 CBL损失

鉴于第2阶段训练中基于难度的采样器与优化目标存在高度耦合，构建一个能够精准表征当前训练难度并有效引导模型优化方向的损失函数至关重要。针对长尾数据集的特点与甲骨字符类内变异显著等问题，本文在LDAM损失的基础上引入类别平衡策略[24]，提出CBL损失。该损失函数融合了边界调整与类别重加权双重机制，有效提升尾部类别的识别效果。Cui等人[24]提出基于有效样本数的类别权重计算方法，有效样本数理论认为类别内的首个样本提供了完整的语义信息，而随着样本数量增加，后续样本的信息含量因语义冗余而逐渐衰减。基于此，对于样本数为n_c的类别c来说，其有效样本数定义为式(5)

$$E_{n_c} = \frac{1-\beta^{n_c}}{1-\beta} \tag{5}$$

其中超参数$\beta \in [0,1)$控制信息重叠程度，当$\beta = 0$时，所有类别权重相等。当β趋近于1时，即假设样本间重叠极少，每个样本都贡献了独立信息，此时权重与类别样本总数成反比。故基于有效样本数的类别权重计算如式(6)所示

$$w_c = \frac{1}{E_{n_c}} = \frac{1-\beta}{1-\beta^{n_c}} \tag{6}$$

本文将该权重机制与LDAM损失集成，定义CBL损失为式(7)。其中w_y为目标类别的平衡权重，Δ_y为目标类别的自适应边际，s为可设置的缩放因子

$$L_{\mathrm{CBL}} = -w_y \cdot \ln \frac{\mathrm{e}^{s\cdot(z_y-\Delta_y)}}{\mathrm{e}^{s\cdot(z_y-\Delta_y)} + \sum_{j\neq y} \mathrm{e}^{s\cdot z_j}} \tag{7}$$

本文所提CBL损失首先利用LDAM损失的边际调整机制实现类别的自适应约束，对头部类别施加较大的分类边际，迫使模型学习更具判别力的鲁棒特征。对尾部类别设定较小的边际，以降低分类难度并扩展其决策边界。在此基础上，引入重加权机制，基于有效样本数重新分配类别权重，显著提升尾部类别在损失函数中的贡献度。配合本文设计的采样策略，本方法不仅确保了尾部类别在特征空间中保留足够的决策区域，更在反向传播中赋予其更强的梯度学习信号，从而实现了尾部类别识别性能的显著提升。

3.4 混合增强策略

针对长尾分布中尾部类别样本稀缺、类内方差不足的问题，本研究引入RandAug[12]策略构建自动数据增强模块，在低计算开销下最大化尾部样本的特征空间。不同于传统固定模式的增强，本文利用随机增强建立了一个包含几何、颜色及图像质量调整的综合变换池。在实现中，该策略通过超参数n控制增强强度，参数m控制变换操作的级联次数，将随机选择的变换序列应用于输入图像。该策略丰富了样本的多样性，并有效地模拟了真实场景下的光照与几何扰动，显著提升了模型在尾部稀缺样本上的鲁棒性，使其能从有限数据中学习到更具泛化能力的特征表示。在提升尾部类别鲁棒性的同时，为缓解头部类别的过拟合现象并防止信息丢失，本文在训练阶段集成CutMix[11]正则化策略，以替代传统的Dropout方法。通过区域替换和软标签机制，模型更加关注图像的全局上下文信息，而非过度依赖头部类别的局部显著特征。具体而言，两个训练样本$(\boldsymbol{x}_{\mathrm{A}}, y_{\mathrm{A}})$和$(\boldsymbol{x}_{\mathrm{B}}, y_{\mathrm{B}})$通过式(8)和式(9)融合生成新的训练样本$(\tilde{\boldsymbol{x}}, \tilde{y})$

$$\tilde{\boldsymbol{x}} = \boldsymbol{M} \odot \boldsymbol{x}_{\mathrm{A}} + (\boldsymbol{1} - \boldsymbol{M}) \odot \boldsymbol{x}_{\mathrm{B}} \tag{8}$$

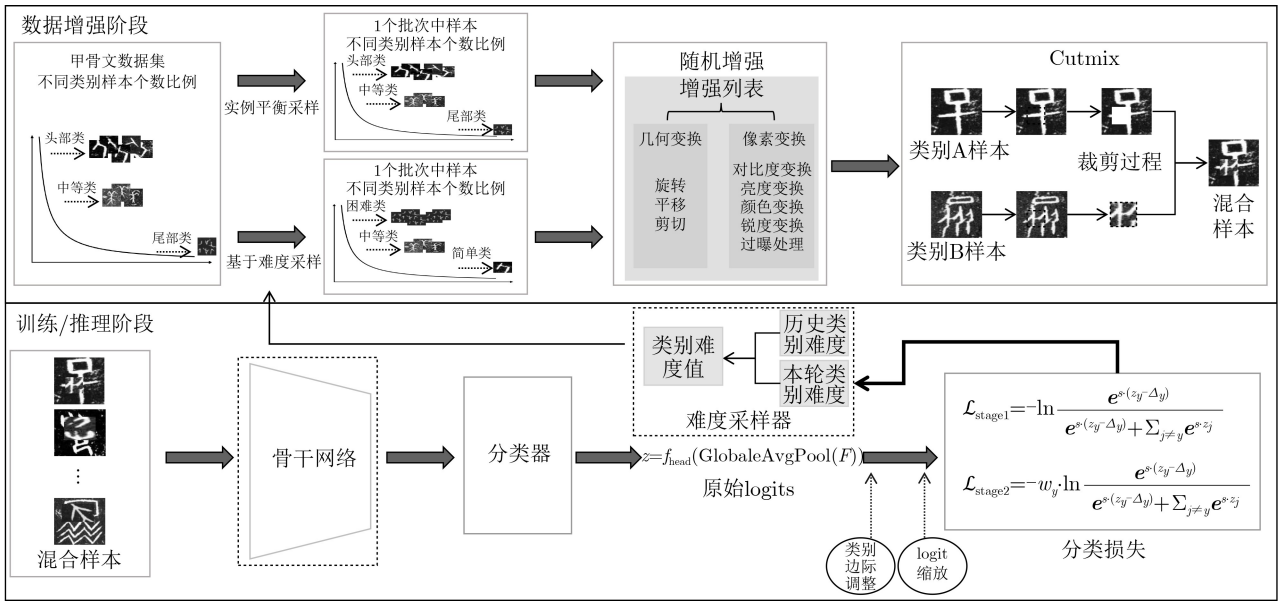
$$\tilde{y} = \lambda y_{\mathrm{A}} + (1-\lambda)\, y_{\mathrm{B}} \tag{9}$$

其中$\boldsymbol{M} \in \{0,1\}^{W\times H}$为二值掩码，指示区域裁剪与填充，$\boldsymbol{1}$是所有元素全为1的二值掩码，$\odot$为全元素乘法符号，$\lambda$表示两样本之间的混合系数，实验中，设定$\lambda$服从$\mathrm{Beta}(\alpha,\alpha)$分布$(\alpha=1)$，即$\lambda \sim \mathrm{Unif}(0,1)$。该策略基于局部区域的替换机制，将另一训练样本的图块嵌入当前图像，生成视觉上更自然、语义上更连贯的混合样本。掩码$\boldsymbol{M}$基于边界框$\boldsymbol{B} = (r_x, r_y, r_w, r_h)$表示裁剪区域，裁剪区域的中心坐标$(r_x, r_y)$服从Unif连续均匀分布，在图像的高$W$和宽$H$范围内进行均匀随机采样，遵循式(10)和式(11)，使裁剪区域比例满足$r_w r_h/(WH) = 1-\lambda$

$$r_x \sim \mathrm{Unif}(0, W), \quad r_w = W\sqrt{1-\lambda} \tag{10}$$

$$r_y \sim \mathrm{Unif}(0, H), \quad r_h = H\sqrt{1-\lambda} \tag{11}$$

上述样本混合的软标签机制防止头部类别在训练过程中主导梯度更新，并为尾部类别生成了大量虚拟样本，有效平滑了类别间的分布差异。该混合增强方法迫使模型从残缺和混合的视觉信息中还原类别特征，显著提升了模型对尾部类别的判别特征提取能力。



注：虚线表示在2阶段训练中，冻结骨干网络参数，只训练分类器，并且只在2阶段训练中启用难度采样器。

图 3 整体方法框架图

4 实验

4.1 数据集

本文在OBC306[4]数据集上进行评估并在Oracle-AYNU[3]数据集上进行验证。OBC306数据集包含306个类别、309 551个样本，具有极端的长尾分布(不平衡率高达25 898:1)。其中29个类别只含有1个样本，无法分割训练集与测试集。因此在实验中去除了这29个类别，最终实验所用的OBC306数据集包含277个类别，309 524个样本。实验中采用3:1的比例划分训练集与测试集。Oracle-AYNU数据集包含2 584个类别，39 072个专家手写样本，实验中采用9:1的比例划分该数据集，划分后的数据集中有662个类别只有1个训练样本。

4.2 实施细节

实验平台为NVIDIA RTX 4090 GPU。本文模型参数设置为：骨干网络选用在ImageNet上预训练的InceptionNeXt-Small[25]；优化器使用AdamW，权重衰减设置为1e-5；第1阶段学习率初始化为0.001，配合余弦退火策略；第2阶段微调分类器时将学习率降至0.000 2。

4.3 评估指标

在图像识别任务中，总识别准确率是衡量模型性能的通用指标[13]，定义为式(12)。然而，在长尾分布场景下，总识别准确率易被占据绝大多数样本的头部类别主导，从而掩盖了模型在尾部类别上的性能缺陷。为了更全面地评估模型在所有类别上的泛化能力，消除类别不平衡带来的评估偏差，本文引入了平均识别准确率[13]，该指标通过计算每个类别的准确率并取其算术平均值，赋予了所有类别相同的权重，平衡类别重要性，定义为式(13)

$$\text{Accuracy}_{\text{total}}=\frac{\sum_{j}^{Y}r_j}{\sum_{j}^{Y}n_j} \tag{12}$$

$$\text{Accuracy}_{\text{avg}}=\frac{1}{Y}\sum_{j=1}^{Y}\frac{r_j}{n_j} \tag{13}$$

其中，Y表示类别总数，n_j表示第j类的样本总数，r_j表示第j类中模型正确分类的样本数。实验中，本文使用平均识别准确率与总识别准确率两项指标综合衡量模型识别性能。

4.4 与其他甲骨文识别方法的对比分析

表1展示了本文所提方法与Inception-v4[16]、Mix-up[7]、对抗性数据增强长尾识别(Adversarial data augmentation Long-Tailed Recognition,

表 1 OBC306数据集上本文方法与其他甲骨文识别方法的对比

方法	识别类数	平均识别准确率(%)	总识别准确率(%)
Inception-v4[4]	306	70.28	×
Mix-up[7]	277	80.16	91.74
AGTGAN[8]	306	85.36	×
CycleGAN[20]	306	83.44	92.02
Decoupled-LTR[10]	277	85.01	92.40
ADA-LTR[21]	277	86.54	93.86
Diff-Oracle[9]	275	88.07	94.12
本文方法	277	**89.89**	**94.34**

ADA-LTR)[21]及Diff-Oracle[9]等主流方法在识别类数、平均识别准确率及总识别准确率上的对比结果。针对识别类数，Inception-v4[16]在构建OBC306数据集时包含全部306个类别，随后的AGTGAN[8]与CycleGAN[20]等生成式方法保留了这一全集设定，并将29个单样本类别直接划入测试集，构成了零样本学习任务。然而，在多数专注于解决长尾分布问题的研究中[7,10,21]，包含0个训练样本的29个类别本质上属于零样本学习范畴，而非纯粹的长尾图像识别任务。为了聚焦于尾部类别的有效训练，通常会剔除这29个单样本类别，仅保留剩余的277类。本文遵循这一实验设置，同样采用277个类别进行评估。而Diff-Oracle[9]受限于跨数据集的字符匹配要求，仅使用了在Oracle-AYNU[3]数据集中能够找到配对字符的275个类别。

本文方法在两项准确率评价指标上均超越了所有对比方法。作为基准的Inception-v4[16]受制于长尾分布的影响，在306个类别上进行评估，平均识别准确率仅为70.28%，表明模型在零样本学习与长尾识别任务上存在性能缺陷。而本文方法将这一指标提升至89.89%，增幅达19.61%，根据识别类数进行性能对齐后，本文方法依旧显著提升识别精度，这一实验结果充分证明本文策略在解决类别不平衡问题上的有效性。其次，相较于Mix-up[7]，CycleGAN[20]和AGTGAN[8]等基于数据增强或生成对抗网络的方法，本文方法同样保持优势。具体而言，虽然CycleGAN[20]和AGTGAN[8]通过生成对抗网络生成样本在一定程度上缓解数据匮乏问题，解决了Mix-up[7]生成样本语义模糊的局限性，将平均识别准确率提升至83%～85%区间，但其仅依靠样本生成的方法可能引入噪声，导致总体准确率受损。值得注意的是，AGTGAN等生成式方法在包含单样本类的306个类别上进行了评估，单样本类别的准确率较低，导致平均识别准确率略低。然而，本文方法与其他同样在277类别上进行评估的方法进行严格对齐的公平比较，本文方法依然取得了提升，这充分证明了本文策略并非仅依赖于类别数的减少，而是稳定的提升了长尾特征的表征能力。ADA-LTR[21]通过生成器与混淆类别样本混合机制的结合，合成虚拟数据补充尾部类别。然而，甲骨文字符拓扑结构复杂、类间差异细微，合成样本可能干扰模型对真实特征的学习。与上述策略相反，本文方法摒弃了不稳定的样本生成策略，采用基于难度的采样策略，通过充分挖掘真实数据中被模型遗漏的困难样本，在不引入合成噪声的前提下，直接增强模型对尾部真实特征的判别能力。相较于解耦长尾识别(Decoupled Long-Tailed Recognition, Decoupled-LTR)[10]方法，本文方法证明了引入难度感知机制的重要性。即使Diff-Oracle[9]借助额外配对信息，本文方法依然在平均识别准确率上取得了1.82%的显著优势，并将总识别准确率提升至94.34%。这表明，运用有效的数据增强方法，并结合合理的采样策略与损失函数加权，仅利用图像自身信息即可构建出比跨域匹配更鲁棒的识别模型。

4.5 消融实验

本文依据消融实验的结果探讨了CutMix[11]，RandAug[12]和本文设计的难度采样器(Difficulty Sampler, DS)的有效性。表2中的第1行作为第1阶段的基线模型。

CutMix方法在不使用采样策略的情况下有效解决了数据集中的极端不平衡问题。具体来说，CutMix使用数据混合，在单个训练步骤中对两个样本进行采样，使模型能够学习两个样本的特征，并更全面地学习样本。这种裁剪混合的增强机制可以防止模型仅根据样本的上下文做出判断，并鼓励模型专注于图像的细节，允许模型即使在数据集中存在严重的噪声时保持更好的判别性能。与基线模型相比，使用CutMix方法后总识别准确率性能从92.33%提高到94.40%，平均识别准确率从78.11%提高到81.55%，这表明了CutMix的有效性。

RandAug[12]方法在综合变换池中随机挑选不同变换方法应用于样本，丰富样本的多样性，提高模型的泛化能力，使模型的分类性能更加稳健。表2第4行显示，总识别准确率从94.40%提高到94.71%，平均识别准确率从81.85%提升至83.87%。

鉴于CutMix与RandAug方法对模型特征提取能力的增强，本文在第2阶段训练中使用以CutMix和RandAug方法训练出的1阶段模型作为2阶段的骨干网络进一步进行微调。

为了更好地解决头部类主导的模型学习问题，在训练的第2阶段设计的难度采样策略通过根据难度分配不同的采样权重来提高尾部类和困难类的识

表 2　基于OBC306数据集的消融研究

CutMix	RandAug	DS	准确率(%)	
			总识别准确率	平均识别准确率
×	×	×	92.23	78.11
√	×	×	94.40	81.85
×	√	×	92.86	79.64
√	√	×	94.71	83.87
×	×	√	93.01	80.57
×	√	√	**94.60**	82.21
√	√	√	94.48	**89.47**

别能力。如表2第7行所示，在使用难度采样策略后，尽管总识别准确率小幅下降，但所提方法的平均精度从83.87%提升至89.47%。由于难度采样策略降低了构成大多数样本的头类的采样权重，为尾类和困难类分配更高权重时，在一定程度上忽略了头类样本，从而导致总识别准确率略有下降。

如表2所示，CutMix和RandAug可以提高模型的整体识别性能，同时有效地平衡头部和尾部类别之间的注意力，显著提高模型的平均准确率。如表2第6行所示，添加RandAug后，平均精度和总识别精度分别上升了1.64%和1.59%，这表明了RandAug在第2阶段中对于模型泛化识别能力的有效性。在表2第7行中，CutMix显著增强了模型对样本的表示学习和泛化性能，实现了更鲁棒的识别性能，并将平均准确率从82.21%显著提高至89.47%。

在Oracle-AYNU上的消融实验效果如表3结果所示，在手写甲骨文识别任务中，常规增强策略存在明显的域适应瓶颈。RandAug所包含的色彩域扰动会向高对比度的二值图像中注入无意义的光度噪声，从而损害准确率。而CutMix方法虽能通过实例拼接增强特征泛化性，但在笔画稀疏的手写图像中，极易产生大面积背景像素的无效替换。这种无效的视觉混合与强制分配的软标签之间产生严重的语义错位，导致模型优化方向偏离，致使其增益效果受限。但难度采样策略通过在2阶段训练中为所有类别动态分配合理的采样权重，将总识别准确率提高至82.37%，平均识别准确率大幅提高至79.75%，证明了该策略的有效性。

4.6 参数敏感性分析

为了深入探究所提难度采样器与损失函数在不同超参数设置下的表现，并验证其鲁棒性与有效性，本文进行了一系列参数敏感性分析实验。

为验证差异化设置平滑系数α_c的必要性，本文设计不同参数组合的实验。如表4所示，若统一采用较高的α_c([0.9, 0.9, 0.9])，难度评估过于迟钝，无法响应尾部类别训练难度变化；若统一采用较低的α_c([0.1, 0.1, 0.1])或采取逻辑反转的设置([0.1, 0.5, 0.9])，则引起头部类别在各轮次间权重剧烈震荡，严重破坏模型收敛。实验结果表明，本文采用的差异化设置α_c =[0.9, 0.5, 0.1]赋予头部类别高平滑系数以保持训练稳定，赋予尾部类别低平滑系数以敏捷响应学习难度，完美契合了长尾数据的分布特性，实现了最高精度的识别能力。

为探究β值如何调控模型对头部和尾部类别的关注权重，确定其最优设置以平衡不同类别的学习效果，本文进行了针对β参数的敏感性分析。如表5所示，若β较小，头部类仍将主导模型的预测和梯度更新，而尾部类则被忽略，平均精度显著降低；若β较大，尾部类的损失将远远超过大多数类的损失，模型由尾部类主导，从而忽略头部类与不属于尾部类的困难类，导致总精度指标的性能不佳。实验结果显示，设置β=0.999可以在尾部类别与头部类别的损失权重中取得良好平衡效果，使模型在总精度和平均精度方面都取得优异的性能。

鉴于缩放因子s在平衡模型置信度方面的重要

表 3 基于Oracle-AYNU数据集的消融研究

CutMix	RandAug	DS	准确率(%)	
			总识别准确率	平均识别准确率
×	×	×	77.32	65.14
√	×	×	78.11	65.35
×	√	×	75.32	63.29
√	√	×	76.89	64.06
×	×	√	**82.37**	**79.75**
×	√	√	79.55	73.42
√	√	√	80.56	75.48

表 4 比较不同α_c参数的性能

α_c参数	准确率(%)	
	总识别准确率	平均识别准确率
[0.1,0.1,0.1]	93.43	86.47
[0.5,0.5,0.5]	94.12	88.12
[0.9,0.9,0.9]	93.67	87.08
[0.1,0.5,0.9]	93.42	86.78
[0.9,0.5,0.1]	**94.34**	**89.89**

表 5 比较不同β参数的性能

β参数	准确率(%)	
	总识别准确率	平均识别准确率
0	94.96	86.02
0.9	94.90	87.00
0.99	**94.80**	88.03
0.999	94.48	**89.47**
0.9999	94.01	89.10

表 6 比较不同缩放因子s的性能

缩放因子s	准确率(%)	
	总识别准确率	平均识别准确率
5	94.48	**89.47**
10	94.60	88.79
20	**94.81**	88.77
30	94.17	89.11

性，本文对其进行了敏感性分析以评估其对长尾分布学习的影响及模型的鲁棒性。如表6所示，缩放因子的大小对本实验的性能影响较为微弱，体现出缩放因子s具有鲁棒性。具体来说，较小的s可以使得Softmax函数输出的概率分布相对平滑。在长尾分布的训练中，这种平滑机制有助于抑制过度自信，缓解模型对头部样本的过拟合倾向，间接地给予了尾部类别更多的优化空间。

考虑到类别间隔最大值对特征空间中决策边界的影响，本文对其进行了敏感性分析以评估其最优设置并理解其对模型性能的权衡。较大的类别间隔最大值会增大各类别之间的间隔，将尾部类别的边界推向头部类别，为稀缺的尾部类别样本在特征空间预留宽阔的安全区域。表7显示，随着最大值从0.1增加到0.5，平均识别准确率小幅上升，最终在最大值为0.5时达到了峰值89.89%。总体上类别间隔最大值参数具有鲁棒性，较大的类别间隔最大值强化了对尾部类别的正则化约束，迫使模型学习更鲁棒的特征区分稀有字形，但由于尾部类别的边界扩张幅度较大，不可避免地挤压了头部类别的特征空间，导致总识别准确率下降。

4.7 难度感知采样机制的动态验证

为验证采样机制是否达到预期，本文追踪训练过程样本实际采样情况进行评估。为排除类别样本基数悬殊带来的统计偏差，本文定义了相对采样频率，即单轮次内某类别的实际被采样的样本总数除以该类别的样本总数。该指标反映类别级别的采样强度：大于1表示被过采样，小于1表示被降采样。结合难度值与采样权重两项指标，全面评估该机制的有效性。

为直观展示难度采样机制的运行逻辑，本文在10个轮次的训练过程中提取了3个代表性对照组：全局最困难组、全局最简单组，尾部最简单组，以验证难度感知采样机制可在训练中根据类别难度动态调整采样权重。如图4(a)和图4(b)所示，难度采样机制的难度值与训练中为类别分配的采样权重一一对应。在训练轮次0时，模型根据该轮训练情况进行评估，并根据难度值为类别赋予采样权重。随训练进行与模型学习，最难组的样本难度上升，而尾部类别中最易组在小幅上升后下降，证实了模型已充分学习到尾部类别中较为简单的类别特征。而最易组的难度下降至趋近于0，在训练过程中为其分配足够小的采样权重，以使模型将重心置于需要学习的困难类别上。在图4(c)中，随训练进行与难度值和采样权重上升，最难组的采样倍率直线后最终在20附近波动，最易组在0.8倍率左右波动，而尾部最易组采样倍率总体上在8附近波动。实验数据有效证明了难度感知采样机制可随训练进行自适应调整采样策略，通过对尾部类别进行学习，从而判断尾部类别难度，对于简单的尾部类别，为其分配较低的采样权重，促使模型将重心置于真正需要模型反复学习的困难类中。

4.8 错误案例分析

本文将错误识别的样本分为4种错误类别：长尾问题、视觉相似、严重噪声与损坏和类内变种。如图5(a)所示，数据集的长尾问题未完全解决，一些尾部类特征缺乏判别性，被误分类为其他类的样本。针对极端少样本的类别，未来可结合可控生成式模型生成高质量样本进行补充；在图5(b)中，两类样本表现出非常相似的视觉特征，导致分类混淆，对于易混淆的困难类别，应同时给予较大学习

表 7　比较不同类别间隔最大值的性能

类别间隔最大值	准确率(%)	
	总识别准确率	平均识别准确率
0.1	94.48	89.47
0.2	94.66	89.71
0.3	94.63	89.67
0.4	**94.80**	89.40
0.5	94.34	**89.89**

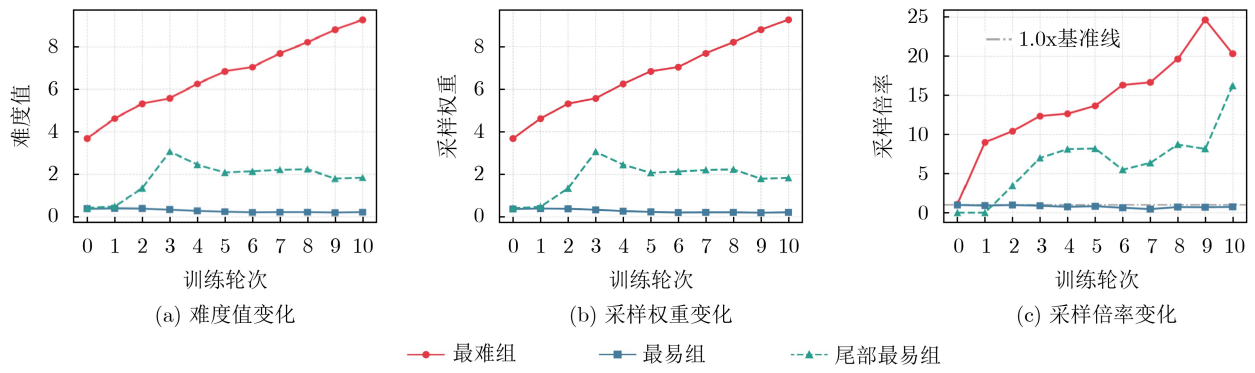


图 4 训练实际情况分析

(a) 长尾问题 (b) 视觉相似

(c) 严重噪声与损坏 (d) 类内变种

图 5 在OBC306数据集上模型识别错误的样本，共分为4种错误类型

权重与更具针对性的数据增强手段以学习类别的判别性特征；如图5(c)所示，由于年代久远及严重腐蚀等原因，样本语义信息受到严重破坏，噪声主导了样本，使其极难正确识别，单纯依赖视觉特征可能很难正确识别，未来可探索多模态联合推理技术，融合甲骨文语境语义信息对残缺部分进行推理；在图5(d)中，样本的某些类内变体之间差异显著，模型难以学习到该类样本的泛化特征，未来可结合图神经网络等针对拓扑结构识别的技术，学习拓扑不变特征，解决类内变种问题。如图6所示，本文抽取了测试集中200个识别错误的样本，并按照错误类别统计错误个数与占比。统计结果显示：残损与严重腐蚀占比最高，其次为视觉相似混淆而单纯因严重长尾和类内变种导致的直接错分仅有20例与14例。这充分证明了本文所提出的难度采样机制可以有效处理类别内变种问题与长尾问题导致的识别错误问题，并将重心移步至残损与腐蚀导致的识别错误中。

5 结束语

针对长尾甲骨文识别场景，本文提出一种通过动态调整类别难度为类别分配采样权重的两阶段式解耦学习框架，以此训练一个在所有类别上识别性能优越的神经网络。在第1阶段中，本文使用实例平衡采样方法，结合CutMix与RandAug训练InceptionNeXt-Small骨干网络与线性分类器。第2阶段设计了一种损失函数，根据训练中不同类别的损失设计难度采样器，为不同难度类别动态分配采样权重；冻结1阶段学习到的骨干网络，只对分类器进行微调，以保持1阶段骨干网络学习到的良好特征。实验证明了本文方法在OBC306数据集上的优越性能，达到了目前的最高性能水平。但是，一些类别由于严重的噪声难以学习到有效特征，识别性能仍然可以进一步提高。未来将继续探索骨干网络的特征学习能力，结合其他方法进一步解决长尾分布场景下的甲骨文识别问题。论文相关开源网址：https://www.scidb.cn/s/iUFz6r

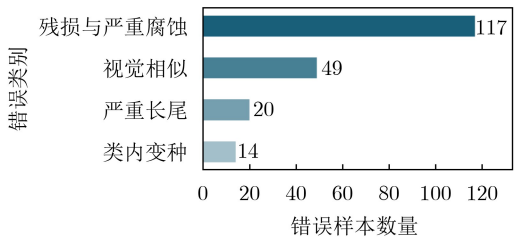


图 6 OBC306测试集上200个识别错误的样本，共分为4种错误类别

参 考 文 献

[1] 葛亮. 一百二十年来甲骨文材料的初步统计[J]. 汉字汉语研究, 2019(4): 33–54, 125. doi: 10.13513/j.cnki.41-1041/h.2019.04.006.
GE Liang. Preliminary statistics of inscribed oracle bones excavated in the past 120 years[J]. *The Study of Chinese Characters and Language*, 2019(4): 33–54, 125. doi: 10.13513/j.cnki.41-1041/h.2019.04.006.

[2] GUO Jun, WANG Changhu, ROMAN-RANGEL E, *et al*. Building hierarchical representations for oracle character and sketch recognition[J]. *IEEE Transactions on Image Processing*, 2016, 25(1): 104–118. doi: 10.1109/TIP.2015.2500019.

[3] ZHANG Yikang, ZHANG Heng, LIU Yongge, *et al*. Oracle character recognition by nearest neighbor classification with deep metric learning[C]. 2019 International Conference on Document Analysis and Recognition, Sydney, Australia,

2019: 309–314. doi: 10.1109/ICDAR.2019.00057.

[4] HUANG Shuangping, WANG Haobin, LIU Yongge, *et al*. OBC306: A large-scale oracle bone character recognition dataset[C]. International Conference on Document Analysis and Recognition, Sydney, Australia, 2019: 681–688. doi: 10.1109/ICDAR.2019.00114.

[5] GUAN Haisu, WAN Jinpeng, LIU Yuliang, *et al*. An open dataset for the evolution of oracle bone characters: EVOBC[OL]. https://doi.org/10.48550/arXiv.2312.13631, 2024.

[6] 韩佳艺, 刘建伟, 陈德华, 等. 深度长尾学习研究综述[J]. 自动化学报, 2025, 51(5): 985–1020. doi: 10.16383/j.aas.c240077.
HAN Jiayi, LIU Jianwei, CHEN Dehua, *et al*. Survey on deep long-tailed learning[J]. *Acta Automatica Sinica*, 2025, 51(5): 985–1020. doi: 10.16383/j.aas.c240077.

[7] LI Jing, WANG Qiufeng, ZHANG Rui, *et al*. Mix-up augmentation for oracle character recognition with imbalanced data distribution[C]. The 16th International Conference on Document Analysis and Recognition, Lausanne, Switzerland, 2021: 237–251. doi: 10.1007/978-3-030-86549-8_16.

[8] HUANG Hongxiang, YANG Daihui, DAI Gang, *et al*. AGTGAN: Unpaired image translation for photographic ancient character generation[C]. The 30th ACM International Conference on Multimedia, Lisboa, Portugal, 2022: 5456–5467. doi: 10.1145/3503161.3548338.

[9] LI Jing, WANG Qiufeng, WANG Siyuan, *et al*. Diff-Oracle: Deciphering oracle bone scripts with controllable diffusion model[OL]. https://doi.org/10.48550/arXiv.2312.13631, 2024.

[10] LI Jing, DONG Bin, WANG Qiufeng, *et al*. Decoupled learning for long-tailed oracle character recognition[C]. The 17th International Conference on Document Analysis and Recognition, San José, USA, 2023: 165–181. doi: 10.1007/978-3-031-41685-9_11.

[11] YUN S, HAN D, CHUN S, *et al*. CutMix: Regularization strategy to train strong classifiers with localizable features[C]. International Conference on Computer Vision, Seoul, South Korea, 2019: 6022–6031. doi: 10.1109/ICCV.2019.00612.

[12] CUBUK E D, ZOPH B, SHLENS J, *et al*. Randaugment: Practical automated data augmentation with a reduced search space[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, USA, 2020: 3008–3017. doi: 10.1109/CVPRW50498.2020.00359.

[13] LI Jing, CHI Xueke, WANG Qiufeng, *et al*. A comprehensive survey of oracle character recognition: Challenges, datasets, methodology, and beyond[J]. *Pattern Recognition*, 2026, 169: 111824. doi: 10.1016/j.patcog.2025.111824.

[14] ZHOU Xinlun, HUA Xingcheng, and LI Feng. A method of Jia Gu Wen recognition based on a two-level classification[C]. 3rd International Conference on Document Analysis and Recognition, Montreal, Canada, 1995: 833–836. doi: 10.1109/ICDAR.1995.602030.

[15] 栗青生, 杨玉星, 王爱民. 甲骨文识别的图同构方法[J]. 计算机工程与应用, 2011, 47(8): 112–114. doi: 10.3778/j.issn.1002-8331.2011.08.033.
LI Qingsheng, YANG Yuxing, and WANG Aimin. Recognition of inscriptions on bones or tortoise shells based on graph isomorphism[J]. *Computer Engineering and Applications*, 2011, 47(8): 112–114. doi: 10.3778/j.issn.1002-8331.2011.08.033.

[16] SZEGEDY C, IOFFE S, VANHOUCKE V, *et al*. Inception-v4, inception-ResNet and the impact of residual connections on learning[C]. The 31st AAAI Conference on Artificial Intelligence, San Francisco, USA, 2017: 4278–4284. doi: 10.1609/aaai.v31i1.11231.

[17] MAI C, PENAVA P, and BUETTNER R. Oracle bone inscription character recognition based on a novel convolutional neural network architecture[J]. *IEEE Access*, 2024, 12: 197021–197034. doi: 10.1109/ACCESS.2024.3521319.

[18] 毕晓君, 毛亚菲. 基于监督对比学习的小样本甲骨文字识别[J]. 智能系统学报, 2024, 19(1): 106–113. doi: 10.11992/tis.202309008.
BI Xiaojun and MAO Yafei. Few-shot oracle bone character recognition based on supervised contrastive learning[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(1): 106–113. doi: 10.11992/tis.202309008.

[19] 刘宗昊, 彭文杰, 代港, 等. 语义增强的零样本甲骨文字符识别[J]. 电子学报, 2024, 52(10): 3347–3358. doi: 10.12263/DZXB.20240286.
LIU Zonghao, PENG Wenjie, DAI Gang, *et al*. Semantic-enhanced zero-shot oracle character recognition[J]. *Acta Electronica Sinica*, 2024, 52(10): 3347–3358. doi: 10.12263/DZXB.20240286.

[20] WANG Wei, ZHANG Ting, ZHAO Yiwen, *et al*. Improving oracle bone characters recognition via a CycleGAN-based data augmentation method[C]. The 29th International Conference on Neural Information Processing, Virtual Event, 2022: 88–100. doi: 10.1007/978-981-99-1645-0_8.

[21] LI Jing, WANG Qiufeng, HUANG Kaizhu, *et al*. Towards better long-tailed oracle character recognition with adversarial data augmentation[J]. *Pattern Recognition*, 2023, 140: 109534. doi: 10.1016/j.patcog.2023.109534.

[22] CAO Kaidi, WEI C, GAIDON A, *et al*. Learning imbalanced datasets with label-distribution-aware margin loss[C]. The 33rd International Conference on Neural Information Processing Systems, Vancouver, Canada, 2019: 140.

[23] YU Sihao, GUO Jiafeng, ZHANG Ruqing, *et al*. A re-balancing strategy for class-imbalanced classification based on instance difficulty[C]. Conference on Computer Vision

and Pattern Recognition, New Orleans, USA, 2022: 70–79. doi: 10.1109/CVPR52688.2022.00017.

[24] CUI Yin, JIA Menglin, LIN T Y, *et al*. Class-balanced loss based on effective number of samples[C]. Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 9260–9269. doi: 10.1109/CVPR.2019.00949.

[25] YU Weihao, ZHOU Pan, YAN Shuicheng, *et al*. InceptionNeXt: When inception meets ConvNeXt[C]. Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2024: 5672–5683. doi: 10.1109/CVPR52733.2024.00542.

孙君为：男，研究方向为甲骨文识别与破译、模式识别.

关素妍：女，研究方向为甲骨文识别与破译.

陈信宇：男，研究方向为甲骨文识别与破译.

王 堃：男，研究方向为甲骨文识别与破译.

蔡院强：男，讲师，研究方向为计算文字学.

责任编辑：余 蓉

Decoupled Learning for Long-tailed Oracle Bone Character Recognition Based on Adaptive Difficulty Sampling

SUN Junwei GUAN Suyan CHEN Xinyu WANG Kun CAI Yuanqiang

(*School of Computer Science* (*National Pilot Software Engineering School*), *Beijing University of Posts and Telecommunications*, *Beijing*, 100876, *China*)

Abstract:

Objective Oracle Bone Character (OBC) recognition is challenged by an extreme long-tailed distribution and substantial intra-class variation. Conventional deep learning methods are often dominated by head classes, whereas existing approaches tend to overfit tail classes or fail to account for differences in learning difficulty across classes. To address these limitations, a two-stage decoupled learning framework is proposed to improve the recognition of tail and difficult classes while preserving the discriminative capability of head classes.

Methods The proposed framework decouples feature representation learning from classifier optimization. In the first stage, the backbone network is trained using a mixed data augmentation strategy that combines CutMix and RandAugment with Label-Distribution-Aware Margin (LDAM) loss to learn robust feature representations and alleviate the effect of intra-class variation. In the second stage, the backbone network is frozen, and only the classifier is optimized. An adaptive difficulty sampling strategy is proposed to dynamically assign sampling weights according to historical and current class-level training difficulty. The classifier is further optimized using a Class-Balanced LDAM (CBL) loss, which combines class-balanced weighting with LDAM to refine decision boundaries for long-tailed classification.

Results and Discussions Experiments on the highly imbalanced OBC306 dataset demonstrate that the proposed method achieves an overall accuracy of 94.34% and an average class accuracy of 89.89%. Compared with the Inception-v4 baseline, the proposed method improves the average class accuracy by 19.61%. Comparisons with representative long-tailed OBC recognition methods further demonstrate superior overall performance. Comprehensive ablation studies verify the effectiveness of the mixed data augmentation strategy and the adaptive difficulty sampling strategy in improving the recognition of rare and difficult characters. Parameter sensitivity analysis and qualitative error analysis further confirm the robustness and effectiveness of the proposed framework.

Conclusions The proposed two-stage decoupled learning framework effectively addresses long-tailed OBC recognition by balancing the learning priorities of head, tail, and difficult classes. The mixed data augmentation strategy improves feature robustness, whereas the adaptive difficulty sampling strategy and the Class-Balanced LDAM loss jointly optimize classifier learning and refine decision boundaries without degrading head-class recognition performance. The proposed framework provides an effective solution for the digital recognition of Oracle Bone Characters and offers technical support for low-resource ancient character recognition.

Key words: Oracle Bone Character(OBC) recognition; Long-tailed distribution; Adaptive difficulty sampling; Decoupled learning