



# 计算机科学

COMPUTER SCIENCE

## 大选择性核双边网络的长尾分布医学图像分类方法

孙汤慧, 赵刚, 郭美倩

引用本文

孙汤慧, 赵刚, 郭美倩. 大选择性核双边网络的长尾分布医学图像分类方法[J]. 计算机科学, 2025, 52(4): 231-239.

SUN Tanghui, ZHAO Gang, GUO Meiqian. Long-tail Distributed Medical Image Classification Based on Large Selective Nuclear Bilateral-branch Networks [J]. Computer Science, 2025, 52(4): 231-239.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于带毒分类器的自监督后门攻击防御方法](#)

Self-supervised Backdoor Attack Defence Method Based on Poisoned Classifier

计算机科学, 2025, 52(4): 336-342. <https://doi.org/10.11896/jsjcx.240100005>

[基于多尺度融合注意力的多视角文档图像篡改检测与定位](#)

Multi-view and Multi-scale Fusion Attention Network for Document Image Forgery Localization

计算机科学, 2025, 52(4): 327-335. <https://doi.org/10.11896/jsjcx.240100142>

[基于YOLO-Unet组合网络的牛只个体识别方法研究](#)

Research on Individual Identification of Cattle Based on YOLO-Unet Combined Network

计算机科学, 2025, 52(4): 194-201. <https://doi.org/10.11896/jsjcx.240100144>

[基于虹膜纹理感知的自适应关联学习方法](#)

Adaptive Contextual Learning Method Based on Iris Texture Perception

计算机科学, 2025, 52(4): 185-193. <https://doi.org/10.11896/jsjcx.250100022>

[基于边缘约束和改进Swin Unetr的复杂器官分割方法](#)

Complex Organ Segmentation Based on Edge Constraints and Enhanced Swin Unetr

计算机科学, 2025, 52(4): 177-184. <https://doi.org/10.11896/jsjcx.240600007>

# 大选择性核双边网络的长尾分布医学图像分类方法

孙汤慧<sup>1</sup> 赵刚<sup>1,2</sup> 郭美倩<sup>1</sup>

1 南昌航空大学数学与信息科学学院 南昌 330063

2 南昌航空大学无损检测技术教育部重点实验室 南昌 330063

(sunth666@163.com)

**摘要** 医学场景下的数据集通常呈现长尾分布的特点,这种不平衡性可能导致模型偏向头部类,而对尾部类的识别性能较差,从而影响模型的准确性。常见的解决方法是对原始数据进行数据增强,使其具备平衡分布的特点,但增强后的尾部类样本质量往往不佳,没有真正改善尾部类的分类精度。针对此问题,提出一种大选择性核双边网络模型(LSKBB)。该模型主要由传统学习分支和重新再平衡分支两部分组成,采用LSK模块来获取关键信息和关注上下文信息,设计了可以使模型由一个关注方向逐渐过渡到另一个关注方向的动态损失函数,从而提高分类精度。在不改变长尾分布特点的医学数据集上进行图像分类实验,与现有方法相比,所提出的LSKBB模型性能在不平衡率为10,50和100时,在BrecaKHis数据集下,准确率分别提高1.41%,1.25%和1.25%;在ChestX-ray数据集下,准确率分别提高6.10%,3.15%和2.47%。实验结果表明,LSKBB模型在不同的不平衡率下性能较好,可用于长尾分布的医学数据集的分类检测。

**关键词**:长尾分布;深度学习;双分支网络;LSK模块;图像分类

中图分类号 TP391.4

## Long-tail Distributed Medical Image Classification Based on Large Selective Nuclear Bilateral-branch Networks

SUN Tanghui<sup>1</sup>, ZHAO Gang<sup>1,2</sup> and GUO Meiqian<sup>1</sup>

1 School of Mathematics and Information Science, Nanchang Hangkong University, Nanchang 330063, China

2 Key Laboratory of Nondestructive Testing Technology of the Ministry of Education, Nanchang Hangkong University, Nanchang 330063, China

**Abstract** In medical scenarios, datasets often exhibit characteristics of a long-tailed distribution, where in the imbalance may cause models to favor head classes, resulting in poorer performance in identifying tail classes and thus affecting model accuracy. Common approaches involve data augmentation to transform original data into a balanced distribution. However, the quality of augmented tail class samples is often inadequate, failing to genuinely improve the classification accuracy of tail classes. Addressing this issue, this paper proposes a large selective kernel bilateral branch network model(LSKBB). The model mainly consists of two parts: the traditional learning branch and the re-balancing branch. It adopts the LSK module to acquire key information and focus on contextual information. Additionally, a dynamic loss function is designed to enable the model to transition gradually from one focus direction to another, thereby enhancing classification accuracy. In image classification experiments conducted on medical datasets with long-tail distributions without altering their characteristics, the proposed LSKBB model shows performance improvements compared to existing methods. When the imbalance ratios are 10, 50, and 100, the accuracy of the LSKBB model increases by 1.41%, 1.25%, and 1.25%, respectively, on BrecaKHis dataset. On ChestX-ray dataset, the accuracy increases by 6.10%, 3.15%, and 2.47%, respectively. The experimental results indicate that the LSKBB model achieves good performance under different imbalance ratios and is suitable for classification and detection on medical datasets with long-tail distributions.

**Keywords** Long-tail distribution, Deep learning, Double branch network, LSK module, Image classification

到稿日期:2024-07-08 返修日期:2024-10-21

基金项目:国家自然科学基金(62366033);无损检测技术教育部重点实验室开放基金(EW202107216);江西省自然科学基金(20242BAB25121)

This work was supported by the National Natural Science Foundation of China(62366033), Open Fund of Key Laboratory of Non-destructive Testing Technology of the Ministry of Education(EW202107216) and Natural Science Foundation of Jiangxi Province, China(20242BAB25121).

通信作者:赵刚(zhaogang0209@163.com)

## 1 引言

近年来,深度学习在学术界和工业界越来越受欢迎,包括语音识别<sup>[1-2]</sup>、计算机视觉<sup>[3-4]</sup>和自然语言处理<sup>[5-6]</sup>在内的各个领域都见证了深度学习的巨大力量,这离不开高质量的呈现均匀分布的大规模数据集,如 ImageNet ILSVRC 2012<sup>[7]</sup>、CIFAR-10<sup>[8]</sup>和 MS COCO<sup>[9]</sup>等。而在现实世界中呈现长尾分布的数据集广泛存在,这类数据集中头部类占据总样本量的绝大部分,尾部类占据总样本量的小部分,样本类别与图片数量的关系如图 1 所示。这种不平衡性使得算法偏向于头部类,从而忽略了对尾部类的学习,导致对尾部类样本的预测出现错误。这种误判可能对患者的健康状况和治疗计划产生严重影响。因此,研究呈现长尾分布的医学数据集的图像分类任务具有重要的意义。

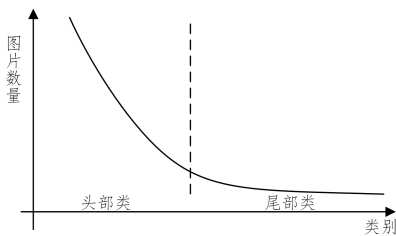


图 1 长尾分布数据的类别与图片数量的关系图

Fig. 1 Diagram of relationship between the categories of long-tail distributed data and the number of pictures

在现有长尾分布图像分类方法中,类重平衡策略是一种主流方法,它主要分为两类,即重新采样和重加权方法。重新采样方法的核心思想是将长尾分布的数据集转化为平衡分布的数据集,其方法有随机过采样、随机欠采样、合成少数过采样技术和直接过采样等<sup>[10]</sup>。例如,Hoyos-Osorio 等提出了基于相关信息的欠采样方法(Relevant Information-based UnderSampling, RIUS),用于从多数类中选择最相关的样本,以提高不平衡数据的分类性能<sup>[11]</sup>。与重新采样相比,重加权方法更为简单且轻量<sup>[12]</sup>。它通常对少数类的训练样本分配较大的权重,对多数类的训练样本分配较小的权重,采用的方法是通过调整训练样本的权重来改变损失函数<sup>[13]</sup>。

还有一类方法是基于改进网络模块的方法解决长尾分布的数据集图像分类任务。例如,Kang 等将分类网络解耦为表征学习和分类器,使用最简单的实例平衡采样学习表征,也可以通过调整分类器来获得较强的长尾识别能力<sup>[14]</sup>。Zhou 等提出了基于累积学习的双侧分支网络模型(Bilateral-Branch Network with Cumulative Learning, BBN)来同时处理表示学习和分类器学习,其中每个分支都单独执行自己的任务<sup>[15]</sup>。它通过累积学习策略先学习通用模式,然后逐渐关注尾部数据。Kang 等提出了 k-positive 对比学习,有效地结合了监督学习和对比学习的优点,可以学习具有判别性的表征,提高了模型的泛化能力<sup>[16]</sup>。Li 等提出的大选择性核网络(Large Selective Kernel Network, LSK),可以动态调整其大空间感受野,更好地模拟遥感场景中各种物体的测距环境<sup>[17]</sup>。

在医学数据集中,图像分类任务可以采用特征工程和

特征提取的通用特征描述符的方法。特征工程方法成功的关键在于找到有代表性的特征,而这些特征通常需由医学专家设计,效率低且计算量大。此外,特征工程的初始化和处理步骤也繁琐耗时。而特征描述符提取的特征不足以描述组织病理图像,如低水平和不具代表性的表面特征,这些特征不适合具有判别分析能力的分类器。

为了更高效地解决这些问题,且更好地处理长尾分布的医学数据集,越来越多的学者尝试其他的方法,如数据增强、代价敏感学习<sup>[18-19]</sup>、数据重加权<sup>[20]</sup>、迁移学习<sup>[21]</sup>和集成学习等。

数据增强是通过对原始数据进行各种变换或合成,生成新的训练样本。常见的数据增强方法包括图像旋转、缩放、裁剪、翻转以及合成新的样本。例如,Galdran 等将流行的 MixUp 正则化技术和改进的训练数据采样策略相结合,提出了 Balanced-MixUp 采样技术,它通过同时执行基于实例和基于类的采样,并将两组样本混合,创建了一个平衡的训练数据分布<sup>[22]</sup>。Ding 等提出了一种基于生成对抗网络(Generative Adversarial Network, GAN)的 IBGAN 模型,该模型主要用于生成类内稀疏样本和类边界样本。第一阶段,设计隔离森林算法和边界样本检测算法,识别类内数据中的稀疏区域样本和边界样本。第二阶段,提出了一种基于支持向量数据描述(Support Vector Data Description, SVDD)的样本评价方法,以滤除生成样本中的噪声数据,保证生成数据的质量<sup>[23]</sup>。数据增强虽然可以提高模型的性能,但是它通常需要消耗额外的计算资源,且会丢失原始数据的一些性能。

集成学习是将多个深度学习模型的预测结果进行集成,通过结合不同模型的优点来改善分类性能。例如,Alqudah 等提出了一种基于卷积神经网络(Convolutional Neural Networks, CNN)和基于集成技术的卷积神经网络组合的全筛选系统,来完成腺癌组织病理学图像的分类任务<sup>[24]</sup>。

对于长尾分布的医学图像分类任务,在不改变数据集原始长尾分布特点的前提下,本文提出了大选择性核双边 LSK-BB(Large Selective Kernel Bilateral Branch)网络模型。LSK-BB 由传统学习分支和重新再平衡分支组成,传统学习分支的采样器为经典的均匀采样器,重新再平衡分支的采样器为反向采样器。将这两个分支得到的输出融合输入到损失函数中。损失函数为新提出的动态损失函数,通过引入自适应的调节函数,可以根据训练过程中的实时情况进行调整,从而自适应地改变模型的学习方向。

## 2 相关知识

### 2.1 LSK 模块

LSK module<sup>[17]</sup>由一个大核卷积序列和一个空间核选择机制组成,它是将大核卷积序列解耦出一系列具有逐渐增长的核和逐渐增加扩张率的深度可分离卷积,再通过空间选择机制,将一系列大深度可分离核处理的特征进行自适应的动态加权并进行空间融合,以根据空间调整每个目标的感受野。LSK 模块结构如图 2 所示。

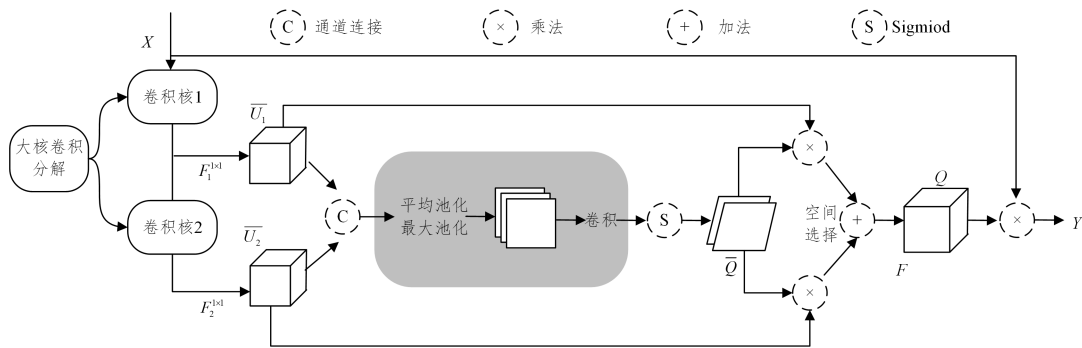


图2 LSK 模块

Fig. 2 LSK module

### 2.1.1 大核卷积

大核卷积被分解为一系列具有逐渐增长的核和逐渐增加扩张率的深度可分离卷积。该序列中第  $i$  次深度卷积的核大小  $k$ 、扩张率  $d$  和感受野  $RF$  定义为:

$$k_{i-1} \leq k_i, d_1 = 1, d_{i-1} < d_i < RF_{i-1} \quad (1)$$

$$RF_1 = k_1, RF_i = d_i(k_i - 1) + RF_{i-1} \quad (2)$$

核大小和扩张率的增加,保证了感受野的扩张速度足够快。设定扩张率的上限,以保证膨胀卷积不会在特征映射之间引入间隙。

为了获得输入  $X$  在不同范围内具有丰富上下文信息的特征,应用一系列具有不同感受野的深度卷积进行分解:

$$U_0 = X, U_{i+1} = F_i^{d_w}(U_i) \quad (3)$$

其中,  $F_i^{d_w}(\cdot)$  是具有核  $k_i$  和扩张率为  $d_i$  的深度卷积。假设有  $N$  个分解的核,每个核通过一个  $1 \times 1$  卷积层  $F^{1 \times 1}(\cdot)$  进一步处理:

$$\bar{U}_i = F_i^{1 \times 1}(U_i), i = 1, 2, \dots, N \quad (4)$$

允许对每个空间特征向量进行通道混合。

### 2.1.2 空间核选择机制

基于多尺度特征对不同目标动态选择核的选择机制的步骤如下。

1) 将从不同核得到的特征与不同的感受野进行拼接,得到:

$$\bar{U} = [\bar{U}_1; \dots; \bar{U}_i] \quad (5)$$

2) 对  $\bar{U}$  应用基于通道的平均池化  $P_{\text{avg}}(\cdot)$  和最大池化  $P_{\text{max}}(\cdot)$  来高效地提取空间关系:

$$Q_{\text{avg}} = P_{\text{avg}}(\bar{U}), Q_{\text{max}} = P_{\text{max}}(\bar{U}) \quad (6)$$

其中,  $Q_{\text{avg}}$  和  $Q_{\text{max}}$  分别是平均和最大池化的空间特征描述符。

3) 为了允许不同空间特征描述符之间的信息交互,将空间池化特征拼接起来,使用一个卷积层  $F^{2 \rightarrow N}(\cdot)$  将池化特征(2个通道)变换成  $N$  个空间注意力块:

$$\bar{Q} = F^{2 \rightarrow N}([Q_{\text{avg}}; Q_{\text{max}}]) \quad (7)$$

4) 对于每个空间注意力块  $\bar{Q}$ ,应用 sigmoid 激活函数获得每个分解大核的单独空间选择掩码:

$$\bar{Q}_i = \sigma(\bar{Q}) \quad (8)$$

其中,  $\sigma(\cdot)$  为 sigmoid 函数。

5) 对分解的大核序列中的特征进行相应的空间选择加权,并通过卷积层  $F(\cdot)$  进行融合,得到注意力特征  $Q$ :

$$Q = F(\sum_{i=1}^N (\bar{Q}_i \cdot \bar{U}_i)) \quad (9)$$

LSK 模块的最终输出  $Y$  是输入特征  $X$  与注意力特征  $Q$  的积<sup>[25]</sup>:

$$Y = X \cdot Q \quad (10)$$

最终输出  $Y$  不仅保留了输入特征图的空间维度和通道数,而且输入特征图中的每个像素值都根据注意力特征图进行了调整,从而在保持原始结构的同时增强了特征表达能力。

## 2.2 Focal Loss

Focal Loss<sup>[26]</sup>是为了解决在训练期间前景和背景类之间存在极端不平衡的单阶段对象检测场景而提出的。它是在标准交叉熵准则中增加了一个调节因子  $(1 - p_i)^\gamma$  和权重因子  $\alpha_i$ 。设置  $\gamma > 0$  减少了分类良好的样本的相对损失 ( $p_i > 0.5$ ),将更多的注意力放在难分类的样本上。Focal Loss 的计算式为:

$$FL(p_i) = -\alpha_i (1 - p_i)^\gamma \log(p_i) \quad (11)$$

其中,  $p_i$  为模型的预测概率分数;  $\alpha_i$  为权重因子,且  $\alpha_i \in [0, 1]$ ;  $(1 - p_i)^\gamma$  为调节因子;  $\gamma$  为参数,当  $\gamma$  为 0 时,式(11)为标准的交叉熵损失公式。Focal Loss 可用于减小易分类样本的损失贡献,即  $p_i$  越大,该样本越容易被区分,调制因子也就越小。这样可以减小易分类样本的损失贡献,从而增加难分类样本的损失比例。当  $p_i$  趋向于 1 时,该样本是易分类样本,此时调节因子  $(1 - p_i)^\gamma$  趋向于 0,说明对损失的贡献较小,即降低了易分类样本的损失比例。当  $p_i$  趋于 0,即某个样本被错误地分到正样本时,调制因子  $(1 - p_i)^\gamma$  趋向于 1,对损失影响不大。因此,通过  $\alpha_i$  可以抑制正负样本数量的失衡,通过  $\gamma$  可以控制难易分类样本数量的失衡。

## 2.3 CE Loss

交叉熵损失(CE Loss)是一种常用的损失函数,用于衡量两个概率分布之间的差异。在分类问题中,CE Loss 通常用于衡量模型的输出概率分布与真实标签的分布之间的差异。CE Loss 在训练神经网络时经常被用作优化目标,以最小化模型输出与实际标签之间的差距,其计算式为:

$$\text{loss} = -y \log(p) \quad (12)$$

其中,  $p$  为模型的预测值,  $y$  为真实标签。

## 2.4 Class-Balanced Softmax Cross-Entropy Loss

CSCE Loss<sup>[27]</sup>通过引入类别平衡因子,对不同类别的样本赋予不同的权重,从而平衡不同类别之间的影响。具体来说,对于类别不平衡问题,CSCE Loss 通过调整损失函数的

权重,使得在训练过程中更加关注数量较少的类别,以减小其误差对整体损失的影响,从而平衡不同类别之间的训练效果。这种权重的调整,通常基于类别的图片数量或者其他衡量类别重要性的指标来确定。

假设所有类别的模型预测输出为  $\mathbf{p}=[p_1, p_2, \dots, p_c]^T$ , 其中  $C$  为类别数量。Softmax 函数将每个类视为互斥的,并计算所有类的概率分布为  $p_i = \frac{\exp(z_i)}{\sum_{j=1}^C \exp(z_j)}$ ,  $\forall i \in \{1, 2, \dots, C\}$ 。给定一个类别标签为  $y$  的样本,假设  $y$  类有  $n_y$  个训练样本,其损失为:

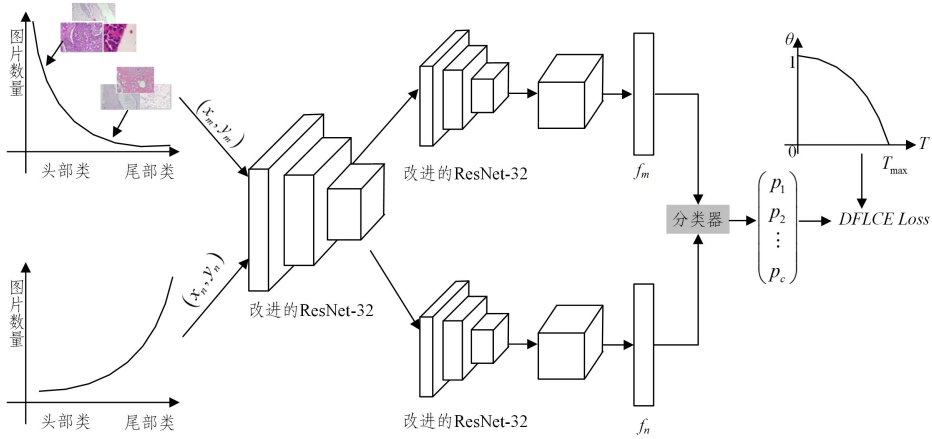


图3 LSKBB模型框架

Fig. 3 LSKBB model framework

### 3.1 总体框架

LSKBB模型如图3所示,设  $\mathbf{x}$  表示训练样本,  $y$  表示其对应的真实标签,  $C$  表示类别总数。传统学习分支的输入数据为  $(x_m, y_m)$ , 重新再平衡分支的输入数据为  $(x_n, y_n)$ , 两个分支通过改进的 ResNet-32 来提取特征,再由全局平均池化将整个特征图分别降维到一维特征向量  $\mathbf{f}_m \in \mathbb{R}^D$  和  $\mathbf{f}_n \in \mathbb{R}^D$ 。然后将  $\mathbf{f}_m$  和  $\mathbf{f}_n$  乘不同的权重,再乘相应的分类器,并通过元素加法得到预测逻辑值:

$$\mathbf{Z} = \lambda \mathbf{W}_m^T \mathbf{f}_m + (1 - \lambda) \mathbf{W}_n^T \mathbf{f}_n \quad (14)$$

其中,  $\mathbf{Z} \in \mathbb{R}^C$  是预测逻辑值,  $\mathbf{Z} = [z_1, z_2, \dots, z_C]^T$ ,  $\lambda$  是权重超参数,  $\mathbf{W}_m \in \mathbb{R}^{D \times C}$  和  $\mathbf{W}_n \in \mathbb{R}^{D \times C}$  是分类器。对每一个类别预测逻辑值并输入到 softmax 分类器中, softmax 的公式为:

$$p_i = p(\bar{y} = i | \mathbf{x}) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \quad (15)$$

其中,  $\mathbf{x}$  是输入,  $\bar{y}$  是类别标签,  $i \in \{1, 2, \dots, C\}$ ,  $p_i$  是样本  $\mathbf{x}$  属于类别  $\bar{y}$  的条件概率。预测输出为  $\mathbf{p} = [p_1, p_2, \dots, p_C]^T$ , 预测值  $P$  可以表示为:

$$P = \arg \max_i p_i, i \in \{1, 2, \dots, C\} \quad (16)$$

LSKBB 损失函数为新的动态损失函数 DFLCE Loss (Dynamic Focal Loss-CE Loss), 其计算式见式(19)。

### 3.2 改进的 ResNet-32

在图像识别任务中,用来提取特征的骨干网络一般选 ResNet 网络,而 ResNet 具有局部性,容易忽略远距离像素间的依赖关系,从而忽视关键信息。为了不错过关键信息,更加关注到上下文信息,在 ResNet 中加入 LSK 模块。

$$\text{CSCE}(z, y) = -\frac{1-\beta}{1-\beta^{n_y}} \log \left( \frac{\exp(z_y)}{\sum_{j=1}^C \exp(z_j)} \right) \quad (13)$$

其中,  $\beta$  为有效样本数。

### 3 大选择性核双边网络 LSKBB

大选择性核双边网络 LSKBB 主要由传统学习分支和重新再平衡分支组成。前者使用均匀采样器采样,后者使用反采样器采样,它们的相同之处是使用改进的 ResNet 来提取特征,并且共享权重,但唯独最后一个残差块不共享权重。LSKBB 模型结构图如图 3 所示。

LSK 模块能够产生具有不同大小的感受野,这有利于后续选择适当的核来提取这些特征,从而更有效地提取有用的信息;而且 LSK 模块使用的序列解耦操作可以有效地减少模型的参数量。因此,在 ResNet-32 网络中的基础模块的跳跃连接处加入了 LSK 模块,得到改进的 ResNet-32。其基础模块结构如图 4 所示,图中灰色方框为 LSK 模块。

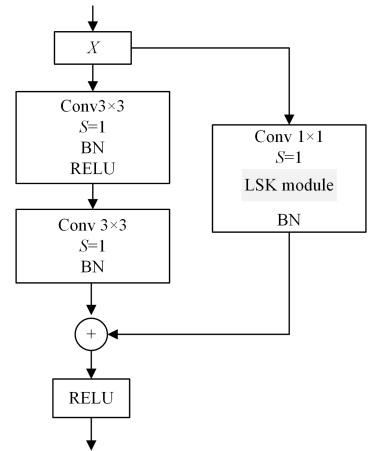


图4 改进的 ResNet-32 的基础模块

Fig. 4 Basic module of improved ResNet-32

### 3.3 新动态损失函数

混合损失函数<sup>[28]</sup>是指通过组合多个不同的损失函数来形成一个综合的损失函数。通过结合不同的损失函数,混合损失函数可以同时考虑两个及两个以上的优化目标,从而促使模型更好地学习特征表示。

而混合损失函数对不同损失函数的权重不同,这使得模型更加关注对整体性能有更大贡献的损失项,从而有效地平衡不同类别对模型训练的影响,提高模型的整体性能。混合损失函数为:

$$loss = \alpha \cdot L_1(p, y) + \beta \cdot L_2(p, y) \quad (17)$$

其中,  $\alpha$  和  $\beta$  是权重,  $L_1$  和  $L_2$  是不同的损失函数。

在实践中,混合损失函数中各个损失函数之间的权重  $\alpha$  和  $\beta$  通常需要进行多次实验,才能找到最适合具体任务和数据集的权重组合。为了优化这一过程,本文提出用调节函数替代混合损失函数的权重,调节函数详细介绍见 3.3.1 小节。

呈现长尾分布的数据集往往是类别不平衡的,针对这类数据集常用的是损失函数 Focal Loss 和 CSCE Loss。又因为每个样本对模型的学习都有一定的贡献度,所以本文还选择了交叉熵损失。

综上所述,本文提出了两个新动态损失函数 DFLCE Loss 和 DFLCSCE Loss (Dynamic Focal Loss-CSCE Loss)。它们可以动态地调整混合损失函数中的权重,使得模型优化的方向从关注一个方向逐渐地转移到另一个方向,从而提高模型的性能和效果。

### 3.3.1 调节函数

调节函数  $\theta$  可以控制混合损失函数的权重,将学习重点由关注一个问题逐渐过渡到关注另一个问题。调节函数的自变量是迭代次数,其公式为:

$$\theta = 1 - \left( \frac{T}{T_{\max}} \right)^2 \quad (18)$$

其中,  $T$  为当前迭代次数,  $T_{\max}$  为总迭代次数。 $\theta$  随  $T$  的增加而逐渐减小。调节函数图像如图 5 所示。

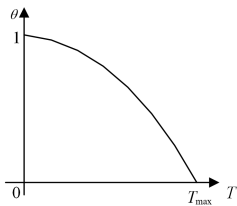


图 5 调节函数的函数图像

Fig. 5 Function image of adjustment function

### 3.3.2 DFLCE Loss

为了使得模型的优化方向可以动态变化,设计模型的优化方向从侧重关注困难样本逐渐地过渡到关注每一个样本。本文设计了新的动态损失函数——DFLCE Loss。Focal Loss 的重心聚焦在那些难区分的样本,而 CE Loss 的重心聚焦在每一个样本,所以 DFLCE Loss 混合了损失函数 Focal Loss 和 CE Loss。此外,引入调节函数,使得混合函数的权重可以进行动态的调整。

将式(11)、式(12)和式(18)代入到式(17)中,得到动态损失函数 DFLCE Loss,其计算式为:

$$loss = \left( \left( \frac{T}{T_{\max}} \right)^2 - 1 \right) \alpha_r (1 - p_r)^{\gamma} \log(p_r) - \left( \frac{T}{T_{\max}} \right)^2 \gamma \log(p_r) \quad (19)$$

### 3.3.3 DFLCSCE Loss

同样地,设计模型的优化方向从侧重关注困难样本逐渐地过渡到侧重关注尾部类别的样本。本文设计了一种新的动态损失函数——DFLCSCE Loss。Focal Loss 的重心聚焦在那些难区分的样本,而 CSCE Loss 的重心聚焦在尾部类样本,所以 DFLCSCE Loss 混合了损失函数 Focal Loss 和 CSCE Loss。此外,引入调节函数,使得混合函数的权重可以进行动态的调整。

将式(11)、式(13)和式(18)代入到式(17)中,得到动态损失函数 DFLCSCE Loss,其计算式为:

$$loss = \left( \left( \frac{T}{T_{\max}} \right)^2 - 1 \right) \alpha_r (1 - p_r)^{\gamma} \log(p_r) - \left( \frac{T}{T_{\max}} \right)^2 \frac{1 - \beta}{1 - \beta^{n_r}} \log \left( \frac{\exp(z_y)}{\sum_{j=1}^C \exp(z_j)} \right) \quad (20)$$

## 4 实验

### 4.1 数据集

BreaKHis 数据集: BreaKHis 数据集由巴西巴拉那州 P&D 实验室的病理解剖学和细胞病理学合作建立,包含了来自 82 名患者的乳腺癌组织病理图像数据。该数据集有 7909 张图像,涵盖了不同放大倍数(40 倍、100 倍、200 倍和 400 倍)的病理解剖图像,包含了 8 个类别,这 8 个类别分为良性样本和恶性样本,其中良性样本 2480 个,恶性样本 5429 个。良性乳腺肿瘤分为 4 种,分别为腺病、纤维腺瘤、叶状瘤和管状腺瘤;恶性乳腺肿瘤分为 4 种,分别为导管癌、小叶癌、粘液腺癌和乳头状癌。图 6 展示了数据集中的 8 个类别的部分组织病理学图像。表 1 列出了 8 类乳腺癌的组织病理图像的数量。

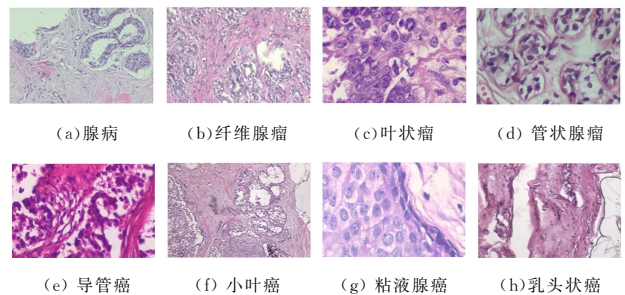


图 6 乳腺癌组织病理学图像

Fig. 6 Breast cancer histopathological image

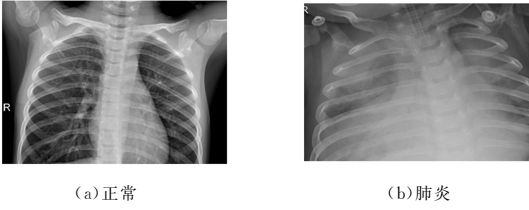
表 1 BreaKHis 数据集中每个类别的数量统计

Table 1 BreaKHis dataset counts for each category

肿瘤的性质	肿瘤类型	数量	总计
良性肿瘤	腺病	444	2480
	纤维腺瘤	1014	
	叶状瘤	453	
	管状腺瘤	569	
恶性肿瘤	导管癌	3451	5429
	小叶癌	626	
	粘液腺癌	792	
	乳头状癌	560	
总计		7909	7909

ChestX-ray 数据集: ChestX-ray 数据集选自广州市妇幼

保健中心的 1 至 5 岁儿科患者的 5856 张胸腔 X 射线透视图, 正常类别有 1583 张, 肺炎类别有 4273 张。图 7 展示了数据集中两类的部分胸腔 X 射线透视图, 左图为正常类别图片, 右图为肺炎类别图片。



(a) 正常 (b) 肺炎

图 7 有/无肺炎的胸部 X 射线透视图

Fig. 7 Chest X-ray radiogram with/without pneumonia

## 4.2 数据预处理

BreaKHis 数据集的图片像素均为  $700 \times 460 \times 3$ , 将每张图片缩放成  $350 \times 230 \times 3$  大小并随机采样成  $224 \times 224 \times 3$  大小, 提供给模型使用。将数据集分为训练集和测试集。训练集为长尾分布的, 每个类别的图片数量如图 8 的蓝色曲线所示。测试集和验证集为平衡分布, 即每个类别图像数量为 80 张。

为了进行不同的不平衡率 ( $\beta = \frac{N_{\max}}{N_{\min}}$ ,  $N_{\max}$  表示最多类别的训练样本数量,  $N_{\min}$  表示最少类别的训练样本数量) 的实验, 将原始数据集自动抽取不同类别的不同数量的照片, 以形成不平衡率为 50 和 100 的数据集, 其中原始数据的不平衡率为 11.59, 可近似看成不平衡率为 10, 这样处理后的数据集仍然符合长尾分布的特点。不同的不平衡率下的训练集数量分布图如图 8 所示。

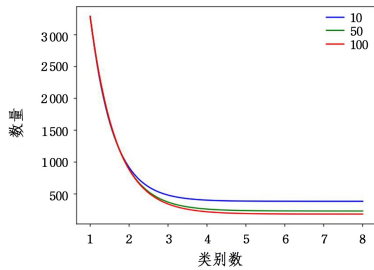


图 8 不同的不平衡率下类别的图像数量分布图  
(电子版为彩图)

Fig. 8 Category of image quantity distribution map under different imbalance ratios

ChestX-ray 数据集的图片大小不统一, 每张图片通过缩放为原来的  $1/4$ , 再沿着图片中心剪裁后缩放为原来的  $1/2$ , 得到了统一的  $224 \times 224 \times 3$  的图片。为了提供不同的不平衡率的数据集, 通过删除训练集的照片得到不平衡率为 10, 50 和 100 的 ChestX-ray 数据集。测试集的图片均为正常类 234 张、肺炎类 390 张。

## 4.3 实验设置

实验软硬件环境为: CPU 为 Intel (R) Core (TM) i9-12900H@2.50 GHz, GPU 为 NVIDIA GeForce RTX 3060 Laptop, 内存为 16 GB, Windows 11 操作系统, CUDA11.4, CuDNN8.9.5.30, PyTorch 深度学习框架, Python 3.11 编程

语言。训练模型的优化算法为随机梯度下降, 学习率为 0.05, 动量为 0.9, 权重衰减系数为  $2 \times 10^{-4}$ 。

## 4.4 实验评价指标

本文实验以准确率 (Accuracy)、精确率 (Precision) 和敏感度 (Sensitivity) 为评价指标。

准确率的表达式为:

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} \quad (21)$$

精确率的表达式为:

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

敏感度的表达式为:

$$Sensitivity = \frac{TP}{TP + FN} \quad (23)$$

其中,  $TP$  (True Positive) 为真正例,  $FN$  (False Negative) 为假反例,  $FP$  (False Positive) 为假正例,  $TN$  (True Negative) 为真正反例。

## 5 实验结果与分析

### 5.1 大选择性核双边网络模型的分类表现

将大选择性核双边网络模型 LSKBB 与现有方法进行对比实验, 使用的数据集为不平衡率为 10, 50 和 100 的 Breakhis 数据集和 ChestX-ray 数据集, 实验得出的分类准确率如表 2 和表 3 所列 (粗体表示每列的最优结果)。

表 2 LSKBB 与各对比方法在 Breakhis 上的准确率

Table 2 Accuracy of LSKBB and comparison methods on Breakhis (%)

数据集 不平衡率	Breakhis			
	10	50	100	平均值
Focal Loss <sup>[26]</sup>	75.51	72.01	71.05	72.86
LDAM-DRW <sup>[29]</sup>	80.80	78.06	76.95	78.60
BBN <sup>[15]</sup>	86.56	82.19	78.59	82.45
MiSLAS <sup>[30]</sup>	83.62	80.45	77.90	79.99
LSKBB	<b>87.97</b>	<b>83.44</b>	<b>79.84</b>	<b>83.75</b>

从表 2 可以看出, 在不同的不平衡率的 Breakhis 数据集中, LSKBB 相较于同样采用双分支结构的 BBN 模型, 平均准确率提升 1.30%。这是因为 LSKBB 在 ResNet 网络中加入了 LSK 模块, 能够捕获关键信息且更关注上下文信息, 从而能更有效地提取有用的信息。此外, DFLCE Loss 的加入, 使得模型的优化方向从侧重关注困难样本逐渐地过渡到侧重关注每一个样本。与 MiSLAS 相比, LSKBB 模型的准确率在不平衡率为 10, 50 和 100 的 Breakhis 数据集上分别高出 4.35%, 2.99%, 1.94%。

表 3 LSKBB 与各对比方法在 ChestX-ray 上的准确率

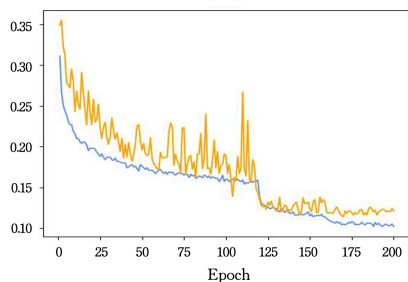
Table 3 Accuracy of LSKBB and comparison methods on ChestX-ray (%)

数据集 不平衡率	ChestX-ray			
	10	50	100	平均值
Focal Loss <sup>[26]</sup>	73.38	70.25	69.51	71.05
LDAM-DRW <sup>[29]</sup>	78.20	77.35	75.95	77.17
BBN <sup>[15]</sup>	81.73	69.07	62.50	71.10
MiSLAS <sup>[30]</sup>	86.37	82.11	81.99	83.49
LSKBB	<b>92.47</b>	<b>85.26</b>	<b>84.46</b>	<b>87.40</b>

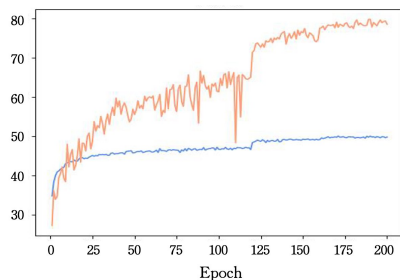
为了验证 LSKBB 模型的泛化性以及鲁棒性,在胸腔 X 射线透视图 ChestX-ray 数据集上进行实验。表 3 中的结果表明,LSKBB 模型在 ChestX-ray 的准确率优于 MiS-LAS 模型,不平衡率为 10,50 和 100 时,准确率分别提高了 6.10%,3.15% 和 2.47%;比 BBN 模型的平均准确率高出了 16.30%。

分析表 2 和表 3 的结果可知,LSKBB 模型的目标函数是动态变化的,其从侧重关注困难样本逐渐地过渡到关注每一个样本,使得 LSKBB 对样本数量不平衡的分析处理能力更强。特别是对于类别较少的数据集,相对于其他模型而言,LSKBB 模型的分析处理能力更好。

不平衡率越高,分类任务越难,选择高度不平衡率更具代表性。当选取不平衡率为 100 时,LSKBB 所对应的 Loss 图和准确率图如图 9 所示。从图 9(a)中可以看出,在 epoch 未达到 124 之前,Loss 迅速下降再到稳定下降;达到 124 后,Loss 又进入到迅速下降再到稳定下降的状态,最后逐渐收敛。从图 9(b)中可以看出准确率整体呈现上升趋势,epoch 为 0~124 的增长态势高于 epoch 为 124~200 的增长态势,这符合 Loss 图的走势。训练集的准确率稳步上升,且低于验证集的准确率,这是由于训练集是长尾分布,而验证集是平衡分布。实验结果表明模型在训练阶段并未倾向于预测多数类别,从而在验证集上表现良好。



(a)不平衡率为 100 的 LSKBB 的 Loss 图



(b)不平衡率为 100 的 LSKBB 的准确率图

图 9 Loss 图和准确率图

Fig. 9 Loss chart and accuracy chart

## 5.2 不同骨干网络的实验结果与分析

为确定最适合呈现长尾分布的数据集的骨干网络,分别使用经典的特征提取器 VGG16, ResNet-18 和 ResNet-32 进行实验对比。同时,对比改进的 ResNet-32, ResNet-32 + Biformer 和 ResNet-32 + SimAM。其中, ResNet-32 + Biformer 和 ResNet-32 + SimAM 是分别在 ResNet-32 的基础模块中引入 Biformer<sup>[31]</sup> 和 SimAM<sup>[32]</sup> 模块而形成的。在不平衡率为

10 的 BraeKHis 数据集进行实验,实验得出的分类准确率如表 4 所列(粗体表示最优结果)。

VGG<sup>[33]</sup> 网络最早由 Visual Geometry Group 提出。该模型通过使用许多  $3 \times 3$  尺寸的卷积核替换更大尺寸的卷积核来实现对 AlexNet 的增强,在 BraeKHis 的二值分类达到了约 83% 的准确率<sup>[34]</sup>, 多值分类准确率约为 76%。ResNet<sup>[35]</sup> 是由 He 等提出的卷积神经网络,在 2015 年的 ImageNet 竞赛中取得了巨大成功。ResNet-18 和 ResNet-32 是 ResNet 的变体,与 VGG16 相比,准确率提高了约为 11%。这表明残差网络更适合长尾分布的数据集分类任务。

表 4 不平衡率为 10 的不同骨干网络的准确率

Table 4 Accuracy of different backbone networks with an imbalance ratio of 10

骨干网络	准确率/%
VGG16	75.96
ResNet-18	85.58
ResNet-32	86.25
ResNet-32+Biformer	84.06
ResNet-32+SimAM	84.61
改进的 ResNet-32	<b>87.33</b>

从表 4 可以看出,残差网络 ResNet-32 的表现优于 VGG16 和 ResNet-18,而改进的 ResNet-32, ResNet-32 + Biformer 与 ResNet-32 + SimAM 都优于 ResNet-32,其中改进的 ResNet-32 准确率最高。

为了验证改进的 ResNet-32 的性能,在不平衡率为 10 时,展示 ResNet-32 的骨干网络的敏感性和精确率性能度量指标,如图 10 所示。从图中可以直观地发现,ResNet-32 和改进的 ResNet-32 敏感性和精确率表现较好,改进的 ResNet-32 在敏感性和精确率方面表现最为出色,达到了较为理想的效果。

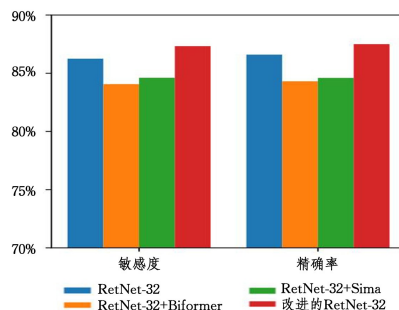


图 10 有关 ResNet-32 骨干网络的敏感性和精确率

Fig. 10 Sensitivity and precision about ResNet-32 backbone network

综上所述,LSKBB 模型选择了改进的 ResNet-32 为骨干网络。

## 5.3 不同损失函数下的实验结果与分析

为了探究不同的损失函数的性能,使用 BraeKHis 数据集选择损失函数 CE Loss, FOCAL Loss, DFLCSCE Loss 和 DFLCE Loss 进行实验。现将网络中其他部分进行固定,在 ResNet-32 和改进的 ResNet-32 下改变损失函数,以找到最优组合。通过多次实验,得到了不同的不平衡率下的准确率和精确率,结果如表 5 所列(粗体表示每列的最优结果)。

表5 不同的不平衡率下不同损失函数的准确率和精确率

Table 5 Accuracy and precision of different loss functions under different unbalance coefficients (%)

骨干网络	损失函数	不平衡率为 10		不平衡率为 50		不平衡率为 100	
		准确率	精确率	准确率	精确率	准确率	精确率
ResNet-32	交叉熵损失	86.25	86.60	82.19	82.41	78.59	79.55
ResNet-32	Focal Loss	85.31	85.66	79.61	79.82	76.25	76.69
ResNet-32	DFLCSCE	87.73	86.93	82.5	82.61	79.14	79.46
ResNet-32	DFLCE	87.81	88.00	82.27	82.51	77.34	78.06
改进的 ResNet-32	交叉熵损失	87.33	86.50	82.73	83.16	78.52	79.29
改进的 ResNet-32	Focal Loss	86.48	86.58	80.7	81.46	76.33	77.45
改进的 ResNet-32	DFLCSCE	87.03	87.19	82.89	83.27	78.67	79.41
改进的 ResNet-32	DFLCE	<b>87.97</b>	<b>88.08</b>	<b>83.44</b>	<b>83.77</b>	<b>79.84</b>	<b>80.06</b>

从实验结果来看,当骨干网络为 ResNet-32,损失函数为 DFLCSCE Loss 和 DFLCE Loss 时,准确率相对较高。这是因为 DFLCSCE Loss 和 DFLCE Loss 相较于交叉熵损失和 Focal Loss,模型优化的方向不是一成不变的,它们都是由重点关注一个方向逐渐过渡到关注另一个方向。在骨干网络为改进的 ResNet-32 下,无论不平衡率是 10,还是 50 或 100 时,损失函数为 DFLCE Loss 的准确率和精确率均最高,说明改进的 ResNet-32 和 DFLCE Loss 更合适。

5.2 节和 5.3 节的实验结果验证了本文提出的改进的 ResNet 以及新提出的两个损失函数的有效性,证明其可以用于长尾分布的图像分类任务,且能有效提升模型分类性能。

**结束语** 在不改变医学数据集原始长尾分布状态的情况下,本文提出了大选择性核双边网络模型,提高了分类能力。通过在不同的医学数据集和不同的不平衡率下进行实验,说明了 LSKBB 模型具有良好的泛化性能和鲁棒性。它能够更好地学习少数类样本,从而提高整体的分类性能。该模型给医生的临床诊断带来一定的技术支持。下一步将对网络结构进一步改进,考虑类内和类间的特征空间关系,克服来自各种组织病理图像的障碍,提高模型的识别准确率。

## 参考文献

- [1] BORSOS Z, MARINIER R, VINCENT D, et al. Audiolm: a language modeling approach to audio generation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31:2523-2533.
- [2] WU J, GAUR Y, CHEN Z, et al. On decoder-only architecture for speech-to-text and large language model integration[C]// 2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). IEEE, 2023:1-8.
- [3] WU J, JIANG Y, LIU Q, et al. General object foundation model for images and videos at scale[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024:3783-3795.
- [4] LI M, LI S, ZHANG X, et al. Univs: Unified and universal video segmentation with prompts as queries[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024:3227-3238.
- [5] SCHICK T, DWIVEDI-YU J, DESSÌ R, et al. Toolformer: Language models can teach themselves to use tools[J]. Advances in Neural Information Processing Systems, 2024, 36:68539-68551.
- [6] KASNECI E, SEBLER K, KÜCHEMANN S, et al. ChatGPT for good? On opportunities and challenges of large language models for education[J]. Learning and Individual Differences, 2023, 103:102274.
- [7] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115:211-252.
- [8] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[J]. Handbook of Systemic Autoimmune Diseases, 2009, 1(4):1-58.
- [9] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context [C]// Computer Vision - ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014:740-755.
- [10] DESUKY A S, HUSSAIN S. An improved hybrid approach for handling classimbalance problem[J]. Arabian Journal for Science and Engineering, 2021, 46:3853-3864.
- [11] HOYOS-OSORIO J, ALVAREZ-MEZA A, DAZA-SANTACOLOMA G, et al. Relevant information undersampling to support imbalanced data classification[J]. Neurocomputing, 2021, 436:136-146.
- [12] CHEN X, ZHOU Y, WU D, et al. Area: adaptive reweighting via effective area for long-tailed classification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023:19277-19287.
- [13] HUANG C, LI Y, LOY CC, et al. Learning deep representation for imbalanced classification [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:5375-5384.
- [14] KANG B, XIE S, ROHRBACH M, et al. Decoupling representation and classifier for long-tailed recognition[EB/OL]. [2024-04-24]. <https://arxiv.org/abs/1910.09217>.
- [15] ZHOU B, CUI Q, WEI X S, et al. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:9719-9728.
- [16] KANG B, LI Y, XIE S, et al. Exploring balanced feature spaces for representation learning [C]// International Conference on Learning Representations. 2021.
- [17] LI Y, HOU Q, ZHENG Z, et al. Large selective kernel network for remote sensing object detection [C]// Proceedings of the

- IEEE/CVF International Conference on Computer Vision, 2023: 16794-16805.
- [18] MIENYE I D, SUN Y. Performance analysis of cost-sensitive learning methods with application to imbalanced medical data [J]. *Informatics in Medicine Unlocked*, 2021, 25: 100690.
- [19] YEUNG M, SALA E, SCHÖNLIEB B, et al. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation [J]. *Computerized Medical Imaging and Graphics*, 2022, 95: 102026.
- [20] HUYNH T, NIBALI A, HE Z. Semi-supervised learning for medical image classification using imbalanced training data [J]. *Computer Methods and Programs in Biomedicine*, 2022, 216: 106628.
- [21] KIM H E, COSA-LINAN A, SANTHANAMN, et al. Transfer learning for medical image classification: a literature review [J]. *BMC Medical Imaging*, 2022, 22(1): 69.
- [22] GALDRAN A, CARNEIRO G, GONZÁLEZ BALLESTER M A. Balanced-mixup for highly imbalanced medical image classification [C] // *Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, Strasbourg, France, September 27-October 1, 2021, Proceedings, Part V 24*. Springer International Publishing, 2021: 323-333.
- [23] DING H, HUANG N, CUI X. Leveraging GANs data augmentation for imbalanced medical image classification [J]. *Applied Soft Computing*, 2024, 165: 112050.
- [24] ALQUDAH A, ALQUDAH A M. Sliding window based deep ensemble system for breast cancer classification [J]. *Journal of Medical Engineering & Technology*, 2021, 45(4): 313-323.
- [25] GUO M H, LU C Z, LIU Z N, et al. Visual attention network [J]. *Computational Visual Media*, 2023, 9(4): 733-752.
- [26] 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*, 2017: 2980-2988.
- [27] CUI Y, JIA M, LIN T Y, et al. Class-balanced loss based on effective number of samples [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 9268-9277.
- [28] ZHANG M, SU H, WEN J. Classification of flower image based on attention mechanism and multi-Loss attention network [J]. *Computer Communications*, 2021, 179: 307-317.
- [29] CAO K, WEI C, GAIDON A, et al. Learning imbalanced datasets with label-distribution-aware margin loss [J]. *Advances in Neural Information Processing Systems*, 2019, 32: 1-12.
- [30] ZHONG Z, CUI J, LIU S, et al. Improving calibration for long-tailed rerecognition [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 16489-16498.
- [31] ZHU L, WANG X, KE Z, et al. Biformer: Vision transformer with bi-level routing attention [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 10323-10333.
- [32] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C] // *International Conference on Machine Learning*. PMLR, 2021: 11863-11874.
- [33] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition. [EB/OL]. [2024-04-24]. <https://arxiv.org/abs/1409.1556>.
- [34] HAN Z, WEI B, ZHENG Y, et al. Breast cancer multi-classification from histopathological images with structured deep learning model [J]. *Scientific reports*, 2017, 7(1): 4172.
- [35] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.



**SUN Tanghui**, born in 1996, postgraduate, is a member of CCF (No. U7258G). Her main research interests include deep learning and computer vision.



**ZHAO Gang**, born in 1976, Ph.D, associate professor, master's supervisor. His main research interests include machine learning and so on.

(责任编辑:柯颖)