

基于高分特征保留网络的中医齿痕舌 图像分类方法

马玉明¹, 王振华^{1,2*}

¹伊犁师范大学网络安全与信息技术学院, 新疆 伊宁

²伊犁智能计算研究与应用重点实验室, 新疆 伊宁

收稿日期: 2026年3月23日; 录用日期: 2026年5月7日; 发布日期: 2026年5月20日

摘要

齿痕舌是反映脾虚、水湿内停等病理状态的重要中医客观体征。现有深度学习模型在提取此类局限于舌体边缘的微小形态特征时, 常因网络过深导致“空间平滑效应”, 且全局平均池化(GAP)易稀释微弱的异常信号, 引发临床漏诊。为此, 本文提出一种基于高分特征保留网络(HRFP-Net)的齿痕舌图像分类方法。该方法引入早期退出(Early-Exit)机制, 在轻量化骨干网络(ConvNeXt-Tiny)浅层进行截断, 以保留高分辨率的物理边缘形态; 同时采用全局最大池化(GMP)构建峰值信号检测器, 精准锁定并提取局部异常激活区域。在公开的Tooth-Marked数据集上的实验表明, HRFP-Net有效克服了边缘特征被掩盖的问题, 准确率达94.42%, 精确率高达97.06%, F_1 分数和AUC分别为93.40%和97.73%, 性能显著优于主流的CNN与Transformer模型。本文方法在降低网络冗余的同时实现了微小齿痕体征的高灵敏度捕捉, 为中医客观化辅助诊断提供了高效、鲁棒的新思路。

关键词

中医舌诊, 齿痕舌识别, 全局最大池化, 图像分类

Image Classification Method for Tooth-Marked Tongue in Traditional Chinese Medicine Based on High-Resolution Feature Preservation Network

Yuming Ma¹, Zhenhua Wang^{1,2*}

*通讯作者。

Abstract

The tooth-marked tongue is a crucial objective sign in Traditional Chinese Medicine (TCM) that reflects pathological states such as spleen deficiency and the internal retention of dampness and fluid. When extracting such subtle morphological features confined to the edge of the tongue, existing deep learning models often suffer from a “spatial smoothing effect” due to excessive network depth. Furthermore, Global Average Pooling (GAP) tends to dilute weak abnormal signals, leading to missed clinical diagnoses. To address these issues, this paper proposes an image classification method for tooth-marked tongues based on a High-Resolution Feature Preservation Network (HRFP-Net). This method introduces an Early-Exit mechanism that truncates the lightweight backbone network (ConvNeXt-Tiny) at a shallow layer to preserve high-resolution physical edge morphologies. Simultaneously, Global Max Pooling (GMP) is employed to construct a peak signal detector, which accurately locks onto and extracts locally abnormal activated regions. Experiments on the public Tooth-Marked dataset demonstrate that HRFP-Net effectively overcomes the problem of edge feature obfuscation. It achieves an accuracy of 94.42% and a remarkably high precision of 97.06%, with an F1-score and AUC of 93.40% and 97.73%, respectively. Its overall performance is significantly superior to mainstream CNN and Transformer models. While reducing network redundancy, the proposed method achieves high-sensitivity capture of subtle tooth-marked signs, providing an efficient and robust new approach for objective computer-aided diagnosis in TCM.

Keywords

Traditional Chinese Medicine (TCM) Tongue Diagnosis, Tooth-Marked Tongue Recognition, Global Max Pooling, Image Classification

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

舌诊是中医“望诊”的核心，齿痕舌作为反映脾虚、水湿内停等病理状态的重要体征，其客观化识别对推动中医现代化具有重要意义。传统舌诊依赖医师主观经验，标准化程度较低，因此引入计算机视觉技术开发辅助诊断系统已成为当前的研究热点[1]。

深度卷积神经网络(CNN)虽已广泛应用于舌象分类，但在齿痕舌识别任务中仍面临显著局限[2]。齿痕表现为舌体边缘局部的微小形态突变(即凹陷与凸起)，这一特性使得齿痕舌识别本质上属于细粒度视觉分类(Fine-Grained Visual Classification, FGVC)的范畴。主流深层网络(如 ResNet、DenseNet 等)过度依赖增加网络深度与多次下采样来获取全局语义，这极易引发“空间平滑效应”，导致细粒度的边缘突变在深层特征中被同化或抹杀[3][4]。此外，由于实际临床数据通常缺乏像素级的病灶标注，该任务同样面临弱监督目标定位(Weakly Supervised Object Localization, WSOL)的挑战。常规网络普遍采用全局平均池化(GAP)进行特征聚合，旨在获取全局上下文，但这在 WSOL 场景下容易将微弱的局部异常信号稀释于

广阔的正常舌体背景中, 造成临床漏诊[5]。尽管部分研究尝试引入复杂的注意力机制或多尺度特征融合来弥补局部感知能力的不足, 却带来了极高的计算开销与参数冗余[6] [7]。

为打破深层网络掩盖局部突变的瓶颈, 本文反其道而行之, 提出了一种基于高分特征保留网络(HRFP-Net)的齿痕舌图像分类方法。本文的主要贡献如下:

1) 提出高分特征保留网络与早期退出(Early-Exit)机制: 针对齿痕易被平滑的痛点, 在轻量化骨干网络(ConvNeXt-Tiny)浅层进行特征截断, 保留高分辨率的空间拓扑节点, 有效避免细微边缘特征在深度前向传播中丢失。

2) 确立基于全局最大池化(GMP)的局部突变捕捉策略: 以 GMP 作为空间峰值检测器替代传统的 GAP, 强制网络跨越正常背景, 精准锁定响应最强烈的局部形态突变, 在维持高精确率的同时大幅降低了漏诊率。

3) 实现优异的辅助诊断性能: 在公开的 Tooth-Marked 数据集上的实验表明, HRFP-Net 在参数量极低的情况下取得了 94.42% 的准确率和 97.06% 的极高精确率, 显著优于主流的深层 CNN 与视觉 Transformer 模型, 具备极高的临床实用价值。

2. 方法

2.1. HRFP-Net 总体架构

针对深层网络易过度平滑齿痕边缘局部形态的问题, 本文提出了高分特征保留网络(HRFP-Net), 旨在精确捕捉舌体边缘的物理凹凸特征。如图 1 所示, HRFP-Net 的总体架构主要由高分特征提取、空间聚合与极简分类三个阶段构成。

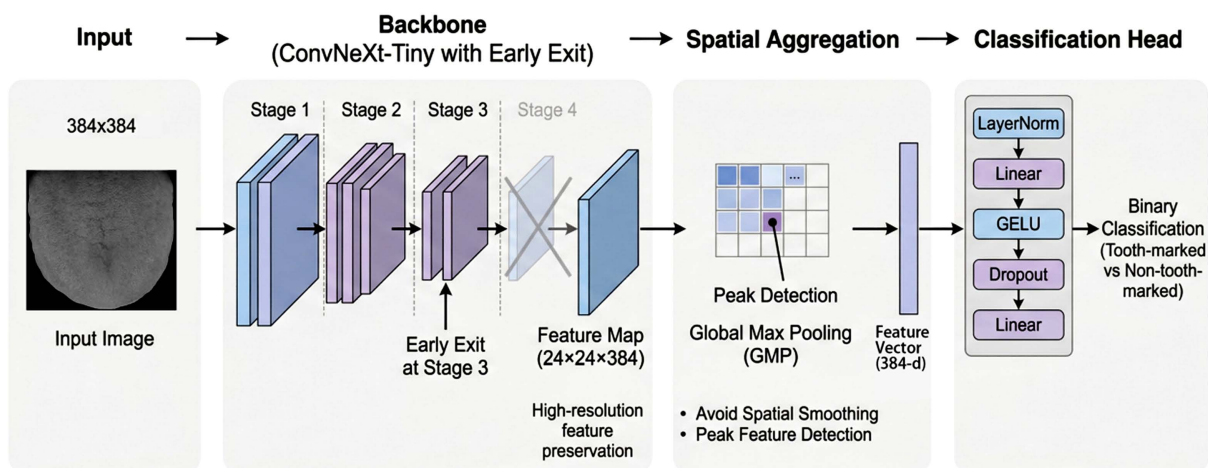


Figure 1. Overall architecture of the HRFP-Net

图 1. 高分特征保留网络总体架构图

1) 高分特征提取: 输入 384×384 的标准化图像, 采用轻量级 ConvNeXt-Tiny 作为骨干网络[8]。通过引入早期退出(Early-Exit)机制, 网络仅保留至 Stage 3 (前 6 个卷积组件), 输出分辨率为 24×24 、通道数为 384 的高分辨率特征图。

2) 空间聚合: 采用全局最大池化(GMP)对 24×24 的特征图进行降维。GMP 充当峰值检测器, 直接从 576 个精细空间节点中提取响应最强烈的异常信号, 将其展平为 384 维的特征向量。

3) 极简分类输出: 特征向量进入轻量化多层感知机(MLP)分类头。该模块依次包含层归一化、全连接层(降至 128 维)、GELU 激活函数、随机失活(Dropout 率为 0.4)及二分类全连接层, 最终输出预测概率。

HRFP-Net 通过“浅层截断”与“峰值聚合”，在大幅减少参数数量的同时，构建了专门面向边缘微小病变的高效识别链路。

2.2. 面向边缘特征保留的早期退出机制

在常规视觉任务中，加深网络常用于获取丰富的全局语义。然而，齿痕舌的病理特征并非大面积纹理改变，而是局限于舌体边缘的细粒度物理形态突变(凹陷与凸起)。传统深层网络(如 ConvNeXt-Tiny)在完整的前向传播中会经历多次空间下采样，其 Stage 4 会将 384×384 的输入急剧压缩至 12×12 。这一过程虽提升了语义表征，却引发了严重的“空间平滑效应”，导致微小且不规则的齿痕边缘被周围正常组织同化，大幅削弱了模型对该体征的敏感度。

为破解高层语义与细粒度形态保留之间的矛盾，HRFP-Net 创新性地引入了早期退出(Early-Exit)机制 [9]。具体而言，模型在特征提取阶段进行截断，舍弃导致分辨率剧降的 Stage 4，仅保留至 Stage 3 结束。该策略直接输出了分辨率为 24×24 的特征图，将空间节点数量由 144 个(12×12)大幅提升至 576 个。这不仅保障了区分舌体与背景所需的语义感受野，更在微观层面上保留了密集的空间拓扑信息，使得舌边缘的物理形态突变得以完整留存，为精准锁定病灶奠定了坚实的特征基础。

2.3. 空间聚合：基于全局最大池化的局部突变捕捉

在获取 24×24 的高分辨率特征图后，网络需通过空间聚合将三维张量压缩为一维向量。主流卷积网络通常采用全局平均池化(GAP)。GAP 在处理全局特征时表现优异，但在齿痕舌识别中存在显著局限：齿痕多散在分布于舌体边缘，面积占比极小。若采用 GAP 计算空间均值，齿痕区域产生的高频异常信号会被广阔的正常舌体背景“稀释”，导致模型对微弱病灶不够敏感，易引发临床漏诊。

为解决局部特征被掩盖的问题，HRFP-Net 摒弃了 GAP，转而引入全局最大池化(GMP) [10]。GMP 在 24×24 的空间节点中进行全局最大值筛选，本质上充当了“空间峰值信号检测器”。它不再关注特征的全局平均状态，而是强制网络越过平缓的正常背景，直接锁定并提取响应最强烈的局部异常激活区域。这一策略与“早期退出机制”形成完美互补： 24×24 的高分特征图提供了密集的细粒度节点，而 GMP 则负责从中精准挑出形态突变的“病灶”，从架构底层保障了模型在高精确率下的高召回率与临床鲁棒性。

3. 实验

3.1. 数据集

本实验采用 Kaggle 公开的 Tooth-Marked 舌象数据集¹。该数据集共包含 1250 张图像，其中齿痕舌 546 张，非齿痕舌 704 张。为保证实验的可复现性，数据集在固定随机种子下按 8:2 的比例划分为训练集与验证集，具体分布如表 1 所示。

Table 1. Sample distribution of the Tooth-Marked dataset

表 1. Tooth-Marked 数据集样本分布

类别	样本总数	训练集(80%)	验证集(20%)
有齿痕(Tooth-marked)	546	436	110
无齿痕(Non-tooth-marked)	704	563	141
总计	1250	999	251

¹<https://www.kaggle.com/datasets/clearhanhui/biyeshheji>

由于 HRFNet 对微小边缘形态高度敏感, 本实验设计了“纯净刚性变换”的数据增强策略。训练阶段, 图像统一缩放至 400×400 后随机裁剪为 384×384 。为避免形态污染与破坏边缘对齐, 实验去除了常规的旋转与仿射变换, 仅使用概率为 0.5 的随机水平与垂直翻转。在验证阶段, 图像直接调整为 384×384 , 刻意避免中心裁剪(CenterCrop), 以防止舌体边缘的齿痕特征被意外切除。最后, 所有图像均进行了标准归一化处理。

3.2. 实验环境与参数设置

本实验基于 PyTorch 深度学习框架, 在单张 NVIDIA A40 (48GB 显存) GPU 上进行。模型训练迭代次数(Epoch)设为 50, 批处理大小(Batch Size)为 32。优化器选用 AdamW, 权重衰减(Weight Decay)设为 1×10^{-4} , 并采用余弦退火算法(Cosine Annealing LR)动态调整学习率。

为兼顾预训练权重的保留与新网络层的收敛, 本研究采用差异化学习率策略: ConvNeXt-Tiny 的早期退出截断层(Stage 3 及之前)学习率设为 1×10^{-5} , 而自定义的空间聚合层与分类器学习率设为 5×10^{-4} 。损失函数采用标准的交叉熵损失(Cross-Entropy Loss), 并赋予正负样本[1.0, 1.15]的类别权重以缓解数据不平衡。

训练阶段引入了 Mixup 数据增强算法[11] ($\alpha = 0.2$)以提升模型泛化能力。此外, 本实验在网络末端分别设置了全局平均池化(GAP)与全局最大池化(GMP)进行对比, 以探究不同空间聚合策略对齿痕形态突变的捕捉能力; 并在验证阶段启用了测试时增强(TTA), 通过对原始图像与翻转图像预测结果求均值, 进一步提升模型分类鲁棒性。

3.3. 评价指标

为了客观且全面地评估模型在齿痕舌二分类任务中的表现, 本实验选取了图像分类领域标准的四个评价指标: 准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F_1 分数(F_1 -Score)。

各评价指标的具体计算公式如下:

1) 准确率(Accuracy): 模型预测正确的样本数占总样本数的比例, 用于直观衡量模型的整体分类能力。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

2) 精确率(Precision): 所有被预测为齿痕舌的样本中, 实际真正为齿痕舌样本的比例。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

3) 召回率(Recall): 所有实际为齿痕舌的样本中, 被模型正确识别出来的比例, 反映了模型对正样本的敏感度。

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

4) F_1 分数(F_1 -Score): 精确率和召回率的调和平均数, 能够综合反映模型在正负样本分布下的整体性能。

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

上述公式中, TP (True Positive)代表真阳性, 即被模型正确预测的齿痕舌样本数; TN (True Negative)代表真阴性, 即被正确预测的非齿痕舌样本数; FP (False Positive)代表假阳性, 即被误判为齿痕舌的非齿痕舌

样本数; FN (False Negative)代表假阴性, 即未被识别出的实际齿痕舌样本数。

3.4. 对比实验与结果分析

为了验证本文提出的高分特征保留网络(HRFP-Net)在齿痕舌识别任务上的有效性与优越性, 本节在公开的 Tooth-Marked 数据集上进行了全面的对比实验。本实验选取了 4 种在图像分类领域具有代表性的主流模型作为基线(Baseline), 包括传统的深度卷积神经网络(ResNet-101 [12]、Inception_V3 [13])、基于自注意力机制的视觉 Transformer(Swin-T) [14], 以及近年来提出的一种结合 CNN 与 Transformer 的先进混合网络[15]。各模型在测试集上的具体评估指标对比结果如表 2 所示。

Table 2. Performance comparison of different models on the Tooth-Marked dataset (%)

表 2. 不同模型在 Tooth-Marked 数据集上的性能对比(%)

方法/模型	精确率(Precision)	召回率(Recall)	F1 分数	准确率(Accuracy)
ResNet-101	89.12	87.97	88.54	88.78
Inception_V3	92.60	90.32	91.44	91.34
Swin-T	91.88	91.81	91.84	91.98
文献[15]方法	93.43	93.15	93.28	93.61
HRFP-Net (本文方法)	97.06	90.00	93.40	94.42

通过对表 2 实验数据的深入分析, 可以得出以下结论:

首先, 相比于具有极深网络层级的 ResNet-101 和 Inception_V3, 本文提出的 HRFP-Net 在各项综合指标上均取得了显著提升, 准确率达到 94.42%, F_1 分数达到 93.40%。这一结果有力地印证了本文的核心假设: 在处理齿痕这类“局部微小物理形态突变”时, 传统的深层卷积架构会因为多次池化和特征下采样, 导致关键的边缘凹凸特征被过度平滑甚至丢失。本文采用的早期退出(Early-Exit)机制成功保留了未经深度压缩的浅层高分辨率特征, 从而更敏锐地锁定了舌体边缘的病变区域。

其次, 与擅长捕捉全局上下文依赖的 Swin-T 模型以及文献[15]的混合网络相比, HRFP-Net 依然保持了明显的性能优势。尤其值得注意的是, 本文方法的精确率(Precision)高达 97.06%, 以绝对优势领先于所有对比模型。极高的精确率意味着模型在预测为“齿痕舌”时具有极高的可靠性, 有效降低了假阳性(误诊)的发生概率。在临床中医辅助诊断场景中, 这一点具有至关重要的实际应用价值, 能够有效避免误报给受检者带来不必要的心理负担。

综合来看, 本文的 HRFP-Net 摒弃了堆叠深层网络的常规思路, 以更加精简的结构和高分辨率特征保留策略, 不仅降低了模型的冗余度, 还实现了更优的分类性能, 充分验证了该算法在中医齿痕舌图像分类任务中的先进性与强鲁棒性。

3.5. 消融实验与分析

为了深入验证本文提出的 HRFP-Net 中核心机制(网络截断深度与空间聚合策略)的有效性, 本节设计了一系列动态变体的消融实验。实验主要考察了三个不同的网络截断退出点(Stage 2、Stage 3、Stage 4)以及两种特征池化方式(GAP、GMP)对模型最终分类性能的影响。消融实验的具体评估结果如表 3 所示。

通过对表 3 实验数据的横向与纵向对比, 可以得出以下两点核心结论:

- 1) 截断深度的权衡: Stage 3 为特征保留的最佳平衡点

网络深度直接决定特征粒度。传统基线模型(Stage 4)因多次下采样导致边缘微小突变被过度平滑, 存在性能瓶颈。反之, 极早期退出模型(Stage 2)虽有极高分辨率, 但因感受野与语义抽象能力不足, 无法区分正常生理起伏与真实病变, F_1 分数跌至 70.24%。相比之下, 本文的 Stage 3 截断点实现了物理凹凸与语义特征的黄金平衡, 在同用 GAP 时精确率即由 92.38% 跃升至 95.96%, 证明了保留浅层高分特征的必要性。

2) 空间聚合策略: GMP 机制更契合局部突变捕捉

采用 GAP 时模型召回率仅为 86.36%, 这是因为均值计算易将极其微弱的齿痕信号稀释于广阔的正常舌体背景中, 导致轻微病变漏诊。而本文的 Stage 3 + GMP 组合(方案 5), 本质上充当了“空间峰值检测器”, 强制网络越过平滑噪声直接锁定局部异常激活区域。该策略不仅取得了 93.40% 的最高 F_1 分数, 且在精确率高达 97.06% 的同时, 将敏感度(召回率)显著提升至 90.00%, 充分证明了 GMP 是处理此类局部形变任务的最优解。

Table 3. Ablation study on core mechanisms of HRFP-Net (%)

表 3. HRFP-Net 核心机制消融实验性能对比(%)

模型变体(Model Variant)	准确率(Acc)	精确率(Prec)	召回率(Rec)	F_1 分数	AUC
1) Baseline (Stage 4 + GAP)	91.63	92.38	88.18	90.23	96.00
2) Baseline (Stage 4 + GMP)	92.03	94.12	87.27	90.57	96.81
3) Early-Exit (Stage 2 + GAP)	75.70	75.79	65.45	70.24	80.37
4) HRFP-Net (Stage 3 + GAP)	92.43	95.96	86.36	90.91	97.49
5) HRFP-Net (Stage 3 + GMP) [Ours]	94.42	97.06	90.00	93.40	97.73

4. 结论

本文针对齿痕舌识别中边缘形态易被深层网络平滑、微弱异常信号易被均值池化稀释的问题, 提出了一种高分特征保留网络(HRFP-Net)。通过在轻量化骨干网络(ConvNeXt-Tiny)浅层引入早期退出(Early-Exit)机制并结合全局最大池化(GMP), 模型成功跨越平滑背景, 精准锁定了局部形态突变。在 Tooth-Marked 数据集上的实验表明, HRFP-Net 以极低的网络冗余取得了 94.42% 的准确率和高达 97.06% 的精确率。极高的精确率有效降低了临床辅助诊断的误诊风险, 证明了该算法在细粒度舌象体征识别上的卓越鲁棒性。

未来, 我们将进一步扩充多中心数据集以提升模型在复杂光照与设备下的泛化表现, 并探索将“浅层截断与峰值聚合”策略迁移至裂纹舌、点刺舌等其他微观舌象体征的识别中。此外, 得益于极低的计算开销, HRFP-Net 极具潜力部署于便携式医疗设备, 为中医基层及远程客观化问诊提供切实可行的智能化支持。

声明

本研究所使用的 Tooth-Marked 舌象数据集为来源于 Kaggle 平台的公开可用数据集。该数据集在公开发布与共享前, 已由数据提供方进行了严格的匿名化与脱敏处理, 完全移除了所有可能识别患者个人身份的敏感信息(如姓名、年龄、联系方式、具体就诊记录等)。本研究的算法设计与实验验证不涉及任何直接的人体干预试验, 亦无侵犯患者隐私的风险。研究全过程严格遵循相关的医学与人工智能科研伦理规范。

基金项目

伊犁智能计算研究与应用重点实验室 2023 年度一般科研项目: 基于多模态学习的新疆地区高发疾病智能中医诊断方法研究(2023WXYLHG003)。

参考文献

- [1] 彭素霞, 杨多, 钟俐芹, 梁昊. 中医舌诊智能化的研究进展[J]. 中医学, 2024, 13(7): 1590-1598
- [2] 卢运西, 李晓光, 张辉, 张菁, 卓力. 中医舌象分割技术研究进展: 方法、性能与展望[J]. 自动化学报, 2021, 47(5): 1005-1016
- [3] 徐雍钦, 杨晶东, 姜泉, 等. 基于多特征融合的中医症候舌象分类方法研究[J]. 智能计算机与应用, 2022, 12(7): 25-34.
- [4] 翟鹏博, 杨浩, 宋婷婷, 等. 融合注意力机制的多阶段舌象分类算法[J]. 计算机工程与设计, 2021, 42(6): 1606-1613.
- [5] Zhou, J., Li, S., Wang, X., Yang, Z., Hou, X., Lai, W., *et al.* (2022) Weakly Supervised Deep Learning for Tooth-Marked Tongue Recognition. *Frontiers in Physiology*, **13**, Article 847267. <https://doi.org/10.3389/fphys.2022.847267>
- [6] 吴欣, 徐红, 林卓胜, 等. 深度学习在舌象分类中的研究综述[J]. 计算机科学与探索, 2023, 17(2): 303-323.
- [7] Tang, W., Gao, Y., Liu, L., Xia, T., He, L., Zhang, S., *et al.* (2020) An Automatic Recognition of Tooth-Marked Tongue Based on Tongue Region Detection and Tongue Landmark Detection via Deep Learning. *IEEE Access*, **8**, 153470-153478. <https://doi.org/10.1109/access.2020.3017725>
- [8] Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T. and Xie, S. (2022) A ConvNet for the 2020s. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 11966-11976. <https://doi.org/10.1109/cvpr52688.2022.01167>
- [9] Teerapittayanon, S., McDanel, B. and Kung, H.T. (2016). BranchyNet: Fast Inference via Early Exiting from Deep Neural Networks. *2016 23rd International Conference on Pattern Recognition (ICPR)*, Cancun, 4-8 December 2016, 2464-2469. <https://doi.org/10.1109/icpr.2016.7900006>
- [10] Oquab, M., Bottou, L., Laptev, I. and Sivic, J. (2015) Is Object Localization for Free? Weakly-Supervised Learning with Convolutional Neural Networks. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 685-694. <https://doi.org/10.1109/cvpr.2015.7298668>
- [11] Zhang, H., Cisse, M., Dauphin, Y.N., *et al.* (2017) Mixup: Beyond Empirical Risk Minimization. <https://doi.org/10.48550/arXiv.1710.09412>
- [12] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/cvpr.2016.90>
- [13] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016) Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 2818-2826. <https://doi.org/10.1109/cvpr.2016.308>
- [14] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 9992-10002. <https://doi.org/10.1109/iccv48922.2021.00986>
- [15] 杨炳乾, 冯秀芳, 董云云, 等. 结合 CNN 和 Transformer 病变信号引导的蜂窝肺 CT 图像识别[J]. 激光与光电子学进展, 2024, 61(14): 457-466.