

基于多模型融合与迁移学习策略的智能舌象识别研究

张茹

陕西国际商贸学院, 中国·陕西 咸阳 712046

摘要: 针对中医舌诊智能化中舌象分类精度不足、数据稀缺的问题, 本研究提出一种多模型融合与迁移学习的舌苔识别网络 Tri-VoteNet。该网络集成 ZhiResNet50、ZhiCNN 与 ZhiEfficientNet-B0 三种架构, 采用分层冻结迁移学习策略, 兼顾特征提取能力与训练效率。基于开源 Tongue+coating+classification 数据集, 经数据校验、标准化、增强及类别均衡化预处理后开展实验。结果表明, Tri-VoteNet 可有效捕捉舌苔颜色、厚薄等细粒度特征, 整体分类性能显著优于单模型, 为中医舌诊客观化与智能化提供了有效技术支持。

关键词: 舌象识别; 迁移学习; 集成模型; 残差网络

Research on intelligent tongue image recognition based on multi-model fusion and transfer learning strategy

Zhang Ru

Shaanxi International Business College, China Shaanxi Xianyang 712046

Abstract: Aiming at the problem of insufficient classification accuracy and data scarcity in the intelligent tongue diagnosis of traditional Chinese medicine, this study proposes a tongue coating recognition network Tri-VoteNet based on multi-model fusion and transfer learning. The network integrates three architectures of ZhiResNet50, ZhiCNN and ZhiEfficientNet-B0, and adopts hierarchical frozen transfer learning strategy, taking into account feature extraction ability and training efficiency. Based on the open source Tongue + coating + classification data set, the experiment was carried out after data verification, standardization, enhancement and category equalization preprocessing. The results show that Tri-VoteNet can effectively capture fine-grained features such as color and thickness of tongue coating, and the overall classification performance is significantly better than that of single model, which provides effective technical support for the objectification and intelligence of tongue diagnosis in traditional Chinese medicine.

Keywords: Tongue image recognition; Transfer learning; Integrated model; Residual network

0 引言

舌诊作为中医“望闻问切”之首, 凭借无创、便捷、实时的优势, 在基层医疗与健康管理中价值显著^[1]。然而, 传统人工舌诊依赖主观经验, 存在标准不一、重复性差等瓶颈, 制约了其客观化与规模化应用。在“中医药现代化”战略背景下, 借助深度学习技术突破舌诊标准化难题, 不仅能提升中医诊疗的精准性与可及性, 更可为传统医学的传承创新与国际传播提供有力支撑。

1 研究现状

舌诊研究覆盖舌体分割、表型分类等全流程。舌体分割是前置关键任务, 核心挑战是背景干扰、反光噪声与边缘模糊。早期 U-Net 模型虽能实现基本分离, 但在复杂场景下精度仍不足。彭辰^[2]等提出了基于卷积神经网络与注意力机制的舌体分割与分类模型 (CRANet), 并设计了卷积残差模块。刘嵘激^[3]等人提出 DeepLabv3-MAC 模型分割

自然环境下采集的舌象。该模型主要是利用 MobileNetv2 网络替换 DeepLabv3+ 模型的主干网络, 另外采用非对称卷积模块增强卷积神经网络的卷积核骨架, 提高卷积信息的利用率; 同时引入 CBAM 注意力机制, 提升模型分割精度。目前舌体分割已较好解决, 舌象分类成为研究重点, 主要面临标注数据稀缺、分类效果欠佳的问题。迁移学习能从海量自然图像中学习通用特征, 仅需少量医学数据适配高级任务, 因此在舌象分类中广受欢迎。目前医学场景中的迁移学习方法主要集中在基于实例的方法和基于模型的方法两大类。基于实例的方法旨在使用特定的权重调整策略, 挑选源域中的实例补充到目标域中。针对肺癌图像, Wang 等^[4]采用预训练的残差神经网络作为迁移学习模型, 在此基础上增加全连接层以增强模型的迁移能力, 并在山东省医院收集的肺癌数据集上进行微调。Demoura 等^[5]基于迁移学习使用不同的特征选择策略, 从目标 OCT 图像

中选择最相关的深度特征子集,实现糖尿病性黄斑水肿患者的分类。基于网络的深度迁移学习主要将经过充分训练的网络重用到目标域。它将源域神经网络的前端层视为多功能的特征提取器,其结构和参数将成为新网络的一部分,但有时会在再训练中冻结。曾丽莉等^[6]人提出基于多源数据的跨任务迁移学习舌象诊断方法 M2T-Net,包括两个阶段:在多源数据的预训练阶段,获取不同任务下的高质图像编码器在跨任务迁移阶段,结合交叉注意力机制,融合不同任务的特征表示,用于疾病分类,并以冠心病和冠心病伴随糖尿病患者分类案例加以验证。Khan 等^[7]使用由自然图像组成的大型基准数据集预训练的 VGG,通过分层调整对网络进行微调,利用一组预定义的层在 MRI 图像上进行训练,以预测阿尔兹海默症。Zhang 等^[8]提出了一种基于实例和基于网络相结合的两阶段深度迁移策略,以解决异常舌象检测中数据不足的问题。第一阶段获取不平衡数据集的可用实例,学习公共基本特征并从更相似的领域转移知识,第二阶段转移结构特征。李俊贤等^[9]人提出一种更加精准易用的舌象点刺识别方法,对传统 CNN 进行了精简设计,并引入全迁移学习方法,采用一种递阶式方法,实现舌象点刺的精准识别,从而提高识别效率。

总之,目前舌象识别领域存在分类精度不足,网络深度训练成本高等问题,因此本研究提出:

(1) 构建三架构加权投票舌象识别网络 Tri-Architecture Weighted Voting Tongue Image Recognition Network(简称:Tri-VoteNet):采用加权投票集成策略集成了三种不同架构的网络——ZhiResNet50、ZhiCNN 和 ZhiEfficientNet-B0,每种网络都具有独特的特征提取能力。(2) 采用了精细化的分层冻结策略,将网络分为冻结层和可训练层。前三层网络(Conv1、MaxPool、Layer1)采用冻结策略,利用预训练模型的通用特征提取能力;后三层网络(Layer2、Layer3、Layer4)设置为可训练,针对舌象识别任务进行微调。这种策略既避免了过拟合,又能充分学习舌象的特定特征。

2 Tri-VoteNet 模型

2.1 ResNet50 网络

ResNet50 网络^[10]基于残差学习机制,通过堆叠 50 个卷积层并引入“跳跃连接”,允许网络在深层训练时将浅层舌体特征(如边缘、颜色)直接传递到深层,从而有效避免梯度消失问题,使网络能够提取舌苔厚薄、腐腻、齿痕、裂纹等复杂高阶特征。其优势在于,残差结构解决了深层网络性能退化问题,能在保留舌象全局形态(如胖瘦、肿胀)的同时精准捕捉局部细节(如瘀斑、点刺),且恒

等映射设计降低了训练难度,即使在中型舌象数据集上也能快速收敛并获得高识别精度,尤其适合中医舌诊中多类别、细粒度特征的自动化分析。

2.2 CNN 网络

CNN 是由 Hubel 和 Wiesel^[11]提出感受野的概念,Fukushima K^[12]在其基础上提出 CNN 模型,是一个集特征自动提取、特征选择、分类为一体的网络模型,其结构如图 1 所示。CNN 由于其参数少,训练快,效果好,易迁移的特点被广泛应用到识别与检测分类模型中。卷积层就是将原始特征向量构成的特征图与卷积核进行点积求和的过程,同时不断移动卷积核的位置,直到遍历完成整个特征图。池化层的目标是选出最具有代表性的特征,减少数据的空间大小,一定程度上控制过拟合。最后通过全连接层实现分类,其核心操作就是矩阵向量乘积,本研究将特征向量输入至全连接层后,通过 softmax 激活函数将其映射到[0,1]范围内的概率值,进而实现舌象舌质分类

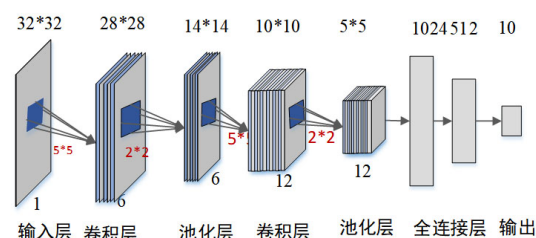


图1 CNN网络模型结构

2.3 EfficientNet-B0 网络

EfficientNet-B0 网络^[13]是一种基于复合缩放方法的高效卷积神经网络,它通过协同调整网络的深度、宽度和输入分辨率,取代传统单一维度的缩放方式,并采用移动翻转瓶颈卷积(MBConv)结合 Squeeze-and-Excitation(SE)注意力机制,在仅 5.3M 参数量下实现高精度识别。将其应用于舌象识别,优势显著:一方面,模型轻量化设计使其易于部署在移动端或边缘设备,适合基层中医诊所的实时辅助诊断;另一方面,复合缩放带来的多尺度感知能力可同时捕捉舌苔细微纹理(如点刺、裂纹)与舌体整体形态(如胖瘦、歪斜),而 SE 注意力模块能自动聚焦舌质颜色、苔质特征等关键区域并抑制背景干扰,从而在保证计算效率的前提下提升舌象分类的准确性与鲁棒性。

2.4 Tri-VoteNet 模型构建

本研究针对舌象数据集识别精度欠佳问题构建了三架构加权投票舌象识别网络 Tri-VoteNet:采用加权投票集成策略集成了三种不同架构的网络——ZhiResNet50、ZhiCNN 和 ZhiEfficientNet-B0。ZhiResNet50 的残差结构能够有效解决梯度消失问题,适合提取深层特征;

ZhiCNN 网络引入注意力机制能够自适应地聚焦于舌象的关键区域; EfficientNet-B0 网络的 MBConv 结构在保证精度的同时具有更好的计算效率。本舌象识别多模型集成系统采用三层架构设计, 包括数据输入层、多模型并行处理层和集成决策层, 系统输入为 $224 \times 224 \times 3$ 的 RGB 舌象图像, 经过三个子网络并行处理后, 通过加权投票机制生成最终分类结果。

数据输入层负责图像的标准化预处理, 包括颜色校正、尺寸归一化、噪声消除等操作。系统采用标准化的图像采集协议, 确保输入图像的质量和一致性; 多模型并行处理层是系统的核心, 由三个子网络组成: ZhiResNet50 基于 ResNet50 架构, 采用分层冻结策略; ZhiCNN 为自定义卷积神经网络, 集成了注意力机制; ZhiEfficientNet-B0 采用轻量化的 MBConv 结构。三个网络并行提取特征, 各自产生分类概率分布; 集成决策层采用加权投票策略, 将三个子网络的预测结果进行融合, 测试多种权重配置结果, 选取最优分类性能的权重配置作为默认权重配置。具体架构见图 2。

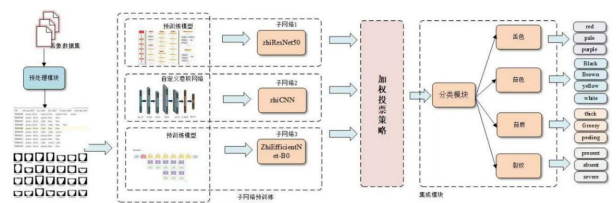


图2 舌象识别预测模型架构图

3 实验验证及分析

3.1 数据集

本研究采用开源数据集 Tongue+coating+classification^[14]作为实验数据集该数据集是面向中医舌诊的舌苔分类专用数据集, 聚焦舌苔颜色、厚薄、质地等核心特征, 用于深度学习模型训练与验证。数据集含 676 张高分辨率图像 (512x512 像素), 其中训练集共 414 张图片, 包含: 黑苔 66 张、棕苔 24 张、地图舌 12 张、紫苔 62 张、红舌黄厚腻苔 108 张、红舌厚腻苔 92 张、白舌厚腻苔 50 张; 验证集共 122 张图片; 测试集共 140 张图片, 适合舌苔分类任务, 为中医舌诊智能化提供数据支撑。

3.2 数据预处理

针对现有数据集, 本研究完成了以下数据预处理工作:

Step 1: 数据整理与校验。核对图像与标签对应关系, 剔除模糊、破损、标注错误的无效样本, 确保训练集、验证集、测试集划分规范。

Step 2: 标签编码。将黑苔、棕苔、地图舌、紫苔、红舌黄厚腻苔、红舌厚腻苔、白舌厚腻苔共 7 类标签转换

为数字编码, 适配模型输入。

Step 3: 图像标准化。统一图像尺寸为 224×224 , 将像素值归一化至 0-1 区间, 消除分辨率与数值差异。

Step 4: 数据增强。对训练集进行水平翻转、小角度旋转、轻微亮度/对比度调整, 增强数据多样性^[15], 缓解小样本过拟合。

Step 5: 数据清洗与增强校验。检查类别分布, 对样本量少的类别 (地图舌、棕苔) 做轻度过采样, 保证模型训练均衡稳定。

3.3 评估指标

本模型性能评价指标包括混淆矩阵、准确率^[16]。混淆矩阵表示如表 1 所示, 其中 TP 表示真阳率, FN 表示假阴率, FP 表示假阳率, TN 表示真阴率。准确率 (Accuracy) 是分类任务最常用的评价指标, 指模型预测正确的样本数占总样本数的比例, 直观反映模型整体分类效果。数值越接近 1, 代表模型对舌苔类别的识别能力越强。其计算方法见式 (1)。

表1 混淆矩阵

预测结果 \ 真实标签	正常	异常
正常	TP	FN
异常	FP	TN

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

(1)

3.4 实验效果

为验证模型有效性, 本研究采用 Tongue+coating+classification 数据集进行的仿真验证, 具体的仿真环境配置如下表 2 所示。

表 2 Tri-VoteNet模型仿真环境配置表

名称	配置
操作系统	Windows10(x64)
处理器	Intel(R) Core(TM) i7-14650HX (2.20 GHz)
机带 RAM	16.0 GB (15.8 GB 可用)
图形卡	NVIDIA GeForce RTX 4060 Laptop GPU (8 GB)
存储	954 GB

将数据预处理后的数据集输入至构建好的模型中进行测试, 舌苔分类混淆矩阵如图 3 所示, 训练过程中的损失函数及准确率如图 2 所示。根据混淆矩阵通过式 (1) 计算得到的准确率指标的测试结果如表 3 所示。由表 3 可知, Tri-VoteNet 模型对于舌质识别的性能明显优于子网络, 表明该集成模型具有明显优势。

表 3 Tri-VoteNet模型仿真测试结果

网络	ACC(%)
zhi_resnet	96.3
zhi_cnn	93.6
zhi_effnet	92.0
Tri-VoteNet	97.2

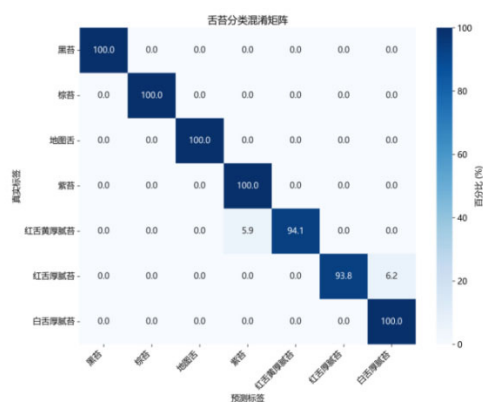


图3 舌苔分类混淆矩阵

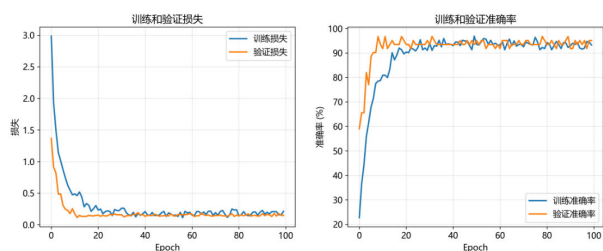


图4 训练和验证过程损失和准确率曲线

4 结语

本研究聚焦中医舌诊智能化需求,构建了基于多模型融合与迁移学习的舌苔分类网络 Tri-VoteNet,实现了七类典型舌苔的精准识别。通过集成残差网络、自定义卷积网络与轻量化高效网络,结合分层冻结迁移学习策略,有效解决了小样本场景下模型泛化能力弱、特征提取不充分的难题。针对 Tongue+coating+classification 数据集,完成数据清洗、标准化、增强与类别均衡化处理,保障了训练数据质量。实验结果显示, Tri-VoteNet 在准确率指标上明显优于各子网络,验证了多模型加权投票融合策略的有效性。该模型兼顾识别精度与部署效率,可为基层中医临床辅助诊断提供技术参考,也为中医药智能化研究提供了新的思路与方法。

参考文献:

- [1] Yu Y. Advances in application of artificial intelligence in medical image analysis. Chinese Journal of Medical Imaging Technology. 2019. 35(12): 1808-1812.
- [2] 彭辰, 李文举, 钱志勤等. CRANet: 舌象图像分割与分类网络[J]. 华东理工大学学报(自然科学版), 2025, 51(04): 564-571.
- [3] 刘嵘激, 辛国江, 张杨等. 基于改进的 DeepLabv3+ 模型的自然环境下舌象分割方法[J]. 计算机与现代化, 2025 (10): 32-36+43.
- [4] Wang S, Dong L, Wang X, et al. Classification of

pathological types of lung cancer from CT images by deep residual neural networks with transfer learning strategy. Open Medicine. 2020. 15(1): 190-197.

[5] De Moura J, Novo J, Ortega M. Deep feature analysis in a transfer learning-based approach for the automatic identification of diabetic macular edema. Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN 2019). 2019.

[6] 曾丽莉, 夏佳楠, 李韶雯等. M2T-Net: 基于多源数据的跨任务迁移学习舌象诊断方法[J]. 计算机科学, 2025, 52(09): 47-53.

[7] KHAN N M, ABRAHAM N, HON M. Transfer learning with intelligent training data selection for prediction of Alzheimer's disease[J]. IEEE Access, 2019, 7: 72726-72735.

[8] ZHANG X, CHEN Z, GAO J, et al. A two-stage deep transfer learning model and its application for medical image processing in Traditional Chinese Medicine[J]. Knowledge-based Systems, 2022, 239: 108060.

[9] 李俊贤, 朱明峰. 融合轻卷积网络 and 全迁移学习的递阶式舌象点刺识别方法[J]. 信息技术与信息化, 2025(11): 16-19.

[10] 张强, 黄鼎, 周思慧. 基于改进 ResNet-50 的起落架多分类识别研究[J/OL]. 现代雷达, 1-11[2026-04-30].

[11] HUBEL DH, WIESEL TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 1962; 160(1): 106-54.

[12] Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern. 1980; 36(4): 193-202.

[13] 李苏, 何巍. 基于改进的 EfficientNet-B0 网络的失能老人表情识别研究[J]. 四川师范大学学报(自然科学版), 2026, 49(1): 97-105.

[14] 舌苔数据集 - 分类完整 512x512 RGB 图像: 高质量 512x512 RGB 舌苔分类数据集, 助力中医智能诊断研究 - AtomGit | GitCode.

[15] 王浩威. 基于多任务模型的舌象特征辨识研究[D]. 大连交通大学, 2025.

[16] 池振飞. 基于深度学习舌象特征的慢性萎缩性胃炎诊断模型构建及相关研究[D]. 辽宁中医药大学, 2025.

作者简介: 张茹(1998-), 女, 汉族, 陕西咸阳人, 硕士, 助教, 研究方向: 图像识别与处理、深度学习。