

文章编号: 1006-3080(2025)04-0564-08

DOI: 10.14135/j.cnki.1006-3080.20240925001

CRANet: 舌象图像分割与分类网络

彭辰¹, 李文举², 钱志勤¹, 徐蕾³, 罗琪¹, 余依婕¹

(1. 华东理工大学机械与动力工程学院, 上海 200237; 2. 郑州市仁济医院南院内二科, 郑州 450004;
3. 上海市嘉定区安亭医院后勤保障部, 上海 201800)

摘要: 中医诊断中, 舌诊是一种重要的诊断方法。然而, 舌象分析的准确性和效率受主观因素影响较大。本文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(CR Block)。该模型能够自动对舌体进行分割与特征识别, 并据此对舌象进行分类, 提高舌诊的客观性和准确性。在分割任务中, 本文模型的平均准确率为 99.43%, 交并比(IoU)为 97.50%, Dice 系数为 98.73%; 在分类任务中, 平均准确率为 88.91%, 精确率为 86.37%, 召回率为 85.32%, F1 分数为 85.84%; 并验证了基于卷积神经网络和注意力机制的舌象图像分割与分类方法在中医舌诊中的应用潜力。

关键词: 卷积神经网络; 注意力机制; 图像分割; 图像分类; 中医舌诊

中图分类号: TP301.6

文献标志码: A

舌诊^[1]是中医四诊(望、闻、问、切)中的一种重要诊断方法, 通过观察舌象(包括舌苔、舌质、舌体形态等), 医生可以获取患者的健康状况和疾病信息。舌诊的原理基于中医理论, 认为舌体的变化可以反映出内脏的功能状态及整体的健康状况^[2]。然而, 传统舌诊依赖于医生的经验和主观判断, 这导致了诊断结果的依赖性和不一致性^[3]。此外, 人工分析舌象的过程既费时又繁琐, 难以满足现代医学快速、准确诊断的需求。因此, 对舌象图像的分割与分类算法进行研究具有重要意义^[4]。

卷积神经网络(CNN)^[5]作为深度学习领域的重要组成部分, 近年来在图像处理任务中取得了显著进展。卷积神经网络的独特结构和强大的特征提取能力, 使得它在图像分割和识别领域得到了广泛应用和高度评价^[6-7]。在深度学习领域, 注意力机制能够使模型对输入数据的特定部分赋予不同的重要性权重, 从而更好地进行决策。图像分割是计算机视觉中的一个基本任务, 其目标是将图像划分为多个具有语义意义的区域, 从而为后续的图像分析和处

理奠定基础。传统的图像分割方法通常依赖于手工设计的特征和复杂的预处理步骤, 这些方法在面对复杂场景和多样化的图像时往往表现不佳。卷积神经网络通过自动学习图像的多层次特征, 显著提升了分割的精度和鲁棒性。典型的卷积神经网络结构如 U-Net^[8]在医学图像分割等领域取得了突出成果, 展示了其强大的应用潜力。目前, 有部分研究将基于卷积神经网络的 U-Net 模型运用于舌象图像分割任务中。陈洪苇等^[9]将 Morphsnakes 算法与传统的 U-Net 模型进行融合, 通过 Morphsnakes 算法对嘴部图像进行预分割, 再通过 U-Net 模型对舌体进行精确分割。注意力机制在自然语言处理和图像处理领域已经取得了显著的成功, 其基本思想是让模型关注输入数据中最重要的部分, 提升模型对细节的捕捉能力。在舌象分割和分类任务中, 注意力机制的引入可以帮助模型聚焦于舌体的关键特征, 忽略背景或无关的区域, 从而提高分割和分类的精度。任思羽等^[10]设计了一种基于度量学习的多分支舌象识别网络。将整个神经网络分为两个部分, 前半部分共

收稿日期: 2024-09-25

基金项目: 上海市“科技创新行动计划”自然科学基金面上项目(23ZR1416200)

作者简介: 彭辰(2001—), 女, 上海人, 硕士生, 主要研究方向为深度学习。E-mail: ceciliapengchen@163.com

通信联系人: 钱志勤, E-mail: qianzhiqin@ecust.edu.cn

引用本文: 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(4): 564-571.

Citation: PENG Chen, LI Wenju, QIAN Zhiqin, et al. CRANet: Tongue Image Segmentation and Classification Network[J]. Journal of East China University of Science and Technology, 2025, 51(4): 564-571.

享权重, 获得精确的特征图, 后半部分具有 4 个舌象识别辅助分支, 构建多标签残差映射以提高准确率。

目前, 针对舌象图像分割与分类的研究较少, 运用于舌象图像分割与分类的模型仍具有抗环境干扰能力差、效率及准确率较低的问题^[1]。本文提出了一种基于卷积神经网络和注意力机制的舌象图像分割与分类方法, 旨在提高分割与分类的精度和效率, 为智能中医舌诊提供可靠的技术基础。

1 神经网络设计

1.1 整体网络结构

本文所设计的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)主要包括两个部分: 舌象分割部分以及舌象分类部分。舌象分割部分能够自动识别并分割出舌象中的舌体部分, 排除背景、面部或其他干扰区域。舌象的准确分割能够确保后续分类任务只基于准确的舌体区域进行。舌象分类部分能够基于分割出的舌体, 根据外观将其划分为 4 类: 薄白苔、厚白苔、黄苔以及黑苔。该神经网络可作为医生的辅助工具, 帮助分析舌象图像, 提供健康状况的初步预测。同时, 所设计的神经网络结合手机或摄像头等设备, 可以实现远程舌象图像采集和自动分析, 为家庭健康监测和远程诊疗提供技术支持。CRANet 整体网络架构如图 1 所示。

神经网络的分割部分由输入层、卷积残差模块、压缩激励模块、注意力门模块以及输出层组成。分类部分由输入层、卷积残差模块、多头注意力模块、注意力门模块以及输出层组成。其中, 分割部分所输出的语义分割图经处理后作为分类部分的输

入, 分类部分则输出舌象的类别以及所属该类别的概率值。

在对整体神经网络架构进行设计的同时设计了卷积残差模块(CR Block), 并引入了压缩激励模块(SE Block)、多头注意力模块(MHSA Block)以及注意力门模块(AG Block)。

1.2 卷积残差模块

本文使用了卷积残差模块, 通过引入残差连接以解决深度网络中的梯度消失问题, 并通过引入随机丢弃以解决网络的过拟合问题, 其结构如图 2(a)所示。从图中可以看出, 输入经过 5 层卷积层, 其中每层卷积层的卷积核尺寸为 3×3 (即 $\text{kernel_size}=3$), 步幅为 2 (即 $\text{stride}=2$), 填充为 1 (即 $\text{padding}=1$)。第 5 层卷积的输出再经 5 层反卷积得到最终输出, 其中反卷积层的卷积核尺寸为 3×3 (即 $\text{kernel_size}=3$), 步幅为 2 (即 $\text{stride}=2$), 填充为 1 (即 $\text{padding}=1$)。模块中具有随机失活层, 随机失活比率为 0.4。本文所设计的卷积残差模块具有通用性和可扩展性, 能够无缝集成到现有神经网络架构中, 提升网络的性能和稳定性。

该模块具体优势如下: (1)通过残差连接, 有效缓解梯度消失问题, 促进网络深层次特征的学习; (2)允许网络在学习过程中更加有效地更新权重, 提高网络的训练效率; (3)在有残差连接的网络中, 信息的多路径流动提高了网络的鲁棒性, 降低其对噪声和输入扰动的敏感性。这些优势使得编码器在舌象图像分割任务中表现出色, 能够提供丰富且稳定的特征图, 从而实现高精度的分割结果。

1.3 多头注意力模块

多头注意力模块^[12]是自注意力机制的扩展, 通

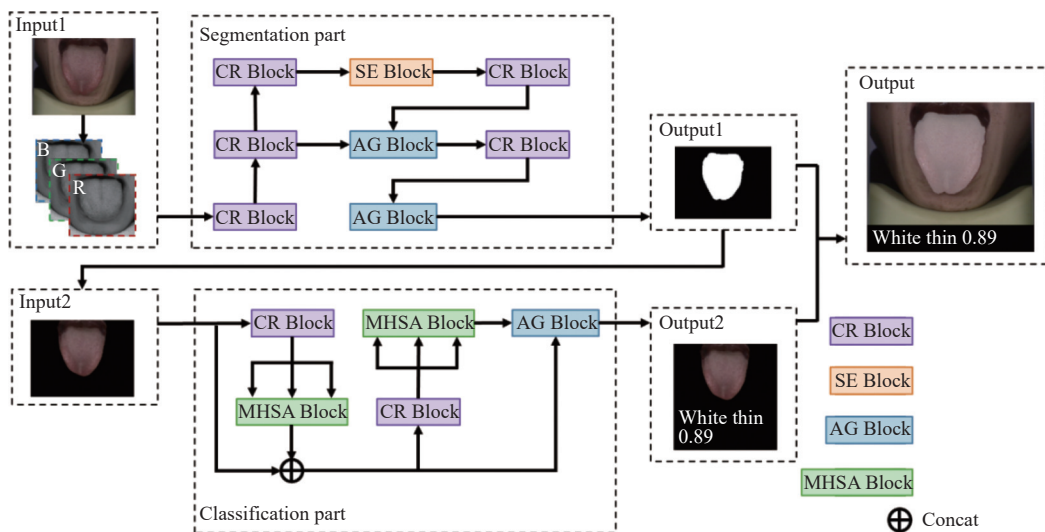


图 1 CRANet 整体结构

Fig. 1 Architecture of CRANet

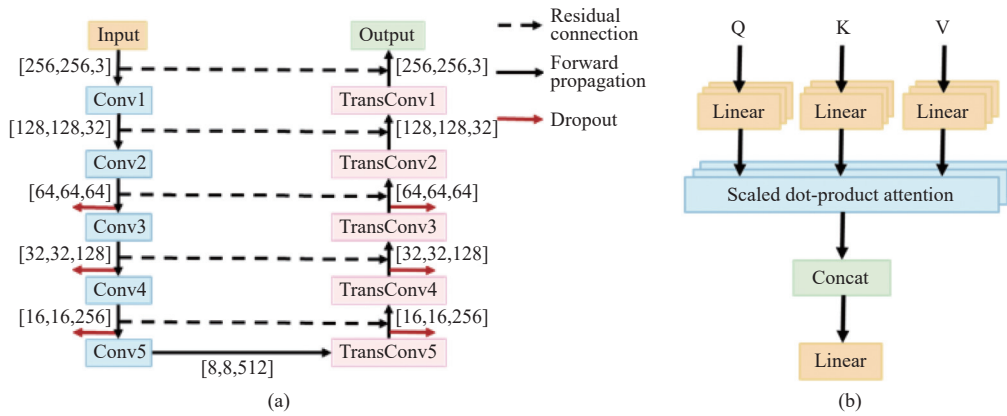


图 2 卷积残差模块(a)及多头注意力模块结构(b)

Fig. 2 Architecture of CR block (a) and MHSA block (b)

过多个独立的注意力头并行处理输入特征,从不同角度捕捉全局上下文信息,其结构如图 2(b)所示。

1.4 压缩激励模块

模型中引入的压缩激励模块^[13]是一种轻量化且有效的注意力机制,其结构如图 3 所示。

压缩激励模块能够根据输入的特征图动态调整每个通道的重要性,使得模型更关注有用的特征,提升模型的表达能力。同时,该模块能够与各类神经网络架构相结合,显著提升神经网络性能。

1.5 注意力门模块

模型在解码部分引入了注意力门模块,用于突出关键特征并抑制无关信息,从而提升分割精度。注意力门能够自动识别图像中的重要区域,并在解码过程中强化其特征表达。结合残差连接,注意力门还能高效融合多层次特征,实现细节与语义信息的互补,增强模型在复杂场景下的分割与分类能力。

舌象图像中不同区域的特征可能存在长距离依赖性,而传统的卷积操作通常只能捕捉局部特征。多头注意力模块能够有效捕捉输入特征中的全局信息,通过并行的多头机制,使得网络能够同时从多个

角度捕捉舌象不同区域之间的相互关系,提升分类精度。

2 数据集与研究方法

2.1 数据集选用

2.1.1 舌象分割数据集 对于中医舌象图像研究,由于临床数据的特殊性以及对患者信息的保密性,目前可供研究的数据集十分有限^[14]。本文选取公开数据集 BioHit 图像数据集,如图 4 所示。BioHit 数据集由 300 例舌体图像组成,涵盖了不同年龄、性别和健康状况的个体。为了支持深度学习模型的训练,BioHit 数据集提供了详细的标注信息。

2.1.2 舌象分类数据集 在舌象分类任务中,本文所采用的数据集为标准环境下所拍摄的高清舌象数据。由专业医师将舌象数据划分为黑棕苔、厚白苔、薄白苔、黄苔 4 类,该数据集如图 5 所示。该数据集中共有 300 张舌象图片,其中黑棕苔舌象图片数量为 47 张,厚白苔舌象图片数量为 80 张,薄白苔舌象图片数量为 112 张,黄苔舌象图片数量为 61 张。

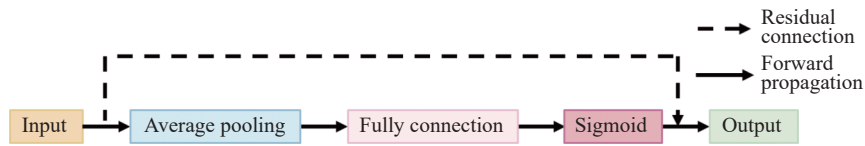


图 3 压缩激励模块结构

Fig. 3 Architecture of SE Block

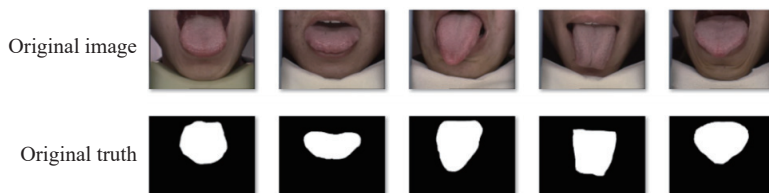


图 4 BioHit 数据集

Fig. 4 BioHit dataset

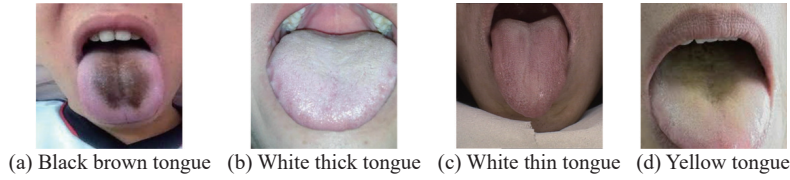


图5 舌象分类数据集

Fig. 5 Tongue classification dataset

为提高模型的泛化能力、鲁棒性和准确性,本文通过旋转、镜像、缩放、裁剪等方法,将分割数据集及分类数据集中的图像都由300张扩充至3000张。扩充完成后对数据集随机划分为训练集、验证集以及测试集(训练集、验证集和测试集分别占数据集的70%、20%和10%,即训练集中包含2100张图像,验证集中包含600张图像,测试集中包含300张图像)。

2.2 参数选用

2.2.1 舌象分割参数 为比较模型对舌象图像分割的能力,本文选用MAE(Mean absolute error)损失、Dice系数、交并比(IoU)、准确率(Acc)这4个参数对模型的分割性能进行评估。

MAE损失是模型在测试数据集上的损失值。它衡量了模型预测值与真实值之间的差距,MAE损失越低,模型性能越好,其计算公式为:

$$\text{MAE} = \frac{1}{n} \sum_{i=0}^n |A_i - B_i| \quad (1)$$

Dice系数是一种用于评估二分类任务或多分类任务中某个类别分割效果的指标,常用于图像分割任务中,其计算公式为:

$$\text{Dice} = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (2)$$

其中, A 是模型的预测结果, B 是真实标签, $|A \cap B|$ 表示预测结果和真实标签的交集。IoU是另一种常用于图像分割和目标检测任务的指标。它计算预测区域和真实区域的交集与并集的比值,其计算公式为:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

其中, $|A \cup B|$ 表示预测结果和真实标签的并集。Acc表示预测正确的样本数量占总样本数量的比例。Dice系数、IoU、Acc的取值范围都在0~1之间,取值越高代表模型性能越好。

2.2.2 舌象分类参数 为比较模型对舌象分类的能力,本文选用Acc、精确率(P)、召回率(R)和F1分数4个参数对模型的性能进行评估。

精确率在多分类任务中通常是对每个类别分别计算的,表示模型预测某个类别为正类时该预测结果的准确性。针对类别 i ,精确率计算公式为:

$$P_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \quad (4)$$

其中, TP_i 为被正确预测为类别 i 的样本数, FP_i 为被错误预测为类别 i 的样本数。计算每个类别的精确率后,精确率为所有类别精确率的平均值。

$$P = \frac{\sum_i P_i}{i} \quad (5)$$

召回率是指在所有实际为正类的样本中,模型能正确预测为正类的比例。它衡量的是模型对正类样本的覆盖程度。针对类别 i ,召回率计算公式为:

$$R_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \quad (6)$$

其中, FN_i 为实际属于类别 i 但被错误预测为其他类别的样本数。对所有类别的召回率取平均值,得到召回率 R 。

$$R = \frac{\sum_i R_i}{i} \quad (7)$$

F1分数是精确率和召回率的调和平均数,是一种综合平衡两者的指标。F1分数在精确率和召回率之间进行了权衡,适用于类别不平衡的数据集。

$$\text{F1} = 2 \times \frac{P \times R}{P + R} \quad (8)$$

准确率、精确率、召回率和F1分数4个参数的取值范围均在0~1之间,1表示最优。

2.3 模型训练

本文的硬件设备采用NVIDIA RTX 3080 GPU对训练过程进行加速,运行内存为32 G。通过Windows 11, Pycharm, Pytorch 1.13.0, CUDA11.7, cuDNN v8.4.0的软件环境对模型进行搭建。本文采用Adam优化器对模型的整体参数进行调整,学习率设置为 1×10^{-4} ,批量大小设置为16。最终预测结果经Sigmoid函数生成。

3 结果与分析

3.1 性能参数对比分析

为验证所设计模型有效性,本文通过实验将所设计模型应用于舌象图像语义分割任务,并对舌象

图像分类任务的效果进行了测试。

在分割任务中, 本文将所设计模型与 FCN^[15]、RNN^[16]、ResNet34^[17]、U-Net^[8]、U-Net++^[18]、DconnNet^[19]、TransUNet^[20] 分别在 BioHit 数据集上进行训练、验证与测试, 并使用 Dice、IoU、Acc、MAE 4 个参数对模型性能进行评估。分割部分的定量比较结果如表 1(左) 所示。其中, Dice、IoU、Acc 的数值越高, 代表模型性能越好; MAE 的数值越低, 代表模型性能越好。

在分类任务中, 本文将所设计模型与 CSSNet^[21]、DDANet^[22]、CAAP-Net^[23]、SANet^[24]、ResUNet++^[25]、UNext^[26]、Swin-unet^[27] 模型分别在分类数据集上进行训练、验证与测试, 并使用 Acc、P、R、F1 分数 4 个参数对模型性能进行评估, 分类部分的定量比较结果如表 1(右) 所示。

由表 1 可知, 与所选取的 5 个模型相比, 在分割与分类数据集上, 本文所设计的模型在所有参数上都取得了最佳值。

在分割任务中, 本文模型的 Dice 系数为 98.73%, 表明模型能够较好地处理舌象图像中的细节部分。本文模型的 IoU 指标达到了 97.50%, 平均

准确率为 99.43%, 相比于传统模型有较大提升。本文模型的 MAE 为 0.0057, 明显低于传统方法的 MAE 值, 表明 CRANet 相比传统模型具有更优表现。

在分类任务中, 本文模型的准确率达到 88.91%, 相较其他模型有较大提升。在精确率、召回率以及 F1 分数上, 本文模型分别达到了 86.37%、85.32% 以及 85.84%。

3.2 可视化结果分析

为了更直观地展示模型的分割性能, 本文对部分测试样本进行了可视化对比分析, 结果如图 6 所示。

图 6(a) 为原始舌象图像, 图 6(b) 为所期望的目标输出, 图 6(c)~6(f) 示出了各个模型所输出的语义分割结果。由图 6 可知, 本文模型能够准确地分割出舌象区域, 尤其在细节处理和边缘识别方面表现优异, 明显优于其他对比方法。在复杂背景下, 本文模型能够有效抑制背景噪声的干扰, 仅保留舌象区域, 显示出对重要特征的强选择性。

为对模型分类性能进行可视化分析, 本文绘制了混淆矩阵, 如图 7 所示, 以明确模型在不同类别上的正确分类和错误分类情况。由混淆矩阵可知, 薄白苔的正确分类率较高, 模型较少将其他类别错误

表 1 分割部分与分类部分的定量结果对比

Table 1 Comparison of quantitative results of the segmentation part and the classification part

Method	Segmentation				Method	Classification			
	Dice/%	IoU/%	Acc/%	MAE		Acc/%	P/%	R/%	F1/%
CRANet	98.73	97.50	99.43	0.0057	CRANet	88.91	86.37	85.32	85.84
FCN ^[15]	85.02	74.28	92.46	0.0754	CSSNet ^[21]	84.07	80.77	75.14	77.85
RNN ^[16]	96.19	94.58	98.75	0.0125	DDANet ^[22]	82.46	77.51	74.51	75.98
ResNet34 ^[17]	97.28	95.66	98.15	0.0120	CAAP-Net ^[23]	87.72	82.69	79.35	80.99
U-Net ^[8]	97.11	94.43	98.73	0.0127	SANet ^[24]	86.21	83.69	82.91	83.30
U-Net++ ^[18]	98.18	96.14	98.82	0.0098	ResUNet++ ^[25]	78.51	76.84	80.36	78.56
DconnNet ^[19]	89.53	90.57	94.38	0.0182	UNext ^[26]	80.53	75.27	70.28	72.69
TransUNet ^[20]	97.62	97.04	99.29	0.0061	Swin-unet ^[27]	82.30	78.19	73.47	75.76

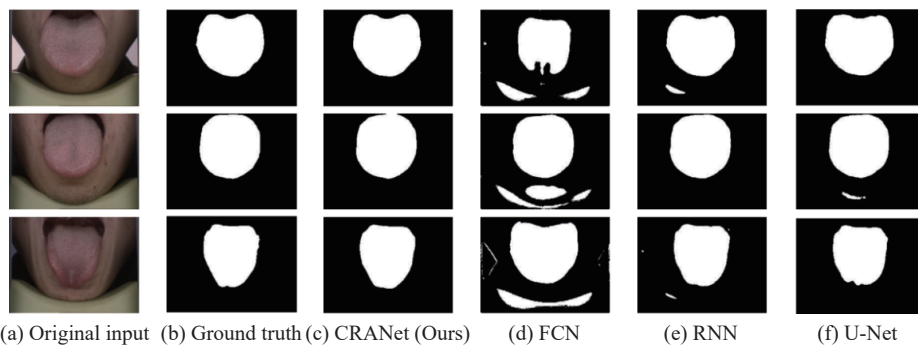


图 6 分割部分可视化结果对比图

Fig. 6 Visualization results comparison of the segmentation part

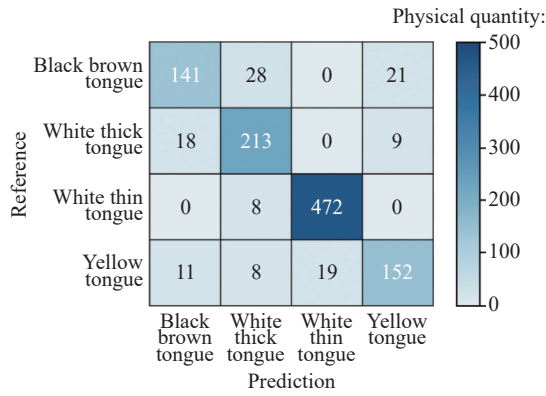
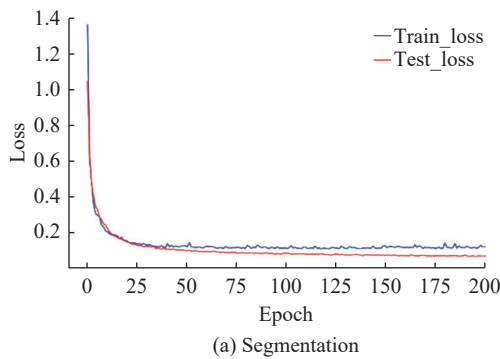
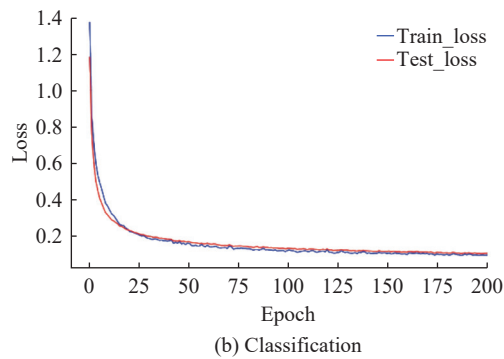


图7 混淆矩阵

Fig. 7 Confusion matrix



(a) Segmentation



(b) Classification

图8 模型训练与验证拟合曲线

Fig. 8 Model training and validation fitting curves

训练损失曲线的快速下降表明模型在不断优化过程中能够有效地拟合训练数据。验证损失曲线的逐步下降并趋于平稳,表明模型在避免过拟合的情况下能够很好地泛化到未见过的数据。这表明本文模型在训练过程中不仅能够很好地拟合训练数据,还具有较好的泛化能力。

3.4 消融实验分析

为证实各个模块的有效性,本文采用消融实验,分别去除卷积残差模块、多头注意力模块以及压缩激励模块,并对模型性能参数进行对比。分割部分消融实验结果如表2所示,分类部分消融实验结果如表3所示。

由消融实验可知,引入卷积残差模块能够使模型在分割任务中的Dice、IoU和Acc分别提升35.45%、36.31%和19.06%,使MAE降低0.0305;在分类任务中的Acc、P、R和F1分别提升20.73%、19.80%、15.42%和17.64%。引入多头注意力模块能够使模型在分割任务中的Dice、IoU和Acc分别提升17.11%、18.17%和11.84%,使MAE降低0.0161;在分类任务中的Acc、P、R和F1分别提升2.55%、

分类为薄白苔,说明该类别的特征较为明显,模型能够较好地识别。黑棕苔和厚白苔之间的分类混淆较为明显,模型在这两类之间的错误率相对较高,可能是因为这两类舌象在某些特征上存在相似性,需要进一步加强特征提取的精度。

3.3 拟合曲线分析

本文记录了训练过程中损失函数的变化情况,通过绘制训练与验证拟合曲线来分析模型的性能表现。在训练过程中,拟合训练损失和验证损失随训练轮数(Epoch)的变化曲线如图8所示。由此可知,训练损失和验证损失均随着训练的进行逐步下降,分割部分在大约25个训练轮数之后趋于平稳,分类部分大约在50个训练轮数后趋于平稳。

2.31%、3.44%和2.78%。引入压缩激励模块能够使模型在分割任务中的Dice、IoU和Acc分别提升

表2 分割部分消融实验结果对比

Table 2 Comparison of ablation experiments results of the segmentation part

CR Block	MHSA Block	SE Block	Dice/%	IoU/%	Acc/%	MAE
√	√	√	98.73	97.50	99.43	0.0057
√	√	×	96.58	95.74	96.32	0.0129
√	×	√	81.62	79.33	87.59	0.0218
×	√	√	63.28	61.19	80.37	0.0362

表3 分类部分消融实验结果对比

Table 3 Comparison of ablation experiments results of the classification part

CR Block	MHSA Block	SE Block	Acc/%	P/%	R/%	F1/%
√	√	√	88.91	86.37	85.32	85.84
√	√	×	83.64	82.16	78.76	80.42
√	×	√	86.36	84.06	81.88	82.96
×	√	√	68.18	66.57	69.90	68.20

2.15%、1.76% 和 3.11%，使 MAE 降低 0.0072；在分类任务中的 Acc、P、R 和 F1 分别提升 5.27%、4.21%、6.56% 和 5.32%。实验结果表明，去除每个模块都会导致模型性能下降，其中卷积残差模块对模型性能的影响尤为显著。

4 结束语

舌象图像分割与分类是医学图像处理中的重要任务，在中医舌诊中有着极大的应用前景。本文基于卷积神经网络与注意力机制设计了一种用于舌象图像分割与分类的神经网络 CRANet，并通过实验验证其性能。主要结论如下：

(1) 本文模型在各项评价指标上均优于传统方法和基准模型。具体而言，模型在分割任务中平均准确率达到了 99.43%，IoU 达到了 97.50%，Dice 系数达到了 98.73%，MAE 损失仅为 0.0057。模型在分类任务中平均准确率达到 88.91%，精确率达到了 86.37%，召回率达到了 85.32%，F1 分数达到了 85.84%。

(2) 通过对分割结果的可视化分析，本文展示了模型在不同舌象图像上的分割效果。模型能够清晰地分割出舌象区域，边缘精确，与真实标签高度一致。这些结果进一步验证了模型在处理复杂背景和细节识别方面的优势。

(3) 消融实验结果表明，本文所设计的卷积残差模块能够使得模型在分割与分类任务中均具有显著的性能提升。

本文所设计的模型 CRANet 在提高分割精度、减少人工干预和辅助医疗决策方面都表现出了显著的潜力和优势。未来的工作将继续致力于模型的优化和改进，以应对更复杂的医学图像分析需求，为中医舌诊提供更精确的辅助诊断工具。

参考文献：

- [1] CAI C, LIU S, WANG L, *et al.* Classification of Chinese herbal medicine using combination of broad learning system and convolutional neural network[C]//2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). Bari, Italy: IEEE, 2019: 3907-3912.
- [2] PAN D, GUO Y, FAN Y, *et al.* Development and application of traditional chinese medicine using AI machine learning and deep learning strategies[J]. *The American Journal of Chinese Medicine*, 2024, 52(3): 605-623.
- [3] BAE H, LEE S, LEE C, *et al.* A novel framework for understanding the pattern identification of traditional Asian medicine from the machine learning perspective[J]. *Frontiers in Medicine*, 2022, 8: 763533.
- [4] LIU Q, LI Y, YANG P, *et al.* A survey of artificial intelligence in tongue image for disease diagnosis and syndrome differentiation[J/OL]. *Digital Health*, (2023-08-01) [2024-09-10]. <https://sage.cnperreading.com/paragraph/article/?doi=10.1177/20552076231191044>.
- [5] SEO H, BADIEI K M, VASUDEVAN V, *et al.* Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications[J]. *Medical Physics*, 2020, 47(5): e148-e167.
- [6] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1440-1448.
- [7] 严计超, 常青. 低对比度医学图像全自动分割算法[J]. *华东理工大学学报(自然科学版)*, 2010, 36(4): 580-584.
- [8] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Munich, Germany: Springer International Publishing, 2015: 234-241.
- [9] 陈洪苇, 石文, 张静宇, 等. 基于 Morphsnakes 和 U-Net 的舌象分割算法研究[J]. *信息与电脑(理论版)*, 2022, 34(10): 57-60.
- [10] 任思羽, 吴瑞, 罗庆林, 等. 基于度量学习的多分支舌象识别网络[J]. *中国医学物理学杂志*, 2024, 41(4): 521-528.
- [11] MA S, LIU J, LI W, *et al.* Machine learning in TCM with natural products and molecules: Current status and future perspectives[J]. *Chinese Medicine*, 2023, 18(1): 43.
- [12] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[C]//31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, CA, USA: Curran Associates, 2017: 5998-6008.
- [13] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.
- [14] XIAN H, XIE Y, YANG Z, *et al.* Automatic tongue image quality assessment using a multi-task deep learning model[J]. *Frontiers in Physiology*, 2022, 13: 966214.
- [15] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [s.l.]: IEEE, 2015: 3431-3440.
- [16] SAK H, SENIOR A W, BEAUFAYS F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling[C]//Proceedings of the Annual Conference of the International Speech Communication Association. Singapore: Interspeech, 2014: 338-342.
- [17] 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. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [18] ZHOU Z, RAHMAN S M, TAJBAKHSH N, *et al.* UNet++: A nested U-NET architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop-DLMIA 2018. Granada, Spain: Springer International Publishing, 2018: 3-11.
- [19] YANG Z, FARSIU S. Directional connectivity-based segmentation of medical images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [s.l.]: IEEE, 2023: 11525-11535.
- [20] CHEN J, MEI J, LI X, *et al.* TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers[J]. *Medical Image Analysis*, 2024, 97: 103280.
- [21] SHAO Y, ZHOU K, ZHANG L. CSSNet: Cascaded spatial shift network for multi-organ segmentation[J]. *Computers in Biology and Medicine*, 2024, 170: 107955.
- [22] TOMAR N K, JHA D, ALI S, *et al.* DDANet: Dual decoder attention network for automatic polyp segmentation[C]//ICPR International Workshops and Challenges: Virtual Event. [s.l.]: Springer International Publishing, 2021: 307-314.
- [23] SAXENA P, BHANDARI A K. CAAP-Net: Context aware automatic polyp segmentation network with mask attention[J]. *IEEE Transactions on Artificial Intelligence*, 2024, 5(7): 3510-3523.
- [24] WEI J, HU Y, ZHANG R, *et al.* Shallow attention network for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Strasbourg, France: Springer International Publishing, 2021: 699-708.
- [25] JHA D, SMEDSRUD P H, RIEGLER M A, *et al.* ResUNet++: An advanced architecture for medical image segmentation[C]//2019 IEEE International Symposium on Multimedia (ISM). San Diego, CA, USA: IEEE, 2019: 225-230.
- [26] VALANARASU J M, PATEL V M. UNeXt: MLP-based rapid medical image segmentation network[C]//International Conference on Medical Image Computing and Computer-assisted Intervention. Cham: Springer Nature Switzerland, 2022: 23-33.
- [27] CAO H, WANG Y, CHEN J, *et al.* Swin-Unet: Unet-like pure transformer for medical image segmentation[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 205-218.

CRANet: Tongue Image Segmentation and Classification Network

PENG Chen¹, LI Wenju², QIAN Zhiqin¹, XU Lei³, LUO Qi¹, YU Yijie¹

(1. School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai 200237, China; 2. Internal Medicine Department II, Zhengzhou Renji Hospital South Campus, Zhengzhou 450004, China; 3. Logistic Support Department, Shanghai Jiading Anting Hospital, Shanghai 201800, China)

Abstract: In Traditional Chinese Medicine (TCM) diagnosis, tongue diagnosis is an important method. However, the accuracy and efficiency of tongue analysis are greatly influenced by subjective factors. This paper proposes a tongue segmentation and classification model based on convolutional neural networks and attention mechanism (CRANet), and designs a Convolutional Residual Block (CR Block). The model can automatically segment the tongue and recognize features, thus classifying tongue images and improving the objectivity and accuracy of tongue diagnosis. The proposed model achieved an average accuracy of 99.43%, an Intersection over Union (IoU) of 97.50%, and a Dice coefficient of 98.73% in the segmentation task. In the classification task, it achieved an average accuracy of 88.91%, a precision of 86.37%, a recall of 85.32%, and an F1 score of 85.84%. This verifies the application potential of the tongue image segmentation and classification method based on Convolutional Neural Networks and Attention Mechanism in TCM tongue diagnosis.

Key words: convolutional neural network; attention mechanism; image segmentation; image classification; traditional chinese medicine tongue diagnosis

(责任编辑:李娟)