

# 融合多注意力的舌象证候分类<sup>①</sup>

宁宏宇<sup>1</sup>, 张魁星<sup>1</sup>, 薛丹<sup>2</sup>, 江梅<sup>1</sup>

<sup>1</sup>(山东中医药大学 智能与信息工程学院, 济南 250355)

<sup>2</sup>(山东第一医科大学第一附属医院 (山东省千佛山医院), 济南 250013)

通信作者: 张魁星, E-mail: zhangkuixing@sducm.edu.cn



**摘要:** 智能舌诊在协助医生诊断病情方面具有重要意义. 当前, 智能舌诊主要集中在单一舌象特征的预测分类, 难以在诊断过程中提供实质性的帮助. 为弥补这一不足, 从舌象证候层面进行精准的预测分类研究, 协助医生诊断病情. 使用 TUNet 对舌体进行分割, 并提出融合多注意力机制的平行残差网络 PMANet 用于舌象证候分类. 在像素准确率、平均交并比和 Dice 系数 3 个评价指标上, TUNet 分别达到 99.7%、98.4%、99.2%, 相较于基线 U-Net, 提高了 3.2%、9.0%、4.8%. 在舌象证候分类研究中, PMANet 的参数总量为 12.34M, 略高于对比实验中的 EfficientNet, 总浮点计算数为 1.021G, 远低于所有对比网络. 在参数量和浮点计算数更少的情况下, 取得了 95.7% 的分类准确率, 实现了精度、参数量和浮点运算数之间的平衡. 这一方法为智能舌诊研究提供了重要支持, 有望推进中医舌诊现代化进程.

**关键词:** 舌诊; 舌体分割; 舌象证候分类; 深度学习; 舌诊智能化

引用格式: 宁宏宇, 张魁星, 薛丹, 江梅. 融合多注意力的舌象证候分类. 计算机系统应用, 2024, 33(10):228-235. <http://www.c-s-a.org.cn/1003-3254/9665.html>

## Tongue Image Syndrome Classification Integrated with Multiple Attention

NING Hong-Yu<sup>1</sup>, ZHANG Kui-Xing<sup>1</sup>, XUE Dan<sup>2</sup>, JIANG Mei<sup>1</sup>

<sup>1</sup>(School of Intelligence and Information Engineering, Shandong University of Traditional Chinese Medicine, Jinan 250355, China)

<sup>2</sup>(The First Affiliated Hospital of Shandong First Medical University (Shandong Provincial Qianfoshan Hospital), Jinan 250013, China)

**Abstract:** Intelligent tongue diagnosis is of great significance in assisting doctors in medical treatment. At present, intelligent tongue diagnosis is mainly focused on the prediction and classification of single tongue image features, making it difficult to provide substantial help in the diagnostic process. To make up for this deficiency, research of accurate prediction and classification is carried out from the level of tongue image syndrome to assist doctors in diagnosing diseases. The TUNet is used to segment the tongue, and a parallel residual network PMANet integrated with the multi-attention mechanism is proposed to classify the syndrome of tongue image. The pixel accuracy (*PA*), mean intersection over union (*MIoU*) and *Dice* coefficient of TUNet reach 99.7%, 98.4%, and 99.2%, respectively, improved by 3.2%, 9.0%, and 4.8% compared with the baseline U-Net. In the research of tongue image syndrome classification, PMANet's total amount of parameters is 12.34M, slightly higher than that of EfficientNet, and its total amount of floating-point calculations is 1.021G, significantly lower than all compared networks. Under the background of a lower amount of both parameters and floating-point calculations, the classification accuracy of PMANet reaches 95.7%, achieving a balance between precision, parameter amount, and floating-point calculations amount. This method provides support for the research of intelligent tongue diagnosis and is expected to promote the modernization of TCM tongue diagnosis.

**Key words:** tongue diagnosis; tongue segmentation; classification of tongue image syndrome; deep learning; intelligent tongue diagnosis

① 基金项目: 山东省自然科学基金 (ZR2020KF013); 山东省中医药科技项目 (Q-2022052); 青岛市科技惠民示范专项 (23-2-8-smjk-2-nsh)

收稿时间: 2024-03-29; 修改时间: 2024-04-23, 2024-05-14; 采用时间: 2024-05-23; csa 在线出版时间: 2024-08-21

CNKI 网络首发时间: 2024-08-22

舌诊,作为中医诊察疾病的重要手段,以其实施便捷和诊断价值显著等优点广泛应用于临床<sup>[1-4]</sup>。然而,传统舌诊易受医生主观经验的影响,且缺乏客观的评判标准,限制了其进一步的发展<sup>[5]</sup>。为减少主客观因素等对诊断结果的影响,许多研究者将舌诊与深度学习技术相结合,以期实现舌诊的客观化和智能化,协助医生诊断、提升工作效率。相关工作主要从舌体分割、舌象特征分类等方面展开<sup>[6]</sup>。

舌体分割是舌诊客观化的基础,准确分割舌体是实现舌诊智能化分析的重要前提。目前,基于深度学习的方法广泛用于舌体分割<sup>[7]</sup>。Liu等<sup>[8]</sup>提出TCCNet网络,并引入Triplet-Loss损失函数,对舌体分割进行研究。结果表明,TCCNet的F1-score和平均精度分别达到88.86%和93.49%,相较于VGG、GoogLeNet等网络,性能显著提升。此外,李宗润等<sup>[9]</sup>使用U-Net<sup>[10]</sup>和DeeplabV3<sup>[11]</sup>进行对比试验,研究表明U-Net在舌体分割方面表现更加出色,其像素准确度和平均交并比为93%和91%。罗思言等<sup>[12]</sup>使用Swin Transformer<sup>[13]</sup>,对哈尔滨工业大学公开的舌象数据集<sup>[14]</sup>进行广泛研究,实验获得了99.47%的平均交并比。上述分割方法还存在分割精度低、效率差等问题,导致舌体边缘不够精细且存在大量噪点,对后续分类研究造成影响。

另一方面,Ni等<sup>[15]</sup>将CapsNet应用于舌色分类研究,并融合残差结构,提出TongueCaps网络,实现了84.56%的分类准确率。卓力等<sup>[16]</sup>将通道注意力机制嵌入轻量级神经网络,提出带有过滤机制的知识蒸馏策略,在教师网络的指导下,显著降低网络复杂度的同时取得93.88%的分类准确度。颜建军等<sup>[17]</sup>采用优化的随机森林算法,在齿痕舌的二分类任务中取得了93%的分类准确度。这些研究虽然都取得了较好的结果,但在实际的医学诊断中,仅依赖单一特征的简单分类,对医生提供实质性帮助有限。

针对上述问题,我们使用改进的TUNet网络进行舌体分割,高效精准地将舌体从面部中分离。不同于以往单一特征(如舌色、齿痕舌等)的分类方法,本研究综合多种舌象特征,以舌象证候为研究对象,为智能舌诊的发展提供了全新视角。我们提出融合多注意力机制的平行残差网络PMANet,通过对舌象证候特征的精准预测分类,协助医生快速诊断患者病情,推动舌诊智能化研究的进展。

## 1 技术背景

### 1.1 U-Net 网络

U-Net是基于全卷积神经网络的分割模型。为解决全卷积神经网络在处理分割任务精度较低的问题,U-Net引入对称的编码器、解码器结构用于生成预测掩膜。编码器挖掘深层次图像信息,解码器融合的特征重新组织生成更精确的输出。同时引入跳跃连接将对应分辨率的特征图合并,更好地保留和利用底层特征,减少网络在传递过程中的信息丢失。

### 1.2 ResNet34 网络

He等<sup>[18]</sup>提出了残差网络(ResNet),通过堆叠残差块来增加网络深度。该设计既解决了精度退化等问题,还显著提高了分类精度。ResNet34是其中一种具体实现形式,其残差块由两条路径组成,分别为 $F(x)$ 和 $X$ 。 $F(x)$ 用于拟合残差,称为残差路径, $X$ 为输入的恒等映射。残差块的输出是两条路径之和,确保网络性能不会下降,具体结构如图1所示。

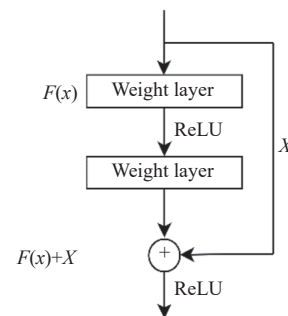


图1 残差模块结构图

### 1.3 通道和空间注意力模块

空间和通道注意力机制模块(convolutional block attention module, CBAM)是一种轻量级的通用模块,能够与任何卷积神经网络兼容<sup>[19,20]</sup>,如图2所示。该模块在通道和空间两个维度上产生注意力信息<sup>[21]</sup>。这些特征信息与对应的输入相乘后自适应调整,生成带有特定注意力的特征图。通过诱导网络聚焦目标对象和精准关注无关杂波降噪,提高模型的性能表现。

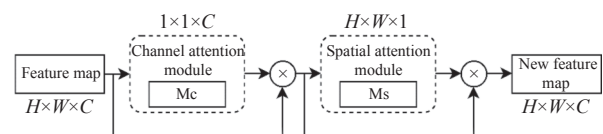


图2 CBAM 整体架构图

## 2 实验方法

### 2.1 平行多注意力分类网络 PMANet

#### 2.1.1 平行多分支模块

提出了一种平行多分支残差模块 PMBR, 由纵向卷积、横向卷积、深度可分离卷积和 SE 注意力模块<sup>[22,23]</sup>融合构成, 具体设计如图 3 所示. Transformer 之所以在众多任务中表现优异, 在于其能够根据全局上下文信息进行建模. 为了扩展感受野和增强网络上下文建模能力, 采用  $11 \times 11$  的大卷积核代替  $3 \times 3$  卷积核. 然而, 随着卷积核的增大也会带来一些值得关注的问题: 尽管大卷积核能够提供更广阔的感受野, 但可能会导致局部信息的丢失, 进而降低网络的表达能力. 同时, 随着网络深度的增加, 计算量和内存需求也会上升, 延长训练和推理时间. 为了解决上述问题, 将  $11 \times 11$  的卷积核拆分为两个并行的深度卷积: 一个  $11 \times 1$  的横向卷积和一个  $1 \times 11$  的纵向卷积. 当特征图输入该模块时, 沿通道维度等量切分. 横向卷积负责提取横向语义信息, 纵向卷积则提取纵向语义信息. 然后, 带有这两种信息的特征图在通道维度进行融合, 以捕捉更广泛的图像信息和更复杂、更抽象的特征. 为了弥补大卷积核带来的局部信息缺失的问题, 融合后的特征图使用  $3 \times 3$  的深度卷积, 保留局部结构和细节信息, 提高网络对局部特征的捕捉能力. SE 模块通过动态调整特征通道的权重, 使网络能够更好地适应不同任务, 提升深度神经网络的性能. 最后, 将这 3 个输出的特征图进行残差融合, 既保留了由大感受野带来的更广泛的图像信息, 又能够很好地保留局部细节. 大卷积核能够减少卷积的计算量和参数量, 而纵向卷积和横向卷积又能够显著降低内存需求. 因此, PMBR 模块不仅具有类似注意力的上下文建模空间, 还能在一定程度上弥补注意力模型在局部信息缺失以及计算、内存需求大的不足.

假设  $X^T \in R^{H \times W \times C}$  表示映射到平行多分支模块的输入特征. 按通道维度将输入划分成两个特征向量, 平行通过纵向卷积和横向卷积层后在通道维度拼接恢复到与输入相同形状.

$$X'_{Dw} = \sigma(Dw_{(1 \times 11)}(X'_a) + Dw_{(11 \times 1)}(X'_b)) \quad (1)$$

其中,  $\sigma$  表示激活函数, Dw 为深度卷积操作, 括号内数字表示卷积核大小,  $a, b$  为特征图通道维度大小.

$$X_{attn} = \sigma(Dw_{(3 \times 3)}(X'_{Dw}) + SE(X'_{Dw})) + X'_{Dw} \quad (2)$$

其中,  $\sigma$  表示激活函数, Dw 为深度卷积操作, 括号内数字表示卷积核大小, SE 为通道注意力操作.

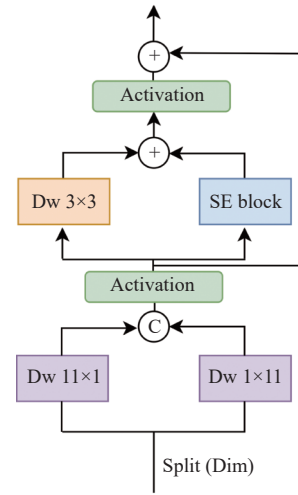


图 3 PMBR 模块架构图

#### 2.1.2 并行多注意力模块

提出了一种并行多注意力模块 PMA, 该模块由平行的 PMBR 模块和 CBAM 模块构成, 分别在获取长效感受野、减少计算内存、降低计算参数和增强聚焦能力等方面提升网络性能, 具体结构如图 4 所示. 在 PMA 模块中, 输入特征图并行流入两个子模块进行处理. 在 CBAM 模块中, 特征图依次通过通道注意力模块和空间注意力模块, 捕捉并整合空间注意力和通道注意力信息, 输出带有两种注意力的特征图; PMBR 模块, 不仅扩大了上下文信息的捕捉范围, 增强局部信息关注能力, 同时节省内存使用和降低网络参数. 两个子模块的输出在通道维度上进行合并后, 经过  $1 \times 1$  的深度卷积, 并与原始特征图进行叠加, 充分融合两模块的建模信息. PMA 能够同时获得通道和空间注意力, 并有效扩展卷积的感受野, 进而在不增加额外计算负担的情况下, 提高网络的性能和效率.

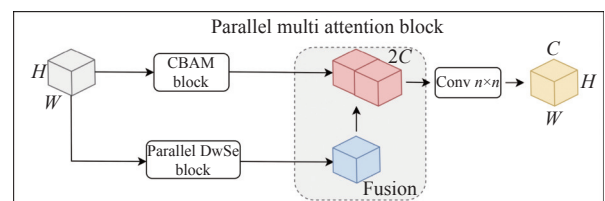


图 4 PMA 模块架构图

在本研究中, 对 ResNet34 架构进行了深度优化, 主要从以下 3 个维度进行改进: (1) 用两个连续的  $3 \times 3$

深度卷积层替代了原始的  $7 \times 7$  卷积层和最大池化层。在提升网络表征能力的同时,减少模型的复杂性和对计算资源的需求。(2) 将 PMA 模块集成到网络中,替换所有残差块,进一步提高网络的效率,增强其泛化能力,并优化注意力机制在特征提取过程中的作用。(3) 在网络中引入了多组具有不同参数设置的 Dropout 层,减

少模型的计算复杂度,消除冗余特征,加快推理速度。基于上述改进,提出了一种全新的金字塔层次网络结构 PMANet。该模型的架构概述在图 5 中展示。PMANet 的不仅能够提升网络处理高维视觉数据时的性能,而且在保持高效率的同时,促进模型在分类任务中的适用性和泛化能力。

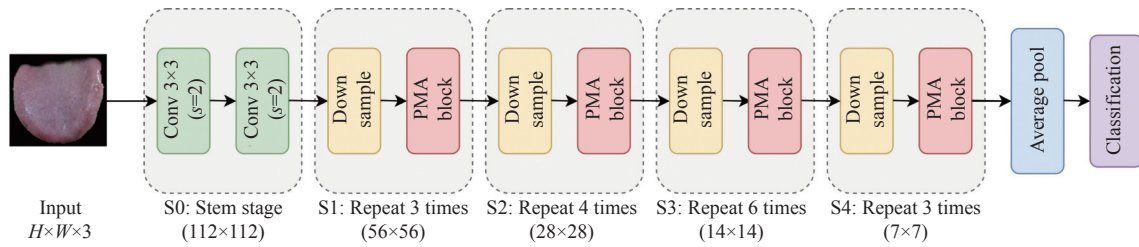


图 5 PMANet 整体架构图

## 2.2 改进的 TUNet 分割网络

尽管 U-Net 在图像分割任务中表现优异,但在舌体分割任务中仍然存在一些不足。首先, U-Net 对数据分布有较高的要求,而舌象数据的分布不均匀,导致分割精度大幅下降。其次, U-Net 难以准确捕捉物体的边缘细节,这导致分割得到的舌体边缘模糊不清,从而对后续研究产生不利影响。针对这些问题,本文在以下 3 个方面对传统 U-Net 进行了改进:(1) 舍弃传统的池化操作,改用卷积下采样,以保留更多的空间信息,增强模型的鲁棒性。(2) 使用空洞卷积以增大感受野,从而提高模型对舌体边缘的感知能力<sup>[24]</sup>。此外,使用批归一化来规范卷积层输出,减小内部协变量偏移,提升网络训练的稳定性。(3) 使用双线性插值方法在上采样中对特征图进行平滑处理,使得上采样后的特征图具有更高的分辨率和更精确的细节描述,进一步提高了分割的精度。经过上述改进的 U-Net 能够更好地适应舌象数据集的分布特征。由于该模型专注于舌体 (tongue) 分割,故将改进后的 U-Net 命名为 TUNet。图 6 展示了 TUNet 的整体架构。

## 3 实验和分析

在舌象证候分类研究中,要综合考虑舌色、舌形(胖瘦)、舌痕等多方面因素的影响。为了使网络更加专注于舌体本身,减少其他因素的干扰,本研究设计了两个独立的网络:用于舌体分割的 TUNet、用于证候分类的 PMANet,两网络均使用交叉熵损失函数 (cross-entropy loss function) 对网络进行训练,但由于是两个

独立的网络,两损失函数之间并无关联。整体实验过程如图 7 所示。

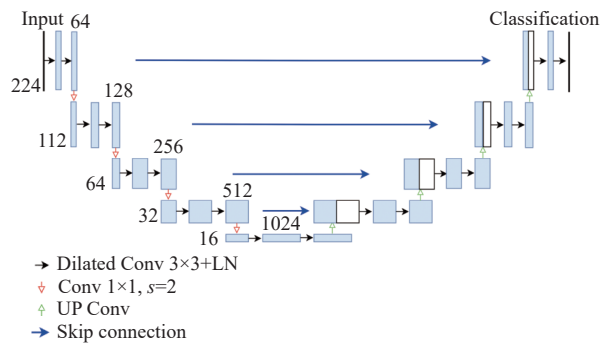


图 6 TUNet 整体架构图

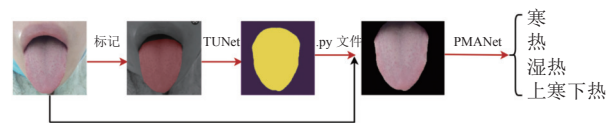


图 7 整体实验过程图

训练分割网络 (TUNet) 的原因:(1) 舌体在面部所占的比例较低,直接进行舌象证候分类不能取得理想的分类结果,需要将舌体从整个面部中分割出来,单独进行分类研究。(2) 手工标注分割出的舌体不精细,且需要消耗巨大的人力、时间成本。通过部分手工标记的标签图,去训练一个分割网络不仅能节省大量的时间成本,而且能够更加精细的分割舌体。通过分割得到舌体图像是后续证候分类实验的基础和前提,确保 PMANet 能够在一个干净的舌体图像上进行训练和评估,从而提高分类的准确性和可靠性。

### 3.1 数据集和实验环境

将收集到的标准舌图像使用 LabelMe 工具进行手工标注, 并按照 8:1:1 的比例划分为训练集、验证集和测试集. 根据专业医师的标注, 将分割得到的舌体图像划分为寒、热、湿热和上寒下热 4 类, 具体数据分布如表 1 所示. 为了增强模型的泛化能力和鲁棒性, 使用随机水平翻转、垂直翻转和随机裁切等数据增强方式, 增加数据集的规模.

表 1 证候分布表

证候	数量
寒	79
热	65
湿热	47
上寒下热	58

实验环境配置如下: 本实验在 Linux 操作系统上进行, 硬件配置为 Intel(R) Xeon(R) Silver 4210R 处理器和 NVIDIA RTX 3090 显卡. 所用的软件版本包括 Python 3.7.3、PyTorch 1.12.1 和 CUDA 11.3.

### 3.2 评价指标

采用像素准确率 (pixel accuracy,  $PA$ ), 平均交并比 ((mean intersection over union,  $MIoU$ )) 和  $Dice$  系数 ( $Dice$  similarity coefficient,  $Dice$ ), 作为衡量舌体分割算法性能的标准

$PA$  是预测类别正确的像素数占总像素数的比例, 计算可描述为:

$$PA = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$MIoU$  衡量预测结果与真实标签的重叠程度, 计算可描述为:

$$MIoU = \frac{1}{k} \sum_{i=1}^k \frac{TP}{TP + FP + FN} \quad (4)$$

$Dice$  系数用于计算两样本的相似度, 结果在  $[0, 1]$  之间, 计算可描述为:

$$Dice = \frac{2 \times TP}{(TP + FP) + (TP + FN)} \quad (5)$$

在分类实验中, 使用精准率 ( $Accuracy$ )、精确率 ( $Precision$ )、召回率 ( $Recall$ ) 和混淆矩阵 (confusion matrix) 作为衡量模型好坏的评价指标.

$Precision$  表示被分类样本中真正例的比例, 是衡量网络预测正例样本的准确性, 计算过程可描述为:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$Recall$  表示所有真正例中被正确识别为正例的比例, 计算过程可描述为:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

混淆矩阵是一个用于评估分类模型性能的表格, 其中行表示实际类别, 每一行的数据总数表示该类别的数据实例数量; 列表示预测类别, 每一列的总数表示被预测为该类别的数据实例数量.

式 (3)–式 (7) 中,  $TP$  (真正例) 表示实际为正且被预测为正的样本数;  $TN$  (真负例) 表示实际为负且被预测为负的样本数;  $FP$  (假正例) 表示实际为负但被预测为正的样本数;  $FN$  (假负例) 表示实际为正但被预测为负的样本数.

### 3.3 舌体分割实验分析

为验证 TUNet 在舌体分割任务中的有效性, 与基准 U-Net 进行对比实验. 表 2 中, TUNet 的像素准确率为 99.7%, 平均交并比为 98.4%,  $Dice$  系数为 99.2%. 相较于 U-Net, TUNet 在这 3 个指标上分别提升了 3.2%、9.0% 和 4.8%.

表 2 各评价指标对照表 (%)

模型	$Accuracy$	$MIoU$	$Dice$
TUNet	99.7	98.4	99.2
U-Net	96.5	90.2	94.4

使用 TUNet 对 4 种证候的舌图像进行随机语义分割, 并将得到的预测掩膜进行黑白二值转换后叠加到原图像上以分割出舌体, 效果如图 8 所示. (a)、(b)、(c) 和 (d) 分别表示寒、热、湿热及上寒下热 4 种证候特征; (1)、(2)、(3) 和 (4) 则分别对应舌图像、预测掩膜、黑白掩膜和舌体分割图. 从图 8 可以看出, 在靠近牙齿、嘴唇以及舌体边缘等较难分割的部分, TUNet 几乎完美地预测, 说明 TUNet 能够更好地适应舌象数据集的分布, 保留更多舌体边缘的细节信息.

### 3.4 舌象证候分类实验分析

#### 3.4.1 PMANet 用于证候分类

为了评估本方法的有效性, 在舌象证候数据集上, 将 PMANet 与基于 CNN 和 Transformer 的先进模型进行比较. 首先将 PMANet 与基线 ResNet34 残差网络进行比较, 两个网络在验证集上的准确率和损失率对比如图 9、图 10 所示.

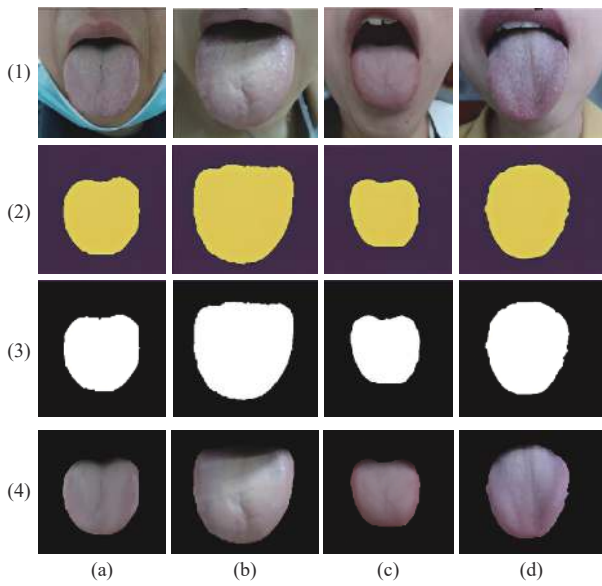


图8 舌体分割效果图

两网络均在 150 个 epochs 内收敛 (如图 9 所示), 但随着 epoch 的增大出现小幅度的震荡, 可能是以下原因导致: 1) 所有照片均为手机拍摄, 光照条件、色彩差异以及计算摄影等因素可能会对图像质量产生影响, 进而影响分类性能. 2) 分类依赖的舌体图像通过分割网络得到的, 部分舌体图像边缘分割存在误差或舌体内部存在预测错误的像素等也会对分类造成影响. 3) 由于网络结构较为庞大复杂, 而收集到的舌象数据相对较少, 可能导致网络在训练时存在一定的过拟合. 针对这些可能存在的问题, 给出以下解决方案: 1) 尽量在光线充足、背景干净的环境中拍摄舌体图像, 以减少光照和色彩的差异. 调整图像的亮度、对比度、色彩平衡等, 尽可能减少手机摄影带来的影响. 2) 继续优化分割算法, 减少边缘分割误差或内部像素预测错误的问题. 3) 扩充训练数据或使用正则化方法限制模型复杂度.

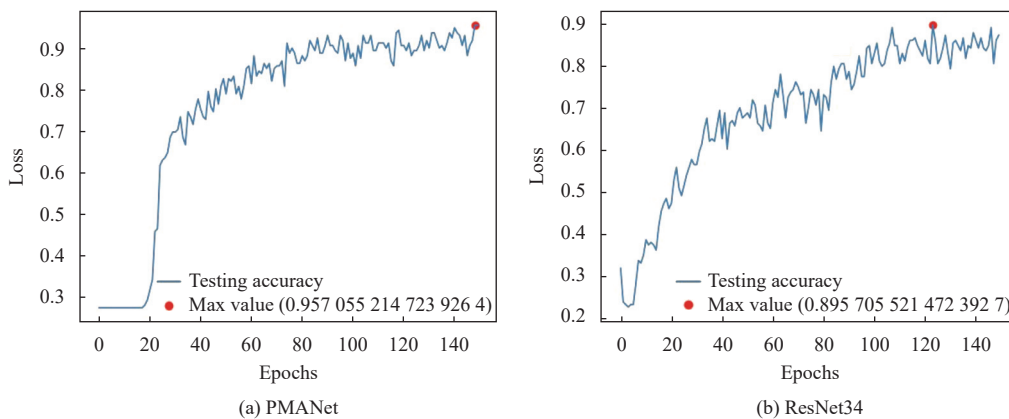


图9 测试集准确率对比图

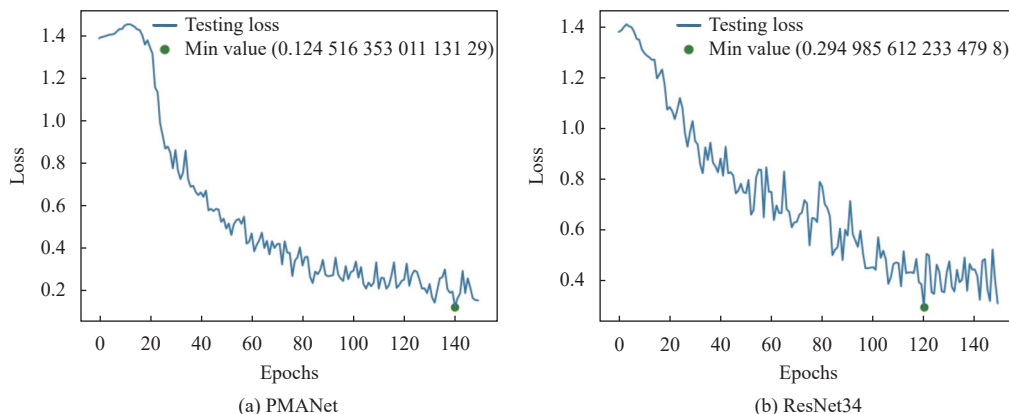


图10 测试集准确率对比图

提出的 PMANet 在验证集上的分类准确率为 95.7%, 相较于 ResNet34 提升了 6.2%, PMANet 的损失为 0.12, 相较于 ResNet34 低了 0.17. 说明 PMANet 在 PMA

模块影响下, 能够更快速地聚焦目标对象, 对证候进行更为精准的分类. 为了进一步探究 PMANet 在证候分类上的有效性, 与另外几种先进的 CNN 和 Transformer

模型进行对比实验。

由于是第 1 个以舌象证候分类作为研究对象的研究,所有对比实验的结果均在上述构建数据集上实验得出.为了保证实验的公平性和对比的有效性,所有对比试验均未使用预训练模型,使用随机水平翻转、垂直翻转随即裁切等进行数据扩充,使用 AdamW 优化器,初始学习率设置为 1E-3,使用余弦退火策略逐步减小学习率,权重衰减设置为 5E-4,批量大小 (batch size) 设置为 128,训练 150 个 epoch,其中前 5 个 epoch 为预热轮次。

表 3 展示了与其他主干网络 (包括基于 Transformer 和基于 ConvNet 的网络) 的比较.与先进的基于 Transformer (即 Swin T) 的架构相比,在 224×224 的输入上,PMANet (95.7%) 比 Swin T (85.4%) 提升了 10.3%.相较于 ViT, PMANet 的分类精度提升了 3.7%,但参数量远低于 ViT,且在浮点运算量上也比 ViT 更少。

表 3 与其他网络对比表

方法	Accuracy (%)	Params (M)	FLOPs (G)
Swin T	85.4	28.27	4.372
EfficientNet	91.4	12.23	1.504
ResNet34	89.7	21.79	3.679
ResNet50	90.2	27.64	4.203
ResNeXt50	90.8	23.87	3.004
ConvNeXt_Tiny	88.3	27.27	4.157
DenseNet	87.1	21.92	3.99
ViT	92.0	82.56	1.251
Ours	95.7	12.34	1.021

与最先进的 ConvNet 相比, PMANet 在分类准确度、算数量和浮点运算数上有着更好的平衡.具体来说,对比精度最高的 EfficientNet, PMANet 虽然参数量略多,但 Top-1 准确率高 4.3%,且浮点运算量低 32%.相比于精度最差的 DenseNet, PMANet 在精度上提升了 8.6%。

### 3.4.2 混淆矩阵分析

图 11 为测试集的混淆矩阵,其中横轴标签表示预测结果,纵轴标签表示真实属性.从图中可以看出,寒、热和上寒下热的错误预测均仅有 1 例,而湿热的错误预测则为 2 例.在寒类中,错误地把寒类证候预测为湿类,湿类预测却有着相反的结果,说明寒湿两类证候在成像方面有着更多类似的地方. PMANet 能够准确地预测舌象证候特征,在寒、热和上寒下热的证候中表现得更好,能够为医生诊断提供有益的参考。

根据混淆矩阵结果可以计算出模型在测试集上的平均准确率 (Precision)、召回率 (Recall) 和精确率

(Specificity),详细结果如表 4 中所示,进一步说明本文的方法的具有一定的可行性。

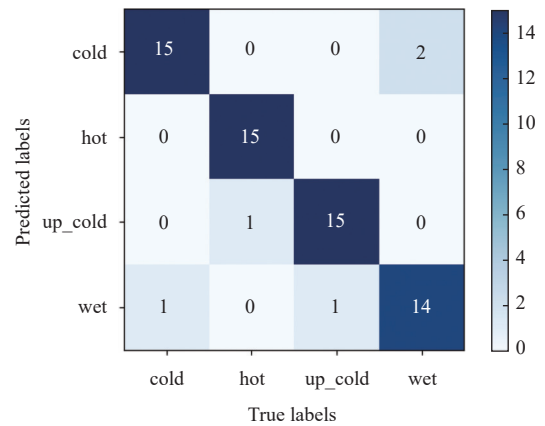


图 11 测试集混淆矩阵图

表 4 混淆矩阵计算表 (%)

评价指标	Mean score
Precision	92.38
Recall	92.23
Specificity	97.38

### 3.4.3 消融实验

为探讨不同模块对模型性能的影响,在证候数据集上进行了消融实验.根据表 5 的数据, PMA 模块带来了 4.5% 的性能提升, PMBR 模块则实现了 3.9% 的提升.对比分析表明, PMA 模块对提高模型准确率的效果更为显著.此外,当将 PMA 与 PMBR 模块一起使用时,模型达到了最高的准确率,这表明这两个模块的协同作用能够更好地适应舌象数据集,从而提供更高的分类精度。

表 5 消融实验模块对比表

Component		Loss	Accuracy (%)
PMA	PMBR		
—	—	0.29	89.6
√	—	0.13	94.5
—	√	0.19	93.3
√	√	0.12	95.7

## 4 结论

本研究的主要目的是通过对舌象证候进行精准的预测分类,从证候的角度去协助医生进行诊断患者病情,弥补当前智能舌诊研究的不足,以求推进舌诊研究的智能化和现代化.研究表明,相较于基准网络, TUNet 能够更加快速准确的分割出舌体,为后续分类实验奠定良好的基础.然后,设计了一种融合多注意力的分类网络 PMANet,在舌象证候数据集上进行实验,通过融

合带有注意力和大卷积野的特征图,增强模型聚焦能力,提升模型性能。PMANet 的分类准确率为 95.7%,相较于其他先进网络,不仅取得了最高的分类准确率,而且实现了参数、准确率和浮点运算数的平衡。说明,本研究能够很好地将各种舌象证候进行精确分类,为协助医生诊疗患者病情提供一种全新的角度和方法。

本研究的不足在于收集的舌象数据较少,可能会对模型泛化能力造成一定影响,在未来研究中,会持续扩大数据集规模。其次,在整个实验设计中,使用类 ResNet34 作为网络主干框架,在未来的工作中可以探索更多不同骨干框架对于分类精度的影响。同时,还会考虑在更多不同数据集下评估 PMANet 的鲁棒性和泛化性。

### 参考文献

- 1 吴欣,徐红,林卓胜,等.深度学习在舌象分类中的研究综述.计算机科学与探索,2023,17(2):303-323.[doi:10.3778/j.issn.1673-9418.2208052]
- 2 Marhamati M, Zadeh AAL, Fard MM, *et al.* LAIU-Net: A learning-to-augment incorporated robust U-Net for depressed humans' tongue segmentation. Displays, 2023, 76: 102371. [doi: 10.1016/j.displa.2023.102371]
- 3 Bhatnagar V, Bansod PP. Challenges and solutions in automated tongue diagnosis techniques: A review. Critical Reviews™ in Biomedical Engineering, 2022, 50(1): 47-63. [doi: 10.1615/CritRevBiomedEng.2022044392]
- 4 Chelluboina SP, Rao KN. CATDSNet: Computer aided tongue diagnosis system for disease prediction using hybrid extreme learning machine. International Journal of Intelligent Engineering & Systems, 2023, 16(1): 1-12.
- 5 汪莉,彭成东,宋婷婷,等.自然光下中医舌色的辨识研究.时珍国医国药,2023,34(2):493-495.[doi:10.3969/j.issn.1008-0805.2023.02.69]
- 6 Liu Q, Li Y, Yang P, *et al.* A survey of artificial intelligence in tongue image for disease diagnosis and syndrome differentiation. Digital Health, 2023, 9: 20552076231191044.
- 7 Shi SK, Wang YQ, Ye QH, *et al.* Ammonia-Net: A multi-task joint learning model for multi-class segmentation and classification in tooth-marked tongue diagnosis. arXiv: 2310.03472, 2023.
- 8 Liu W, Chen JM, Liu B, *et al.* Tongue image segmentation and tongue color classification based on deep learning. Digital Chinese Medicine, 2022, 5(3): 253-263. [doi: 10.1016/j.dcm.2022.10.002]
- 9 李宗润,李炜弘,井淇.基于多种机器视觉技术的舌体分割模型研究.时珍国医国药,2021,32(7):1768-1771.[doi:10.3969/j.issn.1008-0805.2021.07.68]
- 10 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234-241.
- 11 Chen LC, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587, 2017.
- 12 罗思言,王心舟,饶向荣.基于深度学习的舌象自监督聚类.中国医学物理学杂志,2023,40(1):120-125.[doi:10.3969/j.issn.1005-202X.2023.01.020]
- 13 Liu Z, Lin YT, Cao Y, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9992-10002.
- 14 崔振超.计算机舌诊中舌体分割与纹理分类研究[博士学位论文].哈尔滨:哈尔滨工业大学,2015.
- 15 Ni JH, Yan ZZ, Jiang JH. TongueCaps: An improved capsule network model for multi-classification of tongue color. Diagnostics, 2022, 12(3): 653. [doi: 10.3390/diagnostics12030653]
- 16 卓力,孙亮亮,张辉,等.有噪声标注情况下的中医舌色分类方法.电子与信息学报,2022,44(1):89-98.[doi:10.11999/JEIT210935]
- 17 颜建军,李东旭,郭睿,等.基于二级分类器的齿痕舌分类模型研究.中华中医药杂志,2022,37(4):2181-2185.
- 18 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770-778.
- 19 Deeb A, Debow A, Mansour S, *et al.* COVID-19 diagnosis with deep learning: Adjacent-pooling CTScan-COVID-19 classifier based on ResNet and CBAM. Biomedical Signal Processing and Control, 2023, 86: 105285. [doi: 10.1016/j.bspc.2023.105285]
- 20 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3-19.
- 21 陈思卓,赵萌,石凡,等.过参数卷积与CBAM融合的胸腔积液肿瘤细胞团块分割网络.中国图象图形学报,2023,28(10):3243-3254.[doi:10.11834/jig.220848]
- 22 张晓倩,罗建,杨梅,等.基于改进SE-Net网络与多注意力的脑肿瘤分类方法.西华师范大学学报(自然科学版),2024,45(1):93-101.
- 23 Zeng YN, Lv H, Jiang MF, *et al.* Deep arrhythmia classification based on SENet and lightweight context transform. Mathematical Biosciences and Engineering, 2023, 20(1): 1-17.
- 24 李敏,李学俊,廖竟.融合多层次卷积神经网络的知识图谱嵌入模型.计算机工程与应用.<http://kns.cnki.net/kcms/detail/11.2127.TP.20231127.1401.014.html>.(在线出版).

(校对责编:孙君艳)