

面向中医诊疗的一种深度学习舌像分割模型

卢程^{1,2}, 张锦³, 王小菊², 徐新艳⁴, 文建全⁵

(1. 湖南师范大学信息科学与工程学院, 湖南长沙 410081; 2. 湖南中医药大学第一附属医院, 湖南长沙 410007;
3. 长沙理工大学计算机与通信工程学院, 湖南长沙 410006; 4. 飞腾信息技术有限公司, 天津 300459;
5. 湖南创星科技股份有限公司, 湖南长沙 413100)

摘要: 舌像分割一直是舌诊智能化的难点之一, 特别是开放环境带来的复杂背景、不同类拍摄设备以及光照干扰, 增加了舌像分割高精度与高稳定性的实现难度。针对上述问题, 提出了一种基于DeepLabv3+网络架构的轻量化舌像分割模型。首先, 采用轻量级的MobileNetv2作为骨干特征提取网络, 并嵌入了翻译变体卷积降低模型计算复杂度, 同时提升模型处理舌像的实时性; 其次, 通过坐标注意力机制增强对舌像特征的语义理解, 从而实现对舌体区域的更精确定位; 再次, 通过在空洞卷积金字塔池化模块中嵌入条纹池化, 以更高效地捕获舌体边缘的关键信息; 最后, 通过融合多尺度跨轴注意力机制, 使模型能够更好地聚合局部特征, 以适应不同形态的舌体图像。实验结果表明, 与DeepLabv3+等算法相比, 所提模型精确率达到97.67%、MIoU达到96.87%, 综合考虑精确率、MIoU、推理速度、分割精度、参数量等指标, 所提模型具有良好的性能。

关键词: 智能舌诊; 舌像分割; 轻量化; 注意力机制; 条纹池化

DOI: 10.11907/rjdk.241816

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2026)001-0149-09

扫描二维码阅读全文:



A Deep Learning Tongue Image Segmentation Model for Traditional Chinese Medicine Diagnosis and Treatment

LU Cheng^{1,2}, ZHANG Jin³, WANG Xiaojun², XU Xinyan⁴, WEN Jianquan⁵

(1. School of Information Science and Engineering, Hunan Normal University, Changsha 410081, China; 2. The First Hospital of Hunan University of Chinese Medicine, Changsha 410007, China; 3. School of Computer and Communication Engineering, Changsha University of Science & Technology, Changsha 410006, China; 4. Feiteng Information Technology Co., Ltd, Tianjin 300459, China; 5. Hunan Trasen Science & Technology Co., Ltd, Changsha 413100, China)

Abstract: Tongue image segmentation has always been one of the challenges in the intelligentization of tongue diagnosis, especially the complex background brought by the open environment, different types of shooting equipment, and lighting interference, which increase the difficulty of achieving high precision and high stability of tongue segmentation. In response to the above issues, a lightweight tongue image segmentation model based on the DeepLabv3+ network architecture has been proposed. First, the lightweight MobileNetv2 is used as the backbone feature extraction network, and the translation variant convolution is embedded to reduce the computational complexity of the model and improve the real-time performance of the model in processing tongue images; secondly, the semantic understanding of tongue image features is enhanced through the coordinate attention mechanism, so as to achieve more accurate positioning of the tongue area; thirdly, stripe pooling is embedded in the atrous spatial pyramid pooling module to more efficiently capture the key information of the tongue edge; finally, by integrating the multi-scale monvolutional attention mechanism, the model can better aggregate local features to adapt to tongue images of different shapes. The experimental results show that compared with algorithms such as DeepLabv3+, the proposed model has a precision of 97.67% and

收稿日期: 2024-10-11

基金项目: 湖南省中医药管理局重点课题(C2022039)

作者简介: 卢程(1987-), 男, 湖南师范大学信息科学与工程学院硕士研究生, 研究方向为软件工程、人工智能; 张锦(1979-), 男, 博士, CCF杰出会员, 长沙理工大学计算机与通信工程学院教授、博士生导师, 研究方向为软件工程、人工智能等; 王小菊(1989-), 女, 博士, 湖南中医药大学第一附属医院主治医师, 研究方向为心脑血管疾病防治; 徐新艳(1978-), 硕士, 飞腾信息技术有限公司工程师, 研究方向为高性能计算; 文建全(1969-), 男, 湖南创星科技股份有限公司高级工程师, 研究方向为软件工程、智慧医疗。本文通讯作者: 张锦。

MIoU of 96.87%. Considering the precision, MIoU, inference speed, segmentation accuracy, parameter quantity and other indicators, the proposed model has good performance.

Key Words: intelligent tongue diagnosis; tongue image segmentation; lightweight; attention mechanism; stripe pooling

0 引言

舌像是反映人体内在健康状况的重要标志,基于舌像的舌诊一直是中医四诊(望、闻、问、切)之一。舌诊是中医望诊的重要环节,医生通过观察舌像变化以了解人体生理功能和病理变化,具有高效、无痛、无副作用、简便、快速等优点,是中医学中最重要、应用最广泛的诊断方法之一。经过3 000多年的实践,中医师积累了非常丰富的临床经验,如经典的舌诊八则、《皇帝内经》,揭示了舌体各分区与人体器官健康状况之间的密切关系^[1]。然而,传统的中医舌诊高度依赖于医生的临床经验,诊断具有主观性和不可复性的局限^[2]。为了解决这些问题,研究人员逐步推动舌诊量化研究,引入了图像处理、模式识别等技术以实现舌诊自动化、智能化。自动化舌诊通常首先利用图像分割技术从舌像中提取舌体,然后利用特征提取技术计算舌体特征,最后利用分类器实现最终诊断。因此,舌像分割是实现自动舌诊的关键步骤,舌像分割技术旨在区分舌体与背景,对舌诊的客观化研究和临床应用至关重要^[3]。

随着中医现代化的进展,舌诊量化研究取得了显著成果^[4]。目前,大多已有研究是基于分割的结果对受试者的健康状况进行临床辅助决策,但决策的准确性依赖于图像质量,精准的舌像分割是智能化舌诊不可或缺的前提条件^[5-6]。然而,大多数舌像分割模型都是基于使用单个设备在标准环境下以固定环境背景和距离收集的舌像数据训练模型^[7-9]。这些模型的使用场景只能应用于舌诊仪或专用设备上的舌体分割,受空间和设备条件的局限难以应用于开放环境^[9]。为了使用户更实时、更方便地获取健康信息,将智能医疗算法嵌入移动设备端(如广泛使用的移动智能手机)尤为必要且有效^[10-11]。然而,在开放环境下,移动设备端采集的舌像数据常伴有噪声,导致能获取的有效信息较少,这对舌像分割算法提出了更高要求,既需要抑制噪声,又要能够快速准确地提取有效特征信息,以实现开放环境下舌体的精准分割^[12]。传统的舌体分割方法受限于对高质量图像数据和先验知识的依赖,不足以应对开放环境的挑战^[13]。考虑到采集设备的多样性和开放环境的复杂性,模型应从轻量化设计、特征提取和多尺度融合等角度进行综合设计。

基于上述已有研究成果,针对舌诊智能化问题,本文提出了一种面向中医诊疗的深度学习舌像分割模型。该模型通过将轻量化设计的卷积核嵌入到MobileNetV2主干特征提取网络中,以减少参数量,从而降低模型计算复杂度;利用注意力机制增强模型对舌像特征的提取,有效抑

制非舌体区域干扰;通过在空洞空间金子塔池化(ASPP)模块中嵌入条纹池化层以强化多尺度信息融合,提升边缘特征捕捉能力。本文模型在分割精度和实时性上有较好提升,适用于移动设备,且在自建数据集上取得了较好的分割效果。

1 相关工作

舌像分割的主要工作是将舌体部位从人的脸部、唇部和牙齿等背景中分割出来,对分割后的舌体区域进行特征提取,可以有效地提升分割精度。

传统的舌体分割方法主要是利用舌体的形态特征进行分割。Hu等^[14]提出一种使用图像阈值、灰度投影和主动轮廓模型(Active Contour Models, ACM)的舌像分割算法,通过图像阈值化和形态学运算提取舌像的初始区域,然后采用灰度投影方法定位舌体的上根部位,以此细化舌体边缘区域,最后通过ACM对舌体轮廓进行平滑处理,得到了较为精确的分割结果。吴佳等^[15]提出一种基于分水岭变换与主动轮廓模型相结合的舌体分割算法,通过降采样提高分割速度,并采用分水岭和主动轮廓模型提取舌体的边缘区域,达到对舌体区域分割的目的。Zhang等^[16]提出一种极边检测器与主动轮廓模型的相结合的分割算法,首先使用极边检测器对舌体边缘进行提取,然后利用局部自适应边缘双阈值算法对边缘进行二值化操作,最后采用主动轮廓模型对二值化后的舌像进行分割。

传统的图像分割方法通常仅利用舌体的形态特征信息和固有的先验知识对舌体进行分割,容易受到环境背景和拍摄设备干扰,分割结果不稳定。为了解决这一问题,结合深度学习的特征语义信息和稀疏表达等技术方法相继涌现^[17]。SAM(Segment Anything Model)作为一种通用图像分割模型,在自然场景图像分割任务中取得了显著成果^[18]。结合Transformer的分割模型在图像处理领域也表现优异^[19]。深度学习技术被引入中医舌像分割领域,Qu等^[20]采用SegNet实现了舌像小样本训练,并得到了较为准确的分割结果。李渊彤等^[21]结合FCN和CRF结构提出一种结合上下文特征信息的分割网络,保留了更多的特征信息。宋超等^[22]采用ResNet结构与级联分类网络,实现了对舌体的分割和特征分类。张新锋等^[23]提出一种基于条件随机场的舌像分割方法,实现了对舌体上下文特征的精细分割。

Deeplab框架是由谷歌研究团队开源的深度学习网络,该网络框架采用编解码结构,并推出了空洞卷积、条件随机场(CRF)和空洞空间金子塔池化(ASPP)模块。这些模块充分利用了特征图的多尺度信息,增强了模型捕捉细

节的能力,将深度学习语义分割网络的性能提升到了一个新的水平。Deeplabv3+框架是当今最好的通用分割网络之一,其平滑的分割边缘和分割准确性在许多公开数据集中处于领先地位。但还存在一些问题:①Deeplabv3+编码器采用的Xception结构复杂,需要大量的参数计算和存储,导致拟合速度慢,分割效率低;②在图像分割领域,图像轮廓难以精确捕捉,对复杂环境适应差,相似目标容易误判,导致分割精度低。

2 TinyDeepLab 舌像分割模型

2.1 模型框架

本文基于 Deeplabv3+ 框架,提出了一种舌像分割模型 TinyDeepLab。该模型针对开放环境下的分割任务进行了

轻量化设计与优化,包含 Encoder 和 Decoder 两部分,如图 1 所示。

在 Encoder 模块中采用 MobileNetV2 作为主干网络,通过 TVConv 卷积提取输入图像的有效特征,在降低计算成本的同时,也保留了浅层特征的语义信息。浅层特征经 CA 注意力机制模块增强语义表示,为 ASPP 模块提供丰富的中间特征图。ASPP 模块融合不同扩张率的卷积层($r=1, 6, 12, 18$)和 SP 层,适应多尺度特征融合,SP 层因其具有条行卷积核的特点,可用来处理舌体边缘狭长区域,增强特定方向的特征提取。MSCA 模块对 ASPP 输出执行多尺度大核卷积,捕获高维语义信息,抑制非目标区域干扰。MSCA 通过不同尺寸卷积核捕获多尺度上下文信息,采用轻量化条带卷积处理舌体精细分割,增强模型对不同区域和通道的感知。

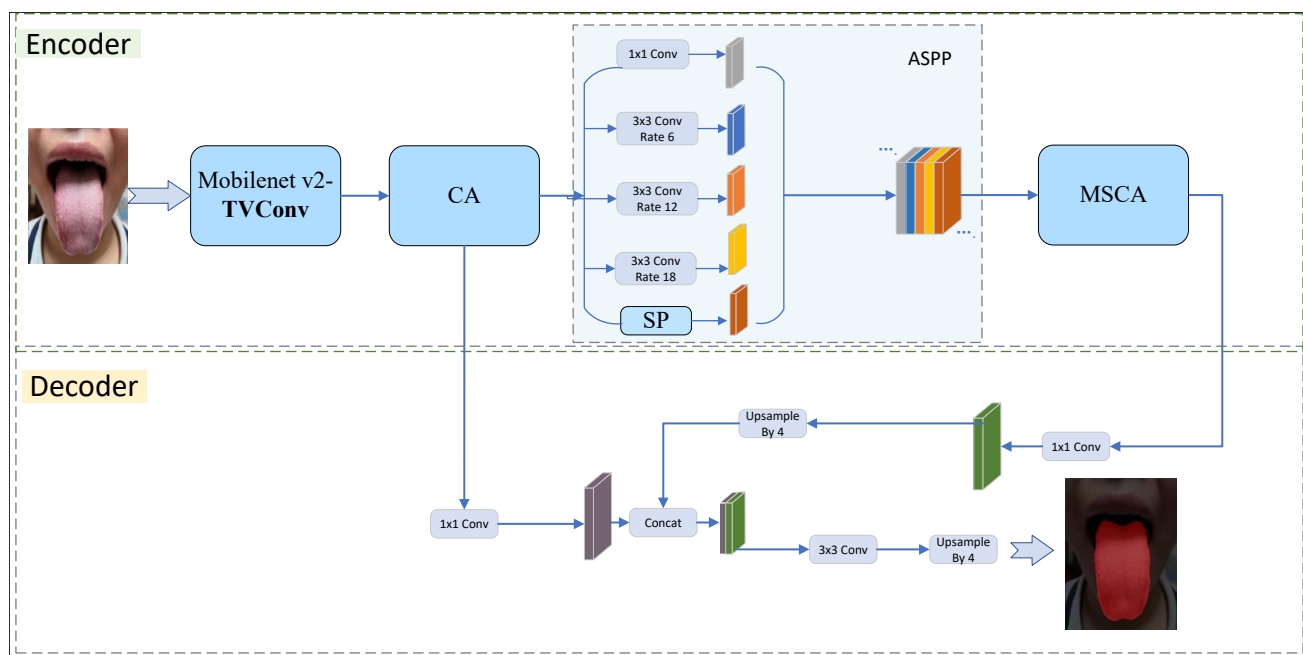


Fig. 1 TinyDeepLab model framework

图 1 TinyDeepLab 模型框架

Decoder 模块使用批量归一化层对 CA 输出特征图进行归一化,通过上采样和卷积操作融合 Encoder 输出与 Decoder 中间特征图,细化语义特征。最终通过上采样操作得到与输入尺寸一致的舌像预测。

2.2 TVConv 卷积

鉴于智能终端对高效轻量化网络设计的迫切需求,MobileNet 系列结构采用深度可分离卷积(DSC)技术,是模型轻量化设计的关键^[24]。MobileNet 通过将标准卷积分解为深度卷积和 1x1 点积,有效减少了模型参数。然而,深度卷积的转换等变性限制了其对不同区域特征的适应性,尤其在开放环境下的舌像分割任务中。这导致模型必须学习大量特定特征匹配过滤器以适应图像特征变化,造成计算资源浪费,不符合开放环境下对模型高效性和适应性的要求^[25]。

为应对这一挑战,本文引入了 TVConv 卷积。与标准卷积相比,它能够自适应地感知图像不同区域,同时显著降低计算复杂度,不影响模型性能。TVConv 卷积的平移方差和跨图像共享特性,使其适用于具有特定布局的应用场景,如开放环境下的舌像分割。

TVConv 的结构设计包括权重生成和应用两个关键过程,如图 2 所示^[26]。权重生成块接收亲和性映射(Affinity Maps),通过卷积层、归一化和激活函数处理,生成能够关注整个图像并建立像素间配对关系的 Affinity Maps。这些映射在不同输入图像间共享参数,减少了模型参数量。TVConv 的权重生成块在自适应性和计算开销间取得平衡,允许模型在不牺牲推理速度的前提下,通过适当参数化增强空间自适应性。

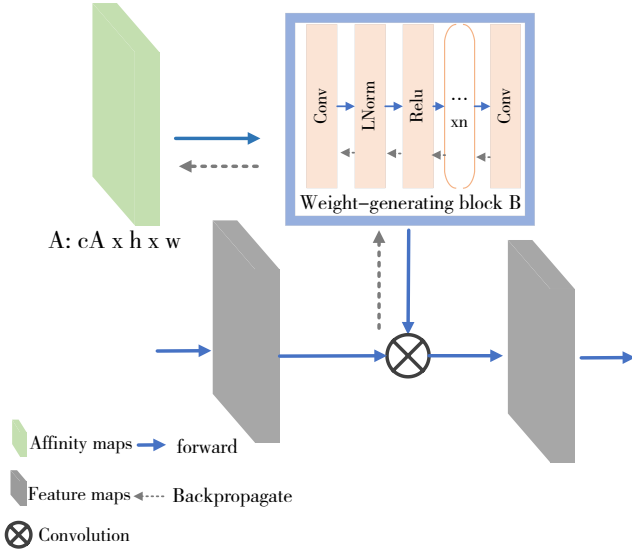


Fig. 2 TVConv structure

图2 TVConv结构

2.3 CA 注意力机制

模型虽在图像分割上表现出色,但特征整合主要在通道维度,限制了对特征重要性及空间关联的深入挖掘,影响了目标空间位置的精确识别。为克服这些局限,本文在DeepLabv3+基础上引入了坐标注意力(Coordinate Attention, CA)机制,以期从通道和空间两个维度同步提取并强化特征的语义信息。

Coord Attention Blocks 在特征提取阶段引入注意力机制,对每个像素点在通道和空间两个维度上赋予动态权重。该模块由 Coord Information Embedding 和 Coord Attention Generation 两部分组成,如图3所示^[27]。

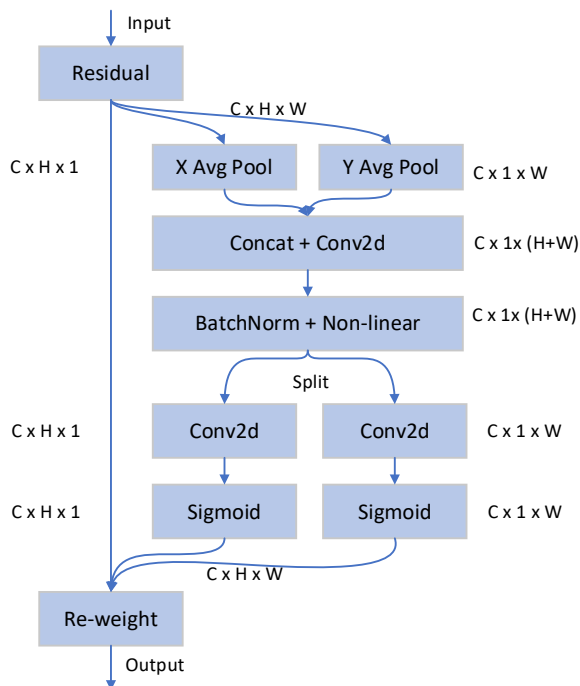


Fig. 3 Coord Attention block

图3 Coord Attention 模块

Coord Information Embedding 负责编码空间位置信息,通过水平方向捕获空间依赖性,同时在垂直方向上保持位置信息的精确性。Coord Attention Generation 则利用这些编码信息,有效识别通道间的相互关系。

CA 机制通过水平和垂直方向的注意力共同作用,并集中于图像关键行和列,以提升舌像空间位置定位准确性,实现更精细的分割效果。

2.4 SP 条纹池化层

在中医舌像分割任务中,舌体往往呈现长条形的形态特征,传统方形池化层难以有效捕捉舌体边缘的狭长语义信息。为解决此问题,本文引入了条纹池化层(Stripe Pooling, SP)^[28]。与基于注意力的模块相比,SP 具有轻量化优势,易于集成,显著增强了模型捕获远程空间依赖关系的能力,并提升了通道间特征利用效率,其结构示意图如图4所示。SP 的优势主要体现在:①SP 采用长条形状的卷积核设计,有利于在独立区域内沿长度方向建立特征关系,从而更有效地捕获舌体边缘信息;②SP 长而窄的结构设计使其能同时关注全局特征与局部细节,减少不相关区域信息干扰,增强了模型对与远程空间依赖关系的建模能力,提升了通道间依赖性特征的提取效率;③在处理复杂背景和多样形态的舌像图像时,SP 模块能使模型更精确地提取出舌体轮廓,尤其在自然环境下采集的数据中优势明显。SP 的集成提高了分割精度,同时保持了计算效率,对实时舌像分割应用具有重要意义。

2.5 MSCA 多尺度跨轴注意力机制

尽管 CA 注意力机制在提升特征语义理解方面效果显著,能够聚焦通道权重和空间位置信息,但在自然环境中,由于舌像边缘信息复杂且存在多种干扰,仅采用 CA 难以完全准确提取边缘信息。为此,本文在模型中引入了多尺度跨轴注意力 MSCA^[29]。

MSCA 通过多尺度特征嵌入,增强了模型对局部特征的聚合能力,能够更有效地捕获上下文信息,尤其在边缘区域。MSCA 包含 3 个核心部分:①深度卷积聚合局部特征,捕获深层语义信息;②多分支深度条带卷积捕获多尺度上下文信息;③1×1 卷积对通道间关系建模,将输出作为注意力权重,对输入特征重新加权。

MSCA 的设计使模型能够更有效地处理不同尺寸和形状的舌像数据,适应多变场景。MSCA 的数学表达式如式(1)、式(2)所示。

$$Att = Conv_{1 \times 1} \left(\sum_{i=0}^3 Scale_i (Conv(F)) \right) \quad (1)$$

$$Out = Att \otimes Out \quad (2)$$

其中, F 表示输入特征, Att 和 Out 分别代表注意力地图和输出特征, $\otimes$ 表示元素乘法运算, $Scale_i$ 表示第几个分支, $i \in \{0, 1, 2, 3\}$ 表示图5中的分支。

在每个分支中,利用两个深度条带卷积(核大小分别为7、11和21)模拟大核深度卷积,以增强对长条状舌体区域的关注,并提升模型识别能力,这种多尺度卷积显著提

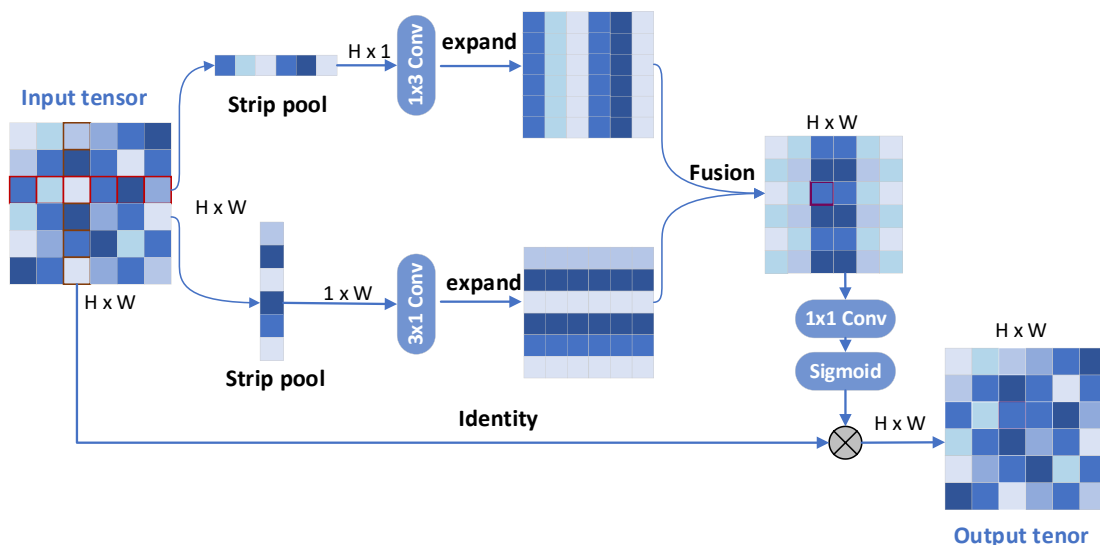


Fig. 4 SP stripe pooling layer structure

图 4 SP 条纹池化层结构

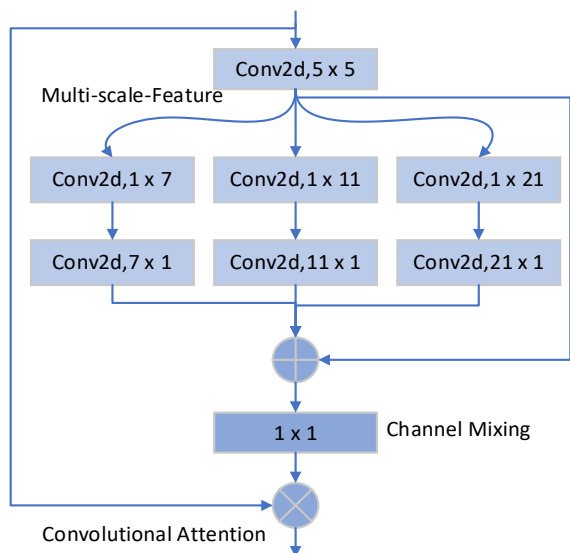


Fig. 5 MSCA block

图 5 MSCA 模块

高了模型在复杂环境下分割舌像的精度和鲁棒性。

3 实验与结果分析

3.1 评价指标

医学图像分割对疾病诊断和治疗计划制定至关重要,全面评价分割算法性能及其临床应用潜力是关键研究问题。本文采用医学图像分割领域广泛认可的评价指标,全面评估所提算法性能。具体评价指标包括:

(1)平均交并比(MIoU)。MIoU 衡量分割结果与真实标注重叠程度,是常用分割精度指标,如式(3)所示。

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (3)$$

(2)类别平均像素精确率(MPA)。MPA 计算每个像素

点分类准确性,评估模型整体性能,如式(4)所示。

$$MPA = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij}} \quad (4)$$

(3)召回率(Recall)。Recall 反映模型识别正样本的能力,对于医学图像分割尤为重要,可确保舌体区域正确识别,如式(5)所示。

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

(4)精确率(Precision)。Precision 衡量模型预测为正样本中实际为正样本比例,反映模型准确性,如式(6)所示。

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

此外,还考虑了浮点运算次数(FLOPs)和参数量(Params)。衡量模型计算复杂度和可训练参数数量是评价模型轻量化的重要指标,对实时应用场景至关重要。本文通过对上述评价指标的定量分析,旨在全面评估分割算法的性能,并探讨其临床应用潜力。

3.2 数据集与环境配置

本文实验数据集来源于湖南中医药大学第一附属医院临床数据和开放环境下不同设备采集的数据。临床数据是在标准化实验环境中,通过舌诊仪采集的 2 000 张临床舌像,设备采集的数据则在非标准化或开放环境下通过手机等移动设备采集的 1 000 张舌像图像,总共 3 000 张。采用 LabelMe 工具在专业医师的指导下逐一标注,将其分为舌体和背景两大类,部分原始图像如图 6 所示,对应的标注图像如图 7 所示。

为确保统计意义和训练有效性,数据集按照 7:2:1 的比例分为训练集、验证集和测试集,具体划分如表 1 所示。所有舌图像经预处理统一调整至 512×512 像素,以满足神经网络输入要求,保证输入图像尺寸的一致性,为后续分析和训练提供标准化基础。

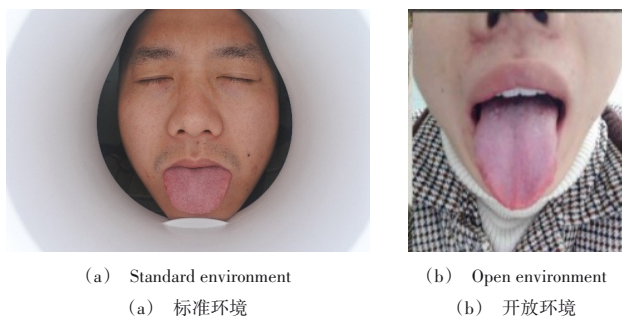


Fig. 6 Original image

图6 原始图像

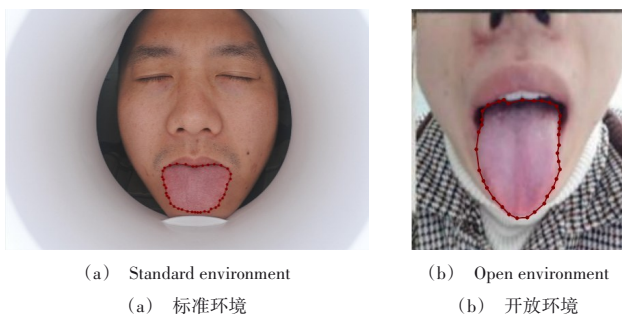


Fig. 7 Annotated images

图7 标注图像

Table 1 Tongue dataset

表1 舌像数据集

采集数量	采集环境	训练集	验证集	测试集
2 000	标准环境	1 400	400	200
1 000	开放环境	700	200	100
3 000(总计)	--	2 100	600	300

实验所依托的硬件配置包括 12 GB 显存的 NVIDIA GeForce GTX 3060 显卡和 Intel i7 处理器,软件环境则搭建在 Ubuntu 22.04 操作系统上,开发工具为 PyCharm 3.7,深度学习框架选用 PyTorch 1.7.0,配合 Cuda 11.6 进行优化。在这样的软硬件条件下,针对中医舌像的语义分割网络进行了训练与性能评估。超参数设置如表 2 所示,根据舌图像特点以及 GPU 性能与网络架构的特点,确定 Batch Size 为 64,优化策略为 Adam 时可获得最优训练和测试效果,学习率为 0.00 005 时,模型损失函数能够得到稳定且快速的收敛。如图 8 所示,模型迭代到 300 epoch 时,模型 MIoU 值基本保持稳定,不在宽幅震荡。

3.3 对比实验结果分析

本文以 Deeplabv3+(xception)为基线算法,分别对比了 UNet、UNet++、PSPNet、HRNet、DeepLabv3、DeepLabv3+

Table 2 Training parameters

表2 训练参数

Parameters	Value
Input Size	512×512
Learning Rate	5×10^{-4}
Batch Size	64
Minbatch Size	8
Optimizer	Adam

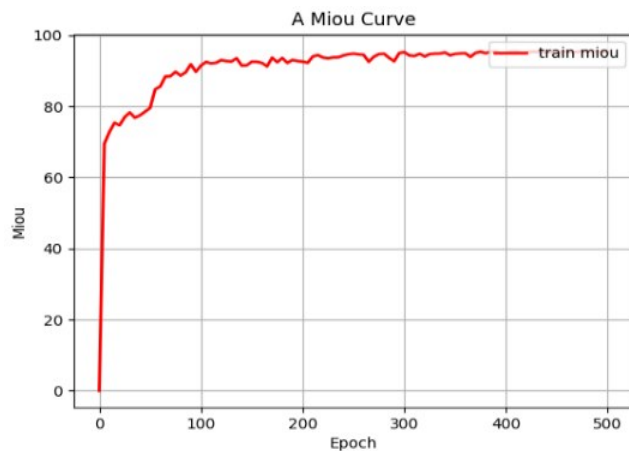


Fig. 8 MIoU curve

图8 MIoU 曲线

(mobilenet)和 Ours 的分割结果,为了更好地对比分割结果,通过将预测图和原图相叠加的方式,对结果进行可视化展示,并对突出区域进行了放大处理,如图 9 所示。

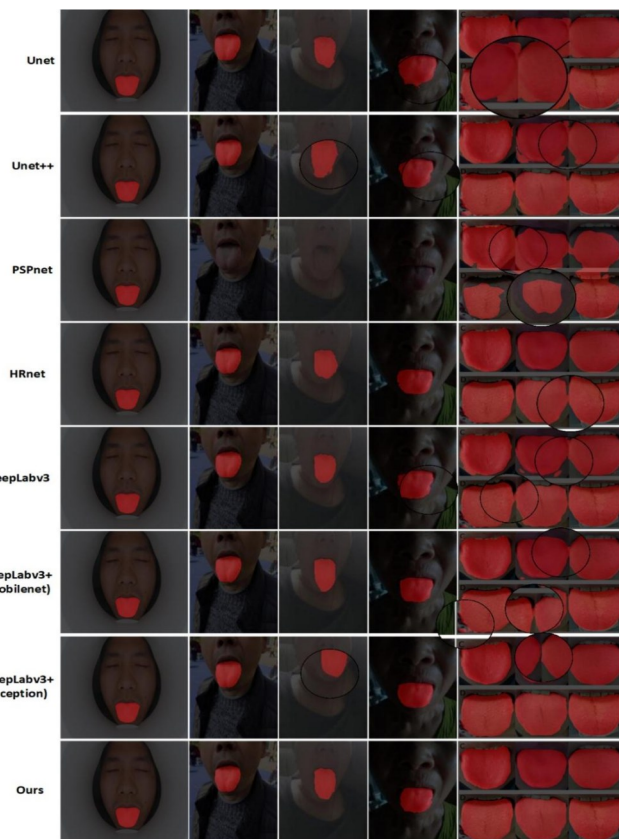


Fig. 9 Comparative experimental segmentation results

图9 对比实验分割结果

UNet在标准和自然场景下表现良好,但在多目标场景下对边缘定位不准确。PSPNet难以精确分割单舌像,HRNet、UNet和UNet++在两种环境下性能相似,但多目标分割存在区域连接问题。DeepLabv3和DeepLabv3+在不同环境下表现优越,但多目标分割仍有过度分割和边缘问题。

通过中医舌像分割的主观效果比较可知,UNet、UN-

et++、PSPNet、HRNet、DeepLabv3 和 DeepLabv3+在边缘区域存在毛糙或粘连噪声,影响了分割质量。而本文所用模型在边缘区域表现优异,有效处理了舌体边缘的毛糙问题。

指标评估结果如表 3 所示,UNet 和 UNet++在召回率上表现优异,分别为 97.54%、97.16%,但计算量过大,不适用于移动设备。PSPNet 的 Params、FLOPs、推理速度表现最佳,但精确率、召回率、MPA、MIoU 均低于 90%,整体性能较差。HRNet 在多类别分割上表现优异,但推理速度仅为 15.34f/s,影响实时性。DeepLabv3 和 DeepLabv3+性能均衡,但计算资源需求仍然较高。然而,DeepLabv3+的两个不同版本有着较大差异,MobileNet 相较于 Xception,在保持模型精度与 MIoU 的同时,降低了参数量与计算复杂度,参数量降为 Xception 的 1/10,计算复杂度降为 1/3,使得模型更加轻量化,有利于部署到移动边缘设备上。

Table 3 Comparative experimental results
表 3 对比实验结果

Algorithm	Precision	Recall	MPA	MIoU	Params/	FLOPs	Detection
	/%	/%	/%	/%	MB	/G	Speed/ f/s
UNet	97.80	97.54	98.55	95.98	24.892	452.310	36.38
UNet++	97.41	97.16	97.87	95.86	47.190	306.547	34.67
PSPNet	74.84	71.16	84.49	76.38	2.376	6.031	88.26
HRNet	97.82	97.40	98.6	97.45	29.538	79.915	15.34
DeepLabv3	96.23	95.37	97.89	95.78	55.813	169.867	60.32
DeepLabv3+ (mobilenet)	97.24	96.66	98.17	96.69	5.813	52.867	86.20
DeepLabv3+ (xception)	97.22	96.64	98.16	96.68	54.709	166.841	62.51
Ours	97.67	96.57	98.15	96.87	6.156	47.333	84.33

在实验数据集上,本文模型实现了 97.67% 的精确率和 96.87% 的 MIoU,虽然分割精度略低于 HRNet,但仍然优于其他算法。HRNet 的推理速度仅为本文模型的 18.19%,实时性最低。与 DeepLabv3+(Xception)相比,尽管召回率和 MPA 分别下降了 0.07% 和 0.01%,但在精确率、MIoU 和推理速度上分别提升了 0.46%、0.2% 和 34.9%,这表明基于 DeepLabv3+的改进增强了模型分割精度和实时处理能力。

本文模型的 Params 和 FLOPs 分别为 6.156 MB 和 47.333 G,尽管在计算复杂度上表现不如 PSPNet,但较其

他算法具有明显优势,且在精确率、召回率、MPA 和 MIoU 上相较于 PSPNet 分别实现了 30.5%、35.7%、16.2% 和 26.8% 的显著提升。与 DeepLabv3+(Xception)相比,Params 和 FLOPs 分别减少了 88.7% 和 71.6%,这表明模型在维持高精度的同时,大幅度降低了计算资源的需求。

3.4 消融实验结果分析

本文验证了所提 CA、MSCA、SP 模块对模型分割精度的影响,使用引入了 TVConv 卷积的 MobileNetV2 作为主干特征提取网络开展实验。结果如表 4 所示(“√”表示所标注的模块用于实验,无“√”表示未添加该模块),分析了不同模块组合以及单一模块对分割效果的影响,并比较了各组合算法的性能。

采用 TVconv 卷积降低模型参数量与计算复杂度,模型在牺牲一定精确度的情况下,将模型参数量与计算复杂度降为 5.813 MB、46.948 G,并将模型检测速度提升至 99.485 f/s。采用 CA 注意力机制的模型在精确率上表现良好,为 97.24%,然而实际分割中仍存在过度分割,以及舌体边缘区域粘连和噪声问题。相比之下,采用 MSCA 注意力机制的模型在召回率上表现更佳,达到了 96.93%,但其 MIoU 值降低了 0.25%。当模型仅嵌入 SP 条纹池化模块时,MIoU 值同样略有下降。

当双重融合 CA 和 MSCA 注意力机制时,模型在精确率、MIoU 值上相较于单一融合模型均略有下降。这表明在特征提取过程中,注意力机制可能学习到了一些干扰信息的特征,导致实际分割过程中舌体边缘区域出现毛糙和噪声,以及舌体与唇部边缘的粘连现象。

在结合 CA 注意力机制和 SP 条纹池化层模块时,精确率、召回率分别提高了约 0.65%、0.29%。SP 条纹池化层与 MSCA 注意力机制结合后,MPA 显著提高。最终,双重融合注意力机制和 SP 条纹池化层模块的模型在精确率上达到了 97.67%,分别比单一融合注意力机制 CA 和 MSCA 提高了 0.44% 和 0.75%。此外,该模型在 MIoU 值方面也表现更佳,达到了 96.87%,优于其他 7 种算法。

综上所述,引入 CA、MSCA、SP 和 TVConv 等技术能够显著提升舌像分割性能,同时保持较低的计算复杂度和较高的实时性。

Table 4 Ablation experimental results
表 4 消融实验结果

CA	MSCA	SP	TVConv	Precision /%	Recall/%	MPA/%	MIoU /%	Params/MB	FLOPs /G	Detection Speed/ f/s
			√	96.26	95.54	97.60	95.70	5.813	46.948	99.485
√			√	97.24	96.28	96.19	95.33	6.009	47.044	87.764
	√		√	96.94	96.93	97.62	95.46	6.101	47.328	79.205
		√	√	96.52	95.93	97.74	95.54	6.230	47.143	85.820
√	√		√	96.84	96.35	97.07	95.16	6.111	47.328	80.402
√		√	√	96.89	95.82	97.01	95.24	6.240	47.144	85.695
	√	√	√	95.74	96.53	97.98	95.52	6.331	47.428	80.302
√	√	√	√	97.67	96.57	98.15	96.87	6.156	47.333	84.335

图 10 展示了 8 种模型变体在不同环境下分割舌像的性能。

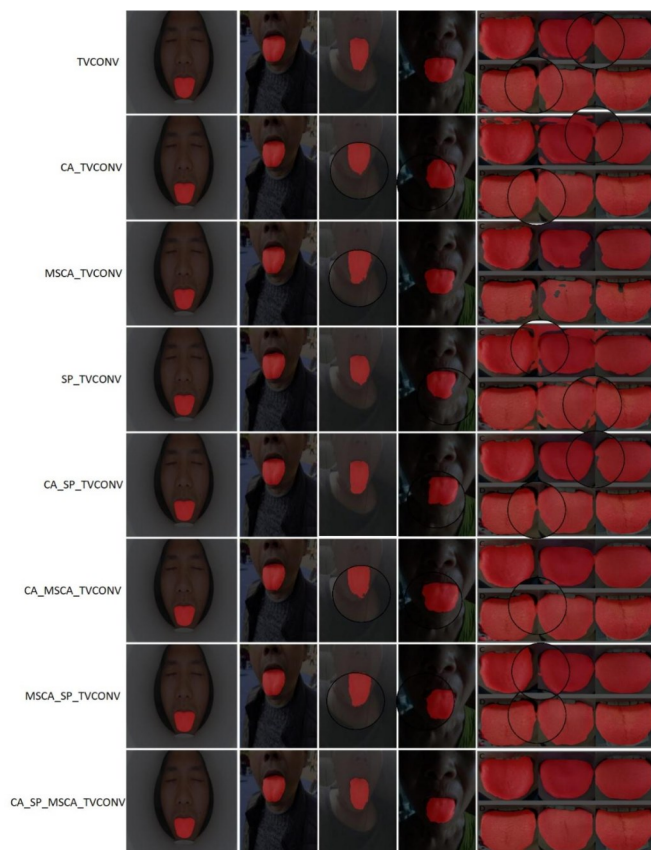


Fig. 10 Ablation experimental segmentation results

图 10 消融实验分割结果

研究表明,尽管 CA 和 MSCA 注意力机制能够捕捉通道权重和空间位置信息,但在处理长距离依赖和条状目标时效果欠佳。在使用了 TVConv 结构的 DeepLabv3+模型的基础上,无论是单独融合 CA 和 MSCA,还是双重融合这两种机制,在多目标、低光照或反光条件下的分割精度均显不足。而仅引入 SP 层,则分割边缘易出现过度分割现象,导致舌体与边缘粘连,无法实现精确分割。因此,本文将 CA、MSCA 和 SP 的优势整合到使用了 TVConv 结构的 DeepLabv3+模型中,从而在开放环境下实现了中医舌像的精确分割。

4 结语

本文提出的 TinyDeepLab 舌像分割模型,在开放环境下具备快速、准确且高鲁棒性的特性。通过消融实验和可视化分析,深入探讨了模型性能提升机制,增强了模型可解释性。该模型在轻量化、推理速度和分割精度方面均表现出色,证明了其在移动设备上实现远程中医智能舌诊的可行性,为中医舌诊现代化和智能化提供了新路径。

由于目前自建数据集规模有限,未来将收集更广泛的数据,以评估模型在实际应用环境中的准确性,并增强其

泛化能力。同时,将模型与中医舌像的证候特征信息相结合,形成初步诊断信息,从而提升医师诊断效率。为了适应移动和边缘设备的部署需求,还将进一步探索模型轻量化途径,以满足日常健康监测需求,助力医疗大数据中心建立,推动中医数字化进程。

参考文献:

- [1] NIU F Q. Research on tongue image classification in traditional Chinese medicine based on deep learning [D]. Chengdu: University of Electronic Science and Technology of China, 2021.
牛富泉. 基于深度学习的中医舌像分类模型研究 [D]. 成都: 电子科技大学, 2021.
- [2] LU Y X, LI X G, ZHANG H, et al. Review on tongue image segmentation technologies for traditional Chinese medicine: methodologies, performances and prospects [J]. Acta Automatica Sinica, 2021, 47(5): 1005-1016.
卢运西, 李晓光, 张辉, 等. 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] GUO J, YANG Y, WU Q, et al. Adaptive active contour model based automatic tongue image segmentation [C]//2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2016: 1386-1390.
- [4] WU K, ZHANG D. Robust tongue segmentation by fusing region-based and edge-based approaches [J]. Expert Systems with Applications, 2015, 42(21): 8027-8038.
- [5] LIN B, XIE J, LI C, et al. Deeptongue: tongue segmentation via resnet [C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018: 1035-1039.
- [6] HU Y, WEN G, LIAO H, et al. Automatic construction of Chinese herbal prescriptions from tongue images using CNNs and auxiliary latent therapy topics [J]. IEEE Transactions on Cybernetics, 2021, 51(2): 708-721.
- [7] HUANG X, ZHANG H, ZHUO L, et al. TISNet-enhanced fully convolutional network with encoder-decoder structure for tongue image segmentation in traditional Chinese medicine [J]. Computational and Mathematical Methods in Medicine, 2020(1): 6029258.
- [8] GAO S, GUO N, MAO D. LSM-SEC: tongue segmentation by the level set model with symmetry and edge constraints [J]. Computational Intelligence and Neuroscience, 2021(1): 6370526.
- [9] LI L, LUO Z, ZHANG M, et al. An iterative transfer learning framework for cross-domain tongue segmentation [J]. Concurrency and Computation: Practice and Experience, 2020, 32(14): e5714.
- [10] ZHANG Q, BAI C, YANG L T, et al. A unified smart Chinese medicine framework for healthcare and medical services [J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2019, 18(3): 882-890.
- [11] TANIA H M, LWIN K, HOSSAIN A M. Advances in automated tongue diagnosis techniques [J]. Integrative Medicine Research, 2019, 8(1): 42-56.
- [12] ZHANG X, ZHAO H. Hyperspectral-cube-based mobile face recognition: a comprehensive review [J]. Information Fusion, 2021, 74: 132-150.
- [13] HU X, LIU Z, YANG X, et al. An unsupervised tongue segmentation algorithm based on improved gPb-OWT-UCM [J]. Journal of Medical Imag-

- ing and Health Informatics, 2021, 11(3): 688–696.
- [14] LIU W X, HU J M, LI Z Y, et al. Tongue image segmentation via thresholding and gray projection[J]. KSII Transactions on Internet and Information Systems, 2019, 13(2): 945–961.
- [15] WU J, ZHANG Y H, BAI J, et al. Tongue contour image extraction using a watershed transform and an active contour model[J]. Journal of Tsinghua University (Science and Technology Edition), 2008, 48(6): 1040–1043. 吴佳, 张永红, 白净, 等. 基于分水岭变换和主动轮廓模型的舌体轮廓图像分割[J]. 清华大学学报(自然科学版), 2008, 48(6): 1040–1043.
- [16] ZHANG H Z, ZUO W M, WANG K Q, et al. A snake-based approach to automated segmentation of tongue image using polar edge detector[J]. International Journal of Imaging Systems and Technology, 2006, 16(4): 103–112.
- [17] RAVI N, GABEUR V, HU Y T, et al. SAM 2: segment anything in images and videos[DB/OL]. <https://arxiv.org/abs/2408.00714>.
- [18] HAN K, WANG Y, CHEN H, et al. A survey on vision transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 87–110.
- [19] YU Y, WANG C, FU Q, et al. Techniques and challenges of image segmentation: a review[J]. Electronics, 2023, 12(5): 1199.
- [20] QU P L, ZHANG H, ZHUO L, et al. Automatic tongue image segmentation for traditional Chinese medicine using deep neural network[C]//Intelligent Computing Theories and Application: 13th International Conference, 2017: 247–259.
- [21] LI Y T, LUO Y S, ZHU Z M. Tongue image analysis in traditional Chinese medicine based on deep learning[J]. Computer Science, 2020, 47(11): 148–158. 李渊彤, 罗裕升, 朱珍民. 基于深度学习的舌象特征分析[J]. 计算机科学, 2020, 47(11): 148–158.
- [22] SONG C, WANG B, XU J T. A tongue image classification method based on deep transfer learning[J]. Computer Engineering and Science, 2021, 43(8): 1488–1496. 宋超, 王斌, 许家伦. 基于深度迁移学习的舌象特征分类方法研究[J]. 计算机工程与科学, 2021, 43(8): 1488–1496.
- [23] ZHANG X F, GUO Y T, CAI Y H, et al. Tongue image segmentation algorithm based on deep convolutional neural network and fully conditional random fields[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(12): 2364–2374. 张新锋, 郭宇桐, 蔡轶珩, 等. 基于 DCNN 和全连接 CRF 的舌图像分割算法[J]. 北京航空航天大学学报, 2019, 45(12): 2364–2374.
- [24] LU G, ZHANG W, WANG Z. Optimizing depthwise separable convolution operations on GPUs[J]. IEEE Transactions on Parallel and Distributed Systems, 2021, 33(1): 70–87.
- [25] WANG Z, MIAO Z, HU J, et al. Adaptive convolutions with per-pixel dynamic filter atom[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 12302–12311.
- [26] HUA K, FANG X, TANG Z, et al. DCAM-NET: a novel domain generalization optic cup and optic disc segmentation pipeline with multi-region and multi-scale convolution attention mechanism[J]. Computers in Biology and Medicine, 2023, 163: 107076.
- [27] ZHU L, XIN G, WANG X, et al. A fast tongue detection and location algorithm in natural environment[J]. Computers, Materials & Continua, 2022, 73(3): 4727–4742.
- [28] LU C, XIA M, LIN H. Multi-scale strip pooling feature aggregation network for cloud and cloud shadow segmentation[J]. Neural Computing and Applications, 2022, 34(8): 6149–6162.
- [29] GUO M H, LU C Z, HOU Q, et al. SegNeXt: rethinking convolutional attention design for semantic segmentation[J]. Advances in Neural Information Processing Systems, 2022, 35: 1140–1156.

(责任编辑: 孙娟)