Software Engineering and Applications 软件工程与应用, 2026, 15(2), 190-204
Published Online April 2026 in Hans. https://www.hanspub.org/journal/sea
https://doi.org/10.12677/sea.2026.152019

Hans汉斯

基于LAE-DeepLabv3+的舌象图像分割方法

李月月[1]，石　文[1*]，曾庆红[1]，张静宇[2]，魏楚婷[1]，王皓宇[1]，孙雅星[1]，郑　涛[1]

[1]天津商业大学信息工程学院，天津
[2]天津市南开医院，天津

收稿日期：2026年3月8日；录用日期：2026年4月8日；发布日期：2026年4月21日

摘　要

针对舌象图像分割中模型参数量大、舌体边缘模糊及易受背景干扰等问题，提出一种轻量化高精度分割模型LAE-DeepLabv3+。首先，采用EfficientViT替换主干网络，大幅降低模型计算复杂度；其次，在低层特征分支引入EMA (Efficient Multi-scale Attention)模块，增强网络对舌体边界及局部细节的感知能力；最后，在ASPP (Atrous Spatial Pyramid Pooling)模块中增加水平与垂直条带池化分支，强化方向性上下文与长程依赖建模。在自建舌象数据集上的实验结果表明，该模型的IoU和Dice分别达到96.47%和98.20%，参数量仅为12.12 M。该方法有效实现了分割精度与轻量化的平衡，为舌象客观化分析提供了可靠支持。

关键词

舌象分割，DeepLabv3+，轻量化网络，EMA模块，多尺度特征融合

Tongue Image Segmentation Method Based on LAE-DeepLabv3+

Yueyue Li[1], Wen Shi[1*], Qinghong Zeng[1], Jingyu Zhang[2], Chuting Wei[1], Haoyu Wang[1], Yaxing Sun[1], Tao Zheng[1]

[1]School of Information Engineering, Tianjin University of Commerce, Tianjin
[2]Tianjin Nankai Hospital, Tianjin

Received: March 8, 2026; accepted: April 8, 2026; published: April 21, 2026

Abstract

To address the challenges of large parameter volumes, blurred tongue edges, and background

*通讯作者。

文章引用: 李月月, 石文, 曾庆红, 张静宇, 魏楚婷, 王皓宇, 孙雅星, 郑涛. 基于 LAE-DeepLabv3+的舌象图像分割方法[J]. 软件工程与应用, 2026, 15(2): 190-204. DOI: 10.12677/sea.2026.152019

interference in tongue image segmentation, a lightweight and high-precision segmentation model, LAE-DeepLabv3+, is proposed. First, the EfficientViT architecture is employed to replace the original backbone network, significantly reducing computational complexity. Second, an EMA (Efficient Multi-scale Attention) module is integrated into the low-level feature branch to enhance the network's perception of tongue boundaries and local details. Finally, horizontal and vertical strip pooling branches are added to the ASPP (Atrous Spatial Pyramid Pooling) module to strengthen directional context and long-range dependency modeling. Experimental results on a self-built tongue image dataset demonstrate that the model achieves an IoU of 96.47% and a Dice coefficient of 98.20%, with a parameter count of only 12.12 M. This method effectively balances segmentation accuracy and model lightness, providing reliable support for the objective analysis of tongue diagnosis.

Keywords

Tongue Image Segmentation, DeepLabv3+, Lightweight Network, EMA Module, Multi-Scale Feature Fusion

1. 引言

中医通过望、闻、问、切四种诊断方法来诊断疾病。其中，舌诊最为重要。中医理论认为，舌通过经络与五脏相连。人体的脏腑、气血、津液的虚实以及疾病的变化，都可以客观地反映在舌象上[1]。医生通过观察患者的舌苔颜色、舌苔质地、舌质颜色、舌质形态，并结合问诊和脉诊，进行辨证论治。因此，舌诊是临床医生辅助诊疗疾病的重要依据[2]。

随着计算机视觉与人工智能技术的不断发展，利用图像分析方法实现舌象客观化、标准化处理，已成为中医智能辅助诊断的重要研究方向。舌象图像分割作为舌象分析中的基础环节，其任务是从原始图像中准确提取舌体区域，为后续的舌色分类、舌苔分析、裂纹识别及疾病辅助诊断提供可靠前提。

舌像分割方法可分为传统分割方法和基于深度学习的分割方法。传统分割方法多基于阈值[3]、边缘检测[4] [5]、活动轮廓模型[6]及颜色空间转换[7] [8]等传统图像处理技术。这类方法通常依赖人工设计特征，虽然在特定场景下具有一定效果，但对光照变化、背景干扰以及舌体与嘴唇、牙齿之间边界模糊等问题较为敏感，鲁棒性和泛化能力有限。基于深度学习的分割效果明显优于传统机器学习。近年来，许多学者对基于深度学习的舌像分割模型进行了研究。Long 等人[9]提出了全卷积网络，首次将深度学习算法引入该领域。为了提高分割精度，相继提出了基于编码器－解码器结构的网络模型，如 U-net [10]和 Segnet [11]。Cai 等人[12]通过改进损失函数来提高模型的分割性能。Huang 等人[13]利用残差网络和扩展残差网络结合特征金字塔来提取多尺度信息，在舌像分割方面也取得了较好的性能。

尽管现有研究已经取得了较好的分割效果，但仍存在一些不足：一方面，部分模型依赖较重的主干网络，参数量较大、计算开销较高，不利于后续轻量化部署与实际应用；另一方面，在舌体边缘与背景过渡区域、舌体上部与牙齿接触区域以及局部细节区域，仍容易出现边界不清晰、轮廓不完整或局部误分的现象。

针对上述问题，本文以 DeepLabv3+为基础，提出一种改进的舌象分割模型 LAE-DeepLabv3+。首先，采用更轻量的 EfficientViT 替换原始主干网络，以降低模型复杂度并提高特征提取效率；其次，在低层特

征分支中引入 EMA 模块，增强浅层特征的细节表达能力，提升模型对舌体边界及局部区域的感知能力；最后，对原有 ASPP 结构进行增强，在保留多尺度空洞卷积分支的基础上增加水平与垂直条带分支，以进一步强化对多尺度上下文信息及长程依赖特征的建模能力。

2. DeepLabv3+基础模型

DeepLabv3+是一种经典的语义分割网络，在 DeepLab 系列模型[14]-[17]的基础上进一步引入了解码器结构，能够在保持较强语义特征提取能力的同时改善目标边界恢复效果。该网络整体采用编码器—解码器框架，主要由主干特征提取网络、空洞空间金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)模块以及解码器三部分组成，其网络结构如图 1 所示。

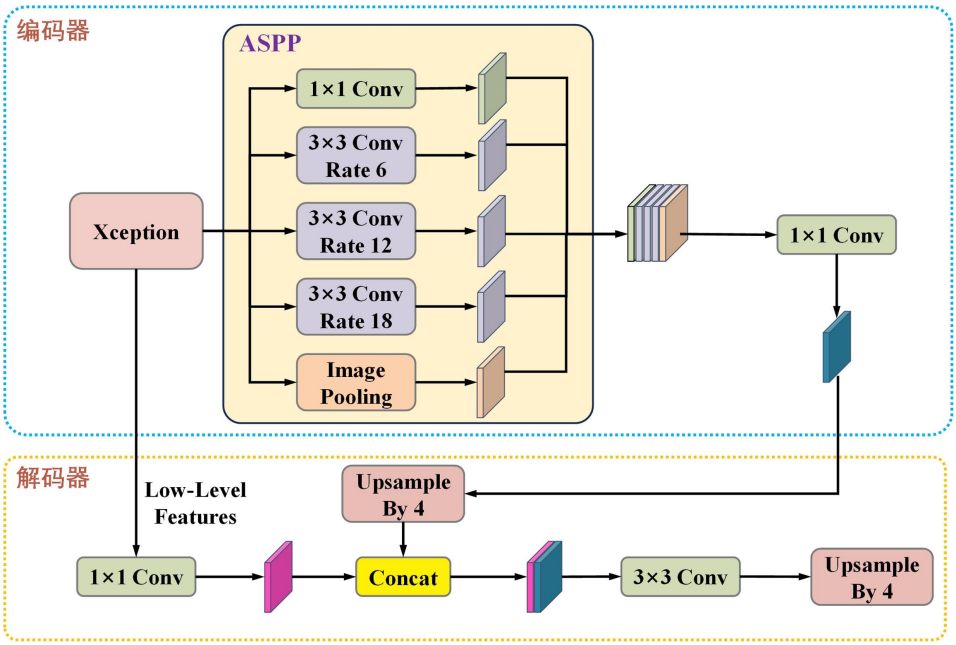

Figure 1. Network structure diagram of DeepLabv3+ model
图 1. DeepLabv3+模型的网络结构图

在编码器部分，DeepLabv3+通过主干网络对输入图像进行逐层特征提取，获得高层语义特征和低层细节特征。其中，高层特征具有更大的感受野和更丰富的语义信息，能够表征目标的整体区域分布；低层特征则保留了更多边缘、纹理和位置信息，有助于后续分割结果的边界恢复。为了在不显著增加参数量的前提下扩大感受野，DeepLabv3+在编码阶段引入了空洞卷积。空洞卷积通过在卷积核元素之间插入空洞，有效扩大了卷积核的感受范围，使模型能够在保持特征图分辨率的同时获取更充分的上下文信息，这对于语义分割任务尤为重要。在高层语义特征提取后，DeepLabv3+采用 ASPP 模块进行多尺度上下文建模。ASPP 模块通常由一个1×1卷积分支、多个不同膨胀率的3×3空洞卷积分支以及一个全局平均池化分支组成。不同膨胀率的空洞卷积可以从不同尺度提取特征信息，从而增强模型对目标大小变化和复杂背景的适应能力；全局平均池化分支则用于补充全局上下文信息，提升网络对整体场景的理解能力。各分支输出特征在通道维度上进行拼接后，再通过卷积进行融合，形成具有较强多尺度表达能力的高层特征表示。

在解码器部分，DeepLabv3+首先对 ASPP 输出的高层特征进行上采样，然后与来自编码器浅层的低层特征进行融合。由于浅层特征通道数通常较多，为降低计算冗余并突出关键信息，网络先通过1×1卷积

对低层特征进行降维，再与上采样后的高层特征进行拼接。随后，经过若干卷积操作进一步融合语义信息与空间细节信息，最后通过再次上采样获得与输入图像大小一致的分割结果。相较于仅依赖编码器输出的分割模型，DeepLabv3+的解码器结构能够更好地恢复目标轮廓和边缘细节，因此在医学图像分割、场景分割等任务中得到了广泛应用。

DeepLabv3+兼顾了高层语义特征提取与低层边界细节恢复，具有较好的分割性能和较强的多尺度建模能力。然而，在舌象分割任务中，原始 DeepLabv3+仍存在一定局限：其主干网络参数量相对较大，计算开销较高；同时，浅层细节特征利用仍不够充分，在舌体边缘、牙齿干扰区域以及复杂背景条件下仍可能出现轮廓不完整或局部误分现象。此外，原始 ASPP 对方向性上下文信息的建模能力仍有进一步提升空间。

3. LAE-DeepLabv3+舌像分割模型

针对 DeepLabv3+在舌象分割任务中存在的主干网络参数量较大、浅层细节特征利用不足以及 ASPP 模块对方向性上下文信息建模能力有限等问题，本文在 DeepLabv3+基础上提出了一种改进的舌象分割模型 LAE-DeepLabv3+，其网络结构如图 2 所示。该模型整体仍沿用 DeepLabv3+的编码器－解码器框架，在保持原有多尺度特征提取与解码融合优势的基础上，分别从主干网络、低层特征分支以及多尺度上下文建模模块三个方面进行改进，以提升模型对舌体区域的特征分割能力。

具体而言，本文首先采用轻量化的 EfficientViT 替换原始主干网络，以降低模型参数量和计算复杂度，同时提高特征提取效率；其次，在低层特征分支中引入 EMA 模块，对经1×1卷积降维后的浅层特征进行增强，以突出边界、纹理等关键信息，改善舌体边缘和局部区域的分割效果；最后，对原始 ASPP 结构进行增强，在保留多尺度空洞卷积分支的基础上增加水平与垂直条带分支，从而强化模型对方向性上下文信息和长程依赖关系的感知能力，进一步提升对舌体整体形状和复杂区域的表征效果。下面将分别对上述三个改进模块进行详细介绍。

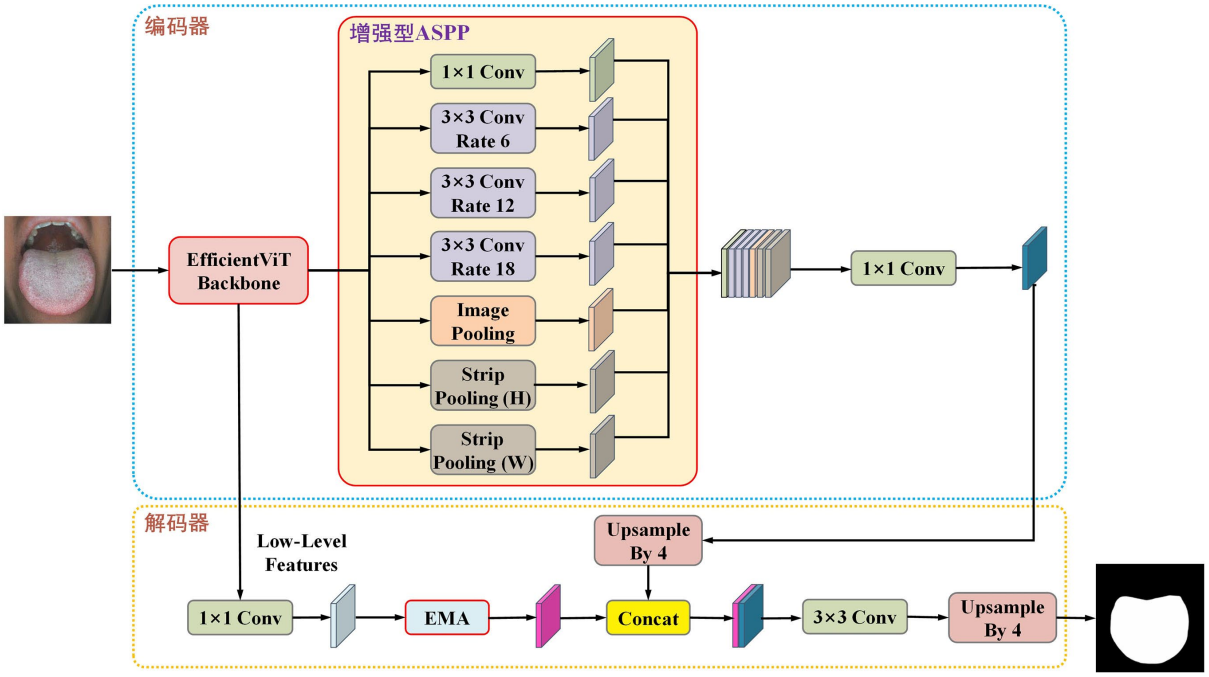

Figure 2. Network structure diagram of LAE-DeepLabv3+ model
图 2. LAE-DeepLabv3+模型的网络结构图

3.1. 基于 EfficientViT 的轻量化主干网络

在原始 DeepLabv3+中，主干网络通常承担输入图像特征提取的任务，其提取能力直接影响后续多尺度上下文建模与解码恢复的效果。然而，传统主干网络在参数规模和计算开销方面相对较大，在舌象分割任务中容易带来较高的计算负担，不利于模型轻量化设计与实际部署。针对这一问题，本文采用 EfficientViT 替换原始 DeepLabv3+的主干网络，以在保证特征表达能力的同时降低模型复杂度。

EfficientViT [18]是一种面向高分辨率密集预测任务设计的高效视觉骨干网络，其核心思想是在保持全局感受野和多尺度特征建模能力的基础上，尽可能减少计算冗余与硬件开销。与传统 Transformer 中计算复杂度较高的标准自注意力不同，EfficientViT 引入了多尺度线性注意力机制，从而以更高效的方式建模长距离依赖关系，并兼顾不同尺度特征的信息交互。该结构在高分辨率视觉任务中具有较好的效率优势，尤其适用于语义分割等密集预测场景。

从舌象分割任务特点来看，舌体区域通常具有较明确的整体结构，但同时也伴随着边缘过渡模糊、牙齿遮挡、背景干扰及局部纹理复杂等问题。因此，主干网络不仅需要提取高层语义信息，还应兼顾浅层空间细节表达。EfficientViT 采用分层特征提取方式，能够输出不同层级的特征表示，为后续解码器中的低层细节恢复和高层语义融合提供支持。相比于较为庞大的传统主干网络，EfficientViT 在减少参数量和运算量的同时，仍能保持较强的特征提取能力，这为后续模型整体轻量化提供了良好基础。

在本文模型中，输入舌象图像首先进入 EfficientViT 主干网络进行特征提取。浅层输出特征作为低层细节特征送入解码器分支，用于保留舌体边缘、轮廓和局部纹理信息；深层输出特征则送入增强型 ASPP 模块，用于进一步提取多尺度上下文语义信息。通过这种方式，EfficientViT 不仅承担了编码器主干的基础功能，还为后续低层特征增强和多尺度特征融合提供了更加高效的特征表示。

总体而言，将 EfficientViT 引入 DeepLabv3+主干网络，能够有效缓解原始模型参数量较大、计算开销较高的问题，同时保持较好的语义表达能力和多尺度特征提取能力。这一改进为本文模型在舌象分割任务中实现轻量化与高精度兼顾奠定了基础，也为后续引入 EMA 模块和增强型 ASPP 模块提供了更加高效的特征输入。

3.2. 融合 EMA 的低层特征增强模块

在 DeepLabv3+的解码阶段，低层特征 X_l 主要包含丰富的边缘轮廓、空间位置信息和局部纹理细节，对于舌体边界恢复具有重要作用。然而，原始 DeepLabv3+通常仅对低层特征进行 1×1 卷积降维后，便直接与上采样后的高层特征 X_h 进行拼接，未能进一步挖掘其中对分割任务更具判别性的细节信息。对于舌象分割任务而言，舌体与嘴唇、牙齿及口腔背景之间常存在边缘模糊、灰度过渡不明显和局部遮挡等问题，若低层特征利用不充分，容易造成边界缺失、局部误分以及轮廓不完整。为此，本文在低层特征分支中引入 EMA (Efficient Multi-scale Attention) [19]模块，其结构如图 3 所示，用于在特征融合前对 X_l 进行自适应增强。

设经过 1×1 卷积降维后的低层特征表示为 $X_l \in R^{C\times H\times W}$，其中，$C$、$H$、$W$ 分别表示特征图的通道数、高度和宽度。为降低计算复杂度并增强不同通道组之间的表征能力，EMA 首先将输入特征按通道划分为 G 组，即

$$X_l = Concat\left(X_l^{(1)}, X_l^{(2)}, \cdots, X_l^{(G)}\right), X_l^g \in R^{\frac{C}{G}\times H\times W} \tag{1}$$

其中，$g = 1,2,\cdots,G$，$Concat(\bullet)$ 表示沿通道维拼接操作。

对于每个分组特征 X_l^g，EMA 采用并行分支机制提取多尺度注意力信息。其中，两条 1×1 感知路径侧重于方向性空间依赖建模，一条 3×3 卷积分支用于补充局部邻域上下文信息。首先，在方向性建模分

支中，分别沿高度方向和宽度方向对特征进行全局平均池化，得到两个一维描述子：

$$h^{(g)} = GAP_H\left(X_l^{(g)}\right) \in R^{\frac{C}{G}\times 1\times W}, w^{(g)} = GAP_W\left(X_l^{(g)}\right) \in R^{\frac{C}{G}\times H\times 1} \tag{2}$$

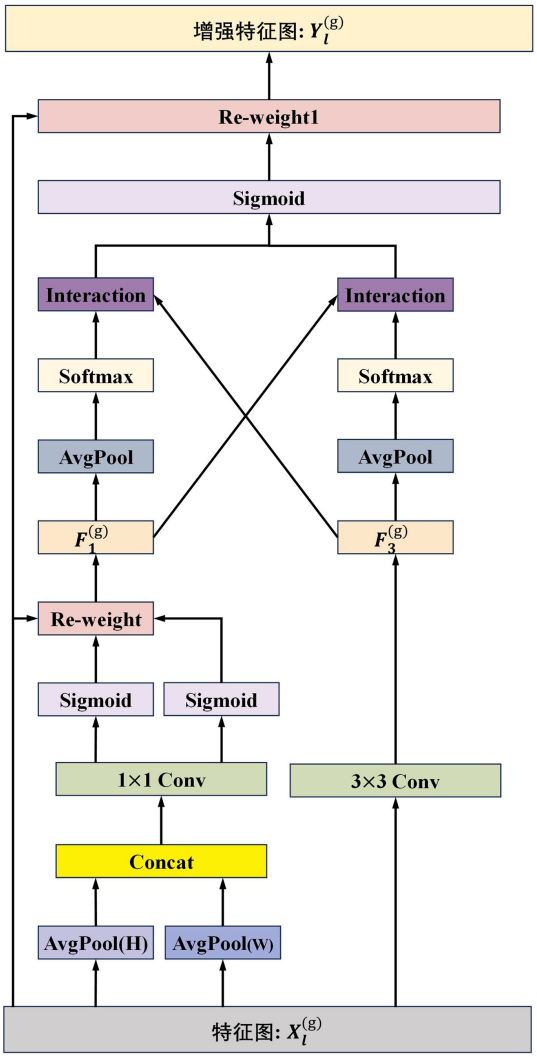

Figure 3. EMA module structure diagram integrated into low-level feature branches
图 3. 融合于低层特征分支的 EMA 模块结构图

其中，$GAP_H(\bullet)$ 和 $GAP_W(\bullet)$ 分别表示沿高度和宽度方向的全局平均池化操作。上述两个描述子分别编码了水平和垂直方向上的空间上下文信息。随后，将二者进行拼接，并通过1×1卷积和 Sigmoid 激活生成两个方向注意力权重：

$$\alpha_h^{(g)}, \alpha_w^{(g)} = \sigma\left(Split\left(Conv_{1\times 1}\left(Concat\left(h^g, w^g\right)\right)\right)\right) \tag{3}$$

其中，$\sigma(\bullet)$ 表示 Sigmoid 激活函数，$Split(\bullet)$ 表示将融合后的特征重新拆分为高度方向和宽度方向两组注意力图。利用方向注意力对原始分组特征进行重标定，可得

$$F_1^g = X_l^{(g)} \otimes \alpha_h^{(g)} \otimes \alpha_w^{(g)} \tag{4}$$

其中，⊗表示逐元素乘法。该过程能够在保留细粒度通道特征的同时，增强特征在水平和垂直两个方向上的位置敏感性，从而更好地突出舌体边缘和轮廓区域。

为了弥补1×1卷积分支感受野有限的问题，EMA同时设置并行的3×3卷积分支，对每个分组特征进行局部上下文建模：

$$F_3^g = Conv_{3\times3}\left(X_l^{(g)}\right) \tag{5}$$

该分支能够有效捕获邻域范围内的局部几何结构和纹理变化信息，对于舌尖、舌边缘以及受牙齿干扰区域的细节表达具有积极作用。

在获得方向增强特征 F_l^g 和局部上下文特征 F_3^g 后，EMA进一步通过跨分支交互机制实现多尺度信息融合。首先，对两条分支特征分别进行全局平均池化，并使用Softmax进行归一化，得到紧凑的全局响应描述子；随后通过交叉重加权方式实现不同尺度信息之间的互补增强，即

$$M^g = \mathrm{Softmax}\left(GAP\left(F_1^{(g)}\right)\right)\otimes F_3^{(g)} + \mathrm{Softmax}\left(GAP\left(F_3^{(g)}\right)\right)\otimes F_1^{(g)} \tag{6}$$

其中，第一项利用方向增强分支生成的全局权重引导局部上下文分支，第二项则利用局部上下文分支生成的全局权重反向增强方向分支。该双向交互机制使模型能够同时关注细粒度边界信息和更大范围的上下文语义信息，从而提高对复杂舌体区域的表征能力。

最后，将融合后的注意力图 M^g 经Sigmoid归一化后作用于原始分组特征，得到第 g 组增强特征：

$$Y_l^{(g)} = X_l^{(g)} \otimes \sigma\left(M^{(g)}\right) \tag{7}$$

将所有分组结果沿通道维拼接后，恢复为完整的增强低层特征表示：

$$Y_l = Concat\left(Y_l^{(1)}, Y_l^{(2)}, \cdots, Y_l^{(G)}\right) \in R^{C\times H\times W} \tag{8}$$

得到的 Y_l 将与上采样后的高层特征 X_h 进行拼接和融合，从而为后续解码器提供更加精细且判别性更强的特征输入。

EMA模块通过方向性注意力建模、局部上下文提取和跨分支交互增强的方式，对低层特征 X_l 进行自适应重标定，使网络在解码融合前能够更加突出与舌体区域相关的边界、轮廓和细节信息，并有效抑制背景噪声和无关响应。该模块弥补了原始DeepLabv3+在浅层特征利用方面的不足，为提高舌象分割结果的完整性、平滑性和边界准确性提供了有力支持。

3.3. 增强型ASPP多尺度上下文模块

为进一步提升网络对舌体区域的全局语义理解能力，本文在高层特征分支中对原始DeepLabv3+的ASPP模块进行改进，引入方向性感受野建模机制，构建增强型多尺度上下文模块。该模块在保留空洞卷积多尺度结构的基础上，融合Strip Pooling分支以建模长距离方向依赖，从而提升对细长结构及边界区域的刻画能力。

3.3.1. 原始ASPP结构

原始的ASPP模块如图2，设主干网络输出的高层语义特征为 $X_h \in R^{C\times H\times W}$，原始ASPP通过不同膨胀率的空洞卷积分支提取多尺度上下文特征，其形式可表示为：

$$F_i = Conv_{3\times3}^{r_i}\left(X_h\right), i = 1,2,3 \tag{9}$$

其中 r_i 表示不同的空洞率。与此同时，ASPP还包括一个1×1卷积分支：

$$F_0 = Conv_{1\times1}\left(X_h\right) \tag{10}$$

以及一个全局平均池化分支：

$$F_g = Up\left(Conv_{1\times1}\left(GAP\left(X_h\right)\right)\right) \tag{11}$$

其中，$Up(\bullet)$表示上采样操作。

最终多分支特征拼接后经融合卷积得到输出：

$$Y_h^{ASPP} = Conv_{1\times1}\left(Concat\left(F_0,F_1,F_2,F_3,F_g\right)\right) \tag{12}$$

然而，空洞卷积虽然能够扩大感受野，但其采样方式呈离散分布，对连续方向结构的建模能力有限；同时，全局池化分支缺乏方向选择性。在舌象图像中，舌体轮廓呈明显横向或纵向延展形态，仅依赖标准 ASPP 难以充分捕获这种方向依赖关系。

3.3.2. 增强型多尺度特征融合

为增强对方向性长程依赖的建模能力，本文在 ASPP 中引入 Strip Pooling 分支，分别沿高度和宽度方向进行条带式池化操作。

设输入特征为 X_h，则水平方向条带池化定义为：

$$S_H = Up\left(Conv_{1\times1}\left(GAP_H\left(X_h\right)\right)\right) \tag{13}$$

其中，$GAP_H(\bullet)$表示沿高度方向的全局平均池化，输出维度为$1\times W$，随后通过1×1卷积进行通道映射，并上采样恢复至$H\times W$。该分支能够在压缩高度维度后建模横向长距离依赖关系。

同理，垂直方向条带池化定义为：

$$S_W = Up\left(Conv_{1\times1}\left(GAP_W\left(X_h\right)\right)\right) \tag{14}$$

其中，$GAP_W(\bullet)$表示沿宽度方向的全局平均池化。该分支用于建模纵向长程依赖。通过条带池化，网络能够获得连续方向上的全局统计信息，相较于空洞卷积的离散采样机制，其上下文聚合更加平滑且具方向选择性。

在引入 Strip Pooling 分支后，增强型 ASPP 模块见图 3，包含以下特征集合：$\left\{F_0,F_1,F_2,F_3,F_g,F_H,F_W\right\}$。

将所有分支特征进行拼接并融合：

$$Y_h = Conv_{1\times1}\left(Concat\left(F_0,F_1,F_2,F_3,F_g,F_H,F_W\right)\right) \tag{15}$$

其中，$Y_h \in R^{C'\times H\times W}$为增强型 ASPP 输出特征。

该结构在原有多尺度空洞卷积基础上，引入方向性全局建模分支，从而形成更完整的空间上下文表达体系。

3.3.3. 对舌象分割任务的作用分析

舌象图像中，舌体轮廓呈弧形延展结构，舌边缘常沿水平方向具有连续边界，而舌尖及舌根区域则表现出明显的纵向结构特征。标准 ASPP 虽能扩大感受野，但难以针对特定方向建立连续响应。引入 Strip Pooling 后，S_H 强化横向边界连续性建模，S_W 强化纵向结构一致性建模，并且与空洞卷积互补，避免单一尺度依赖。因此，增强型 ASPP 模块能够有效提升舌体整体轮廓的完整性和边界光滑度，同时减少因局部遮挡或光照变化导致的误分割现象。

4. 实验结果与分析

本文实验所采用的舌象数据集为自行构建数据集。数据来源主要为《中医舌诊临床图解》[20]《临床

实用舌象图谱》[21]等多本中医舌诊相关专业书籍中的舌象照片。为保证数据质量，本文对采集到的图像进行了筛选与整理，剔除了模糊、遮挡严重以及舌体区域不完整的样本，最终得到 846 张舌象图像作为实验数据。本文利用专业图像标注软件 Labelme 对全部图像进行了像素级精细标注，生成高质量分割标签。实验数据集 8:2 的比例将数据集划分为训练集和测试集。图 4 展示了部分舌象图像及其对应标注样本，从中可以看出，不同样本在舌体颜色、形态及背景条件上均存在一定差异，这也在一定程度上增加了舌象分割任务的复杂性和挑战性。

Figure 4. Partial images of the tongue and their corresponding annotations
图 4. 部分舌象图像及其对应标注

4.1. 实验环境

本文实验基于 PyTorch 2.0.0 深度学习框架实现，开发环境为 Python 3.8，操作系统为 Ubuntu 20.04，CUDA 版本为 11.8。实验硬件平台配置为 1 张 32 GB 显存的 vGPU，CPU 为 24 vCPU AMD EPYC 7T83 64-Core Processor。在模型训练过程中，输入图像尺寸统一裁剪为 512 × 512，批大小(Batch Size)设置为 4，训练迭代次数为 40000 次。模型优化器选用 Adam，初始学习率设置为 0.0001，权重衰减系数为 0.001。此外，网络中 EMA 注意力机制的分组数 G 统一设定为 8。

4.2. 评价指标

为全面评价本文方法在舌象分割任务中的性能，本文选取了交并比(Intersection over Union, IoU)、Dice 系数(Dice coefficient, Dice)、准确率(Accuracy, Acc)、精确率(Precision, Pre)、敏感度(Sensitivity, Sen)、特异度(Specificity, Spe)以及模型参数量作为模型性能评价指标[10] [22] [23]。上述指标能够从区域重叠程度、整体分类准确性以及前景与背景识别能力等多个角度对分割结果进行综合评估。

$$\mathrm{IoU}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}\times 100\% \tag{16}$$

$$\mathrm{Dice}=\frac{2\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TP}+\mathrm{FN}}\times 100\% \tag{17}$$

$$\mathrm{Acc}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}\times 100\% \tag{18}$$

$$\mathrm{Pre}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\times 100\% \tag{19}$$

$$\mathrm{Sen}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\times 100\% \tag{20}$$

$$\mathrm{Spe}=\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}\times 100\% \tag{21}$$

其中，TP 表示模型将舌体区域正确预测为前景的像素数，TN 表示模型将背景区域正确预测为背景的像素数，FP 表示模型将背景误判为舌体区域的像素数，FN 表示模型将舌体区域误判为背景的像素数。

4.3. 与经典方法对比

为验证本文所提出 LAE-DeepLabv3+模型在舌象分割任务中的有效性，选取了多种经典语义分割模型进行对比实验，包括 UNet [10]、UNet++ [24]、SegNet [11]、BiSeNetV2 [25]、Attention U-Net [26]、TransUNet [27]以及 DeepLabv3+ [11]。所有模型均在相同数据集划分和实验环境下进行训练与测试，评价指标包括 IoU、Dice、Acc、Pre、Sen、Spe 以及模型参数量。具体实验结果如表 1 所示。

Table 1. Comparison of various metrics among different algorithms
表 1. 不同算法各指标对比

模型	IoU	Dice	Acc	Pre	Sen	Spe	模型参数量
UNet	95.97	97.43	98.14	97.80	98.09	98.17	7.85 M
UNet++	95.93	97.47	98.14	97.79	97.92	98.23	9.16 M
SegNet	95.72	97.35	98.04	97.72	97.91	98.11	29.44 M
BiSeNetV2	95.20	96.84	97.81	97.84	97.55	97.92	4.95 M
Attention U-Net	95.66	97.23	98.02	97.50	98.07	97.94	57.16 M
TransUNet	96.36	97.69	98.35	97.92	98.21	98.21	192.51 M
DeepLabv3+	96.14	98.03	98.23	97.97	98.09	98.20	43.65 M
LAE-DeepLabv3+	96.47	98.20	98.38	97.86	98.55	98.24	12.12 M

从表 1 可以看出，本文提出的 LAE-DeepLabv3+在多项指标上均取得了较优结果。其中，该模型的 IoU、Dice 和 Acc 分别达到 96.47%、98.20%和 98.38%，均为所有对比方法中的最优值，表明本文方法在舌体区域整体分割精度和预测一致性方面具有较强优势。与基线模型 DeepLabv3+相比，本文模型的 IoU 提高了 0.33 个百分点，Dice 提高了 0.17 个百分点，Acc 提高了 0.15 个百分点，说明通过引入轻量化主干网络、EMA 模块以及改进 ASPP 结构后，模型的特征提取与多尺度上下文建模能力得到了进一步增强，从而有效提升了分割性能。

为了进一步验证 IoU 提升的显著性，我们对 DeepLabv3+和 LAE-DeepLabv3+两种模型的 IoU 值进行了统计检验。具体而言，使用独立样本 t 检验对测试集的 169 张舌象图像的 IoU 差异进行显著性分析，目的是确认模型性能提升是否具备统计学意义。t 检验是一种常用于比较两个样本均值是否存在显著差异的统计方法，能够帮助我们客观判断提升结果是否仅仅是由于偶然因素引起的。在实验中，我们将 DeepLabv3+模型和 LAE-DeepLabv3+模型在相同数据集下的 IoU 均值进行对比，结果如表 2 所示。通过 t 检验结果，我们可以看到，LAE-DeepLabv3+的 IoU 显著高于 DeepLabv3+ (t 值 = 3.45，p 值 = 0.003)。p 值小于 0.05，表明模型性能提升是具有统计显著性的。

Table 2. Test set IoU comparison and t-test results
表 2. 测试集 IoU 比较及 t 检验结果

实验组	IoU 均值(DeepLabv3+)	IoU 均值(DeepLabv3+)	差值	t 值	p 值
测试集	96.14	96.47	0.33	3.45	0.003

在其他评价指标方面，LAE-DeepLabv3+的 Sensitivity 达到 98.55%，同样为所有模型中的最高值，说明该方法对舌体区域具有更强的检测能力，能够更完整地识别目标区域，减少漏分现象。其 Specificity 为 98.24%，也保持了较高水平，表明模型在背景区域判别上同样具有较好的鲁棒性。虽然在 Precision 指标上，本文模型为 97.86%，略低于部分对比方法，但整体差距较小，且结合更高的 Sen、IoU 和 Dice 指标可以说明，本文方法在保证较高预测准确性的同时，更注重舌体区域的完整分割，这对于医学图像分割任务具有更实际的应用价值。

LAE-DeepLabv3+的核心优势在于计算效率的大幅优化。相较于 DeepLabv3+ (参数量为 43.65 M)，本文模型的参数量仅为 12.12 M，减少了约 72.2%。这种显著的计算开销减少为实际部署和应用提供了极大的优势，尤其在资源受限的环境中，能够确保高精度的同时保持较低的计算负担。

4.4. 消融实验

为验证本文所提出各改进模块对模型性能提升的有效性，在基线模型 DeepLabv3+的基础上进行了消融实验。实验分别考察了轻量化主干网络 EfficientViT、低层特征分支中的 EMA 模块以及改进的 Enhanced ASPP 结构对舌象分割性能的影响。各组实验结果如表 3 所示。

由表 3 可以看出，基线模型 DeepLabv3+的 IoU、Dice 和 Acc 分别为 96.14%、98.03%和 98.23%，参数量为 43.65 M。当仅将主干网络替换为 EfficientViT 后，模型 1 的 IoU、Dice 和 Acc 分别提升至 96.28%、98.10%和 98.29%，同时参数量降至 10.81 M。相比基线模型，参数量大幅减少，说明 EfficientViT 在降低模型复杂度方面具有明显优势，同时仍能保持较高的特征提取能力，从而实现分割性能的小幅提升。这表明轻量化主干网络不仅减小了模型规模，也为后续模块的引入奠定了基础。

Table 3. Ablation test results
表 3. 消融实验结果

模型	EfficientViT	EMA	Enhanced ASPP	IoU	Dice	Acc	Pre	Sen	Spe	模型参数量
DeepLabv3+	-	-	-	96.14	98.03	98.23	97.97	98.09	98.20	43.65 M
Model1	√	-	-	96.28	98.10	98.29	97.95	98.26	98.19	10.81 M
Model2	√	√	-	96.38	98.16	98.34	98.08	98.23	98.23	10.81 M
Model3	√	-	√	96.33	98.13	98.32	97.94	98.33	98.22	12.11 M
LAE-DeepLabv3+	√	√	√	96.47	98.20	98.38	97.86	98.55	98.24	12.12 M

在模型 1 基础上进一步加入 EMA 模块后，模型 2 的 IoU、Dice 和 Acc 分别达到 96.38%、98.16%和 98.34%，较模型 1 继续提升。其中，Precision 提升至 98.08%，Specificity 提升至 98.23%。这说明 EMA 模块能够增强网络对低层细节信息的建模能力，使模型在特征融合前更好地关注舌体边缘、轮廓及局部区域，从而提高对前景与背景的区分能力。由于 EMA 模块被引入到 Low-level 分支中，因此其对浅层空间细节的强化作用较为明显。

当仅在 EfficientViT 主干基础上引入 Enhanced ASPP 时，模型 3 的 IoU、Dice 和 Acc 分别达到 96.33%、98.13%和 98.32%，较模型 1 同样有所提升；其中 Sensitivity 达到 98.33%，高于模型 1 和模型 2，表明改进后的 ASPP 结构在多尺度上下文信息建模方面发挥了积极作用。本文在原有 ASPP 基础上增加了两个条带分支，能够分别从水平方向和垂直方向捕获长程依赖信息，从而增强模型对舌体整体形状和区域分布的感知能力，有助于减少漏分现象，提高目标区域识别的完整性。

当三个改进模块同时引入后，最终模型 LAE-DeepLabv3+的各项指标达到最优，其中 IoU、Dice 和 Acc 分别为 96.47%、98.20%和 98.38%，Sensitivity 达到 98.55%，Specificity 达到 98.24%。虽然 Precision 略低于模型 2，但整体提升趋势更加明显，说明三个模块之间具有较好的协同作用。综合来看，EfficientViT 有效降低了参数量，EMA 强化了浅层细节特征表达，Enhanced ASPP 提升了多尺度上下文建模能力，三者结合后能够在保证模型轻量化的同时进一步提升舌象分割性能，验证了本文改进方法的合理性与有效性。

4.5. 可视化分析

为更加直观地验证本文所提 LAE-DeepLabv3+模型在舌象分割任务中的有效性，本文选取了 3 组具有代表性的测试样本进行可视化对比分析，分别展示了原始图像、真实标注以及不同分割模型的预测结果，如图 5 所示。通过对比可以看出，不同方法在舌体区域定位、边界保持以及复杂干扰抑制等方面存在一定差异，而本文方法在整体分割效果上表现出更好的稳定性与准确性。

从第一组样本可以看出，该图像受嘴唇、牙齿以及舌面纹理等因素影响较大，舌体上方区域存在较明显的干扰信息。部分对比方法在预测时受到口腔上部高亮区域的影响，出现了不同程度的误分现象，表现为舌体上边界附近存在额外响应，背景抑制不够充分。相比之下，本文方法能够较好地区分舌体区域与周围干扰区域，分割结果与真实标注更加接近，尤其在舌体上边缘和两侧边界处表现出更好的贴合性，说明该方法在复杂背景下具有较强的抗干扰能力。第二组样本中，舌体姿态存在一定偏斜，且舌面颜色分布不均匀，舌体顶部与口腔区域之间的灰度差异相对较小。这类样本对模型的边界判别能力提出了更高要求。从图中可以看出，一些对比模型在该样本上出现了边缘外扩或局部缺失的问题，说明其在细节刻画和上下文建模方面仍存在局限。本文方法所得分割结果在舌体整体形状保持方面更为自然，能够较准确地恢复舌体轮廓，减少边界模糊造成的误判现象，表明改进后的网络能够更有效地融合浅层细节信息与高层语义信息。第三组样本的成像条件相对较好，舌体区域与背景之间具有较明显区分，因此大多数模型都能够获得较为合理的分割结果。但进一步观察可以发现，不同方法在边缘平滑性和局部细节保留方面仍有差异。部分模型的预测轮廓存在轻微毛刺或边界偏移现象，而本文方法生成的分割区域更加平滑、完整，与真实标注在整体形状上更加一致。这说明本文模型不仅适用于复杂样本，在常规样本上同样具备较好的稳定性和泛化能力。

综合以上可视化结果可以看出，本文提出的 LAE-DeepLabv3+在舌象分割任务中能够更准确地定位舌体区域，并在边界细节恢复、背景干扰抑制以及舌体整体形状保持等方面表现出明显优势。

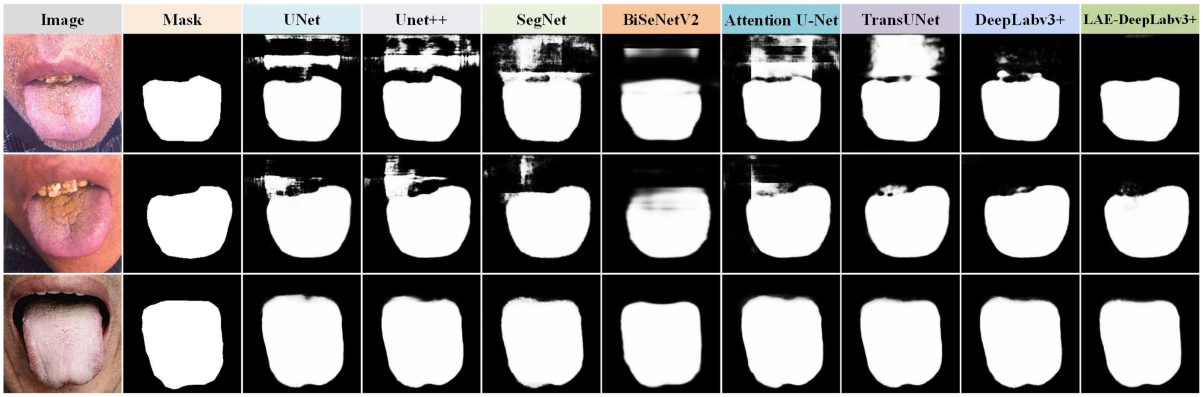

Figure 5. Visualization results
图 5. 可视化结果

5. 讨论

尽管本文提出的 LAE-DeepLabv3+模型在舌象图像分割任务中取得了较为优异的性能，但仍存在一些局限性，需要在未来的研究中进一步改进和优化。

本文使用的舌象数据集规模相对较小，只有 846 张图像，且主要来源于《中医舌诊临床图解》和《临床实用舌象图谱》等书籍中的舌象照片。这种数据集的来源较为单一，可能无法全面涵盖所有舌象类型和复杂的舌体变异，尤其是对于一些特殊病例，模型的表现可能不够理想。因此，未来的研究可以通过扩大数据集规模，引入更多样化的舌象图像，尤其是来自不同地区和不同临床背景的数据，以提升模型的泛化能力和适应性。

本研究提出的 EMA 模块和改进型 ASPP 模块在一定程度上增强了对舌体边界和细节的感知能力，但对于一些复杂舌象的处理能力仍然有限。例如，在舌体与牙齿接触区以及背景噪声较大的情况下，模型仍可能出现误分割现象。因此，未来的研究可以考虑引入更为复杂的上下文建模技术或多模态数据融合策略，以进一步提高在复杂舌象下的分割精度。

尽管本研究在舌象图像分割中取得了一定的进展，但舌象分析的客观化和标准化仍然面临挑战。未来的研究可以结合多种深度学习技术，如自监督学习、迁移学习等，进一步提升模型的泛化能力，并实现更加精准和自动化的舌象分析。此外，舌象与中医诊断的关联性尚待深入探讨，未来研究可以结合舌象分割模型与中医诊断知识库相结合，进一步推动智能中医诊断系统的发展。

6. 结束语

针对舌象图像分割任务中存在的舌体边界模糊、易受复杂背景干扰以及现有模型参数量较大等问题，本文提出了一种基于 LAE-DeepLabv3+的舌象图像分割方法。首先，模型引入 EfficientViT 主干网络，在保证特征提取能力的同时大幅降低了模型复杂度和计算开销；其次，在低层特征分支中引入 EMA 注意力模块，有效增强了模型对舌体边缘、局部纹理等浅层细节特征的表达能力，改善了局部误分现象；最后，设计了增强型 ASPP 模块，通过增加水平和垂直条带池化分支，强化了网络对方向性上下文信息和长程依赖的建模能力，显著提升了舌体整体形状的完整性。实验结果表明，在自建的舌象数据集上，LAE-DeepLabv3+的 IoU、Dice 和 Acc 分别达到了 96.47%、98.20%和 98.38%。与原始 DeepLabv3+及其他主流分割网络相比，本文模型不仅在多项评价指标上取得最优，而且参数量大幅缩减至 12.12 M，实现了分割精度与轻量化的良好平衡。本研究为中医舌诊的客观化与智能化提供了可靠的图像处理基础，具有较强的实际应用价值。

基金项目

2024 年天津商业大学大学生创新创业训练计划项目(202410069094)。

2025 年天津商业大学大学生创新创业训练计划项目(202510069011)。

参考文献

[1] Huang, C., Lin, H., Liao, W., Ceurvels, W. and Su, S. (2019) Diagnosis of Traditional Chinese Medicine Constitution by Integrating Indices of Tongue, Acoustic Sound, and Pulse. *European Journal of Integrative Medicine*, **27**, 114-120. https://doi.org/10.1016/j.eujim.2019.04.001

[2] Pang, B., Zhang, D., Li, N. and Wang, K. (2004) Computerized Tongue Diagnosis Based on Bayesian Networks. *IEEE Transactions on Biomedical Engineering*, **51**, 1803-1810. https://doi.org/10.1109/tbme.2004.831534

[3] 王旭阳, 刘世健. 基于多方向阈值的超分辨率图像噪声识别仿真[J]. 计算机仿真, 2021, 38(12): 132-135, 181.

[4] Wu, K. and Zhang, D. (2015) Robust Tongue Segmentation by Fusing Region-Based and Edge-Based Approaches.

Expert Systems with Applications, **42**, 8027-8038. https://doi.org/10.1016/j.eswa.2015.06.032

[5] Liu, W., Zhou, C., Li, Z. and Hu, Z. (2020) Patch-Driven Tongue Image Segmentation Using Sparse Representation. *IEEE Access*, **8**, 41372-41383. https://doi.org/10.1109/access.2020.2976826

[6] 刘冬梅, 常发亮. 结合 Retinex 校正和显著性的主动轮廓图像分割[J]. 光学精密工程, 2019, 27(7): 1593-1600.

[7] Zhao, Z.X., Wang, A.M. and Shen, L.S. (1999) The Color Tongue Image Segmentation Based on Mathematical Morphology and HIS Model. *Journal of Beijing Polytechnic University*, **25**, 67-71. (In Chinese)

[8] Zhai, X., Lu, H. and Zhang, L. (2009) Application of Image Segmentation Technique in Tongue Diagnosis. 2009 *International Forum on Information Technology and Applications*, Chengdu, 15-17 May 2009, 768-771. https://doi.org/10.1109/ifita.2009.130

[9] Long, J., Shelhamer, E. and Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 3431-3440. https://doi.org/10.1109/cvpr.2015.7298965

[10] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2015, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[11] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. https://doi.org/10.1109/tpami.2016.2644615

[12] Cai, Y., Wang, T., Liu, W. and Luo, Z. (2020) A Robust Interclass and Intraclass Loss Function for Deep Learning Based Tongue Segmentation. *Concurrency and Computation*: *Practice and Experience*, **32**, e5849. https://doi.org/10.1002/cpe.5849

[13] Huang, X., Zhang, H., Zhuo, L., Li, X. and Zhang, J. (2020) TisNet-Enhanced Fully Convolutional Network with Encoder-Decoder Structure for Tongue Image Segmentation in Traditional Chinese Medicine. *Computational and Mathematical Methods in Medicine*, **2020**, Article ID: 6029258. https://doi.org/10.1155/2020/6029258

[14] Chen, L.C., Papandreou, G., Kokkinos, I., *et al*. (2014) Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. arXiv: 1412.7062.

[15] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2018) DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. https://doi.org/10.1109/tpami.2017.2699184

[16] Chen, L.C., Papandreou, G., Schroff, F., *et al*. (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv: 1706.05587.

[17] Chen, L., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49

[18] Liu, X., Peng, H., Zheng, N., Yang, Y., Hu, H. and Yuan, Y. (2023) EfficientViT: Memory Efficient Vision Transformer with Cascaded Group Attention. 2023 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, 17-24 June 2023, 14420-14430. https://doi.org/10.1109/cvpr52729.2023.01386

[19] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., *et al*. (2023) Efficient Multi-Scale Attention Module with Cross-Spatial Learning. *ICASSP* 2023—2023 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*), Rhodes Island, 4-10 June 2023, 1-5. https://doi.org/10.1109/icassp49357.2023.10096516

[20] 许家佗. 中医舌诊临床图解[M]. 北京: 化学工业出版社, 2017.

[21] 王彦晖. 临床实用舌象图谱[M]. 北京: 化学工业出版社, 2012.

[22] Tang, C., Chen, H., Li, X., Li, J., Zhang, Z. and Hu, X. (2021) Look Closer to Segment Better: Boundary Patch Refinement for Instance Segmentation. 2021 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 13921-13930. https://doi.org/10.1109/cvpr46437.2021.01371

[23] Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R. and Jagersand, M. (2020) U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection. *Pattern Recognition*, **106**, Article ID: 107404. https://doi.org/10.1016/j.patcog.2020.107404

[24] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J. (2018) UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., *et al*., Eds., *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1

[25] Yu, C., Gao, C., Wang, J., Yu, G., Shen, C. and Sang, N. (2021) BiSeNet V2: Bilateral Network with Guided Aggregation

for Real-Time Semantic Segmentation. *International Journal of Computer Vision*, **129**, 3051-3068. https://doi.org/10.1007/s11263-021-01515-2

[26] Oktay, O., Schlemper, J., Folgoc, L.L., *et al*. (2018) Attention U-Net: Learning Where to Look for the Pancreas. arXiv: 1804.03999.

[27] Chen, J., Lu, Y., Yu, Q., *et al*. (2021) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv: 2102.04306.