

基于改进U-Net模型的咽拭子图像分割研究

贾 劼

复旦大学工程与应用技术研究院, 上海

收稿日期: 2025年3月12日; 录用日期: 2025年5月15日; 发布日期: 2025年5月26日

摘 要

为提高咽拭子图像中口腔M区域的分割效果, 本文提出一种改进U-Net模型。针对传统U-Net的浅层特征冗余与边缘敏感性不足的问题, 引入循环残差模块增强边缘特征传递能力, 并结合注意力机制抑制背景噪声, 优化多尺度特征融合策略。实验表明, 相比先前研究所采用的U-Net方法, 改进模型的交并比(IoU)、Dice相似系数(DSC)和像素精确率(PA)分别提高3.8%、2.1%和3.5%, 可精准分割复杂背景下低对比度的口腔M目标区域, 为咽拭子采样机器人提供高效的视觉支持, 降低人工采样风险。

关键词

图像处理, 咽拭子图像分割, 循环残差机制, 注意力机制

Research on Pharyngeal Swab Image Segmentation Based on an Improved U-Net Model

Jie Jia

Academy for Engineering and Technology, Fudan University, Shanghai

Received: Mar. 12th, 2025; accepted: May 15th, 2025; published: May 26th, 2025

Abstract

This paper proposes an improved U-Net model to enhance the segmentation performance of the oral M region in pharyngeal swab images. To address the issues of redundant shallow features and insufficient edge sensitivity in traditional U-Net models, a recurrent residual module is introduced to enhance edge feature propagation. Additionally, an attention mechanism is incorporated to suppress background noise and optimize the multi-scale feature fusion strategy. Experimental results demonstrate that compared to the U-Net method used in previous studies, the improved model achieves

increases of 3.8%, 2.1%, and 3.5% in Intersection over Union (IoU), Dice Similarity Coefficient (DSC), and Pixel Accuracy (PA), respectively, which can enable precise segmentation of the low-contrast oral M region in complex backgrounds, providing efficient visual support for pharyngeal swab sampling robots and reducing the risks associated with manual sampling.

Keywords

Image Processing, Pharyngeal Swab Image Segmentation, Recurrent Residual Mechanism, Attention Mechanism

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

咽拭子核酸检测作为新型冠状病毒感染的主要检测手段，其采样过程存在显著的临床挑战：医护人员需与患者密切接触，易暴露于飞沫或气溶胶感染风险；同时，人工采样的操作差异易导致样本质量波动，进而引发假阴性结果，影响诊断准确性。为应对上述问题，咽拭子采样机器人作为一种替代人工采样的有效手段[1]，能够通过自动化操作降低医护人员的感染风险，同时提升采样的标准化水平。在此过程中，口腔内 M 区域的精准分割成为机器人导航与采样的核心前提，其边缘模糊、弱对比度的特性对传统分割方法提出了严峻挑战。

现有研究面向医学图像分割已经取得一定进展，但仍存在局限性。文献[2]提出的全卷积网络(Fully Convolutional Networks, FCN)虽能实现像素级分类，但因池化操作导致细粒度边缘信息丢失，难以满足咽拭子图像中边缘细节的分割需求。基于 U 型架构的模型利用编码器-解码器结构及跳跃连接[3]，实现多尺度特征融合，在 CT/MRI 等医学图像分割任务中表现优异[4]，但对弱边缘区域的敏感性不足。研究[5]通过引入压缩激励(SE)模块优化特征权重分配，虽提升了分割精度，但未针对浅层特征中的背景噪声进行有效抑制。此外，新冠肺炎相关医学图像分割研究多聚焦于 CT 影像的感染区域检测，如文献[6]采用改进的 CENet 模型用于新冠肺炎 CT 图像感染区域分割。但面向咽拭子图像分割的研究较少，近期文献[7]将 U-Net 模型用于咽拭子图像分割。由于咽拭子图像背景复杂、M 目标区域形态多变且对比度低，这种基于传统 U-Net 模型的方法对边界的分割效果仍有待提高。

针对上述挑战，本文提出一种融合循环残差机制与注意力机制的改进 U-Net 模型。首先，通过参数共享的循环卷积单元强化边缘特征的传递能力，结合残差连接缓解梯度消失问题；其次，在跳跃连接中嵌入注意力机制，能够自适应优化特征融合过程，通过动态抑制编码器浅层特征中的冗余背景信息，进一步增强目标区域的特征表示。实验表明，相比先前研究采用的 U-Net 模型方法，该改进方法在保持轻量化特性的同时，显著提升了口腔 M 区域的分割精度与鲁棒性，为咽拭子采样机器人的实时导航与精准操作提供了可靠的技术支撑。

2. 方法

如图 1 所示，基于循环残差机制和注意力机制改进的 U-Net 模型，旨在提升咽拭子 M 区域图像分割的精度和鲁棒性。该模型在先前 U-Net 架构[3]的基础上，通过引入循环残差卷积模块和注意力机制，增强了对复杂背景下的边缘细节捕捉能力，并减少了特征融合过程中的冗余信息。

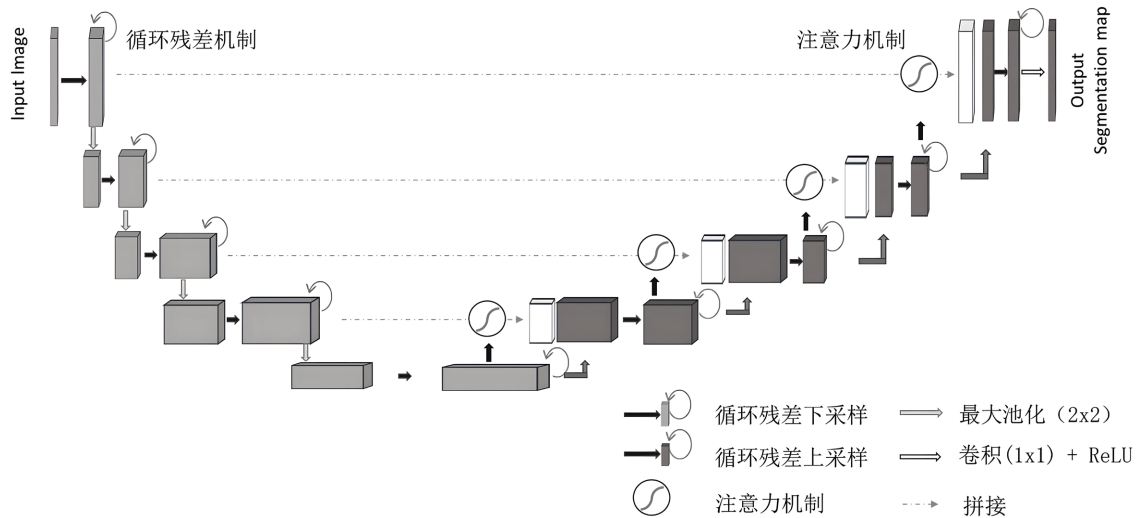


Figure 1. Schematic diagram of improved U-Net model structure

图 1. 改进 U-Net 模型结构示意图

2.1. 基础模型结构

如图 1 所示, 本文采用类似 U-Net 的 U 型架构作为基础模型。该架构的核心包括编码器(下采样路径)、瓶颈层(Bottleneck)和解码器(上采样路径)三大部分, 并通过跳跃连接(Skip Connections)实现多尺度特征融合, 从而显著提升分割性能。

编码器由 4 个编码层构成, 每个编码层依次对输入特征图执行两次 3×3 卷积操作, 卷积过程中均采用 ReLU (Rectified Linear Unit, 修正线性单元)激活函数, 以提取局部纹理和形态特征。随后, 特征图通过 2×2 最大池化(MaxPool2d)操作进行下采样, 将分辨率减半, 从而逐步扩大感受野并捕获更高级的语义信息。降维后的特征图被传递至下一编码层, 最终在编码器末端的瓶颈层中, 特征图的通道数被扩展至 512, 分辨率降至原始输入的 $1/16$ 。瓶颈层通过深度卷积进一步提取全局上下文信息, 为解码器提供丰富的高级语义特征。

解码器由 4 个解码层组成, 每个解码层首先通过上采样操作逐步恢复特征图的分辨率, 随后执行两次 3×3 卷积操作并结合 ReLU 激活函数, 以进一步优化特征表示。模型通过跳跃连接将编码器各层输出的浅层低级语义特征(如边缘、纹理等)与解码器对应层的高级语义特征在通道维度上拼接, 实现多尺度特征融合。这种设计不仅有效缓解了下采样过程中可能引发的信息丢失问题, 还增强了模型对微小解剖结构(如咽拭子区域边缘)等细节特征的捕获能力, 从而在复杂背景下的图像分割任务中表现出色。

2.2. 循环残差卷积模块

在基础模型的编码-解码结构的基础上, 本文进一步引入循环残差机制, 以增强对口腔 M 区域边缘的分割能力, 提升模型在复杂背景下的鲁棒性。该机制主要由两个核心部分构成: 1) 参数共享的循环卷积单元, 2) 残差连接结构, 两者协同作用, 使模型能够在特征提取过程中有效增强边缘信息的捕捉能力。

1) 参数共享的循环卷积单元

循环卷积单元采用参数复用, 通过多次迭代计算, 在不额外增加模型参数量的前提下, 实现短期记忆建模。这一设计能够使网络在特征传递过程中不断强化口腔 M 区域的细节信息, 从而提高分割精度。具体而言, 该模块的每次迭代均使用相同的卷积核权重, 使得输入特征在多个时间步上进行逐步更新,

进而提升模型对局部边缘细节的感知能力。

2) 残差连接结构

为了进一步优化深层网络的训练效果,本文在循环卷积单元基础上引入残差连接结构。该结构在每次循环迭代后,将当前输出特征与原始输入特征进行逐元素相加,从而在深度特征学习过程中有效缓解梯度消失问题。此外,残差连接还能够提供更稳定的梯度流,使网络更易于优化,同时提高模型的收敛速度。

对于输入特征 X , 循环计算的第 t 次迭代 H_t 输出为:

$$H_t = f(H_{t-1}; W) = \text{ReLU}(\text{BN}(\text{Conv}(H_{t-1}; W))) \quad (1)$$

其中, Conv 表示卷积操作, BN 为批归一化(Batch Normalization), ReLU 代表修正线性单元(Rectified Linear Unit)激活函数, W 为共享的卷积核权重参数。最终,残差连接的计算如下:

$$Y = H_T + X \quad (2)$$

其中, T 代表循环次数, Y 为最终的输出特征图。通过这种方式,循环残差机制能够在保持网络轻量化的同时,提高分割精度,增强模型对 M 区域边界的捕捉能力,从而实现更优的医学图像分割效果。

2.3. 注意力机制

在传统 U-Net 模型中,跳跃连接通常将编码器各层提取的特征图直接与解码器对应层的特征图拼接。然而,编码器提取的高分辨率浅层特征中往往包含大量冗余信息。为了减少浅层特征与深层特征融合时不相关信息和噪声的干扰,本文在跳跃连接中引入了注意力机制(如图 1 所示)。该机制通过计算注意力权重,自适应地优化特征融合过程,抑制无关区域的特征传递,从而增强目标区域的特征表示。

具体而言,在编码器输出的高分辨率浅层特征 X_{enc} (包含丰富空间细节但混杂背景噪声)与解码器传递的深层语义特征 X_{dec} (具备高级语义但空间信息粗糙)交互时,首先对 X_{dec} 进行 1×1 卷积实现通道维度对齐,使其与 X_{enc} 保持相同通道数。随后将二者沿通道维度拼接,通过 $1 \times 11 \times 1$ 卷积结合 ReLU 激活生成中间门控信号,再经 1×1 卷积与 Sigmoid 函数处理,最终输出空间注意力权重图 $\alpha \in [0,1]^{H \times W}$,其数学表达为:

$$\alpha = \sigma(W_2 \cdot \delta(W_1 \cdot [X_{enc}, X_{dec}])) \quad (3)$$

其中, W_1 、 W_2 为可学习的 1×1 卷积核参数, δ 与 σ 分别代表 ReLU 和 Sigmoid 激活函数。该权重图通过逐元素相乘作用于原始编码器特征 X_{enc} , 实现 $X_{att} = \alpha \odot X_{enc}$ 的加权增强,在保留目标区域(如口腔 M 区)边缘细节的同时,显著抑制背景噪声的干扰。

2.4. 损失函数

在训练神经网络的过程中,需要在每次前向推理后使用损失函数计算预测值与真实值的损失(loss),从而利用梯度下降法(Gradient Descent)对网络参数进行迭代优化。基于不同的任务,选用适合的损失有利于模型的训练记忆最终表现。

在语义分割任务中,本质上是对输入图像中的每一个像素点进行类别分类,因此损失函数的选择需要适应像素级分类的特性。交叉熵损失函数(Cross-Entropy Loss)是分类任务中常用的损失函数,在语义分割模型训练中具有广泛应用。

其核心思想在于评估模型预测的类别概率分布与真实标签的差异,从而引导模型参数的优化调整。交叉熵损失的数学表达式如下:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(p_{i,c}) \quad (4)$$

其中, N 表示图像像素总数, C 为类别数量, $y_{i,c} \in \{0,1\}$ 代表第 i 个像素的真实类别标签(One-hot 编码), $p_{i,c}$ 为模型预测的该像素属于类别 c 的概率值。交叉熵通过惩罚预测概率与真实分布的偏离, 迫使网络学习更具判别性的特征表示。

3. 实验验证

3.1. 实验设置

数据集准备: 本文数据集采用飞桨口腔图像识别比赛公开数据集, 数据集及标注的 M 区域如图 2 所示。此数据集总共包含 835 张图像, 训练集和验证集按照 4:1 比例划分。

数据集预处理: 在模型训练前, 对所有图像数据进行预处理, 并将其统一调整至 512×512 的标准尺寸, 以保证输入数据的规范性和一致性。

模型训练: 模型训练使用交叉熵损失函数进行优化。实验基于 Python 实现 Pytorch 框架提供 GPU 加速的张量计算功能; 使用 Adam 优化器进行训练。

对比方法: 本文对所构建的改进 U-Net 模型与基本 U-Net 模型[3]在验证集进行预测效果对比, 以验证改进方法的有效性。

3.2. 评价指标

为了对分割模型的实验结果进行定量分析和对比, 本文在测试集上分别采用交并比系数(Intersection over Union, IoU)、Dice 相似系数(Dice Similarity Coefficient, DSC)以及精确率(Precision Accuracy, PA)三种医学图像分割常用评价指标, 以衡量图像分割的性能。

交并比系数(IoU): IoU 是评估分割模型性能的常用指标, 用于衡量模型预测结果与真实标注之间的重叠程度。其计算公式如下:

$$\text{IoU} = \frac{|\mathcal{P} \cap \mathcal{G}|}{|\mathcal{P} \cup \mathcal{G}|} = \frac{TP}{TP + FP + FN} \quad (5)$$

其中, $\mathcal{P}$ 为预测区域, $\mathcal{G}$ 为真实标注, TP 为正确分割的 M 区域像素, FP 为背景被误判为 M 区域的像素, FN 为漏分割的 M 区域像素。

Dice 相似系数(Dice Similarity Coefficient, DSC): DSC 衡量预测结果与真实标注的像素级相似性。其计算公式如下:

$$\text{DSC} = \frac{2|\mathcal{P} \cap \mathcal{G}|}{|\mathcal{P}| + |\mathcal{G}|} = \frac{2TP}{2TP + FP + FN} \quad (6)$$

像素精确率(Pixel Accuracy, PA): PA 计算正确分类像素占总像素的比例。其计算公式如下:

$$\text{PA} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

其中, TN 为正确识别的背景像素。

3.3. 结果与分析

图 2 展示了不同模型在 M 区域分割任务中的效果对比。各列依次为: 原始输入(包含 M 区域的咽拭子图像)、真实标注(由专家标记的 M 区域轮廓)、改进 U-Net 模型的分割结果以及传统 U-Net 模型的分割

结果。由于咽拭子图像中 M 区域与背景的对比较低，对 M 区域的边界分割存在边界不明确的挑战。对比实验结果可以看出，在 M 区域的边缘部分，改进模型的分割效果更好，能够更精确地捕捉 M 区域的轮廓信息。

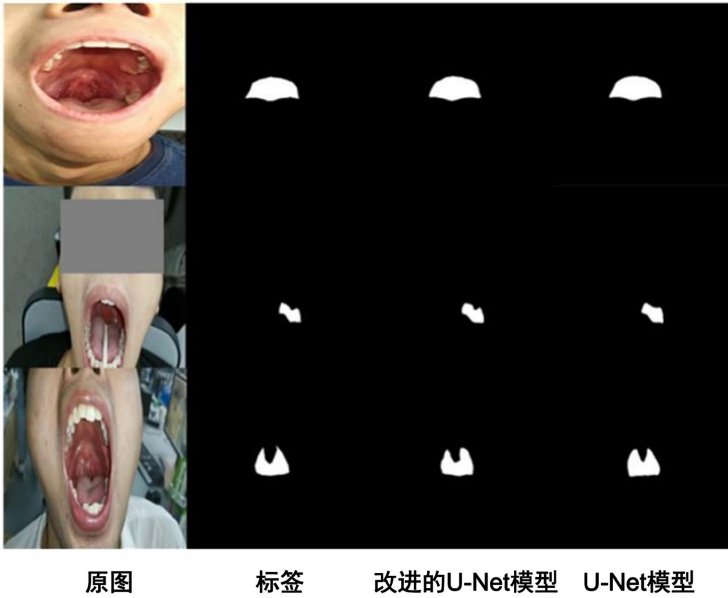


Figure 2. Visualization of segmentation results
图 2. 分割结果可视化

Table 1. Comparison of experimental results
表 1. 对比实验结果

方法	IoU	DSC	PA
U-Net	0.812	0.896	0.923
Proposed	0.843	0.915	0.955

定量结果：如表 1 实验结果所示，在 IoU、DSC 和 PA 三个评价指标中，所改进 U-Net 模型相比传统 U-Net 模型在评估指标上表现更优，与真实标注的吻合度相比传统 U-Net 模型方法更好。具体而言，改进模型的平均交并比(MIOU)提升 3.82%，Dice 相似系数(MDSC)提升 2.12%，像素准确率(PA)提升 3.47%。这一结果表明，通过引入的改进机制，模型在复杂区域(如口腔图像 M 区域)的细节捕捉能力和分割精度得到有效增强。同时，改进后的 U-Net 模型仍保持了原有模型轻量化的特性，能够实现对口腔图像 M 区域的实时精确分割，完全满足实际核酸采样机器人对图像分割精度的实时需求，为实际应用提供了可靠的技术保障。

4. 结论

本文针对咽拭子图像中口腔 M 区域弱边缘分割的挑战，提出一种改进 U-Net 模型，通过循环残差机制与注意力机制的协同优化，显著提升复杂背景下的分割精度与鲁棒性。实验验证表明，改进模型在 IoU、DSC 等指标上均优于先前 U-Net 方法，该方法满足咽拭子采样机器人系统对口腔目标 M 区域的分割需求，为咽拭子采样机器人提供高效的视觉支持，降低人工操作的风险代价。未来工作将探索将本方法与咽拭子采样机器人路径规划模块结合，进一步验证机器人系统在实际环境中的鲁棒性。

参考文献

- [1] Li, S., Guo, W., Liu, H., Wang, T., Zhou, Y., Yu, T., *et al.* (2020) Clinical Application of an Intelligent Oropharyngeal Swab Robot: Implication for the COVID-19 Pandemic. *European Respiratory Journal*, **56**, Article ID: 2001912. <https://doi.org/10.1183/13993003.01912-2020>
- [2] 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>
- [3] 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., *Lecture Notes in Computer Science*, Springer International Publishing, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [4] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., *et al.* (2017) A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, **42**, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- [5] Roy, A.G., Navab, N. and Wachinger, C. (2018) Concurrent Spatial and Channel ‘Squeeze & Excitation’ in Fully Convolutional Networks. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C. and Fichtinger, G., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 421-429. https://doi.org/10.1007/978-3-030-00928-1_48
- [6] 邱纯乾, 陈建森, 郑茜颖. 基于改进 CENet 的新冠肺炎 CT 图像感染区域分割[J]. 传感器与微系统, 2023, 42(11): 139-142+146.
- [7] 刘柏男. 视觉引导咽拭子机器人智能采集关键技术研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2021.