

DOI:10.19651/j.cnki.emt.2518539

基于双编码器的鼻中隔医学图像分割模型^{*}

周宝康¹ 曹爽¹ 高洪涌¹ 宋维波¹ 崔树林²

(1. 大连海洋大学信息工程学院 大连 116023; 2. 大连理工大学附属中心医院 大连 116033)

摘要: 鼻中隔解剖结构的精准分割对病症评估及手术规划具有重要临床价值,然而现有基于卷积神经网络(CNN)的方法在全局特征表征方面存在局限性。为此,本研究创新性地构建 CTA-Net 模型,通过双分支编码架构实现局部—全局特征协同学习;CNN 分支捕捉解剖结构细节特征,Transformer 分支建模长距离空间依赖关系,并设计特征融合模块实现特征信息交互。特别在瓶颈层引入多尺度特征注意力机制,通过自适应感受野调整增强模型对复杂解剖结构的表征能力。实验验证采用自主标注的鼻中隔临床数据集及 ISIC 2018、Kvasir 共 3 个医学数据集进行对比实验,结果显示本模型在鼻中隔分割任务中 IoU 和 Dice 分别达到 90.38% 和 94.94%。在跨数据集对比实验中,胃肠镜图像分割任务上的 IoU 精度达 76.17%,显著优于其他现有模型,证实了模型在特征学习能力和泛化适应性方面的优势。本研究为医学图像分析提供了一种融合局部感知与全局建模的创新性解决方案,在耳鼻喉科智能诊疗领域具有重要应用前景。

关键词: CNN; 鼻中隔; 多尺度特征; 医学图像分割; 深度学习

中图分类号: TP391; TN911 文献标识码: A 国家标准学科分类代码: 520.6040

Dual encoder-based nasal septum medical image segmentation model

Zhou Baokang¹ Cao Shuang¹ Gao Hongyong¹ Song Weibo¹ Cui Shulin²

(1. College of Information Engineering, Dalian Ocean University, Dalian 116023, China;

2. Dalian Municipal Central Hospital, Dalian 116033, China)

Abstract: Accurate segmentation of the nasal septum anatomical structure holds significant clinical value for disease assessment and surgical planning. However, existing methods based on Convolutional Neural Network (CNN) exhibit limitations in global feature representation. To address this issue, this study innovatively constructs the CTA-Net model, which achieves local-global feature collaborative learning through a dual-branch encoding architecture: the CNN branch captures fine anatomical details, while the Transformer branch models long-distance spatial dependencies, and a feature fusion module is designed to enable effective information exchange. Particularly, a multi-scale feature attention mechanism is introduced in the bottleneck layer to enhance the model's capability to represent complex anatomical structures through different receptive fields. Experiments were conducted on three medical datasets—a self-annotated clinical dataset of the nasal septum, ISIC 2018, and Kvasir. The results demonstrate that, in the nasal septum segmentation task, the model achieves IoU and Dice coefficients of 90.38% and 94.94%, respectively. In cross-dataset experiments, the IoU accuracy for gastrointestinal endoscopic image segmentation reached 76.17%, significantly outperforming other existing models, thereby confirming the model's advantages in feature learning and generalization. This study provides an innovative solution for medical image analysis by integrating local perception with global modeling, and it holds significant promise for intelligent diagnosis and treatment in the otolaryngology field.

Keywords: CNN; nasal septum; multi-scale feature; medical image segmentation; deep learning

0 引言

维持鼻腔结构稳定、分隔左右鼻腔及引导气流的重要功能,同时还参与调节局部湿度和温度等生理过程。鼻中隔偏曲^[1-3]是指鼻中隔偏离正常中线位置的病理状态,常因遗

鼻中隔位于鼻腔中央,由骨与软组织构成,不仅承担着

传、外伤或发育异常等因素引起,导致一侧鼻腔气道狭窄,并可能引发呼吸障碍、慢性鼻炎及其他并发症。近年来,随着临床对精细解剖结构评估要求的不断提高,如何在影像数据中准确定位并定量分析鼻中隔偏曲程度,已成为改善诊疗策略、降低手术风险的重要技术难题。

传统的诊断方法主要依赖鼻内镜检查和 CT 影像评估,但在实际应用中往往面临多重挑战。首先,CT 影像中鼻中隔与周围软组织之间的灰度值差异较小,加之部分病例存在图像噪声、低对比度及边缘模糊问题,使得传统分割方法(基于阈值分割^[4]及边缘检测^[5]方法)难以准确提取鼻中隔的完整结构。其次,由于鼻中隔的解剖结构存在细微且复杂的曲率变化,单一的传统方法往往难以同时兼顾全局形态与局部细节的提取,进而影响临床对偏曲程度的定量评估及手术规划的精准性。

近年来,深度学习技术^[6-8]在医学图像分割领域取得了显著进展。基于 CNN 的分割模型如 UNet^[9] 和 UNet 十十^[10] 在局部特征提取方面表现优异,UNet 的多种变种^[11-16] 模型在改进特征融合与增强感兴趣区域聚焦方面取得了进一步提升。尽管这些方法在捕捉局部细节上表现出色,但其固有的卷积操作限制了全局上下文和长距离依赖关系的建模能力。难以捕捉鼻中隔整体结构中的长距离依赖关系。

为了解决 CNN 在全局特征语义提取上的不足,Transformer^[17] 架构逐步被引入医学图像分割领域。得益于自注意力机制^[18] 在捕捉全局依赖性方面的优势,ViT^[19] 和 Swin Transformer^[20] 通过将图像划分为多个图块并引入分层设计,展现了处理复杂场景的潜力。然而,单一 Transformer 结构在局部细节捕捉上存在不足,且通常依赖大规模数据进行训练,这在医学影像数据有限的情况下成为另一种问题。

为了充分利用 CNN 在局部特征提取上的优势和 Transformer 在全局信息建模上的能力,各类混合模型相继涌现。例如,TransUNet^[21] 首次将 Transformer 嵌入 UNet 框架中,通过跳跃连接实现了两种机制的有效融合;随后,TransClaw U-Net^[22-23] 和 DS-TransUNet^[24] 在多尺度特征融合和细粒度特征提取上进行了进一步优化;而 HiFormer^[25] 与 MAXFormer^[26] 则在平衡全局语义与局部细节方面提供了新的思路。近期,Petit 等^[27] 提出的双编码器结构通过交叉注意力机制与增强特征融合进一步提升了模型性能。尽管这些工作展示了混合架构在医学图像分割中的巨大潜力,但在应对解剖结构复杂、形态多变的任务(如鼻中隔分割)时,仍存在局部信息提取不足、特征交互不充分以及多尺度特征整合效率不高等问题。

针对上述问题,本文提出了一种新型双编码器分割模型 CTA-Net。该模型采用双分支编码器架构,一方面通过 CNN 分支有效捕捉鼻中隔的局部纹理与细节信息,另一方面利用 Transformer 分支构建长距离空间依赖关系,实现

全局语义的精准建模。特别是在网络瓶颈层中,本文引入了多尺度特征注意力提取模块,通过并行空洞卷积结合空间注意力策略实现不同尺度上下文信息的自适应捕捉,从而提升模型在处理复杂解剖结构时的鲁棒性。此外,设计的特征融合模块实现了两分支间的深度信息交互,进一步优化了分割效果。实验部分在自主标注的鼻中隔临床数据集以及 ISIC 2018、Kvasir 公开数据集上进行验证,结果显示 CTA-Net 在 IoU 和 Dice 系数等关键指标上均显著优于现有主流模型,为鼻中隔的精细分割提供了有力技术支持。本文的工作内容如下:

- 1) 提出了一种双分支编码模型 CTA-Net,结合了 CNN 的局部特征提取能力和 Transformer 的全局特征提取能力。实验结果表明,双编码器有效的结合了 CNN 和 Transformer 的优势,充分提高特征提取能力,提高分割性能。
- 2) 在 CNN 编码器的特征提取阶段使用改进的残差卷积块,以缓解梯度消失问题,并促进更深层特征的有效学习。
- 3) 为了增强模型提取多尺度特征的能力,在网络的瓶颈层引入了 MSFA(多尺度特征注意力),通过并行的空洞卷积和空间注意力捕捉不同尺度的上下文信息并关注目标区域,适合处理鼻中隔结构存在尺度变化的问题。
- 4) 标注了一个鼻中隔数据集,填补了该领域高质量标注数据的空白。

1 本文模型

1.1 整体概述

本节首先对模型整体架构进行介绍,如图 1 所示,本模型整体由双编码器,特征融合模块,MSFA 以及解码器构成,并在编码器与解码器之间设置多层跳跃连接,实现高层语义信息与低层细节信息的有效融合。CNN 分支通过一系列残差卷积编码块提取局部纹理、边缘及细节特征。在每个编码单元中,使用改进的残差卷积以缓解梯度消失问题,确保深层网络训练的稳定性。Transformer 分支利用自注意力机制对输入特征进行全局建模,捕捉长距离依赖关系,从而弥补传统 CNN 在全局信息提取上的不足。在编码过程中,Transformer 分支提取的全局特征表示与 CNN 分支的局部特征经过设计的特征融合模块进行特征融合,从而获得具有更全面表征能力的综合特征表示。融合模块有效实现了双分支特征信息的互补。网络瓶颈层引入了 MSFA,通过设置不同空洞率的卷积核和空间注意力并行提取多尺度上下文信息,从而丰富了特征表示。这一设计不仅提高了模型对结构尺度变化的适应能力,也为后续解码阶段提供了更加丰富的特征基础。在解码阶段,模型采用转置卷积操作逐步对特征图进行上采样,并通过与编码阶段相应层的跳跃连接进行特征融合,从而有效恢复分辨率,并保证细节信息得以充分保留。

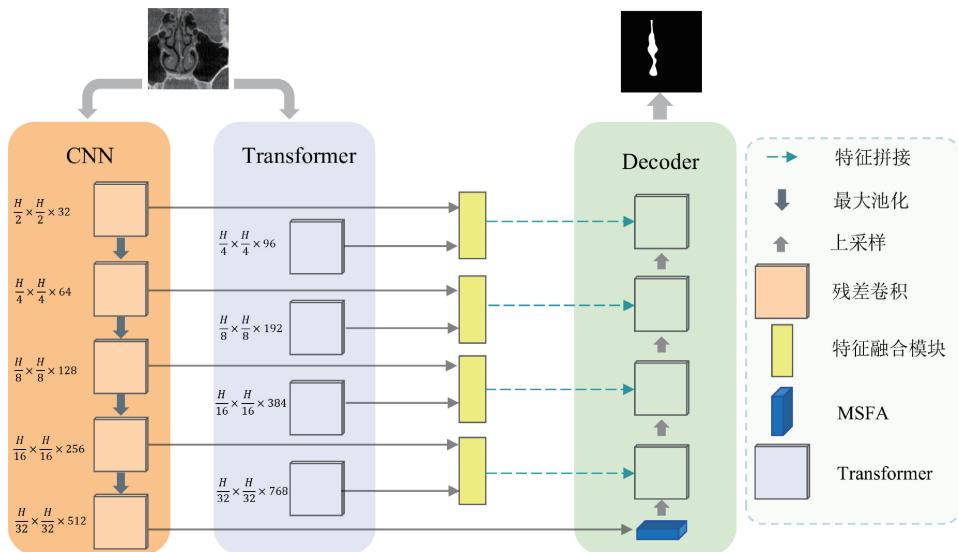


图 1 CTA-Net 模型结构

Fig. 1 Network architecture of CTA-Net

后续章节将对残差卷积模块、MSFA 以及特征融合模块的结构和具体实现进行详细介绍,以进一步揭示它们在提升模型分割性能中的关键作用。

1.2 Transformer 编码器

Swin Transformer 在传统 Transformer 的基础上,通过引入局部窗口自注意力机制和层次化设计,实现了从局部细节到全局语义的逐层的全局特征表示,从而有效降低计算复杂度。Swin Transformer 架构如图 2 所示。本文采用 Swin Transformer 作为双分支编码器中的一支,用以提取输入图像的全局特征,其具体处理流程如下:设输入图像为 X ,首先,将图像划分为大小为 4×4 的不重叠 patch,每个 patch 展平后得到维度为 48 的向量,构成矩阵通过线性映射,将每个 patch 映射到预设的目标维度 $C=96$,得到

的初始特征图 $S_1 = \frac{H}{4} \times \frac{W}{4} \times 96$,为了构建层次化的特征表示,采用块合并(Patch Merging)操作降低空间分辨率并提高特征维度。当前层特征图为 S_1 ,在从 S_1 至 S_2 的过程中,将相邻 2×2 的 patch 的特征进行拼接,得到维度为 $4C=384$ 的向量,然后通过线性变换将其压缩至 $2C=192$ 。重复此过程,即可依次获得更深层次的特征图 S_3 和 S_4 ,从而构成完整的特征分支。在每个 Transformer 层中,Swin Transformer 采用局部窗口自注意力(Window-based Self-Attention)。将特征图划分为大小为 $M \times M$ 的局部窗口,在每个窗口内分别计算查询(Q)、键(K)和值(V),其计算公式如式(1)所示。

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

其中, d_k 表示键向量的维度。局部窗口自注意力能有效捕获局部上下文信息,并通过窗口平移机制实现跨窗口

信息交互,从而逐步构建全局语义表达。

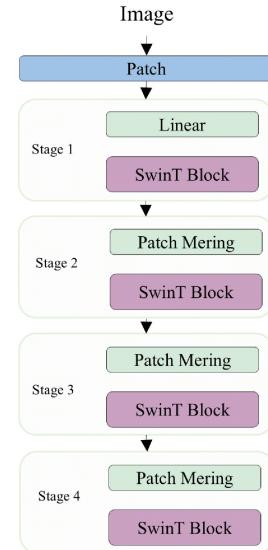


图 2 Swin Transformer 架构

Fig. 2 Swin Transformer architecture

1.3 CNN 编码器

如图 3 所示,本研究设计了一种改进的残差卷积^[28]模块,旨在增强深度特征提取能力,同时有效缓解梯度消失和梯度爆炸问题。该模块采用两层标准 3×3 卷积,并结合批量归一化(Batch Normalization, BN)和 ReLU 激活函数,以保证特征分布的稳定性和非线性表达能力。此外,为了确保不同通道数的输入仍然能够通过残差路径进行有效传播,本文引入了可选的 1×1 卷积分支,用于对输入进行通道调整。输入特征图的通道数为 C_{in} ,输出通道数为 C_{out} ,则该模块的主要计算过程如下:对输入特征 X 进

行第一层 3×3 卷积, 经过 BN 归一化后, 应用 ReLU 激活函数得到特征 X_1 , 在第一层变换得到的特征基础上再次进行 3×3 卷积得到 X_2 , 此时, 若 $C_{in} = C_{out}$, 直接采用恒等映射进行跳跃连接, 若 $C_{in} \neq C_{out}$, 则采用 1×1 卷积调整通道数, 使输入与输出匹配, 最终, 对加权特征 Y 进行 ReLU 激活, 增强非线性表达能力。残差模块在 CNN 编码器的多个层级中被广泛应用, 为后续特征融合模块提供了丰富的局部特征表达能力。计算过程如式(2)~(5)所示。

$$X_1 = \text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3}(X))) \quad (2)$$

$$X_2 = \text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3}(X_1))) \quad (3)$$

$$Y = \begin{cases} X_2 + X_1, & C_{in} = C_{out} \\ X_2 + \text{BN}(\text{Conv}_{1 \times 1}(X_1)), & C_{in} \neq C_{out} \end{cases} \quad (4)$$

$$Y = \text{ReLU}(Y) \quad (5)$$

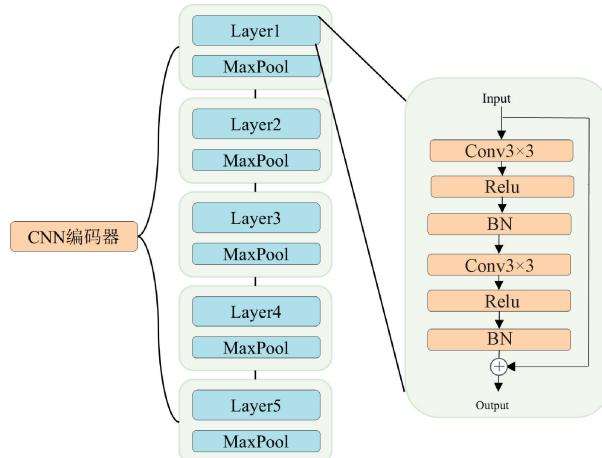


图 3 残差卷积模块

Fig. 3 Residual convolution module

1.4 特征融合模块

本文提出了一个特征融合模块, 用于结合双分支编码器提取的特征。Transformer 编码器分支的输出特征 Y 由于其层级特征维度与 CNN 编码器不同, 因此需要通过转置卷积调整, 使其与 CNN 编码器前 4 层的特征具有相同的维度。经过转置卷积调整后的特征记作 Y' , 采用归一化对 Y' 调整特征分布, 使不同通道的特征均匀化, 提高训练稳定性, 由于特征融合后的特征通道数增加了一倍, 因此使用 1×1 卷积恢复原始通道数, 降低计算复杂度的同时去除冗余特征, 最后, 使用批量归一化和 ReLU 激活函数, 为融合特征引入非线性表达能力, 增强模型的特征表示。特征融合模块如图 4 所示, 该模块保证了来自双分支编码器提取的特征能够高效融合, 提升了分割性能。运行流程如式(6)所示。

$$Out = \text{ReLU}(\text{BN}(\text{Conv}(\text{cat}(X, Y')))) \quad (6)$$

为充分发挥 Transformer 分支与 CNN 分支各自的优势, 在对应层级($S_1 \sim S_4$ 和 $C_1 \sim C_4$)中引入特征融合模块。设 Transformer 分支第 i 层特征为 F_T^i , 而 CNN 分支相应层级特征为 F_C^i , 则特征融合操作如式(7)所示。

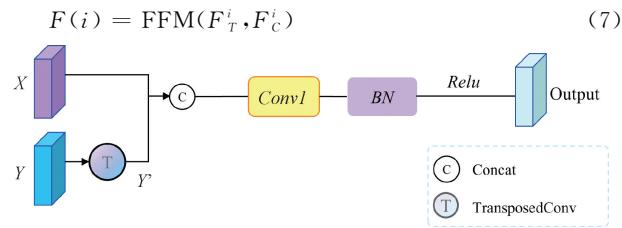


图 4 特征融合模块

Fig. 4 Feather fusion module

1.5 多尺度特征注意力模块 MSFA

为了在鼻中隔分割任务中更高效地融合多尺度上下文信息并精准捕捉关键图像区域, 设计了一个多尺度特征注意力模块, 受 DeepLabv3^[29] 的启发。该模块在保留传统空洞卷积多尺度特征提取优势的同时, 在每个空洞卷积分支中引入了空间注意力机制, 从而实现对目标区域的自适应增强。其详细流程如图 5 所示: 设输入 $X \in R^{C \times H \times W}$ 首先, 利用 1×1 卷积对输入特征进行低维映射和全局信息提取, 生成基础特征, 为后续的多尺度融合提供稳固的初始表示。然后针对不同的扩张率设置, 每个分支采用不同空洞率的空洞卷积, 分别提取具有不同感受野的局部特征。接着, 针对每个分支的卷积输出, 利用空间注意力模块进行特征加权。为了增强全局语义信息, 模块还采用自适应平均池化将输入特征压缩为全局描述, 再通过 1×1 卷积映射、批归一化和 ReLU 激活, 最后利用双线性插值上采样至原始尺寸, 得到全局特征分支。最后, 将上述各分支(1×1 卷积分支、多尺度空洞卷积分支(含空间注意力)以及全局池化分支)的输出在通道维度上进行级联, 并通过 1×1 卷积、批归一化、ReLU 和 Dropout 进行融合, 形成最终的特征表示。

运行流程如式(8)~(12)所示。

$$Y_1 = \text{ReLU}(\text{BN}(W_1 * X)) \quad (8)$$

$$Z_r = \text{ReLU}(\text{BN}(W_r * X)) \quad (9)$$

$$Y_r = \odot(W_s * (\text{Avg}(Z_r), \text{Max}(Z_r))) \oplus Z_r \quad (10)$$

$$Y_g = \text{Up}(\text{ReLU}(\text{BN}(W_g * \text{Gap}(X)))) \quad (11)$$

$$Y_{out} = \text{ReLU}(\text{BN}(\text{Concat}(Y_1, Y_r, Y_g))) \quad (12)$$

其中, X 是输入特征图, W_1, W_r, W_s, W_g 表示卷积核, ReLU 表示 ReLU 激活函数, $*$ 表示卷积, \odot 表示 Sigmoid 激活函数, \oplus 表示逐元素乘积, BN 是批归一化, Avg、Max 为平均池化、最大池化。

2 数据集与实验

2.1 数据集

鼻中隔数据集: 鼻中隔数据集的 CT 图像采集时间为 2023 年 1 月 ~ 2023 年 2 月。该数据集的手动标注工作于 2023 年 3 月 ~ 2023 年 4 月之间完成。该数据集由大连某三甲医院提供的头部锥形束计算机断层扫描(CBCT)图像组成。在临床诊断中, 医生通常会对 CBCT 图像的冠状

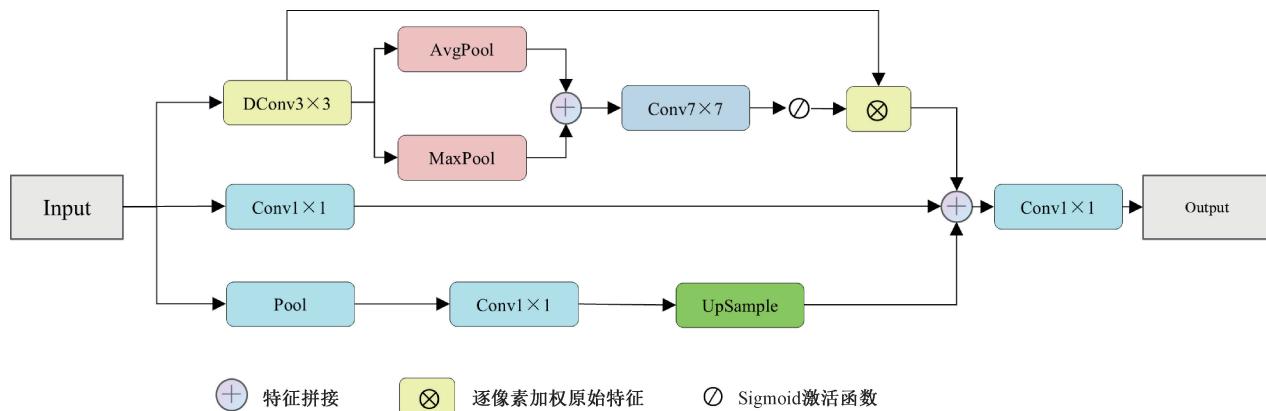


图 5 多尺度特征注意力模块
Fig. 5 Multi-scale features attention model

面、矢状面和横断面进行诊断。冠状面通常可以直接观察到鼻中隔的形态,因此本实验的数据标注是在冠状面上进行的。由于每个 CBCT 图像由 400~500 张切片组成,只有在中间范围的 150~230 张切片中,才能清晰地看到鼻中隔的具体形态。共选择了 69 例患者的头部 CBCT 进行数据标注。每位患者的冠状面切片首先被保存为 2D 图像。临床医生亲自勾画出鼻中隔的具体形态,然后使用 Labelme 标注工具完成数据标注。最终获得了 681 个数据集用于实验。鼻中隔数据集的示例如图 6 所示。从左至右第 1、2 张图片为原数据,第 3、4 张图片为标签数据。



图 6 鼻中隔数据集
Fig. 6 Nasal septum dataset

ISIC2018^[30-31]: ISIC 2018 数据集是一个开放的数据集,专注于皮肤病变图像。该数据集包含超过 2 500 张不同类型的皮肤病变图像,在医学图像处理领域有广泛的应用。许多研究者已使用该数据集来测试所提模型的有效性,并为皮肤病研究的进展做出了重要贡献。

Kvasir^[32]: Kvasir 数据集是一个开放的医学图像分析数据集,专注于胃肠内窥镜图像。该数据集包含大量带标签的图像,特别适用于计算机视觉任务,如病变检测和图像分割。该数据集由挪威 Simula 研究实验室发布,旨在推动机器学习技术在胃肠疾病检测中的应用。

2.2 实验环境与评价指标

实验环境的主要配置:实验操作系统为 Windows 11, GPU 使用 NVIDIA GeForce RTX 3080 Ti, GPU 运行内存为 12 GB, 软件环境使用 Python 3.9 和 Pytorch 1.12.1 框架进行实验。在实验过程中,将批量大小设置为 8, 输入图像大小统一调整为 256×256 的分辨率, 损失函数采用 BCEDiceLoss 损失函数, 学习率计划为 0.000 1, 优化器使

用 Adam 优化器训练所有模型。训练的总轮数为 100。在实验中,所有数据集将按 8 : 2 的比例划分为训练集和验证集。本文实验均采用五折交叉验证方式,保证本文中的所有实验均使用相同的参数设置进行训练,并且在训练过程中不会使用任何预训练权重来训练模型。

在本文实验中,采用了 4 种方法,即 Dice 系数(Dice)、交并比(IoU)、召回率(Recall)和准确率(Acc),用以与其他医学图像分割模型进行比较。评价指标公式如式(13)~(16)所示。

$$Dice = \frac{2TP}{2TP + FN + FP} \quad (13)$$

$$IoU = \frac{TP}{TP + FN + FP} \quad (14)$$

$$Rec = \frac{TP}{TP + FN} \quad (15)$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (16)$$

其中, TP(true positive)为真正例,表示被正确预测为目标类别的像素或区域。FN(false negative)为假负例,表示被错误预测为不属于目标类别的像素或区域。TN(true negative)为真负例,表示被正确预测为不属于目标类别的像素或区域。FP(false positive)为假正例,表示被错误预测为属于目标类别的像素或区域。

2.3 实验结果

为验证 CTA-Net 模型的优势,本研究在鼻中隔数据集开展了详细的消融实验。表 1 展示了在原始双分支编码器基础模型(Base)上,通过逐步加入残差卷积和 MSFA 的改进模块,模型性能的提升情况。其中,单独添加残差连接后,模型 IoU 和 Dice 系数分别提升了约 0.12% 和 0.11%,表明残差卷积模块在缓解深层网络中的梯度消失和特征丢失问题方面具有显著效果;而引入 MSFA 后,性能进一步提升,最终 Base+MSFA+Residual 组合实现了在数据集达到了 IoU 为 90.38 和 Dice 为 94.94 的分割精度。证明了 MSFA 对提升模型分割精度的积极作用,能在

鼻中隔边界多变的结构中更加关注目标区域。消融实验证明缺少任意模块的改进都将导致模型分割性能的下降。

表 1 消融实验

Table 1 Ablation study

模型	IoU	Dice	Recall	Acc	%
Base	89.89	93.57	97.10	99.49	
Base+Residual	90.01	94.67	97.19	99.54	
Base+MSFA	90.24	94.87	97.28	99.71	
Base+MSFA+Residual	90.38	94.94	97.28	99.71	

2.4 对比试验

本文将 CTA-Net 模型与当前主流模型进行对比实验,包括(U-Net、UNet++、Att-Unet、ResUnet^[33]、Medical-T^[34]、Segformer^[35]和ResUnet++^[36])共7种模型在鼻中隔数据集上进行对比实验。如表2所示,本研究提出的CTA-Net在鼻中隔影像分割任务中取得了最优异的综合性能表现。在关键评价指标方面,本方法以90.38%的IoU和94.94%的Dice相似系数达到当前最优水平。相较于UNet,CTA-Net在IoU和Dice指标上分别实现0.43%和0.25%的显著提升。在与其他先进模型的横向对比中,本研究方法展现出持续优势,相比性能仅次于本研究的ResUnet++模型,CTA-Net仍保持0.22%的性能增益。特别值得注意的是,本模型在Recall指标上取得97.32的最高值,较次优模型UNet++提升0.09%,同时维持与其他模型相当的Acc指标。实验结果充分验证了CTA-Net网络架构创新在鼻中隔分割任务中的有效性。

表 2 在鼻中隔数据集实验结果

Table 2 The experimental results on the nasal septum dataset

模型	IoU	Dice	Recall	Acc	%
UNet	89.95	94.69	97.16	99.71	
UNet++	89.66	94.54	97.23	99.70	
Att-Unet	89.59	94.50	96.90	99.70	
ResUnet	89.89	94.66	97.02	99.71	
Medical-T	88.85	94.09	96.79	99.67	
Segformer	89.01	94.18	96.78	99.68	
ResUnet++	90.16	94.82	97.11	99.71	
本文方法	90.38	94.94	97.28	99.71	

2.5 推理时间对比

结合图7可看出本文模型的平均推理时间为0.8823 s,模型推理复杂度仅高于Unet模型,由于Unet模型仅有编码器和解码器以及跳跃连接三部分组成,所以推理速度较快。反观CTA-Net由于加入了残差连接缓解梯度下降,

以及MSFA模块在瓶颈层获取多尺度特征。虽然在推理速度上较高于UNet。但是以牺牲一点推理时间为代价来换取分割精度上的显著优势是很有必要的。同时CTA-Net的推理速度与分割精度优于其他5个对比模型。依然能体现出CTA-Net的优势。

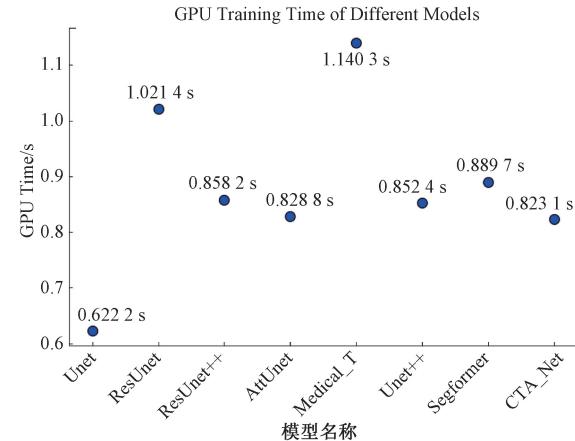


图 7 模型推理时间

Fig. 7 Model inference time

2.6 泛化试验

为了体现CTA-Net在其他数据集上的表现,本文在ISIC2018数据集和Kvasir数据集进行泛化性实验。表3(ISIC2018数据集)展示了上与其他模型的实验对比数据。CTA-Net在分割任务中取得了82.17%的IoU和89.84%的Dice分数。与实验中表现最好的ResUnet模型相比,CTA-Net模型在IoU上提高了0.8%,在Dice分数上提高了0.4%。在Kvasir数据集上(表4)CTA-Net的IoU和Dice系数分别达到76.17%和86.19%,显著优于所有对比模型。相较于UNet++,IoU提升1.13%,Dice系数提升0.73%。这一性能提升表明,CTA-Net对息肉尺寸多变性和边界模糊性具有更强的鲁棒性。在两个医学公开数据集上的优秀表现,证明了CTA-Net具有强大的泛化能力。

表 3 在 ISIC 2018 数据集上的实验结果

Table 3 The experimental results on the

模型	IoU	Dice	Recall	Acc	%
UNet	80.84	89.08	93.07	95.54	
UNet++	81.07	89.21	93.27	95.60	
Att-Unet	81.32	89.42	93.41	95.63	
ResUnet	81.50	89.50	93.17	95.74	
Medical-T	81.01	89.17	92.77	95.65	
Segformer	80.94	88.98	93.88	95.79	
ResUnet++	80.87	89.09	93.62	95.48	
本文方法	82.17	89.84	93.36	95.90	

上述 3 个数据集的实验结果证明了提出的 CTA-Net 模型具有出色的泛化能力。表 1 中的消融实验证明了提出的残差卷积提高了模型的稳定性,而提出的特征融合模块可以有效的融合双分支编码器的特征。此外,添加到瓶颈层的 MSFA 通过学习多尺度特征进一步提升了模型的分割性能。表 2 中关于鼻中隔数据集的对比实验结果证明了本文模型在鼻中隔图像分割中的有效性,这为未来医生研究鼻中隔病例提供了一定的借鉴作用。在表 3 和表 4 中与其他模型在两个公共医学图像数据集上的泛化实验也取得了最好的结果,证明了提出的模型为医学图像分割提供了另一种有效的方法。

图 8 展示了本文 CTA-Net 模型在 3 个数据集上的视觉分割结果,并与其他公开模型进行了比较。前两行是本研究标注的鼻中隔数据集,鼻中隔数据集比其他公共医学数据集局部结构可能并不更复杂但是与周围区域的对比度更低,分割难度较大。局部区域可能会发生过度分割或欠分割现象,这进一步提高了模型对局部特征提取能力的要求。从图中可以看出本文模型在特征提取方面取得了不错的结果,值得注意的是最后两行在 Kvasir 数据集的实

验中,由于数据集的背景和噪声更加复杂导致一些模型的分割效果差强人意,如在 Medical-T 和 ResUnet++ 上甚至出现了乱分割现象。但是本文模型依然取得了最好的分割性能。更加体现了加入残差连接对模型的稳定性起到的良好效果。

表 4 在 Kvasir 数据集上的实验结果

Table 4 The experimental results on the Kvasir dataset

模型	IoU	Dice	Recall	Accuracy	%
UNet	73.55	84.43	90.23	95.44	
UNet++	75.04	85.46	90.75	95.69	
Att-Unet	74.58	85.12	90.63	95.52	
ResUnet	73.24	84.17	90.21	95.34	
Medical-T	59.28	73.85	84.51	92.39	
Segformer	67.82	80.31	88.34	94.30	
ResUnet++	70.14	81.92	89.18	94.55	
本文方法	76.17	86.19	91.60	95.88	

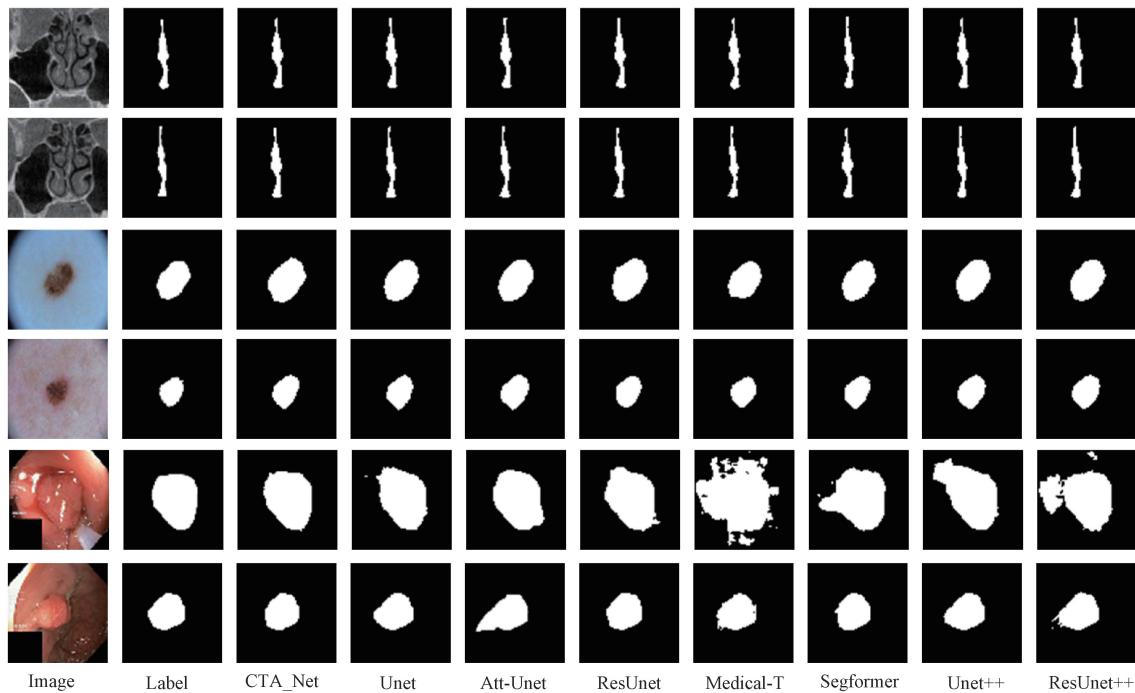


图 8 与其他模型分割效果展示

Fig. 8 Comparison of segmentation results with other models

3 结论

本文针对鼻中隔医学图像分割任务中提出了一种基于双编码器架构的创新模型 CTA-Net。该模型通过 CNN 分支与 Transformer 分支的协同设计,分别捕获解剖结构的局部细节特征和长距离空间依赖关系,并引入多尺度特

征注意力 MSFA 模块与特征融合机制,有效解决了现有方法在复杂解剖结构分割中的局限性。消融实验揭示了双编码器架构的核心贡献,残差卷积模块缓解了深层网络训练中的梯度退化问题,而 MSFA 通过多尺度空洞卷积与注意力加权机制,显著增强了模型对模糊边界的鲁棒性。实验结果表明,CTA-Net 在鼻中隔临床数据集上的分割性能

显著优于主流模型。在泛化实验中体现了模型在跨数据集泛化适应性方面的优势。值得注意的是,设计合适的特征交互将是一项有意义的研究工作。此外,还可以探索其他特征融合方法,以在保留多尺度特征信息的同时,进一步优化模型性能。还可以探索轻量化模型部署方案,满足临床实时性需求。

参考文献

- [1] 董怿,崔顺九,黄谦,等.鼻型NK/T细胞淋巴瘤与慢性鼻窦炎的鉴别诊断要点[J].中国耳鼻咽喉头颈外科,2024,31(10):642-646.
- [2] DONG Y, CUI SH J, HUANG Q, et al. Key points in differential diagnosis of nasal NK/T-cell lymphoma and chronic sinusitis [J]. Chinese Journal of Otorhinolaryngology Head and Neck Surgery, 2024, 31(10): 642-646.
- [3] 朱广田.鼻中隔偏曲手术矫正治疗体会[J].国际感染病学(电子版),2020,9(1):85-86.
- [4] ZHU G T. Experience in surgical correction of deviated nasal septum [J]. International Infectious Diseases(Electronic Edition), 2020, 9(1): 85-86.
- [5] AZIZ T, BIRON V L, ANSARI K, et al. Measurement tools for the diagnosis of nasal septal deviation: A systematic review [J]. Journal of Otolaryngology-Head & Neck Surgery, 2014, 43(1): 11.
- [6] AL-AMRI S S, KALYANKAR N V. Image segmentation by using threshold techniques[J]. ArXiv preprint arXiv:1005.4020, 2010.
- [7] MA W Y, MANJUNATH B S. EdgeFlow: A technique for boundary detection and image segmentation [J]. IEEE Transactions on Image Processing, 2000, 9(8): 1375-1388.
- [8] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [9] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(6):1229-1251.
- [10] ZHOU F Y, JIN L P, DONG J. A survey of convolutional neural networks[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.
- [11] LIU X B, SONG L P, LIU SH, et al. A review of deep-learning-based medical image segmentation methods[J]. Sustainability, 2021, 13(3): 1224.
- [12] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, 2015: 234-241.
- [13] ZHOU Z, RAHMAN S M 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, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer International Publishing, 2018: 3-11.
- [14] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning where to look for the pancreas[J]. ArXiv preprint arXiv:1804.03999, 2018.
- [15] 邹凯鑫,张自嘉,孙伟,等.改进U型网络的路面缺陷图像分割算法[J].电子测量与仪器学报,2024,38(8):15-25.
- [16] ZOU K X, ZHANG Z J, SUN W, et al. Pavement defect image segmentation algorithm based on improved U-Net [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(8): 15-25.
- [17] 陈法法,董海飞,何向阳,等.改进U-Net网络的钢结构表面锈蚀图像分割方法[J].电子测量与仪器学报,2024,38(2):49-57.
- [18] CHEN F F, DONG H F, HE X Y, et al. Image segmentation method for steel structure surface corrosion based on improved U-Net[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(2): 49-57.
- [19] ALOM M Z, YAKOPCIC C, HASAN M, et al. Recurrent residual U-Net for medical image segmentation[J]. Journal of Medical Imaging, 2019, 6(1): 014006.
- [20] CAI S J, TIAN Y X, LUI H, et al. Dense-UNet: A novel multiphoton *in vivo* cellular image segmentation model based on a convolutional neural network [J]. Quantitative Imaging in Medicine and Surgery, 2020, 10(6): 1275.
- [21] HUANG H M, LIN L F, TONG R F, et al. UNet 3+: A full-scale connected unet for medical image segmentation[C]. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 1055-1059.
- [22] VASWANI A, SHAZEEB N, PARMAR N, et al. Attention is all you need[J]. ArXiv preprint arXiv: 1706.03762, 2017.
- [23] 刘文婷,卢新明.基于计算机视觉的Transformer研究进展[J].计算机工程与应用,2022,58(6):1-16.
- [24] LIU W T, LU X M. Research progress of transformer based on computer vision[J]. Computer Engineering and Applications, 2022, 58(6): 1-16.

- [19] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [J]. ArXiv preprint arXiv:2010.11929, 2020.
- [20] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
- [21] CHEN J N, LU Y Y, YU Q H, et al. TransUNet: Transformers make strong encoders for medical image segmentation[J]. ArXiv preprint arXiv:2102.04306, 2021.
- [22] CHANG Y, MENGHAN H, GUANGTAO Z, et al. Transclaw U-Net: Claw U-Net with transformers for medical image segmentation [J]. ArXiv preprint arXiv:2107.05188, 2021.
- [23] YAO CH, TANG J Y, HU M H, et al. Claw U-Net: A Unet-based network with deep feature concatenation for scleral blood vessel segmentation [J]. ArXiv preprint arXiv:2010.10163, 2020.
- [24] LIN AI L, CHEN B ZH, XU J Y, et al. DS-transUNet: Dual swin transformer U-Net for medical image segmentation [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-15.
- [25] HEIDARI M, KAZEROUNI A, SOLTANY M, et al. Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation[C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2023: 6202-6212.
- [26] LIANG ZH W, ZHAO K, LIANG G, et al. MAXFormer: Enhanced transformer for medical image segmentation with multi-attention and multi-scale features fusion[J]. Knowledge-Based Systems, 2023, 280: 110987.
- [27] PETIT O, THOME N, RAMBOUR C, et al. U-Net transformer: Self and cross attention for medical image segmentation[C]. Machine Learning in Medical Imaging: 12th International Workshop, 2021: 267-276.
- [28] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [29] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
- [30] CODELLA N, ROTEMBERG V, TSCHANDL P, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration(isic)[J]. ArXiv preprint arXiv: 1902.03368, 2019.
- [31] WEN D, KHAN S M, XU A J, et al. Characteristics of publicly available skin cancer image datasets: A systematic review [J]. The Lancet Digital Health, 2022, 4(1): e64-e74.
- [32] POGORELOV K, RANDEL K R, GRIWODZ C, et al. Kvadir: A multi-class image dataset for computer aided gastrointestinal disease detection[C]. 8th ACM on Multimedia Systems Conference, 2017: 164-169.
- [33] DIAKOGIANNIS F I, WALDNER F, CACCETTA P, et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 162: 94-114.
- [34] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical transformer: Gated axial-attention for medical image segmentation [C]. Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, 2021: 36-46.
- [35] XIE EN Z, WANG W H, YU ZH D, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers [J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.
- [36] 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). IEEE, 2019: 225-2255.

作者简介

周宝康,硕士研究生,主要研究方向为医学图像分割、计算机视觉。

E-mail:zhoumostu@163.com

曹爽,硕士研究生,主要研究方向为信息智能检测。

E-mail:3168465755@qq.com

高洪涌,硕士研究生,主要研究方向为医学图像分割。

E-mail:17806248628@163.com

宋维波(通信作者),博士,讲师,硕士生导师,主要研究方向为计算机视觉、医学影像分析。

E-mail:swb@dlou.edu.cn

崔树林,硕士,主任医师,硕士生导师,主要研究方向为耳鼻喉科。

E-mail:57053332@qq.com