

中图法分类号: TP37 文献标识码: A 文章编号: 1006-8961(2026)01-0303-17

论文引用格式: Zhu Z Q, Sun M W, Qi G Q, Li Y Y, Yang M J, Cheng J and Liu Y. 2026. Frequency adaptation and feature transformation network for medical image segmentation. Journal of Image and Graphics, 31(1):0303-0319(朱智勤, 孙梦薇, 齐观秋, 李媛源, 杨梦杰, 程俊, 刘羽. 2026. 融合频率自适应和特征变换的医学图像分割. 中国图象图形学报, 31(1):0303-0319)[DOI:10.11834/jig.250100]

融合频率自适应和特征变换的医学图像分割

朱智勤¹, 孙梦薇¹, 齐观秋², 李媛源^{1*}, 杨梦杰³, 程俊⁴, 刘羽⁵

1. 重庆邮电大学自动化学院, 重庆 400065; 2. 纽约州立大学布法罗州立学院, 布法罗 14222, 美国;
3. 盛云科技有限公司, 昆明 650000; 4. 中国科学院先进技术研究院, 深圳 518000;
5. 合肥工业大学生物医学工程学院, 合肥 230009

摘要: 目的 医学图像中病灶的精准分割对临床诊断和治疗制定至关重要。虽然现有方法多利用深度特征、注意力机制或多尺度结构来提升性能,但对高低频特征的显式建模不足,下采样过程中易产生信息丢失。本文提出一种融合频率自适应与特征变换的分割网络(frequency adaptation and feaTNet),通过动态平衡高低频分量并优化下采样策略,以提升分割的准确性。**方法** 该方法是基于编码器—解码器结构实现的,主要包括频率自适应编码模块(frequency adaptation encoding module, FAE)、特征分解变换模块(feature decomposition transformation module, FDT)和空间通道信息重构模块(spatial-channel information reconstruction module, SCIR)。频率自适应编码模块用于特征表示学习中的高低频分量平衡,在特征编码过程采用动态扩张率和自适应核策略;特征分解变换模块应用于下采样过程,通过并行最大池化和离散小波变换平衡操作,实现细节感知信息的恢复和保持;空间通道信息重构模块采用分离重构策略及分离变换融合操作实现解码过程的信息特征交叉重建并抑制空间通道维度特征冗余。**结果** 在ISIC2017(international skin imaging collaboration 2017)、ISIC2018和DSB2018(data science bowl challenge 2018) 3个公开数据集中进行训练和验证。在ISIC2017和ISIC2018数据集上的Dice得分为89.42%和89.84%,分别优于其他对比方法0.04%~7.83%和0.13%~5.95%。在ISIC2018基准上的敏感性(sensitivity, SE)指标达到91.19%,具有明显优势。在DSB2018数据集上,本文模型的Dice系数(Dice coefficient)、准确度(accuracy, ACC)和敏感性(SE)指标分别为91.52%、97.65%和91.88%,相较于较先进的算法分别提升2.59%、2.29%和2.86%。**结论** 本文提出的分割模型,充分利用了图像中的高频和低频信息,有效提升了分割性能。

关键词: 深度学习;医学图像分割;卷积神经网络(CNN);频率自适应;特征分解变换

Frequency adaptation and feature transformation network for medical image segmentation

Zhu Zhiqin¹, Sun Mengwei¹, Qi Guanqiu², Li Yuanyuan^{1*}, Yang Mengjie³, Cheng Jun⁴, Liu Yu⁵

1. College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;
2. Computer Information Systems Department, State University of New York at Buffalo State, Buffalo 14222, USA;
3. Sheng Yun Technology Co., Kunming 650000, China; 4. Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518000, China; 5. Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China

收稿日期: 2025-03-20; 修回日期: 2025-07-12; 预印本日期: 2025-07-19

* 通信作者: 李媛源 liyy@cqupt.edu.cn

基金项目: 国家自然科学基金项目(62576062, 62276037, 62576132, U23A20294, 62176081); 重庆青年创新人才项目(CSTB2024NSCQ-QCXMX0067); 重庆市教育委员会科学技术研究项目(KJZD-M202500604); 重庆市技术创新与应用发展专项重大项目(CSTB2024TIAD-STX0023); 云南省科技人才与平台计划项目(202305AF150152)

Supported by: National Natural Science Foundation of China(62576062, 62276037, 62576132, U23A20294, 62176081)

Abstract: Objective Precise segmentation of lesions in medical images is of great significance for clinical diagnosis and treatment planning. The development of automated, accurate, and robust medical image segmentation methods is critical for the advancement of computer-aided diagnosis. In recent years, the U-Net architecture based on convolutional neural networks (CNNs) has become a mainstream approach for medical image segmentation. However, the limited receptive fields of CNNs constrain their ability to capture widely distributed low-frequency features in images, including background information and large-scale structures. Inspired by the attention mechanism in natural language processing, Transformer, a method that can focus on global context information, was introduced into the medical image segmentation task. Medical image segmentation methods based on CNNs and Transformer have achieved good segmentation performance, but they mainly focus on extracting global semantic features and ignore low-frequency features such as background information and large-scale structures, as well as high-frequency detailed features widely distributed in images. This study proposes a medical image segmentation network (i. e., FAFTNet) that integrates frequency adaptation and feature transformation, aiming to achieve accurate segmentation through the dynamic balance of high- and low-frequency components and the optimization of downsampling strategies. **Method** This method is implemented on the basis of the U-Net encoder-decoder architecture and mainly consists of frequency adaptation encoding (FAE), feature decomposition transformation (FDT) module, and spatial-channel information reconstruction (SCIR) modules. The FAE module is used to balance high- and low-frequency components in feature representation learning and adopts adaptive dilation rates and kernel strategies in the feature encoding process. During the downsampling process, the FDT module is utilized to perform parallel max pooling and discrete wavelet transform balancing operations, aiming to restore and enhance the preservation of fine-grained details. Features processed through the downsampling stage are passed to the decoder stage, in which the first three stages are symmetrically aligned with the encoder. In the last three stages, the SCIR module is adopted, using a separation-based reconstruction strategy and separate transformation fusion operations to achieve cross reconstruction of information features while suppressing redundant spatial-channel dimension features. In the decoding phase, implicit features containing local and global information are decoded through multiple upsampling steps, gradually generating full-resolution segmentation maps. **Result** The proposed method was trained and validated on three public datasets: international skin imaging collaboration (ISIC) 2017, ISIC 2018, and data science bowl (DSB) 2018. Four evaluation metrics, namely, Dice coefficient, accuracy (ACC), specificity, and sensitivity (SE), were used to evaluate the effectiveness of the proposed method. On the ISIC 2017 and 2018 datasets, the proposed method indicated Dice scores of 89.42% and 89.84%, respectively, outperforming other comparison methods by 0.04%~7.83% and 0.13%~5.95%, respectively. Additionally, the proposed method achieved an SE score of 91.19 on the ISIC 2018 benchmark, demonstrating clear advantages. On the DSB dataset, the model achieved Dice coefficient, ACC, and SE scores of 91.52%, 97.65%, and 91.88%, respectively, presenting improvements of 2.59%, 2.29%, and 2.86%, respectively, over state-of-the-art methods. The proposed method also demonstrated superior performance in the visualization results, i. e., it effectively predicted the target segmentation area in images. Furthermore, under various interference factors, the proposed method still obtained better segmentation results than other methods, and the results closely matched the ground truth. Ablation experiments verified the effectiveness of the proposed model and its individual modules. **Conclusion** The segmentation model proposed in this paper fully utilizes the high- and low-frequency information in images and effectively improves the segmentation performance. In future work, the focus will be on enhancing model adaptability across datasets and domains through transfer and multitask learning. The performance and application value of the network will be further improved by combining multitask learning with other tasks such as image classification and feature detection.

Key words: deep learning; medical image segmentation; convolutional neural network (CNN); frequency adaptation; feature decomposition transformation

0 引言

随着临床医生用于诊断、治疗和监测病情目的的医疗数据量的增加,开发自动、准确和鲁棒的医学图像分割方法对计算机辅助诊断的发展至关重要。医学图像分割的目标是从医学图像中准确定位解剖或病理结构,利用产生的像素定位信息辅助诊断疾病、计划治疗干预或监测病情进展(Conze等,2023)。受益于计算资源进行的精确、快速、可重复和客观的测量,医学图像自动分割技术不仅能够减少手工标注时间人力成本、提高工作效率,而且能够更清晰地识别和分析病变区域以增强诊断准确性。然而,由于图像采集过程存在噪声伪影和质量差异问题以及目标区域本身存在的结构复杂、形态差异和边缘模糊问题,医学图像分割面临技术层面、数据层面和应用层面等多种挑战。

传统医学图像分割方法通常基于阈值分割、区域生长和边缘检测。然而,这些传统方法通常需要人工干预或超参数微调才能有效选择特征,达到良好的效果。随着深度学习的发展,许多应用于医学图像分割任务的模型相继提出。卷积神经网络(convolutional neural network, CNN)由于强大的特征表示能力,广泛应用于医学图像分割任务(Yao等,2024)。其中,全卷积网络(fully convolutional network, FCN)(Long等,2015)能够在任意尺寸图像上进行处理。基于全卷积的编码器—解码器模型U-Net(Ronneberger等,2015)用于医学图像分割任务并逐渐扩展了许多变体结构,包括UNet++(Zhou等,2018)、EGEUNet(efficient group enhanced U-Net)(Ruan等,2023)。Ruan等人(2022)提出基于U形结构和注意力模块的MALUNet(multi-attention and light-weight U-Net)。基于CNN的模型在医学图像分割中取得了良好的表现。然而上述方法由于CNN侧重关注局部特征学习一定程度忽略了背景信息和大尺度结构等低频特征(Park和Kim,2022),无法有效识别病变区域与正常区域的过渡区域,从而使性能停滞不前。

受到自然语言处理中的注意机制(Vaswani等,2017)的启发,Transformer这一能够聚焦全局上下文信息的方法被引入医学图像分割任务中,在密集预测任务中具有明显的优势。DC-Net(dual context network)(Xu等,2021)、TransFuse(fusing Transform-

ers and CNNs for medical image segmentation)(Zhang等,2021)、Swin-UNet(UNet-like pure Transformer for medical image segmentation)(Cao等,2023)、TransAS-UNet(regional segmentation of breast cancer Swin Transformer and of UNet algorithm)(徐旺旺等,2024)和DGLCANet(dual-encoder global-local cross-attention network)(李赫等,2024)等一些基于纯粹Transformer或混合Transformer与CNN的编解码网络能够有效实现全局特征建模。Chen等人(2021)首次提出TransUNet(Transformers make strong encoders for medical image segmentation)这一Transformer和CNN的混合架构来探索Transformer在医学图像分割中的潜力。然而,简单的Transformer模型由于缺少图像相关的先验知识会导致模型收敛慢,实际应用性能比较低。Wang等人(2021)使用边界感知Transformer(boundary-aware Transformer, BAT)将边界先验知识纳入分割模型中,以处理皮肤病变图像中的模糊边界。然而医学图像中目标区域可能出现不同的尺度、不规则的形态和不同的颜色分布,注意力机制通过加权平均的方式聚焦于重要特征,容易导致高频细节特征处理不足,丢失边缘纹理等细节信息。同时由于复杂计算策略,注意力机制在内存占用方面存在一定局限性(Zhu等,2023)。

状态空间模型(state space model, SSM)引起了研究界的极大关注。现有SSM方法如Mamba(linear-time sequence modeling with selective state spaces)(Gu和Dao,2023)不仅能够捕获长距离依赖关系,而且支持并行训练,从而大幅降低模型参数的复杂性。Mamba算法模型创新地将时变参数引入SSM。与传统的Transformer相比,Mamba以更少的参数展示了卓越的性能。一系列基于Mamba的医学图像分割模型相继提出,例如VM-UNet(vision Mamba UNet)(Ruan等,2024)、Mamba-UNet(UNet-like pure visual Mamba for medical image segmentation)(Wang等,2024)和H-vmunet(high-order vision Mamba UNet for medical image segmentation)(Wu等,2025)。Li等人(2025)提出可以根据病变特点动态调整局部和全局特征的选择性融合Mamba(SF-Mamba)和跨尺度捕捉全局信息的多尺度融合Mamba(MF-Mamba)。Zhu等人(2025)提出一种具有强归纳偏差的视觉稳定Mamba U型网络VSMU-Net用于脑肿瘤分割。尽管基于Mamba的方法在医

学图像分割中展现出良好的潜力,但其由于为序列建模设计,在处理复杂医学图像时,空间感知能力可能存在局限。此外,Mamba的状态更新过程在高分辨率图像中可能导致额外计算开销。

U-Net的编解码架构极大促进了包括皮肤病变分割等医学图像分割任务性能的提升。在编解码网络中,为了提取深层次特征、压缩信息并在一定程度减少内存需求,执行下采样这一关键操作逐步减少特征图空间分辨率。在FCN和U-Net等网络中主要采用最大池化、平均池化、随机池化和跨度卷积等操作来提取深层次特征并逐步减少特征图空间分辨率(周涛等,2021)。Chen等人(2022)提出学习分数下采样的卷积块。尽管一些方法(Sarker等,2021; Roy等,2023)从多尺度上下文信息以及特征增强的角度恢复部分信息,但它们仅作用于局部区域特征,这可能仍无法避免关联细节信息的丢失(Xu等,2023)。

基于CNN和Transformer的医学图像分割方法取得了良好的分割性能,但它们大多都侧重于全局语义特征的提取,对图像中广泛分布的低频特征,包括背景信息和大尺度结构以及高频细节特征的显式建模仍显不足。同时,如何减少编解码过程中下采样操作导致的信息丢失对于提高分割质量具有重要意义。本文提出一种基于编码器—解码器架构的频率自适应分割网络,旨在通过高低频分量动态平衡和下采样策略优化,实现更精准的分割。该方法主要包括频率自适应编码模块(frequency adaptation encoding module, FAE)、特征分解变换模块(feature decomposition transformation module, FDT)和空间通道信息重构模块(spatial-channel information reconstruction module, SCIR)。频率自适应编码模块在特征编码过程采用动态扩张率和自适应核策略,实现特征表示学习中的高低频分量平衡。下采样过程引入特征分解变换模块通过并行最大池化和离散小波变换平衡操作处理每一级编码特征,实现细节感知信息的恢复和保持。在解码阶段的空间通道信息重构模块采用分离重构策略及分离变换融合操作,实现信息特征交叉重建并抑制空间通道维度特征冗余。通过以上设计,本文提出的分割框架可以有效地提取和融合高低频特征并减少信息丢失,在ISIC2017(international skin imaging collaboration 2017)、ISIC2018和DSB2018(data science bowl challenge 2018)3个公开数据集上进行了大量实验来验

证本文方法的有效性。

本文工作主要有3个方面的贡献:1)提出一种频率自适应分割网络用于医学图像分割,可以动态平衡高低频分量,实现感受野扩张来捕获有效特征。构建频率自适应编码模块,采用动态扩张率和自适应核策略进行卷积加权分解,通过空间变化的更新加权平衡特征表示中的高频和低频分量。2)提出特征分解变换模块,实现下采样过程中的细节感知信息保持。通过通道分解操作将编码特征转换为两个分支,分别进行最大池化和离散小波变换平衡,在减少信息丢失的同时进一步获取全局和局部细节特征。3)提出空间通道信息重构模块,利用分离重构策略实现信息特征交叉重建,抑制空间维度特征冗余,通过分离变换和融合操作减少通道维度冗余,降低计算开销和复杂度的同时增强特征表示学习。

1 研究方法

1.1 模型概述

如图1所示,本文提出的频率自适应分割网络架构建立在对称编码器—解码器框架之上,主要由频率自适应编码模块、特征分解变换模块和空间通道信息重构模块组成。频率自适应编码模块作为前3个阶段编码器采用动态扩张率和自适应核策略动态平衡高频和低频分量,后3个阶段利用组多轴哈达玛卷积模块(Ruan等,2023)从不同角度提取表示信息。在下采样过程利用特征分解变换模块实现并行最大池化和离散小波变换平衡操作,以恢复和增强保留细粒度细节。经过下采样过程处理的特征传递到解码器阶段,前3个阶段与编码器相对称,最后3个阶段利用空间通道信息重构模块,采用分离重构策略及分离变换融合操作实现信息特征交叉重建并抑制空间通道维度特征冗余。解码阶段经过多个上采样步骤对包含局部和全局信息的隐式特征进行解码,逐步生成全分辨率分割图。

1.2 频率自适应编码模块

如图2所示,给定输入图像, H 和 W 表示高度和宽度, C 为通道维度,经过最大池化和GeLU(Gaussian error linear unit)激活、组归一化、 1×1 卷积操作对输入特征图进行非线性处理和特征转换。通过频率自适应编码模块动态平衡高低频分量,实现感受野扩张来捕获有效特征。其中,高频分量中包含边缘细节信

息,增强高频分量能够帮助提取关键细节特征;低频分量显示图像基本形状和较大的结构变化,可以减少背景对病变区域分割的干扰,低频信息有助于保持全局一致性和结构信息。首先利用离散傅里叶变换将通道特征映射 $Y \in \mathbf{R}^{H \times W}$ 变换到频域,可以表示为

$$Y_F(m, n) = \frac{1}{HW} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} Y(h, w) e^{-2\pi j(mh + nw)} \quad (1)$$

式中, $Y_F \in \mathbf{R}^{H \times W}$ 表示经过傅里叶变换后的复数输出数组, h, w 表示特征图 Y 的坐标。高度和宽度维度的归一化频率由 $|m|$ 和 $|n|$ 给出。将低频移至中心后, m 取自集合 $\{-H/2, -(H+1)/2, \dots, (H-1)/2\}$, 而 n 取自 $\{-W/2, -(W+1)/2, \dots, (W-1)/2\}$ 。

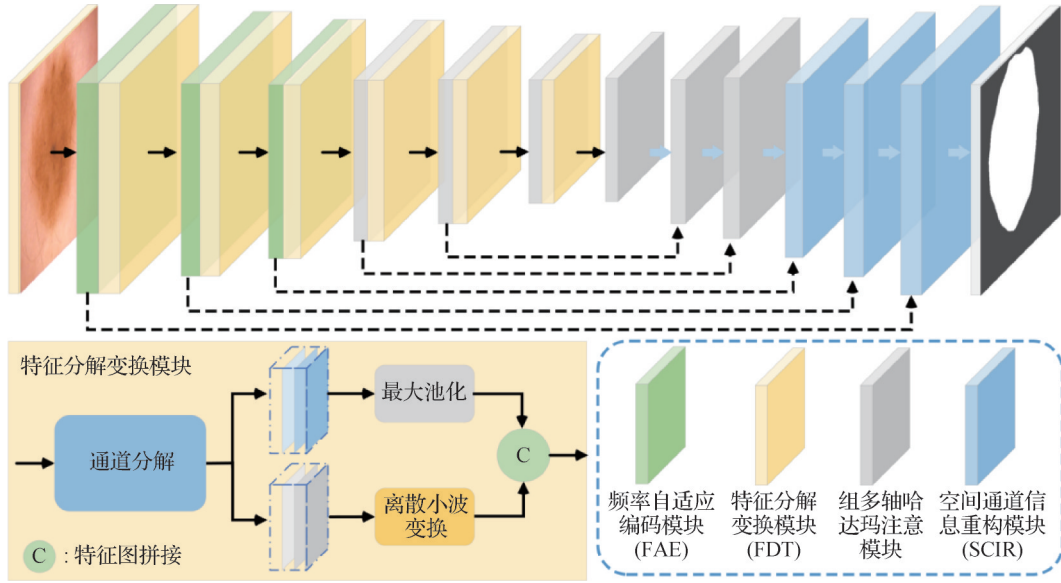


图1 频率自适应分割网络架构

Fig. 1 Overall structure of frequency adaptation segmentation network

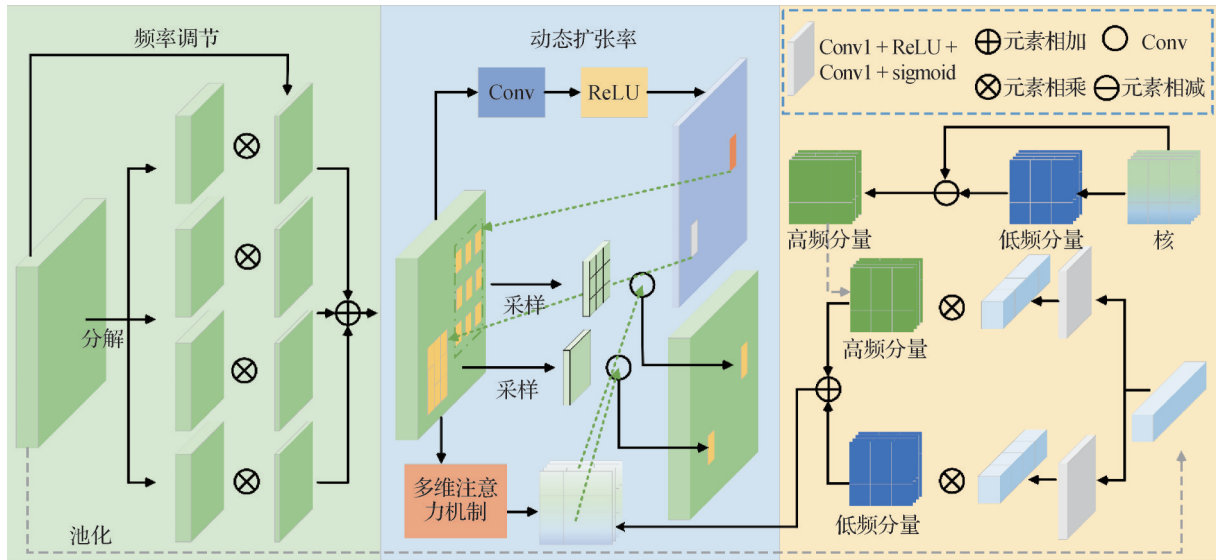


图2 频率自适应编码模块

Fig. 2 Frequency adaptation encoding module

为了权衡感受野和有效带宽,本文方法采用动态扩张率策略为每个像素分配不同扩张率。具体为

$$Z(p) = \sum_{i=1}^{K \times K} W_i Y(p + \Delta p_i \times \hat{D}(p)) \quad (2)$$

式中, $Z(p)$ 表示输出特征图中位置 p 处的像素值, K

是核大小, W_i 表示核的权重参数, 而 $Y(p + \Delta p_i)$ 表示输入特征图中与 p 偏移 Δp_i 相对应的位置处的像素值。变量 Δp_i 表示预定义采样的第 i 个位置 $\{(-1, -1), (-1, 0), (-1, +1), \dots, (+1, +1)\}$ 。 $\hat{D}(p)$ 表示位置 p 处的扩张率, 与感受野相关联, 可以用参数为 θ 的卷积层进行预测。采用扩张调节机制 (Zhu 等, 2019) 最大化感受野, 最小化每个像素丢失的频率信息。对于以 p 为中心、窗口大小为 s 的局部特征 $Y^{(p,s)}$, 其感受野 $\Phi(p) = (K - 1) \times \hat{D}(p) + 1$ 与 $\hat{D}(p)$ 呈正相关。一组 $\tau_{\hat{D}(p)}^+$ 中的频率难以被准确捕获, 因此, 可以通过计算高频功率 $HP(p) = \sum_{\tau_{\hat{D}(p)}^+} |Y_F^{(p,s)}(m, n)|^2$ 测量丢失的频率信息。 θ 的优化可以表示为

$$\theta = \max_{\theta} \left(\sum \Phi(p) - \sum HP(p) \right) \quad (3)$$

然而, 由于频率集合 $\tau_{\hat{D}(p)}^+$ 的离散性质以及 $HP(p)$ 的计算不可微, 选择直接优化 $\hat{D}(p)$, 即在 $HP(p)$ 较低的位置增加扩张率以扩大感受野, 而在 $HP(p)$ 较高的位置通过抑制扩张率减少频率信息的损失, 优化过程表示为

$$\theta = \max_{\theta} \left(\sum_{p \in HP^+} \hat{D}(p) - \sum_{p \in HP^-} \hat{D}(p) \right) \quad (4)$$

式中, HP^+ 和 HP^- 分别表示具有最高和最低高频像素。

通过动态扩张率实现有效带宽和感受野之间的平衡优化后, 为了进一步提高有效带宽, 采用自适应核策略进行卷积加权分解, 将卷积核参数分解为低频和高频分量, 在多通道上动态地操作这两个分量, 增加高频核权重会增强对高频分量的响应, 进而增加有效带宽, 通过引入动态加权调整频率响应。对于静态卷积核, 其权重 ω 可以分解为

$$\omega = \bar{\omega} + \hat{\omega} \quad (5)$$

式中, $\bar{\omega} = \frac{1}{K \times K} \sum_{i=1}^{K \times K} \omega_i$ 表示逐核平均 ω 。其功能为低通 $K \times K$ 均值滤波器, 然后用 $\bar{\omega}$ 定义参数进行 1×1 卷积。较高的平均值更有可能衰减高频分量。 $\hat{\omega} = \omega - \bar{\omega}$ 表示残差部分, 能够捕捉局部差异并提取高频分量。经过分解后, 动态核策略动态调整高频和低频分量, 可以表示为

$$\omega' = \delta_l \bar{\omega} + \delta_h \hat{\omega} \quad (6)$$

式中, δ_l, δ_h 是每个通道的动态权重, 通过全局池化和

卷积层得到。根据输入上下文动态调整 δ_l/δ_h 的比例, 期望网络关注特定频段并适应复杂特征, 通过动态频率自适应增强低频上下文和高频局部细节的捕获能力。

传统卷积操作生成的特征图往往包含较高比例的高频分量, 导致需要采用较小的总扩张率保持高有效带宽, 然而这种权衡会限制感受野。为了优化特征表示中高频与低频分量的平衡, 并增强感受野的有效学习, 引入了频率调节策略。频率调节策略将特征图从低到高分解为多个频率通道。然后, 使用调节特征图对不同频率通道进行重新加权, 以平衡频率功率, 从而有效地学习更大的感受野。首先在傅里叶域中应用不同的掩码将特征分解为不同频段, 即

$$Y_b = F^{-1}(M_b Y_F) \quad (7)$$

$$M_b(m, n) = \begin{cases} 1 & \vartheta_b \leq \max(|m|, |n|) < \vartheta_{b+1} \\ 0 & \text{其他} \end{cases} \quad (8)$$

式中, F^{-1} 表示快速傅里叶逆变换。 M_b 是用于提取相应频率的二进制掩码, $\vartheta_b, \vartheta_{b+1}$ 来自 $B + 1$ 个预定义的频率阈值 $\{0, \vartheta_1, \vartheta_2, \dots, \vartheta_{B+1}, 1/2\}$ 。然后, 在空间上对不同频段的频率分量进行动态重新加权, 可以表示为

$$\hat{Y}(i, j) = \sum_{b=0}^{B-1} A_b(i, j) Y_b(i, j) \quad (9)$$

式中, $\hat{Y}(i, j)$ 是在频率调节之后学习的频率平衡特征, A_b 表示第 b 频带的选择映射。

通过自适应频率编码操作, 可以将输入的特征图进行有效的频率特征平衡, 将经过频率特征平衡阶段的特征图输入到可变形卷积中进行进一步的深层多尺度特征提取, 实现对深层语义特征的学习。在通过 FAE 模块实现频率特征平衡输出阶段特征图后, 应用多轴注意融合机制在增强特征交互的同时提取深层次语义信息。将经过下采样操作的特征图输入沿通道维度进行分组, 对每组子特征计算空间轴和通道轴注意力权重, 通过哈达玛乘积操作实现逐元素相乘的操作, 用于融合不同轴上的注意力权重。通过哈达玛乘积操作替代传统矩阵乘法, 在减少计算复杂度的同时保留局部特征细节。最后将不同分组特征沿通道维度连接起来, 执行深度可分离卷积以整合来自不同角度的信息特征并得到编码阶段输出的特征图。

1.3 特征分解变换模块

在分割任务中,下采样操作广泛用于聚合局部特征和扩大接受域。然而,在局部区域上聚集特征可能会导致重要空间信息的丢失,这不利于逐像素预测。为了实现下采样过程中的细节感知信息保持,提出特征分解变换模块,通过通道分解操作将编码输出特征转换为两个分支,分别进行最大池化操作和离散小波变换平衡。离散小波变换平衡利用 Haar 小波变换层,在保留细节信息的同时有效降低特征映射的空间分辨率。Haar 小波变换是一种基于正交函数的变换,广泛应用于信号处理、图像编码和边缘提取。1 级一维 Haar 变换的小波基函数用于提取图像信号中的高频分量,尺度函数用于图像信号的平滑处理和近似表示,可定义为

$$\begin{cases} \varphi_1(x) = \frac{1}{\sqrt{2}} \varphi_{1,0}(x) + \frac{1}{\sqrt{2}} \varphi_{1,1}(x) \\ \psi_1(x) = \frac{1}{\sqrt{2}} \varphi_{1,0}(x) - \frac{1}{\sqrt{2}} \varphi_{1,1}(x) \end{cases} \quad (10)$$

式中, $\varphi_{i,j}(x)$ 定义为

$$\varphi_{i,j}(x) = \sqrt{2^i} \varphi(2^i x - j), j = 0, 1, \dots, 2^i - 1 \quad (11)$$

式中,参数 i 和 j 分别表示 Haar 基函数在图像处理域中的尺度和在二维图像中表示的方向,定义为

$$\varphi_{0,0}(x) = \varphi_0(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x < 1 \\ 0 & x \geq 1 \end{cases} \quad (12)$$

因此,1 级 Haar 变换可以用 0 级 Haar 基函数表示为

$$\begin{cases} \varphi_1(x) = \varphi_0(2x) + \varphi_0(2x - 1) \\ \psi_1(x) = \varphi_0(2x) - \varphi_0(2x - 1) \end{cases} \quad (13)$$

式中, φ_1 和 ψ_1 分别代表低通和高通滤波器,能够从图像中提取低频信息和高频信息。式(13)将原始图像信号分解为两个长度减半的子带信号,分别对应不同频段的图像特征。如图 3 所示,将 Haar 小波变换应用于二维图像信号时,进一步分解产生 4 个分量,对应为在水平、垂直和对角线方向上的低频分量和高频分量。每个分量的分辨率降低到原来的 1/4,而特征映射的通道数增加了 4 倍。Haar 小波变换可以将部分空间维度信息编码到通道维度,而不会丢失信息。最后通过 1×1 卷积调整特征映射的通道数并过滤冗余信息,与经过最大池化操作后的特征合并输出。

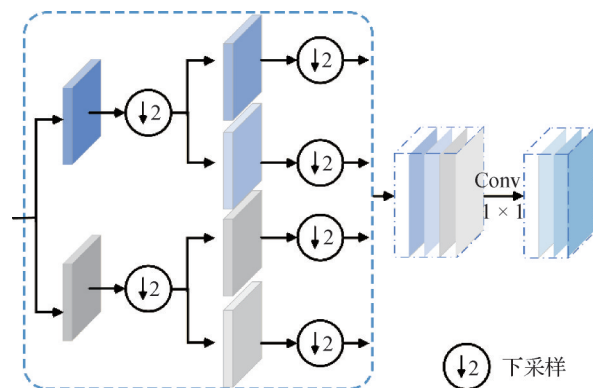


图3 离散小波变换

Fig. 3 Discrete wavelet transform

1.4 空间通道信息重构模块

在特征映射的空间和通道维度存在特征冗余,可能导致计算资源浪费、训练效率降低或增加过拟合风险。为了在降低特征冗余的同时提升模型效率,本文提出空间通道信息重构模块,应用于特征解码过程。对于编码器处理的中间特征 X_m ,利用空间特征冗余,采用分离重构策略实现信息特征交叉重建。分离操作将信息内容差异较大的特征图分离开。首先标准化中间特征,利用 GN (group normalization) 层中的可训练参数 r 作为衡量每个批次和通道的空间像素方差的方法,呈正相关。对参数进行归一化得到权重 ω_r ,将经 ω_r 重新加权后的特征映射的权值通过 sigmoid 函数映射到 $(0, 1)$ 范围,并通过阈值进行门控。通过权重设置分别得到信息权重 ω_1 ,和非信息权重 ω_2 。获取 ω 的过程表示为

$$\omega = \text{Gate} \left(\text{sigmoid} \left(\omega_r \left(\text{GN} (X_m) \right) \right) \right) \quad (14)$$

然后将输入特征 X_m 分别乘以 ω_1 和 ω_2 ,得到不同加权特征,分为具有丰富信息的空间内容和信息较少内容。如图 4 所示,通过采用交叉重构运算,将加权后的两个不同的信息特征相结合,加强信息流。最后将交叉重构的特征进行拼接,得到空间细化特征映射 X_s 。

为了抑制通道维度特征冗余,采用分离变换和融合策略沿通道维度处理空间细化特征映射 X_s 。首先将 X_s 的通道分割为 ηC 通道和 $(1 - \eta)C$ 通道两部分,其中 $\eta \in [0, 1]$ 为分割比。利用 1×1 卷积压缩特征映射的通道以提高计算效率。在分割压缩操作之后,将 X_s 划分为 X_h 和 X_l ,分别进行转换操作。采用组卷积和逐点卷积运算来提取 X_h 中的高级代表

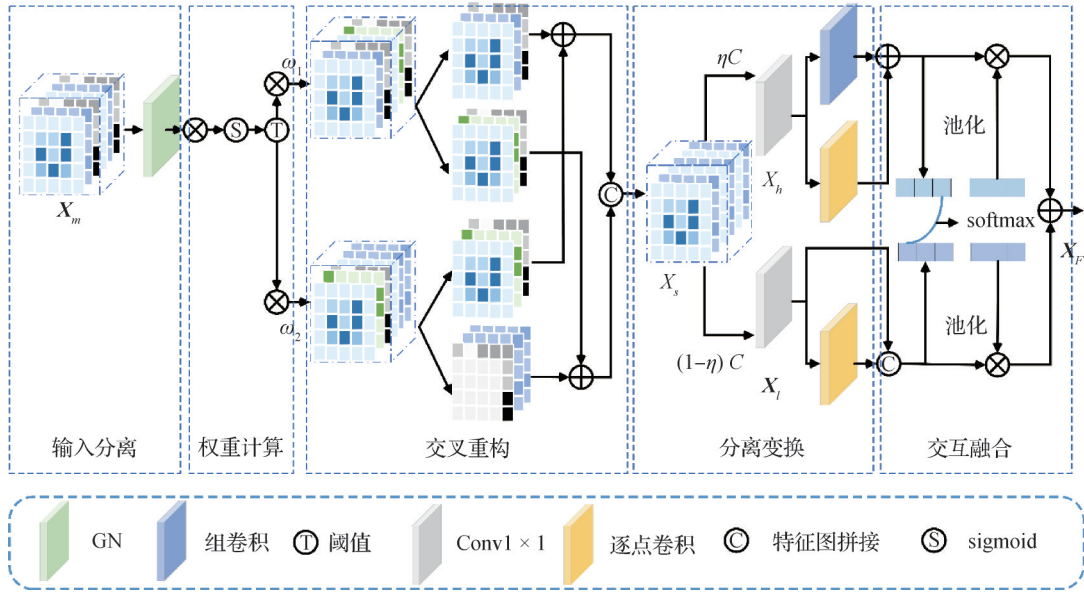


图4 空间通道信息重构模块

Fig. 4 Spatial-channel information reconstruction module

性信息,并降低计算成本。对 X_l 执行简单的 1×1 逐点卷积操作来生成具有浅层隐藏细节的特征映射,作为高级特征提取的补充。

在执行分离转换操作之后,采用全局平均池化,将上下特征叠加实现全局空间信息整合。使用softmax运算生成特征重要性向量后,在特征向量的指导下按通道方向合并特征,得到空间通道细化特征 X_F 。

1.5 损失函数

本文提出的分割网络采用鲁棒损失函数、二进制交叉熵损失和骰子(Dice)损失的组合损失函数进行训练(Sudre等,2017)。具体而言,解码器中每个阶段的输出使用二进制交叉熵损失和骰子损失的混合损失函数。其中,二进制交叉熵损失函数作为损失函数中的一个分量应用于像素级二进制分类问题,骰子损失旨在解决类不平衡问题。混合损失函数通过赋予不同阶段不同权重平衡最终预测性能,其计算为

$$l_i = L_{\text{BCE}}(g, y) + L_{\text{Dice}}(g, y) \quad (15)$$

$$L_{\text{BD}} = \sum_{i=0}^5 \lambda_i \times l_i \quad (16)$$

式中, L_{BCE} 表示二进制交叉熵损失函数, L_{Dice} 表示骰子损失, i 表示解码器输出阶段, g 和 y 分别表示像素真实的分类和网络输出的预测结果, λ_i 是不同阶段的权重。在本文中,将 λ_i 默认设置为1、0.5、0.4、

0.3、0.2、0.1。

为了有效降低数据中噪声和异常值存在的影响,通过鲁棒损失函数这一基于Student-t分布的负对数似然,通过单个参数动态调节损失函数对异常值的灵敏度,从而有效地处理数据中的异常值,具体为

$$L_r = \frac{1}{N} \sum_{j=1}^N -\log p(y_j | f_b(g_j), \Sigma; v) \quad (17)$$

式中, g_j 和 y_j 分别表示给定图像真实的分类和网络输出的预测掩码,具有参数 b 的函数 f_b 能够使 $f_b(x)$ 近似于给定图像的精确二值分割掩码。启发式的假设误差项遵循Student-t分布, y_j 的Student-t分布是由概率密度函数定义的。同时,由于均方误差(mean squared error, MSE)损失是通过最小化正态分布的负对数似然得到的,应用相同的转换为

$$-\log p(y_j | \mu, \Sigma; v) = -\log \Gamma\left(\frac{v+D}{2}\right) + \log \Gamma\left(\frac{v}{2}\right) + \frac{1}{2} \log |\Sigma| + \frac{D}{2} \log(\pi v) + \kappa \quad (18)$$

$$\kappa = \frac{v+D}{2} \log \left[1 + \frac{(y-u)^T \Sigma^{-1} (y-u)}{v} \right] \quad (19)$$

式中, $\Gamma(\cdot)$ 代表伽马函数; μ 和 Σ 分别为相关多元正态分布的均值和协方差矩阵, v 为自由度数, $|\cdot|$ 为行列式,近似认为 $u = f_b(g_i)$, κ 是中间变量。最终分割损失函数由鲁棒损失函数、二进制交叉熵损失和

Dice 损失的组合得到, 具体为

$$L_{\text{seg}} = \alpha \cdot L_{\text{BD}} + \beta \cdot L_T \quad (20)$$

式中, α 和 β 是两个可调整的加权系数, 用于平衡两个损失函数的影响。

2 实验

2.1 实验数据集

在实验中, 为了增强模型鲁棒性和泛化能力, 选取皮肤损伤分割的公共数据集和细胞核分割数据集作为该模型的训练和测试数据集。图 5 展示了实验数据集中的部分样本图像和对应标签图像。其中, 皮肤损伤分割公共数据集为 ISIC 发布的大规模的皮肤镜图像数据集 ISIC2017 和 ISIC2018。ISIC2017 和 ISIC2018 分别具有 2 150 幅和 2 694 幅数字皮肤病变图像数据和来自世界各地的专家注释的分割掩膜标签。由于 ISIC2017 和 ISIC2018 的分辨率并不统一, 实验中将图像缩放为 256×256 像素。数据集以 7:3 的比例随机划分作为实验的训练集和测试集。ISIC2017 训练集中有 1 500 幅图像, 测试集中有 650 幅图像。ISIC2018 训练集中有 1 886 幅图像, 测试集中有 808 幅图像。细胞核分割数据集使用来自 Data Science Bowl Challenge 的 DSB2018 数据集, 取自多种显微成像条件, 具有细胞类型、放大倍率和成像模式多样化的特点, 图像尺寸为 256×256 像素。数据集共有 670 幅图像, 以 8:2 的比例随机划分作为实验的训练集和测试集, 训练和测试图像分别为 536 幅和 134 幅。

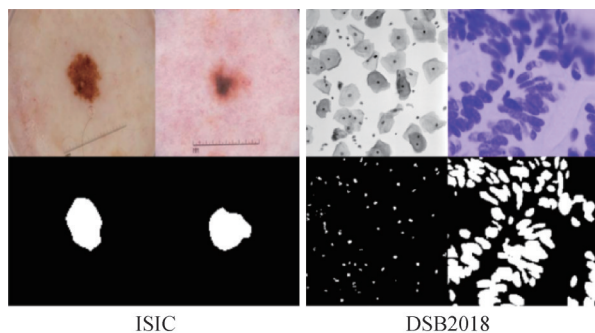


图 5 ISIC 和 DSB2018 数据集中的部分样本

Fig. 5 Some samples from the ISIC and DSB2018 datasets

2.2 实验设置

本文模型是在 PyTorch 框架下实现的。实验是在单个 NVIDIA RTX 4090 GPU 上进行的。在训练

过程中, 实验动态地应用水平翻转、垂直翻转和随机旋转等数据增强技术提高模型的泛化能力。实验中使用 AdamW 优化器对模型进行训练。初始学习率设置为 0.001, 最大迭代次数为 50, 最小学习率为 $1e-5$, 模型训练中 CosineAnnealingLR (Loshchilov 和 Hutter, 2017) 用做调度器。训练批次大小和总 epoch 分别设置为 8 和 300。

为了衡量分割性能, 本文采用 5 个常用的指标评估提出的模型和应用于数据集上的其他方法 (Müller 等, 2022), 包括混淆矩阵、Dice 系数 (Dice)、准确率 (accuracy, ACC)、敏感性 (sensitivity, SE) 和特异性 (specificity, SP)。

2.3 参数验证

在第 1.5 节中, 损失函数中涉及到可调参数 α 和 β 。在本小节中, 主要分析式 (20) 中的参数对所提出方法获得的分割性能的影响。参数的功能是平衡语义分割损失和鲁棒损失, 以确保两者都有足够的贡献。在实验中, 由于 $\alpha + \beta = 1$, 本文通过调节 β 控制两个损失函数的相对重要性。设置 β 为一组值, 包括 1, 0.5, 0.1 和 0.05, 参数调节后相应的分割性能结果显示在表 1 中。

从表 1 所示的实验结果可以看出, 在 ISIC2018 数据集中, 当 β 值增大时, Dice 和准确率有所下降。这是因为 β 值增大, 二进制交叉熵损失和 Dice 损失构成的语义分割损失比重下降, 一定程度影响模型的训练方向。由于 ISIC2018 数据集存在噪声伪影和设备标记标尺等干扰, 当 β 值为 0.1 时, 鲁棒损失函数发挥其最佳效果, 模型整体性能达到较好的平衡。在 DSB2018 数据集中, 当 β 值增大时, Dice、特异性和敏感性指标逐渐下降, 综合所有性能评价指标, 鲁棒损失函数的影响占比较小。基于上述分析, 本文实验默认将 β 值设置为 0.1, 以在不同度量的性能之间实现良好的权衡。

2.4 对比实验

为了验证所提出的方法在医学图像分割方面的优越性, 将提出的分割网络与在 ISIC2017、ISIC2018 和 DSB2018 基准上测试过的先进分割方法及经典分割方法进行比较。包括 U-Net、TransFuse、EGEUNet (efficient group enhanced U-Net)、MALUNet (multi-attention and light-weight U-Net)、TransUNet (Transformers make strong encoders for medical image seg-

表1 在ISIC2018和DSB2018基准上调节损失函数参数值后的分割性能

Table 1 Segmentation performance after adjusting loss function parameter values on ISIC2018 and DSB2018 benchmarks

β 值	ISIC2018/%				DSB2018/%			
	Dice	ACC	SP	SE	Dice	ACC	SP	SE
1	86.08	92.11	94.65	89.31	87.71	94.25	95.13	89.96
0.5	88.13	94.52	96.44	88.57	89.54	95.97	96.38	90.60
0.1	89.84	94.98	96.20	91.19	91.52	97.65	98.50	91.88
0.05	89.66	94.26	95.52	90.07	91.79	97.83	97.22	91.35

注:加粗字体表示各列最优结果。

mentation)、SCR-Net (spatial and channel-wise reinforcement network)、MedT (meta Transformer for medical image segmentation)、FAT-Net (feature adaptive Transformers for accurate and efficient medical image segmentation)、C²SDG (cross-scale cross-modality semantic decoupling guidance network)、LTMSegnet (light-weight multi-scale medical image segmentation combining Transformer and MLP)、VM-UNet、Rolling-Unet (revitalizing MLP’s ability to efficiently extract long-distance dependencies)、UNeXt (MLP-based rapid medical image segmentation network) 和 TransAttUNet (multi-level attention-guided U-Net with Transformer)。

鉴于很多现有的医学图像分割方法的源代码未公开,同时为了避免潜在训练模型偏差的引入,查阅相关文献,部分方法直接参考相关出版物显示的定量结果进行比较验证。针对部分对比模型训练,为减少偏差,设置相同通道数和参数以保证对比效果。

2.4.1 皮肤病变分割结果

表2和表3分别展示了不同方法在ISIC2017和ISIC2018基准数据集的定量评价结果。

图6(a)(b)展示了不同方法在皮肤病变分割上Dice、ACC、SP、SE指标的表现,在每种情况下,执行最佳的方法用星号标记。为了更直观地展现模型在区分前景(病灶区域)与背景方面的性能差异,在ISIC2017和ISIC2018数据集上分别计算了混淆矩阵。混淆矩阵能够清晰反映模型在像素级别上的分类准确性,从而揭示模型对目标区域和背景区域的识别与误判情况。

图7展示了混淆矩阵的可视化结果,进一步验

表2 分割方法在ISIC2017基准上的客观评价结果

Table 2 Objective evaluation results of different segmentation methods on the ISIC2017 benchmark

方法	/%			
	Dice	ACC	SP	SE
U-Net(Ronneberger等,2015)	81.59	91.64	96.80	81.72
TransFuse(Zhang等,2021)	88.40	96.17	97.98	87.14
SCR-Net(Wu等,2021)	88.98	95.88	98.53	84.97
FAT-Net(Wu等,2022)	85.00	93.26	97.25	83.92
MALUNet(Ruan等,2022)	88.13	96.18	98.47	84.78
EGEUNet(Ruan等,2023)	88.47	96.3	98.41	85.48
C ² SDG(Hu等,2023)	89.38	95.88	97.65	88.59
VM-UNet(Ruan等,2024)	89.03	96.29	97.58	89.90
Rolling-Unet(Liu等,2024)	87.42	95.88	97.98	85.45
本文	89.42	96.48	98.00	88.91

注:加粗字体表示各列最优结果。

表3 分割方法在ISIC2018基准上的客观评价结果

Table 3 Objective evaluation results of different segmentation methods on the ISIC2018 benchmark

方法	/%			
	Dice	ACC	SP	SE
U-Net(Ronneberger等,2015)	85.45	94.04	96.97	88.00
MedT(Valanarasu等,2021)	83.89	93.58	96.37	82.52
MALUNet(Ruan等,2022)	89.04	94.62	96.19	89.74
EGEUNet(Ruan等,2023)	88.73	94.52	96.44	88.57
C ² SDG(Hu等,2023)	88.06	95.06	96.43	89.70
VM-UNet(Ruan等,2024)	89.71	94.91	96.13	91.12
Rolling-Unet(Liu等,2024)	88.26	94.50	97.58	84.92
本文	89.84	94.98	96.20	91.19

注:加粗字体表示各列最优结果。

证了本文所提出方法在减少误分割和漏分割方面的优势,尤其在复杂边界与遮挡情况下仍保持了较高的像素区分能力。综合实验结果可知,与目前最先进的分割网络相比,所提出的方法在ISIC2017和ISIC2018数据集上表现出了显著的综合性能。具体而言,本文方法在ISIC2017和ISIC2018基准上的Dice得分为89.42%和89.84%。同时,本文方法在ISIC2018基准上的SE指标达到了91.19%,具有明显的优势。通过分析Dice和SE指标,相比于Trans-

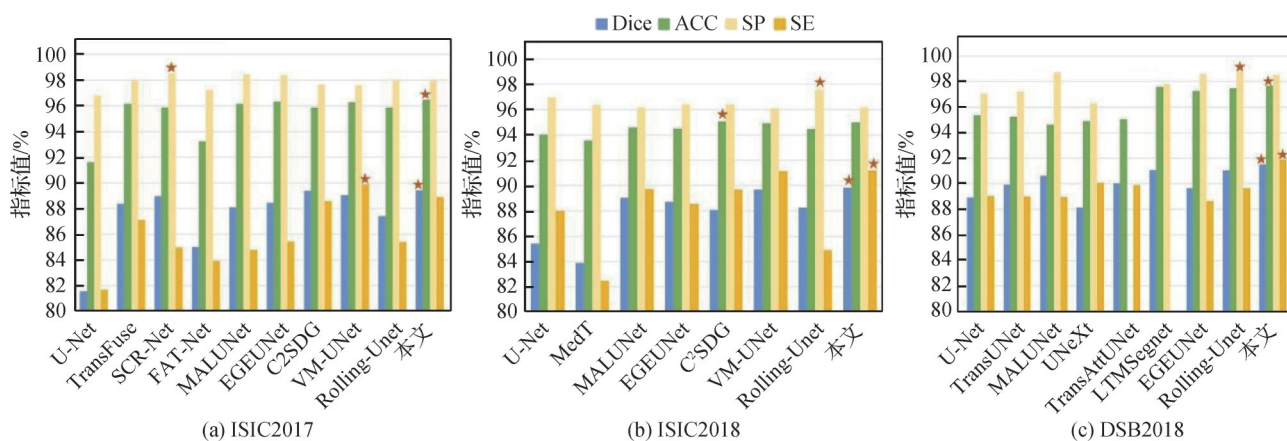


图6 不同分割方法在ISIC2017、ISIC2018和DSB2018数据集上Dice、ACC、SP、SE指标比较

Fig. 6 Performance comparison of different segmentation methods on the ISIC2017, ISIC2018, and DSB2018 datasets in terms of Dice, ACC, SP, and SE metrics((a)ISIC2017;(b)ISIC2018;(c)DSB2018)

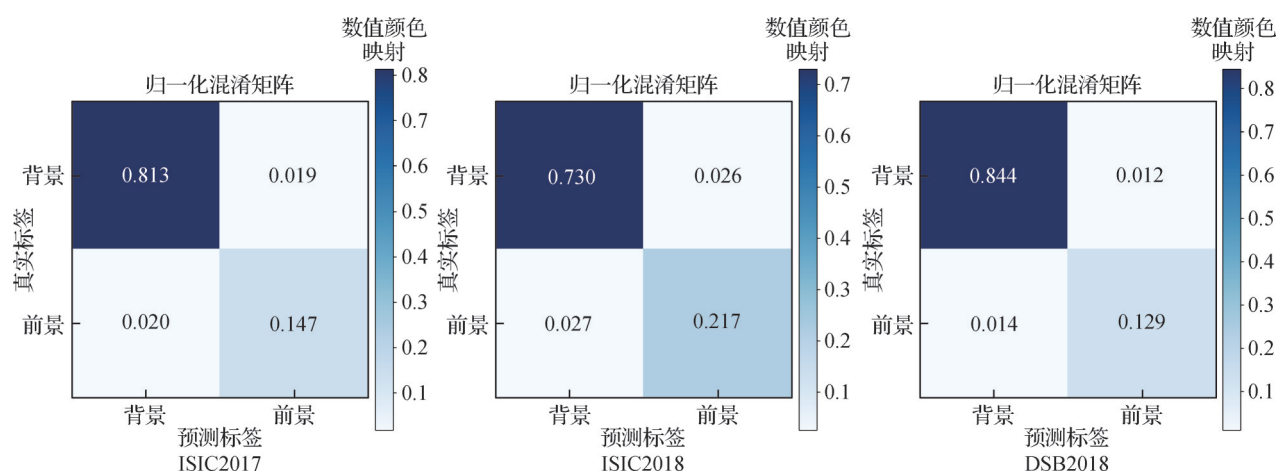


图7 ISIC2017、ISIC2018和DSB2018数据集上的混淆矩阵可视化

Fig. 7 Visualization of confusion matrix on the ISIC2017, ISIC2018, and DSB2018 datasets

Fuse 这一由 Transformer 和 CNN 组成的并行分支结构进行分割的方法,本文方法在整体分割和病变区域识别性能提升最为明显。与最新提出的应用不同方向核心多层感知机(multilayer perceptron, MLP)模块来融合局部特征和远程依赖的编解码网络 Rolling-Unet 相比,本文网络在整体分割质量和病变区域识别方面都取得了明显的改进,也产生了更具竞争力的结果。

由于 CNN 接受域范围局限以及 Transformer 聚焦于关键区域的特性,一般网络在识别病变区域与皮肤的过渡区域方面受到限制,从而一定程度影响分割性能。图8所示的具有挑战性的病例中病变出现不同的规模、不同的颜色分布以及不同的人体位置。通过不同方法得到的分割结果的视觉效果对比,U-Net通常不能准确识别边缘模糊和具有一定干

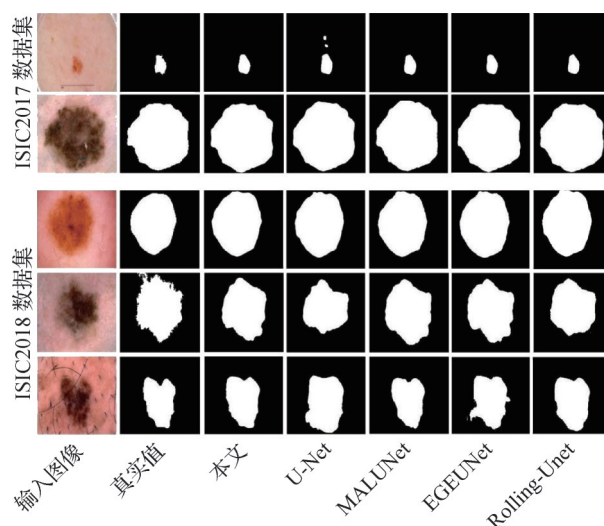


图8 不同方法获得的皮肤病分割结果的视觉效果比较

Fig. 8 Visual comparison of skin lesion segmentation results obtained by different methods

扰存在的复杂边界,同时出现正常区域判断错误的情况,部分网络所分割的整体区域以及病灶边界与 Ground Truth 存在较大差异。当皮肤病变图像中存在毛发血管自然伪影以及病变区域特点变化等干扰时(如图8第3列),相较于其他方法,本文方法实现了更高的皮肤病变分割精度。通过采用频率自适应编码模块和特征分解变换模块,整体区域分割和病变边界识别更准确,特别是对于边界不规则、尺度变化或颜色分布复杂的挑战性病例。

根据图9所示的高频、低频特征图和频谱图对比可以看出,UNet++整体表现较为模糊,边缘信息不清晰,难以突出关键的细节特征。TransFuse通过注意力机制在一定程度上增强了边缘响应,但在目标边界处仍存在信息缺失的问题。而本文方法在高频分量的提取上效果更好,能够清晰地刻画出病变区域边界。

在边缘模糊的情况下,也能保留丰富的纹理与结构细节,说明模型具备一定的低频增强能力。在低频特征图中,UNet++存在明显的过度平滑现象,病灶区域与背景区分不明显,TransFuse的低频特征较为稳定,但在局部细节保留上仍有欠缺。相比之下,本文方法不仅病变区域的整体轮廓和结构能实现进一步优化,还能有效提升分割的细节效果。本文方法的频谱图表现出较为集中的能量分布,低频区域轮廓清晰,同时高频环状区域依然保持有显著信号,说明模型在获取全局信息的同时,保留了边缘频率特征。相比之下,UNet++的频谱能量较为分散,中心频率衰减不明显,特征提取不够精准。TransFuse在频谱上有所收缩,但高频段能量偏弱。这一对比结果进一步验证了频率自适应机制对高低频分量的有效建模能力。

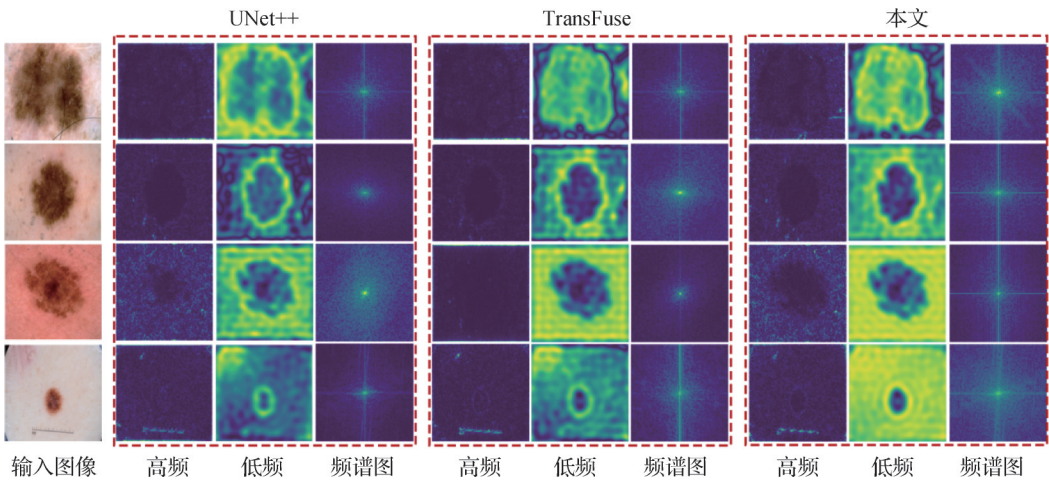


图9 不同方法的高频、低频特征图和频谱图对比

Fig. 9 Comparison of high-frequency, low-frequency feature maps, and spectrograms across different methods

2.4.2 细胞核分割结果

表4展示了不同对比方法在 DSB2018 数据集进行分割的定量评价结果。

图6(c)展示了不同方法在细胞核分割上 Dice、ACC、SP、SE 指标的表现。本文方法在 DSB2018 数据集的 Dice 得分为 91.52%, 优于其他对比方法 0.44%~2.59%, 在细胞核分割方面的性能得到了提升。同时,本文方法在 DSB2018 基准上的 SE 指标达到了 91.88%, 具有明显的优势。通过图10的可视化结果,相比于 U-Net 等模型仅在总体上划分出了细胞核,本文模型能够精确分割每个细胞核区域及边缘识别。在图像中存在大量小面积核且细胞核形态大小差异显著导致分割困难的条件下,

本文方法能够实现细胞核独立分割,抑制核间相连影响。如图10第4行绿色虚线框标注,其他方法在分割后均产生了噪声影响分割效果,而本文方法能够排除背景颜色干扰,实现更高的细胞核分割精度。

2.5 消融实验

为了验证并评估提出的模型以及每个模块的有效性,本文进行了相应的消融研究,分析了不同模块的分割效果,包括频率自适应编码模块、特征分解变换模块和空间通道信息重构模块。实验将最终模型与几种组合不同模块的模型进行比较,分别从定量和定性的角度验证不同模块的效果。为了充分验证模型及设计模块的分割效果,本文在 ISIC2017 和

表 4 分割方法在 DSB2018 基准上的客观评价结果
Table 4 Objective evaluation results of different segmentation methods on the DSB2018 benchmark

方法	Dice	ACC	SP	SE
U-Net(Ronneberger等,2015)	88.93	95.36	97.06	89.02
TransUNet(Chen等,2021)	89.92	95.23	97.23	89.01
MALUNet(Ruan等,2022)	90.63	94.62	98.72	88.97
UNeXt(Valanarasu和Patel,2022)	88.13	94.9	96.31	90.07
TransAttUNet(Chen等,2024a)	90.04	95.06	—	89.9
LTMSegnet(Huang等,2024)	91.08	97.57	97.82	—
EGEUNet(Ruan等,2023)	89.67	97.25	98.59	88.66
Rolling-Unet(Liu等,2024)	91.05	97.47	98.78	89.64
本文	91.52	97.65	98.50	91.88

注:加粗字体表示各列最优结果。“—”表示未提供相关数据。

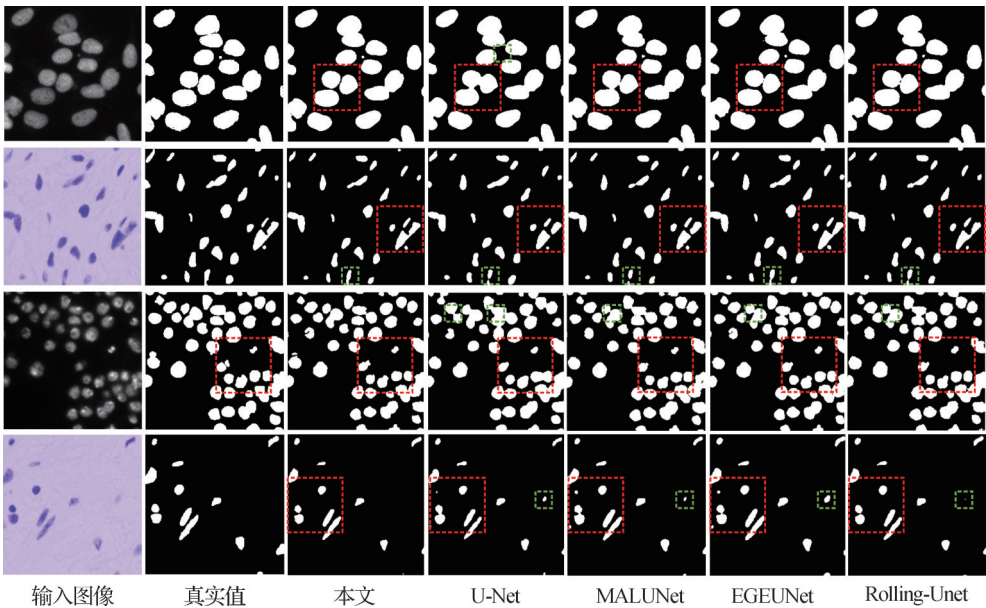


图 10 不同方法获得的细胞核分割结果的视觉效果比较

Fig. 10 Visual comparison of nuclear segmentation results obtained by different methods

ISIC2018数据集上进行了消融实验。其中,定量结果如表 5 所示,Dice、ACC、SP 和 SE 评价指标数值的变动一定程度上从整体分割质量和病变区域识别角度直观表明了模块的性能提升效果。FLOPs(floating point operations per second)表示每秒浮点运算次数。定性分析的结果如图 11 所示,能够进一步可视化各个模块和最终模型的分割效果。

本实验中使用基于六阶段 U 形架构的编码器—解码器网络作为基线模型,然后逐步添加不同的模块组合来验证其效果。具体而言,实验比较了以下模型的效果:

- 1)基线模型。在综合考虑性能和模型大小的前提下,利用通道数为{8, 16, 24, 32, 48, 64}的六阶 U-Net 作为基线模型。每个阶段包含一个简单的卷积操作,内核大小为 3。
- 2)基线模型+FAE。在基线模型的基础上将使用 FAE 模块作为前 3 个阶段的编码器,从而实现高低频特征动态平衡和有效特征的提取,进一步增强皮肤病分割的背景结构信息和细节特征的学习。
- 3)基线模型 + FAE + FDT。包括应用自适应核策略的 FAE 模块以及联合最大池化和离散小波变换操作的特征分解变换模块 FDT 作为下采样模块减

表 5 在 ISIC2017 和 ISIC2018 基准上消融研究的客观评价结果
Table 5 Objective evaluation results of the ISIC2017 and ISIC2018 baseline ablation study

方法	参数量/K	GFLOPs	ISIC2017				ISIC2018			
			Dice/%	ACC/%	SP/%	SE/%	Dice/%	ACC/%	SP/%	SE/%
基础网络	24.597	2.260	88.47	96.30	98.41	85.48	88.95	94.55	95.99	90.06
基础网络+FAE	27.708	2.495	88.81	96.26	97.80	88.63	89.13	94.69	96.39	89.40
基础网络+FDT	26.459	2.403	88.69	96.35	97.92	88.21	89.36	94.93	96.44	89.95
基础网络+FAE+FDT	30.242	2.539	89.25	96.46	98.22	87.13	89.67	94.91	96.26	90.68
基础网络+FAE+FDT+SCIR	39.278	2.701	89.42	96.48	98.00	88.91	89.84	94.98	96.20	91.19

注:加粗字体表示各列最优结果。

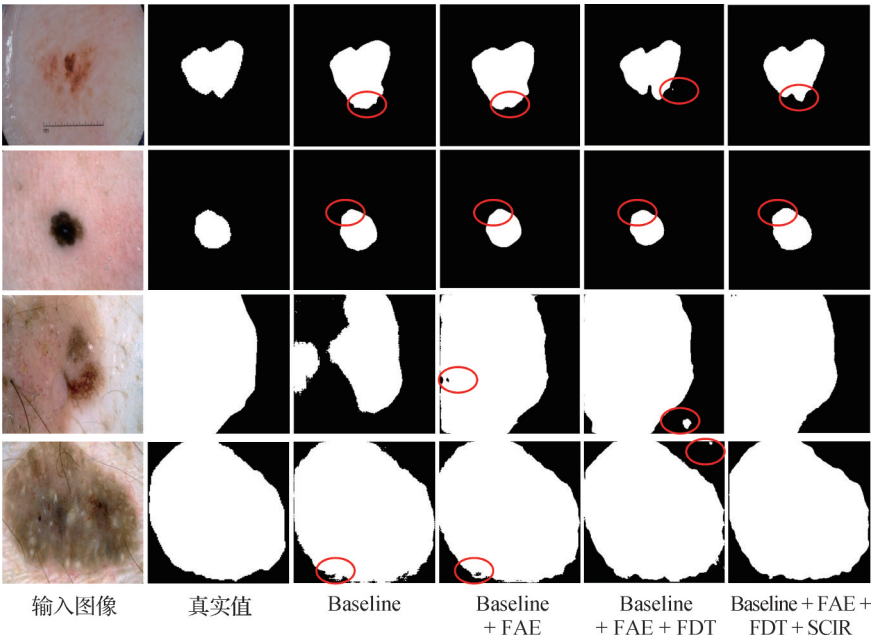


图 11 消融研究中不同模型分割结果的视觉效果比较

Fig. 11 Visual comparison of the segmentation results of different models in the ablation study

少局部和全局信息丢失。

4)基线模型 + FAE + FDT + SCIR。提出的最终模型包括 FAE 模块构成的编码器、FDT 构成的下采样模块以及 SCIR 构成的解码器实现空间通道信息的有效传播感知。

实验通过对基线模型与基线模型添加不同模块的比较评估添加模块的效果。如表 5 第 2 行所示,与基本模型相比,将 FAE 代替普通卷积作为编码器后,在 ISIC2017 和 ISIC2018 数据集上 Dice 指标均有提高。这一定程度克服了类别不平衡并提升了分割性能。表中第 3、4 行的定性结果对比体现了使用改进下采样模块 FDT 在细节信息捕获方面相较于池化操作的优势。同时结合定性结果即图 11 中第 5 列,相较于第 3 列基线模型及第 4 列,在边缘模糊且存在

干扰的条件下,对于病变区域的整体形状以及边缘分割更加清晰。添加了 FAE、FDT 和 SCIR 的最终模型相较于基线模型,分割效果显著增强。当目标皮肤病变区域内颜色变化差异较大时,有效抑制了空间通道特征冗余,减少了标记、毛发和皮肤色素沉积所产生的噪声干扰。最终模型不仅可以克服类内差异,还可以处理皮肤病变形状和边界不规则性,使得整体区域分割和病灶边界识别更准确。

为进一步验证所提出模块的有效性,在空域与频域两个方面对 FAE 和 FDT 模块的输出特征进行了频域和特征图可视化分析,如图 12 所示。图中展示了输入图像、基线及分别引入 FAE、FDT 以及两个模块组合作用下的特征图和频谱图。从特征图可以观察到,基线模型提取的特征边界模糊,区域响应不

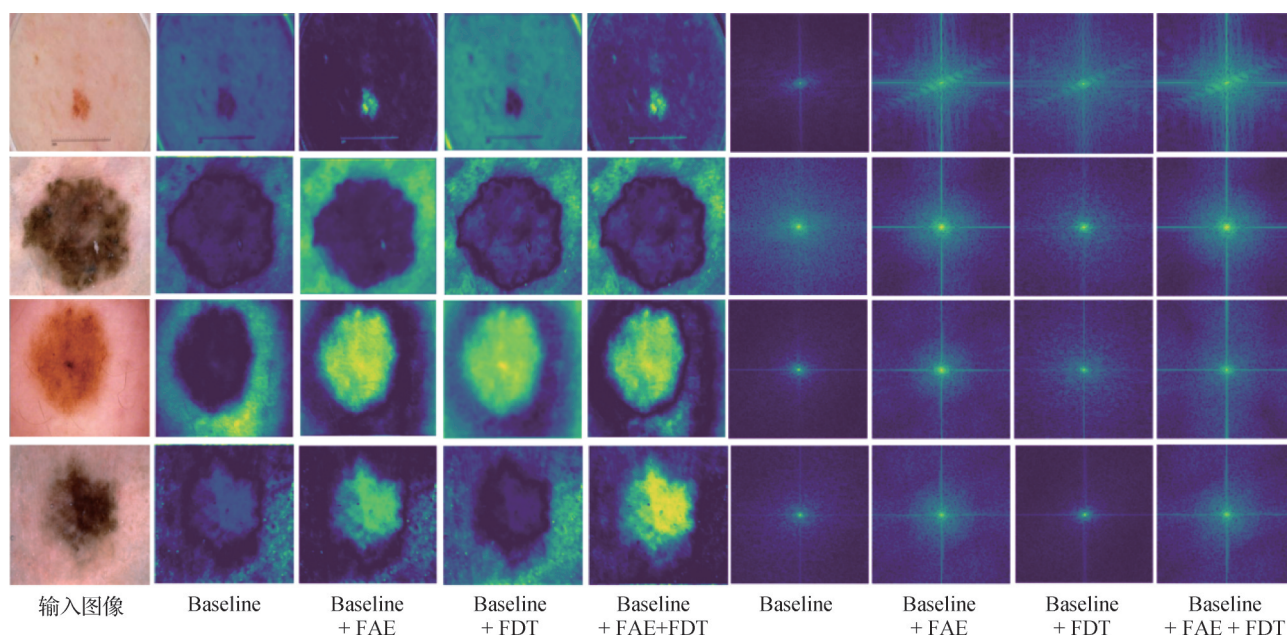


图12 消融研究中FAE和FDT分割结果的特征图和频域可视化

Fig. 12 Feature maps and frequency domains visualizations of segmentation results in ablation studies for FAE and FDT

够集中,难以准确描绘目标结构。在添加FAE后,特征图中的边缘信息显著增强,目标边界更加清晰,说明该模块有效提升了模型的边缘感知能力。应用FDT作为下采样操作后,特征图目标区域的边界和细节结构更为清晰。联合FAE和FDT后,特征图在边缘与区域信息感知方面均表现出显著提升,呈现出完整、清晰的目标结构,表明两模块在空间和频率信息建模上具有良好的互补性。

同时,频域特征图也进一步体现了模块在频率信息和特征提取方面的提升。基线模型的频谱图主要集中于低频区域,缺乏高频细节,这表示其提取的特征较为粗糙。添加FAE后,频谱中出现更多高频成分,说明网络进一步关注图像中的边缘与纹理特征。FDT相较于基线模型中的池化操作,其频谱结构更清晰。而在FAE与FDT联合作用下,频谱图呈现低频与高频共同增强的趋势,能量分布集中且层次分明,说明网络学习到了更丰富、全面的频率特征,进一步增强了模型对边缘结构与语义区域的感知能力,从而有效提升了分割性能。

3 结论

本文针对图像高频和低频特征利用不充分以及下采样操作容易产生信息丢失问题,提出一种基于编码器—解码器架构的频率自适应分割网络,通过

频率自适应编码模块动态平衡高低频分量,实现更精准的分割。同时将特征分解变换模块应用于下采样过程,通过并行最大池化和离散小波变换平衡操作处理编码特征,实现细节感知信息的恢复和保持。在解码过程中引入空间通道信息重构模块,采用分离重构策略及分离变换融合操作实现信息特征交互并有效抑制空间通道维度特征冗余。实验结果表明了所设计的关键部件的有效性。此外,在ISIC2017、ISIC2018和DSB2018这3个数据集中,将模型与先进的医学图像分割算法进行了比较,从实验结果可知,本文模型在主观视觉和客观评价指标方面都有优秀的表现,取得了更好的性能。

在未来的工作中,将聚焦于通过迁移学习和多任务学习提升模型在跨数据集和跨领域上的适应性,使其不仅在特定数据集上表现出色,还能在分布不同的其他数据集上获得更好的效果。此外,将通过多任务学习结合其他任务,包括图像分类和特征检测等,进一步提升网络的性能和应用价值。

参考文献(References)

- Cao H, Wang Y Y, Chen J, Jiang D S, Zhang X P, Tian Q, et al. 2023. Swin-unet: unet-like pure transformer for medical image segmentation//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 205-218 [DOI: 10.1007/

- 978-3-031-25066-8_9]
- Chen B Z, Liu Y S, Zhang Z, Lu G M and Kong A W K. 2024a. TransAttUnet: multi-level attention-guided U-Net with transformer for medical image segmentation. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(1): 55-68 [DOI: 10.1109/TETCI.2023.3309626]
- Chen J N, Lu Y Y, Yu Q H, Luo X D, Adeli E, Wang Y, et al. 2021. TransUNet: transformers make strong encoders for medical image segmentation [EB/OL]. [2025-03-20]. <https://arxiv.org/pdf/2102.04306.pdf>
- Chen L H, Bampis C G, Li Z, Chen C and Bovik A C. 2022. Convolutional block design for learned fractional downsampling//Proceedings of the 56th Asilomar Conference on Signals, Systems, and Computers. Pacific Grove, USA: IEEE: 640-644 [DOI: 10.1109/IEEECONF56349.2022.10052104]
- Conze P H, Andrade-Miranda G, Singh V K, Jaouen V and Visvikis D. 2023. Current and emerging trends in medical image segmentation with deep learning. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 7(6): 545-569 [DOI: 10.1109/TRPMS.2023.3265863]
- Gu A and Dao T. 2023. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. [2025-03-20]. <https://arxiv.org/pdf/2312.00752.pdf>
- Hu S S, Liao Z H and Xia Y. 2023. Devil is in channels: contrastive single domain generalization for medical image segmentation//Proceedings of the 26th International Conference on Medical Image Computing and Computer-Assisted Intervention. Vancouver, Canada: Springer: 14-23 [DOI: 10.1007/978-3-031-43901-8_2]
- Huang X, Tang H X, Ding Y, Li Y Y, Zhu Z Q and Yang P. 2024. LTMSegnet: lightweight multi-scale medical image segmentation combining transformer and MLP. *Computers in Biology and Medicine*, 183: #109259 [DOI: 10.1016/j.combiomed.2024.109259]
- Li G J, Huang Q H, Wang W and Liu L Z. 2025. Selective and multi-scale fusion Mamba for medical image segmentation. *Expert Systems with Applications*, 261: #125518 [DOI: 10.1016/j.eswa.2024.125518]
- Li H, Liu J J and Xiao L. 2024. Dual-encoder global-local cross-attention network for medical image segmentation. *Journal of Image and Graphics*, 29(11): 3462-3475 (李赫, 刘建军, 肖亮). 2024. 融合交叉注意力与双编码器的医学图像分割. *中国图象图形学报*, 29(11): 3462-3475 [DOI: 10.11834/jig.230705]
- Liu Y T, Zhu H J, Liu M T, Yu H Y, Chen Z H and Gao J. 2024. Rolling-Unet: revitalizing MLP's ability to efficiently extract long-distance dependencies for medical image segmentation//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 3819-3827 [DOI: 10.1609/aaai.v38i4.28173]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Loshchilov I and Hutter F. 2017. SGDR: stochastic gradient descent with warm restarts//Proceedings of the 5th International Conference on Learning Representations. Toulon, France: ICLR
- Müller D, Soto-Rey I and Kramer F. 2022. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1): #210 [DOI: 10.1186/s13104-022-06096-y]
- Park N and Kim S. 2022. How do vision transformers work?//Proceedings of the 10th International Conference on Learning Representations, [s.l.]: ICLR
- Ronneberger O, Fischer P and Brox T. 2015. U-net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Roy S, Koehler G, Ulrich C, Baumgartner M, Petersen J, Isensee F, et al. 2023. MedNeXt: transformer-driven scaling of ConvNets for medical image segmentation//Proceedings of the 26th International Conference on Medical Image Computing and Computer-Assisted Intervention. Vancouver, Canada: Springer: 405-415 [DOI: 10.1007/978-3-031-43901-8_39]
- Ruan J C, Li J C and Xiang S C. 2024. VM-UNet: vision Mamba UNet for medical image segmentation [EB/OL]. [2025-02-17]. <https://arxiv.org/pdf/2402.02491v2.pdf>
- Ruan J C, Xiang S C, Xie M Y, Liu T and Fu Y Z. 2022. MALUNet: a multi-attention and light-weight UNet for skin lesion segmentation//Proceedings of 2022 IEEE International Conference on Bioinformatics and Biomedicine. Las Vegas, USA: IEEE: 1150-1156 [DOI: 10.1109/BIBM55620.2022.9995040]
- Ruan J C, Xie M Y, Gao J S, Liu T and Fu Y Z. 2023. EGE-UNet: an efficient group enhanced UNet for skin lesion segmentation//Proceedings of the 26th International Conference on Medical Image Computing and Computer-Assisted Intervention. Vancouver, Canada: Springer: 481-490 [DOI: 10.1007/978-3-031-43901-8_46]
- Sarker M M K, Rashwan H A, Akram F, Singh V K, Banu S F, Chowdhury F U H, et al. 2021. SLSNet: skin lesion segmentation using a lightweight generative adversarial network. *Expert Systems with Applications*, 183: #115433 [DOI: 10.1016/j.eswa.2021.115433]
- Sudre C H, Li W Q, Vercauteren T, Ourselin S and Jorge Cardoso M. 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations//Proceedings of the 3rd International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Québec City, Canada: Springer: 240-248 [DOI: 10.1007/978-3-319-67558-9_28]
- Valanarasu J M J, Oza P, Hacıhaliloglu I and Patel V M. 2021. Medical transformer: gated axial-attention for medical image segmentation//Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention. Strasbourg,

- France: Springer: 36-46 [DOI: 10.1007/978-3-030-87193-2_4]
- Valanarasu J M J and Patel V M. 2022. UNeXt: MLP-based rapid medical image segmentation network//Proceedings of the 25th International Conference on Medical Image Computing and Computer-Assisted Intervention. Singapore, Singapore: Springer: 23-33 [DOI: 10.1007/978-3-031-16443-9_3]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A Z, et al. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6000-6010
- Wang J C, Wei L, Wang L S, Zhou Q C, Zhu L and Qin J. 2021. Boundary-aware transformers for skin lesion segmentation//Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention. Strasbourg, France: Springer: 206-216 [DOI: 10.1007/978-3-030-87193-2_20]
- Wang Z Y, Zheng J Q, Zhang Y C, Cui G and Li L. 2024. Mamba-UNet: unet-like pure visual Mamba for medical image segmentation [EB/OL]. [2025-03-20]. <https://arxiv.org/pdf/2402.05079.pdf>
- Wu H, Chen S, Chen G, Wang W, Lei B and Wen Z. 2022. FAT-Net: feature adaptive transformers for automated skin lesion segmentation. *Medical Image Analysis*, 76: #102327 [DOI: 10.1016/j.media.2021.102327]
- Wu H S, Zhong J F, Wang W, Wen Z K and Qin J. 2021. Precise yet efficient semantic calibration and refinement in ConvNets for real-time polyp segmentation from colonoscopy videos//Proceedings of the 35th AAAI Conference on Artificial Intelligence. [s.l.]: AAAI: 2916-2924 [DOI: 10.1609/aaai.v35i4.16398]
- Wu R K, Liu Y H, Liang P C and Chang Q. 2025. H-vmunet: high-order vision Mamba UNet for medical image segmentation. *Neurocomputing*, 624: #129447 [DOI: 10.1016/j.neucom.2025.129447]
- Xu G P, Liao W T, Zhang X, Li C, He X W and Wu X L. 2023. Haar wavelet downsampling: a simple but effective downsampling module for semantic segmentation. *Pattern Recognition*, 143: #109819 [DOI: 10.1016/j.patcog.2023.109819]
- Xu R T, Wang C W, Xu S B, Meng W L and Zhang X P. 2021. DC-Net: dual context network for 2D medical image segmentation//Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention. Strasbourg, France: Springer: 503-513 [DOI: 10.1007/978-3-030-87193-2_48]
- Xu W W, Xu L F, Li B K, Zhou X, Lyu N and Zhan S. 2024. TransAS-UNet: regional segmentation of breast cancer Swin Transformer and of UNet algorithm. *Journal of Image and Graphics*, 29(3): 741-754 (徐旺旺, 许良凤, 李博凯, 周曦, 律娜, 詹曙. 2024. TransAS-UNet: 融合 Swin Transformer 和 UNet 的乳腺癌区域分割. *中国图象图形学报*, 29(3): 741-754) [DOI: 10.11834/jig.230130]
- Yao W J, Bai J J, Liao W, Chen Y H, Liu M J and Xie Y. 2024. From CNN to transformer: a review of medical image segmentation models. *Journal of Imaging Informatics in Medicine*, 37(4): 1529-1547 [DOI: 10.1007/s10278-024-00981-7]
- Zhang Y D, Liu H Y and Hu Q. 2021. TransFuse: fusing transformers and CNNs for medical image segmentation//Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention. Strasbourg, France: Springer: 14-24 [DOI: 10.1007/978-3-030-87193-2_2]
- Zhou T, Dong Y L, Huo B Q, Liu S and Ma Z J. 2021. U-Net and its applications in medical image segmentation: a review. *Journal of Image and Graphics*, 26(9): 2058-2077 (周涛, 董雅丽, 霍兵强, 刘珊, 马宗军. 2021. U-Net 网络医学图像分割应用综述. *中国图象图形学报*, 26(9): 2058-2077) [DOI: 10.11834/jig.200704]
- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-net architecture for medical image segmentation//Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada, Spain: Springer: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]
- Zhu X Z, Hu H, Lin S and Dai J F. 2019. Deformable ConvNets V2: more deformable, better results//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 9300-9308 [DOI: 10.1109/CVPR.2019.00953]
- Zhu Z Q, He X Y, Qi G Q, Li Y Y, Cong B S and Liu Y. 2023. Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI. *Information Fusion*, 91: 376-387 [DOI: 10.1016/j.inffus.2022.10.022]
- Zhu Z Q, Wang Z D, Qi G Q, Zhao Y X and Liu Y. 2025. Visually stabilized Mamba U-shaped network with strong inductive bias for 3-D brain tumor segmentation. *IEEE Transactions on Instrumentation and Measurement*, 74: #2518511 [DOI: 10.1109/TIM.2025.3551581]

作者简介

朱智勤,男,教授,主要研究方向为图像处理、机器学习、智能控制和区块链技术。E-mail: zhuzq@cqupt.edu.cn

李娜源,通信作者,女,正高级工程师,主要研究方向为图像处理和目标识别。E-mail: liyy@cqupt.edu.cn

孙梦薇,女,硕士研究生,主要研究方向为医学图像分割。E-mail: s220333016@stu.cqupt.edu.cn

齐观秋,男,教授,主要研究方向为图像处理和机器学习。E-mail: qig@buffalostate.edu

杨梦杰,男,高级工程师,主要研究方向为医学影像处理和医学影像系统。E-mail: yangmj68@163.com

程俊,男,研究员。主要研究方向为计算机视觉、机器人和人机交互。E-mail: Jun.cheng@siat.ac.cn

刘羽,男,教授,主要研究方向为多源图像融合、图像复原、物体识别和医学图像分析。E-mail: yuliu@hfut.edu.cn