

MCNet:融合多层感知机和卷积的轻量级病变区域分割网络

中华磊^{a,b,c},上官国庆^a,袁成雨^a,陈艳浩^{a,b,c},刘栋^{a,b,c}

(河南师范大学 a.计算机与信息工程学院;b.河南省教育人工智能与个性化学习重点实验室;

c.教学资源与教育质量评估大数据河南省工程实验室,河南 新乡 453007)

摘要:针对现有医学图像分割网络存在计算量大、对硬件资源要求高和推理速度慢等不足,提出一种轻量级快速分割网络 MCNet. MCNet 采用编码器-解码器架构,使用多层感知机(MLP)和卷积分别提取并融合医学图像的全局特征和局部特征,以减少网络参数量并提高分割精度.在编码阶段使用卷积分支和多层感知机分支分别提取多尺度的局部特征和全局特征.通过跳跃连接融合这些特征并送入解码器.在解码阶段使用注意力门控机制进行特征增强.在 BUSI 和 ISIC2018 数据集上进行实验.和当前最优方法相比,MCNet 的 Dice 相似系数和均交并比在 BUSI 数据集上分别提高 0.11% 和 0.09%,在 ISIC2018 数据集上分别提高 0.64% 和 0.95%.同时,MCNet 显著减少了网络参数量、降低了浮点运算次数并缩短了 CPU 推理时间.

关键词:医学图像分割;深度神经网络;多层感知机(MLP);轻量级网络

中图分类号:TP391.41

文献标志码:A

文章编号:1000-2367(2025)03-0096-08

深度学习近年来在医学图像处理和分析领域取得显著进展,特别是在医学图像分割领域.卷积神经网络(convolutional neural network, CNN)^[1]具有出色的局部信息提取能力,但对全局特征的捕获能力不强.基于注意力机制的网络能够有效地建模全局上下文,但对局部信息建模能力弱,并且计算成本较高.最近,结合卷积运算的多层感知机(multi-layer perceptron, MLP)使用线性移位操作和全连接层提取全局上下文信息,取得了与注意力机制相当的效果,且显著减少了参数量和计算复杂度.

本文提出一种用于病变区域分割的轻量级快速深度神经网络 MCNet. MCNet 以 U-Net 为基本架构,在编码阶段使用 MLP 分支和 Convolution 分支分别提取医学图像的全局和局部特征,同时使用融合模块将二者融合.和现有方法相比,MCNet 显著降低了网络的参数量和计算复杂度并提高了分割精度.本文主要贡献如下:(1)MCNet 对经典的编码器-解码器网络架构进行改进,在编码阶段引入 MLP 分支和 Convolution 分支,分别提取医学图像的全局特征和局部特征;(2)在编码阶段使用 BiFusion 融合模块充分融合 MLP 提取的全局特征和 Convolution 提取的局部特征,以增强 MCNet 的特征表达能力;(3)在 BUSI 和 ISIC2018 数据集上进行了充分实验.和现有方法相比,MCNet 取得了最好的分割结果、显著减少了参数量并缩短了 CPU 推理时间.

收稿日期:2024-01-18;**修回日期:**2024-02-25.

基金项目:国家自然科学基金(62072160);河南省科技攻关项目(232102211024).

作者简介:中华磊(1981—),男,河南西平人,河南师范大学副教授,博士,研究方向为深度学习、计算机视觉, E-mail: shenhualei@htu.edu.cn.

通信作者:陈艳浩(1980—),男,河南舞钢人,河南师范大学高级工程师,研究方向为人工智能、智慧教育、医学信息系统, E-mail: cyh@htu.edu.cn.

引用本文:中华磊,上官国庆,袁成雨,等. MCNet:融合多层感知机和卷积的轻量级病变区域分割网络[J]. 河南师范大学学报(自然科学版), 2025, 53(3): 96-103. (Shen Hualei, Shangguan Guoqing, Yuan Chengyu, et al. MCNet: a lightweight lesion segmentation network integrating multilayer perceptrons and convolutions[J]. Journal of Henan Normal University(Natural Science Edition), 2025, 53(3): 96-103. DOI: 10.16366/j.cnki.1000-2367.2024.01.18.0001.)

1 相关工作

不同于自然图像分割,医学图像分割具有以下特点:(1)分割难度更大,受成像设备、患者个体和成像条件之间差异的影响,不同患者的组织或器官成像时呈现出较强的类内差异和类间相似性,这导致医学图像分割任务的难度更大。(2)精度要求更高,虽然人体组织或器官的类别有限,但对它们的准确分割至关重要。因为上述原因,医学图像分割与自然图像分割常采用不同的网络架构。目前,大多数基于深度神经网络的医学图像分割模型源于经典 U-Net 架构:通过编码器提取图像的高级语义特征,再由解码器融合浅层与深层特征生成分割结果。近年来,这类模型往往融入了各种卷积或注意力模块,旨在提升分割精度和模型鲁棒性。

基于 CNN 的方法,RONNEBERGER 等^[2]开创性地提出用于医学图像分割的 U-Net,该网络已经成为医学图像分割的基础网络架构。XIAO 等^[3]提出的 ResUNet 给 U-Net 的跳跃连接加入残差以避免梯度消失并直接进行特征融合。ZHOU 等^[4]设计的 U-Net++ 使用稠密块替代标准卷积以更高效地学习多尺度特征。基于 CNN 的方法使用卷积提取医学图像的局部信息,显著提高了医学图像的分割精度,但对全局信息的提取能力不足。

基于 Transformer 的方法,DOSOVITSKIY 等^[5]提出基于 Transformer 的视觉模型 ViT,ViT 将输入图像等分为多个图像块,对每个图像块进行线性嵌入和位置编码。LIU 等^[6]提出一种新的视觉模型 Swin Transformer,该模型引入窗口机制将图像划分为非重叠的窗口,并且只在窗口内使用自注意力机制。VALANARASU 等^[7]采用局部全局训练方法融合图像块和图像整体的关联信息,TransFuse 构建了包含 CNN 和 Transformer 的双分支网络结构^[8]。基于 Transformer 的方法通过表征图像蕴含的全局信息提高图像分割精度。但是,这类方法的计算复杂度高、计算开销大。

基于 MLP 的方法,MLP 使用线性移位操作和全连接层提取特征之间的远程依赖,这显著减少了网络参数并提高了推理速度。TOLSTIKHIN 等^[9]提出一种完全基于 MLP 的架构 MLP-Mixer,该架构借助通道混合和 Token 混合机制实现自注意力机制。UNeXt^[10]是第一个基于 MLP 和 CNN 的轻量级医学图像分割混合网络。MSS-UNet^[11]使用双空间移位 MLP 模块实现不同空间位置之间的通信。与基于 Transformer 的方法相比,基于 MLP 的网络显著减少了参数量并提高了推理速度,但对局部信息的提取能力不足。

受现有工作启发,本文设计了一个新的医学图像病变区域分割网络 MCNet。不同于现有方法,MCNet 在编码阶段包含 Convolution 分支、MLP 分支和特征融合模块,有效提取并融合医学图像的局部特征与全局特征,以此增强编码器的特征表达能力。特别地,MLP 分支使用线性移位操作实现同注意力机制相当的特征提取效果并显著减少了网络的参数量和计算复杂度。

2 MCNet

2.1 网络结构

MCNet 的网络结构如图 1 所示。该网络基于经典的编码器-解码器架构,包括上下 5 层。在编码阶段,每一层都将特征图的分辨率减少至 $1/2$ 、特征通道数增加两倍;在解码阶段,每一层将特征图的分辨率增加两倍、特征通道数减少至 $1/2$ 。在编码器阶段,MLP 分支和 Convolution 分支分别生成独立的特征图。这两种特征图经过 BiFusion 模块进行融合,得到一个充分融合全局信息和局部信息的特征图。MCNet 的跳跃连接引入了 AG^[12]模块。AG 模块可以突出显著特征并以较少的计算开销提高模型的分割精度。在解码阶段,AG 模块生成的特征图与对应解码层输出的特征图沿通道拼接;之后进行两个核为 3×3 的卷积运算和一个核为 1×1 的卷积运算。此外每个解码层还引入 Dropout 以防止过拟合。解码器的最后一层使用双线性插值将特征图的分辨率恢复至输入图像的原始分辨率大小;然后经过核为 1×1 的卷积运算将特征图的通道数降至 1。最后使用 sigmoid 函数将特征图的每个像素值映射到 $[0, 1]$,从而得到分割结果。

2.2 MLP 分支

MLP 分支包含 5 个 MLP 模块,每个 MLP 模块的输出被用作下一层 MLP 模块的输入。MLP 模块包含 PatchEmbed 和 ShiftBlock 两个子模块。PatchEmbed 包含卷积层、reshape 层和 LayerNorm 层。其中卷积层

的核大小为 3×3 、步长为 2、padding 为 1。PatchEmbed 通过卷积提取特征,同时使图像的分辨率降低至 $1/2$ 。ShiftBlock 包含 LayerNorm 层、ShiftedMLP 层和 DropPath 层。

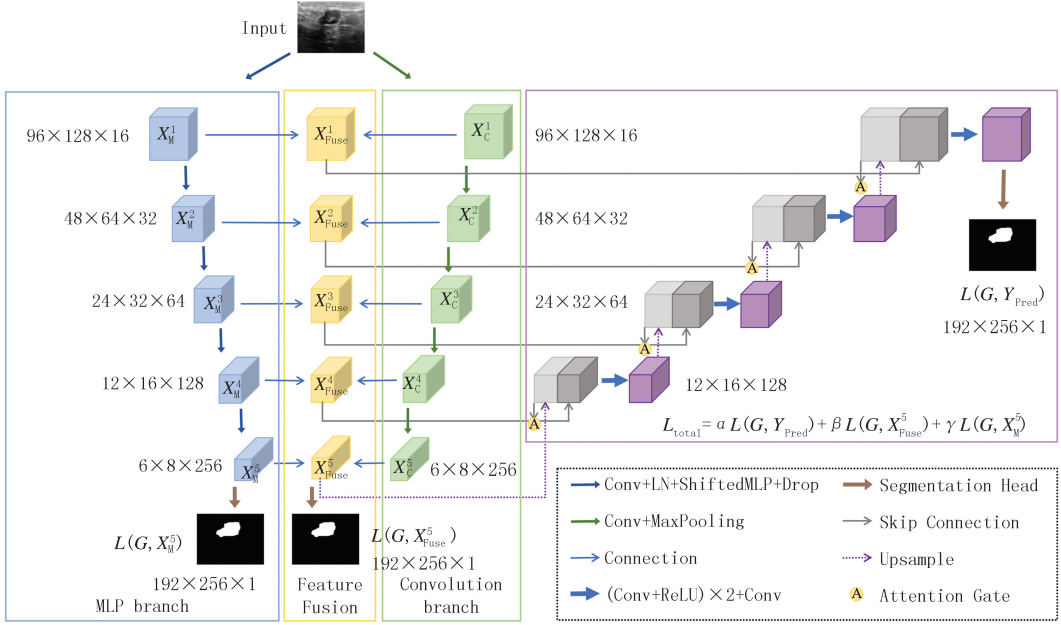


图1 MCNet网络结构

Fig.1 Architecture of the proposed MCNet

ShiftedMLP 通过线性移位操作获取特征之间的远程依赖.首先对特征图进行填充,填充大小为 $k/2$ (k 为超参数).然后沿通道方向将特征图平均划分为 k 组,并沿高度方向进行平移,平移步长为 $(-k/2, -k/2+1, \dots, 0, \dots, k/2-1, k/2)$.接着将 k 组平移后的特征沿通道方向拼接.最后将拼接后的特征图送入全连接层、深度方向卷积层、GeLU 激活函数和 Dropout 函数.沿宽度方向重复上述分组和平移操作.MLP 下采样模块的计算式为:

$$X_p^i = \varphi_L(\text{Conv}(X_M^{i-1})), \quad (1)$$

$$X_M^i = X + \sigma(\phi(\varphi_L(X_p^i))), \quad (2)$$

其中,输入特征 $X_M^{i-1} \in R^{C_{i-1} \times H \times W}$,输出特征 $X_M^i \in R^{C_i \times \frac{H}{2} \times \frac{W}{2}}$, i 表示下采样的次数,从 1 开始.当 $i=1$ 时, X_M^0 对应初始输入图像 X . C_i 表示特征图的通道数, H 、 W 分别表示特征图的高度和宽度. $\text{Conv}(\cdot)$ 表示卷积, $\varphi_L(\cdot)$ 表示层归一化, $\sigma(\cdot)$ 表示随机丢弃, $\phi(\cdot)$ 表示 ShiftedMLP.

2.3 Convolution 分支

Convolution 分支包含 5 个 Convolution 模块,每个 Convolution 模块包含一个卷积层、一个归一化层、一个池化层以及一个激活函数.卷积核的大小为 3×3 、步幅和 padding 均为 1.池化层的窗口大小为 2×2 .激活函数选用 ReLU,以提高模型逼近复杂函数的能力.CNN 下采样模块的计算式为:

$$X_C^i = R_e(M_P(\varphi_B(\text{Conv}(X_C^{i-1})))), \quad (3)$$

其中,输入特征 $X_C^{i-1} \in R^{C_{i-1} \times H \times W}$,输出特征 $X_C^i \in R^{C_i \times \frac{H}{2} \times \frac{W}{2}}$,当 $i=1$ 时, X_C^0 对应初始输入图像 X . $\varphi_B(\cdot)$ 表示批次归一化, $M_P(\cdot)$ 为最大池化, $R_e(\cdot)$ 为 ReLU 激活函数.

2.4 融合模块

MCNet 使用 BiFusion 模块^[8]融合 MLP 分支得到的全局特征与 Convolution 分支获取的局部特征.Bi-Fusion 使用压缩激励机制^[13]以增强对通道特征的感知能力.首先使用全局平均池化算子对特征图进行压缩.接着借助两个全连接层表征通道间的相关性,并输出与输入特征尺寸大小相同的权重以实现激励.再使用一个全连接层将通道特征维度降低到输入的 $1/N$,其中 N 为超参数,分别取 1、1、1、2、4.然后使用 ReLU 激活函数和另一个全连接层将特征图的大小上调至原始尺寸.采用多组全连接层的收益在于提供更多非线性操作以更好地拟合通道之间的相关性,并显著减少参数量.使用 Sigmoid 函数获得标准化的权重.最后通

过逐元素相乘将通道权重与原特征图相乘,以生成最终的特征图.压缩注意力子模块的计算式为:

$$X_{\text{MSE}}^i = X_{\text{M}}^i \odot (S(F_{\text{C}}(A_{\text{P}}(X_{\text{M}}^i))))), \quad (4)$$

其中, X_{MSE}^i 表示经过压缩注意力机制之后得到的特征图, $A_{\text{P}}(\cdot)$ 表示平均池化, F_{C} 为全连接层, $S(\cdot)$ 为 sigmoid 函数, $\odot$ 表示点乘操作.

为增强 Convolution 分支对局部细节的建模能力,引入空间注意力子模块^[14].首先沿输入特征图的通道维度分别进行最大池化和平均池化操作.之后将这两个池化操作的结果沿通道方向拼接得到新的特征图.接着引入一个额外的 1×1 卷积层,将特征图的通道数降至 1 以便进行特征加权与融合.卷积层的输出结果经 sigmoid 激活函数映射到 $0 \sim 1$, 以表示每个位置的注意力权重.最后将这些注意力权重与原始特征图进行逐元素相乘,生成强调关键信息的新特征表示.这种结合最大池化和平均池化的空间注意力子模块能够有效聚焦关键特征,实现自适应的特征重新加权.该子模块的计算式为:

$$X_{\text{CSA}}^i = X_{\text{C}}^i \odot (S(\text{Conv}(C_{\text{t}}(M_{\text{P}}(X_{\text{C}}^i), A_{\text{P}}(X_{\text{C}}^i))))), \quad (5)$$

其中, X_{C}^i 是 Convolution 分支的输出, X_{CSA}^i 表示经过空间注意力操作之后得到的特征图, $C_{\text{t}}(\cdot)$ 为拼接操作.

融合 MLP 模块和 Convolution 模块生成的特征图,以增强网络的特征表达能力.具体而言,通过卷积提取特征,之后将这两组特征逐元素相乘,然后通过核为 1×1 的卷积减少特征图的通道数,最后使用双线性池化减小特征图的尺寸.这样得到综合全局上下文和局部上下文两种信息的低维特征表达.对应的计算式为:

$$X_{\text{BP}}^i = B_{\text{P}}(\text{Conv}(\varphi_{\text{B}}(\text{Conv}(X_{\text{M}}^i)) \odot \varphi_{\text{B}}(\text{Conv}(X_{\text{C}}^i)))), \quad (6)$$

其中, X_{BP}^i 为得到的特征图, $B_{\text{P}}(\cdot)$ 表示双线性池化.

融合先前获取 3 个特征图以产生更丰富的信息.首先将两个分支网络的输出 X_{MSE}^i 、 X_{CSA}^i 和双线性池化的输出 X_{BP}^i 沿通道维度拼接.接着通过卷积层和 ReLU 激活层从拼接的特征图学习到更高级的特征表达.然后使用核为 1×1 的卷积减少特征图的通道数.最后引入残差连接,将上述第 2 步与第 3 步的结果连接.残差结构有助于缓解梯度消失,并使网络在更深层次学习到新的特征表示.这种多分支的结构设计充分融合了不同分支的输出特征,从而增强了网络的表征能力.上述操作的计算式为:

$$X_{\text{F}}^i = C_{\text{t}}(X_{\text{MSE}}^i, X_{\text{CSA}}^i, X_{\text{BP}}^i), \quad (7)$$

$$X_{\text{Fuse}}^i = \text{Conv}(X_{\text{F}}^i) + \text{Conv}(R_{\text{e}}(\varphi_{\text{B}}(X_{\text{F}}^i))), \quad (8)$$

其中, X_{F}^i 为沿通道拼接 X_{MSE}^i 、 X_{CSA}^i 、 X_{BP}^i 得到的特征图, X_{Fuse}^i 是融合模块的输出.

2.5 Loss 函数

使用加权二元交叉熵与交并比的和作为损失函数 $L(\cdot)$ 对网络进行训练.其定义为:

$$L_{\text{total}} = \alpha L(G, Y_{\text{Pred}}) + \beta L(G, X_{\text{Fuse}}^5) + \gamma L(G, X_{\text{M}}^5), \quad (9)$$

其中, Y_{Pred} 是网络的最终预测结果, X_{Fuse}^5 是 MLP 分支与 Convolution 分支最深层融合的输出结果, X_{M}^5 是 MLP 分支最深层的输出结果, α, β, γ 是超参数, 实验中分别设置为 0.5、0.3、0.2. G 为真实标签, L_{total} 为总损失函数.

3 实验结果与分析

分别在 BUSI^[15] 和 ISIC2018^[16] 数据集上进行实验,对比本文模型和其他先进方法.

3.1 数据集

BUSI 数据集包含 780 幅超声图像和对应的分割掩码,这些图像被分为 3 类:正常、良性和恶性.我们只使用良性和恶性图像,共 647 幅.ISIC2018 数据集包含 2 594 幅皮肤癌图像与对应的分割图.使用 5 折交叉验证进行实验.

3.2 实验设置

实验在配置有 Intel Xeon E5-2620 CPU 和 NVIDIA A100-PCIE-80GB GPU 的工作站上进行.设置学习率为 0.001,使用动量为 0.9 的 Adam^[17] 优化器.训练总轮数为 300,批次大小为 8.若总损失函数经过连续 100 个 epoch 不下降则停止训练.数据增强策略为图像水平翻转、图像垂直翻转和图像对比度随机改变.

3.3 评价指标

使用参数量、GFLOPs、平均 CPU 推理时间、Dice 相似系数(D_{sc})和均交并比(U_m)作为评价指标。 D_{sc} 的定义为:

$$D_{sc}=\frac{2(P\cap T)}{P\cup T},\tag{10}$$

其中, P 为预测标签, T 为真实标签.

U_m 的定义为:

$$U_m=\frac{1}{c+1}\sum_{i=0}^c\frac{TP}{TP+FP+FN},\tag{11}$$

其中, c 为类别数, TP 为被正确预测的前景像素个数; FP 为被错误预测为前景的像素个数; FN 为被错误预测为背景的像素个数.

3.4 结果分析

将 MCNet 与经典的 CNN 方法比较,包括 U-Net、UNet++ 和 ResUNet;同时将其与基于 Transformer 的方法 TransFuse 和 MedT、基于 MLP 的方法 UNeXt 进行比较.实验结果如表 1 所示,其中加粗字体对应每列最优值.可以看出,本文提出的 MCNet 在 BUSI 和 ISIC2018 数据集上都取得了最好的结果.

由表 1 可见,MCNet 在 BUSI 数据集上的 D_{sc} 为 79.32%,比 U-Net、UNet++、ResUNet、MedT、TransFuse 和 UNeXt 分别提升了 5.65%、5.15%、14.90%、12.08%、0.11%、3.27%.其 U_m 为 70.84%,较以上 6 种模型分别提高 6.22%、10.76%、17.04%、13.99%、0.09%和 8.34%.在 ISIC2018 数据集上,MCNet 的 D_{sc} 为 89.62%,比以上 6 种模型分别提高 10.17%、0.79%、2.99%、4.75%、0.64%和 0.30%.MCNet 的 U_m 为 82.97%,比以上 6 种模型分别提高 11.94%、2.61%、3.71%、6.23%、0.95%和 1.86%.在参数量方面,MCNet 的参数量为 3.66 M,虽然比参数量最少的 UNeXt 多 1.19 M,但比 U-Net、UNet++、ResUNet 和 TransFuse 分别减少 88.21%、60.04%、71.93%、86.01%.在计算复杂度方面,MCNet 的 FLOPs 为 1.21 G,比 UNeXt 仅多出 0.63 G;比 U-Net、UNet++、ResUNet、MedT 和 TransFuse 分别减少 97.83%、96.53%、98.51%、94.30%和 86.01%.在 Intel Xeon E5-2620 CPU 上对分辨率为 192×256 的图像进行推理:MCNet 的平均 CPU 推理时间为 1 034 ms,分别比 U-Net、UNet++、ResUNet、MedT 和 TransFuse 减少 77.12%、82.64%、84.11%、91.57%和 35.47%.综合考虑分割精度、参数数量和平均 CPU 推理时间等评价指标,MCNet均优于现有先进网络.

表 1 不同模型的实验结果对比
Tab. 1 Comparison of experimental results of different models

模型	参数量/M	FLOPs/ G	平均 CPU 推 理时间/ms	BUSI		ISIC2018	
				D_{sc}	U_m	D_{sc}	U_m
U-Net ^[2]	31.04	55.84	4 519	0.736 7±0.023 7	0.646 2±0.023 2	0.794 5±0.018 8	0.710 3±0.021 8
UNet++ ^[4]	9.16	34.91	5 957	0.741 7±0.022 3	0.600 8±0.028 7	0.888 3±0.015 2	0.803 6±0.022 9
ResUNet ^[3]	13.04	80.98	6 509	0.644 2±0.032 3	0.538 0±0.034 0	0.866 3±0.006 1	0.792 6±0.008 8
MedT ^[7]	1.60	21.24	12 264	0.672 4±0.038 6	0.568 5±0.043 8	0.848 7±0.022 0	0.767 4±0.027 1
TransFuse ^[8]	26.17	8.65	1 609	0.792 1±0.016 6	0.707 5±0.021 3	0.889 8±0.004 4	0.820 2±0.005 6
UNeXt ^[10]	1.47	0.58	298	0.760 5±0.003 7	0.625 0±0.003 9	0.893 2±0.007 0	0.811 1±0.011 1
MCNet	3.66	1.21	1 034	0.793 2±0.010 7	0.708 4±0.012 4	0.896 2±0.004 2	0.829 7±0.005 3

3.5 消融实验

在 BUSI 和 ISIC2018 数据集上进行消融实验以验证 Convolution 分支和 MLP 分支的有效性.更改 MCNet 的网络结构,分别关闭 MLP 分支(-MLP)和 Convolution 分支(-CNN)并测试其分割性能,结果如表 2 所示,其中加粗字体对应每列最优值.从表 2 可以看出,编码器仅使用 Convolution 分支时,网络(-MLP)在 BUSI 数据集上的 D_{sc} 为 75.62%,比 MCNet 低 3.7%;网络的 U_m 为 66.5%,比 MCNet 低 4.34%.网络(-MLP)在 ISIC2018 数据集上的 D_{sc} 为 89.15%,比 MCNet 低 0.47%;网络的 U_m 为 82.25%,比 MCNet 低

0.72%.显然,编码器仅使用 Convolution 分支容易导致全局信息丢失,从而降低分割精度.编码器仅使用 MLP 分支时,网络(-CNN)在 BUSI 数据集上的 D_{sc} 为 78.95%,比 MCNet 降低 0.37%; U_m 为 70.40%,比 MCNet 降低 0.44%.网络在 ISIC2018 数据集上的 D_{sc} 为 89.46%,比 MCNet 降低 0.16%; U_m 为 82.84%,比 MCNet 降低 0.13%.这表明 MLP 分支获取的全局特征对提高网络分割精度起重要作用.

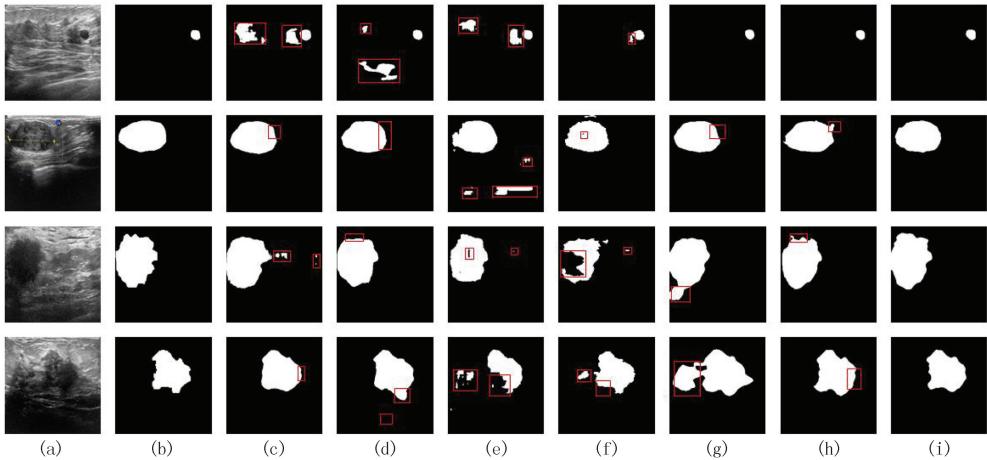
表 2 MLP 分支和 Convolution 分支对分割性能的影响

Tab. 2 Effect of MLP branch and Convolution branch on segmentation performance

模型	BUSI		ISIC2018	
	D_{sc}	U_m	D_{sc}	U_m
-MLP	0.756 2±0.019 9	0.665 0±0.024 8	0.891 5±0.014 6	0.822 5±0.004 1
-CNN	0.789 5±0.018 2	0.704 0±0.022 4	0.894 6±0.002 4	0.828 4±0.003 1
MCNet	0.793 2±0.010 7	0.708 4±0.012 4	0.896 2±0.004 2	0.829 7±0.005 3

3.6 分割结果可视化

为了展示 MCNet 的优越性,将其与其他网络在 BUSI 和 ISIC2018 数据集上的分割结果进行可视化对比,结果如图 2 和图 3 所示.



(a) 列为原始图像; (b) 列为真实标签; (c) ~ (h) 列分别为U-Net、U-Net++、ResUNet、MedT、TransFuse和UNetXt的分割结果; (i) 列为MCNet的分割结果. 图3同.

图2 不同模型在BUSI数据集上的分割结果对比

Fig.2 Comparison of segmentation results of different models on BUSI dataset

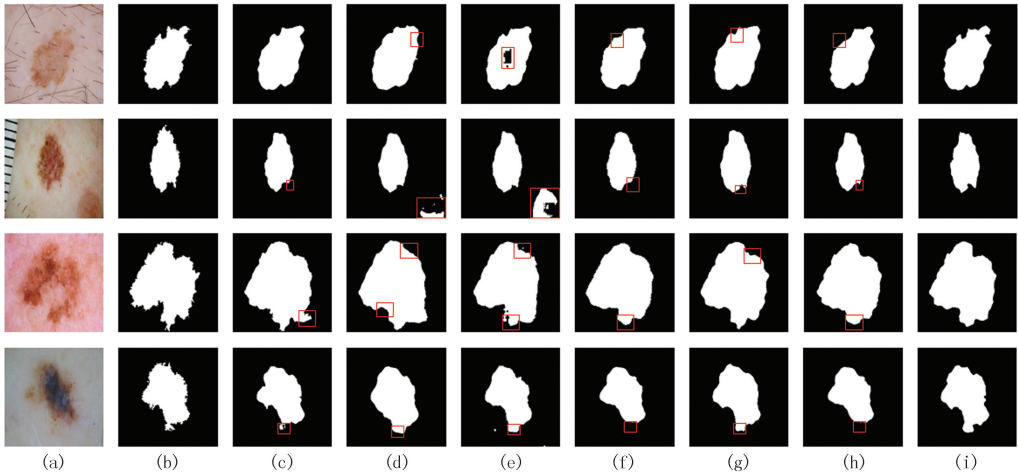


图3 不同模型在ISIC2018数据集上的分割结果对比

Fig.3 Comparison of segmentation results of different models on ISIC2018 dataset

图 2 展示了 BUSI 数据集上的分割结果.首先,U-Net、U-Net++、ResUNet 和 MedT 的分割结果较差、出现误分割的现象,这是因为这些方法仅关注局部特征或全局特征.特别是在从上到下第 1 行和第 4 行,这些方法无法准确地分割出目标对象.其次,U-Net 和 TransFuse 在第 2 行和第 3 行的分割结果也不理想.尽管 TransFuse 融合了局部特征和全局特征,但其分割结果与真实值之间仍存在较大的差异,这表明其对全局特征的提取不够充分.MedT 在第 3 行生成了大面积的欠分割结果,这表明仅提取全局特征不能有效解决这类复杂的分割问题.本文提出的 MCNet 在所有样例上都取得了最好的分割结果,这表明其融合全局和局部特征的能力最强,分割效果最好.

图 3 展示了 ISIC2018 数据集上的分割结果.从上到下,第 1 行的结果中,ResUNet 出现了中心区域欠分割的现象.这是由于 ResUNet 在处理中心区域时无法准确地捕捉到局部细节信息,导致目标区域的欠分割.在第 2 行,U-Net++和 ResUNet 的分割结果都出现严重的误分割现象.这是因为这两种方法在处理复杂场景时不能很好地捕捉全局上下文信息.在第 3 行和第 4 行,现有方法不能很好地处理分割结果的边缘细节.与其他网络相比,MCNet 仍取得了最好的分割结果.

4 结 论

研究快速准确的医学图像分割方法具有重要的理论意义和应用价值.使用多层感知机代替 Transformer 能够更有效地提取全局特征、减少参数数量并提高推理速度.本文提出一种融合多层感知机与卷积的双分支网络 MCNet,实现了轻量快速准确的医学图像分割.MCNet 采用编码器-解码器架构,在编码阶段包含卷积和多层感知机两个分支,以分别提取医学图像的局部特征和全局特征;在解码阶段通过特征融合模块有效融合这两种特征,从而提高分割精度并减少参数数量.与其他主流方法相比,MCNet 在 BUSI 和 ISIC2018 两个数据集上都取得了最好的分割结果.MCNet 的分割结果仍然存在分割边缘不准确、不平滑等不足.未来将针对这些问题开展研究,以进一步优化其分割性能.

参 考 文 献

[1] LONG J,SHELHAMER E,DARRELL T.Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition.Boston:IEEE,2015:3431-3440.

[2] RONNEBERGER O,FISCHER P,BROX T.U-Net:convolutional networks for biomedical image segmentation[M]//Lecture Notes in Computer Science.Cham:Springer International Publishing,2015:234-241.

[3] XIAO X,LIAN S,LUO Z M,et al.Weighted res-UNet for high-quality retina vessel segmentation[C]//2018 9th International Conference on Information Technology in Medicine and Education.Hangzhou:IEEE,2018:327-331.

[4] ZHOU Z W,SIDDIQUEE M M R,TAJBAKSH N,et al.UNet++:redesigning skip connections to exploit multiscale features in image segmentation[J].IEEE Transactions on Medical Imaging,2020,39(6):1856-1867.

[5] DOSOVITSKIY A,BEYER L,KOLESNIKOV A,et al.An image is worth 16x16 words:Transformers for image recognition at scale [C]//2021 International Conference on Learning Representations.[S.l.:s.n.],2021:1-22.

[6] LIU Z,LIN Y T,CAO Y,et al.Swin Transformer:hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision.Montreal:IEEE,2021:10012-10022.

[7] VALANARASU J M J,OZA P,HACIHALILOGLU I,et al.Medical transformer:gated axial-attention for medical image segmentation [M]//Lecture Notes in Computer Science.Cham:Springer International Publishing,2021:36-46.

[8] ZHANG Y D,LIU H Y,HU Q.TransFuse:fusing transformers and CNNs for medical image segmentation[M]//Lecture Notes in Computer Science.Cham:Springer International Publishing,2021:14-24.

[9] TOLSTIKHIN I O,HOULSBY N,KOLESNIKOV A,et al.MLP-mixer:An all-MLP architecture for vision[C]//2021 International Conference on Neural Information Processing Systems.Cambridge:MIT Press,2021:24261-24272.

[10] VALANARASU J M J,PATEL V M,UNeXt:MLP-based rapid medical image segmentation network[M]//Lecture Notes in Computer Science.Cham:Springer Nature Switzerland,2022:23-33.

[11] ZHU W H,TIAN J Y,CHEN M Z,et al.MSS-UNet:a multi-spatial-shift MLP-based UNet for skin lesion segmentation[J].Computers in Biology and Medicine,2024,168:107719.

[12] SCHLEMPER J,OKTAY O,SCHAAP M,et al.Attention gated networks:learning to leverage salient regions in medical images[J].Medical Image Analysis,2019,53:197-207.

[13] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132-7141.

[14] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 3-19.

[15] AL-DHABYANI W, GOMAA M, KHALED H, et al. Dataset of breast ultrasound images[J]. Data in Brief, 2020, 28: 104863.

[16] CODELLA N C F, GUTMAN D, CELEBI M E, et al. Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)[C]//2018 IEEE 15th International Symposium on Biomedical Imaging. Washington: IEEE, 2018: 168-172.

[17] KINGMA DP, BA J. Adam: A method for stochastic optimization[C]//2015 3rd International Conference on Learning Representations. San Diego: [s.n.], 2015.

MCNet: a lightweight lesion segmentation network integrating multilayer perceptrons and convolutions

Shen Hualei^{a,b,c}, Shangguan Guoqing^a, Yuan Chengyu^a, Chen Yanhao^{a,b,c}, Liu Dong^{a,b,c}

(a. School of Computer and Information Engineering; b. Henan Key Laboratory of Educational Artificial Intelligence and Personalized Learning; c. Big Data for Teaching Resources and Educational Quality Evaluation
Henan Engineering Laboratory, Henan Normal University, Xinxiang 453007, China)

Abstract: To address the shortcomings of existing medical image segmentation networks, such as high computational demands, significant hardware resource requirements, and slow inference speeds, a lightweight and fast segmentation network named MCNet is proposed. MCNet adopts an encoder-decoder architecture, utilizing both multilayer perceptron (MLP) and convolutions to extract and fuse global and local features of medical images, respectively, thereby reducing network parameters and improving segmentation accuracy. During the encoding stage, convolutional branches and MLP branches are used to extract multi-scale local and global features. These features are fused via skip connections and passed to the decoder. In the decoding stage, an attention gating mechanism is employed to enhance feature representation. Experiments were conducted on the BUSI and ISIC2018 datasets. Compared with state-of-the-art methods, MCNet achieves improvements in Dice similarity coefficient and mean Intersection over Union of 0.11% and 0.09% on the BUSI dataset, and 0.64% and 0.95% on the ISIC2018 dataset, respectively. Additionally, MCNet significantly reduces the number of network parameters, decreases the number of floating-point operations, and shortens CPU inference time.

Keywords: medical image segmentation; deep neural network; multi-layer perceptron (MLP); lightweight network

[责任编辑 赵晓华 刘洋]