

LM-UNet: 横向 MLP 用于增强 U-Net 的医学图像分割^①



邱海韬, 史 操

(青岛科技大学 信息科学技术学院, 青岛 266061)

通信作者: 史 操, E-mail: caoshi@yeah.net

摘要: 卷积神经网络 (CNN) 作为医学图像分割领域中 U-Net 基线网络的重要组成部分, 其主要作用是处理局部特征信息之间的关系. 而 Transformer 是一种能够有效强化特征信息之间的远距离依赖关系的视觉模型. 目前的研究表明, 结合 Transformer 和 CNN 可以在一定程度上提高医学图像分割的准确性. 但是, 由于医学图像的标注数据较少, 而且训练 Transformer 模型需要大量数据, 这使得 Transformer 模型面临耗时长和参数量大的挑战. 基于这些考虑, 本文在 UNeXt 模型的基础上, 结合多尺度混合 MLP 和 CNN, 提出了一种新型的基于混合 MLP 的医学图像分割模型——LM-UNet. 这种模型能够有效地增强局部与全局信息之间的联系, 并加强特征信息间的融合. 在多个数据集上的实验表明, LM-UNet 模型在皮肤数据集上的分割性能明显提升, 平均 Dice 系数达到 92.58%, 平均 IoU 系数达到 86.52%, 分别比 UNeXt 模型提高了 3% 和 3.5%. 在软骨和乳腺数据集上的分割效果也有显著提升, 平均 Dice 系数分别比 UNeXt 提高了 2.5% 和 1.0%. 因此, LM-UNet 模型不仅提高了医学图像分割的准确性, 还增强了其泛化能力.

关键词: 医学图像分割; MLP; 多尺度横向连接; U-Net

引用格式: 邱海韬, 史操. LM-UNet: 横向 MLP 用于增强 U-Net 的医学图像分割. 计算机系统应用, 2024, 33(5): 110-117. <http://www.c-s-a.org.cn/1003-3254/9510.html>

LM-UNet: Lateral MLP Augmented U-Net for Medical Image Segmentation

QIU Hai-Tao, SHI Cao

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: Convolutional neural network (CNN), as an important part of U-Net baseline networks in the field of medical image segmentation, is mainly used to deal with the relationships among local feature information. Transformer is a visual model that can effectively strengthen the long-distance dependency among feature information. The previous study shows that Transformer can be combined with CNNs to improve the accuracy of medical image segmentation to a certain extent. However, labeled data in medical images are rarely available while a large amount of data is required to train the Transformer model, exposing the Transformer model to the challenges of high time consumption and a large number of parameters. Due to these considerations, this paper proposes a novel medical image segmentation model based on a hybrid multi-layer perception (MLP) network by combining the multi-scale hybrid MLP with a CNN based on the UNeXt model, namely, the LM-UNet model. This model can effectively enhance the connection between local and global information and strengthen the fusion between feature information. Experiments on multiple datasets reveal significantly improved segmentation performance of the LM-UNet model on the international skin imaging collaboration (ISIC) 2018 dataset manifested as an average Dice coefficient of 92.58% and an average intersection over union (IoU) coefficient of 86.52%.

① 基金项目: 国家自然科学基金 (61806107, 61702135, 62201314); 山东省外国专家团队项目 (WST2021020)

收稿时间: 2023-11-02; 修改时间: 2023-12-04, 2024-01-10; 采用时间: 2024-01-18; csa 在线出版时间: 2024-04-01

CNKI 网络首发时间: 2024-04-03

which are 3% and 3.5% higher than those of the UNeXt model, respectively. The segmentation effects of the proposed model on the osteoarthritis initiative-zuse institute Berlin two-dimensional (OAI-ZIB 2D) and the breast ultrasound image (BUSI) datasets are also substantially superior, represented as average Dice coefficients 2.5% and 1.0% higher than those of the UNeXt counterpart, respectively. In summary, the LM-UNet model not only improves the accuracy of medical image segmentation but also provides better generalization performance.

Key words: medical image segmentation; MLP; multi-scale lateral connection; U-Net

1 引言

随着医学影像设备的快速发展和普及、成像技术已广泛应用于临床实践,成为疾病诊断、手术规划和预后评估等过程中不可或缺的辅助工具.医学影像分割是医学影像分析的重要领域,也是计算机辅助诊断、监测、干预和治疗的必要组成部分,其关键任务之一就是分割医学图像中的目标对象(如器官或病变)^[1,2],从而为下游任务提供如详细分析^[3]、合理诊断、预测和预防疾病等方面的有价值信息.但目前医学影像分割仍存在数据掩码少^[4]、组织边界信息弱^[5]以及分割精度低等问题.为了探索这些问题,Ben-Cohen等人^[6]首次尝试将全卷积网络(FCN)应用于计算机断层扫描(CT)检查中的肝脏分割和肝转移检测任务,对一个相对较小的数据集上进行研究,结果表明FCN优于基于补丁的卷积神经网络(CNN)和基于稀疏性的分类方案.U-Net由Ronneberger等人^[7]首次提出,该架构由一个用于捕捉上下文的收缩路径和一个用于精确定位的对称扩展路径组成,因此这种网络可以通过几张图像进行端到端^[8]训练,这使得U-Net成为医学图像分割任务的首选网络.Zhou等人^[9]发现了一种编码器和解码器子网络通过一系列嵌套的密集跳跃路径连接的深度监督编码器-解码器网络,该网络中重新设计的跳跃路径的主要目的就是缩小编码器和解码器子网络特征图之间的语义差距^[7].NnU-Net是由Isensee等人提出的^[10],它是一种基于二维和三维vanilla U-Net的鲁棒性自适应框架,在Isensee等人在研究过程中摒弃了拟议网络设计中多余的花哨设计,重点关注在影响方法性能和通用性的关键方面.

近年来,随着网络的更新壮大,神经网络架构的轻量化得到了改善,但提高医学图像分割的准确性仍然是一个巨大的挑战.Valanarasu等人^[11]就提出了一种门控轴向注意模型,通过在自我注意中引入额外的控制机制以扩展现有的架构.基于多层感知机(MLP)的网

络,如MLP-Mixer(完全基于MLP的架构)^[12]、gMLP^[13]也被发现是可以较好地完成计算机视觉下游任务的重要手段.UNeXt^[14]是一种将CNN和MLP相结合的网络,其特点是模型的参数量较小、推理速度较快.因此在医学图像分割中的应用得到研究者越来越广泛的关注.但UNeXt仍存在忽略网络体系结构设计(例如跳跃连接)^[15],导致其在分割准确度方面还有待提升.

基于此,本文提出了一个可以有效增强特征间融合的多级多层感知机模块的LM-UNet(lateral MLP augmented U-Net for medical image segmentation)模型,该模型网络底部使用混合MLP块的对称结构.LM-UNet模型改进了U-Net^[7]的跳跃连接,在编码器端的每个层级上的一组卷积和降采样后得到的特征图进行卷积-维度卷积操作并特定标记化,然后分别在通道信息和空间信息对特征信息进行全连接操作,最后将增强后的特征融合到解码器中.LM-UNet模型包含3个阶段:卷积阶段、多层感知机阶段和横向多层感知机增强阶段.其中卷积阶段用于提取图像的局部特征,多层感知机阶段用于提取全局特征和增强边界信息,横向多层感知机增强阶段用于加强特征信息的融合^[16]以及增强局部特征信息和全局信息之间的联系.

2 方法

图1表示的是LM-UNet模型的总体架构.LM-UNet模型是一个分层的编码器-解码器架构,包括横向MLP块、混合MLP块和卷积模块,其中编码器执行的初始操作由卷积和降采样、卷积和维度卷积组成.每个横向MLP块由一个令牌化特征模块和多个混合MLP块组成,横向MLP块的输入大小为 x_i ,输出特征的尺寸与同层级输入特征的尺寸相同.与一般的跳跃连接不同,本文提出的LM-UNet模型设计了多级横向MLP块连接,将特征图标记化,并将其投射到混合MLP块中.该模型的前3层的横向连接被横向MLP块取代,

作为特征增强模块. 编码器和解码器端的特征图在横向 MLP 块的作用下, 使其空间信息和信道信息之间的联系得到了加强, 最后将处理后的特征图送入高层网络中进行处理. 此外, 在该模型的第 4 层和瓶颈层使用了 4 个混合 MLP 块, 目的是保持 U-Net 结构的对称性

和网络本身的稳定性. 区别于瓶颈层, 在模型的第 4 层添加了跳跃连接, 保持与前 3 层在形式上的一致性. 解码器包括编码器提取的特征及融合多层级上采样后的特征. 在整个 U-Net 架构中, 结合了卷积和 MLP, 两者相辅相成^[7,12].

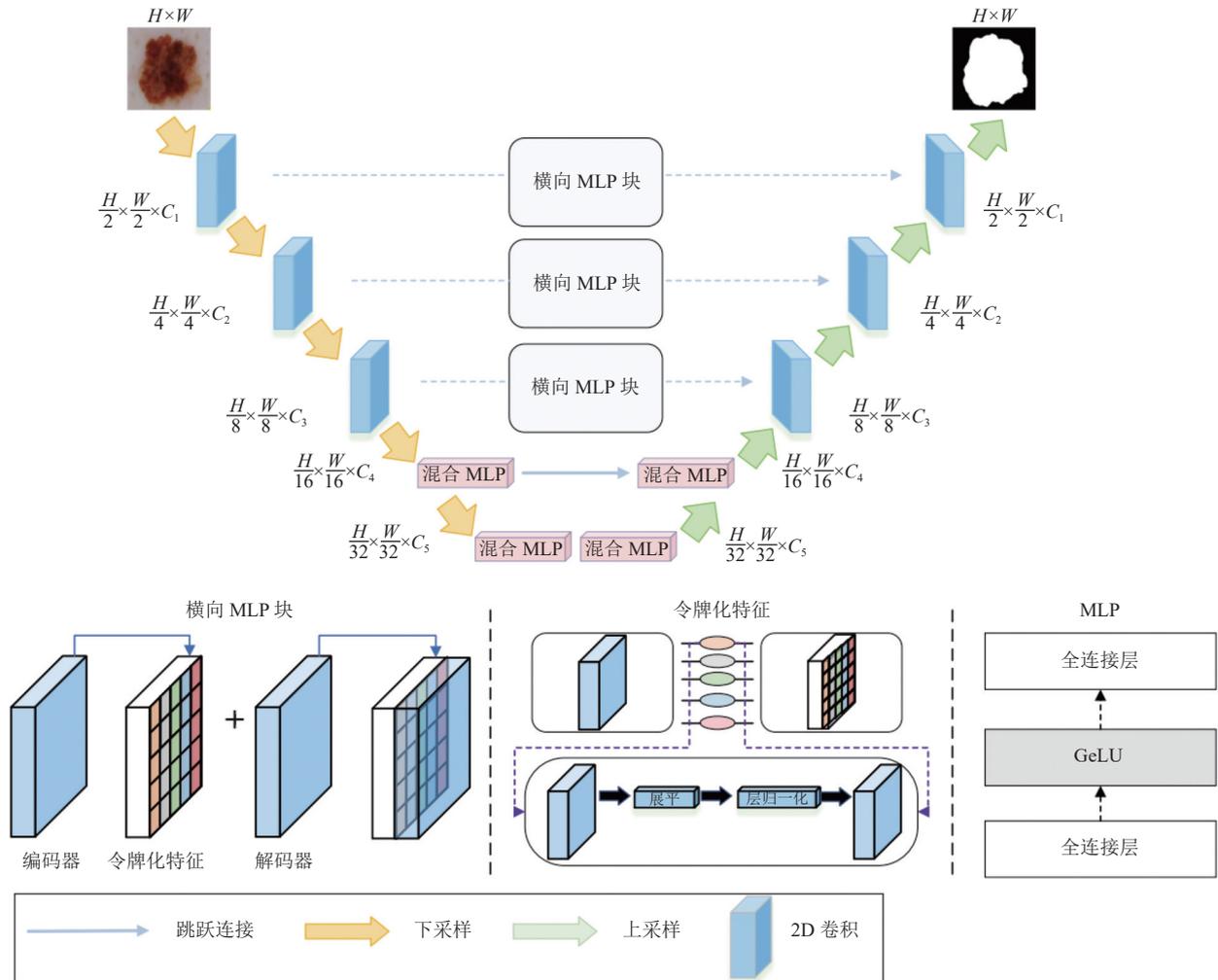


图 1 LM-UNet 的总体架构

2.1 卷积块

输入图像为 $x \in R^{H \times W \times C}$, 分辨率为 $H \times W$, 输入通道为 C . 每个卷积块都包含一个卷积层、一个批处理归一化层和一个 ReLU 非线性层. 使用的卷积核大小为 3×3 , 步长为 1, 填充为 1. 还使用了大小为 2×2 的最大池化层. 解码器端的路径包括 3 个用于特征提取的卷积块和 3 个上采样层. 每个降采样层的输出特征图分别为: $x_i \in R^{\frac{H}{k} \times \frac{W}{k} \times C_i}$, 其中参数的组合为 $(i, k) = \{(1, 2), (2, 4), (3, 8), (4, 16), (5, 32)\}$, 输出特征图由式 (1) 所得.

$$x'_i = \text{MaxPool}(\text{BN}(\text{Conv}(x_i))), i \in \{1, 2, 3\} \quad (1)$$

其中, x_i 表示第 i 输入特征图, x'_i 表示输出特征图. $\text{Conv}()$ 表示卷积操作, $\text{BN}()$ 表示批量归一化. $\text{MaxPool}()$ 是最大池化层.

在多层级令牌化特征图的输入端, 本文采用了双卷积块 (卷积-维度卷积) 操作. 如图 2 所示, 首先, 通过卷积操作将特征图的通道数做升维操作, 然后经过维度卷积块 (本质是卷积层), 其目的是降低在低层中的特征图维度, 因为参数过大的特征融合到混合 MLP 模块中容易造成过拟合现象. 因此本文添加了维度卷积

块,在令牌化特征图的输入端尽量缩减通道维度并且通道数是有规律的依次进行升维和降维。

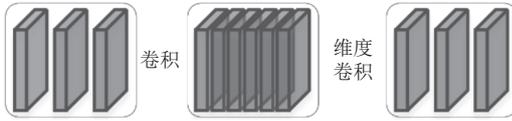


图2 多层次令牌化特征图的输入

2.2 混合 MLP 块

混合器将输入作为 S 个非重叠图像片段序列,并将每个片段序列投射到所需的隐藏维度 C 上,从而得到一个二维实值输入 $x \in R^{S \times C}$ 。如图3所示,混合 MLP 块包含令牌混合 MLP 和通道混合 MLP。令牌混合 MLP 操作后的输出和通道混合 MLP 操作后的输出分别如式(2)、式(3)所示。

$$u'_i = x'_i + W_2 \sigma(W_1 LN(x'_i)), i = 1, \dots, C \quad (2)$$

$$v'_j = u'_j + W_4 \sigma(W_3 LN(u'_j)), j = 1, \dots, S \quad (3)$$

其中, W_1 、 W_2 、 W_3 、 W_4 为权重。原始输入图像的分辨率为 $H \times W$, 每个补丁的分辨率为 $P \times P$, 补丁的数量为 $S = H \times W / P^2$ 个。 $LN()$ 表示归一化层。这里的 $\sigma()$ 是一个非线性函数 GELU^[17]。深层阶段的对称混合 MLP 块符合编码器-解码器结构,与 ViT^[18] 相比,它的参数数量更少,精度更高。

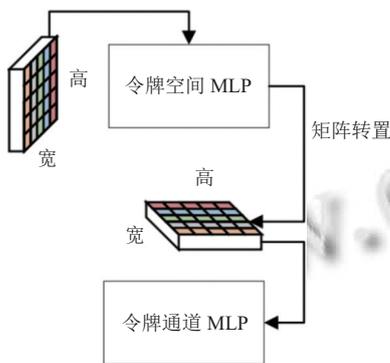


图3 混合 MLP 模块核心思想

2.3 横向 MLP 块

根据卷积块操作得到的特征输出, F_i 是每个输出特征图的维度, 让 $X = [x'_1, x'_2, \dots, x'_i, \dots, x'_N] \in R^{F_i \times N}$ 。随后, 通过归一化层将每个补丁展平并将维度扩充为 C_i , 每个补丁的大小为 3×3 。通过合并补丁融合特征图的空间和通道信息并作为混合 MLP 块的输入。标记化后的输入特征图、维度变换后的标记化输入特征图、包含混

合 MLP 块信息的输出特征图及横向 MLP 块获得特征图与解码器相应层获取的特征图的相加见式(4)–式(7)。

$$x'_i = Tok(x'_i) \quad (4)$$

$$FM_i = Resize(LN(x'_i)) \quad (5)$$

$$FM'_i = MixerLayer(FM_i) \quad (6)$$

$$Out_i = FM'_i + y'_i \quad (7)$$

其中, x'_i 表示第 i 层的输入特征图, $Tok()$ 是标记化特征图的过程。 FM'_i 表示包含混合 MLP 块信息的输出特征图。 $Resize()$ 用于调整特征图的尺寸, 以满足混合 MLP 块的输入需求。 y'_i 是由 3 个上采样层的卷积块提取的特征图。 Out_i 表示将横向 MLP 块获得的特征图与解码器相应层获得的特征图相加。

2.4 损失函数

交叉熵损失 (cross entropy loss) 和骰子损失 (Dice loss) 都被广泛用于一般的分割任务中。由于交叉熵损失在类别分布均匀的情况下表现较好, 而骰子损失则更适用于大尺寸的目标对象, 因此本文使用二元交叉熵 (binary cross entropy) 和骰子损失的线性组合来训练网络。损失函数如式(8)所示。

$$L = aBCE(Y, y) + bDice(Y, y) \quad (8)$$

其中, L 表示预测值 Y 和目标值 y 之间的损失函数, 参数 a 和 b 为可调的超参数。

2.5 数据规格和数据处理

为了评估 LM-UNet 模型, 在 ISIC 国际皮肤成像数据集^[19]、OAI-ZIB 软骨数据集^[20]和 BUSI 乳腺超声数据集^[21]这 3 个不同的数据集上进行实验。ISIC 2018 数据集由 2594 幅图像组成, 其中既有相机获取的皮肤病图像, 也有相应的皮损区域分割图, 所有图像的分辨率被调整为 512×512 。OAI-ZIB 数据集最初由 507 个带分割的三维 MR 图像组成, 使用的每张三维图像包含 160 个切片, 大小为 384×384 , 随机选择 407 张图像用于训练, 50 张用于验证, 50 张用于测试。BUSI 乳腺超声数据集使用良性和恶性图像, 总共有 630 张图像, 每张分辨率的大小为 256×256 。

3 实验结果

3.1 模型对比分析

表1展示的是 U-Net、U-Net++、ResUNet、Trans-

UNet、UNeXt 和 LM-UNet 这 6 种模型在 ISIC 2018 数据集上分割性能指标。由表 1 中的指标可以看出, 本文提出的模型分割性能最佳, ISIC 2018 数据集上的平均 IoU 系数最大值对应模型为 LM-UNet, 其平均 IoU 系数为 0.8652, 较 UNeXt 提高 3.2%。与 U-Net++, ResUNet、TransUNet 和 U-Net 模型相比, LM-UNet 的分割性能也有不同程度的提高, 平均 IoU 系数分别比前四者高 10.8%、9.8%、5.3% 和 2.8%。

表 1 ISIC 2018 数据集不同模型的实验分割性能

模型	参数量 (M)	GFLOPs	Average IoU
U-Net	7.85	55.84	0.8307
U-Net++	9.16	34.65	0.7577
ResUNet	62.74	94.56	0.7673
TransUNet	105.32	38.52	0.8123
UNeXt	1.47	0.57	0.8337
LM-UNet	14.14	53.79	0.8652

GFLOPs 是用来衡量神经网络模型计算速度和能力的指标, 从表 1-表 3 中看出, LM-UNet 模型展现出了较高的计算速度和推理能力。相比较 TransUNet 模型, 对于医学图像数据而言, 基于 Transformer 的模型耗时长, 这里从数据中看出, LM-UNet 模型比基于 Transformer 的模型能够更快地处理数据, 因此在训练和推理任务中更高效。

表 2 OAI-ZIB 2D 数据集不同模型的实验分割性能

模型	参数量 (M)	GFLOPs	Average IoU
U-Net	7.85	55.84	0.8195
U-Net++	9.16	34.65	0.7568
ResUNet	62.74	94.56	0.7659
TransUNet	105.32	38.52	0.7878
UNeXt	1.47	0.57	0.8058
LM-UNet	14.14	53.79	0.8470

表 3 BUSI 数据集不同模型的实验分割性能

模型	参数量 (M)	GFLOPs	Average IoU
U-Net	7.85	55.84	0.7850
U-Net++	9.16	34.65	0.8474
ResUNet	62.74	94.56	0.8009
TransUNet	105.32	38.52	0.8202
UNeXt	1.47	0.57	0.8355
LM-UNet	14.14	53.79	0.8502

LM-UNet 在跳跃连接部分添加了横向 MLP 块并二次加强了特征信息间的融合, 使模型具有多层级强化特征融合和边界信息处理的特点, 这也是 LM-UNet 在 ISIC 数据集上获得更高的分割性能的根本原因。

U-Net、U-Net++、ResUNet、TransUNet、UNeXt 和 LM-UNet 这 6 种模型在 OAI-ZIB 2D 数据集上分割

性能指标如表 2 所示。分析表 2 不难看出 OAI-ZIB 2D 数据集上分割效果最好的模型为 LM-UNet, 平均 IoU 系数 0.8470, 两个指标均为 6 种模型中的最大值。LM-UNet 具有最大平均 IoU 系数 0.8470, 比 UNeXt 高 4.2%。与 U-Net++、ResUNet、TransUNet 和 U-Net 模型相比, LM-UNet 在 OAI-ZIB 2D 数据集上分割性能也均较前四者有不同程度的提高, 平均 IoU 系数提高 2.8%–9.1%。LM-UNet 在 OAI-ZIB 2D 数据集上的高分割性能再次印证了模型的合理性。

表 3 是 BUSI 数据集上 6 种模型的分割性能指标。

从表 3 中分析可得, LM-UNet 模型在 BUSI 数据集上同样达到了最优分割性能, 平均 IoU 系数为 0.8470, 为 6 种模型中的最优值。LM-UNet 模型比 UNeXt 模型的平均 IoU 系数提高 1.5%。与 U-Net++、ResUNet、TransUNet 和 U-Net 模型相比, LM-UNet 的分割指标提高率如下: 平均 IoU 系数提高 1.5%–6.8%。BUSI 数据集上的实验结果表明, 在 2D 医学图像分割任务中, 横向 MLP 块具有优化性能和数据泛化的潜力, 以及更高的捕捉高水平语义信息能力。

图 4 表示的是 6 个模型在 3 个不同数据集上得到的分割性能指标平均 Dice 系数的可视化结果。其平均 Dice 系数 0.9258, 为 6 种模型中的最高值。相比之下, 模型 UNeXt 的平均 Dice 系数仅为 0.9051, LM-UNet 模型较 UNeXt 模型提高 2%。与 U-Net++、ResUNet、TransUNet 和 U-Net 模型相比, LM-UNet 的分割性能也有不同程度的提高, 平均 Dice 系数较前四者依次提高 7.4%、7.2%、3.8% 和 1.9%; OAI-ZIB 2D 数据集上分割效果最好的模型为 LM-UNet, 平均 Dice 系数为 0.9170, 并且, LM-UNet 模型比 UNeXt 模型的平均 Dice 系数提高 2.5%。与 U-Net++、ResUNet、TransUNet 和 U-Net 模型相比, LM-UNet 在 OAI-ZIB 2D 数据集上分割性能也均较前四者有不同程度的提高, 平均 Dice 系数提高 1.7%–5.7%。在 BUSI 数据集上, 平均 Dice 系数为 0.9170, 为 6 种模型中的最优值。LM-UNet 模型比 UNeXt 模型的平均 Dice 系数提高 0.8%, 与 U-Net++、ResUNet、TransUNet 和 U-Net 模型相比, LM-UNet 的分割指标提高率如下: 平均 Dice 系数提高 0.8%–4.8%。

相比于其他模型, LM-UNet 模型在 2D 医学图像分割应用中具备更优的数据泛化性和鲁棒性。不同特性的医学图像数据集展现出了各自独特的语义信息和

分割挑战. 而 LM-UNet 模型在 3 种数据集上均具备较高的分割性能, 这说明该模型可以更好地在训练过程中学习到更多不同类别图像特性和丰富的语义信息, 并能够通过横向 MLP 块对多层次特征的加强融合以及对多层次特征边界信息的处理, 最终在多个数据集上获得最优分割性能和最接近真实分割值的分割结果.

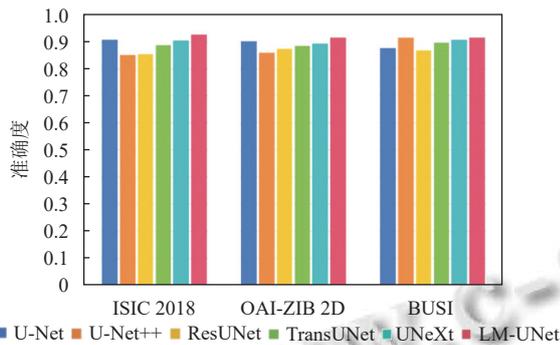


图 4 不同模型在 3 种数据集上的平均 Dice 系数对比

3.2 消融实验

选取 ISIC 2018 数据集和 OAI-ZIB 2D 数据集进行消融实验, 消融实验结果如表 4、表 5 所示.

表 4 ISIC 2018 数据集上的消融实验

SC	LMB	LL	HL	FTL	Average IoU	Average Dice
√	—	—	—	—	0.8147	0.8954
—	√	√	—	—	0.7954	0.8793
—	√	—	√	—	0.8092	0.8893
—	√	—	—	√	0.8652	0.9258

注: 跳跃连接 (SC), 横向 MLP 块 (LMB), 网络低层 (LL), 网络高层 (HL), 前 3 层 (FTL)

表 5 OAI-ZIB 2D 数据集上的消融实验

SC	LMB	(32, 256)	(256, 2048)	FTL	Average IoU	Average Dice
√	—	√	—	—	0.7558	0.8605
√	—	—	√	—	0.7564	0.8606
—	√	√	—	√	0.8470	0.9170
—	√	—	√	√	0.7530	0.8571

注: 跳跃连接 (SC), 横向 MLP 块 (LMB), 前 3 层 (FTL); (X_i, Y_i) 中 X_i 表示令牌混合维度, Y_i 表示通道混合维度

将混合 MLP 块中的令牌混合维度 (T_d) 初始参数设置为 32, 通道混合维度 (C_d) 初始为 256. 现有的 MLP 网络模型通常在宽和高两个方向上进行改进或在模型的瓶颈处进行创新. FF Only^[22]在分类任务中的结果表明, 当模型扩大时, 前馈式会有很好的表现. 相比于一个轻量级的模型架构 UNeXt, LM-UNet 模型的参数量在相对合理的范围内增加, 其本质是因为 MLP 的全连接特性, 横向 MLP 块能够较好地建立模型低层级的局部特征间的联系, 并且还能够加强不同层级的特

征之间的融合. 横向 MLP 块对特征图的令牌化分块处理后, 将每块的空间信息和通道信息均使用了全连接操作来进行充分的融合, 可以更好地对特征信息进行学习以及对分割图像的边界进行了加强处理. 因此, 本研究建立了由低层级的横向 MLP 块、4 个对称的混合 MLP 块和 U-Net 组成的混合模型. 本文提出的 LM-UNet 模型适当地扩大了模型参数量并优化了跳跃连接 (SC). 但是, 对于这项分割任务来说, 参数量不宜过大, 因为过大的参数量可能会导致基于 MLP 的模型出现过拟合. 对不同的参数进行实验, 发现参数越大, 效果可能越差. 实验结果表明, 在一般的跳跃连接中, 模型的平均 Dice 系数为 89.54%, 参数为 7.44M.

由表 4 中可以看出, 在前 3 层 (FTL) 添加横向 MLP 块 (LMB) 操作时, 当参数增加到 14.14M, 平均 Dice 系数增加了 3%. 若只在低层 (LL) 或者只在高层 (HL) 添加横向 MLP 块, 并不能够提高分割精度, 如表 4 所示, 尝试只在低层添加横向 MLP 块, 发现实验结果参数量相对较多, 而且无法对后续的特征图进行特征融合, 最终导致对图像语义信息的获取不够充分, 空间信息和信道信息的特征间的联系也没有得到增强. 为了验证 LM-UNet 模型的稳定性和通用性, 在几个具有不同特征的医学图像数据集上分别进行测试, 结果均显示取得了良好的分割效果. 这说明 LM-UNet 模型的确提高了医学图像分割的精确度.

表 5 显示, 如果 T_d 和 C_d 值过大, 模型很容易出现过拟合. 通过对比实验结果, 发现当 (T_d, C_d) 的组合为 (32, 256) 时, 分割效果最好, 在 OAI-ZIB 2D 数据集上的平均 Dice 系数达到 91.70%, 平均 IoU 系数达到 84.70%.

3.3 混合 U-Net 架构和 MLP 的优势

多阶段设计是 CNN 模型的标准设计, 但目前基于 Transformer 或基于 MLP 的模型设计中, 多阶段设计的重要性却在很大程度上被忽视了. 研究发现, 无论单独选择 CNN、Transformer 或 MLP 这 3 种网络结构中的哪一种作为单级框架, 其性能都远不及多级框架^[23,24]的性能.

通过轻量级深度卷积, 卷积模型在执行计算机视觉下游任务过程中具备较高性能. 当模型规模较小时, MLP 具有很强的性能, 但当模型规模增大时, MLP 就会出现严重的过拟合问题, 过拟合是 MLP 实现 SOTA 性能的主要障碍^[25], 因此, 控制模型的参数量以避免过拟合问题是决定模型优劣的关键. U-Net 已被广泛应用于医疗分割, 因此沿用这一架构, 使用 CNN 获取特征的局部关系, 而全局关系则利用 MLP 获取, 通过 MLP

融合空间信息和信道信息可以有效避免过多计算量和全局信息替换. 在 CNN 结合 MLP 基础上获得的 LM-UNet 模型保留了跳跃连接, 并且能够通过横向 MLP 块进一步优化多层次特征信息融合及边缘信息提取, 增强编码器和解码器之间的特征信息传递.

3.4 实验结果

6 种模型在 3 个不同数据集上的分割效果如图 5 所示. 与其他 5 种模型相比, 由于横向 MLP 块加强了通道特征信息和空间特征信息之间的关系, LM-UNet 模型让图像的边界信息分割的更加明确. 与基准模型 U-Net 一样, 图像分割的模糊混浊边界得到了改善, 分

割出的轮廓更加平滑、轮廓更加清晰. 不同数据上的结果显示 LM-UNet 模型可以针对不同类型的数据集进行分割, 并获取更接近标签值 (ground truth, GT) 的分割效果.

皮肤数据集 (A) 上分割效果的可视化结果显示, LM-UNet 模型在边缘的平滑度、边界轮廓的完整性和分割的准确性方面都取得了最佳结果. 对于医学影像分割来说, 从不同尺度的特征图中获得的信息是不同的, 横向 MLP 块在建立空间和通道信息联系方面发挥着关键作用, 不同的权重矩阵有助于提高 LM-UNet 模型的分割性能.

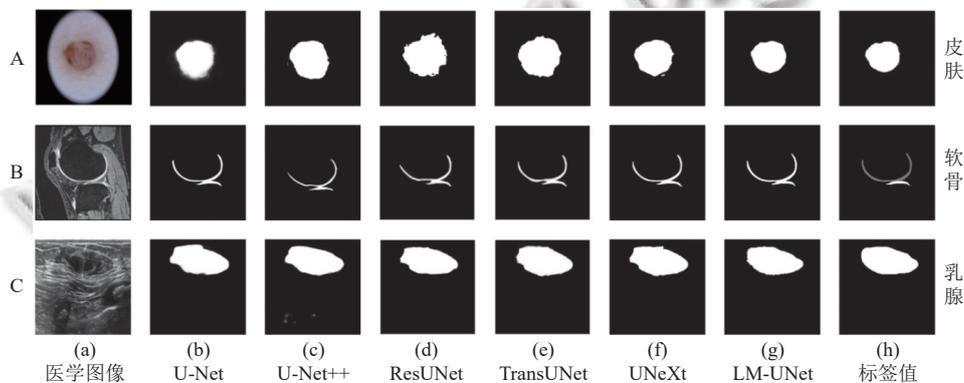


图 5 ISIC 2018 数据集 (A)、OAI-ZIB 2D 数据集 (B) 和 BUSI 数据集 (C) 上不同模型的分割效果

从软骨图像 (B) 的分割图中可以看出, LM-UNet 模型不论是对于上软骨组织还是下软骨组织都可以清晰分割出来并最接近标签真实值.

乳腺超声图像 (C) 的特点是背景和待分割区域不太明显, 但该模型仍能捕捉局部信息范围之间的联系和差异, 在合理的参数范围内获得更完整的语义信息, 分割结果也更加与标签真实值接近.

3.5 讨论

通过图 5 的对比, 可以更加直观地看到, 在横向 MLP 块的作用下, LM-UNet 模型实现的图像分割结果拥有更清晰、平滑的边界和轮廓, 其分割准确率显著高于其他模型. 横向 MLP 块通过混合空间信息和通道信息, 有效改善了边界特征的分割并提升了边界分割精度. 此外, 横向 MLP 块在模型的前 3 层进行二次特征提取, 增强了多层次特征中局部与全局特征间的联系, 从而使 LM-UNet 模型获得更精确的分割结果, 得到的分割图更加平滑、清晰, “云状”轮廓的结构更少.

4 结论与展望

本文提出了基于卷积和 MLP 混合的深度学习模型 LM-UNet, 该模型通过增加横向 MLP 块有效改进了一般的跳跃连接. 在 ISIC 2018、OAI-ZIB 和 BUSI 这 3 个不同的数据集上进行实验, 对比实验结果表明, LM-UNet 模型相比于 U-Net、U-Net++、ResUNet、TransUNet 和 UNeXt 模型有着更接近于标签值的分割结果, 且 LM-UNet 模型在 ISIC 2018、OAI-ZIB 和 BUSI 这 3 个数据集上的平均 Dice 系数相比于 UNeXt 模型分别提高 0.8%、2.0%、2.5%, 平均 IoU 系数分别提高 1.5%、2.9%、4.2%. 消融实验的结果也再次印证 LM-UNet 模型具有最佳的分割性能. 虽然本研究中 LM-UNet 模型具有较高的医学图像分割效果, 但 MLP 本身存在的过拟合问题仍未完全解决, 因此在后续的研究中应该重点关注如何精简 MLP 的全连接操作.

参考文献

- Spanhol FA, Oliveira LS, Petitjean C, *et al.* A dataset for breast cancer histopathological image classification. IEEE

- Transactions on Biomedical Engineering, 2016, 63(7): 1455–1462. [doi: [10.1109/TBME.2015.2496264](https://doi.org/10.1109/TBME.2015.2496264)]
- 2 Shi C, Xu CH, He JF, *et al.* Graph-based convolution feature aggregation for retinal vessel segmentation. *Simulation Modelling Practice and Theory*, 2022, 121: 102653. [doi: [10.1016/j.simpat.2022.102653](https://doi.org/10.1016/j.simpat.2022.102653)]
 - 3 Guan H, Liu MX. Domain adaptation for medical image analysis: A survey. *IEEE Transactions on Biomedical Engineering*, 2022, 69(3): 1173–1185. [doi: [10.1109/TBME.2021.3117407](https://doi.org/10.1109/TBME.2021.3117407)]
 - 4 Despotović I, Goossens B, Philips W. MRI segmentation of the human brain: Challenges, methods, and applications. *Computational and Mathematical Methods in Medicine*, 2015, 2015: 450341.
 - 5 Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: Overview, challenges and the future. In: Dey N, Ashour AS, Borra S, eds. *Classification in BioApps*. Cham: Springer, 2018. 323–350.
 - 6 Ben-Cohen A, Diamant I, Klang E, *et al.* Fully convolutional network for liver segmentation and lesions detection. *Proceedings of the 1st International Workshop on Deep Learning in Medical Image Analysis*. Athens: Springer, 2016. 77–85.
 - 7 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 234–241.
 - 8 Xu CH, Shi C, Chen YN. End-to-end dilated convolution network for document image semantic segmentation. *Journal of Central South University*, 2021, 28(6): 1765–1774. [doi: [10.1007/s11771-021-4731-9](https://doi.org/10.1007/s11771-021-4731-9)]
 - 9 Zhou ZW, Rahman Siddiquee M, Tajbakhsh N, *et al.* UNet++: A nested U-Net architecture for medical image segmentation. *Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis*. Granada: Springer, 2018. 3–11.
 - 10 Isensee F, Petersen J, Klein A, *et al.* Abstract: NnU-Net: Self-adapting framework for U-Net-based medical image segmentation. In: Handels H, Deserno TM, Maier A, *et al.*, eds. *Bildverarbeitung für die Medizin 2019*. Wiesbaden: Springer, 2019. 22.
 - 11 Valanarasu JMJ, Oza P, Hacihaliloglu I, *et al.* Medical Transformer: Gated axial-attention for medical image segmentation. *Proceedings of the 24th International Conference on Medical Image Computing and Computer-assisted Intervention*. Strasbourg: Springer, 2021. 36–46.
 - 12 Tolstikhin I, Houlsby N, Kolesnikov A, *et al.* Supplementary materials for: MLP-Mixer: An all-MLP architecture for vision. *Proceedings of the 35th Conference on Neural Information Processing Systems*. 2021. 24261–24272.
 - 13 Liu HX, Dai ZH, So DR, *et al.* Pay attention to MLPs. *Proceedings of the 35th Conference on Neural Information Processing Systems*. 2021. 9204–9215.
 - 14 Valanarasu JMJ, Patel VM. UNeXt: MLP-based rapid medical image segmentation network. *Proceedings of the 25th International Conference on Medical Image Computing and Computer-assisted Intervention*. Singapore: Springer, 2022. 23–33.
 - 15 Li H, Xu Z, Taylor G, *et al.* Visualizing the loss landscape of neural nets. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Montréal: Curran Associates Inc., 2018. 6391–6401.
 - 16 陈珊, 郑琴, 秦绪佳. 基于二维经验模式分解的医学图像融合方法. *计算机系统应用*, 2014, 23(6): 105–110.
 - 17 Hendrycks D, Gimpel K. Gaussian error linear units (GELUs). arXiv:1606.08415, 2016.
 - 18 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929, 2020.
 - 19 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). arXiv:1902.03368, 2019.
 - 20 Liu Q, Xu ZL, Jiao YN, *et al.* iSegFormer: Interactive segmentation via Transformers with application to 3D knee MR images. *Proceedings of the 25th International Conference on Medical Image Computing and Computer-assisted Intervention*. Singapore: Springer, 2022. 464–474.
 - 21 Al-Dhabyani W, Gomaa M, Khaled H, *et al.* Dataset of breast ultrasound images. *Data in Brief*, 2020, 28: 104863. [doi: [10.1016/j.dib.2019.104863](https://doi.org/10.1016/j.dib.2019.104863)]
 - 22 Melas-Kyriazi L. Do you even need attention? A stack of feed-forward layers does surprisingly well on ImageNet. arXiv:2105.02723, 2021.
 - 23 王金祥, 付立军, 尹鹏滨, 等. 基于 CNN 与 Transformer 的医学图像分割. *计算机系统应用*, 2023, 32(4): 141–148. [doi: [10.15888/j.cnki.csa.009010](https://doi.org/10.15888/j.cnki.csa.009010)]
 - 24 张潮, 宋亚林, 袁明阳. 基于 MultiResUNet-SMIS 的皮肤黑色素瘤图像分割. *计算机系统应用*, 2023, 32(6): 221–230. [doi: [10.15888/j.cnki.csa.009113](https://doi.org/10.15888/j.cnki.csa.009113)]
 - 25 Zhao YC, Wang GT, Tang CX, *et al.* A battle of network structures: An empirical study of CNN, Transformer, and MLP. arXiv:2108.13002, 2021.

(校对责编: 张重毅)