

文章编号: 1671-7449(2024)04-0448-07

BiUNet: 基于双层路由注意力的轻量化医学分割网络

王莹¹, 吴本阳², 郭晋川², 张萌², 原锌蕾³

(1. 山西大学 电力与建筑学院, 山西 太原 030006; 2. 山西大学 物理电子工程学院, 山西 太原 030006;
3. 山西财经大学 管理科学与工程学院, 山西 太原 030006)

摘要: 针对视觉 Transformer 骨干提取网络计算开销大, 模型训练缓慢的问题, 同时为了进一步提升 Transformer 结构在医学图像领域的分割性能, 提出一种名为 BiUNet 的新型轻量级 U 型架构的医学图像分割网络。将输入医学图像整切成若干图像块后, 送入一种基于双层路由动态稀疏注意力机制的 BiFormer 转换器中, 通过组合下采样和特定块数的 BiFormer 模块, 构建多级金字塔结构实现特征提取。随后通过组合上采样和卷积模块, 相应构建多级金字塔结构进行特征解码, 进而实现像素级语义分割。该模型在 3 个医学数据集上依次取得了 90.2%, 93.7% 和 85.6% 的 mIoU 值以及 5.55 G 的 Flops 和 28.10 M 的参数量。结果表明, BiUNet 能够以轻量化的效果有效提升医学图像分割的精度。

关键词: 双层路由注意力机制; Transformer 结构; 医学图像分割; 轻量级; U 型结构

中图分类号: TP911.73 **文献标识码:** A **doi:** 10.3969/j.issn.1671-7449.2024058

引用格式: 王莹, 吴本阳, 郭晋川, 等. BiUNet: 基于双层路由注意力的轻量化医学分割网络[J]. 测试技术学报, 2024, 38(4): 448-454.

WANG Ying, WU Benyang, GUO Jinchuan, et al. BiUNet: lightweight medical segmentation network based on Bi-level routing attention[J]. Journal of Test and Measurement Technology, 2024, 38(4): 448-454.

BiUNet: Lightweight Medical Segmentation Network Based on Bi-Level Routing Attention

WANG Ying¹, WU Benyang², GUO Jinchuan², ZHANG Meng², YUAN Xinlei³

(1. School of Electric Power, Shanxi University, Taiyuan 030006, China;

2. School of Physical and Electronic Engineering, Shanxi University, Taiyuan 030006, China;

3. School of Management Science & Engineering, Shanxi University of Finance and Economics, Taiyuan 030006, China)

Abstract: To solve the problems of high computing cost and slow model training in vision Transformer backbone extraction network, and to further improve the performance of Transformer structure in the field of medical image segmentation, a new lightweight U-architecture medical image segmentation network named BiUNet was proposed. The input medical image was cut into several blocks, and then the blocks were fed into the BiFormer based on the dynamic sparse attention mechanism of Bi-level routing. By combining downsampling and BiFormer modules with a specific number of blocks, a multi-level pyramid structure was constructed to achieve feature extraction. Subsequently, the feature map output from the encoder was decoded by a multi-level pyramid structure which was constructed by combining the upsampling and

收稿日期: 2023-08-21

作者简介: 王莹(1987-), 女, 讲师, 博士, 主要从事图像处理的研究。E-mail: wangying52@sxu.edu.cn.

convolution modules, and pixel-level semantic segmentation was realized. The model achieved 90.2%, 93.7% and 85.6% mIoU values as well as 5.55 G Flops and 28.10 M parameters on the three medical datasets sequentially. The results show that BiUNet can effectively improve the accuracy of medical image segmentation with a lightweight effect.

Key words: Bi-level routed attention mechanism; transformer structure; medical image segmentation; lightweight; U-shaped structure

至今为止,黑色素瘤,器官息肉等疾病仍然是导致人类痛苦甚至死亡的棘手问题,对于上述疾病的预防和治疗,相当依赖于先进的医学图像处理技术^[1]。自深度学习算法迅速发展以来,基于人工智能的计算机工作方式已经在各种领域广泛应用^[2]。而在医学领域,依托于医学图像分割结果的CAD(Computer-aided Diagnosis)^[3]方式也借助深度学习模型发生了重大变革。相较于分类和检测任务,通常医疗专业及相关人员更关注于语义分割的结果^[4]。

在语义分割任务上,尤其是医学图像分割任务,现存的分割模型结构重度依赖于UNet的U型网络结构^[5]。而U型网络结构是由编码器、解码器、跳跃连接3个核心模块组成。在ViT^[6](Vision Transformer)结构出现之前,在计算机视觉领域,由CNN(Convolutional Neural Networks)模块主导实现特征提取的功能。UNet本身的编码器和解码器模块核心部件由特定大小卷积核的卷积算子实现。研究者们利用CNN结合U型框架,提出了种种改进UNet的网络模型。这一类结构主要有UNet++^[7], UNet+++^[8], AACN-UNet^[9], DCSAU-Net^[10], Attention-UNet^[11]等。而ViT结构出现之后,将图片切割成图像块送入Transformer结构进行注意力计算的思想激励了研究者们针对这一新的特征提取骨干模块的探索。在TransUNet^[12]中,作者先利用CNN生成特征图,将其送入堆叠好的12个连续的Transformer块实现特征提取,之后,利用CNN实现特征解码,在多器官分割任务上取得了较好的分割结果。而在SwinUNet^[13]中,作者设计了独特的补丁扩展层,使用纯Transformer模块搭建的U型结构取得了比TransUNet在同一数据集上更好的分割结果。此外,利用Transformer结构进行医学图像分割的模型还有TransFuse^[14], CoTr^[15], DS-TransUNet^[16], TransClaw UNet^[17]等。虽然Transformer结构带来了优秀的特征提取能力,但是它所依赖的传统自注意力机制却带来了很大的计算开销。针对这一问题,有研究者开始探索轻量化的Transformer结构,如LeViT^[18], Swin Transformer等^[19]。而Zhu等^[20]提出

的BiFormer,在保证自身轻量化的同时,在图像分类、目标检测、语义分割等下游视觉任务上取得了超越Swin Transformer的效果。

为了解决现存医学图像分割网络计算开销大、训练缓慢的问题,更出于对实现更高分割精度的目标,本文在BiFormer的启发下,针对二维医学图像分割任务,提出了BiUNet。BiUNet在编码器阶段充分利用BiFormer优秀的特征提取能力,构建了金字塔式分级结构,实现分级特征图提取,利用双层路由机制保证模型轻量化,在解码器阶段利用传统的卷积操作和独特的上采样模块,实现特征解码和逐级分辨率恢复,输出分割结果。

1 方法

设计的BiUNet整体结构如图1所示。受典型U型架构的影响,BiUNet也由编码器、解码器、跳跃连接构成。编码器的核心特征提取模块由基于双层路由注意力机制的BiFormer构成。

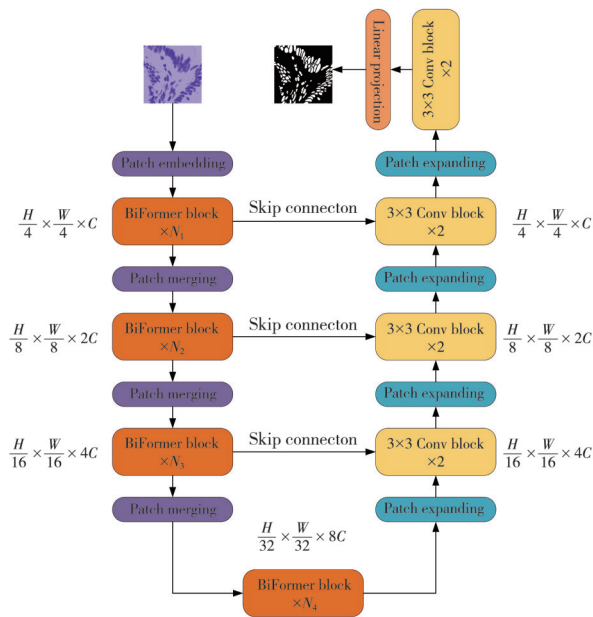


图1 BiUNet模型结构

Fig. 1 The Structure of BiUNet model

如图1所示,对于输入到网络里的一张皮肤病图片,送入Patch Embedding模块,处理成图像序列。

将处理后的图像序列送入特定深度的BiFormer块与Patch Merging模块构成的层级结构中学习多尺度的特征表示；随后送入由卷积模块和Patch Expanding, 将已学到的高级语义表示进行解析和分辨率逐级恢复；在解码之前, 将编码器多级结构里输出的不同分辨率特征图与对应解码器上采样模块输出的同分辨率大小特征图进行跳跃连接, 最后恢复至输入大小分辨率得到预测分割图。

1.1 编码器

在编码器阶段, 构建了4级金字塔式结构, 实现特征提取。在第1层的Patch Embedding中, 首先使用2个3*3大小的卷积核, 步长为2的卷积层将3通道的输入($H \times W \times C$)映射至通道维度 C (本文设置为64), 同时分辨率大小变至 $\frac{H}{4} \times \frac{W}{4}$, 再送入 N_1 个连续的BiFormer模块。在之后的3层里, 均为Patch Merging模块和BiFormer模块的组合, 第2层、第3层、第4层的BiFormer深度分别为 N_2 , N_3 和 N_4 (N_1, N_2, N_3, N_4 的取值在后文BiUNet-T和BiUNet-S配置中给出)。Patch Merging模块本质上是1个卷积核大小为 3×3 , 步长为2的卷积操作, 从而对特征图分辨率进行2倍下采样, 再通过对输出通道数的设置, 将通道数扩展为原先的2倍。经过这样的分层结构, 实现多尺度特征提取。

1.1.1 BiFormer

BiFormer的具体结构由图2所示。

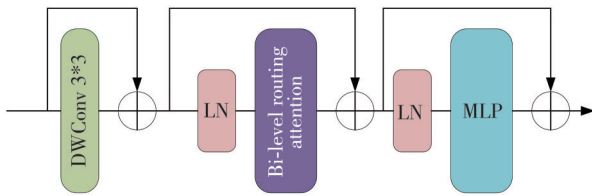


图2 BiFormer模块结构

Fig.2 The structure of BiFormer module

对于每一个具体的BiFormer结构, 均由一个 3×3 的深度卷积模块、一个层归一化模块和BRA (Bi-level Routing Attention)模块的组合, 以及一个层归一化模块和MLP()的组合以残差连接的方式组合而成。对于输入的特征图, 通过深度堆叠的BiFormer模块, 分辨率和通道数均维持不变。

1.1.2 双层路由注意力模块 (BRA)

BRA模块是一种动态的, 具有查询感知特性的稀疏注意力机制。它的核心思想是只保留一小

部分路由区域, 而在粗粒度上过滤掉最不相关的键值对。在细粒度上, 在所保留路由区域的并集中进行标记到标记的注意力的计算。以单个输入的单头自注意力为例, 如式(1), 首先将一个 $X \in R^{H \times W \times C}$ 的输入特征图划分成 $S \times S$ 个互不重叠的区域, 每个区域包含 $\frac{HW}{S^2}$ 个特征向量。这一步也可表述为将 X 重构成 $X_r \in R^{s^2 \times \frac{HW}{s^2} \times C}$ 。再对 X_r 线性投影得到 $Q, K, V \in R^{s^2 \times \frac{HW}{s^2} \times C}$ 。

$$Q = X^r W^q, K = X^r W^k, V = X^r W^v, \quad (1)$$

式中: $W^q, W^k, W^v \in R^{C \times C}$ 分别代表 Q, K, V 的映射权重。

第二步是先构造出一个有向图, 找到每个给定区域该关注的区域。具体实现是先通过平均每个区域的 Q 和 K , 来得出区域级的 Q_r, K_r 。如式(2), 通过 Q^r 和 K^{rT} 矩阵相乘得到区域到区域亲和力图的邻接矩阵 A^r 。 A^r 中的元素值实际上衡量了两个区域语义上的相关程度。

$$A^r = Q^r (K^r)^T. \quad (2)$$

接下来在每个区域仅保留 $top-k$ 个连接来裁剪亲和力图, 通过行向 $topk$ 算子构建一个路由索引矩阵 I^r 实现。如式(3)

$$I^r = topkIndex(A^r). \quad (3)$$

I^r 的第 i 行代表了第 i 个区域关联度最高区域的 k 个索引值。

最后一步根据索引矩阵 I^r , 进行细粒度的注意力计算。即对于第 i 个区域中每个查询标记, 将以 $I_{(i,1)}^r, I_{(i,2)}^r, \dots, I_{(i,k)}^r$ 的索引方式, 来关注 k 个路由区域并集区域中的所有键值对, 如式(4), 聚集键张量和值张量得到聚集后的 K^g 和 V^g 。如式(5), 对 K^g 和 V^g 作注意力计算。LCE()函数实际上是通过一个 5×5 核大小的深度卷积进行参数化。

$$K^g = gather(K, I^r), V^g = gather(V, I^r), \quad (4)$$

$$O = Attention(Q, K^g, V^g) + LCE(V). \quad (5)$$

整个BRA计算复杂度的计算如式(6)

$$\begin{aligned} FLOPs &= FLOPs_{proj} + FLOPs_{routing} + FLOPs_{attn} = \\ &= 3HWC^2 + 2(s^2)C + 2HWk \frac{HW}{s^2} C = \\ &= 3HWC^2 + C \left(2S^4 + \frac{k(HW)^2}{S^2} + \frac{k(HW)^2}{S^2} \right) \geq \\ &= 3HWC^2 + 3C \left(2S^4 \frac{k(HW)^2}{S^2} \frac{k(HW)^2}{S^2} \right)^{\frac{1}{3}} = \end{aligned}$$

$$3HWC^2 + 3Ck^{\frac{2}{3}}(2HW)^{\frac{4}{3}}, \quad (6)$$

式中: H, W 为输入的特征图分辨率; C 为特征图通道数; k 为参与注意力计算的数目。等号仅在式(7)成立时实现

$$S = \left(\frac{k}{2} (HW)^2 \right)^{\frac{1}{6}}. \quad (7)$$

最终, BRA 实现了 $O(HW)^{\frac{4}{3}}$ 的计算复杂度。

1.2 解码器

对于编码器输出的多尺度特征图。在解码器阶段相应构建4级金字塔式结构。每一层级均由两个 3×3 核大小, 步长为1的卷积模块和上采样模块组成。这里的上采样模块, 本文主要参考了 SwinUNet 中的 Patch Expanding 模块, 对于前3层的上采样, 它的核心实现是先将特征图通道图扩展为之前的2倍, 然后经过矩阵重排的操作实现分辨率扩展和通道数减半, 在最后一个上采样块, 将倍率调整为4。并且将跳跃连接设置在 $1/16, 1/8, 1/4$ 分辨率处。最后通过 1×1 核大小的卷积层将特征通道数映射成类别数, 得到输出分割图。

2 实验

2.1 数据集

本文所使用的数据集为 CVC-ClinicDB, 2018 Data Science Bowl, ISIC-2018 数据集。CVC-ClinicDB 数据集是从结肠镜检查视频中提取的帧数据库, 主要用于息肉分割任务, 曾作为 MICCAI 2015 自动息肉分割的训练数据库。2018 Data Science Bowl 数据集主要用于自动化细胞核检测, 来源于2018数据科学碗竞赛中细胞显微镜图像。ISIC-2018 数据集来源于国际皮肤成像协作组在2018年发布的一个皮肤镜图像数据集, 曾作为2018皮肤病分类挑战赛训练数据库, 本文采用了 task1 皮肤病分割数据集。相关细节规模设置如表1所示。

表1 数据集规模设置
Tab.1 Dataset size setting

数据集	图片数	训练集	验证集	测试集
ISIC-2018	2 594	1 868	467	259
2018 Data Science Bowl	670	483	120	67
CVC-ClinicDB	612	441	110	61

2.2 实现细节

本文所有实验均在 pytorch1.10 和 python3.8 下完成, 所有网络遵循同样的实验配置, 所使用的显卡型号为 Nvidia RTX 3090, 显存大小为 24 GB。在训练过程中, 对3个数据集都统一裁剪到 224×224 的大小, batch size 设置为 16, 学习率设置为 $1e-4$, 使用交叉熵损失函数搭配 Adam 优化器进行优化, 以5折交叉验证的方式共训练 150 个 epoch。其中 BiUNet, TransUNet, LeViT-UNet^[21], HiFormer^[22] 还需要加载各自预训练权重。

2.3 实验结果及分析

2.3.1 评估指标

采用 F1 分数 (F1)、准确率 (Acc)、召回率 (Recall)、平均交并比 (mIoU)、精确率 (Precision) 5 个评价指标评估网络性能。在医学图像分割领域, 通常重点关注 mIoU 和 F1 值。相关公式为

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, \quad (8)$$

$$Precision = \frac{TP}{TP + FP}, \quad (9)$$

$$Recall = \frac{TP}{TP + FN}, \quad (10)$$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}, \quad (11)$$

$$mIoU = \frac{TP}{TP + FP + FN}, \quad (12)$$

式中: TP 代表实际为正样本, 预测也为正样本的样本数; TN 代表实际为负样本, 预测也为负样本的样本数; FP 代表实际为负样本, 预测为正样本的样本数; FN 代表实际为正样本, 预测为负样本的样本数。mIoU 值和 F1 值越大, 代表预测的图像与标签图像重合度越高, 准确率越高。对于计算开销的评估, 本文采用计算量 GFlops 和参数量 (Param) 进行衡量。

2.3.2 实验结果分析

表2展示了 BiUNet 与其他分割算法在 ISIC-2018 上的对比结果。其中 BiUNet-T 是指 BiFormer 块数组合为 $[N_1=2, N_2=2, N_3=8, N_4=2]$, 并且使用 BiFormer-T 的预训练权重。BiUNet-S 是指 BiFormer 块数组合为 $[N_1=4, N_2=4, N_3=18, N_4=4]$, 并且使用 BiFormer-S 的预训练权重。BiUNet-S 以 90.2% 的 mIoU 值和 94.6% 的 F1 分数值取得最佳的表现。BiUNet-S

超越了 TransUNet 约 3% 以及 LeViT-UNet 约 1.5% 的 mIoU 值。

表 2 在 ISIC-2018 数据集上的对比结果
Tab. 2 Comparative results on the ISIC-2018 dataset

模型	<i>Acc</i>	<i>Precision</i>	<i>Recall</i>	<i>mIoU</i>	F_1
Attention-UNet	0.954±0.044	0.927±0.123	0.915±0.130	0.855±0.134	0.914±0.113
UNet	0.948±0.063	0.943±0.112	0.890±0.120	0.844±0.131	0.908±0.104
TransUNet	0.956±0.047	0.918±0.108	0.950±0.073	0.872±0.112	0.927±0.086
HiFormer	0.958±0.033	0.951±0.051	0.924±0.102	0.879±0.093	0.932±0.062
LeViT-UNet	0.959±0.030	0.949±0.063	0.935±0.082	0.887±0.087	0.938±0.055
DCSA-UNet	0.954±0.049	0.923±0.108	0.923±0.121	0.855±0.128	0.915±0.098
BiUNet-T	0.953±0.047	0.945±0.107	0.891±0.127	0.846±0.133	0.909±0.104
BiUNet-S	0.962±0.027	0.958±0.048	0.941±0.076	0.902±0.076	0.946±0.047

表 3 是 BiUNet 与其他分割算法在 CVC-ClinicDB 上的对比结果。BiUNet-S 也以 93.7% 的 mIoU 值和 96.7% 的 F_1 分数值取得最佳的分割精度。BiUNet-S 超越了 HiFormer 2.6% 以及 DCSA-UNet 约 3.7% 的 mIoU 值。

表 3 在 CVC-ClinicDB 数据集上的对比结果
Tab. 3 Comparative results on the CVC-ClinicDB dataset

模型	<i>Acc</i>	<i>Precision</i>	<i>Recall</i>	<i>mIoU</i>	F_1
Attention-UNet	0.988±0.012	0.916±0.176	0.899±0.186	0.855±0.182	0.904±0.177
UNet	0.989±0.012	0.922±0.177	0.909±0.178	0.870±0.177	0.914±0.175
TransUNet	0.990±0.011	0.961±0.042	0.940±0.058	0.905±0.065	0.949±0.038
HiFormer	0.991±0.121	0.945±0.127	0.945±0.126	0.911±0.127	0.945±0.125
LeViT-UNet	0.989±0.011	0.919±0.138	0.940±0.128	0.881±0.136	0.927±0.128
DCSA-UNet	0.991±0.011	0.951±0.125	0.929±0.133	0.900±0.128	0.939±0.126
BiUNet-T	0.990±0.012	0.955±0.042	0.939±0.073	0.900±0.081	0.945±0.050
BiUNet-S	0.993±0.010	0.969±0.029	0.967±0.031	0.937±0.040	0.967±0.022

表 4 在 2018 Data Science Bowl 数据集上的对比结果
Tab. 4 Comparative results on the 2018 Data Science Bowl dataset

模型	<i>Acc</i>	<i>Precision</i>	<i>Recall</i>	<i>mIoU</i>	F_1
Attention-UNet	0.967±0.030	0.882±0.070	0.906±0.088	0.807±0.095	0.890±0.060
UNet	0.969±0.029	0.910±0.064	0.887±0.082	0.816±0.096	0.896±0.060
TransUNet	0.977±0.017	0.909±0.062	0.927±0.048	0.850±0.077	0.917±0.047
HiFormer	0.977±0.022	0.884±0.136	0.913±0.149	0.831±0.149	0.896±0.139
LeViT-UNet	0.962±0.034	0.859±0.096	0.901±0.100	0.787±0.113	0.875±0.080
DCSA-UNet	0.971±0.025	0.897±0.066	0.910±0.069	0.825±0.086	0.901±0.054
BiUNet-T	0.975±0.018	0.908±0.066	0.916±0.053	0.840±0.083	0.911±0.052
BiUNet-S	0.979±0.015	0.922±0.056	0.919±0.072	0.856±0.091	0.919±0.059

综合表 2、表 3、表 4 的结果，除上述表述外，事实上，BiUNet-S 在 3 个数据集上 5 项指标均达到了最佳的表现，充分说明了 BiUNet 优越的性能。部分主流分割算法的对比可视化结果如图 3 所示。图 3 中，在 CVC-ClinicDB 数据集上，BiUNet-S 的优势相对突出，对息肉缺口部分信息预测准确率明显优于 U-Net，LeViT-UNet 以及 HiFormer。在 2018 Data Science Bowl 数据集上，BiUNet-S 对于细胞核边界信息的预测精度也明显优于 Attention-UNet 和 LeViT-UNet。

表 5 是各个模型计算开销的统计结果。其中，BiUNet-T 以 3.44G 的 Flops 取得了最小的计算复杂

度。而 BiUNet-S 以 5.55 G 的 Flops 和 28.10 M 的参数数量在 ViT 结构 (HiFormer, LeViT-UNet, TransUNet) 里取得了最小的计算开销。

表 4 是 BiUNet 与其他分割算法在 2018 Data Science Bowl 上的对比结果。BiUNet-S 依旧以 85.6% 的 mIoU 值以及 91.9% 的 F_1 分数值取得了最高的分割成绩。BiUNet-S 超越了 TransUNet 约 0.6% 以及 UNet 约 4% 的 mIoU 值。

表 5 各模型计算开销对比结果

Tab. 5 Comparison results of computational overhead of each model

模型	<i>GFlops/G</i>	<i>Param/M</i>
Attention-UNet	103.62	57.16
UNet	41.91	31.04
TransUNet	24.67	93.23
HiFormer	16.06	31.50
LeViT-UNet	20.36	29.84
DCSA-UNet	5.29	2.60
BiUNet-T	3.44	15.1
BiUNet-S	5.55	28.10

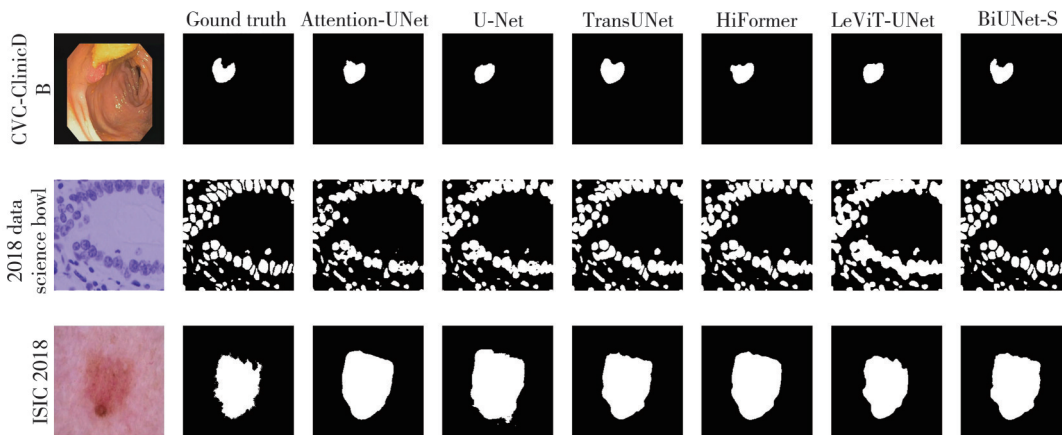


图 3 BiUNet-S与其他算法分割结果可视化对比图

Fig. 3 Visual comparison of BiUNet-S and other algorithms' segmentation results

3 消融实验

为了探索 BiUNet 网络性能细节, 本文在 3 个

数据集上设置了 2 组消融实验, 结果如表 6、表 7 所示。为简化数据集名称表述, 将 2018 Data Science Bowl 简记为 cell, ISIC-2018 简记为 skin, CVC-ClinicDB 简记为 cvc。

表 6 上采样层对 BiUNet 的影响结果对比

Tab. 6 Comparison of results of the impact of the upsampling layer on BiUNet

Dataset	模型	Acc	Precision	Recall	mIoU	F ₁
skin	CZ-BiUNet	0.958±0.033	0.948±0.065	0.927±0.093	0.878±0.093	0.932±0.059
	BiUNet-S	0.962±0.027	0.958±0.048	0.941±0.076	0.902±0.076	0.946±0.047
cvc	CZ-BiUNet	0.991±0.011	0.956±0.041	0.957±0.050	0.918±0.065	0.956±0.038
	BiUNet-S	0.993±0.010	0.969±0.029	0.967±0.031	0.937±0.040	0.967±0.022
cell	CZ-BiUNet	0.976±0.018	0.900±0.078	0.923±0.054	0.840±0.093	0.910±0.063
	BiUNet-S	0.979±0.015	0.922±0.056	0.919±0.072	0.856±0.091	0.919±0.059

CZ-BiUNet 是指将 BiUNet-S 中上采样层由 Patch Expanding 替换为双线性插值, 表 6 结果表

明 Patch Expanding 层作为上采样能够取得更好的分割效果。

表 7 损失函数对 BiUNet 的优化影响结果对比

Tab. 7 Comparison of the results of the impact of the loss function on the optimization of BiUNet

Dataset	模型	Acc	Precision	Recall	mIoU	F ₁
skin	DEBiUNet	0.956±0.050	0.948±0.062	0.923±0.109	0.874±0.105	0.928±0.080
	BiUNet-S	0.962±0.027	0.958±0.048	0.941±0.076	0.902±0.076	0.946±0.047
cvc	DEBiUNet	0.991±0.012	0.963±0.037	0.948±0.051	0.916±0.061	0.955±0.036
	BiUNet-S	0.993±0.010	0.969±0.029	0.967±0.031	0.937±0.040	0.967±0.022
cell	DEBiUNet	0.975±0.018	0.892±0.072	0.935±0.041	0.842±0.082	0.912±0.052
	BiUNet-S	0.979±0.015	0.922±0.056	0.919±0.072	0.856±0.091	0.919±0.059

DE-BiUNet 是指使用 Dice Loss 损失函数替代交叉熵损失函数对 BiUNet 进行优化, 在 skin, cvc, cell 3 个数据集上的实验结果表明, 交叉熵损失函数更加契合 BiUNet-S。

骨干结构的 U 型医学图像分割网络。充分利用双层路由稀疏注意力机制, 不仅有效地减小了计算代价, 还更高效地提升了分割性能。但是在实验中未能以更小的计算代价超越快速 CNN 模型 DCSAU-Net 的性能, 所以在未来的研究中, 将继续探索更加高效的轻量化视觉 Transformer 分割模型, 实现以更小的计算代价获取更好的分割能力。

4 结 语

为了提升现有以 Transformer 为主干特征提取网络的分割算法的精度和解决 Transformer 结构计算开销大的问题, 本文提出了一个以 BiFormer 为

参考文献:

[1] MA X, NIU Y, GU L, et al. Understanding adversarial

- attacks on deep learning based medical image analysis systems[J]. *Pattern Recognition*, 2021, 110: 107332.
- [2] NARAYAN V, AWASTHI S, FATIMA N, et al. Deep learning approaches for human gait recognition: a review [C]//2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023: 763-768.
- [3] HE S, MINN K T, SOLNICA-KREZEL L, et al. Deeply-supervised density regression for automatic cell counting in microscopy images [J]. *Medical Image Analysis*, 2021, 68: 101892.
- [4] ZHOU S, NIE D, ADELI E, et al. High-resolution encoder-decoder networks for low-contrast medical image segmentation [J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2019. doi: 10.119/TIP.2019.2919937.
- [5] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]//NAVAB N, HORNEGGER J, WELLS W M, et al. *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2015: 234-241.
- [6] DOSOVITSKIV A, BEYER L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale [EB/OL]. *arXiv*: 2010.11929, 2020.
- [7] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHS N, et al. U-net++: A nested U-net architecture for medical image segmentation [C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer International Publishing, 2018: 3-11.
- [8] HUANG H, LIN L, TONG R, 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), 2020: 1055-1059.
- [9] WU H, WANG J, WANG X, et al. Attention augmented ConvNeXt UNet for rectal tumour segmentation [EB/OL]. <http://arxiv.org/abs/2210.00222v2>.
- [10] XU Q, MA Z, HE N, et al. DCSAU-Net: a deeper and more compact split-attention U-Net for medical image segmentation [J]. *Computers in Biology and Medicine*, 2023, 154: 106626.
- [11] 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.
- [12] CHEN J, LU Y, YU Q, et al. TransUNet: transformers make strong encoders for medical image segmentation [EB/OL]. <http://arxiv.org/abs/2102.04306v1>.
- [13] CAO H, WANG Y, CHEN J, et al. Swin-UNet: Unet-like pure transformer for medical image segmentation [C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2023: 205-218.
- [14] ZHANG Y, LIU H, HU Q. TransFuse: fusing transformers and CNNs for medical image segmentation [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 14-24.
- [15] XIE Y, ZHANG J, SHEN C, et al. CoTr: efficiently bridging CNN and transformer for 3D medical image segmentation [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 171-180.
- [16] LIN A, CHEN B, XU J, et al. DS-TransUNet: dual swin transformer U-net for medical image segmentation [J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 4005615.
- [17] YAO C, HU M, ZHAI G, et al. TransClaw U-net: claw U-net with transformers for medical image segmentation [EB/OL]. <http://arxiv.org/abs/2017.05188v1>.
- [18] GRAHAM B, EL-NOUBY A, TOUVRON H, et al. LeViT: a vision transformer in ConvNet's clothing for faster inference [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 12239-12249.
- [19] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 19992-10002.
- [20] ZHU L, WANG X, KE Z, et al. BiFormer: vision transformer with bi-level routing attention [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 10323-10333.
- [21] XU G, WU X, ZHANG X, et al. LeViT-UNet: make faster encoders with transformer for medical image segmentation [EB/OL]. <http://arxiv.org/abs/2019.08623v1>.
- [22] HEIDARI M, KAZEROUNI A, SOLTANY M, et al. HiFormer: hierarchical multi-scale representations using transformers for medical image segmentation [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023: 6191-6201.