

文章编号 1004-924X(2024)04-0609-13

融入交叉注意力编码的皮肤病变分割网络

李大湘, 杨福杰*, 刘颖, 唐 垚

(西安邮电大学 通信与信息工程学院, 陕西 西安 710121)

摘要: 由于卷积操作的局限性, 现有的皮肤病变图像分割网络无法对图像中的全局上下文信息建模, 导致其无法有效捕获图像的目标结构信息, 本文设计了一个融入交叉自注意力编码的 U 型混合网络, 用于皮肤病变图像分割。首先, 将设计的多头门控位置交叉自注意力编码器引入到 U 型网络的最后两个层级中, 使其能够在图像中学习语义信息的长期依赖关系, 弥补卷积操作全局建模能力的不足; 其次, 在跳跃连接部分引入一个新的位置通道注意力机制, 用于编码融合特征的通道信息并保留位置信息, 提高网络捕获目标结构的能力; 最后, 设计一个正则化 Dice 损失函数, 使网络能够在假阳性和假阴性之间权衡, 提高网络的分割结果。基于 ISBI2017 和 ISIC2018 数据集的对比实验结果表明, 本文网络的 Dice 分别为 91.48% 和 91.30%, IoU 分别为 84.42% 和 84.12%, 分割精度在整体上优于其他网络, 且具有较低的参数量和计算复杂度, 即本文网络能够高效地分割皮肤病变图像的目标区域, 可为皮肤疾病辅助诊断提供帮助。

关键词: 医学图像分割; 皮肤病变; 交叉自注意力编码; 位置通道注意力

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.37188/OPE.20243204.0609

Skin lesion segmentation network with cross-attention coding

LI Daxiang, YANG Fujie*, LIU Ying, TANG Yao

(School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

* Corresponding author, E-mail: 729378215@qq.com

Abstract: Owing to the limitations of convolutional operations, existing skin lesion image segmentation networks are unable to model the global contextual information in images, resulting in their inability to effectively capture the target structural information of images. In this paper, a U-shaped hybrid network with cross-self-attention coding was designed for skin lesion image segmentation. Firstly, the designed multi-head gated position cross self-attention encoder was introduced in the last two layers of the U-shaped network to enable it to learn the long-term dependencies of semantic information in images and to compensate for the lack of global modelling capability of the convolutional operation; Secondly, a novel position channel attention mechanism was implemented in the skip connection part to encode the channel information of the fused features and retain the positional information to improve the network's ability to capture the target structure; finally, a regularised dice loss function was designed to enable the network to trade off between false positives and false negatives to improve the network's segmentation results. Experimental results on ISBI2017 and ISIC2018 datasets show that the network presented in this paper achieves Dice

收稿日期: 2023-08-24; 修订日期: 2023-10-10.

基金项目: 国家自然科学基金项目 (No. 62071379); 陕西省自然科学基金项目 (No. 2019JM-604); 西安邮电大学创新基金 (No. CXJJYL2022002)

score of 91.48% and 91.30%, and IoU of 84.42% and 84.12%, respectively. The network outperforms other networks in terms of segmentation accuracy with fewer parameters and lower computational complexity. Therefore, it can efficiently segment the target region of skin lesion images and aid in the adjunctive diagnosis of skin diseases.

Key words: medical image segmentation; skin lesion; cross-self-attention coding; position channel attention

1 引 言

皮肤癌作为美国乃至全世界最常见的癌症,其发病率正在逐年快速上升,给全球健康带来了巨大的威胁与挑战^[1]。临床上皮肤癌主要分为恶性黑色素瘤(Malignant Skin Melanoma, MSM)和非黑色素瘤皮肤癌(Non Melanoma Skin Cancer, NMSC),其中90%以上的皮肤癌死亡病例均由MSM导致。临床上,医生通常采用“ABCDE”法则来辨别MSM,其中A,B,C,D和E分别表示痣的形状不对称、边缘不规则、颜色不均匀、直径改变、大小改变以及进展性改变^[2]。通过早期发现,MSM的5年生存率可达99%,而延迟诊断导致生存率下降至23%^[3]。因此,运用图像分割技术对每张皮肤病变图像进行目标区域的分割,具有较高的临床意义。

目前,皮肤病变图像分割方法可分为传统图像分割方法和基于深度学习的分割方法。其中,传统分割方法包括阈值处理、活动轮廓模型和监督学习方法^[4],但由于皮肤病变图像中的病变区域与周围区域对比度低,且病变的位置、形状、颜色和边界各不相同,传统方法难以准确的分割皮肤病变区域。随着深度学习的发展,卷积神经网络(Convolutional Neural Networks, CNN)已广泛应用于各类分割任务,其中基于全卷积神经网络(Fully Convolutional Network, FCN)^[5]和U-Net^[6]的一系列扩展已取得良好的分割性能。例如:SpineParseNet^[7]将用于三维粗分割的图卷积分割网络与用于二维分割细化的残差U型网络组合,实现了对体积MR图像的脊椎解析。UNet++^[8]设计一系列嵌套且密集的跳跃连接,减少了编码器和解码器之间的语义差距;Attention U-Net^[9]提出一种新型的注意力门控机制,并将其与U-Net相结合,使网络能够自动学习并关注

具有不同形状和大小的目标结构,同时抑制不相关的特征响应。该网络在腹部CT的多标签分割问题上取得了良好的分割效果;CA-Net^[10]提出了一个全面的基于注意力的CNN网络,即通过向U-Net网络中引入多个注意力模块,同时了解最重要的空间位置、通道和尺度信息,实现准确且具有可解释性的医学图像分割网络。虽然上述基于CNN的方法能有效捕捉局部细节信息,但由于卷积结构存在局部感受野,该类网络通常无法对皮肤损伤图像中的目标形状和边界信息进行建模^[11]。为了解决这一问题,文献[12]提出了一种基于视觉Transformer(Vision Transformer, ViT)的网络,其通过自注意力机制捕获长期依赖关系并编码形状表示,解决了卷积结构缺乏对图像中全局上下文理解的问题。受此启发,许多工作将Transformer引入到医学图像分割任务中。例如:TransUNet^[13]首先在编码部分使用CNN提取底层特征图谱,然后通过Transformer对全局上下文信息进行建模,在解码部分对编码后的特征进行上采样,并采用跳跃连接技术,在CT多器官分割任务中创造了新的记录。TransFuse^[14]通过并行的方式将CNN与Transformer结合,使网络以更浅的方式有效地捕捉全局依赖性和低级空间细节,提高了网络建模全局上下文的效率。但由于自注意力机制的计算复杂度随图像大小的平方倍增加,导致大多数基于Transformer的网络具有较高的计算复杂度和参数量,且其将输入视为1维序列,仅关注于全局上下文建模,导致低分辨率特征图谱缺乏详细的定位信息。而这种信息并不能通过直接上采样进行有效地恢复,导致了粗略的分割结果。

针对上述网络存在的问题,本文提出一种融入交叉注意力编码的U型混合网络(U-shaped hybrid network, UH-Net),用于皮肤病变图像分

割。首先,本文构造了一个高效的多头门控位置交叉自注意力(Multi-head Gated Position Cross Self-Attention, MhGPCSA)编码器,并将其引入到U型网络的最后两个层级中,用于建模全局上下文信息,使得网络能够捕获目标区域的形状和边界信息;其次,在跳跃连接中引入一个新的位置通道注意力(Position Channel Attention, PosCA)机制,以使网络在建模通道关系的同时保留精确的位置信息,提高其捕获病变区域的空间结构能力;最后,针对Dice损失无法平衡假阳性(False Positive, FP)与假阴性(False Negative, FN)的问题,设计了一个正则化Dice损失函数,使网络在推理过程中同时关注于输入图像的前景与背景区域,提高网络的分割性能。

2 网络构造

2.1 网络整体架构设计

针对皮肤病变图像中存在目标区域形状变化大,边界不明显的问题,本文提出一种融入交叉注意力编码的皮肤病变图像分割网络(UH-Net),其网络结构如图1所示,共设计了五层U型混合网络,其中前三层级采用的是卷积操作,以捕获浅层特征而使网络前端更注重图像的局部细节信息,而在网络的最后两个层级采用的是MhGPCSA编码器,以弥补卷积操作建模全局信息能力的不足;在跳跃连接中,向融合后的特征图谱中引入PosCA模块,用于编码通道信息,并将结果输入到解码器中。最后通过正则化Dice损失函数实现皮肤病变图像中目标区域的分割。

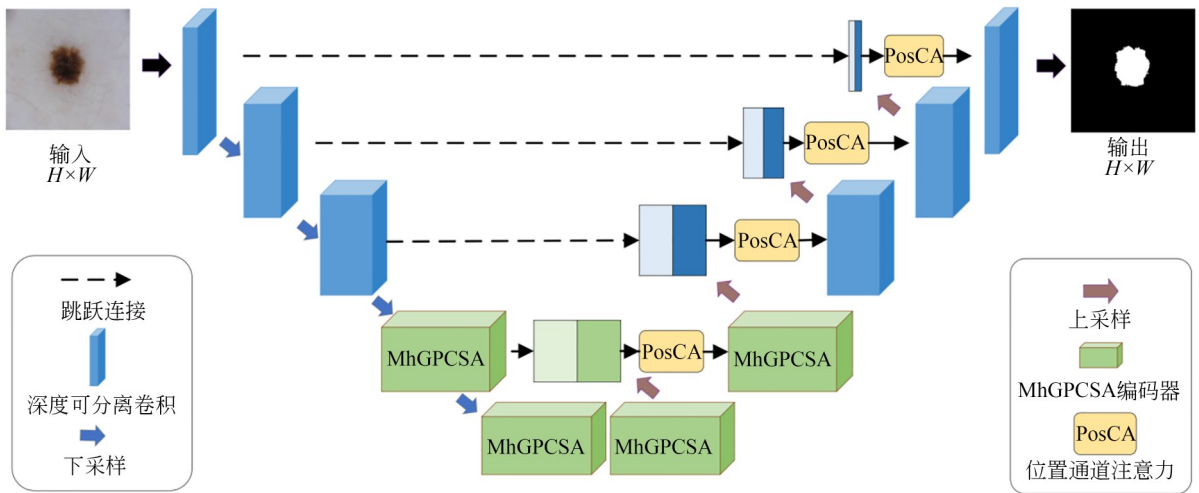


图1 UH-Net网络整体架构设计示意图

Fig. 1 Schematic of overall architecture design of UH-Net network

2.2 MhGPCSA 编码器

为了在图像中学习语义信息的长期依赖关系,本节设计了一个新的门控位置交叉自注意力(Gated Position Cross Self-Attention, GPCSA)机制,且基于它构造了MhGPCSA编码器,并将其融入UH-Net的最后两个层级,使网络能够高效地捕获图像的全局上下文信息,以缓解卷积操作全局建模能力的不足,提高其分割性能。相关原理如下:

2.2.1 GPCSA 机制

给定任意输入特征图谱 $F \in \mathbb{R}^{C \times H \times W}$, 其中 C, H 和 W 分别表示 F 的通道数、高度和宽度。

首先,使用三个 1×1 卷积来投影特征图谱 F , 分别生成三个新的特征图谱 Q, K 和 V , 即:

$$\begin{cases} Q = f_0^{1 \times 1}(F) \\ K = f_1^{1 \times 1}(F), \\ V = f_2^{1 \times 1}(F) \end{cases} \quad (1)$$

其中: $f^{1 \times 1}$ 表示 1×1 卷积, 设 $f^{1 \times 1}$ 卷积核个数均为 C' , 则 Q, K 与 $V \in \mathbb{R}^{C' \times H \times W}$, 即它们通道数均为 C' ; 然后, 在 Q 中的任意空间位置 $i (i = 1, 2, \dots, H \times W)$ 上沿通道维度抽取特征向量 $q_i \in \mathbb{R}^{C' \times 1}$, 并在 K 中抽取与位置 i 处于同一行或同一列(即十字形邻域)的所有特征向量, 获得特

征向量集合记为 $K_i \in R^{C' \times (H+W-1)}$, 同理在 V 中抽取与位置 i 处于同一行或同一列的特征向量, 获得特征向量集合 $V_i \in R^{C' \times (H+W-1)}$, 则可得到 $q_i \in R^{C' \times 1}$ 的交叉注意力输出 $z_i \in R^{C' \times 1}$, 即:

$$z_i = \sum_{u=1}^{H+W-1} \text{SoftMax}(q_i^T K_{i,u}) V_{i,u}, \quad (2)$$

其中: $K_{i,u} \in R^{C' \times 1}$ 与 $V_{i,u} \in R^{C' \times 1}$ 分别表示 K_i 与 V_i 中的第 u 个列向量, q_i^T 表示 q_i 的转置。

虽然上述交叉注意力机制可使网络在图像中学习语义信息的长期依赖关系, 但由于该操作

$$z_i = \sum_{u=1}^{H+W-1} \left(\text{SoftMax}(q_i^T K_{i,u} + G_i^q(q_i^T r_i^q) + G_{i,u}^k(K_{i,u}^T r_{i,u}^k)) \right) V_{i,u}, \quad (3)$$

其中: $r_i^q \in R^{C' \times 1}$, $r_i^k \in R^{C' \times (H+W-1)}$ 表示可学习的相对位置编码, $G_i^q \in R$ 与 $G_{i,u}^k \in R^{H+W-1}$ 表示可学习的门控参数, 用于控制相对位置编码项对自注意力权值的影响, 即较之未准确学习的相对位置编码项, 准确学习的位置编码项将被门控机制赋予更大的权值。

最后, 对于 Q 中的所有空间位置 i 沿通道抽取的 $q_i \in R^{C' \times 1}$, 按式(3)进行 GPCSA 计算重组, 则输入的特征图谱 F 被重构为一个新的特征图谱 Z , 该过程记为:

$$Z = \text{GPCSA}(F), \quad (4)$$

其中, Z 表示 GPCSA 重构的结果。

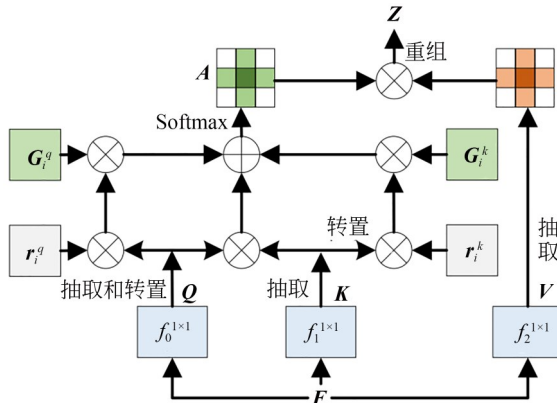


图2 门控位置交叉注意力(GPCSA)机制示意图

Fig.2 Schematic of Gated position cross self-attention mechanism

2.2.2 MhGPCSA 编码器

由于单个 GPCSA 机制只能捕获每个像素对应的交叉路径的上下文信息, 导致其无法与周围

专注于建模全局上下文信息, 导致低分辨率特征图谱缺乏详细的位置信息, 而在图像分割任务中, 不同目标的相对位置关系通常利于捕捉目标结构。因此, 为了提高网络捕获目标结构的能力, 受文献[15]的启发, 设计了一个新的 GPCSA 机制(如图2所示), 即在计算注意力权值时, 添加相对位置编码项与门控项, 以使其对位置信息更加敏感, 进而提高网络的分割精度, 则公式(2)可改写为:

不在交叉路径中的像素建立联系, 为了解决这一问题, 可通过连续使用两层 GPCSA 机制来对特征图谱的全局上下文信息建模, 并在 GPCSA 的基础上设计 MhGPCSA 编码器, 以使网络能够关注来自同一位置的不同通道间的子空间信息。图3为设计的 MhGPCSA 编码器结构示意图, 对于输入的特征图谱 $F \in R^{C \times H \times W}$, 对其编码过程可描述为:

$$\begin{cases} Z' = \text{MhGPCSA}(\text{BN}(F)) \\ Z'' = F + \text{MhGPCSA}(Z') \\ \bar{F} = Z'' + \text{Conv}(\text{ReLU}(\text{BN}(Z''))) \end{cases}, \quad (5)$$

其中: $\text{BN}()$, $\text{ReLU}()$ 与 $\text{MhGPCSA}()$ 分别表示层归一化、ReLU 激活函数与 MhGPCSA 操作, $\text{Conv}()$ 表示 C 个不同 1×1 核的卷积操作, $\bar{F} \in R^{C \times H \times W}$ 表示编码器的最终输出。MhGPC-

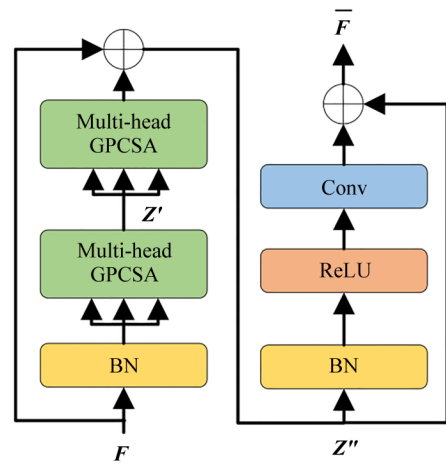


图3 多头门控位置交叉自注意力编码器

Fig.3 Multi-head gated position cross self-attention encoder

SA是GPCSA的扩展,即并行地运行8个GPCSA操作,每个注意力头将从不同的视角捕获特征图谱 F 的长期依赖关系而挖掘其全局上下文信息,并将它们的输出串联起来作为最终重构结果,即:

$$Z' = \text{MhGPCSA}(F) =$$

$$\text{Concat}(\text{GPCSA}_1(F), \dots, \text{GPCSA}_8(F)), \quad (6)$$

其中: $\text{Concat}[\cdot]$ 表示将每个注意力头的输出结果沿通道维度进行拼接。为了使 F 经MhGPCSA重构之后,其输出 Z' 与输入 F 保持相同的维度,式(3)所示的每个GPCSA头输出的维度 C' 设置为输入元素维度的8分之一,即 $C' = C/8$ 。

2.3 PosCA 机制

在SE-Net^[16]等传统的通道注意力(Channel

$$\begin{cases} F_X^{\text{avg}} = \left[\frac{1}{W} \sum_{j=1}^W F(c, h, j) \mid c=1, \dots, C, h=1, \dots, H \right] \\ F_Y^{\text{avg}} = \left[\frac{1}{H} \sum_{i=1}^H F(c, i, w) \mid c=1, \dots, C, w=1, \dots, W \right] \end{cases}, \quad (8)$$

$$\begin{cases} F_X^{\text{max}} = \left[\max_{j \in (1 \dots W)} (F(c, h, j)) \mid c=1, \dots, C, h=1, \dots, H \right] \\ F_Y^{\text{max}} = \left[\max_{i \in (1 \dots H)} (F(c, i, w)) \mid c=1, \dots, C, w=1, \dots, W \right] \end{cases}, \quad (9)$$

其中: $F_X^{\text{avg}}, F_X^{\text{max}} \in R^{C \times H \times 1}$ 表示沿宽度方向的一维全局平均与最大池化结果, $F_Y^{\text{avg}}, F_Y^{\text{max}} \in R^{C \times 1 \times W}$ 表示沿高度方向的一维池化结果。然后,基于上述四组一维池化结果,本节设计了一个PosCA机制,具体原理如图4所示。

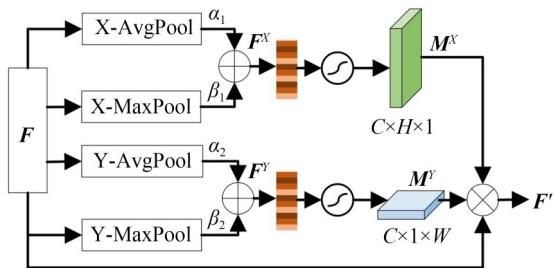


图4 位置通道注意力机制

Fig. 4 Position channel attention mechanism

对于输入特征图谱 F ,先用式(8)和式(9)所示一维池化分别沿 F 的宽度轴和高度轴对其进行信息聚合,生成四种上下文描述符 $F_X^{\text{avg}}, F_X^{\text{max}}, F_Y^{\text{avg}}$ 和 F_Y^{max} ;然后,由于平均和最大池化在特征聚

Attention, CA)机制中,对于输入特征图谱 $F \in R^{C \times H \times W}$ (其中 C, H 与 W 分别表示其通道数、高度与宽度),通常采用“二维全局平均池化”将其综合为 $C \times 1 \times 1$ 的特征描述符,即:

$$F_{\text{GAP}} = \left[\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F(c, i, j) \mid c=1 \dots C \right], \quad (7)$$

其中: $F_{\text{GAP}} \in R^{C \times 1 \times 1}$ 表示输入经过池化后的结果。显然,式(7)所示的二维池化(即同时沿高与宽二个维度进行池化)将每个通道的 C 个数压缩为一个数,导致生成的CA权值不带有位置信息,若直接将其用在分割网络中,则不利于像素的精确定位而影响分割精度。所以,本节分别沿高与宽维度定义了“一维全局平均与最大池化”用于建模通道关系且兼顾空间位置信息,即:

合时发挥的作用存在差异,则不对它们进行简单的相加融合。因此,引入两组自适应机制,分别对高、宽度方向的池化结果进行自适应融合,二个不同方向的融合图谱记为 $F^X \in R^{C \times H \times 1}$ 与 $F^Y \in R^{C \times 1 \times W}$;最后,在通道信息交互时,针对SENet采用全连接层会存在通道降维而带来信息丢失的问题,PosCA机制使用快速一维卷积实施通道交互,再结合Sigmoid激活函数得到二个方向的注意力权重 $M^X \in R^{C \times H \times 1}$ 与 $M^Y \in R^{C \times 1 \times W}$,即:

$$M^X = \sigma \left(\text{C1D}_{k \times 1 \times 1} (\alpha_1 \otimes F_X^{\text{avg}} + \beta_1 \otimes F_X^{\text{max}}) \right), \quad (10)$$

$$M^Y = \sigma \left(\text{C1D}_{k \times 1 \times 1} (\alpha_2 \otimes F_Y^{\text{avg}} + \beta_2 \otimes F_Y^{\text{max}}) \right), \quad (11)$$

其中: $\alpha_1, \beta_1, \alpha_2$ 和 β_2 表示自适应机制加权系数,作为可学习参数在网络训练过程中通过随机梯度下降法进行训练, $\otimes$ 表示广播元素乘法, σ 表示Sigmoid函数。 $\text{C1D}_{k \times 1 \times 1}$ 表示核大小为 k 的一维卷积, k 值使用ECA^[17]中的经验公式计

算,即:

$$k = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}}, \quad (12)$$

其中: $\lfloor t \rfloor_{\text{odd}}$ 表示离 t 最近的奇数, C 为通道数。 γ 和 b 均为超参数, 本文在所有的实验中将其分别设置为 2 和 1, 因此一维卷积核大小由通道数 C 自适应地确定; 最后, 将输入特征图谱 F 与 M^x , M^y 进行广播乘法, 得到 PosCA 模块的输出。其表达式为:

$$F' = (F \otimes M^y) \otimes M^x, \quad (13)$$

其中, $\otimes$ 表示广播元素乘法。由式(13)可知, PosCA 机制可给 F 中不同空间位置处的数据赋予不同的注意力, 即可有差别地为图像中不同目标给予相应的注意力而提高其定位能力, 从而提高整个网络分割精度。

2.4 损失函数及正则化

在医学图像分割任务中, 由于 FN 较之 FP 将付出更大的代价, 则在网络训练过程中, 希望所有“前景(病变区域)”都能被分割出来, 所以传统的分割网络采用的损失函数均为 Dice 损失函数^[18], 其表达式为:

$$L_{\text{Dice}}(p_i, \hat{p}_i) = 1 - \frac{2 \sum_{i=1}^N p_i \hat{p}_i}{\sum_{i=1}^N p_i + \sum_{i=1}^N \hat{p}_i}, \quad (14)$$

其中, $\hat{p}$ 和 p 分别表示预测结果和 Ground-Truth, N 为像素总数。由式(14)可知, 但当 $p_i = 0$ 时, 无论 $\hat{p}_i$ 取何值, $p_i \hat{p}_i$ 恒等于 0, 即 Dice 损失只关注于对前景区域的挖掘, 虽降低了训练过程中 FN, 但忽略了背景区域, 将会导致 FP 过高, 即更多的背景区域被分割为前景。所以, 本节对 Dice 损失函数进行了改进, 设计一个“正则化 Dice 损失函数”, 即:

$$L = L_{\text{Dice}}(p_i, \hat{p}_i) - \frac{\beta}{N} \sum_{i=1}^N (1 - p_i) \ln(1 - \hat{p}_i), \quad (15)$$

其中, β 为可学习的正则化参数, 用于控制正则化项在网络训练过程中所占比重, 其随网络的不断迭代自动地调整优化。分析式(15)所示的损失函数可知, 由于 $(1 - p_i)$ 的制约, 若 $p_i = 1$ (即前景)则损失为 0, 所以正则化只关注背景区域被误分所造成的损失, 即网络在训练时, 若更多的背

景被分割为前景(FP 过高)时, 则更多的 $\hat{p}_i$ 趋向于 1, 导致正则化项中的 $\ln(1 - \hat{p}_i)$ 将更趋向于负无穷大, 再与前面的“负号”结合, 将导致损失变大, 损失在梯度下降反向传播时, 就会一定程度上约束更多的 $\hat{p}_i$ 趋向于 1 的情况, 降低 FP。这样一来, 本文网络在训练过程中, 经过不停地优化迭代, 就会在 FP 与 FN 之间进行权衡, 从而实现更好的分割结果。

3 实验结果及分析

3.1 实验数据及参数设置

本文的实验数据为 ISBI2017 数据集^[19] 和 ISIC2018 数据集^[3,20], 均为面向黑色素瘤检测的皮肤病变分析挑战赛使用的公共数据集, 其中 ISBI2017 数据集包含 2150 张 RGB 图片, ISIC2018 数据集包含 2 594 幅 RGB 图片, 均采用二进制标注。由于数据集中图像的分辨率各不相同, 且样本数量较少, 为了增加样本的多样性而缓解网络训练时的过拟合问题^[21], 采用以下方法对数据集中的每幅图像进行扩充: 首先, 在角度区间 $[0, 360)$ 之内以 45 度为步长, 对原始图像进行 8 种不同方向的旋转; 然后, 对旋转后的 8 幅图像分别进行高斯模糊与 RGB 平移处理, 此时单幅图像可以扩充为 24 幅不同的图片 (即旋转的 8 幅、旋转的 8 幅经高斯模糊后的 8 幅以及旋转的 8 幅经过 RGB 平移后的 8 幅); 最后, 为了避免直接缩放而破坏原始图像的形状和纹理等信息, 对增扩后的每幅图像进行中心裁剪, 并将裁剪后的图像缩放为 256×256 大小作为最终的样本图像。这样以来, ISBI2017 和 ISIC2018 数据集经扩充之后样本数分别为 51 600 和 62 256 幅不同的图片, 然后再分别按 8:1:1 比例随机划分为训练集、验证集和测试集, 用于网络的训练与测试。

实验中网络的具体参数: 输入图像大小设置为 256×256 , 批量大小为 16, 初始学习率为 10^{-3} , 网络优化器为自适应矩估计 (Adaptive Moment estimation, Adam), 一阶矩估计指数衰减率设置为 0.9, 权重衰减系数为 10^{-4} , 训练周期为 200。在训练过程中使用余弦退火策略对学习率进行

调整,并将最小学习率设置为 10^{-5} 。实验的硬件环境为Intel(R) Xeon(R) Platinum 8358P CPU,工作频率为2.60 GHz,内存为80 GB,NVIDIA GeForce RTX 3090 显卡。操作系统为Ubuntu 18.04,编程语言为Python 3.8,所有程序均在Pytorch框架下实现。

3.2 评价指标

本文使用交并比(Intersection over Union, IoU)、Dice系数(Dice)、Hausdorff_95(HD95)距离、准确率(Precision)和召回率(Recall)作为评价指标,用于衡量各种网络分割结果的好坏。其中,前4种评价指标的计算公式如式(16)~式(19)所示:

$$IoU(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}, \quad (16)$$

$$Dice = \frac{2TP}{2TP + FP + FN}, \quad (17)$$

$$Precision = \frac{TP}{TP + FP}, \quad (18)$$

$$Recall = \frac{TP}{TP + FN}, \quad (19)$$

其中: X 为预测结果中皮肤病变区域的像素集合, Y 为原皮肤病变图片真值中病变区域的像素集合, TP 为分割结果中被正确分割的像素个数, FN 为分割结果中被错误预测为背景的皮肤病变区域像素个数, FP 为分割结果中被错误分割的前景像素个数。

Hausdorff(HD)距离计算公式为^[22]:

$$HD(\tilde{Y}, Y) = \max(h(\tilde{Y}, Y), h(Y, \tilde{Y})), \quad (20)$$

$$\begin{cases} h(\tilde{Y}, Y) = \max_{\tilde{y} \in \tilde{Y}} \{ \min_{y \in Y} \|\tilde{y} - y\| \} \\ h(Y, \tilde{Y}) = \max_{y \in Y} \{ \min_{\tilde{y} \in \tilde{Y}} \|y - \tilde{y}\| \} \end{cases}, \quad (21)$$

其中: $\tilde{Y}$ 为预测结果, Y 为GroundTruth, $\|\cdot\|$ 为点集 $\tilde{Y}$ 和点集 Y 之间的 L_1 距离范式。为了排除一些离群点造成的不合理距离,保持整体数值稳定性,本文选择从小到大排名为前95%的HD距离(即HD95)作为评价指标。

3.3 综合对比实验及分析

3.3.1 不同网络对比实验

为了进一步验证本文网络UH-Net的有效性,将其与近年来在医学图像分割领域的典型方法进行综合对比,对比网络选择了U-Net^[6],UNet++^[8],TransUNet^[13],TransFuse^[14],MedT^[23]以及轻型医学图像分割网络UNeXt^[24]。其中U-Net由编码器和解码器组成,在编码器的每个层级中使用最大池化来降低特征图谱大小并通道数翻倍,在解码器的每个层级中通过双线性插值来增大特征图谱减少通道数,恢复病变区域尺寸;UNet++通过使用一系列嵌套且密集的跳跃连接,降低了编码器和解码器之间的语义差距;TransUNet和TransFuse通过将Transformer引入到卷积网络中,提高了网络建模全局上下文信息的能力;MedT和UNeXt对注意力机制进行改进,提高了网络建模长期依赖关系的效率。在同样的实验环境下对上述六种网络进行实验,对比结果如表1、表2所示,其中测试最优性能指标加粗表示。

表1 ISBI2017数据集上不同网络结果

Tab. 1 Results of different networks on ISBI2017 dataset

网络	IoU	Dice	HD95	Precision	Recall	Params/M	GFLOPs/G
U-Net ^[6]	75.35±0.27	83.34±0.23	41.38±0.19	93.17±0.12	75.12±0.23	31.13	55.84
UNet++ ^[8]	76.65±0.16	84.06±0.52	39.63±0.31	93.50±0.45	76.35±0.14	9.16	34.65
TransUNet ^[13]	79.82±0.18	89.01±0.35	35.47±0.25	97.24±0.56	82.06±0.43	105.32	38.52
TransFuse ^[14]	80.11±0.42	89.56±0.14	34.68±0.22	98.13±0.15	82.36±0.14	26.27	11.53
MedT ^[23]	78.35±0.21	87.01±0.43	36.14±0.16	94.28±0.31	80.78±0.22	1.60	21.24
UNeXt ^[24]	82.30±0.23	90.21±0.54	33.79±0.27	96.59±0.47	84.62±0.74	1.47	0.57
UH-Net	84.42±0.11	91.48±0.32	32.67±0.19	97.37±0.21	86.26±0.15	0.78	1.44

表 2 ISIC2018数据集上不同网络结果

Tab. 2 Results of different networks on ISIC2018 dataset

网络	IoU	Dice	HD95	Precesion	Recall	Params/M	GFLOPs/G
U-Net ^[6]	74.55±0.96	84.03±0.87	41.15±0.43	82.78±0.31	85.32±0.12	31.13	55.84
UNet++ ^[8]	76.12±0.65	84.96±0.71	40.08±0.51	83.56±0.13	86.41±0.26	9.16	34.65
TransUNet ^[13]	80.51±0.72	88.91±0.63	35.21±0.47	90.01±0.36	87.84±0.23	105.32	38.52
TransFuse ^[14]	80.63±0.32	89.27±0.13	34.42±0.33	89.61±0.11	88.93±0.18	26.27	11.53
MedT ^[23]	79.54±0.26	87.35±0.28	35.79±0.17	87.83±0.32	86.88±0.43	1.60	21.24
UNeXt ^[24]	81.70±1.53	89.70±0.96	34.18±0.69	88.95±0.19	90.46±0.25	1.47	0.57
UH-Net	84.12±0.15	91.30±0.24	33.24±0.21	92.19±0.42	90.52±0.37	0.78	1.44

表 1 和表 2 给出了本文网络与其他六种网络在 ISBI2017 和 ISIC2018 数据集上的性能指标、参数量(Params)和计算复杂度(GFLOPs),综合对比本文 U 型混合网络(UH-Net)性能较优,且在计算复杂度和参数量方面具有明显的优势。与基于 Transformer 的网络相比,UH-Net 网络参数是 TransUNet 网络参数的 0.74%,计算复杂度仅是其 3.74%。可见对注意力机制进行改进,可有效地提高网络分割效率。与卷积网络 U-Net 相比,UH-Net 在分割性能、参数量和计算复杂度方面均具有明显的优势,表明 UH-Net 具有良好的分割结果。

在表 1 和表 2 中本文网络 UH-Net 召回率均最高,两个数据集上的召回率分别为 86.26% 和 90.52%,召回率表示正确分割为前景像素数量与金标准中的前景像素数量之比,表明 UH-Net 可有效地挖掘皮肤病变区域信息,降低欠分割率。IoU 和 Dice 可用于衡量真值与分割图像的相似程度,本文网络 UH-Net 的 IoU 和 Dice 均最高, IoU 达到 84.42% 和 84.12%, Dice 达到 91.48% 和 91.30%,表明 UH-Net 分割结果更接近于真实标签。HD95 常用于衡量边缘分割精度,本文网络 UH-Net 在两个数据集上的 HD95 均最低,分别为 32.67 和 33.24,表明本文网络 UH-Net 可以有效地分割病灶区域的边缘信息。因此,与其他网络相比,UH-Net 具有良好的分割结果和较低的欠分割。

图 5 给出了不同网络在 ISBI2017 和 ISIC2018 数据集上的分割结果,图 5 中(a)~图 5(d)来源于 ISBI2017 数据集,图 5 中(e)~图 5(h)

来源于 ISIC2018 数据集,从上到下依次为原图、真值(Ground Truth, GT),UNet, UNet++, TransUNet, TransFuse, MedT, UNeXt 和 UH-Net 分割结果。由图 5 可知,与其他网络的分割结果相比,本文网络 UH-Net 的分割结果更优。在图 5(f)中,由于皮肤病变区域与其他区域对比度较高,本文网络与其他网络均能有效地分割出目标区域。在图 5(h)中,皮肤病变区域边缘不明显,U-Net 和 UNet++ 作为卷积网络捕获目标结构的能力不足,UH-Net 在网络的最后两个层级中引入 MhGPCSA 编码器以及在跳跃连接中引入 PosCA 机制,提高了网络捕获目标结构的能力,使得病变区域的边缘更加清晰。在图 5(c)中,由于毛发干扰,皮肤病变区域不明显,导致网络在分割目标区域时容易过分割或欠分割,UH-Net 通过将正则化 Dice 损失替换传统的 Dice 损失,以使网络同时关注与病变区域的前景和背景信息,提高了网络分割结果。对比结果表明本文网络 UH-Net 可以有效捕获病变区域,提高网络的分割结果。

3.3.2 MhGPCSA 编码器分析

为验证 MhGPCSA 编码器各部分对实验结果的影响,将本文网络 UH-Net 进行如下调整:(a)使用卷积操作代替 UH-Net 的 MhGPCSA 编码器,记为 Network_0;(b)将门控和位置信息从 MhGPCSA 编码器中移除,记为 Network_1;(c)仅将门控从 MhGPCSA 编码器中移除,记为 Network_2;(d)对 UH-Net 不做任何更改。在 ISIC2018 数据集上进行实验,结果如表 3 所示。

由表 3 所示的实验结果可知,Network_1 与

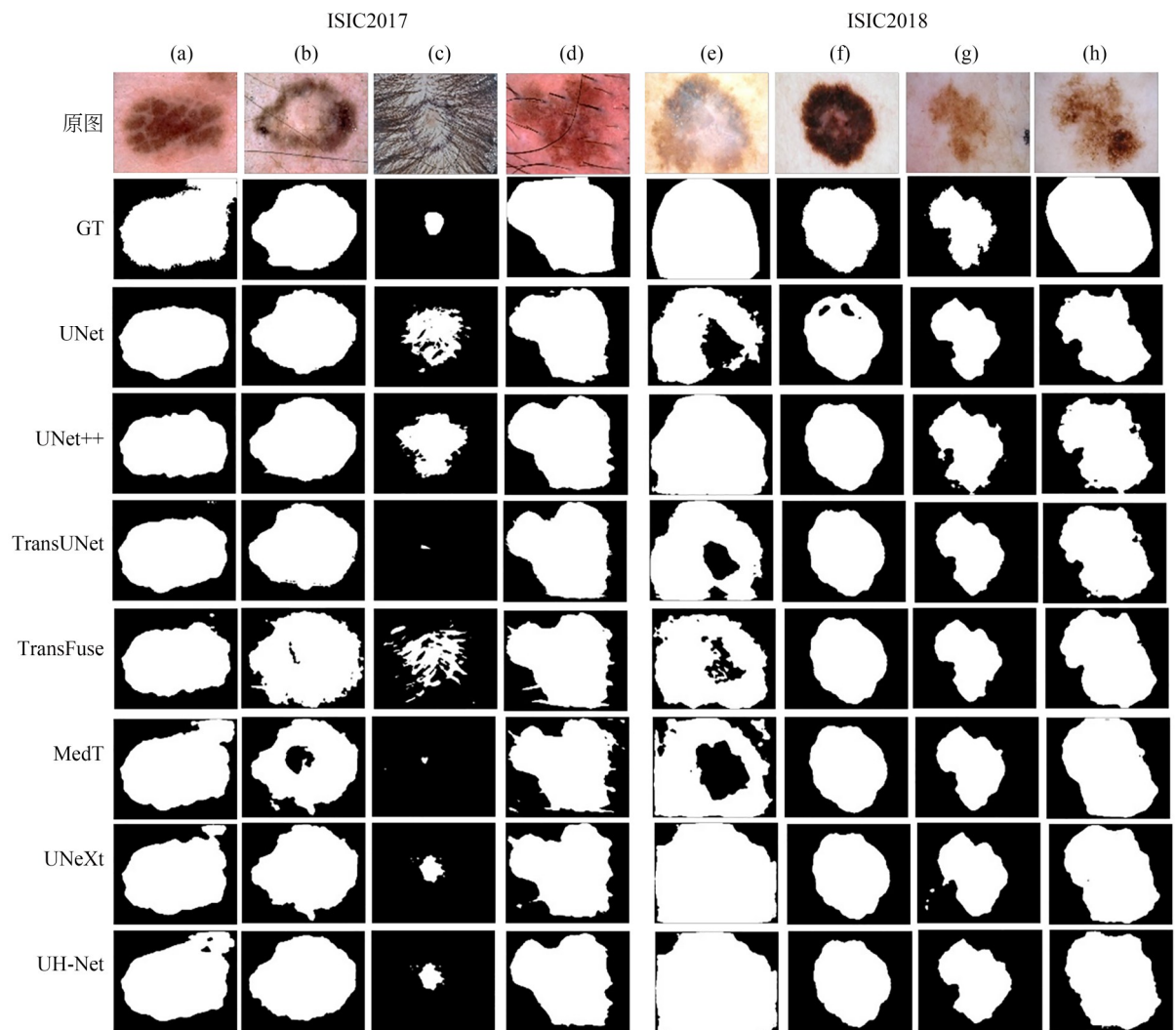


图 5 在 ISBI2017 和 ISIC2018 数据集上不同网络的分割结果

Fig. 5 Segmentation results for different networks on ISBI2017 and ISIC2018 datasets

表 3 MhGPCSA 编码器的不同设置对实验结果的影响

Tab. 3 Effect of different MhGPCSA encoder settings on experimental results

网络	分割性能/%		Params/M	GFLOPs/G
	Dice	IoU		
Network_0	90.19±0.18	82.34±0.21	1.42	1.86
Network_1	90.72±0.32	83.12±0.45	0.78	1.42
Network_2	91.07±0.16	83.81±0.16	0.78	1.44
UH-Net	91.30±0.24	84.12±0.15	0.78	1.44

Network_0相比,Dice和IoU分别提升了0.53%,0.78%,表明通过GPCSA机制来建模输入特征图谱的全局上下文信息,可使网络的分割结果得到明显提升;Network_2与Network_1相比,Dice和IoU分别提高了0.35%,0.69%,表明向交叉自注意力机制中添加位置信息可提高网络捕获目标结构的能力,从而有效地提高网络的分割结果。UH-Net与Network_2相比,Dice和

IoU 分别提高了 0.23%, 0.31%, 表明在编码器中同时使用门控和位置信息, 更加能够捕获图像的全局上下文信息, 可进一步提高网络的分割结果。

3.3.3 PosCA 机制分析

为了验证本文提出的 PosCA 机制的有效性, 分别将 UH-Net 的 PosCA 机制舍弃(无)、仅保留 PosCA 机制中沿宽度(+宽度注意力)或高度方向(+高度注意力)的一维全局池化以及将两组一维池化更改为对应的全局池化(+全局注意力), 以了解在建模通道信息的过程中保留位置信息的重要性。在 ISIC2018 数据集上的分割结果如表 4 所示。与全局注意力相比, 添加沿任意

方向的注意力对网络分割性能的影响相当, 但当同时考虑宽度注意力和高度注意力时, 即 UH-Net, Dice 和 IoU 分别提升了 0.47% 和 1.26%。因此, 在增加可学习参数和计算成本相当的情况下, 本文提出的 PosCA 机制对网络的分割结果更有帮助。

3.4 消融实验

为评估本文所设计的 MhGPCSA 模块、PosCA 机制以及正则化 Dice 损失在图像分割任务中的有效性, 本文在 ISIC2018 数据集上进行消融实验。其中, 使用卷积核数量较少的 U-Net^[5]作为基线网络, 然后在此基础上分别加入或换上相应部件, 消融实验结果如表 5 所示。

表 4 PosCA 机制的不同设置对实验结果的影响

Tab. 4 Effect of different PosCA mechanism settings on experimental results

网络	分割性能/%		Params/M	GFLOPs/G
	Dice	IoU		
无	90.59±0.32	82.76±0.20	0.77	1.42
+宽度注意力	90.87±0.12	82.91±0.24	0.77	1.43
+高度注意力	90.91±0.26	82.93±0.16	0.77	1.43
+全局注意力	90.83±0.18	82.86±0.29	0.77	1.43
UH-Net	91.30±0.24	84.12±0.15	0.78	1.44

表 5 消融实验结果

Tab. 5 Ablation experiment

方法	MhG-PCSA	Pos-CA	正则化	IoU/%	Dice/%	HD95/%	Precesion/%	Recall/%	Params/M	GFLOPs/G
基线网络	—	—	—	80.59±0.23	88.71±0.12	35.01±0.15	82.78±0.31	86.21±0.12	1.42	1.84
	✓	—	—	81.86±0.16	89.87±0.18	34.24±0.18	92.54±0.18	87.35±0.33	0.77	1.42
	—	✓	—	81.38±0.35	89.31±0.98	34.81±0.27	91.76±0.52	86.99±0.26	1.42	1.86
	—	—	✓	81.72±0.26	89.77±0.11	34.56±0.41	93.67±0.34	85.18±0.43	1.42	1.84
	✓	✓	—	82.46±0.12	90.35±0.21	34.01±0.26	89.94±0.20	90.76±0.19	0.78	1.44
	✓	—	✓	82.76±0.20	90.59±0.32	33.93±0.18	92.38±0.16	88.87±0.21	0.77	1.42
	—	✓	✓	82.34±0.21	90.19±0.18	34.27±0.50	92.32±0.52	88.16±0.42	1.42	1.86
UH-Net	✓	✓	✓	84.12±0.15	91.30±0.24	33.24±0.24	92.19±0.42	90.52±0.37	1.44	1.44

从表 5 实验结果可知, 在 ISIC2018 数据集上, 使用 MhGPCSA 编码器替换基线网络最后两个层级的卷积, Dice 和 IoU 分别提升了 1.16% 和 1.27%, HD95 减小了 0.77, 准确率和召回率分别

提升了 9.76% 和 1.14%, 表明使用 GPCSA 机制来学习图像语义信息的长期依赖关系, 可有效提高网络捕获图像目标结构的能力, 减少过分割; 向基线网络中引入 PosCA 机制, Dice 和 IoU 分别

提升了0.60%和0.79%,HD95减小了0.20,准确率和召回率分别提升了8.98%和0.78%,而网络的计算复杂度和参数量增加较少,表明PosCA机制可有效地保留图像中的位置信息;向Dice损失中引入正则化项来训练基线网络,Dice和IoU分别提升了1.06%和1.13%,HD95减小了0.45,准确率提升了10.89%,而召回率略有下降,表明通过向Dice损失中引入一个正则项,提高了网络在推理过程中对背景区域的关注度,可有效地降低过分割率。显然,本文设计的三个部件是有效的,能够有效地提高网络的分割精度。同时也可看出,与仅使用一个部件相比,三个部件进行两两组合可进一步提高网络的分割精度,且同时使用三个部件(即UH-Net),网络的分割性能最高,这证实了三个部件可以相互补充。

参考文献:

- [1] 陈东宇,杨晓雨,王红心,等. 基于Joinpoint回归和年龄-时期-队列模型分析中国皮肤癌的长期发病趋势[J]. 中华疾病控制杂志, 2022, 26(7): 756-765.
CHEN D Y, YANG X Y, WANG H X, *et al.* Analysis of secular trends in incidence of skin cancer in China based on Joinpoint regression and age-period-cohort model [J]. *Chinese Journal of Disease Control & Prevention*, 2022, 26(7): 756-765. (in Chinese)
- [2] 林千里,张文俊,汪江,等. 皮肤黑色素瘤流行病学及防治研究进展[J]. 中国医药导报, 2019, 16(3): 28-32.
LIN Q L, ZHANG W J, WANG H, *et al.* Research progress on epidemiology, treatment and prevention of malignant melanoma of skin [J]. *China Medical Herald*, 2019, 16(3): 28-32. (in Chinese)
- [3] CODELLA N, ROTEMBERG V, TSCHANDL P, *et al.* Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC) [J]. *ArXiv e-Prints*, 2019: arXiv: 1902.03368.
- [4] 梁礼明,周琬颂,冯骏,等. 基于高分辨率复合网络的皮肤病变分割[J]. 光学精密工程, 2022, 30(16): 2021-2038.
LIANG L M, ZHOU L S, FENG J, *et al.* Skin lesion segmentation based on high-resolution composite network [J]. *Opt. Precision Eng.*, 2022, 30(16): 2021-2038. (in Chinese)
- [5] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640-651.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C] *Proceedings of the Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18, Springer, 2015: 234-241.
- [7] PANG S M, PANG C L, ZHAO L, *et al.* SpineParseNet: spine parsing for volumetric MR image by a two-stage segmentation framework with semantic image representation[J]. *IEEE Transactions on Medical Imaging*, 2021, 40(1): 262-273.
- [8] ZHOU Z W, SIDDIQUEE M M R, TAJBAKSH

4 结 论

针对皮肤病变图像中目标区域与周围区域对比度低,且病变结构和位置变化大等问题,本文提出一种融入交叉自注意力编码的U型混合网络,用于皮肤病变图像分割。首先,网络利用MhGPCSA编码器对输入特征的全局上下文信息建模,弥补了卷积操作全局建模能力不足,提高网络的分割结果;其次,在跳跃连接中引入PosCA机制来提高网络捕获目标空间结构的能力;最后通过一个新的正则化Dice损失函数,使得网络能够在FP与FN之间进行权衡,减少过分割和欠分割。在ISBI2017和ISIC2018数据集上IoU分别为84.42%和84.12%,且网络的参数量仅为TransUNet参数量的0.74%,分割效率和性能优于现有网络,为辅助医生诊断提供了有效的帮助。

- 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.
- [9] OKTAY O, SCHLEMPER J, LE FOLGOC L, *et al.* Attention U-net: learning where to look for the pancreas [J]. *ArXiv e-Prints*, 2018: arXiv: 1804.03999.
- [10] GU R, WANG G T, SONG T, *et al.* CA-net: comprehensive attention convolutional neural networks for explainable medical image segmentation [J]. *IEEE Transactions on Medical Imaging*, 2021, 40(2): 699-711.
- [11] FEYJIE A R, AZAD R, PEDERSOLI M, *et al.* *Semi-Supervised Few-Shot Learning for Medical Image Segmentation* [EB/OL]. 2020: arXiv: 2003.08462. <http://arxiv.org/abs/2003.08462.pdf>
- [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An image is worth 16x16 words: transformers for image recognition at scale[J]. *ArXiv e-Prints*, 2020: arXiv: 2010.11929.
- [13] CHEN J N, LU Y Y, YU Q H, *et al.* TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation [EB/OL]. 2021: arXiv: 2102.04306. <http://arxiv.org/abs/2102.04306.pdf>
- [14] ZHANG Y D, LIU H Y, HU Q. *TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation*[M]. *Medical Image Computing and Computer Assisted Intervention-MICCAI 2021*. Cham: Springer International Publishing, 2021: 14-24.
- [15] WANG H Y, ZHU Y K, GREEN B, *et al.* *Axial-DeepLab: Stand-Alone Axial-Attention for Panoptic Segmentation* [M]. VEDALDIA, BISCHOFH, BROXT, alet, eds. *Computer Vision-ECCV 2020*. Cham: Springer International Publishing, 2020: 108-126.
- [16] HU J, SHEN L, ALBANIE S, *et al.* Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.
- [17] WANG Q L, WU B G, ZHU P F, *et al.* ECA-Net: efficient channel attention for deep convolutional neural networks[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA. IEEE, 2020: 11531-11539.
- [18] MILLETARI F, NAVAB N, AHMADI S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation [C]. 2016 *Fourth International Conference on 3D Vision (3DV)*. Stanford, CA, USA. IEEE, 2016: 565-571.
- [19] GUTMAN D, CODELLA N C F, CELEBI E, *et al.* Skin lesion analysis toward melanoma detection: a challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC) [J]. *ArXiv e-Prints*, 2016: arXiv: 1605.01397.
- [20] TSCHANDL P, ROSENDAHL C, KITTLER H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions[J]. *Scientific Data*, 2018, 5: 180161.
- [21] 徐昌佳, 易见兵, 曹锋, 等. 采用 DoubleUNet 网络的结直肠息肉分割算法[J]. *光学精密工程*, 2022, 30(8): 970-983.
- XU C J, YI J B, CAO F, *et al.* Colorectal polyp segmentation algorithm using DoubleUNet network [J]. *Opt. Precision Eng.*, 2022, 30(8): 970-983. (in Chinese)
- [22] 孙军梅, 葛青青, 李秀梅, 等. 一种具有边缘增强特点的医学图像分割网络[J]. *电子与信息学报*, 2022, 44(5): 1643-1652.
- SUN J M, GE Q Q, LI X M, *et al.* A medical image segmentation network with boundary enhancement [J]. *Journal of Electronics & Information Technology*, 2022, 44(5): 1643-1652. (in Chinese)
- [23] VALANARASU J M J, OZA P, HACIHALILOGLU I, *et al.* *Medical Transformer: Gated Axial-Attention for Medical Image Segmentation* [M]. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2021*. Cham: Springer International Publishing, 2021: 36-46.
- [24] VALANARASU J M J, PATEL V M. UNeXt:

Mlp-Based Rapid Medical Image Segmentation Network [C]. *Proceedings of the Medical Image Computing and Computer Assisted Intervention -*

MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part V, Springer, 2022: 23-33.

作者简介:



李大湘(1974—),男,湖南怀化人,博士,硕士生导师。主要从事医学图像分割、目标检测与跟踪、医学图像识别、多实例学习、深度学习与计算视觉等方面的研究。E-mail: www_ldx@163.com

通讯作者:



杨福杰(1996—),女,河南周口人,硕士研究生。主要从事医学图像分割、医学图像识别、目标检测与跟踪、深度学习等方面的研究。E-mail: 729378215@qq.com