控制与决策
Control and Decision

面向医学图像分割的双通道特征感知网络

武相虎, 束鑫, 范燕, 黄树成, 史金龙

引用本文:
武相虎, 束鑫, 范燕, 等. 面向医学图像分割的双通道特征感知网络[J]. 控制与决策, 2025, 40(3): 871-879.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2024.0360

面向医学图像分割的双通道特征感知网络

武相虎[1]，束 鑫[1,2†]，范 燕[1]，黄树成[1]，史金龙[1]

(1. 江苏科技大学 计算机学院，江苏 镇江 212100；2. 发育与妇儿疾病四川省重点实验室，成都 610000)

摘 要: 卷积神经网络 (CNN) 在医学图像分析领域得到了广泛应用, 但是受其固定感受野的局限性, 传统的 CNN 模型难以建立图像中的长距离依赖关系. Transformer 通过自注意力机制能够建立图像全局视角下的信息依赖, 拥有更强的序列建模能力. 然而, Transformer 难以捕获图像的局部细节特征. 为了解决上述问题, 提出一种基于 CNN 与 Transformer 的混合模型 DC-TransNet, 用于医学图像分割. 首先, DC-TransNet 采用双解码器结构建立图像局部和长距离依赖, 以捕获局部和全局特征; 然后, 考虑到基于编码器-解码器结构的网络模型在不同深度提取到的特征图大小不一致, 设计两种特征感知注意力机制 CFP 和 SFP, 以合理分配局部和全局特征的权重; 最后, 在多个医学数据集上进行实验. 实验结果表明: DC-TransNet 在 2D 医学图像单类别分割任务中取得了有竞争力的结果, mIoU 和 mDice 等系数均得到了显著提升.

关键词: 深度学习；卷积神经网络；Transformer；注意力机制；多尺度特征提取；医学图像分割

中图分类号: TP391.41 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2024.0360

引用格式: 武相虎, 束鑫, 范燕, 等. 面向医学图像分割的双通道特征感知网络 [J]. 控制与决策, 2025, 40(3): 871-879.

A dual-channel feature perception network for medical image segmentation

WU Xiang-hu[1]，*SHU Xin*[1,2†]，*FAN Yan*[1]，*HUANG Shu-cheng*[1]，*SHI Jin-long*[1]

(1. School of Computer Science， Jiangsu University of Science and Technology， Zhenjiang 212100， China；
2. Development and Related Diseases of Women and Children Key Laboratory of Sichuan Province， Chengdu 610000，China)

Abstract: Convolutional neural networks (CNNs) have been widely used in the field of medical image analysis. However, due to the limitation of its fixed receptive field, the traditional CNN model makes it difficult to establish long-distance dependencies in images. Transformer can establish the information dependence in the global perspective of the image through the self-attention mechanism and has stronger sequence modeling ability. However, Transformer makes it difficult to capture the local detailed features of images. To solve the above problems, a hybrid model DC-TransNet based on the CNN and Transformer is proposed for medical image segmentation. DC-TransNet uses a dual-decoder structure to establish local and long-distance dependencies in the image and capture local and global features. Considering that the size of the feature maps extracted by the network model based on the encoder-decoder structure is inconsistent at different depths, we design two feature perception attention mechanisms, channel feature percep-tion attention (CFP) and spatial feature perception attention (SFP), to reasonably allocate the weight of local and global features. Experiments are conducted on multiple medical datasets and the results show that DC-TransNet is effective in 2D medical images. Competitive results are achieved in single-category segmentation tasks, and coefficients such as mIoU and mDice are significantly improved.

Keywords: deep learning； convolutional neural network； Transformer； attention mechanism； multi scale feature extraction； medical image segmentation

0 引 言

随着医学成像技术的不断演进, 医学影像分析已成为疾病诊断的常见手段, 精确分割医学图像的重要性日益凸显[1]. 医学图像分割的目的是在影像中准确定位目标区域, 辅助医生进行临床诊断[2-3]. 传统的医学图像分割算法主要包括阈值分割[4]、边缘检测[5] 和模糊聚类[6] 等, 这些方法在处理复杂数据时往往表现出较差的泛化能力. 此外, 一些病理图像可

收稿日期: 2024-04-02；**录用日期:** 2024-07-24.

基金项目: 江苏省研究生科研与实践创新计划项目 (SJCX24_2541)；发育与妇儿疾病四川省重点实验室开放课题项目 (2023003)；国家自然科学基金面上项目 (62276118).

†通信作者. E-mail: shuxin@just.edu.cn.

能会受到噪声干扰, 存在前景与背景对比度低、组织背景复杂等现象, 导致传统分割算法难以取得理想的分割效果.

近年来, 卷积神经网络 (CNN) 已广泛应用于医学图像检测、分割和分类等任务, 成为提高临床诊断效率和准确性的关键方法之一[7-9]. 目前, 大多数CNN分割网络均基于编码器-解码器结构[10], 其中最具代表性的是U-Net[11]和全卷积神经网络 (FCN)[12]. 该体系结构采用端到端的设计方式, 由编码器连续卷积和下采样提取图像的高级语义特征, 再由解码器逐层上采样得到预测的分割掩码, 实现了高精度的像素级分类. 虽然CNN在医学图像分割中的优势已经确立, 但是受其固定感受野的局限性, CNN模型难以建立图像长距离依赖关系. 视觉Transformer (ViT)[13]凭借其基于自注意力机制结构从全局视角捕获图像特征, 具有更好的序列建模能力, 克服了CNN固定感受野的局限性. 然而, 为了计算图像序列间的相似关系, Transformer模型需要较大的计算量和参数量, 往往需要在大型数据集上进行预训练后才可准确提取图像的全局位置信息. 此外, Transformer注重在全局视角下提取图像特征, 不能充分地捕获图像的局部细节特征. 为了发挥CNN和Transformer各自的优势, 本文提出一种基于CNN-Transformer混合的双通道特征感知网络, 简称为DC-TransNet, 用于生物医学图像分割. 本文内容具体如下.

1) DC-TransNet采用双解码器结构, 利用Transformer自注意力机制弥补CNN固定感受野的局限性, 同时利用CNN补偿Transformer无法捕捉图像局部细节特征的不足, 通过空间特征感知注意力 (SFP) 和通道特征感知注意力 (CFP) 有效感知图像局部和全局特征.

2) 提出一种空间特征感知注意力SFP, 并将其嵌入至特征图通道数少且分辨率较大的浅层解码器部分, 用于感知浅层特征的空间重要性, 并对双解码器中同一深度的特征进行有效融合.

3) 提出一种通道特征感知注意力CFP, 并将其嵌入至特征图通道数较多的深层解码器部分, 对双解码器的高级语义特征进行自主选择.

4) 在4个不同类型的医学图像数据集上进行性能测试, 实验结果表明DC-TransNet的多项测评指标均取得了最优值.

1 相关工作

1.1 卷积神经网络 (CNN)

近年来, CNN被广泛应用于计算机视觉任务, 研究人员提出了许多端到端架构的编码器-解码器结构模型, 如U-Net[11]、DeepLab[14]等, 这些模型可辅助医生在手动标注前进行高精度的预分割处理. U-Net是一个完全对称的U型结构网络, 采用跳跃连接来融合编码器部分的低级特征和解码器部分的高级特征, 取得了显著的性能提升. 由于U-Net结构简单且固定, 在此基础上衍生出了一系列变体模型. Zhou等[15]认为, 不管是浅层特征还是深层特征, 对于最终的目标分割均是重要的, 于是提出了U-Net ++模型. U-Net ++嵌套了不同深度的子网络, 建立了密集连接来整合不同层次的图像特征, 使得网络的分割精度更高. 同时, U-Net ++搭配深监督机制, 在满足精度要求的情况下大幅减少了网络参数量. DoubleUNet[16]使用两个并行的U-Net架构: 一个U-Net用于接收图像并生成粗糙的分割结果; 另一个U-Net同时接收原始输入图像和前者生成的粗糙掩码, 并利用这些信息生成细化的分割结果. Xu等[17]指出, U-Net的编码器部分存在一些设计上的限制, 包括每个下采样层的结构相似以及简单的卷积层堆叠, 这些限制可能会影响U-Net从不同深度提取丰富的特征信息. 因此, 提出了DCSAU-Net, 这是一种更深、更紧凑的U型网络, 其依赖于主要特征守恒和紧凑分裂注意力模块, 以有效地融合底层和高层的语义信息. 曹飞道等[18]提出了一种视网膜血管分割模型ResDBTAGU-Net, 该网络在解码阶段引入三端注意力模块抑制背景噪声, 并在网络中加入了不同尺度的空洞卷积来扩大感受野, 在视网膜数据集上取得了优异的分割性能. 王维等[19]提出了一种自适应多目标进化卷积神经架构搜索算法AdaMo-ECNAS, 可构建灵活的分割模型, 通过多目标进化算法找到合适的网络架构和超参数, 最大限度地减少了计算成本. 齐咏生等[20]构建了一种三通道语义表征模型MCDWA-Net, 该模型使用3个通道分别提取图像中3类互补的语义特征, 包括图像的局部信息、过渡信息和类别语义信息, 在保证高分割精度的同时实现了较快的推理速度.

虽然上述方法在医学图像分割任务中取得了不错的成果, 但是, 基于CNN的网络模型仍然受到固定感受野的限制, 不能充分地提取图像全局特征. 而视觉Transformer可通过其自注意力机制很好地处理序列信息、捕获全局视角下的图像特征, 弥补了CNN中存在的这一局限性.

1.2 CNN-Transformer混合网络

Transformer早期应用于自然语言处理 (NLP),

凭借其自注意力机制, 可以很好地捕获全局序列信息. 如今, 随着视觉 Transformer 技术的不断完善, 基于 Transformer 的医学图像分割模型得到了快速发展. 为了有效利用图像的局部和全局特征, Chen 等[21] 提出了第 1 个 CNN-Transformer 混合模型 TransUNet, 该网络在编码器中使用 ResNet50 来提取图像的局部特征, 然后使用 Transformer 建立图像的长距离依赖关系, 在多器官分割等多项医学图像分割任务中取得了更好的性能; Wang 等[22] 提出的 UCTrans 采用 Ctrans 模块取代传统 U-Net 的跳跃连接, 以弥补低级特征与高级特征间的语义鸿沟和分辨率差异, 能够捕获复杂的通道依赖关系; Huang 等[23] 提出了 TDD-UNet, 用于 COVID-19 病灶区域分割, TDD-UNet 是一种带有双解码器的网络, 其将 Transformer 的多头注意力机制引入至 U-Net 编码层来提取全局上下文信息, 并通过对背景的预测和深监督机制来改善前景分割的效果; Zhang 等[24] 认为, 基于 Transformer 的分割网络中位置编码模块不能充分地编码位置信息, 串行解码器不能有效地利用上下文信息, 因此提出了 APT-Net, 该网络引入用于融合多接受域位置信息的自适应位置编码模块, 为 Transformer 提供更充分的位置信息; Yang等[25] 提出了 MMViT-Seg, 该模型采用双编码器路径, 既能够捕获图像特征的全局依赖关系, 又能捕获图像浅层细节特征, 在多种病例分割中取得了优秀的表现.

综上, CNN 和 Transformer 在医学图像分割中能够优势互补. 然而, 上述网络模型只是简单地将图像的全局与局部语义特征进行融合, 未能充分考虑两者之间存在的语义鸿沟和不同深度的特征图分辨率差异. 为了克服这些问题, 本文利用视觉 Transformer 来解决 CNN 固有的局限性, 并提出两种不同的注意力机制, 用于感知双解码器通道的特征权重, 从而更好地融合局部和全局信息, 提高模型对医学图像的分割性能.

2　方　法

2.1　DC-TransNet 整体架构

本文提出一种带有双解码器的 CNN-Transformer 混合网络 DC-TransNet, 用于 2D 医学图像单类别分割任务, 其整体架构如图 1 所示. DC-TransNet 由左侧编码器和右侧双解码器组成, 网络中均采用3×3大小的双卷积操作来完成主要的特征提取功能. 对于输入图像$x \in R^{H\times W\times C}$, 由编码器进行逐层下采样得到特征图$C_i(i=1,2,3,4)$, 瓶颈层特征图$C_4$作为解码器 A 和解码器 B 的输入. Transformer 需要将特征映射平面化来计算注意力矩阵, 这在一定程度上会导致通道信息丢失, 因此, 利用单独的一条解码器 (解码器 A) 建立图像的全局上下文依赖联系, 并指导解码器 B 融合局部和全局特征. 具体而言, 解码器 A 将特征图C_4序列化后由 Transformer 提取全局序列特征, 再通过卷积运算和逐层上采样来增强得到的全局信息, 并生成特征图$V_i(i=1,2,3,4)$. 解码器 B 为传统的 CNN 子网络, 其任务是建立图像的局部信息依赖关系, 提取细节特征, 从而得到特征图$D_i(i=1,2,3,4)$. 首先, 在逐层上采样后执行跳跃连接, 并使用三重注意力模块 (TA)[26] 建立跨维度的信息依赖; 然后, 使用所提出 SFP 和 CFP 模块有效感知当前局部特征D_i以及解码器 A 捕获的同一深度的全局特征V_i, 作为解码器 B 下一次上采样操作的输入; 最后, 由解码器 B 生成预测的分割掩码y.

2.2　ViT 编码器

图像序列化: 首先, ViT 将输入特征图$x\in R^{H\times W\times C}$分割为$n=HW/p^2$个边长为$p$的正方形块, 有序组成一个图像序列$\{x_p^i \in R^{p\times p\times C} | i=1,2,\ldots,n\}$, 并

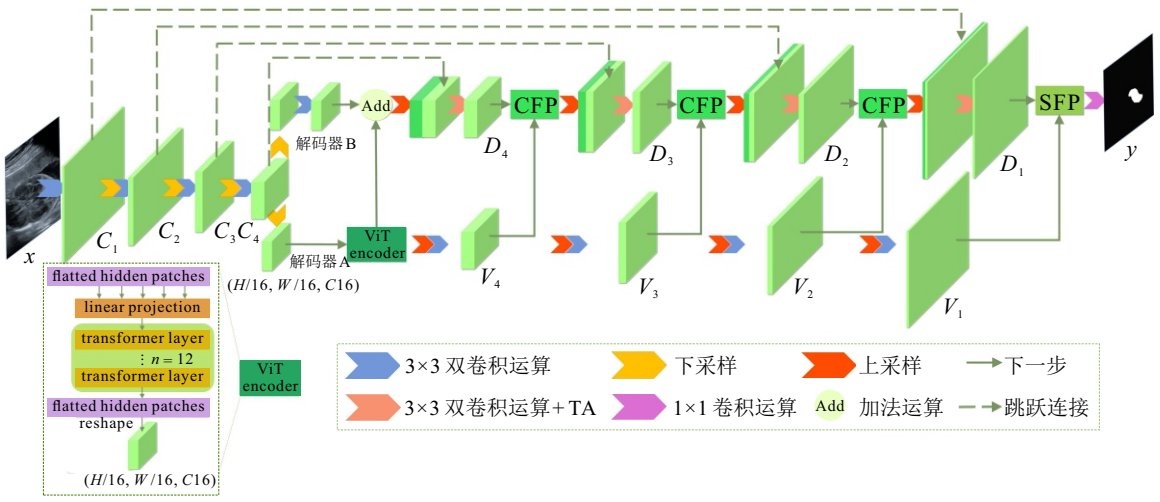


图1　DC-TransNet 网络架构

使用一个可训练的向量$E \in R^{(p^2\times C)\times d}$将$x_p^i$线性投影到$d$维; 然后, 加入位置编码向量$E_{\text{pos}} \in R^{n\times d}$来保留空间位置信息; 最后, 得到向量$z$, 即

$$z = [x_p^1 E; x_p^2 E; \dots; x_p^n E] + E_{\text{pos}}. \tag{1}$$

transformer layer: transformer layer 包含多头注意力机制 (MHA)、多层感知器 (MLP)、归一化和残差连接. 其核心在于多头注意力机制 MHA, 通过计算序列间的相似性, 有效弥补了 CNN 在感受野方面的限制. 图 2 为多头注意力机制. 如图 2 所示: 首先, ViT 根据多头注意力的头数将输入的图像序列z等分为$z_i(i=1,2,\dots,h)$. 然后, 通过将z_i与 3 个可训练参数矩阵w_q、w_k、w_v做点积运算, 生成注意力权重矩阵Q_i、K_i、V_i, 有

$$Q_i = w_q \cdot z_i, \tag{2}$$

$$K_i = w_k \cdot z_i, \tag{3}$$

$$V_i = w_v \cdot z_i. \tag{4}$$

通过点积运算计算注意力权重Q_i与K_i间的相关性, 经归一化后生成如下矩阵$A_i(i=1,2,\dots,h)$:

$$A_i = \text{soft max}\left(\frac{Q_i \cdot K_i^{\text{T}}}{\sqrt{d_{hd}}}\right), \tag{5}$$

其中d_{hd}为向量K的维度, 用于对点积后的值$(Q_i \cdot K_i^{\text{T}})$进行缩放, 以防止数值过大导致的梯度爆炸等情况发生. 接着, 利用得到的A_i和V_i计算最终的注意力权重$O_i(i=1,2,\dots,h)$, 有

$$O_i = A_i \cdot V_i. \tag{6}$$

最后, 将得到的权重O_i拼接后通过一个线性层压缩通道数, 得到 MHA 的输出特征图$y_{\text{MHA}} \in R^{n\times d}$.

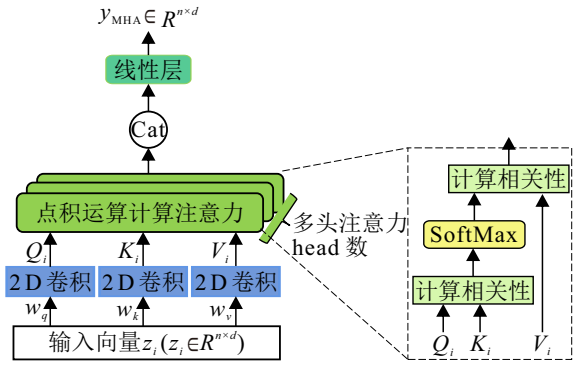


图2 多头注意力机制

2.3 双通道特征感知注意力

大多数双解码器网络在融合两部分特征时通常只执行简单的拼接或相加操作, 并未考虑不同特征图间的语义鸿沟和分辨率不匹配等问题. 此外, 上采样过程中特征图的分辨率和通道数在不断发生变化, 如 U-Net 中, 瓶颈层的高级语义特征图大小为$\left(\frac{H}{16}\right)\times\left(\frac{W}{16}\right)\times(C\times 16)$, 解码器会逐层上采样将特征图恢复至$H\times W\times C$的输入大小, 这表明需要使用不同方法计算不同深度的特征权重. 为了有效融合双解码器通道的特征表示, 本文提出了两种不同的特征感知注意力机制 CFP 和 SFP 来获取网络在不同深度的特征权重.

CFP 为通道特征感知注意力, 用于通道数较多且特征图相对较小的深层解码器, 其结构如图 3 所示. CFP 的输入包括来自解码器 A 的全局特征$V_i(i=2,3,4)$和来自解码器 B 的局部特征$D_i(i=2,3,4)$. 其核心思想在于建立高效的深层网络中通道间的相互依赖关系, 自适应地校准通道维度的特征响应, 从而获取每个特征通道的重要程度, 即对特征分布较多的通道赋予更大的权重因子, 并利用$V_i(i=2,3,4)$从图像全局角度指导局部特征的通道权重分配, 再通过逐点卷积运算生成最后的通道权重, 这也避免了通道与权重间存在间接的映射关系, 可以更真实地反映深层网络中特征图的通道重要性. 具体而言, CFP 机制的工作流程如下.

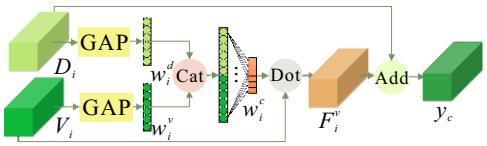


图3 CFP 结构

首先, 对输入的V_i和D_i进行平均池化, 得到权重向量$W_i^v \in R^{1\times 1\times C}$和$W_i^d \in R^{1\times 1\times C}$, 并将$W_i^v$和$W_i^d$拼接在一起, 通过卷积操作生成通道权重$W_i^c \in R^{1\times 1\times C}$. 然后, 将$V_i$与$W_i^c$做点积运算生成特征图$F_i^v$, 并将$F_i^v$与$D_i$相加, 得到输出特征图$y_c$. 简单来说, CFP 可表示为

$$W_i^c = \sigma(w_0[g(D_i); g(V_i)]) = \sigma(w_0(W_i^d; W_i^v)), \tag{7}$$

$$F_i^v = V_i \cdot W_i^c, \tag{8}$$

$$y_c = D_i + F_i^v. \tag{9}$$

其中: $g(\alpha) = \dfrac{1}{WH}\sum_{i=1,j=1}^{W,H}\alpha_{ij}$为对特征图$\alpha$做通道平均池化, σ为 ReLU 激活函数, w_0为1×1大小的卷积核.

SFP 为多尺度空间特征感知注意力, 用于通道数较少且特征图相对较大的浅层解码器, 其结构如图 4 所示. 与先前的空间注意力不同, SFP 可建立全局图像级与像素级的特征联系, 其核心在于综合考量图像的全局结构信息和局部细节特征, 通过压缩

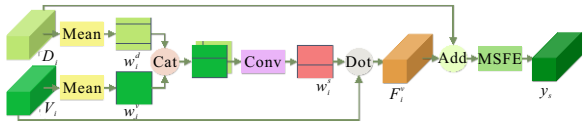


图4　SFP 结构

通道维度的方式来获取各自的空间重要性, 并使用线性运算将全局和局部特征直接映射为最后的综合空间权重. 其全局特征权重与局部细节权重的结合使得网络可以同时关注病灶区域的整体空间定位和边缘细节特征. 具体而言, SFP 机制的工作流程如下.

首先, SFP 分别对V_i和D_i的通道计算均值, 得到空间特征权重$W_i^v \in R^{H\times W\times 1}$和$W_i^d \in R^{H\times W\times 1}$, 并将$W_i^v$与$W_i^d$拼接在一起, 通过卷积操作来压缩通道数, 生成空间权重W_i^s; 然后, 将V_i与W_i^s做点积运算, 生成特征图F_i^v; 最后, 将F_i^v与D_i相加, 输入至 MFSE 模块来提取不同尺度的特征信息, 最终输出特征图y_s. 简单来说, SFP 可表示为

$$W_i^s = \sigma(w_0[g(D_i); g(V_i)]) = \sigma(w_0(W_i^d; W_i^v)), \tag{10}$$

$$F_i^v = V_i \cdot W_i^s, \tag{11}$$

$$y_s = \mathrm{MSFE}(F_i^v + D_i). \tag{12}$$

其中: $g(\alpha) = \dfrac{1}{C}\sum_{i=1}^{C}\alpha_i$为对特征图$\alpha$的通道计算均值, σ为 ReLU 激活函数, w_0为1×1大小的卷积核.

2.4　多尺度特征提取模块

受到 ResNet[27] 的启发, 本文设计了一种基于 2D 卷积操作的多尺度特征提取模块 (MSFE), 其结构如图 5 所示. MSFE 模块由两组并行的卷积操作组成: 在第 1 个分支中, 通过使用1×1大小的卷积核对每个像素进行独立的线性变换, 生成特征图y_1; 在第 2 个分支中, 通过两个3×3大小的卷积以及批处理归一化, 并经 ReLU 函数激活得到特征图y_2; 最后, 将y_1与y_2相加得到 MSFE 的输出特征图y_{MSFE}. 从结构上看, MSFE等同于大小分别为1×1、3×3、5×5的 3 个卷积核并联的轻量级特征提取模块, 这一设计允许其能够有效地捕获不同尺度的特征信息, 有助于提高网络性能. MSFE 可表示为

$$y_1 = \sigma(w_1 x), \tag{13}$$

$$y_2 = \sigma(w_2 x) + \sigma(w_2\sigma(w_2 x)), \tag{14}$$

$$y_{\mathrm{MSFE}} = y_1 + y_2. \tag{15}$$

其中: x为输入特征图, σ为 ReLU 激活函数, w_1、w_2分别为1×1和3×3卷积操作.

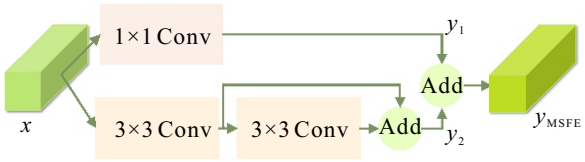


图5　多尺度特征提取模块 (MSFE)

2.5　损失函数

本文采用 PPA 损失函数[28] 来训练模型, PPA 损失函数包括 wBCE 损失 (L_{wBCE}^s) 和 wIoU 损失 (L_{wIoU}^s), 其中L_{wBCE}^s可以很好地解决 BCE 损失独立计算每个像素的损失、忽略图像全局结构等问题. 与 BCE 损失相比, L_{wBCE}^s更关注硬像素, 可对局部结构信息进行编码, 有助于模型关注更大的感受野. 而L_{wIoU}^s可应对传统 IoU 损失忽略像素间差异的问题, 通过引入权重因子, L_{wIoU}^s可将更多的权重赋予硬像素来强调它们的重要性. L_{wIoU}^s和L_{wBCE}^s分别表示为

$$L_{\mathrm{wIoU}}^s = 1 - \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(g_{ij}^s \times p_{ij}^s)\times(1+\gamma\alpha_{ij}^s)}{\sum_{i=1}^{H}\sum_{j=1}^{W}(g_{ij}^s + p_{ij}^s - g_{ij}^s \times p_{ij}^s)\times(1+\gamma\alpha_{ij}^s)}, \tag{16}$$

$$L_{\mathrm{wBCE}}^s = -\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(1+\gamma\alpha_{ij})\sum_{l=0}^{1}\mathbf{1}(g_{ij}^s = l)\log\Pr(p_{ij}^s = l|\Psi)}{\sum_{i=1}^{H}\sum_{j=1}^{W}\gamma\alpha_{ij}}. \tag{17}$$

其中: $\mathbf{1}(\cdot)$为指示函数; γ为一个超参数; $l\in\{0,1\}$为真、假两种标签; p_{ij}^s和g_{ij}^s分别为图像中坐标(i,j)处像素的预测值和真值; s为空间维度; Ψ为模型的所有参数; $\Pr(p_{ij}^s = l|\Psi)$为预测的概率; α为像素重要性的指示符, 可根据中心像素与其周围像素间的差值计算, 如下所示:

$$\alpha_{ij}^s = \left|\frac{\sum_{m,n\in A_{ij}} g_{mn}^s}{\sum_{m,n\in A_{ij}} 1} - g_{ij}^s\right|. \tag{18}$$

这里: α_{ij}为像素坐标(i,j)周围的区域, 对于所有像素, $\alpha_{ij}^s\in[0,1]$.

PPA 损失L为

$$L = L_{\mathrm{wIoU}}^s + L_{\mathrm{wBCE}}^s. \tag{19}$$

将L_{wIoU}^s与L_{wBCE}^s相结合, PPA 损失函数为所有像素生成不同的权重来指导网络学习, 这可以使得网络

在获取结构信息和特征表示方面取得更好的效果, 并产生清晰的细节.

3 实 验

3.1 数 据 集

为了评估 DC-TransNet 的有效性和泛化能力, 在 4 个数据集上进行性能测试, 包括 JSUAH-Cerebellum[8]、2018 DSB[29]、ISIC 2018[30] 和 CVC-ClinicDB[31]. 具体如表 1 所示.

表1 各数据集的详细信息

数据集	图像数量	训练集	验证集	测试集
JSUAH-Cerebellum	959	699	68	192
2018 DSB	670	536	67	67
ISIC 2018	2 594	2 076	259	259
CVC-ClinicDB	612	441	110	61

3.2 实验环境

实验在装备有 AMD 5900 x 处理器和 NVIDIA GeForce RTX 1080 的设备上进行. 为了解决医学图像数据集规模较小的问题, 本文对所用数据集使用随机垂直翻转($p = 0.5$)、随机水平翻转($p = 0.5$)和随机旋转进行数据增强处理, 图像分辨率设置为 224×224. 模型的初始学习率均为 1E-4, 训练周期设置为 500, batch size 设置为 6, 学习率使用余弦退火策略进行动态调整, 最大迭代次数为 40. 所有实验均采用相同的训练策略和超参数设置.

3.3 性能评估标准

本文使用 IoU[32]、Dice[33]、Precision[34]、Recall[31] 和 F1[35] 五项性能评估指标来测试网络性能, 指标分数越高, 预测的结果越接近真实标签.

3.4 性能评价

3.4.1 与其他算法对比

为了验证 DC-TransNet 在医学图像分割中的有效性, 本文进行一系列对比实验, 包括一些先进的分割网络: U-Net[11]、U-Net ++[15]、SCIF-Net[36]、FCRB U-Net[37]、DoubleUNet[16]、CSCA U-Net[9]、TransUNet[21]、UCTrans[22] 以及 DCSAU-Net[17].

表 2 为各模型在 JSUAH-Cerebellum 数据集上的实验结果. 从实验数据来看, DC-TransNet 模型的分割性能优于其他网络, 5 项测评指标分别达到了 87.89 %、93.31 %、93.33 %、93.91 % 和 93.49 %. 与 U-Net 相比, DC-TransNet 的各项指标分别提升了 1.4 %、0.97 %、0.55 %、1.02 % 和 0.96 %. 相比于 DCSAU-Net, 各项指标分别提高了 2.4 %、1.63 %、0.68 %、2.11 % 和 1.55 %. 这是由于 DC-TransNet 在 CNN 基础上很好地适配了 Transformer, 弥补了 CNN 无法建立全局依赖的不足. 与首个 CNN-Transformer 混合模型 TransUNet 相比, 尽管 mPrecision 略有降低, 但是, 其余指标分别提高了 0.31 %、0.23 %、0.84 % 和 0.19 %, 表明 DC-TransNet 在整体上具有更好的分割性能.

表2 在 JSUAH-Cerebellum 数据集上的实验结果 %

model	mIoU	mDice	mPrecision	mRecall	mF1
U-Net[11]	86.49	92.34	92.78	92.89	92.53
U-Net ++[15]	86.69	92.50	92.83	93.15	92.69
SCIF-Net[36]	86.98	92.65	92.49	93.66	92.88
FCRB U-Net[37]	86.72	90.45	91.88	93.56	92.23
DoubleUNet[16]	87.14	92.70	93.08	93.35	92.95
CSCA U-Net[9]	87.19	93.07	91.89	**94.63**	93.24
TransUNet[21]	87.58	93.08	**93.83**	93.07	93.30
UCTrans[22]	85.67	91.97	92.68	93.05	92.65
DCSAU-Net[17]	85.49	91.68	92.65	91.80	91.94
ours	**87.89**	**93.31**	93.33	93.91	**93.49**

在生物医学图像分析领域, 细胞核的分割是一项重要的任务. 为了评估 DC-TransNet 与其他对比网络在这一任务上的性能, 本文使用 2018 DSB 数据集, 并将各模型的评估结果列于表 3 中. 实验结果显示: DC-TransNet 的 mIoU 达到了 88.70 %, 相较于 U-Net ++高出了 1.07 %;而在 mDice 指标上, DC-TransNet 达到了 92.63 %, 比 DCSAU-Net 高 1.05 %. 这些结果表明, DC-TransNet 在细胞核分割任务中表现出色, 其较高的 mIoU 和 mDice 表明 DC-TransNet 对细胞核边界具有优秀的分割能力.

表3 在 2018 DSB 数据集上的实验结果 %

model	mIoU	mDice	mPrecision	mRecall	mF1
U-Net[11]	87.44	91.90	94.16	92.51	92.96
U-Net ++[15]	87.63	92.01	93.40	93.66	93.14
SCIF-Net[36]	87.67	89.58	93.50	93.53	93.21
FCRB U-Net[37]	86.52	89.08	84.58	**95.89**	89.49
DoubleUNet[16]	87.85	91.91	93.20	94.09	93.30
CSCA U-Net[9]	87.50	91.61	93.81	93.01	93.00
TransUNet[21]	85.20	85.45	83.65	95.43	88.44
UCTrans[22]	82.02	89.37	92.28	92.08	91.86
DCSAU-Net[17]	87.23	91.58	92.59	93.91	92.93
ours	**88.70**	**92.63**	**94.50**	93.54	**93.92**

表 4 为各模型在皮肤病变挑战赛 ISIC 2018 上的评估结果, 其中 mIoU 为该挑战赛的官方评估标准. 实验结果表明: DC-TransNet 在多项指标上取得了最优值, 特别是 mIoU 达到了 87.37 %, 相较于 UCTrans 提升了 1.53 %,比 FCRB U-Net 提升了 1.42 %. 这突显了 DC-TransNet 在分割任务中对病灶区域的预测结果与相应标签间更具相似性, 对于皮肤病变

的自动化识别具有重要价值.

表4　在 ISIC 2018 数据集上的实验结果　%

model	mIoU	mDice	mPrecision	mRecall	mF1
U-Net[11]	85.34	90.38	93.50	91.44	91.77
U-Net ++[15]	85.47	90.56	94.00	91.15	91.88
SCIF-Net[36]	86.19	91.60	93.54	92.41	92.28
FCRB U-Net[37]	85.95	91.34	93.29	92.10	91.59
DoubleUNet[16]	85.11	89.45	94.92	89.82	91.58
CSCA U-Net[9]	86.34	91.99	93.61	92.54	92.35
TransUNet[21]	87.34	91.55	94.89	92.22	**93.04**
UCTrans[22]	85.84	91.78	**95.08**	90.48	92.11
DCSAU-Net[17]	85.29	90.65	94.24	90.73	91.64
ours	**87.37**	**91.98**	94.69	**92.47**	92.63

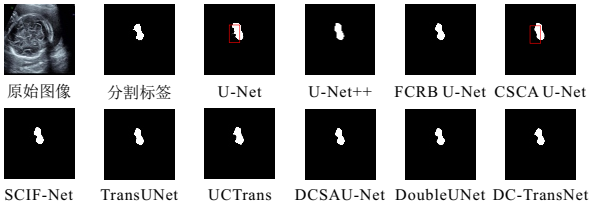
CVC-ClinicDB 数据集的定量结果如表 5 所示. DC-TransNet 在各项性能指标上表现出色, 分别达到了 88.01 %、92.36 %、93.04 %、94.41 % 和 92.77 %. 与 DCSAU-Net 相比, mIoU 和 mRecall 取得了显著提升, 分别增加了 3.99 % 和 4.07 %. 尽管 UCTrans 在 mRecall 指标上略高于 DC-TransNet, 但是, 所提出模型在其余各项测评指标上均优于 UCTrans, 这表明 DC-TransNet 的综合分割性能更好. 与其他算法相比, DC-TransNet 实现了明显的性能提升, 在处理前景与背景差异小的息肉数据时取得更佳的分割效果.

表5　在 CVC-ClinicDB 数据集上的实验结果　%

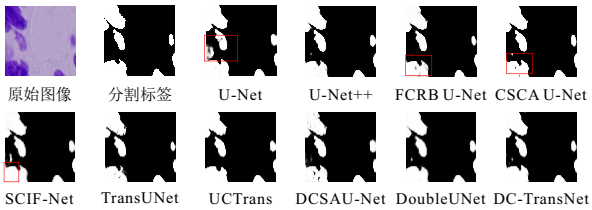
model	mIoU	mDice	mPrecision	mRecall	mF1
U-Net[11]	77.54	83.47	86.75	85.25	84.05
U-Net ++[15]	83.80	88.82	91.59	88.19	89.17
SCIF-Net[36]	86.23	91.08	92.19	91.51	91.47
FCRB U-Net[37]	86.82	91.76	**93.62**	92.13	92.12
DoubleUNet[16]	81.23	86.91	93.08	85.68	87.35
CSCA U-Net[9]	87.17	91.76	91.86	93.16	92.02
TransUNet[21]	82.32	87.49	88.49	88.32	87.79
UCTrans[22]	83.62	89.22	89.82	91.25	89.50
DCSAU-Net[17]	84.02	89.59	90.30	90.34	90.05
ours	**88.01**	**92.36**	93.04	**94.41**	**92.77**

3.4.2　可视化分割结果

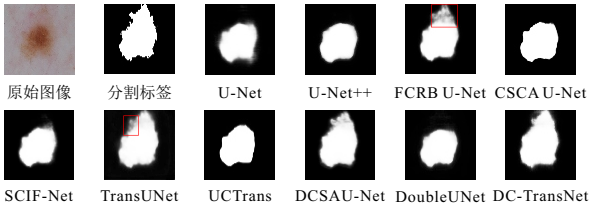
图 6 为所有对比网络在各数据集上的部分可视化分割结果. 针对 A 组样本: U-Net 存在分割边界不够平滑和欠分割等问题, CSCA U-Net 在目标区域的左侧分割不够精准. 在 B 组样本中: CSCA U-Net 和 FCRB U-Net 出现了过度分割, 而 SCIF-Net 和 U-Net 受到了一些噪声的影响. C 组样本显示: 对于前景与背景差异小的区域, U-Net 等 6 个网络无法完整识别, 而 FCRB U-Net 和 TransUNet 分别存在分割不精细和边界模糊的问题. D 组样本中: 其他网络存在明显的分割缺陷, 而 DC-TransNet 的分割结果最接近真实标签.



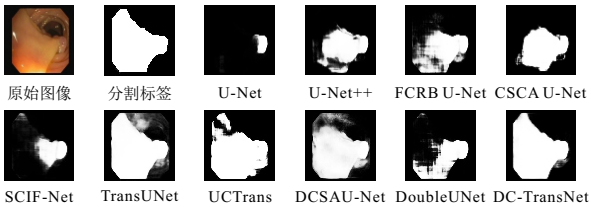
(a)　JSUAH-Cerebellum 数据集分割结果(A 组样本)



(b)　2018 DSB 数据集分割结果(B 组样本)



(c)　ISIC 2018 数据集分割结果(C 组样本)



(d)　CVC-ClinicDB 数据集分割结果(D 组样本)

图6　可视化分割结果

总体来看, DC-TransNet 不仅能够实现高精度的定位, 而且其分割边界更加平滑.

3.4.3　消融实验

为了进一步验证 DC-TransNet 模型中每个模块的有效性, 在 JSUAH-Cerebellum 数据集上进行了详细的消融实验, 结果如表 6 所示.

表6　在 JSUAH-Cerebellum 数据集上的消融实验结果　%

实验设置	mIoU	mDice	mPrecision	mRecall	mF1
U-Net (编码器 + 解码器B)	86.49	92.34	92.78	92.89	92.53
U-Net + 解码器A	87.27	92.87	93.17	93.40	93.09
U-Net + 解码器A + CFP	87.53	93.05	93.35	93.51	93.26
U-Net + 解码器A + SFP	87.65	93.12	**93.45**	93.56	93.33
U-Net + 解码器A + CFP + SFP	**87.89**	**93.31**	93.33	**93.91**	**93.49**

1) U-Net + 解码器 A: 在 U-Net 的基础上引入一个带有 Transformer 的解码器 B, 将瓶颈层的特征图传送至 Transformer, 建立图像中的长距离依赖关系, 通过上采样生成特征图V_i. 然后, 将V_i与图 1 中解码器 B 生成的特征图D_i进行简单的相加运算. 实验结

果显示, 加入解码器 A 后, 网络的各项指标分别提高了 0.78 %、0.53 %、0.39 %、0.51 % 和 0.56 %. 这表明, 利用解码器 A 来提取图像的全局特征能够有效提升网络性能.

2) U-Net + 解码器 A + CFP: 在 1) 的基础上引入 CFP 模块, 它能够有效地计算解码器 A 和解码器 B 中特征图的通道重要性. 通过引入 CFP 模块, 各项分割指标均得到了提升, 这表明 CFP 模块在提升网络的分割性能方面发挥了积极的作用.

3) U-Net + 解码器 A+SFP: 在 1) 的基础上加入 SFP 模块, 用于感知双通道中特征图的空间重要性. 实验结果表明, 加入 SFP 模块后网络的各项性能分别提升了 0.38 %、0.25 %、0.28 %、0.16 % 和 0.24 %.

4) U-Net + 解码器 A + CFP + SFP: 最后, 将各模块进行融合, 组成了 DC-TransNet, 在 5 个测评指标中有 4 项取得了最高值, 表明所提出各模块间的协同配合对于整体网络性能的提升具有重要作用, 以及使用不同的方法对网络中不同深度的特征图计算注意力权重是至关重要的.

4 结 论

本文提出了一种面向医学图像的分割网络, 称为 DC-TransNet. 该模型巧妙地结合了 Transformer 和 CNN 两种架构, 建立了强大的上下文依赖关系, 缓解了不同特征图间的语义鸿沟和分辨率不匹配问题. DC-TransNet 通过 CFP 和 SFP 计算解码器在不同深度时的双解码器通道特征权重, 从而有效地融合了图像的局部信息和全局特征表示. 在 4 个医学图像数据集上对网络性能进行了广泛评估, 实验结果表明, DC-TransNet 在分割任务中表现出色, 多项评估指标均取得了最好值. DC-TransNet 将 CNN 与 Transformer 相结合, 有效地克服了各自的局限性, 为医学图像分割任务提供了更好的解决方案, 在临床医学和科学研究领域具有一定的潜力和应用价值.

参考文献 (References)

[1] Ma X J, Niu Y H, Gu L, et al. Understanding adversarial attacks on deep learning based medical image analysis systems[J]. Pattern Recognition, 2021, 110: 107332.

[2] Wu X H, Huang S C, Shu X, et al. MPFC-Net: A multi-perspective feature compensation network for medical image segmentation[J]. Expert Systems with Applications, 2024, 248: 123430.

[3] 罗凌, 薛定宇, 冯兴隆. 基于紧凑混合网络的视网膜血管自动分割[J]. 控制与决策, 2022, 37(2): 353-360.
(Luo L, Xue D Y, Feng X L. Automatic segmentation of retinal vessel via compact mixed network[J]. Control and Decision, 2022, 37(2): 353-360.)

[4] Otsu N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.

[5] Muthukrishnan R, Radha M. Edge detection techniques for image segmentation[J]. International Journal of Computer Science & Information Technology, 2011, 3(6): 259-267.

[6] Dhanachandra N, Manglem K, Chanu Y J. Image segmentation using *K*-means clustering algorithm and subtractive clustering algorithm[J]. Procedia Computer Science, 2015, 54: 764-771.

[7] Minaee S, Boykov Y, Porikli F, et al. Image segmentation using deep learning: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence., 2022, 44(7): 3523-3542.

[8] Shu X, Chang F, Zhang X, et al. ECAU-Net: Efficient channel attention U-Net for fetal ultrasound cerebellum segmentation[J]. Biomedical Signal Processing and Control, 2022, 75: 103528.

[9] Shu X, Wang J S, Zhang A P, et al. CSCA U-Net: A channel and space compound attention CNN for medical image segmentation[J]. Artificial Intelligence in Medicine, 2024, 150: 102800.

[10] 赖晓婷, 张静. 语义扩散对齐的多尺度感知医学图像分割方法[J]. 计算机辅助设计与图形学学报, DOI: 10.3724/SP.J.1089.2023-00604.
(Lai X T, Zhang J. Semantic diffusion alignment-based multi-scale perception for medical image segmentation[J]. Journal of Computer-Aided Design & Computer Graphics, DOI: 10.3724/SP.J.1089.2023-00604.)

[11] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. Medical Image Computing and Computer-Assisted Intervention. Munich, 2015: 234-241.

[12] 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.

[13] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[J/OL]. 2020, arXiv: 2010.11929.

[14] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[15] Zhou Z W, Siddiquee M M R, Tajbakhsh N, et al. UNet ++: A nested U-net architecture for medical image segmentation[J]. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018, 11045: 3-11.

[16] Jha D, Riegler M A, Johansen D, et al. DoubleU-net: A deep convolutional neural network for medical image segmentation[C]. Proceedings of the 33rd IEEE International Symposium on Computer-Based Medical Systems. Rochester, 2020: 558-564.

[17] Xu Q, Ma Z C, 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.

[18] 曹飞道, 赵怀慈. 基于三端注意力机制的视网膜血管分割算法[J]. 控制与决策, 2022, 37(10): 2505-2512.
(Cao F D, Zhao H C. Improved U-Net based on three-terminal attention mechanism for retinal vessel segmentation[J]. Control and Decision, 2022, 37(10): 2505-2512.)

[19] 王维, 王显鹏, 宋相满. 基于自适应多目标进化CNN的图像分割方法[J]. 控制与决策, 2024, 39(4): 1185-1193.
(Wang W, Wang X P, Song X M. An image segmentation method based on adaptive multi-objective evolutionary CNN[J]. Control and Decision, 2024, 39(4): 1185-1193.)

[20] 齐咏生, 陈培亮, 高学金, 等. 高精度实时语义分割算法框架: 多通道深度加权聚合网络[J]. 控制与决策, 2024, 39(5): 1450-1460.
(Qi Y S, Chen P L, Gao X J, et al. High precision real-time semantic segmentation algorithm: Multi-channel deep weighted aggregation network[J]. Control and Decision, 2024, 39(5): 1450-1460.)

[21] Chen J N, Lu Y Y, Yu Q H, et al. TransUNet: Transformers make strong encoders for medical image segmentation[J/OL]. 2021, arXiv: 2102.04306.

[22] Wang H N, Cao P, Wang J Q, et al. UCTransNet: Rethinking the skip connections in U-Net from a channel-wise perspective with transformer[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(3): 2441-2449.

[23] Huang X P, Chen J X, Chen M Z, et al. TDD-UNet: Transformer with double decoder UNet for COVID-19 lesions segmentation[J]. Computers in Biology and Medicine, 2022, 151: 106306.

[24] Zhang N, Yu L, Zhang D Z, et al. APT-Net: Adaptive encoding and parallel decoding transformer for medical image segmentation[J]. Computers in Biology and Medicine, 2022, 151: 106292.

[25] Yang Y, Zhang L, Ren L, et al. MMViT-Seg: A lightweight transformer and CNN fusion network for COVID-19 segmentation[J]. Computer Methods and Programs in Biomedicine, 2023, 230: 107348.

[26] Misra D, Nalamada T, Arasanipalai A U, et al. Rotate to attend: Convolutional triplet attention module[C]. IEEE Winter Conference on Applications of Computer Vision. Waikoloa, 2021: 3138-3147.

[27] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770-778.

[28] Wei J, Wang S H, Huang Q M. F³Net: Fusion, feedback and focus for salient object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12321-12328.

[29] Caicedo J C, Goodman A, Karhohs K W, et al. Nucleus segmentation across imaging experiments: The 2018 data science bowl[J]. Nature Methods, 2019, 16(12): 1247-1253.

[30] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions[J/OL]. 2018, arXiv: 1803.10417.

[31] Bernal J, Sánchez F J, Fernández-Esparrach G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians[J]. Computerized Medical Imaging and Graphics, 2015, 43: 99-111.

[32] Jaccard P. The distribution of the flora in the alpine zone[J]. New Phytologist, 1912, 11(2): 37-50.

[33] Lee R D C. Measures of the amount of ecologic association between species[J]. Ecology, 1945, 26(3): 297-302.

[34] Zhang Q L, Yang Y B. SA-Net: Shuffle attention for deep convolutional neural networks[C]. IEEE International Conference on Acoustics, Speech and Signal Processing. Toronto, 2021: 2235-2239.

[35] Taha A A, Hanbury A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool[J]. BMC Medical Imaging, 2015, 15(1): 29.

[36] Tan D Y, Yao Z Y, Peng X, et al. Multi-level medical image segmentation network based on multi-scale and context information fusion strategy[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, 8(1): 474-487.

[37] Shu X, Gu Y Y, Zhang X, et al. FCRB U-Net: A novel fully connected residual block U-Net for fetal cerebellum ultrasound image segmentation[J]. Computers in Biology and Medicine, 2022, 148: 105693.

作者简介

武相虎 (1998−), 男, 硕士生, 主要研究方向为医学图像处理与分析, E-mail: 221210701225@stu.just.edu.cn;

束鑫 (1979−), 男, 教授, 博士, 主要研究方向为图像增强与去噪、基于图像的缺陷检测与计算机视觉, E-mail: shuxin@just.eud.cn;

范燕 (1978−), 女, 副教授, 主要研究方向为计算机视觉与人工智能, E-mail: jsjxy_fy@just.edu.cn;

黄树成 (1969−)，男，教授，博士，主要研究方向为机器学习、表情识别与计算机视觉，E-mail: schuang@just.edu.cn;

史金龙 (1976−), 男, 教授, 博士, 主要研究方向为模式识别、协同视觉定位与 GPU 并行计算, E-mail: jsjxy_sjl@just.edu.cn.