

融合改进Transformer和卷积通道注意力模块的U-Net用于双心室分割

陈沐萱,袁金丽,郭志涛,卢成钢

河北工业大学电子信息工程学院,天津 300401

【摘要】设计一种融合改进Transformer和卷积通道注意力模块的U-Net用于MRI图像双心室分割。通过在U-Net的高层卷积部分基础融合改进Transformer,有效增强全局特征信息的提取能力以应对右心室复杂的形态变化造成低分割性能的难题。改进的Transformer在自注意力模块部分中加入固定窗口注意力进行位置定位,随后对其输出特征图进行聚合以缩小特征图尺寸;同时通过改进多层感知器来加深网络深度以提高网络学习能力。为解决组织边缘模糊造成的分割性能不理想问题,引入特征聚合模块进行多层次底层特征的融合,利用卷积通道注意力模块对底层特征进行重标定,实现自适应地学习特征权重。此外,针对编解码结构中通道衰减造成特征丢失导致的低分割性能,网络集成一个即插即用的特征增强模块,保证空间信息同时增加有用通道信息的比重。在ACDC数据集对本文算法进行测试,结果表明本文方法对左右心室的分割精度均优于近年其他算法,尤其是右心室分割结果,相比于其他方法,DSC系数提高至少2.83%,证明本文方法对双心室分割的有效性。

【关键词】双心室分割;图像处理;Transformer;注意力机制;特征提取

【中图分类号】R318;TP391

【文献标志码】A

【文章编号】1005-202X(2024)01-0032-11

Biventricular segmentation using U-Net incorporating improved Transformer and convolutional channel attention module

CHEN Muxuan, YUAN Jinli, GUO Zhitao, LU Chenggang

School of Electronic Information Engineering, Hebei University of Technology, Tianjin 300401, China

Abstract: A U-Net incorporating improved Transformer and convolutional channel attention module is designed for biventricular segmentation in MRI image. By replacing the high-level convolution of U-Net with the improved Transformer, the global feature information can be effectively extracted to cope with the challenge of poor segmentation performance due to the complex morphological variation of the right ventricle. The improved Transformer incorporates a fixed window attention for position localization in the self-attention module, and aggregates the output feature map for reducing the feature map size; and the network learning capability is improved by increasing network depth through the adjustment of multilayer perceptron. To solve the problem of unsatisfactory segmentation performance caused by blurred tissue edges, a feature aggregation module is used for the fusion of multi-level underlying features, and a convolutional channel attention module is adopted to rescale the underlying features to achieve adaptive learning of feature weights. In addition, a plug-and-play feature enhancement module is integrated to improve the segmentation performance which is affected by feature loss due to channel decay in the codec structure, which guarantees the spatial information while increasing the proportion of useful channel information. The test on the ACDC dataset shows that the proposed method has higher biventricular segmentation accuracy, especially for the right ventricle segmentation. Compared with other methods, the proposed method improves the DSC coefficient by at least 2.83%, proving its effectiveness in biventricular segmentation.

Keywords: biventricular segmentation; image processing; Transformer; attention mechanism; feature extraction

前言

根据世界卫生组织数据显示,心脏病是人类的主要致死疾病之一^[1]。准确有效的早期诊断对心脏病的治疗以及降低死亡率有重要的意义。在早期诊断中,依据心脏磁共振图像(Magnetic Resonance Imaging, MRI)的心室功能评估技术起着至关重要的

【收稿日期】2023-08-06

【基金项目】河北省教育厅重点项目(ZD2022115)

【作者简介】陈沐萱,硕士研究生,研究方向:计算机视觉、机器学习、医疗图像处理,E-mail: cmxstudent@163.com

【通信作者】袁金丽,博士,副教授,研究方向:智能信息处理、计算机视觉、机器学习,E-mail: jinli_yuan@hebut.edu.cn

作用^[2]。为获取心室容积、射血系数、心肌质量等临床心脏指标以进行心室功能评估,准确的左右心室分割是十分必要的^[3]。然而,由于心室形状不规则、边界模糊不清且尺寸变化大,双心室分割仍是一项复杂且具有挑战性的任务^[4-5]。

近年来,基于编解码器结构的网络已广泛应用于心脏分割领域。Ronneberger等^[6]通过跳跃连接将编码器的特征图与解码器中具有相同分辨率的特征图连接起来,提出U-Net,为医学图像分割领域提供一种有前途的架构。随后,Baumgartner等^[7]使用各种超参数测试2D U-Net和3D U-Net以及训练Dice损失与交叉熵损失的影响,结果表明在ACDC心脏MRI数据集上,2D U-Net在交叉熵损失训练下表现更好。自此,基于U-Net的改进在心脏分割领域中受到了许多研究者的青睐。改进方法包括重新设计跳跃连接^[8-9]、将残差结构引入U-Net^[10]、增加注意力模块^[11-12]以及重新设计U-Net型网络^[13-14]。改进的方法在一定程度上改善心脏分割性能,却忽略卷积运算固有局限性对性能的影响。堆叠卷积、残差密集结构先后被用于扩大感受野以打破卷积固有局限,但造成较大的计算成本,进而导致计算效率低下和网络优化困难^[15-16]。有研究使用扩张卷积来扩大感受野,然而,单一的扩张率可能只对一些大目标有效果,锯齿状结构的扩张率仍会导致计算成本增加^[10,17];另外,扩张卷积仍无法从全局角度捕捉上下文信息。这些方法能在一定范围内扩大感受野以提取全局信息,改善分割性能,但仍无法完全跳出卷积的固有局限。

Transformer结构能建模全局上下文信息,并且不受局部运算约束,在自然语言处理任务上有着出色的表现。Dosovitskiy等^[18]提出ViT(Vision Transformer),成功地将Transformer引入计算机视觉领域,在下游任务中表现出优于卷积神经网络(Convolutional Neural Network, CNN)方法的性能。随后,Zheng等^[19]根据基础ViT结构,实现将语义分割任务转化为序列形式的预测任务;但由于输出特征图的分辨率普遍较低且单一,模型丢失了大量局部信息。在心脏分割任务中,局部特征和全局特征同样重要,局部信息中物体的几何信息更充裕,而全局信息中物体的空间位置信息丰富。最近的研究倾向于结合CNN和Transformer,以充分发挥它们的优势,提高心脏分割性能。有研究在保持U-Net结构大致不变的情况下引入Transformer,造成较大的计算成本,进而影响图像处理速度^[20-23]。Swin-UNet使用基于窗口的自注意力机制完全替换U-Net中的卷积模块,降低原本自注意力模块机制的复杂度,以改善

心脏分割性能^[24]。Swin-UNet依赖于完备的预训练权重,导致实时应用受限,另外,滑动窗口存在重叠会造成特征信息冗余。

基于上述问题,本研究构建一种融合改进Transformer和注意力模块的U-Net用于双心室分割。针对心室尺寸变化大的问题,提出采用改进的Transformer模块(i-Trans模块)取代原有U-Net中的高层次卷积模块,以平衡提取全局及局部信息,应对心室的变化大的问题。i-Trans模块首先令特征图进行固定窗口注意力计算以增加其位置定位能力,其输出特征图根据窗口大小聚合以减小特征图尺寸,随后进行自注意力计算,提取全局特征;另外,为增加网络深度,调整Transformer中多层感知器(Multilayer Perceptron, MLP)深度。其次,采用特征聚合模块(Feature Aggregation Module, FAM)提取边缘特征信息,增强不同底层特征信息之间的信息交互。卷积通道注意(Convolutional Channel Attention, CCA)模块增强了底层特征信息中的有用信息,为网络提供信息过滤能力,提高了底层特征的利用率。FAM与CCA模块的结合可以有效地解决由边缘模糊造成分割结果不理想的问题。另外,为解决编解码器结构在特征恢复过程中由于信道减半造成的信息丢失问题,在网络中加入特征增强模块(Feature Enhancement Module, FEM)。

1 心脏分割算法

本研究构建一个融合改进Transformer和注意力模块的U-Net结构用于MRI双心室分割。整体网络结构如图1所示,主要包含i-Trans模块、FAM、CCA模块以及FEM。该网络基于编解码结构,在解码时使用跳跃连接来增强底层特征。i-Trans模块仅用于较小分辨率的高层特征图,以降低计算成本,而底层仍然使用卷积运算。卷积部分使用FAM关注局部信息,其中包含更多高分辨率细节,同时向模型引入一些结构先验,这有助于处理较小尺寸的医学图像数据集。跳跃连接加入CCA模块,以增强底层特征信息中的边缘信息,提高底层特征的利用率。FEM用于级联上采样,以缓解解码器恢复特征时由于信道减半造成的信息丢失对分割性能的影响。

对于给定的输入特征图 $F \in R^{H \times W \times C}$,网络的编码器部分每次通过卷积或i-Trans模块后,空间分辨率减半,信道数量加倍,最后得到 $\text{skip1} \in R^{H \times W \times C}$ 、 $\text{skip2} \in R^{H/2 \times W/2 \times 2C}$ 、 $\text{skip3} \in R^{H/4 \times W/4 \times 4C}$ 。编码器底层特征图被馈送到CCA模块,用于特征融合及增强,随后将其传输到相同分辨率的解码器以指导分割。解码器部分由i-Trans模块和级联的FEM组成。

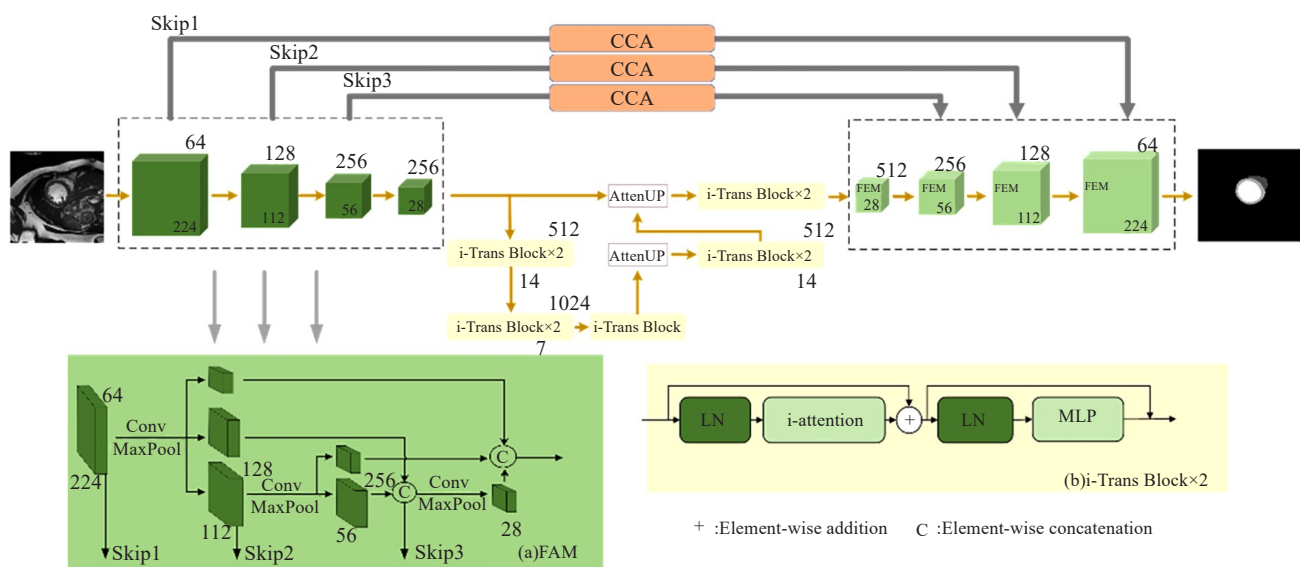


图1 网络结构模型

Figure 1 Network structure model

模块右上角为通道数,右下角为特征图尺寸

1.1 U-Net 结构

全局信息中物体的空间位置信息丰富,而局部信息中物体的几何信息更充裕,因此将局部信息与全局信息进行融合能提高分割准确性。本研究结合 i-Trans 模块与 U-Net,设计一种新的骨干网络,能有效提取全局信息及局部信息以应对心室尺寸变化大造成的分割性能不理想。鉴于之前诸多方法已经证明卷积结构提取底层特征信息的能力,因此本研究采用 i-Trans 模块替换 U-Net 的高层卷积部分,实现底层卷积模块提取底层纹理特征信息以及 i-Trans 模块提取高级语义信息。

传统 Transformer 模块包含两个主要部分,即自注意力模块和 MLP。自注意力模块用来提取全局信息,首先对输入图像进行线性映射后得到 query(*q*)、key(*k*)、value(*v*),其次根据 *q* 和 *k* 计算权重系数,最后根据权重系数对 *v* 进行加权求和。自注意力模块计算如下:

$$\text{Attention}(q,k,v) = \text{softmax}\left(\frac{qk^T}{\sqrt{d_k}}\right) \cdot v \tag{1}$$

其中, k^T 表示*k*的转置, d_k 表示*k*的维度。

当特征图尺寸较大时,传统的自注意力模块计算成本大。另外,单独的自注意力模块对特征图整体提取特征信息缺乏位置定位;而基于滑动窗口的自注意力模块存在窗口重叠,造成特征信息冗余。本研究基于传统 Transformer 模块设计 i-Trans 模块,其中 i-Attention 模块是 i-Trans 模块的核心(图 2)。i-Attention 模块融合固定窗自注意力模块以提供位置定位同时避免滑动窗口造成的特征冗余,按照固定窗口尺寸(win_size=4)对特征图进行聚合以减小特征图尺寸,随后进行全局自注意力计算。窗自注意力模块能为模型增加位置定位能力,有助于提取丰富有用信息。此外,i-Attention 模块利用窗口融合方式,在一定程度上减小自注意力模块的计算复杂度,进而缩短推理时间。传统的多头自注意力模块复杂度^[25]和 i-Attention 模块的复杂度见式(2)、(3):

$$\Omega(\text{MSA}) = 4hwC^2 + 2(hw)^2C \tag{2}$$

$$\Omega(\text{i-Attention}) = 4C^2\left(hw + \frac{hw}{M^2}\right) + 2C\left(M^2hw + \frac{hw}{M^4}\right) \tag{3}$$

其中,*h*、*w*为特征图尺寸,*C*表示通道数,*M*表示窗口大小。自注意力模块的计算通常是巨大的,会导致

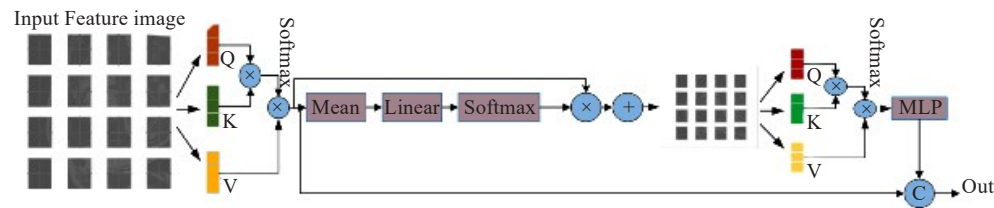


图2 i-Attention 模块

Figure 2 i-Attention module

模型推理速度缓慢,而 i-Attention 模块能在一定程度上降低计算复杂度,缓解计算压力。例如,针对尺寸为 14×14 以及通道数为 512 的特征图而言, i-Attention 模块的复杂度比传统的自注意力模块复杂度下降 10%。

本研究的 MLP 与传统 Transformer 不同,使用了 4 层全连接层,实现非线性和深度叠加,使得模型能保持传统深度网络的优势。由于上述优势,与传统自注意力模块相比, i-Attention 模块使模型在功能和计算性能方面更加友好,更有利于实际应用。

整个网络结构为编解码对称结构,编码过程中使用卷积模块实现下采样,并使用注意力上采样模块 (AttenUp 模块) 实现在解码过程中其对应的上采样。AttenUp 模块能有效增加特征交互能力(图 3)。将跳跃连接得到的 Skip 线性映射得到 Query,将上一级 i-Trans 模块结果经过线性映射得到 Key、Value,并且进一步使用自注意力计算过程,以此实现上采样。

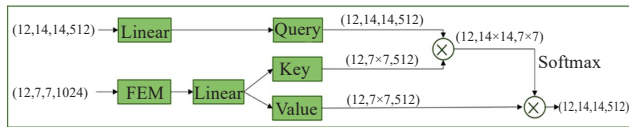


图 3 AttenUp 模块
Figure 3 AttenUp module

1.2 底层特征处理

针对心脏 MRI 图像边缘模糊导致的分割性能不理想问题,本研究使用 FAM 和 CCA 模块对包含大量边缘信息的编码器底层特征图进行有效利用。FAM 能提取边缘特征信息,同时实现不同底层特征信息之间的信息交互;CCA 模块增强底层特征信息中的纹理信息,为跳跃连接提供信息过滤能力,以更好地指导编码器恢复分辨率。

1.2.1 FAM 在卷积神经网络中,底层卷积提取的物体的几何信息中包含大量的轮廓边缘信息,对分割任务来说至关重要,充分利用底层纹理特征能有效解决边缘模糊造成的分割不理想问题。本研究的编码器底层卷积部分利用 FAM 有效融合不同大小的底层特征图,减小卷积层的参数误差,保留大量的边缘信息。另外,与传统卷积模块相比, FAM 可以通过最大池化来弥补由于不同特征映射而产生的特征偏移。最大池化类似特征选择,选择辨识度更好的特征能为模型增加一定量的非线性映射能力。

FAM 通过不同池化窗口的池化操作,保留了不同尺寸的特征信息。其中第一层使用的池化窗口大小分别为 2、4、8,第二层为 2、4,第三层则为 2。随后,采用拼

接操作对相同分辨率的特征图进行融合,以保存不同深度特征信息,进一步通过跳跃连接辅助分割。

为提取更丰富的底层纹理信息,编码器底层特征提取部分加入 FAM,通过对特征图进行不同尺度的池化以便进行特征融合。在特征融合过程中, FAM 能加强不同层次的特征交互,以更好地保留并提取纹理信息。

1.2.2 CCA 传统的跳跃连接结构将编码器提取的特征直接连接到解码器对应的同分辨率特征层,而由于缺乏多级信息交互以及特征筛选能力,因此在面对器官大小及形状多变的问题时,其性能有限。鉴于通道和空间注意力机制在特征提取上的良好表现,本研究在跳跃连接部分使用 CCA 模块,利用不同尺寸的卷积核进行多尺度特征融合,随后,使用注意力模块对特征进行通道维度的特征增强,让网络更关注有用的特征信息。

如图 4 所示, CCA 模块输入一个特征图 F_i , 该特征图来自底层编码器,包含丰富的纹理信息。为使解码器能充分利用有用的纹理信息进行特征恢复, CCA 模块通过对输入特征图进行不同尺寸的卷积以融合多尺度空间信息,随后进行通道注意力计算实现特征增强。首先,对输入特征图 F_i 分别进行 3×3 和 5×5 的卷积以获得多尺度特征信息。卷积操作之后利用填充操作调整输出尺寸,使输出与输入尺寸一致,随后将卷积结果与输入进行像素级加法操作。其结果初步融合了多尺度底层纹理信息,如式(4)所示:

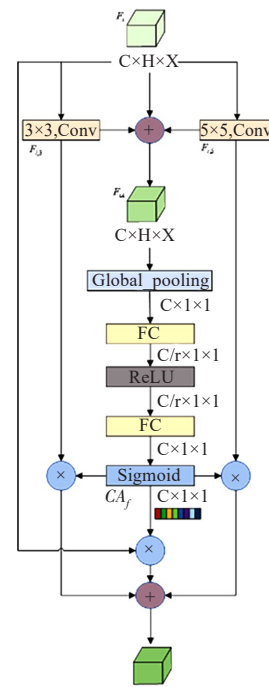


图 4 卷积通道注意力模块
Figure 4 Convolutional channel attention module

$$F_{i4} = F_i + F_{i,3} + F_{i,5} \quad (4)$$

其中, F_i 表示编码器第 i 层特征图, $F_{i,3}$ 、 $F_{i,5}$ 分别表示输入特征进行 3×3 、 5×5 卷积的结果, F_{i4} 表示特征进行像素级加法的结果。

其次,对 F_{i4} 进行通道注意力计算。先利用全局平均池化对输入沿着通道维度进行压缩,紧接着利用两个具有缩减比率 r 的全连接(Fully Connected, FC)层来限制模型复杂性并改进泛化,并且第一个 FC 层后跟随 ReLU 激活函数。FC 层输出即为 CA_f , 经过 Sigmoid 函数后分别与卷积结果和输入相乘,将得到的 3 个结果进行加法操作:

$$F_{oi} = CA_f \times F_i + CA_f \times F_{i,3} + CA_f \times F_{i,5} \quad (5)$$

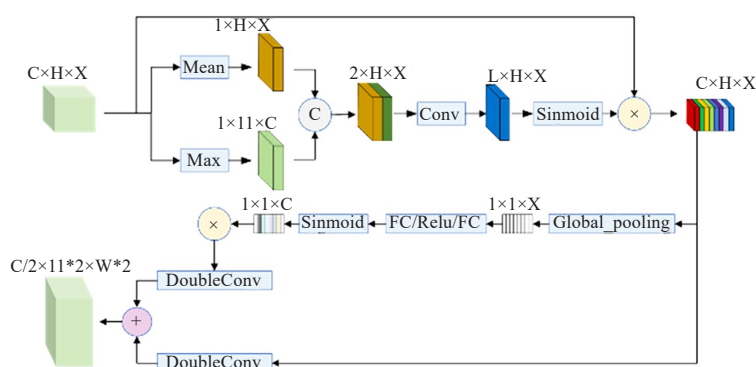


图5 特征增强模块

Figure 5 Feature enhancement module

FEM 能在保证空间信息的基础上,减小通道减半造成信息丢失对分割性能的影响。FEM 先对特征图的空间位置信息进行增强,以减少通道减半对空间信息的影响;进行通道信息增强后输入到卷积模块中进行通道减半。

首先,输入特征图通过使用平均池化和最大池化操作来聚合特征图的空间信息,随后将融合特征图进行卷积,得到空间注意力图,其尺寸为 $(1, H, W)$ 。对空间注意力图使用 Sigmoid 函数得到空间增强后的输出特征图。空间注意力增强过程见式(6):

$$F_s = \text{Sigmoid}(\text{Conv}(\text{Mean}(F) + \text{Max}(F))) \times F \quad (6)$$

其中, F 表示输入特征图, F_s 表示空间注意力增强后的输出特征图。

其次,对 F_s 使用两个并行操作。一方面直接进行上采样;另一方面,进行通道维度增强,强调有用特征以减少通道减半造成的特征损失对分割性能的影响。在通道维度增强过程中, F_s 先进行全局平均池化,再通过两层全连接层得到通道注意力图,其尺寸为 $(C, 1, 1)$ 。对通道注意力图使用激活函数得到通道增强后的输出特征图。最后,对 F_c 进行卷积实

通过对融合的不同尺度底层特征图进行通道注意力增强, CCA 为编码器提供更有辨别能力的特征信息,有利于提高心室分割的性能。

1.3 FEM

编解码结构网络试图在编码阶段提取尽可能多的特征信息,同时在解码阶段保留尽可能多的提取特征,并将它们恢复到输入图像的分辨率。然而,传统的编解码器结构在解码阶段由于信道数减半而导致特征丢失,进而影响心室分割结果。参考 CBAM 结构设计^[26],并结合心室分割实际需求,本研究使用 FEM 以减少信道减半造成的特征损失,如图 5 所示。

现上采样。其中通道注意力增强过程见式(7):

$$F_c = \text{Sigmoid}(\text{FC}(\text{Global_Pooling}(F_s))) \times F_s \quad (7)$$

其中, F_c 表示空间注意力增强后的输出特征图。

FEM 的输出即两个并行操作结果相加:

$$F_o = \text{DoubleConv}(F_s) + \text{DoubleConv}(F_c) \quad (8)$$

其中, F_o 为 FEM 模块的输出, DoubleConv 为串联两层 3×3 卷积操作。

FEM 利用空间注意力关注空间位置信息,同时,使用通道注意力强调高级语义信息。FEM 基于空间和通道维度对特征细化后的结果进行上采样,能在保持开销较小的同时实现分割性能改进。

1.4 损失函数

损失函数由 Dice 相似系数损失和交叉熵损失两部分组成,其中, Dice 损失是为了增强医学图像中细小结构的分割能力,而交叉熵损失函数能针对每个像素点进行损失计算。

总的损失函数表达式为:

$$L = 0.5 \times L_{CE} + 0.5 \times L_{Dice} \quad (9)$$

交叉熵损失函数相对于传统的均方差损失而言,能避免 Sigmoid 导数造成的参数更新缓慢问题。

另外,交叉熵损失函数逐像素地计算预测值与真值的距离,更适合像素级分类的语义分割问题。交叉熵损失函数定义为:

$$L_{CE} = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic})$$

(10)

其中, M 表示类别数; y_{ic} 表示符号函数,如果 i 的类别数为 c , 取 1, 否则取 0; p_{ic} 表示观测样本 i 属于类别 c 的概率。

Dice 相似系数损失能缓解样本中不平衡造成的负面影响。使用 Dice 相似系数损失训练时,模型更关注前景区域而忽略背景,有导致损失饱和的可能。因此 Dice 相似系数损失单独使用不能取得理想性能,常常与其他损失函数结合使用,本研究结合 Dice 相似系数损失与交叉熵损失函数。Dice 相似系数损失函数定义为:

$$L_{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|}$$

(11)

其中, X 为真值, Y 为预测值。

2 实验及分析

所有实验都在 NVIDIA GeForce RTX 2080Ti GPU 上的 Pytorch 环境下运行。对模型均使用 Adam 优化器进行优化,其中学习率设置为 $1e-4$, batch_size 为 6, epoch 为 100。实验结果分析包括对损失函数以及模型相关参数的分析、本文网络与其他网络结构进行性能对比测试以及不同结构下的网络性能比较。

2.1 数据集及评价指标

ACDC 2017 数据集采用两台不同磁强(1.5T 和 3.0T)的 MRI 扫描仪进行采集^[27]。其中,一系列短轴切片覆盖心脏,从左心室基部到心尖,切片厚度为 5~8 mm,片间间隙为 5~10 mm。短轴平面内空间分辨率为 $0.83 \sim 1.75 \text{ mm}^2/\text{像素}$ 。该测试集包含 100 例患者的心脏 MRI 公共数据集,包括 3 个标签,分别为右心室、心肌、左心室。实验中,将 70 例患者(1304 张图像)用于训练,10 例(182 张图像)用于测试,其余 20 例(404 张图像)用于验证。输入图像分辨率裁剪为 224×224 。网络在训练过程中进行数据增强,包括随机翻转、随机旋转以及图像的比例缩放等操作,增强后的训练数据量为 5216 张。

为准确地评估分割的准确性,本研究使用 Dice 相似系数(Dice Similarity Coefficient, DSC)^[28]和豪斯多夫距离(Hausdorff Distance, HD)^[20]作为主要评价指标,定义如下:

$$HD = \max \{ h(X, Y), h(Y, X) \}$$

(12)

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

(13)

其中, X 为真值, Y 为预测值。

计算 HD 时,首先将点集合 X 中的每个点与集合 Y 中离该点最近的点之间的距离排列,然后取该距离的最大值作为 HD 的值。HD 对分割边界敏感,因此主要用于图像分割任务。HD 越小,分割效果越好。为排除一些离群点造成的不合理距离,保持整体数值稳定性,本研究使用 HD95 值(从小到大排名前 95% 的距离)进行性能评估。

DSC $\in[0, 1]$ 是医学图像分割中常用来评价地面真相与分割结果相似性的评价矩阵。DSC 值越高,图像分割效果越好。

2.2 消融实验

为评估本文方法并验证其在不同设置下的性能,设置一系列的消融实验,包括:CCA 模块的数量及位置、i-Trans 模块数量、i-Trans 结构有效性以及各个模块对于整体框架的必要性。

2.2.1 不同位置及数量的跳跃连接 跳跃连接部分加入 CCA 模块,试图以特征增强的方式为模型添加特征筛选能力,进一步削弱医学图像中器官形状及大小多变对分割带来的影响。为测试不同位置的跳跃连接对模型的影响,在 1/2、1/4、1/8 分辨率处依次添加 CCA 模块,进一步组合这 3 个位置的 CCA,探索 CCA 的最优设置。表 1 展示了不同位置以及不同组合下的分割结果。结果表明分割结果与 CCA 模块的设置密切相关,改变其设置均呈现不同的变化。作为分割任务,DSC 是衡量结果好坏的主要参考指标,在 1/2、1/4、1/8 分辨率处同时添加 CCA 模块时,分割性能表现最佳。底层特征图中包含大量的边缘信息,然而边缘信息所占比重较小,为增大边缘信息比重,模型加入 CCA 模块能有效增加有用的边缘信息的比重,进一步帮助还原心室边界,从而取得良好的分割结果。

表 1 不同位置加入 CCA 模块的实验结果
Table 1 Experimental results of adding CCA modules at different locations

序号	位置	DSC/%	HD95/mm
1	Skip1	91.38	1.24
	Skip2	91.18	1.44
	Skip3	91.48	1.24
2	Skip1,2	91.69	1.23
	Skip1,3	91.53	1.23
	Skip2,3	91.50	1.28
3	Skip1,2,3	91.82	1.22

2.2.2 i-Trans模块的数量 i-Trans模块被用来替代U-Net高级卷积部分。为设置合理的i-Trans模块数量,本研究进行对应的消融实验,如表2所示。结果表明两层i-Trans模块取得最佳性能,因此本实验中均使用两层i-Trans模块。

表 2 i-Trans 模块不同数量实验结果
Table 2 Experimental results with different quantities of i-Trans modules

数量	DSC/%	HD95/mm
1	91.03	2.12
2	91.82	1.22
3	91.80	1.25
4	76.38	3.21

2.2.3 i-Trans模块的不同结构 i-Trans模块由3部分组成,即基于窗的自注意力模块、自注意力模块以及MLP。对3个模块进行消融实验评估其性能,如表3所示。结果发现本研究的i-Trans模块取得了最佳的心脏分割性能。相比于传统的MLP,包含4个全连接层的MLP能取得更好的分割性能。全连接层能加深i-Trans模块的网络深度,进一步增加网络学习能力,但是全连接层的数量直接导致计算成本的增大,因此必须探索合理结构以达到最佳性能。针对i-Trans模块各个模块实验后发现,完整的i-Trans模块能取得最佳性能,证明了该模块的有效性。

表 3 i-Trans 模块的不同结构实验结果(%)
Table 3 Experimental results with different structures of i-Trans modules (%)

方法	DSC
W+S+DM	90.87
W+M	91.09
S+M	91.48
W+S+M	91.82

W表示基于窗的自注意力模块,S代表自注意力模块,M代表本研究提出的包括4个全连接层的MLP模块,DM代表传统的包括两个全连接层的MLP模块

2.2.4 各个模块对于整体框架的作用 以U-Net为基础,依次加入i-Trans模块、FAM、CCA模块以及FEM,评估各个模块的必要性,如表4所示。随着各个模块的依次加入,分割结果有所改善,完整的方法取得最佳的平均DSC及HD结果,证明了各个模块对于改善分割性能的有效性。

表 4 各个模块对整体网络的有效性实验
Table 4 Effectiveness experiments of each module on the whole network

模块	DSC/%	HD95/mm
ViT Encoder ^[20]	89.38	2.54
i-Trans	91.33	2.11
i-Trans+FAM	91.58	1.35
i-Trans+FAM+CCA	91.71	1.28
All	91.82	1.22

2.3 实验结果分析

2.3.1 不同损失函数的对比分析 本文模型的训练效果与损失函数的设置有关,因此为获得最合适的损失函数以及最好的训练效果,本研究针对交叉熵损失函数、Dice相似系数损失函数和组合损失函数进行对比实验。

图6第1行展示了在3种不同损失函数下的模型训练效果,采用组合损失函数可以达到最优的训练结果。其原因在于心室面积相对于MRI图像面积,占比较小,而使用单一的损失函数进行优化时,心室部分对应的梯度变化很容易受到其他背景区域梯度的影响,从而导致网络训练困难。组合损失函数综合了两种损失函数的特点,在网络反向传播过程中能对难以学习的样本进行稳定且有针对性的优化,从而缓解正负样本不平衡的问题,提高训练性能。图6第2行展示了3种损失函数情况下取得的验证数据集的平均Dice分数与迭代次数变化关系,进一步证明组合损失函数能取得更好的训练效果。在组合损失函数训练的情况下,验证数据集的平均Dice分数随着迭代次数增加而增大;而交叉熵损失函数以及Dice相似系数损失函数训练下,验证集的平均Dice分数趋于稳定。

2.3.2 不同参数设置的对比分析 为了寻找最佳参数设置,对不同窗口大小的窗自注意力模块进行分析,比较每秒10亿次的浮点运算数(Giga Floating-point Operations Per Second, GFLOPs)、DSC以及HD值,其结果如表5所示。

当窗口大小为4时,模型取得相对较小的参数量,同时HD值最小。相对于窗口为5时的结果,窗口为4时的模型参数量下降了0.09 GFLOPs,且HD95值下降了0.09 mm,同时DSC相差不大。参照实验环境,在保证计算量和分割性能的基础上,本研究选择大小为4的窗口进行实验。

2.3.3 底层特征处理结果分析 为验证FAM以及CCA模块对底层特征信息加强的效果,本研究对底层特征图进行可视化对比,如图7所示。第1行为未

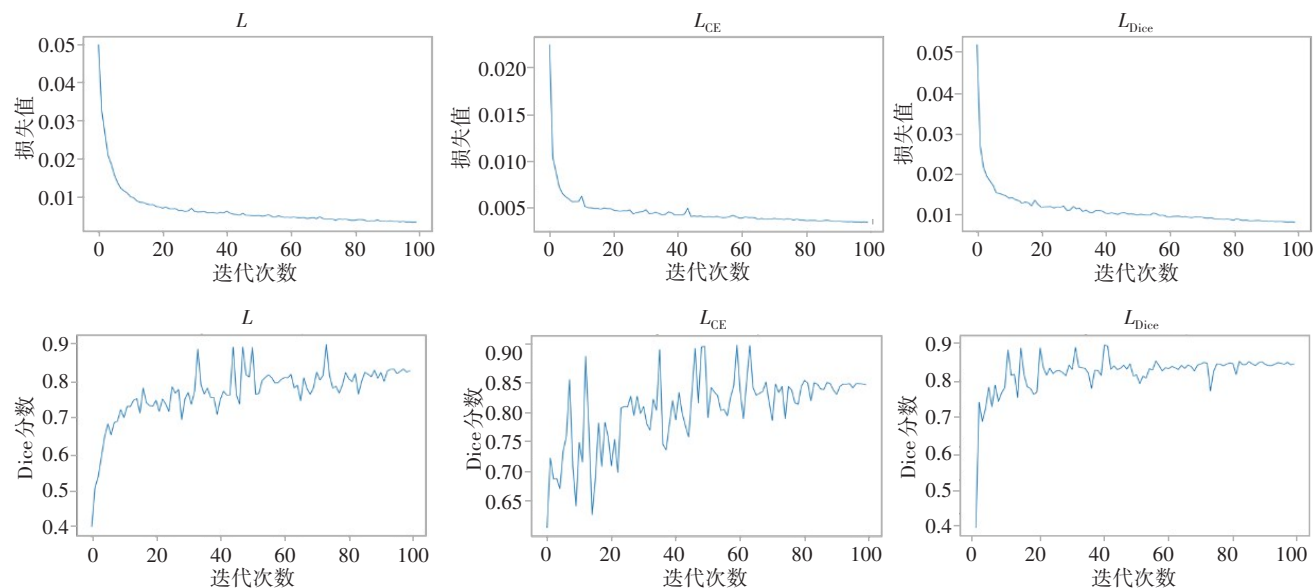


图6 损失值和验证集Dice分数与迭代次数变化

Figure 6 Changes in loss value and validation dataset Dice score over epochs

表5 不同窗口大小结果对比

Table 5 Comparison of the results at different window sizes

窗口大小	GFLOPs	DSC/%	HD95/mm
2	103.68	91.89	1.40
3	103.80	91.87	1.24
4	103.75	91.82	1.22
5	103.84	91.91	1.26
6	104.06	91.51	2.12

使用底层特征处理的特征图,第2行则为本文方法提取的特征图。经过特征处理后,纹理特征更丰富清晰,证明特征处理能明显加强边缘特征信息。对于较深层次的特征图(第3列),特征处理能有效加强边缘特征对比度,增强特征可辨识度。因此,FAM 和 CCA 模块结合可有效地增强边缘细节特征,以缓解由边缘模糊造成的分割结果不理想。

2.3.4 分割性能对比及分析 为验证本文方法在心脏图像上的分割性能,将本文方法与现有的方法(R50 UNet^[6]、R50 AttnUNet^[11]、ViT-CUP^[18]、R50 ViT^[18]、TransUNet^[20]、Swin-UNet^[24])进行对比。其对比结果如表6所示(对比方法的数据来源于MT-UNet^[29])。

本文方法在验证集上取得的平均 DSC 为 91.82%,相比之前提出的一系列医学图像分割网络性能均有较大的提升。与原始的 R50 UNet 相比,平均 DSC 提高 4.22%,其中右心室、心肌、左心室的 DSC

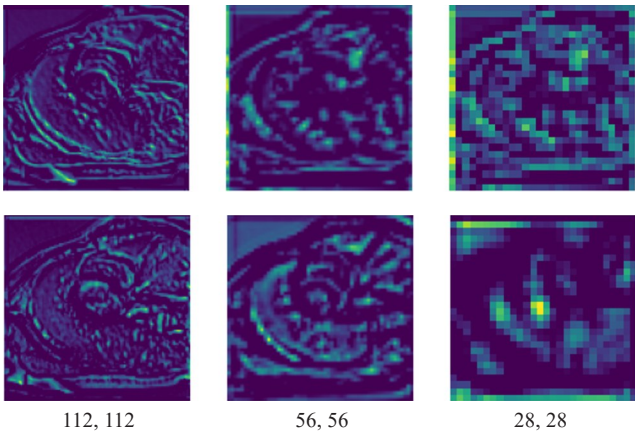


图7 底层特征处理前(上行)后(下行)特征图对比

Figure 7 Comparison of feature maps before and after underlying feature processing

表6 不同分割方法DSC对比(%)

Table 6 Comparison of DSC among different segmentation methods (%)

方法	右心室	心肌	左心室	平均DSC
R50 UNet	84.62	84.52	93.68	87.60
R50 AttnUNet	83.27	84.33	93.53	86.90
ViT-CUP	80.93	78.12	91.17	83.41
R50 ViT	82.51	83.01	93.05	86.19
TransUNet	86.67	87.27	95.18	89.71
Swin-UNet	85.77	84.42	94.03	88.07
MT-UNet	86.64	89.04	95.62	90.43
本文	89.47	90.01	95.90	91.82

分别提升 4.85%、5.49%、2.22%；对比 MT-UNet 方法也有一定的提升,平均 DSC 提升 1.39%,右心室、心肌及左心室分别提升 2.83%、0.97%、0.28%。实验结果表明本文方法能改善对心脏双心室的分割结果;对于较难分割的右心室而言,本文方法相比于 MT-UNet 方法提高 2.83%,右心室分割精度明显改善。

为证明本文方法对于形状、尺寸多变的右心室的分割效果,选择 6 个不同的模型分割图像的可视化

结果,如图 8 所示。图 8a 是医生标注的心脏结构轮廓金标准,图 8g 是使用本文方法对心脏分割的结果。与之前方法相比,本文方法的 DSC 达到 91.82%,说明本文方法在心脏分割上更精确。在可视化结果图中,特别是第 1 行和第 4 行表明本文方法对于右心室的分割结果更完整,能有效解决右心室尺度变化大的问题,提高右心室分割精度。

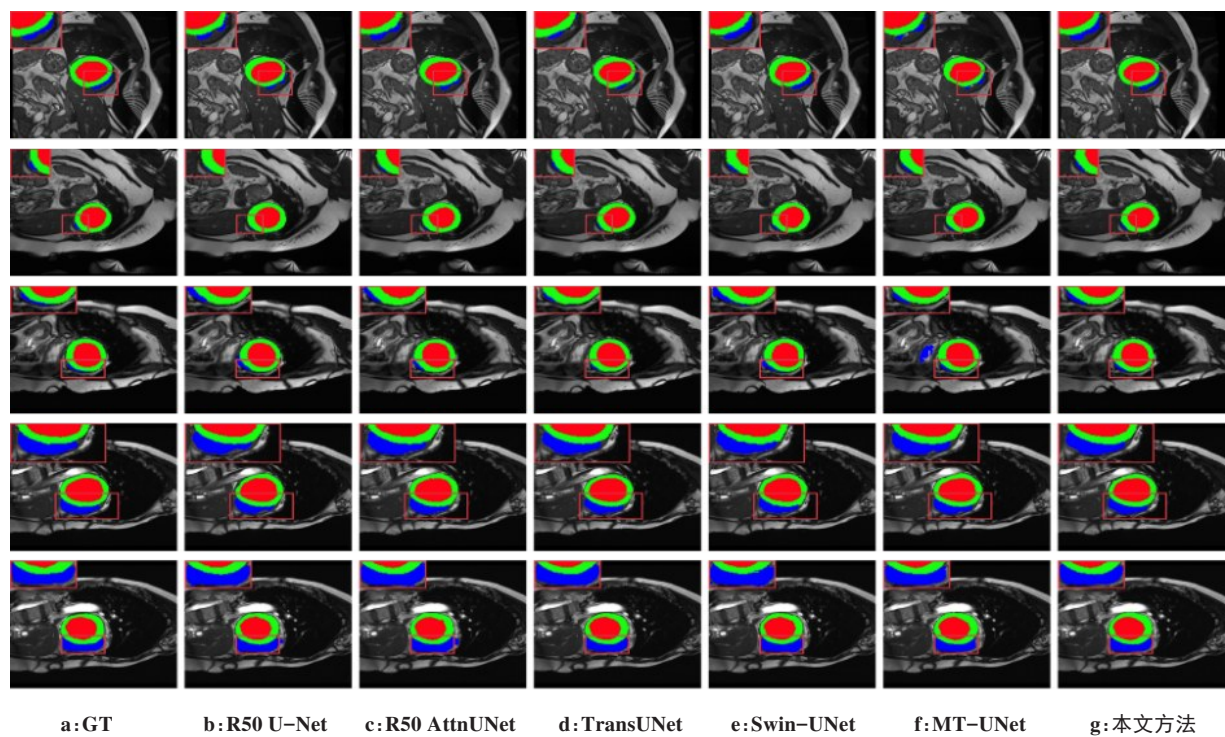


图 8 不同尺度的右心室分割结果对比

Figure 8 Comparison of right ventricular segmentation results

蓝色区域表示右心室,红色区域表示左心室,绿色区域表示心肌

为验证本文方法对存在边缘模糊的心脏 MRI 图像的分割效果,对不同模型结果进行可视化对比,如图 9 所示。可以看出,几种方法对于存在边缘模糊的 MRI 图像有不同程度的欠分割和过分割现象。相比之下,本文方法能更有效地区分心室与其他组织区域,对心室的分割更精确。另外,观察可视化结果可以发现本文方法得到的分割边界更加准确,能有效缓解边缘模糊造成的分割性能不理想现象。

3 结 论

针对心室形状尺寸多变以及组织边界模糊等难点问题,本研究提出一种融合改进 Transformer 和注意力模块的 U-Net 用于双心室分割。该算法结合 i-Trans 模块和 U-Net 网络,以提取丰富的全局特征和

局部特征。i-Trans 模块中让特征图依次进行固定窗口注意力计算和全局自注意力计算以增加位置定位能力,同时,提取全局特征信息。另外,该模型融合了注意力机制,在增大信息提取能力和编码多尺度特征信息的同时对特征进行重标定,使网络学习关注有用特征,以增强对不同尺寸心室图像的特征提取能力,进而提高心室分割性能。在 ACDC 数据集上验证和评估本文方法的分割性能,结果表明与已广泛用于医学图像分割的 Swin-UNet 相比,其 DSC 提高了 3.75%,也高于最新的心脏分割网络 MT-UNet。本研究为解决双心室分割中的难点问题提供了有价值的研究方向,并具有一定的临床应用价值。

尽管本文方法在心室分割任务中取得了一些优势,但仍然存在一些不足,例如 Transformer 结构中的

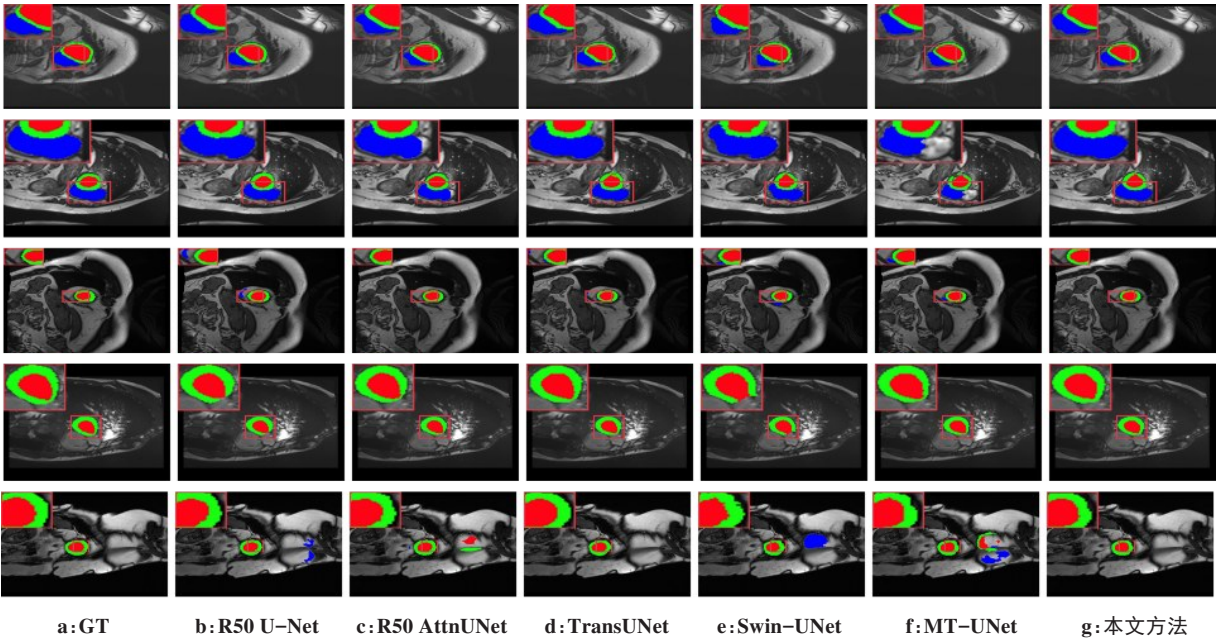


图9 存在边缘模糊的心脏MRI分割结果对比

Figure 9 Comparison of MRI cardiac segmentation results in the presence of blurred edges

蓝色区域表示右心室,红色区域表示左心室,绿色区域表示心肌

大量计算增加了模型的推理时间,在未来的工作中,网络会继续优化非目标区域的计算量,从而提升模型的推理速度;其次,下一步还将聚焦于增强模型泛化能力,通过改进本文方法来解决其他医学图像分割数据集中存在的问题。

【参考文献】

[1] Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke statistics-2023 update: a report from the American Heart Association [J]. Circulation, 2023, 147(8): e93-e621.

[2] 刘从军, 徐佳陈, 肖志勇, 等. 基于深度学习的心脏核磁共振图像自动分割算法[J]. 计算机工程与科学, 2022, 44(9): 1646-1654.

Liu CJ, Xu JC, Xiao ZY, et al. An automatic cardiac magnetic resonance image segmentation algorithm based on deep learning[J]. Computer Engineering and Science, 2022, 44(9): 1646-1654.

[3] Li FY, Li W, Gao X, et al. A novel framework with weighted decision map based on convolutional neural network for cardiac MR segmentation[J]. IEEE J Biomed Health Inf, 2022, 26(5): 2228-2239.

[4] 苏新宇, 王丽嘉, 聂生东, 等. 基于心脏磁共振短轴电影图像的右心室分割新进展[J]. 波谱学杂志, 2019, 36(3): 377-391.

Su XY, Wang LJ, Nie SD, et al. Progress of right ventricle segmentation from short-axis images acquired with cardiac cine MRI [J]. Chinese Journal of Magnetic Resonance, 2019, 36(3): 377-391.

[5] Chen C, Qin C, Qiu H, et al. Deep learning for cardiac image segmentation: a review[J]. Front Cardiovasc Med, 2020, 7(25): 1-33.

[6] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.

[7] Baumgartner CF, Koch LM, Pollefeys M, et al. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation [C]//Statistical Atlases and Computational Models of the Heart. ACDC and MMWS Challenges: 8th International Workshop, 2018: 111-119.

[8] Zhou Z, Siddiquee MM, Tajbakhsh N, et al. UNet++: redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Trans Med Imaging, 2019, 39(6): 1856-1867.

[9] 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.

[10] Zhang H, Zhang W, Shen W, et al. Automatic segmentation of the cardiac MR images based on nested fully convolutional dense network with dilated convolution[J]. Biomed Signal Proces Control, 2021, 68: 102684.

[11] Cui H, Yuwen C, Jiang L, et al. Multiscale attention guided U-Net architecture for cardiac segmentation in short-axis MRI images[J]. Comput Meth Prog Biomed, 2021, 206: 106142.

[12] Wang X, Yang S, Fang Y, et al. SK-UNet: an improved U-Net model with selective kernel for the segmentation of LGE cardiac MR images [J]. IEEE Sensors J, 2021, 21(10): 11643-11653.

[13] Shi J, Ye Y, Zhu D, et al. Automatic segmentation of cardiac magnetic resonance images based on multi-input fusion network[J]. Comput Meth Prog Biomed, 2021, 209: 106323.

[14] Du X, Xu X, Liu H, et al. TSU-Net: two-stage multi-scale cascade and multi-field fusion U-Net for right ventricular segmentation [J]. Comput Med Imaging Grap, 2021, 93: 101971.

[15] Bi L, Kim J, Kumar A, et al. Stacked fully convolutional networks with multi-channel learning: application to medical image segmentation[J]. Visual Comput, 2017, 33: 1061-1071.

[16] Li C, Song X, Zhao H, et al. An 8-layer residual U-Net with deep supervision for segmentation of the left ventricle in cardiac CT angiography[J]. Comput Meth Prog Biomed, 2021, 200: 105876.

[17] Ahmad F, Hou W, Xiong J, et al. Fully automated cardiac MRI segmentation using dilated residual network[J]. Med Phys, 2022, 50 (4): 2162-2175.

[18] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16× 16 words: transformers for image recognition at scale [J]. arXiv preprint arXiv:2010.11929, 2020.

[19] Zheng S, Lu J, Zhao H, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 6881-6890.

[20] Chen J, Lu Y, Yu Q, et al. TransuNet: transformers make strong encoders for medical image segmentation[J]. arXiv preprint arXiv:

- 2102.04306, 2021.
- [21] 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: 24th International Conference, 2021: 14-24.
- [22] 李擎, 皇甫玉彬, 李江昀, 等. UConvTrans: 全局和局部信息交互的双分支心脏图像分割[J/OL]. 上海交通大学学报, 2022, 56(9): 1-12. Li Q, Huangfu FB, Li JY, et al. UConvTrans: global and local information interaction for two-branch cardiac image segmentation [J/OL]. Journal of Shanghai Jiaotong University, 2022, 56(9): 1-12.
- [23] Wang H, Cao P, Wang J, et al. UCTransNet: rethinking the skip connections in U-Net from a channel-wise perspective with transformer [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(3): 2441-2449.
- [24] Cao H, Wang Y, Chen J, et al. Swin-Unet: UNet-like pure transformer for medical image segmentation [C]//Computer Vision-ECCV 2022 Workshops, 2023: 205-218.
- [25] 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: 10012-10022.
- [26] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [27] Bernard O, Lalande A, Zotti C, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? [J]. IEEE Trans Med Imaging, 2018, 37(11): 2514-2525.
- [28] Abdollahi A, Pradhan B, Alamri A. VNet: an end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data [J]. IEEE Access, 2020, 8: 179424-179436.
- [29] Wang H, Xie S, Lin L, et al. Mixed transformer u-net for medical image segmentation [C]//ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing, 2022: 2390-2394.

(编辑:谭斯允)