基于混洗特征编码与门控解码的医学图像分割网络

雷 涛[1,2],张峻铭[1,2],杜晓刚[1,2*],闵重丹[1,2],杨子瑶[1,2]
(1. 陕西科技大学电子信息与人工智能学院,陕西西安 710021;2. 陕西省人工智能联合实验室(陕西科技大学),陕西西安 710021)

摘 要: 针对医学图像分割领域长期存在的多目标尺度变化大和边界模糊以致分割困难的问题,提出了一种新型的基于混洗特征编码和门控解码的双分支混合网络框架用于多器官精准分割. 为了充分利用卷积神经网络(Convolutional Neural Network,CNN)在局部信息提取方面和Transformer在长程依赖关系建模方面的优势,采用U-Net和Swin-Unet构建双分支网络. 该方法的创新之处在于对不同网络分支的多个阶段学习到的高维特征进行混洗操作,通过双支路通道交叉融合的方式实现局部信息与全局信息的高效融合,加强了双分支网络在不同阶段间的信息交互,从而解决了图像目标轮廓模糊引起的分割精度受限的问题. 此外,为了解决多器官尺度变化大的问题,进一步提出了一种全新的基于多尺度特征图的门控解码器(Gated Decoder based on Multi-scale Feature,GDMF). 该解码器能够学习网络不同阶段的多尺度高维特征并进行自适应特征增强,采用注意力机制和特征映射来辅助获取精准目标信息. 实验结果表明,与现有主流医学图像分割方法相比,所提方法在ACDC(Automated Cardiac Diagnosis Challenge)和FLARE21(Fast and Low GPU memory Abdominal oRgan sEgmentation challenge 2021)数据集上均表现出更优的性能,有效解决了医学图像中多目标尺度变化大和边界模糊问题.

关键词: 医学图像分割;CNN-Transformer混合架构;混洗特征编码;门控解码

基金项目: 国家自然科学基金(No.62271296,No.62201334);陕西省杰出青年科学基金(No.2021JC-47);陕西省重点研发计划(No.2022GY-436,No.2021ZDLGY08-07);陕西省创新能力支撑计划(No. 2020SS-03);陕西省教育厅科学研究计划项目(No.23JP014,No.23JP022)

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2024)12-4142-11

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20231011

Medical Image Segmentation Network Based on Shuffled Feature Encoding and Gated Decoding

LEI Tao[1,2], ZHANG Jun-ming[1,2], DU Xiao-gang[1,2*], MIN Chong-dan[1,2], YANG Zi-yao[1,2]
(1. *School of Electronic Information and Artificial Intelligence*, *Shaanxi University of Science and Technology*, *Xi'an*, *Shaanxi* 710021, *China*;
2. *Shaanxi Joint Laboratory of Artificial Intelligence*, *Shaanxi University of Science and Technology*, *Xi'an*, *Shaanxi* 710021, *China*)

Abstract: To solve the long-standing problems of the great scale variation in target sizes and blurred boundaries that make segmentation difficult in medical image segmentation, we propose a novel dual-branch hybrid network framework based on feature encoding and gated decoder based on multi-scale feature for accurate multi-organ segmentation. In order to fully exploit the strengths of convolutional neural network (CNN) in local information extraction and transformers in modeling long-range dependency, we employ U-Net and Swin-Unet to construct the dual-branch network. The innovation of this method lies in the shuffling operation of high-dimensional features extracted at multiple stages from different branches of the network. It efficiently integrates local and global information by means of a dual-branch channel cross-fusion, enhancing information interaction between the dual-branch network at different stages. This addresses the limitation in segmentation accuracy caused by the blurring of object contours in images. Additionally, to address the challenge of great scale variation among multiple organs, we introduce a new gated decoder based on multi-scale feature (GDMF) to extract multi-scale high-dimensional features at different stages of the network and perform adaptive feature enhancement, and adopts the attention mechanisms and feature mappings to assist in acquiring accurate target information. The experimental results on automated cardiac diagnosis challenge (ACDC) and fast and low GPU memory abdominal organ segmentation challenge 2021

收稿日期:2023-10-27;修回日期:2024-08-07;责任编辑:朱梅玉
*通讯作者:杜晓刚

(FLARE21) datasets demonstrate that our proposed method outperforms existing mainstream medical image segmentation methods and effectively solves the problems of the great scale variation in target sizes and blurred boundary in medical images.

Key words: medical image segmentation; CNN-Transformer architecture; shuffle feature encoding; gated decoding

Foundation Item(s): National Natural Science Foundation of China (No.62271296, No.62201334); Shaanxi Province Outstanding Youth Science Fund (No. 2021JC-47); Shaanxi Province Key Research and Development Program (No.2022GY-436, No.2021ZDLGY08-07); Shaanxi Province Innovation Capability Support Program (No.2020SS-03); Scientific Research Program Funded by Shaanxi Provincial Education Commission (No.23JP014, No.23JP022)

1 引言

图像分割作为医学图像分析的重要组成部分,对于计算机辅助诊断、治疗以及手术等具有重要意义. 然而,由于成像方法的局限性,医学图像普遍存在对比度低、边界模糊、噪声强和目标尺度变化大等问题. 这些问题导致传统的图像分割方法分割精度较差,且受噪声影响较大,分割效果很难令人满意. 随着深度学习的不断发展,深度神经网络能有效提取图像特征,并且对于噪声具有较强的鲁棒性,因此被广泛应用于医学图像分割领域. 目前,基于卷积的神经网络取得了优异的医学图像分割性能,例如U-Net[1]、V-net[2]等. 然而,传统卷积使用滑动窗口策略,局限于提取图像的局部特征,难以有效捕捉全局语义信息,在处理医学图像时,使用有限感受野提取的特征往往存在偏差. 因此,一些研究通过使用残差机制[3,4],注意力机制[5~8],金字塔结构[9~12],或者更复杂的卷积形式(空洞卷积[13~15]、深度可分离卷积[16,17]、动态卷积[18,19]和可变形卷积[20~22]等)以及更加密集的跳跃连接[23~26]等改进措施,有效提高了模型长距离建模能力,弥补了传统卷积在处理全局关系方面的局限性.

然而,这些改进方法依然存在感受野受限的问题,Transformer[27]的提出为探索全局关系提出了新的方法. 在Vision Transformer[28]中,图像以具有位置信息的2D图像块作为输入,转换成token序列后通过多头自注意力机制和前馈神经网络对token进行全局上下文关系建模. 但在医学图像分割方面,使用纯Transformer网络却无法得到准确的结果,主要原因在于Transformer的每一层只关注全局上下文关系,缺乏图像局部的细节信息,如边界、纹理等信息的提取,难以解决医学图像边缘模糊、目标尺度变化较大等问题. 因此,许多研究工作探索如何将Transformer与其他网络结构相结合,从而实现精准的医学图像分割. TransUnet[29]首次将Transformer集成到U-Net架构中用于处理医学图像,该架构在保留Transformer全局建模能力的同时保留了卷积神经网络(Convolutional Neural Network,CNN)归纳偏置的特性. Swin-Unet[30]是首次使用纯Transformer的U型结构,将原本U-Net中的编解码部分替换为Swin-Transformer[31],类似卷积采用了滑动窗口操作,在窗口内计算自注意力,有效减少了Transformer的计算量. 这些混合架构将Transformer与U型结构结合,有效地提高了模型的分割能力. 这些网络在医学图像分割任务中的成功证明了Transformer在视觉领域的应用潜力.

医学图像本身存在对比度低、边界模糊、噪声较大的问题,而多器官分割又存在不同目标之间大小、分布差异较大的问题. Transformer在处理医学图像的长程依赖关系和建立全局关系方面表现出显著优势,但缺少局部细粒度信息的提取,导致目标分割边界模糊. 同时,卷积神经网络由于其强大的局部信息提取能力而被广泛应用,但仍存在缺少长距离关系建模的局限性. 因此,为了能够保证计算效率,同时能够将CNN的细节信息提取能力与Transformer的全局建模能力结合起来,本文提出了一种用于医学图像分割的CNN-Transformer混合架构网络,主要贡献如下:

(1)提出了基于混洗特征编码的CNN-Transformer融合策略. 与传统双分支网络显著不同,提出的双支路交叉融合在不同尺度上融合了CNN与Transformer分支的信息,通过通道混洗编码不仅减少了高级语义信息与细节信息的损失,而且对医学图像的局部细节特征与全局语义信息进行了更好地提取融合.

(2)提出了一种基于多尺度特征图的门控解码器(Gated Decoder based on Multi-scale Feature,GDMF)对混合结构进行解码,在恢复图像分辨率的过程中补充丢失的细节信息,提取网络不同阶段的多尺度高维特征并对特征图进行自适应增强,为逐像素预测提供可靠信息. 与常见CNN或Transformer解码器相比,GDMF解码器充分利用了不同尺度特征图内的有效信息,在混合结构中表现出良好的性能.

(3)所提方法在ACDC(Automated Cardiac Diagnosis Challenge)心脏数据集与FLARE21(Fast and Low GPU memory Abdominal oRgan sEgmentation challenge 2021)多器官数据集上的DSC(Dice Similarity Coefficient)分割指标分别高达92.13%和88.96%,均优于当前的主流方法,验证了所提方法的优越性.

2 相关工作

2.1 基于卷积神经网络的医学图像分割

随着深度神经网络的发展,基于卷积的神经网络

在医学图像分割任务中表现出了优异的性能. 通过对卷积核进行滑动窗口操作,以固定的步长对局部信息进行编码,对卷积核内参数进行共享,降低了计算复杂度和参数数量,并为神经网络带来了归纳偏置. 然而,使用滑动窗口的计算方式也导致了有限的感受野的问题,模型由于缺少对长距离关系的建模,难以捕获全局上下文信息. 为解决这一问题,通过使用多层卷积层和下采样层逐渐增大感受野,提取更高级的语义信息. 这种方法带来的优势使得卷积神经网络能够高效地处理高分辨率图像,使得网络在较少数量的数据集上能够快速收敛并取得较好的性能.

Ronneberger等人[1]提出的U-Net结构在全卷积网络的基础上增加了上采样,通过跳跃连接弥补下采样造成的细节信息损失,在各种医学分割任务中取得了巨大的成功,成为医学图像分割领域的一个重要里程碑. 随后,许多学者在U-Net的基础上对网络架构进行改进,Li等人[3]提出的ResU-Net在编码阶段增加了残差连接,有效地解决了网络深层梯度消失的问题. Zhou等人[23]与Huang等人[32]分别提出了U-Net++与U-Net3+,采用更密集的跳跃连接,增加不同层级间的交流,进一步减少了下采样造成的信息损失,减少编码器与解码器之间特征映射的语义差距.

此外,许多工作将U-Net和注意力机制结合. 常见的注意力主要可以分为3类:通道注意力、空间注意力以及通道和空间注意力的结合. Oktay等人[6]提出的Attention Unet使用空间自注意力机制在跳跃连接阶段对特征图进行增强,有效地增强了特征图中的关键特征并抑制不相关区域的特征. Guo等人[33]提出的CAR-Unet将通道注意力引入残差块,集中于具有充足信息的通道. Woo等人[34]将通道与空间注意力相结合,提出了CBAM(Convolutional Block Attention Module),通过将通道与空间注意力串行计算来生成注意力图,对重要的通道进行增强并细化局部区域内的信息.

2.2 基于Transformer的医学图像分割

注意力机制在医学图像分割中发挥了越来越重要的作用. Transformer作为自注意力机制,在自然语言处理方面有着优异的性能. 受U-Net结构的启发,Cao等人[30]提出了Swin-Unet,采用Swin-Transformer作为主干网络替换U-Net中的编码与解码部分,使用Patch Merging执行下采样操作,通过跳跃连接来弥补下采样造成的空间信息损失,最后由Patch Expanding操作将图像恢复到输入的分辨率,显著降低了Transformer的参数量与计算量. Peiris等人[35]提出的VT-UNet设计了分层的纯Transformer网络,采用并行的自注意力和交叉注意力来进行边界的细化. Tragakis等人[36]提出的FCT(Fully Convolutional Transformer)使用卷积替代Transformer中的自注意力,设计了Convolution Attention模块来学习全局上下文关系,以及全卷积Wide-Focus模块来学习局部到全局的关系,提高了网络的细粒度信息特征提取能力. He等人[37]设计了三分支解码器分别用于细胞核分割、核边缘分割和聚类边缘分割,并使用注意力共享策略来降低参数数量.

研究者也在寻求通过卷积的形式达到与Transformer相近的效果. Liu等人[38]在ConvNeXt中模仿Swin-Transformer的优化策略,达到了优于Swin-Transformer的效果. Lee等人[39]提出的3D UX-Net使用三维的大内核卷积作为主干网络处理腹部医学图像.

2.3 基于卷积神经网络和Transformer的医学图像分割

由于自注意力的计算复杂度随着输入图像大小的增大成二次方增加,因此,Transformer将图像裁剪成patch,通过丢弃补丁中的细节信息减少计算量,这对于分类任务是可以接受的,但是对于像素级的分割任务,丢失细节信息是不可接受的,因此许多工作研究如何将CNN的优势与Transformer的优势结合起来. Chen等人[29]提出的TransUNet首次在医学图像分割领域将Transformer与卷积神经网络以串行的方式结合,将Transformer架构集成到CNN中,利用Transformer来提取全局信息并通过U型结构保留卷积神经网络低级的边缘信息. 同样,Gao等人[40]提出的UTNet在Unet结构的基础上集成了Transformer,设计了Transformer解码块并在跳跃连接处应用了Transformer用于恢复图像分辨率,显著降低了计算复杂度.

在每个卷积层后引入Transformer可以在不同尺度建立长程依赖关系,然而串行连接的方式会造成信息损失,因此一些工作使用并行的双分支Transformer结构来同时学习全局与局部依赖关系. Lei等人[41]提出的CiT-Net在CNN分支专注于对空间特征的自适应提取,Transformer分支使用滑窗自适应互补注意力模块增加通道与空间的关系,采用并行连接的方式同时获取两者的优势.

此外,有工作将Transformer结构以堆叠金字塔的方式代替卷积的编码层或解码层. Xu等人[42]提出的LeViT-UNet使用Transformer作编码器,CNN作解码器,并设计注意力下采样机制,提高网络的计算效率. Gong等人[43]提出的CTranS网络采用混合CNN-Transformer架构在不同分辨率下相互连接,在无需预训练的情况下获得了与Transformer相似的性能. Hatamizadeh等人[44]设计了一种以Transformer作编码器,CNN作解码器的U型架构UNETR(UNEt TRansformers)进行脑部三维医学图像分割. 此后,Hatamizadeh等人[45]根据Swin-Transformer的思想提出了Swin UNETR,将Transformer

替换为Swin-Transformer,以提取脑部特征并进行脑部信息分割.

3　提出的方法

3.1　总体结构

给定输入图像的高度H、宽度W以及通道C,期望得到预测大小为$H\times W$的相应像素标签图. 与传统的神经网络架构区别在于,本文设计网络采用CNN-Transformer并行的混合架构,并通过双分支交叉融合操作对图像特征进行编码,此外还引入了基于多尺度特征图的门控网络解码器来还原图像分辨率,进一步提升分割精度.

在3.2节中,本文将详细介绍这一混合结构的特点以及如何使用混洗特征编码将CNN与Transformer分支提取到的信息充分融合. 在3.3节中,将详细介绍基于多尺度特征图的门控解码器(GDMF)如何选取,以及如何融合来自不同尺度特征图的特征来生成高分辨率的输出.

3.2　混洗特征编码

编码器主要由3个部分组成,包括Transformer分支、CNN分支以及混洗特征编码模块,以下是对这3个部分的详细描述. (1)Transformer分支. 选用Swin-Unet作为主要编码结构,Swin-Unet相比于PVT(Pyramid Vision Transformer)、ViT(Vision Transformer)等其他Transformer方法具有更高效的计算和参数资源利用率. 这意味着在相同的计算资源下,Swin-Unet能更好地处理大规模图像和复杂的任务. 此外,Swin-Unet还具备强大的多尺度上下文表示和全局特征提取能力,使其成为长距离信息建模的理想选择. (2)CNN分支. 为了弥补Transformer全局建模中忽视的局部信息,引入CNN分支对细节信息进行特征提取. U-Net被广泛认可为一种出色的网络结构,特别适用于局部特征的细粒度提取. 在这一分支中提供有关图像结构的重要细节,这种局部信息的有效利用可以增强模型对图像中微小目标或局部特征的感知能力,从而提高图像分割的精确性. (3)混洗特征编码融合模块. 为了将来自Transformer分支和CNN分支的特征进行有效特征编码,将2个分支输出结果调整为相同尺寸和通道数,使得2个分支特征图能够在通道维度上进行拼接融合. 在混洗特征编码融合的过程中,首先将拼接合成后的特征图分为2个通道组,再将这些通道组交叉排列,以增加不同分支之间的信息交流,最后将交叉排列后的通道组合成为新的特征图进行后续处理. 在对通道重排列的过程中,网络不同分支之间的信息得以充分交流,综合利用全局与局部信息,弥补由于边界模糊导致的难以分割的问题,以此提高模型的特征表达能力.

如图1所示,首先将通道平均分成2个部分,分别进行卷积与Transformer计算. 在卷积分支中,采用U-Net网络对局部信息进行细粒度提取,捕获图像中的纹理、边缘、形状等局部特征,从而增强网络对局部信息的感知能力;在Transformer分支中,利用Swin-Unet强大的多尺度上下文和全局特征提取能力来处理具有全局联系的区域和特征,实现对全局信息的有效提取,帮助网络理解图像整体结构以及捕获不同区域之间的关系.

具体地说,在Transformer分支中,与Swin-Unet相对应,执行Patch Partition、Linear Embedding以及Swin-Transformer Block后,输出的特征序列为3 136×96,再将输出特征序列还原为$56\times56\times96$. 对于CNN分支而言,本文使用卷积操作对源输入数据进行局部细节信息提取. 在此过程中,使用了2次无损下采样和升维操作,使得最终得到的特征图尺寸同样为$56\times56\times96$. 此时,Transformer分支和CNN分支输出的特征图大小和通道数均相同,可以进行Concat操作进行通道融合. 通过上述操作,不仅保证了2个分支的输出数据尺寸一致,而且确保了后续的交叉融合操作也能顺利进行. 此外,通过这样的方式将2个分支进行融合,能够捕获到图像更加丰富的局部与全局信息.

如图2左侧所示,蓝色代表经过Transformer块输出的特征图,橙色代表经过CNN块输出的特征图. 对2个分支的输出结果进行Concat操作相互融合,创建一个更丰富的特征表示. 将拼接后的特征图进行分组,并使用通道重排列操作对通道进行混洗重排,这有助于加强不同通道间的交流通信并相互补充,从而更好地融合全局与局部信息. 与传统的并行双支路网络不同,本文网络在每一个阶段都执行信息融合,以确保每个阶段都能充分受益于全局和局部信息. 最后,将融合后的特征图重新划分并分配给2个分支,继续进行卷积操作与全局自注意力计算,这有助于进一步提取和整合特征信息,为准确的分割结果提供了可靠的基础. 通过采用本文提出的通道交叉融合操作,有利于2个分支信息之间的流动与交互,部分由CNN提取出的局部特征会进入Transformer分支进行全局信息建模,而部分由Transformer提取出的全局特征进入CNN分支进行细节信息捕获,从而使特征图包含更为丰富的互补信息,使模型学到更加复杂和抽象的特征表示.

编码器部分可以表示如下:

$$F_i=\mathrm{Concat}(F_{i,\mathrm{conv}},F_{i,\mathrm{trans}}) \tag{1}$$

$$F_{i,\mathrm{conv}}=F_i[:,1::C/2,:,:] \tag{2}$$

$$F_{i,\mathrm{trans}}=F_i[:,C/2::C,:,:] \tag{3}$$

其中,F_i为每一层编码器的输入,$F_{i,\mathrm{conv}}$表示原特征图通道数一半进入卷积分支,$F_{i,\mathrm{trans}}$表示另一半通道进入

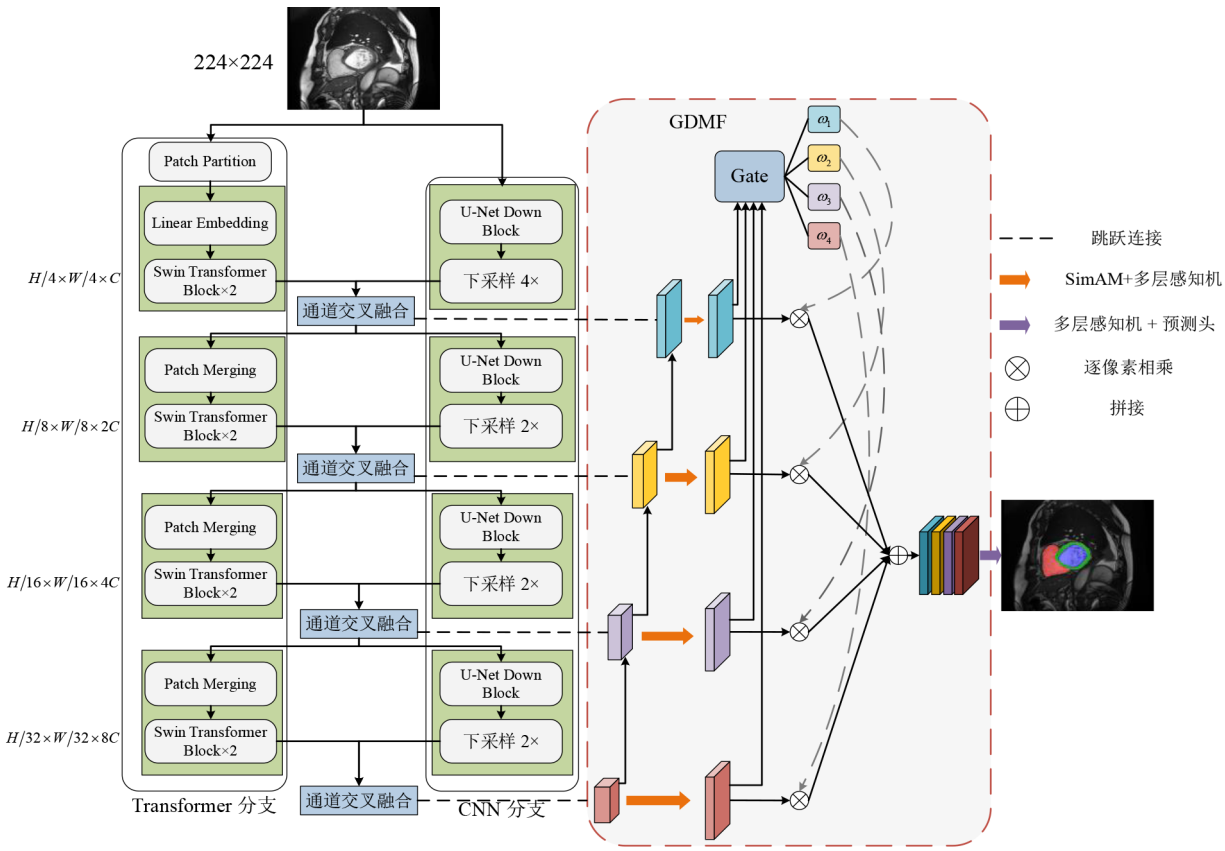


图 1 提出的混合 CNN-Transformer 网络架构图

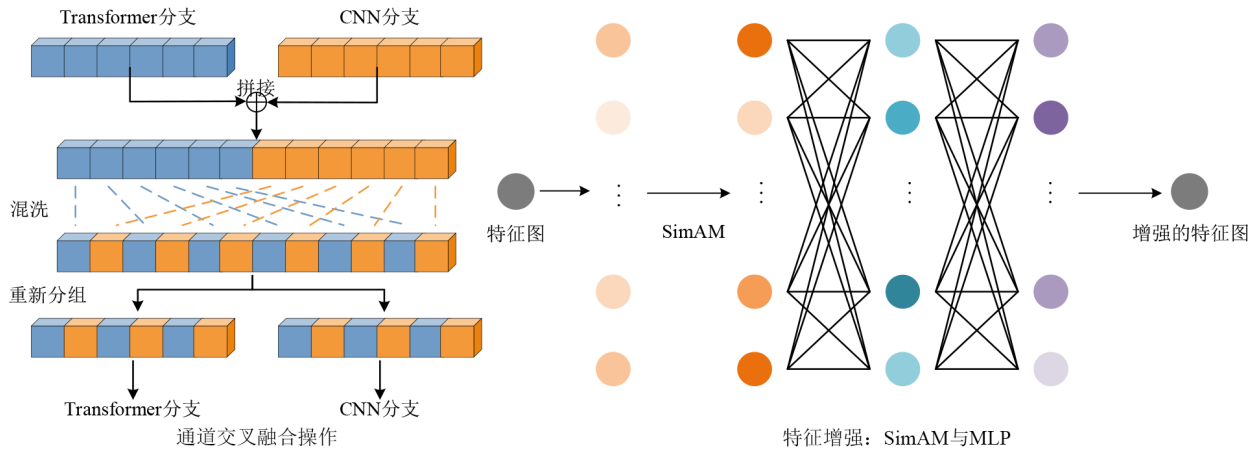


图 2 编码器:通道交叉融合模块与特征增强模块

Transformer 分支.

$$F_i = \text{Concat}(F_{i,\text{conv}}, F_{i,\text{trans}}) \tag{4}$$

$$F_{i+1,\text{trans}} = \text{Attention}(F_{i,\text{trans}}) \tag{5}$$

$$F_{i+1} = \text{Concat}(F_{i+1,\text{conv}}, F_{i+1,\text{trans}}) \tag{6}$$

其中，$F_{i+1,\text{conv}}$ 为经过卷积后的输出，$F_{i+1,\text{trans}}$ 为经过 Transformer 分支的输出. 将 2 个分支的输出结果在通道维度进行拼接，从而得到第 $i+1$ 层的输入. 其中自注意力计算公式如下：

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^{\text{T}}}{\sqrt{d}} + B\right)V \tag{7}$$

其中，Q、K、V 分别代表 Query、Key、Value，d 表示他们相应的维度，B 为偏移值.

3.3 基于多尺度特征图的门控解码器

在完成局部和全局特征提取后，对于双分支混合编码器，简单地采用 CNN 或 Transformer 解码器可能无法很好地还原特征图. 此外，在图像预测的过程中，需

要同时捕捉到局部边缘信息和全局位置信息,以便能够更有效地理解图像内容. 因此,本文提出了一个基于多尺度特征图的门控解码器,该解码器能够自适应地学习网络不同阶段的高维信息进行分割结果预测.

在医学图像分割的任务中,不同器官具有的尺度和复杂性差异较大,因此需要一个高效的解码器处理不同尺度的目标. 传统的方法往往依赖于对瓶颈层进行上采样,这在处理不同尺度目标时存在明显的不足. 为了克服这一问题,引入多尺度特征图的概念,以更好地捕获目标的结构和细节信息. 同时引入门控机制,以便在解码器中灵活地选择和整合特征图信息,控制不同层级特征图的权重,从而选择性地融合不同尺度和分辨率的特征. 这种灵活性有助于提高模型对局部边缘信息和全局信息的感知和利用. 本文提出的GDMF解码器能够利用每一层特征图的差异性信息,而不仅仅依赖于瓶颈层进行预测.

如图3所示,针对不同层的输出结果,首先进行上采样,使其大小与前一层特征图相匹配并进行特征融合,通过综合不同层级的特征信息,以补充边缘轮廓等细节信息. 特征增强部分如图2右侧所示,对于每一层的特征图,首先将图片尺寸上采样至$H/4 \times W/4$,使用SimAM注意力机制[46]对图像特征进行增强,在不增加模型复杂度的前提下,自适应地突出特征图中的关键信息. 接着使用多层感知器(MultiLayer Perceptron, MLP)进一步对特征图增强并选择重要特征,以上操作能够为门控解码器预测提供可信信息. 将增强后的同一尺度的4张特征图F_1、F_2、F_3、F_4输入门控网络. 首先,在通道维度上进行拼接操作,经过卷积层与Softmax层生成大小为$H/4 \times W/4$的权重图ω_1、ω_2、ω_3、ω_4,每个像素位置的权重总和为1,即$\omega_1+\omega_2+\omega_3+\omega_4=1$. 通过学习权重来决定保留重要信息,抑制不重要的信息. 这些计算出的权重应用于原始特征图,得到合成特征图. 最后,通过MLP层将融合后的特征图上采样到原始图像大小并用于分割结果的预测. 基于多尺度特征图的门控解码器部分公式如下:

$$F'_{i+1}=\mathrm{Upsample}(F_{i+1}) \tag{8}$$

$$F_e=\mathrm{MLP}(\mathrm{SimAM}(\mathrm{Concat}(F_i+F'_{i+1}))) \tag{9}$$

$$W_i=\mathrm{Softmax}(F_e) \tag{10}$$

$$\mathrm{OUT}=\sum \omega_i \odot F_i \tag{11}$$

其中,F_i表示混合架构每一层的输出,F_e表示增强后的特征图. 具体来说,首先对跳跃连接增强后的图像进行MLP映射,使用SimAM注意力机制对图像进行增强,再通过MLP得到最终增强后的图像F_e. ω_i表示经过门控网络生成的权重,OUT表示最终图像的输出. 这样的设计充分利用了特征图中的信息,有效地避免了细节信息的丢失. 网络通过权衡全局特征与局部特征的重要性,对不同尺度目标选择适当的特征组合,进行最终的预测.

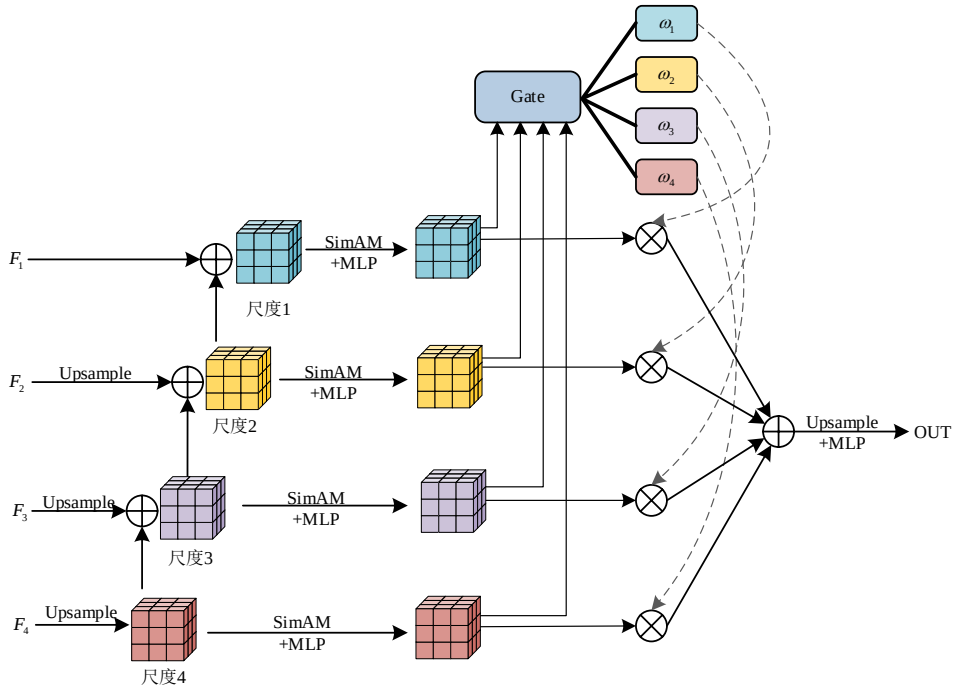


图3　基于多尺度特征图的门控解码器

4 实验结果与分析

4.1 实验数据集与评估指标

为证明提出方法的有效性，在2个公开医学图像数据集ACDC[47]、FLARE21[48]上进行了实验. 各数据集介绍如下.

ACDC：包含了从100名患者收集的心脏短轴磁共振成像数据. 对于每个数据实例，有左心室（Left Ventricle，LV），右心室（Right Ventricle，RV）和心肌（MYOglobin，MYO）3类标签. 在实验中，将数据集划分为80例训练样本和20例测试样本，使用Dice Similarity Coefficient（DSC）、95 Hausdorff Distance（95HD）、Average Symmetric Surface Distance（ASSD）来评估实验结果.

FLARE21：包含来自2个数据中心的361例病例. 每一例数据包含肝脏、肾脏、脾脏和胰腺4类标签. 在实验中，将数据集划分为289例训练样本和72例测试样本，使用DSC来评估实验结果.

4.2 实验细节

实验中所有的网络均基于NVIDIA Geforce RTX3090 with 24GB memory、Ubuntu 18.04、PyTorch 1.7的服务器实现. 实验采用五折交叉验证，数据集处理时调整至相同间距，由于没有使用预训练模型，本文采用随机初始化权重，从头开始训练，因此对数据进行了增强，增强方法包括缩放范围（最大0.3）、旋转（180°）、增加高斯噪声（0.02）. 训练时，从原图中随机裁剪224×224大小的图片，batchsize设置为8，使用交叉熵结合骰子损失训练500 epochs，使用AdamW优化器，初始学习率设置为0.000 1，权重衰减为0.000 1.

4.3 实验结果分析

4.3.1 不同方法的结果对比

将所提出的方法与9种最先进的医学图像分割方法进行比较. 表1显示了在ACDC数据集上的DSC分割结果，其中加粗数字表示获得了最优结果，下划线数字表示获得了次优结果，表2~表4同理. 实验结果表明，本文提出的方法获得了最佳性能，其分割精度DSC达到92.13%，与传统CNN方法如U-Net与nnUNet相比，提出的方法分别提高了4.58个百分点与0.52个百分点，这表明通过添加全局位置信息，有效提高了网络的性能. 与Swin-Unet相比，所提方法的DSC提高了2.13个百分点，验证了为Transformer引入局部细节信息有助于提高网络对器官边缘的分割能力. 此外，通过与其他主流方法对比，证明了本文提出的混合结构以及门控解码器能够更有效地提取和筛选图像特征，从而更精准预测最终的分割结果.

表2显示了ACDC数据集上的95HD和ASSD结果. 所提方法在右心室、心肌、左心室的ASSD指标取得了最好的结果，分别达到0.63 mm、0.37 mm、0.41 mm. 与nnUNet相比，本文方法在MYO分割提高了0.33 mm，在RV、LV的95HD获得了次优结果，这表明本文方法在边缘预测方面具有更出色的性能. 图4中第1~3行显示了不同方法在ACDC数据集上的分割可视化结果，其中蓝色表示左心房，绿色表示心肌，红色表示右心房. 从第1行可以观察到，当前主流方法在识别较细的边缘区域上存在一定的局限性，导致出现漏分割的现象. 所提方法有效提取了关键信息，提高了模型对较细边缘的识别能力. 第2行为较简单病例的分割结果，分割结果的主要误差在于边缘区域，所提方法与nnUnet效果相近，与Transformer方法相比，边缘信息更加精确. 第3行则为较为困难的病例，本文方法明显避免了漏检的情况，这主要是因为门控解码器使用多尺度特征图提供的全局与局部信息进行预测，有效地为解码过程提供了丰富信息，从而避免了漏检情况的发生，在较为复杂的情况下能更准确地预测心脏结构.

表1 ACDC数据集中的DSC结果 单位：%

方法	DSC	右心室	心肌	左心室
R50 U-Net[1]	87.55	87.10	89.63	94.92
R50 Att-Unet[7]	86.75	87.58	79.20	93.47
ViT[28]	81.45	81.46	70.71	92.18
TransUNet[29]	89.71	88.86	84.53	<u>95.73</u>
Swin-Unet[30]	90.00	88.55	85.62	**95.83**
LeViT-UNet$_{384}$[42]	90.32	89.55	87.64	93.76
MISSFormer[49]	90.86	89.55	88.04	94.99
nnUNet[50]	91.61	<u>90.24</u>	89.24	95.36
nnFormer[51]	<u>91.78</u>	90.22	<u>89.53</u>	95.59
本文方法	**92.13**	**91.78**	**89.59**	95.01

表2 ACDC数据集上的95HD以及ASSD结果

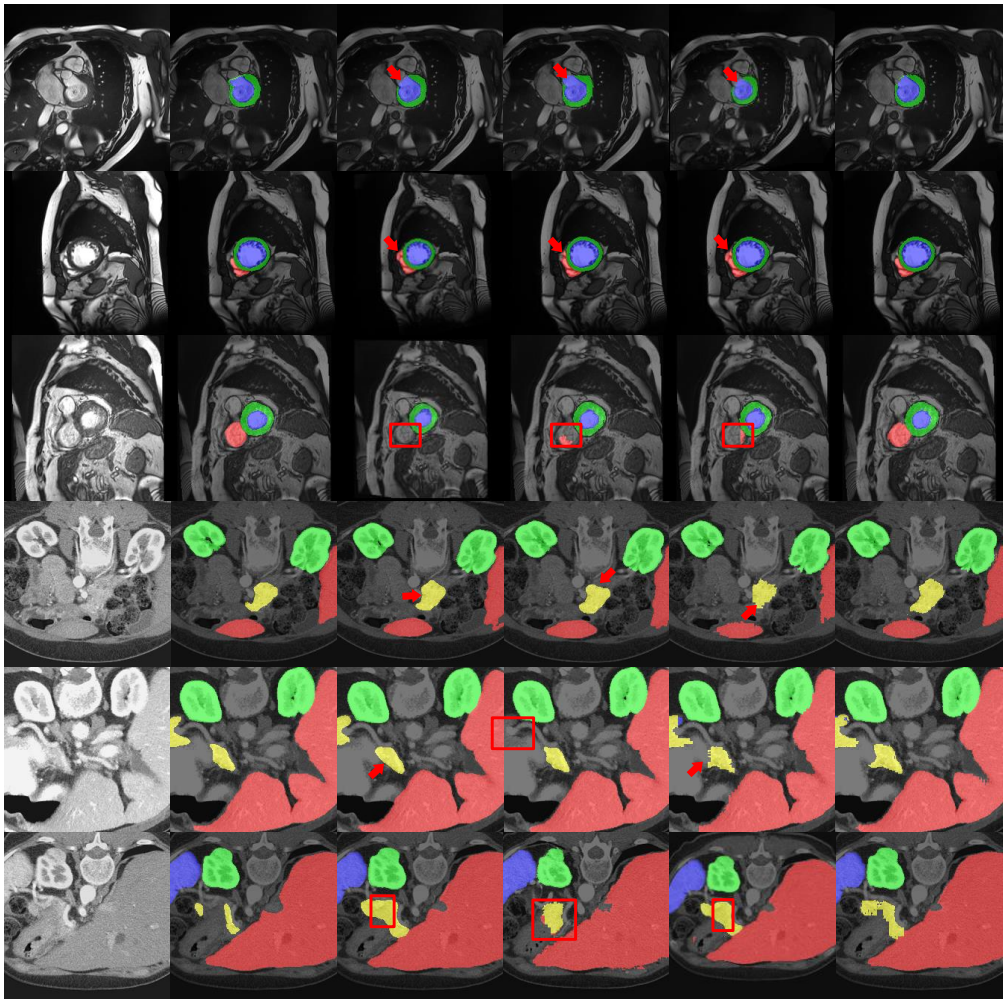
方法	DSC/%	右心室/mm		心肌/mm		左心室/mm	
		95HD	ASSD	95HD	ASSD	95HD	ASSD
CE-Net[15]	89.01	6.76	1.01	5.45	0.62	4.79	0.69
TransUNet[28]	89.73	7.61	1.13	5.03	0.57	6.75	0.84
Swin-Unet[29]	90.30	5.75	0.87	3.78	0.47	5.70	0.55
nnUNet[50]	<u>91.62</u>	**4.47**	<u>0.63</u>	<u>2.88</u>	<u>0.39</u>	**3.70**	<u>0.43</u>
本文方法	**92.13**	<u>4.56</u>	**0.63**	**2.55**	**0.37**	<u>3.79</u>	**0.41**

表3显示了在FLARE21数据集上的DSC结果. 为了验证所提网络的泛化性能，本文在FLARE21数据集上进行了腹部器官分割的测试. 由表3的实验结果可以得出，该方法的平均分割精度DSC达到88.96%，与Swin-Unet和nnUNet相比，本文方法分别提高了2.44个百分点和0.4个百分点，优于当前主流的分割算法，在肝脏与脾脏的分割中取得了最好的结果，分别达到93.96%和92.98%. 图4中第4~6行显示了不同方法在FLARE21数据集上的分割可视化结果，其中红色表示肝脏，绿色表示肾脏，黄色表示胰腺. 第4行为正常病

例的分割结果,分割的主要误差在于边缘区域黏连导致的边界难以分割,本文方法使用全局信息对特征进行定位,更加精确地完成了腹部器官分割任务. 第5行为相对困难的病例,nnUNet缺少全局信息的补充,存在漏分割的现象,Swin-Unet则缺少局部信息,因而存在分割结果不平滑的问题. 第6行是较难分割的病例,胰腺器官形态变化较大且目标较小,所有方法分割结果都不理想. 本文所设计的方法考虑到了全局与局部关系,对胰腺区域能更加精确地进行分割.

表3　FLARE21数据集上的DSC结果　　单位:%

方法	DSC	肝脏	肾脏	脾脏	胰腺
Att-UNet[6]	87.98	93.40	93.45	90.38	74.68
Swin-Unet[29]	86.52	92.87	91.30	89.85	72.08
TransUNet[28]	88.31	93.29	93.43	90.50	76.02
nnUNet[50]	88.56	93.06	**93.59**	91.24	**76.32**
本文方法	**88.96**	**93.96**	92.98	**92.98**	75.81



(*a*) 原始图像　(*b*) 真实标签　(*c*) 本文方法　(*d*) nnUNet　(*e*) Swin-Unet　(*f*) TransUnet

图4　不同方法的可视化结果

4.3.2　消融实验

本文提出的混合架构网络利用混洗特征编码策略促进CNN分支与Transformer分支之间的信息交流,从而实现局部特征与全局信息的融合. 此外,还设计了一种基于多尺度特征图的门控解码器用于还原图像分辨率并预测最终分割结果. 为了验证所提混合架构的有效性,进行了一系列消融实验,结果如表4所示. 首先,保持U-Net和Swin-Unet编码器不变,仅将解码器替换为GDMF. 与之对应,将编码器修改为CNN-Transformer的混合结构,将解码器分别替换为CNN与GDMF. 实验结果表明,基于混洗编码的CNN-Transformer混合结构有助于提高网络性能,GDMF有助于保留特征图中有效信息的同时又能对不同维度的特征进行自适应增强,提高分割结果的准确率. 消融实验充分证明了所提出的CNN-Transformer混合策略以及GDMF解码器在医学多器官分割任务中的高效性.

为了进一步探究解码器各部分的有效性,进行了消融实验,结果见表5. “√”说明使用了该模块,“—”说

明该模块没有被使用,以DSC系数为指标测试了这些方法. 实验结果表明,在跳跃连接处使用MLP对于网络性能有较大提升,再增加SimAM注意力机制后,经过MLP进行全局特征归纳映射,通过这2个部分协同作用,网络实现了出色的性能表现.

表4 消融实验:不同编码器与解码器对于实验结果的影响

编码器	解码器	DSC/%	Flops/G	Params/M
Swin-Unet	Swin-Unet	90.00	8.693	41.342
Swin-Unet	GDMF	90.18	5.901	16.949
U-Net	U-Net	89.56	36.963	28.949
U-Net	GDMF	90.59	28.664	22.736
CNN-Transformer	U-Net	91.40	43.563	47.388
CNN-Transformer	GDMF	**92.13**	7.668	28.230

表5 消融实验:GDMF内部各模块对分割结果的影响 单位:%

多层感知机	SimAM	DSC	右心室	心肌	左心室
—	—	90.27	89.58	86.89	94.32
—	√	90.35	89.42	87.07	94.37
√	—	91.73	91.06	89.53	94.59
√	√	92.13	91.78	89.59	95.01

5 结论

本文提出了一种基于混洗特征编码和门控解码的医学图像分割网络. 该网络在编码部分采用混洗策略,在不同阶段有效地融合双分支网络提取到的高维特征,从而捕捉多尺度、全局与局部的特征信息,有效地解决了由于成像导致的轮廓模糊而引起的分割精度受限的问题. 此外,针对器官尺度变化大的问题,设计了基于多尺度特征图的门控解码器. 具体地说,编码器提取到了图像的全局与局部特征,先通过SimAM注意力机制和MLP来筛选特征,最后通过门控机制来确定所需特征,从而得到准确的分割结果. 通过在2个常用的多器官医学图像分割数据集上的充分实验,进一步验证了所提方法的有效性和优越性. 后续将探索更多不同网络的融合策略和特征增强技术,在提高分割精度的同时降低对计算资源的需求,以推动医学图像分割领域的更快发展.

参考文献

[1] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[M]// Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.

[2] 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). Piscataway: IEEE, 2016: 565-571.

[3] LI R, ZHENG S Y, DUAN C X, et al. Multistage attention ResU-net for semantic segmentation of fine-resolution remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 8009205.

[4] BI R R, JI C L, YANG Z P, et al. Residual based attention-Unet combing DAC and RMP modules for automatic liver tumor segmentation in CT[J]. Mathematical Biosciences and Engineering, 2022, 19(5): 4703-4718.

[5] CHENG Z M, QU A P, HE X F. Contour-aware semantic segmentation network with spatial attention mechanism for medical image[J]. The Visual Computer, 2022, 38(3): 749-762.

[6] OKTAY O, SCHLEMPER J, LE FOLGOC L, et al. Attention U-net: Learning where to look for the pancreas[EB/OL]. (2018-05-20)[2023-10-27]. https://arxiv.org/abs/1804.03999v3.

[7] 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.

[8] HE H Y, CAI J F, LIU J, et al. Pruning self-attentions into convolutional layers in single path[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(5): 3910-3922.

[9] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 6230-6239.

[10] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 936-944.

[11] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-12-05)[2023-10-27]. https://arxiv.org/abs/1706.05587v3.

[12] ZHOU Z X, HE Z S, JIA Y Y. AFPNet: A 3D fully convolutional neural network with atrous-convolution feature pyramid for brain tumor segmentation via MRI images[J]. Neurocomputing, 2020, 402: 235-244.

[13] YU F, KOLTUN V, FUNKHOUSER T. Dilated residual networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 636-644.

[14] XIE F, HUANG Z, SHI Z J, et al. DUDA-Net: A double U-shaped dilated attention network for automatic infec-

tion area segmentation in COVID-19 lung CT images[J]. International Journal of Computer Assisted Radiology and Surgery, 2021, 16(9): 1425-1434.

[15] GU Z W, CHENG J, FU H Z, et al. CE-net: Context encoder network for 2D medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281-2292.

[16] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2023-10-27]. https://arxiv.org/abs/1704.04861v1.

[17] LEI T, SUN R, DU X G, et al. SGU-net: Shape-guided ultralight network for abdominal image segmentation[J]. IEEE Journal of Biomedical and Health Informatics, 2023, 27(3): 1431-1442.

[18] YANG B, BENDER G, LE Q V, et al. CondConv: Conditionally parameterized convolutions for efficient inference[EB/OL]. (2020-09-04)[2023-10-27]. https://arxiv.org/abs/1904.04971v3.

[19] LEI T, ZHANG D, DU X G, et al. Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network[J]. IEEE Transactions on Medical Imaging, 2023, 42(5): 1265-1277.

[20] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 764-773.

[21] YANG X, LI Z Q, GUO Y Q, et al. DCU-net: A deformable convolutional neural network based on cascade U-net for retinal vessel segmentation[J]. Multimedia Tools and Applications, 2022, 81(11): 15593-15607.

[22] LEI T, WANG R S, ZHANG Y X, et al. DefED-net: Deformable encoder-decoder network for liver and liver tumor segmentation[J]. IEEE Transactions on Radiation and Plasma Medical Sciences, 2022, 6(1): 68-78.

[23] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHSH 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.

[24] CAI S J, TIAN Y X, LUI H, et al. Dense-UNet: A novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network[J]. Quantitative Imaging in Medicine and Surgery, 2020, 10(6): 1275-1285.

[25] CAI Z T, XIN J M, SHI P W, et al. DSTUNet: UNet with efficient dense SWIN transformer pathway for medical image segmentation[C]//2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). Piscataway: IEEE, 2022: 1-5.

[26] 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.

[27] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. (2017-06-12) [2023-10-27]. https://arxiv.org/abs/1706.03762.

[28] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [EB/OL]. (2021-06-03) [2023-10-27]. https://arxiv.org/abs/2010.11929v2.

[29] CHEN J N, LU Y Y, YU Q H, et al. TransUNet: Transformers make strong encoders for medical image segmentation[EB/OL]. (2021-02-08) [2023-10-27]. https://arxiv.org/abs/2102.04306v1.

[30] CAO H, WANG Y Y, CHEN J, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation[M]// Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023: 205-218.

[31] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 9992-10002.

[32] HUANG H M, LIN L F, TONG R F, 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). Piscataway: IEEE, 2020: 1055-1059.

[33] GUO C L, SZEMENYEI M, HU Y T, et al. Channel attention residual U-net for retinal vessel segmentation[C]// ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2021: 1185-1189.

[34] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 3-19.

[35] PEIRIS H, HAYAT M, CHEN Z L, et al. A robust volumetric transformer for accurate 3D tumor segmentation[M]// Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022: 162-172.

[36] TRAGAKIS A, KAUL C, MURRAY-SMITH R, et al. The fully convolutional transformer for medical image segmentation[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2023: 3649-3658.
[37] HE Z Q, UNBERATH M, KE J, et al. TransNuSeg: A lightweight multi-task transformer for nuclei segmentation[M]//Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023: 206-215.
[38] LIU Z, MAO H Z, WU C Y, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 11966-11976.
[39] LEE H H, BAO S X, HUO Y K, et al. 3D UX-Net: A large kernel volumetric ConvNet modernizing hierarchical transformer for medical image segmentation[EB/OL]. (2023-03-02)[2023-10-27]. https://arxiv.org/abs/2209.15076v4.
[40] GAO Y H, ZHOU M, METAXAS D N. UTNet: A hybrid transformer architecture for medical image segmentation[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021: 61-71.
[41] LEI T, SUN R, WANG X, et al. CiT-net: Convolutional neural networks hand in hand with vision transformers for medical image segmentation[C]//Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2023: 1017-1025.
[42] XU G P, ZHANG X, HE X W, et al. LeViT-UNet: Make faster encoders with transformer for medical image segmentation[M]//Lecture Notes in Computer Science. Singapore: Springer Nature Singapore, 2023: 42-53.
[43] GONG Z D, FRENCH A P, QIU G P, et al. CTranS: A multi-resolution convolution-transformer network for medical image segmentation[C]//2024 IEEE International Symposium on Biomedical Imaging (ISBI). Piscataway: IEEE, 2024: 1-5.
[44] HATAMIZADEH A, TANG Y C, NATH V, et al. UNETR: Transformers for 3D medical image segmentation[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2022: 1748-1758.
[45] HATAMIZADEH A, NATH V, TANG Y C, et al. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022: 272-284.
[46] QIN X Y, LI N, WENG C, et al. Simple attention module based speaker verification with iterative noisy label detection[C]//ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2022: 6722-6726.
[47] 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 Transactions on Medical Imaging, 2018, 37(11): 2514-2525.
[48] MA J, ZHANG Y, GU S, et al. Fast and low-GPU-memory abdomen CT organ segmentation: The FLARE challenge[J]. Medical Image Analysis, 2022, 82: 102616.
[49] HUANG X H, DENG Z F, LI D D, et al. MISSFormer: An effective medical image segmentation transformer[EB/OL]. (2021-12-19)[2023-10-27]. https://arxiv.org/abs/2109.07162v2.
[50] ISENSEE F, JAEGER P F, KOHL S A A, et al. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation[J]. Nature Methods, 2021, 18(2): 203-211.
[51] ZHOU H Y, GUO J S, ZHANG Y H, et al. nnFormer: Volumetric medical image segmentation via a 3D transformer[J]. IEEE Transactions on Image Processing, 2023, 32: 4036-4045.

作者简介

雷 涛 男,1981 年 11 月出生,陕西大荔人. 2011 年在西北工业大学获得博士学位,现为陕西科技大学教授,博士生导师. 主要研究方向为图像处理、模式识别和计算机视觉等.
E-mail: leitao@sust.edu.cn

张峻铭 男,1998 年 9 月出生,江苏无锡人. 陕西科技大学硕士. 主要研究方向为计算机视觉.
E-mail: zhangjm922@outlook.com

杜晓刚 男,1985 年出生,陕西宝鸡人. 现为陕西科技大学电子信息与人工智能学院副教授. 主要研究方向为机器学习、计算机视觉、医学图像处理.
E-mail: du423@sina.com