文章编号：1672-3961(2024)06-0008-11 DOI：10.6040/j.issn.1672-3961.0.2023.174

基于双解码器的医学图像分割模型

刘全金[1]，嵇文[1]，胡浪涛[1]，黄汇磊[1]，杨瑞[1]，李翔[2,3]，高泽文[2,3]，魏本征[2,3*]
（1.安庆师范大学电子工程与智能制造学院，安徽 安庆 246133；2.山东中医药大学医学人工智能研究中心，山东 青岛 266112；3.山东中医药大学青岛中医药科学院，山东 青岛 266112）

摘要：针对医学图像目标区域尺度不一及有标签医学图像样本少的问题，提出一种基于双解码器的医学图像分割模型（dual-decoding Swin-Unet，DDS-UNet）。DDS-UNet 模型以 Swin Transformer 模块构建编码器，提取医学图像多尺度特征；解码器 1 利用 Swin Transformer 模块全局和远程语义特征提取优势，在上采样过程中逐级恢复并聚合编码器输出的对应尺度特征信息；解码器 2 利用卷积神经网络（convolutional neural networks，CNN）的局部特征提取优势，在上采样过程中逐级恢复医学图像空间信息；特征融合模块利用空洞卷积分解编码器输出的深层语义特征信息，并在上采样过程中协同融合双解码器输出的多尺度特征信息，重建医学图像目标区域的空间细节信息。脊柱和脑胶质瘤图像分割试验结果表明，DDS-UNet 模型对目标区域具有优异的特征提取和分割能力。消融试验进一步验证 DDS-UNet 模型对医学图像分割的有效性。

关键词：医学图像分割；双解码器；Swin Transformer；空洞卷积；多尺度特征融合
中图分类号：TP391 **文献标志码**：A
引用格式：刘全金，嵇文，胡浪涛，等. 基于双解码器的医学图像分割模型[J].山东大学学报（工学版），2024，54(6)：8-18.
LIU Quanjin，JI Wen，HU Langtao，et al. Medical image segmentation model based on double decoder[J]. Journal of Shandong University（Engineering Science），2024，54(6)：8-18.

Medical image segmentation model based on double decoder

LIU Quanjin[1]，JI Wen[1]，HU Langtao[1]，HUANG Huilei[1]，YANG Rui[1]，LI Xiang[2,3]，GAO Zewen[2,3]，WEI Benzheng[2,3*]
（1. School of Electronic Engineering and Intelligent Manufacturing，Anqing Normal University，Anqing 246133，Anhui，China；2. Center for Medical Artificial Intelligence，Shandong University of Traditional Chinese Medicine，Qingdao 266112，Shandong，China；3. Qingdao Academy of Chinese Medical Sciences，Shandong University of Traditional Chinese Medicine，Qingdao 266112，Shandong，China）

Abstract：Since the target area scales of medical images were different，and samples of labeled medical images were few，a dual decoder medical image segmentation model DDS-UNet was proposed. To be more specific，the DDS-UNet model used Swin Transformer module to construct the encoder，to extract multi-scale features of medical images. The decoder 1 took advantage of Swin Transformer module for global and remote semantic feature extraction to recover and aggregate the corresponding scale feature information of the encoder output step by step during the upsampling process. The decoder 2 made use of the local feature extraction advantage of convolutional neural networks（CNN）to recover the spatial information of medical images step by step during the upsampling process. The feature fusion module used the cavity convolution to decompose the deep semantic feature information output by the encoder，and collaboratively fused the multi-scale feature information output by the double decoders in the upsampling process，so as to reconstruct the spatial details of the target region of the medical image. The experimental results of spine and brain glioma image segmentation showed that the DDS-UNet model had significant abilities on feature extraction and segmentation for the target region. The ablation experiment further verified the effectiveness of the DDS-UNet model for medical image segmentation.
Keywords：medical image segmentation；double decoder；Swin Transformer；atrous convolution；multi-scale feature fusion

收稿日期：2023-07-21
基金项目：国家自然科学基金资助项目（62372280，61872225）；山东省自然科学基金资助项目（ZR2020KF013，ZR2020ZD44，ZR2019ZD04，ZR2020QF043）；山东省高校青创引才育才计划资助项目（2019-173）；青岛市科技惠民示范专项资助项目（23-2-8-smjk-2-nsh）
第一作者简介：刘全金（1971— ），男，安徽寿县人，教授，硕士生导师，博士，主要研究方向为机器学习、医学图像处理、智能无线通信。E-mail：liuquanjin@aqnu.edu.cn
*通信作者简介：魏本征（1976— ），男，山东临沂人，教授，博士生导师，博士，主要研究方向为医学人工智能、计算医学、机器学习等。E-mail：wbz99@sina.com

0　引言

医学影像分割是医学领域的研究重点,为医生提供更准确和可靠的诊断工具,也为研究人员理解疾病发展机制提供更深入的手段[1]。本研究以脊柱图像和结构相对复杂的脑胶质瘤图像为例,研究医学图像分割问题。

X 光片、计算机断层扫描(computed tomography, CT)和磁共振成像(magnetic resonance imaging, MRI)等脊柱影像在脊柱疾病诊断中不可或缺,医生根据脊柱影像识别、定位脊柱畸形、椎间盘突出等异常情况[2]。

脑胶质瘤起源于大脑和脊髓中的胶质细胞癌变[3],分为低级别胶质瘤(low grade glioma, LGG)和高级别胶质瘤(high grade glioma, HGG)[4]。HGG 具有强烈的浸润性和侵袭性,患者生存期较短;LGG 生长缓慢,患者通常具有较长的生存期。精准分割脑胶质瘤病灶图像有助于临床医生准确诊断肿瘤类别,提升患者生存期。

基于卷积神经网络(convolutional neural networks, CNN)的 U 型架构利用跳跃连接[5],将解码器子网络的边缘特征和语义特征与来自编码器子网络的多尺度特征相结合,对恢复图像细节方面非常有效。U 型架构在医学图像分割中取得了成功,但仍有提升空间。由于 CNN 感受野有限,卷积核只能关注图像中的局部区域,不利于建立特征间远程依赖关系,无法捕获全局上下文信息[6]。文献[7]尝试使用注意力机制进行远程依赖性建模,但在图像分割时仍存在较大的局限性。U-Net 的跳跃连接和多尺度特征融合方法对于小尺度目标分割效果有限。研究人员利用基于 CNN 的双解码器提高图像分割质量:文献[8]提出双解码网络,在编码器的最后一层引入 Transformer 的自注意力机制,以便捕获全局上下文信息,采用双解码结构进一步解析语义特征,实现对新冠肺炎病灶的精准分割;文献[9]提出一种双解码 U 型卷积神经网络,将低级空间信息和高级语义信息融合,提高分割网络对特征信息的有效利用;文献[10]提出一种基于双解码器的脑肿瘤图像分割模型,通过上下文语义解码路径对图像进行初步语义信息分割,利用空间解码路径将粗粒度的语义分割图与空间信息相结合,提取更丰富的信息,提高分割的准确度;文献[11]提出一种双解码器神经网络用于遥感图像特征提取,采用双解码器逐层融合语义特征,用于弥补传统 UNet 上采样信息丢失过多的问题。

视觉 Transformer(vision Transformer, ViT)是第一个纯粹基于 Transformer 的图像识别模型,性能与基于卷积的优秀方法相当[12-13]。可变形 Transformer(deformable Transformer, DETR)是基于 Transformer 构建的第一个完全端到端的对象检测模型[14],利用 Transformer 的多头自注意力(multi-head self-attention, MSA)机制分析特征图序列之间的相关性,提取图像全局特征信息。微软研究院提出 Swim Transformer 算法[15],基于移位窗口的多头自注意力(shifted window multi-head self-attention, SW-MSA)机制提取图像上下文信息,实现层级式多尺度特征提取,在图像检测任务中取得了显著的突破,超越了以往最先进的方法。Transformer 模型也应用在医学图像分割研究中:文献[16]利用 CNN 提取图像特征,基于 Transformer 构建特征间的远程依赖关系;文献[17]基于 Transformer 和 CNN 图像特征提取优势,提出一种图像特征融合方法,提高模型的分割性能;医学 Transformer(medical Transformer, MedT)方法探索了小规模数据集情况下应用 Transformer 的可行性[6];Swin-UNet 基于 Transformer 构建 U 型编解码结构,取得了良好的性能[18]。上述方法的图像分割结果表明 Transformer 在医学图像分割中存在巨大潜力。

CNN 的归纳偏置特性使之适于小数据集,且在捕获局部信息方面能力较强。相较于 CNN,Swin Transformer 在捕获全局特征和上下文信息方面能力较强,只是需要的训练样本特别多[19]。鉴于有标注的医学图像样本相对较少,本研究将具有全局和远程语义特征提取优势的 Swin Transformer 与具有局部特征提取优势的 CNN 相结合,基于 U 型架构提出一种基于双解码器的医学图像分割模型(dual-decoding Swim-UNet, DDS-UNet)。DDS-UNet 的编码器利用 Swin Transformer 模块提取图像多尺度特征;解码器 1 基于 Swin Transformer 恢复空间特征,解码器 2 基于 CNN 恢复空间特征;特征融合模块在分解编码器深层语义特征基础上,协同融合 2 个解码器输出的多尺度特征,恢复医学图像细节信息,实现医学图像分割。在脊柱数据集 VerSe[20-21]和脑胶质瘤数据集 BraTs[22]上的分割结果表明,DDS-UNet 算法的语义分割性能优于 UNet 和 Swin-UNet 算法。

1　医学图像分割算法 DDS-UNet

本研究基于 U 型架构提出一种双解码器的端

到端医学图像分割模型 DDS-UNet。鉴于有标注医学图像样本相对较少，本研究提出分别利用具有全局和远程语义特征提取优势的 Swin Transformer 和训练样本需要较少且具有局部特征提取优势的 CNN 构建 2 个解码器，通过多尺度特征融合模块（multi-scale feature fusion module，MFFM）协同 2 个解码器实现多尺度特征融合。

双解码 DDS-UNet 模型如图 1 所示，编码器基于 Swin Transformer 模块提取分辨率为 1/4、1/8、1/16和 1/32 的 4 个尺度特征图。

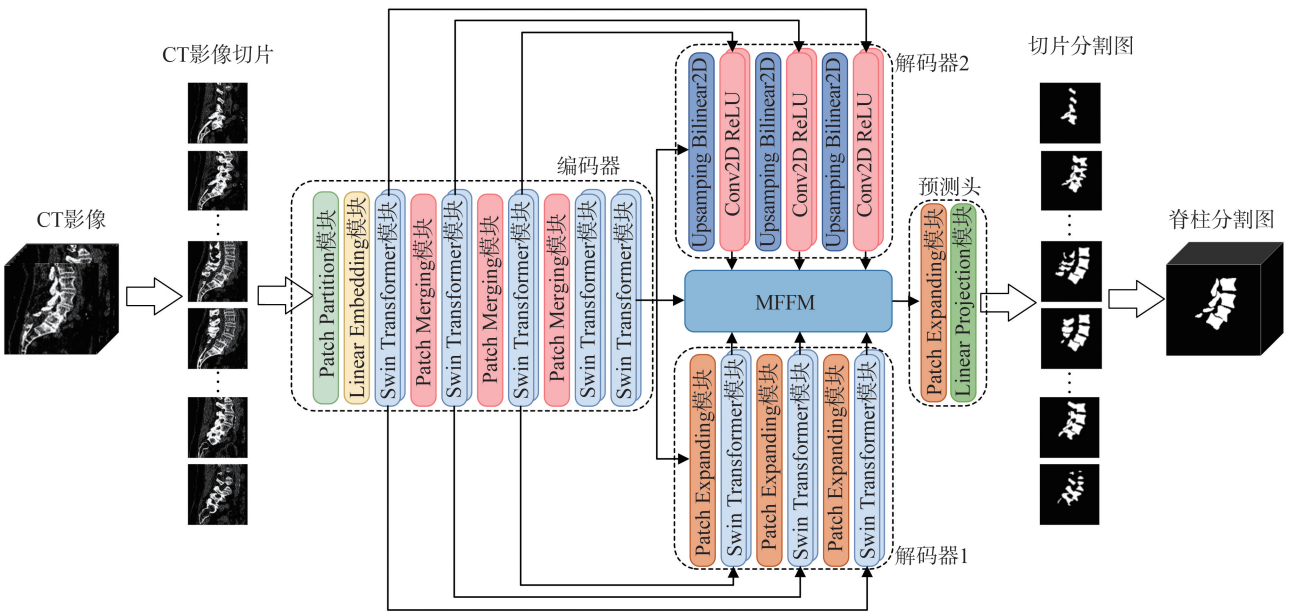

图 1 双解码 DDS-UNet 模型

Fig.1 Double decoding DDS-UNet model

解码器 1 利用 Patch Expanding 模块和 Swin Transformer 模块逐层恢复图像空间特征信息；解码器 2 利用 Upsampling 和卷积模块逐层恢复空间图像特征信息；2 个解码器通过跳跃连接分别与 3 个编码器层相接，减少因编码器下采样导致的空间信息损失。MFFM 模块协同解码器 1 和解码器 2，基于编码器输出的分辨率为 1/32 深层语义特征图，逐层融合 2 个解码器输出多尺度特征图，预测头通过线性映射层完成医学图像分割。

图 1 以脊柱 CT 图像分割为例，展示了 DDS-UNet 模型的医学图像分割流程：医学图像横断面的切片图像输入 DDS-UNet 模型；编码器提取脊柱切片图像 4 个尺度特征图；解码器 1 和解码器 2 分别基于 Swin Transformer 和 CNN 恢复切片图像空间特征信息；MFFM 模块融合编码器和 2 个解码器输出特征信息；预测头利用线性映射层分割脊柱切片图像；依次整合分割好的脊柱切片图像，生成完整的三维分割图。

1.1 编码器模块

与 CNN 处理图像的思路不同[23]，Swin Transformer 利用移动窗口自注意力获取上下文特征信息，能更好地关注医学图像关键区域的特征。

对于维度为 $H\times W\times C$ 的医学图像，编码器经 4 次下采样分别提取 4 个尺度特征图，其中，H 为图像像素的行数，W 为图像像素的列数，C 为图像的通道数。利用 Patch Partition 模块和 Linear Emebeding 模块作下采样[15]，由 Swin Transfomrer 模块提取图像浅层特征；由 Patch Merging 模块和 Swin Transfomrer 模块提取 3 个深层特征图。

以维度为 320 × 320 × 3 的图像为例，Patch Partition 模块将图像以 4×4 区域分块，R、G、B 通道数据展平后是 48 个通道，得到的数据结构为 80×80×48。Linear Embeding 模块通过线性层将通道数据映射到 Swin Transfomrer 模块嵌入特征（维度为 96），得到数据结构为 80×80×96。

所有 Swin Transformer 模块结构相同，包括窗口多头自注意力（windows multi-head self-attention，W-MSA）层、SW-MSA 层、多层感知机（multi-layer perceptron，MLP）和层归一化（layer normalization，LN），如图 2 所示[15]。

W-MSA 层将图像块或特征块划分为互不重叠的局部区域，计算每个区域内自注意力，同时做跳跃连接：

$$\hat{z}^{l}=\text{W-MSA}(\text{LN}(z^{l-1}))+z^{l-1}, \tag{1}$$

式中：W-MSA( ) 为计算窗口数据的自注意力函数；z^{l-1} 为第 $l-1$ 块 Swin Transformer 模块输出；LN( ) 为

层归一化函数,对输入数据进行标准化。

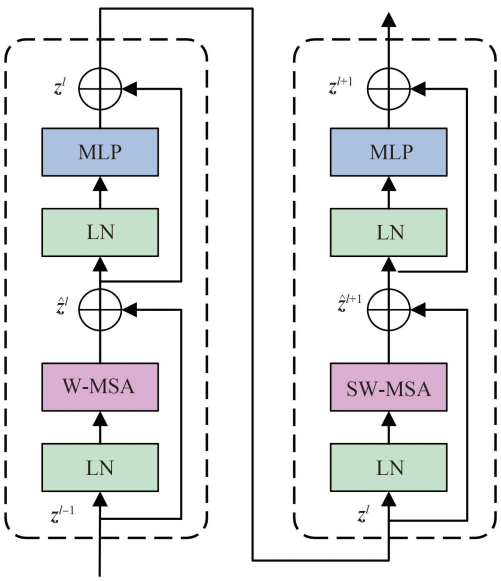

图 2　Swin Transformer 模块对结构图

Fig.2　The structure of Swin Transformer block

自注意力

$$\boldsymbol{A}_{\text{ttention}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d}}\right)\boldsymbol{V}, \tag{2}$$

式中,$\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\in\mathbf{R}^{M^2\times d}$ 分别为查询矩阵、键矩阵和值矩阵,d 为查询或键的维数。

LN 层输出数据经过展平后送入 MLP 层。MLP( )为多层感知机,包括 2 个全连接层,全连接层激活函数为 GELU。MLP 层的输出

$$z^l=\text{MLP}(\text{LN}(\hat{z}^l))+\hat{z}^l。 \tag{3}$$

SW-MSA 层通过窗口移动计算不同窗口之间的信息交互,获取上下文语义特征,同时做跳跃连接:

$$\hat{z}^{l+1}=\text{SW-MSA}(\text{LN}(z^l))+z^l, \tag{4}$$

式中 SW-MSA( )为计算移动窗口数据的自注意力。

$\hat{z}^{l+1}$ 经 MLP 层和跳跃连接,得到第 l+1 块 Swin Transformer 模块输出

$$z^{l+1}=\text{MLP}(\text{LN}(\hat{z}^{l+1}))+\hat{z}^{l+1}。 \tag{5}$$

编码器通过 4 次下采样降低特征空间维度,利用 Swin Transfomrer 模块先后提取 4 个尺度特征图。Swin Transformer 模块的 W-MSA 层和 SW-MSA层组对完成对图像或特征的注意力分析。W-MSA 和 SW-MSA 相互配合,实现窗口内部和窗口之间块信息的传递和交互,提高模型提取图像全局特征信息的能力。

1.2　双解码模块

Swin-UNet 采用跳跃连接将编码器提取的特征与解码器解析的特征在通道方向合并、融合,提升分割网络的性能[18]。但是,Swin-UNet 对处理小尺寸目标分割任务的提升效果有限。同时,深层语义特征在上采样过程中与低级特征融合时,特征之间跨度较大,导致低级特征信息丢失[24]。另外,基于 Transformer 的 Swin-UNet 算法对训练样本的需要量较大。因此,在医学图像的小尺度目标分割中,Swin-UNet 的表现不佳[25]。

针对上述问题,本研究提出基于双解码器的 DDS-UNet 模型。利用 Patch Expanding 模块和 Swin Transformer 模块构建 DDS-UNet 模型的解码器 1,旨在利用 Swin Transformer 全局上下文建模优势,捕获更丰富的全局特征信息。基于上采样和卷积模块构建解码器 2,旨在利用卷积模块的局部感知性和参数共享优势,在提取更丰富局部细节信息的同时,提升模型效率和泛化能力。

如图 1 所示,解码器 2 通过卷积模块和双线性插值逐层解析来自编码器 Swin Transformer 模块提取的语义特征。解码器 2 利用卷积模块的局部感知性,恢复出比解码器 1 更多的医学图像局部细节和边缘特征。解码器 1 第 i 层输出特征图

$$\boldsymbol{F}_i^{\text{decoder1}}=\text{Swin-T}(\text{Cat}(\boldsymbol{F}_i^{\text{coder}},\text{Pe}(\boldsymbol{F}_{i-1}^{\text{decoder1}}))),\ i=1,2,3, \tag{6}$$

式中,$\boldsymbol{F}_i^{\text{coder}}$ 为编码器第 i 层输出特征图,Pe( )为 Patch Expanding 上采样操作,Cat( )为将特征图沿通道方向合并,Swin-T( )为 SwimTransformer 模块。解码器 2 第 i 层输出特征图

$$\boldsymbol{F}_i^{\text{decoder2}}=\text{Conv}(\text{Cat}((\boldsymbol{F}_i^{\text{coder}},\text{Up}(\boldsymbol{F}_{i-1}^{\text{decoder2}}))),\ i=1,2,3, \tag{7}$$

式中,Up( )为上采样操作,Conv( )为卷积操作。双解码的初始特征图均来自编码器的深层特征图,即:

$$\boldsymbol{F}_0^{\text{decoder2}}=\boldsymbol{F}_0^{\text{decoder1}}=\boldsymbol{F}_4^{\text{coder}}, \tag{8}$$

式中 $\boldsymbol{F}_4^{\text{coder}}$ 为编码器输出的最后一层特征图。

DDS-UNet 算法的 2 个解码器功能互补,利用 Swin Transformer 模块和卷积模块在上采样过程中分别恢复图像空间特征信息。一方面,每层解码模块将上采样模块得到高级特征信息与跳跃连接引入编码器提取的低级特征信息融合;另一方面,2 个解码器输出特征图维度相同,便于特征融合时实现二者之间协同工作。这种设计既能利用 Swin Transformer 建立全局的特征关联和上下文信息的交互,也能采用 CNN 捕捉图像的局部细节和边缘特征。

1.3　MFFM 模块

针对双解码器学习到的特征图语义特性差异,本研究基于特征金字塔架构[26]设计一种新多尺度特征融合模块 MFFM,如图 3 所示。MFFM 模块基

于编码器的第 4 层特征图逐级接收、融合双解码器的 3 层特征图。

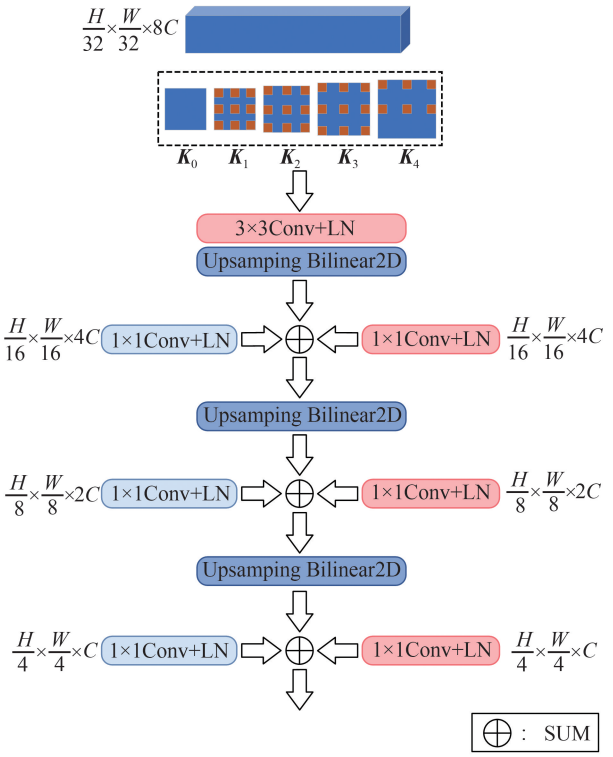

图 3 多尺度特征融合模块

Fig.3 Multi-scale feature fusion module

MFFM 模块将多级空洞卷积应用于具有丰富语义特征信息的编码器最后一层特征图，自适应地学习多尺度目标的语义特征，提取医学图像的全局先验表征信息，并融合多尺度特征信息。

对于维度 $H\times W\times C$ 的医学图像，编码器输出 4 个尺度的特征图，其中，最后一层特征图隐含丰富的深层语义特征信息。如图 3 所示，MFFM 模块用膨胀率为 1、3、5 和 7 的 4 个空洞卷积从最后一层特征图提取多尺度语义特征，扩大卷积核的感受野，以提高特征解析质量，保持分辨率不受损失。$\boldsymbol{K}_0$ 为该特征图，$\boldsymbol{K}_1$、$\boldsymbol{K}_2$、$\boldsymbol{K}_3$ 和 $\boldsymbol{K}_4$ 分别为 4 种空洞卷积结果。将 $\boldsymbol{K}_1$、$\boldsymbol{K}_2$、$\boldsymbol{K}_3$、$\boldsymbol{K}_4$ 和 $\boldsymbol{K}_0$ 这 5 个特征图沿通道方向合并，使用 3×3 卷积和层归一化操作将通道数统一为 D。上述过程可表述为：

$$\boldsymbol{K}_i=\mathrm{Conv}_{k_i}(\boldsymbol{F}_4^{\mathrm{coder}})+b,\ i=1,2,3,4, \tag{9}$$

$$\boldsymbol{F}_1=\mathrm{LN}(\mathrm{Conv}_{3\times 3}(\mathrm{Cat}(\boldsymbol{K}_0,\boldsymbol{K}_1,\boldsymbol{K}_2,\boldsymbol{K}_3,\boldsymbol{K}_4))), \tag{10}$$

式中，k_i 为空洞卷积膨胀率，$\mathrm{Conv}_{k_i}(\ )$ 为空洞卷积，$\boldsymbol{K}_i$ 为空洞卷积产生的特征图，$\boldsymbol{F}_1$ 为 MFFM 模块的第 1 层特征图。

双解码器输出 3 层特征图的维度如图 3 所示，利用 1×1 卷积将 3 层特征图的通道数统一为 D。解码器 1 输出第 i 层的特征层 $\boldsymbol{F}_i^{\mathrm{decoder1}}$ 和解码器 2 输出第 i 层的特征层 $\boldsymbol{F}_i^{\mathrm{decoder2}}$ 分别为：

$$\boldsymbol{F}_i^{\mathrm{decoder1}}=\mathrm{LN}(\mathrm{Conv}_{1\times 1}(\boldsymbol{F}_i^{\mathrm{decode1}})),\ i=1,2,3, \tag{11}$$

$$\boldsymbol{F}_i^{\mathrm{decoder2}}=\mathrm{BN}(\mathrm{Conv}_{1\times 1}(\boldsymbol{F}_i^{\mathrm{decode2}})),\ i=1,2,3, \tag{12}$$

式中 BN( ) 为批量归一化操作。

MFFM 模块通过与双解码器输出特征图相加，让解码器 1 和解码器 2 协同工作。MFFM 模块以特征图 $\boldsymbol{F}_1$ 为基础，逐层融合双解码器输出的特征图，获取整体和局部信息。MFFM 模块第 j 层生成的特征图

$$\boldsymbol{F}_j=\mathrm{Conv}(\boldsymbol{F}_{j-1}+\boldsymbol{F}_j^{\mathrm{decoder1}}+\boldsymbol{F}_j^{\mathrm{decoder2}}),\ j=2,3,4。 \tag{13}$$

MFFM 模块对编码器输出的第 4 层深层语义特征进行多膨胀率的空洞卷积，捕获多级粗、细粒度特征信息，避免单一卷积核因匹配度过低造成的细节缺失；通过协同解码器 1 和解码器 2，逐层融合 2 个解码器输出多尺度语义特征，由线性映射层实现医学图像分割。

2 试验及分析

为客观验证本研究提出的 DDS-UNet 医学图像分割算法性能，将 U-Net[5]、Swin-UNet[18]、FCN[27]、Deeplabv3[28] 和 DDS-UNet 等 5 种图像分割算法分别在内容相对简单的脊柱 CT 图像和内容相对复杂的脑胶质瘤 MRI 图像上进行分割试验对比。为确保试验对比的公平，4 种对比算法均使用作者公开的原始代码，5 种分割算法均在相同的脊柱和脑胶质瘤图像训练集和测试集中进行五折交叉校验。

所有试验均基于深度学习框架 PyTorch1.7 实现。试验操作环境为运行在 Ubuntu 18.04 版本的深度学习服务器，服务器配备 Intel Xeon Platinum 8268 CPU，拥有 32 GB 内存，搭载 NVIDIA Tesla V100 32 GB 显卡作为 GPU 加速设备。

2.1 试验数据

本研究使用的数据集为国际顶级人工智能医学影像学术会议 MICCAI 提供的 VerSe19、VerSe20 脊柱数据集[20-21] 和 BraTs2018 脑胶质瘤数据集[22]。脊柱数据集 VerSe 包含 355 例脊柱 CT 和人工标注标签，图像分辨率为 512×512×L，矢状面切片数 $L\in[121,525]$。剔除不含有标注的切片，以包含脊柱标注的切片作为试验数据。最终，本研究试验数据集由 10 460 张脊柱 CT 切片图像构成。

BraTs2018 脑胶质瘤数据集由 19 个机构采用不同的核磁共振扫描仪获得，为了确保数据的一致性，统一对数据集进行初步预处理。数据集包括 75 例 LGG 患者和 210 例 HGG 患者病例，每个病例包

含 T1 加权(T1)、T1 序列(T1ce)、T2 加权(T2)和液体衰减反转序列(FLAIR)模态的 MRI 图像,分辨率为 240×240×155。图像数据由专家手动标注,标签值与区域的对应关系为:0 对应背景,1 对应坏死和非增强肿瘤核心(necrotic and non-enhancing tumor, NCR/NET),2 对应水肿区域(edema, ED),4 对应增强肿瘤核心(enhancing tumor, ET)。脑肿瘤 MRI 图像及专家标注如图 4 所示,其中图 4(a)~(d)为患者 Brats18_CBICA_ASW_1_85 的脑肿瘤 MRI 图像,图 4(e)中伪彩图对应专家标注,3 个不同颜色区域表示 3 个类别。为评估分割算法的客观性,本试验将每个病例的 4 种模态叠加,对叠加后的数据进行分割试验。本研究试验数据集由 10 624张脑胶质瘤 MRI 切片图像构成。DDS-UNet 医学图像分割算法的实现细节和源代码见文献[29]。

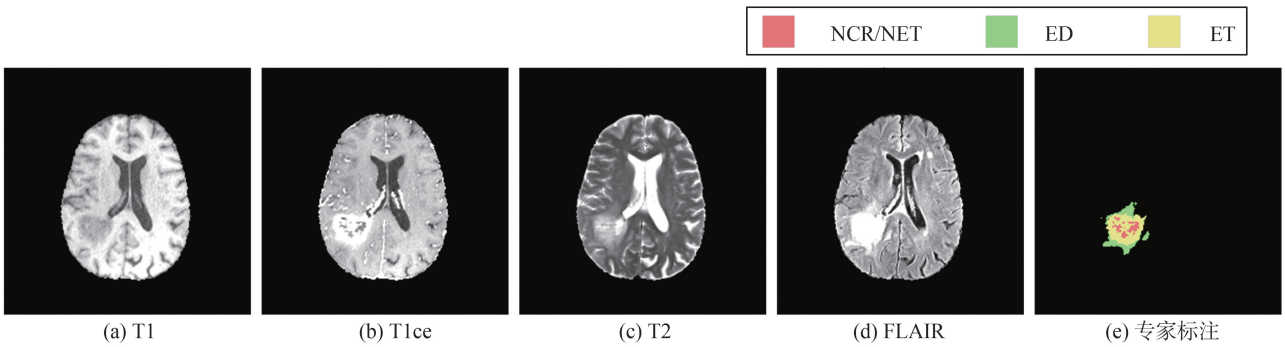

图 4　脑肿瘤 MRI 图像及专家标注

Fig.4　Brain tumor MRI images and expert annotation

2.2　图像分割评价指标

为验证所提医学图像分割算法性能,本研究采用在语义分割和影像分割算法性能评价中常用的骰子系数 D_{ice}、交并比 M_{IoU}和准确率 M_{ACC}指标对脊柱图像和脑胶质瘤图像分割效果进行定量分析,具体公式如下:

$$D_{ice}(A,B)=\frac{2S(A\cap B)}{S(A)+S(B)}, \tag{14}$$

$$M_{IoU}=\frac{S(A\cap B)}{S(A\cup B)}, \tag{15}$$

$$M_{ACC}=\frac{T_P+T_N}{T_P+F_N+F_P+T_N}, \tag{16}$$

式中,A 为图像分类算法分割区域,B 为专家标注区域,$S(\ )$为面积,T_P为预测为正类样本中实际属于正类的数量,F_P为预测为正类样本中实际属于负类的数量,F_N为预测为负类样本中实际属于正类的数量,T_N为预测为负类样本中实际属于负类的数量。

2.3　脊柱图像分割试验结果及分析

将 U-Net、Swin-UNet、FCN、Deeplabv3 和 DDS-UNet 算法在 VerSe19 和 VerSe20 组成的训练集上进行图像分割训练。5 种图像分割算法在测试集上的脊柱分割检测试验结果如表 1 所示。由表 1 可知,本研究提出的 DDS-UNet 算法在 D_{ice}、M_{IoU}和 M_{ACC}指标上均取得了最优的分割结果。相较于 U-Net算法,DDS-UNet 的 D_{ice}、M_{IoU}和 M_{ACC}分别提高了 0.89 百分点、5.96 百分点和 1.87 百分点;DDS-UNet 算法更优于 FCN 和 Deeplabv3 算法的分割结果,表明 DDS-UNet 算法比这 3 个算法提取到更全面的全局上下文信息和局部信息。

表 1　5 种算法在 VerSe 测试集上的分割结果

Table 1　Segmentation result of the five image segmentation algorithms on VerSe testsets　　单位:%

算法	D_{ice}	M_{IoU}	M_{ACC}
U-Net	90.28±1.25	86.27±0.83	88.75±1.34
Swin-UNet	86.12±0.37	84.86±0.77	85.66±0.83
FCN	84.48±1.56	83.56±1.17	86.49±1.32
Deeplabv3	89.67±1.42	85.13±1.34	88.47±1.54
DDS-UNet	**91.17±0.53**	**92.23±0.37**	**90.62±0.74**

相较于 Swin-UNet 算法,DDS-UNet 算法的 D_{ice}、M_{IoU}和 M_{ACC}提升效果更为显著,分别提升了 5.05百分点、7.37百分点和 4.96 百分点,表明 DDS-UNet 算法既利用了解码器 2 中卷积模块的局部信息提取优势,也通过 MFFM 模块有效聚合了双解码器传递的语义特征。试验结果表明,本研究算法能有效提升脊柱图像分割精度,可以辅助临床医生进一步提升脊柱疾病诊断效率和准确率。

U-Net、Swin-UNet 和 DDS-UNet 图像分割算法对 VerSe 数据集脊柱不同位置切片图像的分割效果对比如图 5 所示,其中红框区域为误分割和漏分割情况。图 5(a)(f)(k)(p)是 4 幅脊柱切片原图,图 5(b)(g)(l)(q)是专业标注的真实标签,图 5(c)(h)(m)(r)是 U-Net 算法分割效果图,图 5(d)(i)(n)(s)是 Swin-UNet 算法分割效果图,图 5(e)(j)(o)(t)是 DDS-UNet 算法分割效果结果图。

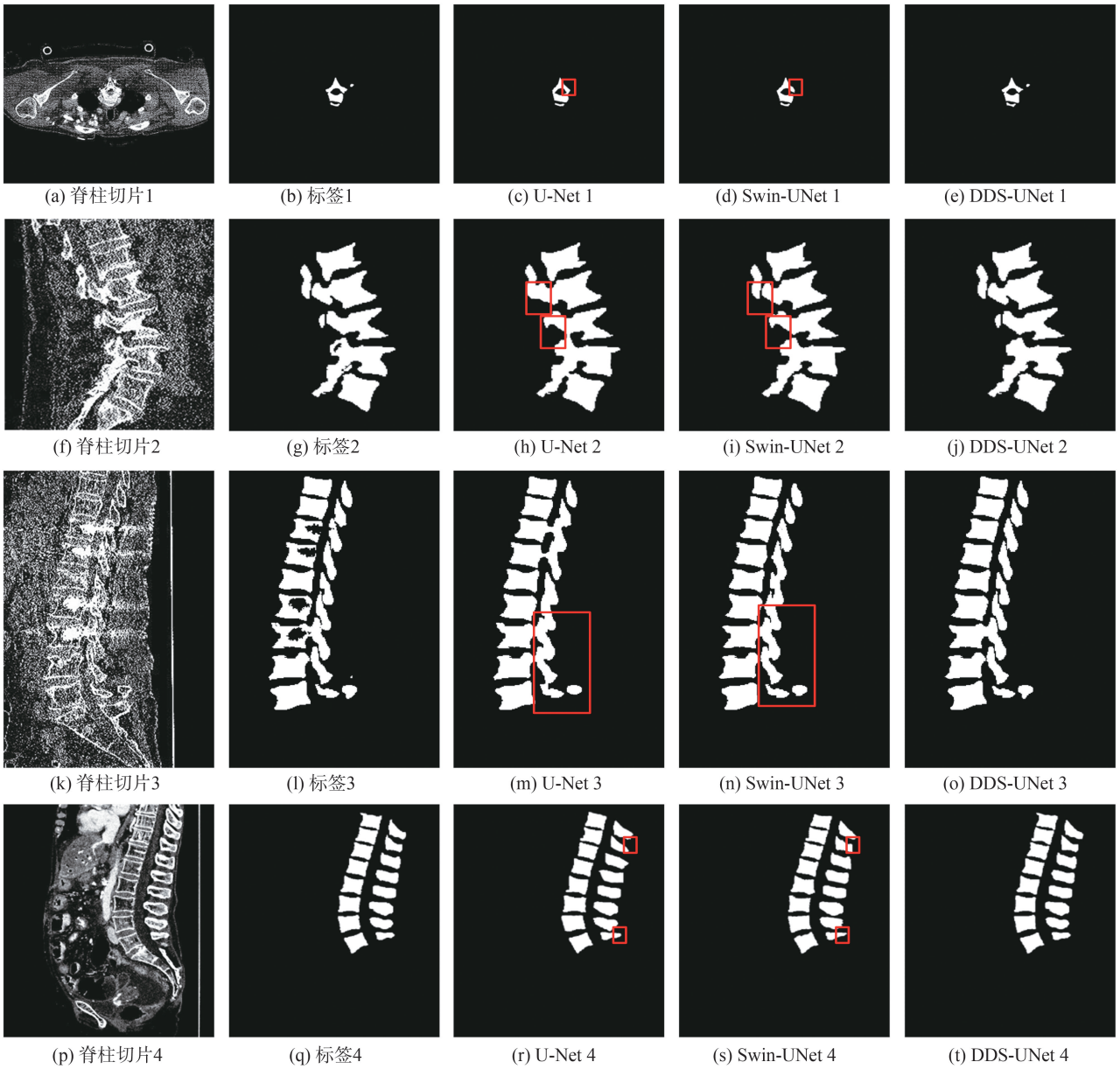

图 5 脊柱图像分割结果

Fig.5 Spinal image segmentation results

在图 5(a)脊柱切片图像中,脊柱与周围身体组织对比度低。由图 5(c)~(e)可知:U-Net 和 Swin-UNet 算法有明显的误分割和漏分割情况,DDS-UNet 算法能够捕获更多的空间细节特征,在定位脊柱位置方面表现出较好的性能,分割结果更接近真实标签。图 5(f)中脊柱切片图像受噪声干扰严重,背景灰度与待分割的脊柱几乎一致。由图 5(h)~(j)可知:U-Net和 Swin-UNet 算法出现误分割的情况,而 DDS-UNet 算法在整体和细节方面与真实标签吻合度更高,脊柱边界能平滑过度。图 5(k)中脊柱切片图像同时受到噪声和伪影的双重干扰。由图 5(m)~(o)可知:U-Net 和 Swin-UNet 算法抑制背景噪声干扰能力较差,分割结果存在漏分割和误分割现象,而 DDS-UNet 算法仍能较好定位脊柱边界。图 5(p)中脊柱切片图像受干扰相较少,3 种算法分割效果判别不大。

由图 5 中 3 种图像分割算法的脊柱图像分割效果可知,DDS-UNet 算法能构建脊柱图像的全局与局部联系,细化边缘特征,较好定位脊柱整体和细节。DDS-UNet 算法的特征提取和边界处理能力使其即便在图像受严重干扰的情况下也能精准定位脊柱,减少误分割和漏分割。分割效果对比进一步说明 DDS-UNet 算法的优越性。

2.4 脑胶质瘤图像分割试验结果及分析

将 U-Net、Swin-UNet、FCN、Deeplabv3 和 DDS-UNet 算法在 BraTs 训练集训练,损失函数选择四分类交叉熵。

5 种图像分割算法在测试集上的脑胶质瘤分割检测结果如表 2 所示。

表 2　5 种图像分割算法在 BraTs 测试集上的分割结果
Table 2　Segmentation results of the five image segmentation algorithms on BraTs testset

单位：%

算法	D_{ice}	M_{IoU}	M_{ACC}
U-Net	89.28±1.36	84.66±0.43	89.75±1.48
Swin-UNet	85.61±0.48	83.46±0.67	86.73±0.76
FCN	81.81±2.63	79.41±1.76	81.13±1.96
Deeplabv3	86.27±1.72	84.13±1.82	87.54±1.33
DDS-UNet	**92.37±0.39**	**91.43±0.77**	**92.57±0.63**

由表 2 可知，本研究所提 DDS-UNet 算法在 D_{ice}、M_{IoU} 和 M_{ACC} 3 项指标上均取得了最优的分割结果。对比 U-Net 算法，DDS-UNet 的 D_{ice}、M_{IoU} 和 M_{ACC} 分别提高了 3.09 百分点、6.77 百分点和 2.82 百分点；同样，DDS-UNet 算法也优于 FCN 和 Deeplabv3 算法分割结果，进一步表明 DDS-UNet 算法基于 Swin Transformer 模块和 MFFM 模块基于卷积模块能比 3 种算法提取到更丰富的全局上下文信息和局部信息。相较于 Swin-UNet 算法，DDS-UNet 算法的 D_{ice}、M_{IoU} 和 M_{ACC} 有更大提升，分别提升了 6.76百分点、7.97 百分点和 5.84 百分点，表明 DDS-UNet 算法既利用了双解码器中卷积模块的局部信息提取优势，也通过 MFFM 模块使双解码器协同工作，有效聚合了双解码器送来的特征信息。脑胶质瘤分割试验结果表明，本研究算法同样能有效提升脑胶质瘤图像分割精度，可以辅助临床医生进一步提升脑胶质瘤诊断效率和准确率。

本研究选择 3 个患者病例，其脑肿瘤的大小、形状和位置有较大差异。图 6 将分割结果叠加在对应的 FLAIR 切片上展示，图 6（a）（f）（k）分别为 FLAIR 轴向切片图像，图 6（b）（g）（l）为真实标注标签，图 6（c）（h）（m）为 U-Net 算法的分割结果，图 6（d）（i）（n） 为 Swin-UNet 算法的分割结果，图 6（e）（j）（o）为 DDS-UNet 算法的分割结果。由图 6 可知，相比于 U-Net 和 Swin-UNet 算法，DDS-UNet 分割效果更好，即使对于较难分割的 NCR/NET 区域（浅红色）与 ET 区域（黄色），仍能较为完整地保留特征信息。对于病例 1（图 6（a）~（e）），U-Net 和 Swin-UNet 算法的 ED 区域和 NCR/NET 区域误分割较多（红框区域），而本研究算法极大改善了这一问题；对于病例 2（图 6（f）~（j）），U-Net 和 Swin-UNet 算法的 ET 区域、ED 区域和 NCR/NET 区域误分割较多，将很多 ED 区域分割为 ET 区域，DDS-UNet 较准确地分割出了整个肿瘤区域；DDS-UNet 分割效果优于其他 2 种算法，分割的 ED 区域几乎与真实标签重合。总体来说，DDS-UNet 比 U-Net 和 Swin-UNet 算法分割更精细，预测结果更接近真实标签。

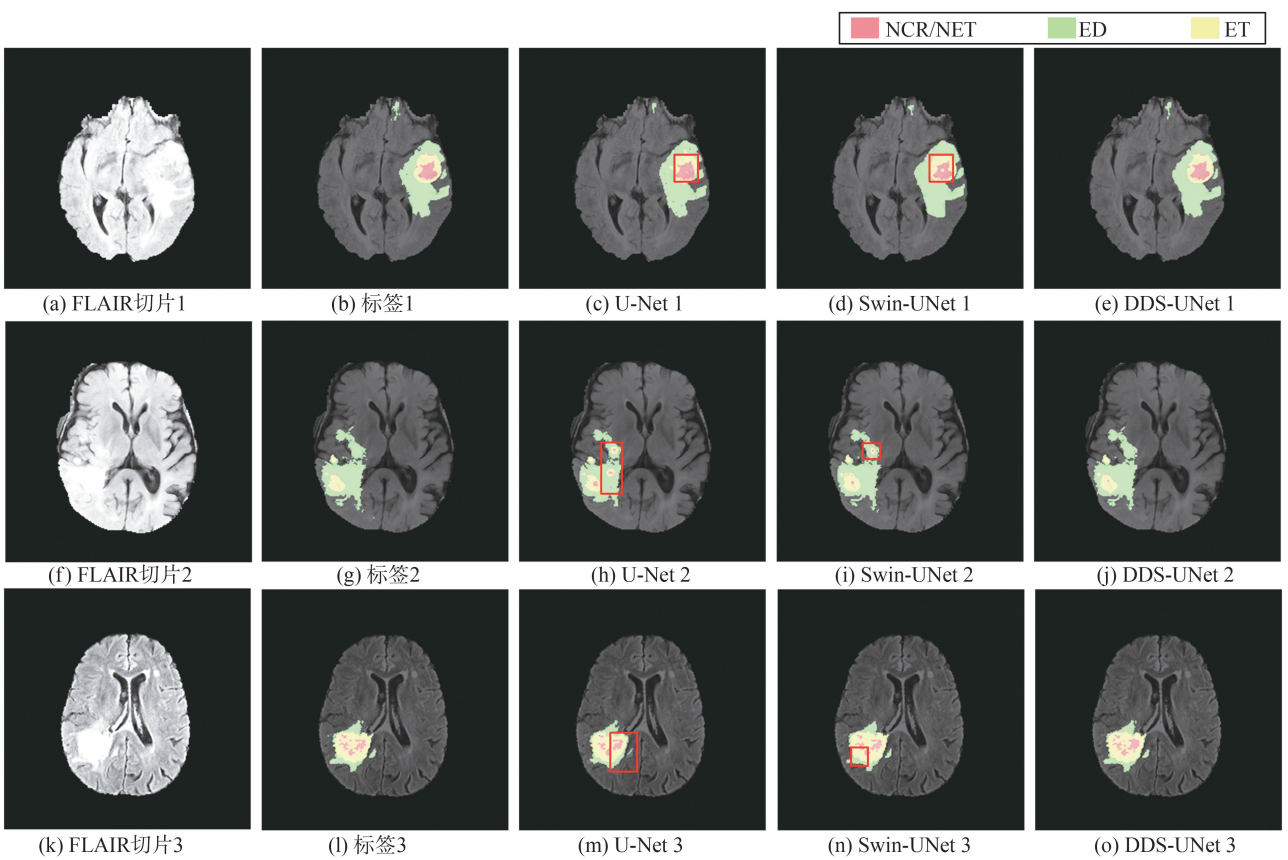

图 6　脑肿瘤图像分割结果
Fig.6　Brain tumor image segmentation results

比较5种图像分割算法在VerSe和BraTs测试集上的分割结果，擅长分割自然图像的Swin-UNet算法却在2个医学图像分割上表现欠佳。这是因为Swin-UNet算法需要海量的图像训练样本才能达到需要的分割性能。

2.5 消融试验

为验证所提双解码器和MFFM模块对医学图像分割算法的作用，本研究使用控制变量方法在VerSe和BraTs测试集上进行消融试验。定量消融试验结果如表3所示，“Swin-T+解码器1”表示U型架构下以Swin-T为编码器骨干网，以DDS-UNet模型的解码器1为解码器，即Swin-UNet模型；“Swin-T+解码器1+解码器2”表示U型架构下以Swin-T为编码器骨干网，以DDS-UNet模型的双解码器为解码器；“Swin-T+解码器1+解码器2+MFFM”表示U型架构下以Swin-T为编码器骨干网，DDS-UNet模型的双解码器为解码器，并用MFFM模块进行特征融合，即DDS-UNet模型。对于BraTs数据，将编码器的骨干网Swin-T换为Swin-B。

表3 VerSe和BraTs数据集上的消融试验结果

Table 3 Results of ablation experiments on VerSe and BraTs datasets 单位：%

数据集	Swin-T+解码器1	解码器2	MFFM	D_{ice}	M_{IoU}	M_{ACC}
VerSe	√	×	×	86.12±0.37	84.86±0.77	85.66±0.83
	√	√	×	88.47±0.67	87.13±0.26	87.41±0.57
	√	√	√	**91.17±0.53**	**92.23±0.37**	**90.62±0.74**
BraTs	√	×	×	85.61±0.48	83.46±0.67	86.73±0.76
	√	√	×	89.21±0.86	87.49±0.43	88.39±0.24
	√	√	√	**92.37±0.39**	**91.43±0.77**	**92.57±0.63**

表3展示了解码器2模块和MFFM模块对医学图像分割精度的影响。对于脊柱图像数据，Swin-UNet在添加解码器2模块后，D_{ice}、M_{IoU}和M_{ACC}分别提高了2.35百分点、2.27百分点和1.75百分点，表明解码器2模块能构建局部与全局的联系，进一步细化边缘信息。在双解码器模块加持MFFM模块后，D_{ice}、M_{IoU}和M_{ACC}又分别提高了2.70百分点、5.10百分点和3.21百分点，表明MFFM模块能够协同2个解码器的多尺度语义特征，融合上下文信息，增强特征的表示能力。

对于脑胶质瘤图像数据，Swin-UNet在添加解码器1和MFFM 2个模块后，D_{ice}、M_{IoU}和M_{ACC}同样得到提升。消融试验证明，DDS-UNet提出的双解码器在MFFM调控下协同工作，可有效提取图像数据的局部和全局信息，建立图像信息间远程依赖关系；同时也验证了MFFM模块能有效融合、学习多尺度语义特征信息，对于边缘模糊或形状不规则区域的分割具有一定的优化效果，可进一步提高算法的分割性能。

3 结论

从脊柱和脑胶质瘤图像的分割结果及消融试验结果可知，DDS-UNet模型在医学图像目标区域尺度不一时，特征金字塔结构保证其可以接收不同尺度目标的特征信息。一方面，层级式Swin Transformer能够提取到不同尺度目标的关键信息；另一方面，基于Swin Transformer和卷积模块的双解码器在上采样过程从不同角度聚合了丰富的多尺度语义信息。MFFM基于编码器输出的深层语义特征，协同融合双解码器的多尺度语义信息，进一步提升了多尺度目标分割能力。

图像分割结果表明，本研究所提双解码器模块和MFFM模块有效提高了图像分割算法的泛化性能和分割精准度。同时，解码器2和MFFM模块的引入使基于Swin Transformer模块的图像分割模型在没有海量训练样本的情况下得到较好的分割效果。

整合DDS-UNet模型得到的所有切片分割图像，生成脊柱和脑胶质瘤三维分割图像。这种三维分割图像为临床医生提供直观准确的病灶立体场景，能有效辅助临床医生诊断脊柱和脑胶质瘤疾病，提升诊断效率。

综上，DDS-UNet模型既能满足医学图像分割的精准性需求，也为临床医生提供一种可选择的辅助诊断工具。但DDS-UNet模型仍有提升空间。未来将研究通过无监督学习提升编码器特征提取性能，以期在有限医学图像样本情况下提高模型分割性能。

参考文献:

[1] SHAMSHAD F, KHAN S, ZAMIR S W, et al. Transformers in medical imaging: a survey[J]. Medical Image Analysis, 2023, 88: 102802.

[2] BAUR D, KROBOTH K, HEYDE C, et al. Convolutional neural networks in spinal magnetic resonance imaging: a systematic review [J]. World Neurosurgery, 2022, 166: 60-70.

[3] AHIR B K, NGELHARD H H, AKKA S S. Tumor development and angiogenesis in adult brain tumor: glioblastoma[J]. Molecular Neurobiology, 2020, 57: 2461-2478.

[4] LOUIS D N, PERRY A, WESSELING P, et al. The 2021 WHO classification of tumors of the central nervous system: a summary[J]. Neuro-Oncology, 2021, 23(8): 1231-1251.

[5] OLAF R, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]// Proceedings of the 18th Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: MICCAI, 2015: 234-241.

[6] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical Transformer: gated axial-attention for medical image segmentation[C]//Proceedings of the 24th Medical Image Computing and Computer Assisted Intervention. Strasbourg, France: MICCAI, 2021: 36-46.

[7] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: learning where to look for the pancreas [EB/OL]. (2018-05-20) [2023-07-08]. https://arxiv.org/abs/1804.03999.

[8] HUANG X, CHEN J, CHEN M, et al. TDD-UNet: Transformer with double decoder UNet for COVID-19 lesions segmentation [J]. Computers in Biology and Medicine, 2022, 151: 106306.

[9] 毕秀丽, 陆猛, 肖斌, 等. 基于双解码 U 型卷积神经网络的胰腺分割[J]. 软件学报, 2022, 33(5): 1947-1958.
BI Xiuli, LU Meng, XIAO Bin, et al. Pancreas segmentation based on dual-decoding U-Net[J]. Journal of Software, 2022, 33(5): 1947-1958.

[10] 苏赋, 方东, 王龙业, 等. 基于双解码路径 DD-UNet 的脑肿瘤图像分割算法[J]. 光电子激光, 2023(3): 328-336.
SU Fu, FANG Dong, WANG Longye, et al. Brain tumor image segmentation algorithm based on dual decoding path DD-UNet[J]. Journal of Optoelectronics Laser, 2023(3): 328-336.

[11] 侯月武, 刘兆英, 张婷, 等. 基于改进的 DUNet 遥感图像道路提取[J]. 山东大学学报(工学版), 2022, 52(4): 29-37.
HOU Yuewu, LIU Zhaoying, ZHANG Ting, et al. Road extraction from remote sensing images based on improved DUNet[J]. Journal of Shandong University (Engineering Science), 2022, 52(4): 29-37.

[12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: NIPS, 2017: 6000-6010.

[13] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. (2021-06-03) [2023-07-08]. https://arxiv.org/abs/2010.11929.

[14] ZHU X, SU W, LU L, et al. Deformable DETR: deformable Transformers for end-to-end object detection [EB/OL]. (2021-05-18) [2023-07-08]. https://arxiv.org/abs/2010.04159.

[15] LIU Z, LIN Y, CAO Y, et al. Swin Transformer: hierarchical vision Transformer using shifted windows [C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 10012-10022.

[16] CHEN J, LU Y, YU Q, et al. TransUNet: Transformers make strong encoders for medical image segmentation [EB/OL]. (2021-02-08) [2023-07-08]. https://arxiv.org/abs/2102.04306.

[17] ZHANG Y, LIU H, HU Q. TransFuse: fusing transformers and CNNs for medical image segmentation [C]//Proceedings of the 24th Medical Image Computing and Computer Assisted Intervention. Strasbourg, France: MICCAI, 2021: 14-24.

[18] CAO H, WANG Y, CHEN J, et al. Swin-Unet: Unet-like pure Transformer for medical image segmentation [C]//European Conference on Computer Vision. Tel Aviv, Israel: Springer, 2022: 205-218.

[19] D′ASCOLI S, TOUVRON H, LEAVITT M L, et al. ConViT: improving vision Transformers with soft convolutional inductive biases[C]//Proceedings of the 38th International Conference on Machine Learning. [S.l.]: PMLR, 2021: 2286-2296.

[20] LOFFLER M, SEKUBOYINA A, JAKOB A, et al. A vertebral segmentation dataset with fracture grading[J]. Radiology: Artificial Intelligence, 2020, 2(4): e190138.

[21] SEKUBOYINA A, HUSSEINI M, BAYAT A, et al. VerSe: a vertebrae labelling and segmentation benchmark for multi-detector CT images[J]. Medical Image Analysis, 2021, 73: 102166.

[22] BAKAS S, REYES M, JAKAB A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge[EB/OL]. (2021-04-23) [2023-07-08]. https://arxiv.org/abs/1811.02629.

[23] 刘方旭, 王建, 魏本征. 基于多空间注意力的小儿肺炎辅助诊断算法[J]. 山东大学学报(工学版), 2023, 53(2): 135-142.
LIU Fangxu, WANG Jian, WEI Benzheng. Auxiliary diagnosis algorithm for pediatric pneumonia based on multi-spatial attention[J]. Journal of Shandong University (Engineering Science), 2023, 53(2): 135-142.

[24] 陆猛. 基于全卷积神经网络的医学图像分割[D]. 重庆: 重庆邮电大学, 2021.
LU Meng. Medical image segmentation based on full convolutional neural network[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2021.

[25] HE X, ZHOU Y, ZHAO J, et al. Swin Transformer embedding UNet for remote sensing image semantic segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-15.

[26] LIN T, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 2117-2125.

[27] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 3431-3440.

[28] CHEN L, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-12-05) [2023-11-08]. https://arxiv.org/abs/1706.05587.

[29] DDS-UNet 医学图像分割算法源代码[EB/OL]. (2023-07-20) [2023-11-08]. https://github.com/DeepPicEI/DDS-UNet.

(编辑:孙亚彤)

(上接第 7 页)

[34] WIEHMAN S, KROON S, DE VILLIERS H. Unsupervised pre-training for fully convolutional neural networks[C]// Proceedings of the 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference. Stellenbosch, South Africa: IEEE, 2016: 1-6.

[35] SUN X, YANG Z, ZHANG C, et al. Conditional gaussian distribution learning for open set recognition [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 13480-13489.

[36] CHOUDHURI N, GHOSAL S, ROY A. Nonparametric binary regression using a Gaussian process prior[J]. Statistical Methodology, 2007, 4(2): 227-243.

[37] SONG H, THIAGARAJAN J J, SATTIGERI P, et al. Optimizing kernel machines using deep learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(11): 5528-5540.

[38] MAMAT N, OTHMAN M F, ABDULGHAFOR R, et al. Enhancing image annotation technique of fruit classification using a deep learning approach[J]. Sustainability, 2023, 15(2): 1-19.

[39] SAMBATURU B, GUPTA A, JAWAHAR C V, et al. ScribbleNet: efficient interactive annotation of urban city scenes for semantic segmentation[J]. Pattern Recognition, 2023, 133: 109011.

(编辑:孙亚彤)