

基于全局多分支融合匹配注意力的医学图像分割模型

郭宇辰, 罗会兰

(江西理工大学 信息工程学院, 江西 赣州 342100)

摘要: 为了解决医学图像分割中模型忽略同一解剖结构在不同尺度下存在相似表征的问题, 提出基于全局多分支融合匹配注意力的模型(MFP Net)。多分支融合(MFF)模块将模型编码器不同阶段下不同尺度的特征进行统一, 并搜集相似的特征信息; 全局匹配注意力(GPA)模块将特征进行空间和通道上的对齐, 增强网络对于不同尺度结构的表征能力; 带注意力门控模块的解码器通过抑制冗余特征, 提升模型对关键特征的识别能力。实验结果表明, MFP Net可以通过学习图像中的相似表征, 提升图像分割精度。其在Synapse与ACDC数据集上的Dice指标达到79.28%和89.96%, 在Synapse数据集上的HD95指标达到18.28%, 显示出较先前方法具有更高的分割精度, 并能对内部结构进行更精确地识别。

关键词: 医学图像分割; 多分支融合; 全局匹配注意力

DOI: 10.11907/rjdk.241205

中图分类号: TP391.41

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)004-0147-07



Medical Image Segmentation Model Based on Global Multi-Branch Fusion Paired Attention

GUO Yuchen, LUO Huilan

(School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 342100, China)

Abstract: To address the issue of models in medical image segmentation overlooking similar representations of the same anatomical structure at different scales, a model based on global multi-branch fusion paired attention (MFP Net) is proposed. The Multi-branch Feature Fusion (MFF) module unifies features of different scales from various stages of the model encoder and collects these similar feature informations; the Global Paired Attention (GPA) module aligns features both spatially and across channels, enhancing the network's ability to represent structures at different scales; a decoder with attention-gated modules suppresses redundant features, boosting the model's recognition of key features. Results demonstrate that MFP Net, by learning similar representations within images, enhances segmentation accuracy, achieving Dice scores of 79.28% and 89.96% on the Synapse and ACDC datasets, respectively, and an HD95 metric of 18.28 on the Synapse dataset, indicating higher segmentation precision and more accurate internal structure identification compared to previous methods.

Key Words: medical image segmentation; multi-branch feature fusion; global paired attention

0 引言

医学图像分割旨在精确分割图像内的结构和病变区域, 因此在计算机辅助诊断和治疗规划方面发挥着重要作用。

随着计算机性能的提升, 深度学习技术极大地推动了医学图像分割技术的进步。其通过自动学习大量数据的复杂特征, 显著提升了分割的准确性和效率。

深度卷积神经网络(Deep Convolutional Neural Network, DCNN)已在一些任务上获得了成功, 但在图像分割

收稿日期: 2024-03-05

扫描二维码阅读全文:



基金项目: 国家自然科学基金项目(61862031); 江西省自然科学基金重点项目(20232ACB202011); 江西省主要学科科学技术带头人培养计划——领军人才项目(20213BCJ22004)

作者简介: 郭宇辰(1998-), 男, 江西理工大学信息工程学院硕士研究生, 研究方向为计算机视觉; 罗会兰(1973-), 女, 博士, 江西理工大学信息工程学院教授, 研究方向为计算机视觉。本文通讯作者: 郭宇辰。

任务中,由于全连接层信息丢失而限制了其应用^[1-4]。Long等^[5]通过去掉卷积网络后的全连接层适应图像分割任务,但是网络在上采样阶段仍存在特征丢失等问题。Ronneberger等^[6]提出U-Net,通过在上采样阶段拼接来自对称下采样阶段的特征,以解决信息丢失问题。此类对称的编码器—解码器结构应用于医学图像分割领域取得了较好效果^[7]。很多研究者在U-Net基础上继续深入研究,并提出了U-Net++^[8]、R2U-Net^[9]等方法。然而,以上基于CNN方法的模型无法建立全局和远程依赖的关系^[10]。为了解决此类问题,Chen等^[11]在解码器中引入深度可分离卷积,以增强感受野。Gu等^[12]提出密集空洞卷积模块和残差多核池化,以学习更多高级特征,保留更多空间信息。Oktay等^[13]在跳跃连接上使用注意力门控来突出显著特征,抑制不相关的噪声和特征冗余。但是,由于卷积计算自身的局限性,卷积网络感受野受限的问题依然没有得到解决。

Vaswani等^[14]在2017年首次提出自注意力机制(Self-Attention)替换循环神经网络,该方法在自然语言处理领域获得了巨大成功。Dosovitskiy等^[15]提出ViT(Vision in Transformer),并首次在大规模图像识别任务中应用了Transformer架构。其通过将2D图像转化成1D的图像序列,使模型建立了全局图像特征之间的依赖关系。但是,该方法依赖于大量预训练,计算复杂度高。为了解决这些问题,Liu等^[16]提出一种基于窗口限制和滑动窗口的Swin Transformer模型,有效降低了ViT模型的计算复杂度,解决了需要大量训练数据的问题。在此基础上,Chen等^[17]结合ResNet^[2]和ViT进一步提出对称编解码器架构的TransUnet,并获得成功。Wang等^[18]使用通道交叉融合注意力,进一步整合了通道维度的全局特征,增强了有效信息。然而,这些模型将重点放在提升网络对全局特征的关注上,却忽视了图像中不同解剖结构之间的相似性,导致出现网络在处理不同尺度结构时表征能力较弱的问题。

综上所述,本文提出一种基于全局多分支融合匹配注意力机制的医学图像分割模型(Multi-branch Feature Fusion Paired Attention Net, MFP Net)。具体内容包括:

(1)设计了多分支融合全局匹配注意力模块(Multi-branch Feature fusion global Paired Self-Attention, MFPSA)。通过多分支融合模块(Multi-branch Feature Fusion block, MFF)对编码器不同阶段的特征进行尺度和通道对齐,再进行拼接融合,以获取相似表征。该设计增强了模型对同一解剖结构在不同尺度下的特征提取能力,并解决了模型在分割不同尺度解剖结构时表征能力弱的问题。

(2)提出全局匹配注意力模块(Global Paired Multi-Head Self-Attention, GPMSA),利用空间和通道维度的自注意力机制对融合特征进行对齐学习,以更好地理解编码器不同阶段的信息,解决了融合后特征在空间和通道上存在偏差导致的模型对关键信息识别与利用能力差的问题。

(3)由于融合后特征冗余造成有用特征信息被稀释,在解码器添加注意力门控,以抑制冗余特征,从而增强边界信息,解决了因融合后特征被稀释导致的边缘分割精度下降的问题。

1 方法介绍

1.1 网络结构设计

本文提出基于全局多分支融合注意力的医学图像分割网络MFP Net结构,如图1所示。网络下采样阶段使用Swin Transformer对模型进行建模,获取全局特征。对不同Swin Transformer阶段的特征,在编码器底部使用全局多分支融合匹配注意力模块提取不同解剖结构之间的相似表征。在解码器部分,通过ResNet Block将模型对编码器不同阶段的结果还原至二维特征图。在模型底端,将编码器前3个阶段输出的特征张量输入至多分支特征融合注意力模块MFP SA。经过该模块计算后,与经过还原的维度为 $\frac{H}{32} \times \frac{W}{32} \times 8C$ 的张量进行拼接。然后,模型使用注意力门控对来自下一层和当前阶段的特征冗余进行抑制,使用反卷积对下一层特征进行上采样,并对其进行拼接。最后,模型经过一个输出层输出最终的预测结果。

1.2 多分支特征融合模块

为了提升网络对不同解剖结构的分割能力,以及对不同解剖结构之前存在相似局部特征的理解能力,本文提出一种多分支融合模块(Multi-branch Feature Fusion Block, MFF Block)。如图2所示,该模块将来自不同阶段的Swin Transformer编码器的特征进行汇集。针对第一阶段的特征,使用三层卷积最大池化下采样模块,将特征下采样至 $\frac{H}{32} \times \frac{W}{32} \times 8C$ 维度,使用卷积批归一化模块减少特征通道数量;针对第二阶段的特征,使用两层卷积最大池化下采样,再利用卷积批归一化模块压缩特征通道;针对第三阶段的特征,使用一层卷积最大池化下采样,利用卷积批归一化模块减少通道数量;针对最后阶段的特征,特征图大小已经相等,因此只需使用卷积批归一化模块减少特征通道数。最终将融合后的特征分别进行拼接,经过MFF-Out block输出,得到拼接后的特征 M_{out} 。此时,特征的维度依然是 $\frac{H}{32} \times \frac{W}{32} \times 8C$ 。MFF-Out block由基础的ResNet block组成。该模块的具体公式如下:

$$F^{n(i)}(x) = \sigma(\text{Max}(\text{Conv}(x)))_n = \begin{cases} 3\text{if } i = 1 \\ 2\text{if } i = 2 \\ 1\text{if } i = 3 \\ 0\text{if } i = 4 \end{cases} \quad (1)$$

$$G(x) = \sigma(\gamma(\text{Conv}(x))) \quad (2)$$

$$\text{MFF}_{out}(x_i) = x_{i-1} + \text{Conv}(x_i) \quad (3)$$

$$X_i = G(F^{n(i)}(\text{stage}_i))i \in [1, 2, 3, 4] \quad (4)$$

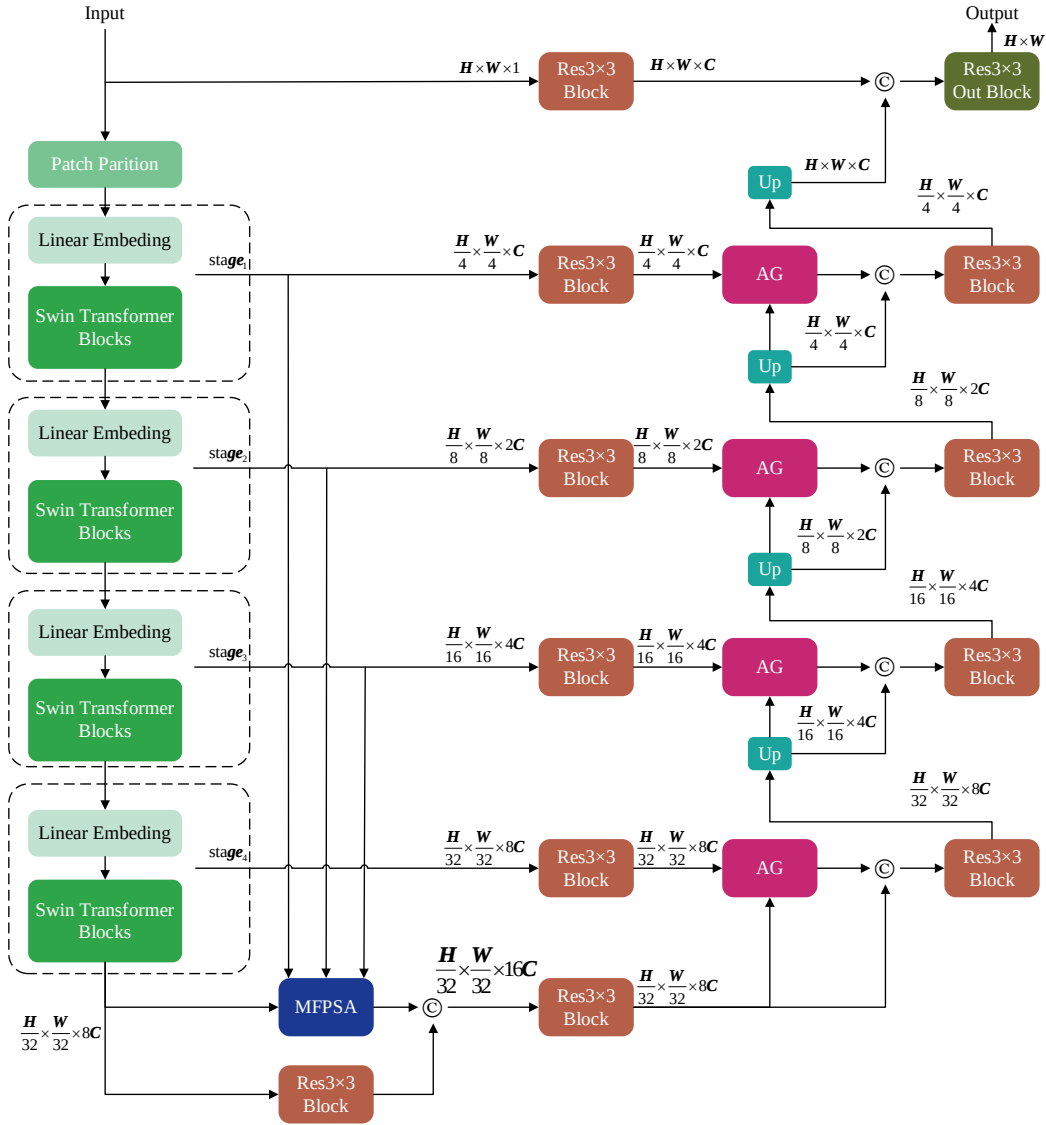


Fig. 1 Overview of model structure

图 1 模型结构总览

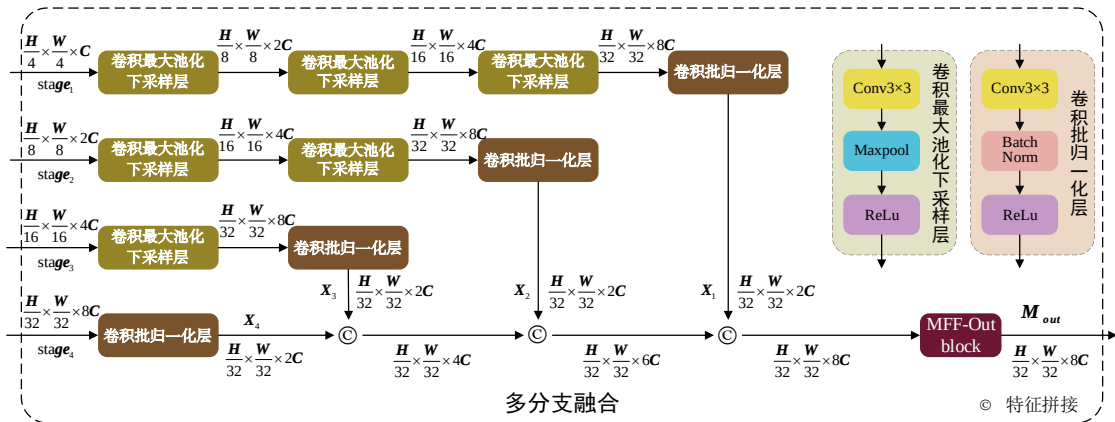


Fig. 2 Multi-branch feature fusion block

图 2 多分支特征融合模块

$$M_{out} = MFF_{out}([X_1, \dots, X_i]) \quad i \in [1, 2, 3, 4] \quad (5)$$

在公式(1)中, $F^{n(i)}(x)$ 定义为在 i 阶段时函数一共使

用 $n(i)$ 次。当 $i=1$ 时, 函数使用次数为 3 次; 当 $i=2$ 时, 函数使用次数为 2 次; 当 $i=4$ 时, 函数则不使用。 σ 表示 ReLu 激活函数, Max 表示最大池化下采样, γ 表示批归一化。对于

不同阶段的特征 $stage_i$, 使用不同次数的 $F^{n(i)}$ 函数对特征进行下采样。之后, 使用 ConvBNblock, 即 $G(x)$ 函数对特征进行通道压缩, 如公式(2)所示。接下来, 在由公式(1)和公式(2)结合所得的公式(4)中, 通过计算编码器各个阶段的特征 $stage_i$ 得到各阶段 $stage_i$ 对应的特征 X_i 。最后, 将 X_1, X_2, X_3, X_4 特征进行拼接, 使用残差模块 MFF_{out} 输出最终的特征 M_{out} , 如公式(5)所示。其中, 残差模块 MFF_{out} 计算原理如公式(3)所示。

1.3 全局匹配注意力模块

为了增强网络对融合后特征的匹配能力, 以更好地理解编码器不同阶段的特征, 本文设计了全局匹配注意力模块, 以增强网络对于不同尺度结构的表征能力。如图3所示, 该模块由全局匹配空间自注意力和全局匹配通道自注意力组成。首先, 全局匹配空间自注意力将混合后的特征 M_{out} 使用 W_s 矩阵线性映射成空间查询向量 Q_s 、空间键值向量 K_s 和空间权值向量 V_s 。然后将 Q_s 与 K_s 的转置相乘, 并使用 softmax 函数生成对应的空间自注意力权重矩阵, 将该矩阵与空间权值向量 V_s 相乘, 得到最终的注意力输出 $PSMA_{out}$ 。以上操作将混合后的特征在空间上进行自注意力计算, 使模型关注到特征图中不同位置之间的关系, 进而捕捉图像内部的空间与结构信息。此外, 全局匹配通道自注意力通过 W_c 矩阵将混合特征 M_{out} 分别生成通道查询向量 Q_c 、通道键值向量 K_c 和通道权值向量 V_c , 将 Q_c 的转置

与 K_c 相乘, 使用 softmax 函数生成对应的通道自注意力权重向量。最后, 将该向量与 V_c 相乘, 得到最终的通道自注意力输出 $PCMA_{out}$ 。该操作使模型关注到特征中不同通道之间的依赖关系, 从而有效地对特征进行选择并增强表述。全局匹配注意力模块具体公式如下:

$$Q_s, K_s, V_s = W_s(M_{out}) \quad (6)$$

$$Q_c, K_c, V_c = W_c(M_{out}) \quad (7)$$

$$PSMA_{out} = softmax(\theta \frac{Q_s K_s^T}{\sqrt{d}}) V_s \quad (8)$$

$$PCMA_{out} = softmax(\theta \frac{Q_c^T K_c}{\sqrt{d}}) V_c \quad (9)$$

$$GPMSA_{out} = [PSMA_{out}, PCMA_{out}] + M_{out} \quad (10)$$

$$G_{out} = GPMSA_{out \ l-1} + Conv(GPMSA_{out \ l}) \quad (11)$$

其中, 公式(6)、公式(7)中的 W_s 和 W_c 代表两个不同的线性映射操作。首先, 对映射的结果 K_s 和 K_c 乘以相同的可学习系数 θ , 用来对各自路径下空间自注意力和通道自注意力进行动态加权, 并对注意力效果进行调节。其次, 分别求出匹配后的空间注意力 $PSMA_{out}$ 和通道注意力 $PCMA_{out}$, $PSMA_{out}$ 和 $PCMA_{out}$ 计算公式如公式(8)、公式(9)所示。再次, 对这两个注意力结果进行拼接, 得到全局匹配注意力结果 $GPMSA_{out}$, 如公式(10)所示。最后, 使用一个卷积和残差模块对 $GPMSA_{out}$ 进行计算, 得到最终结果 G_{out} , 如公式(11)所示。

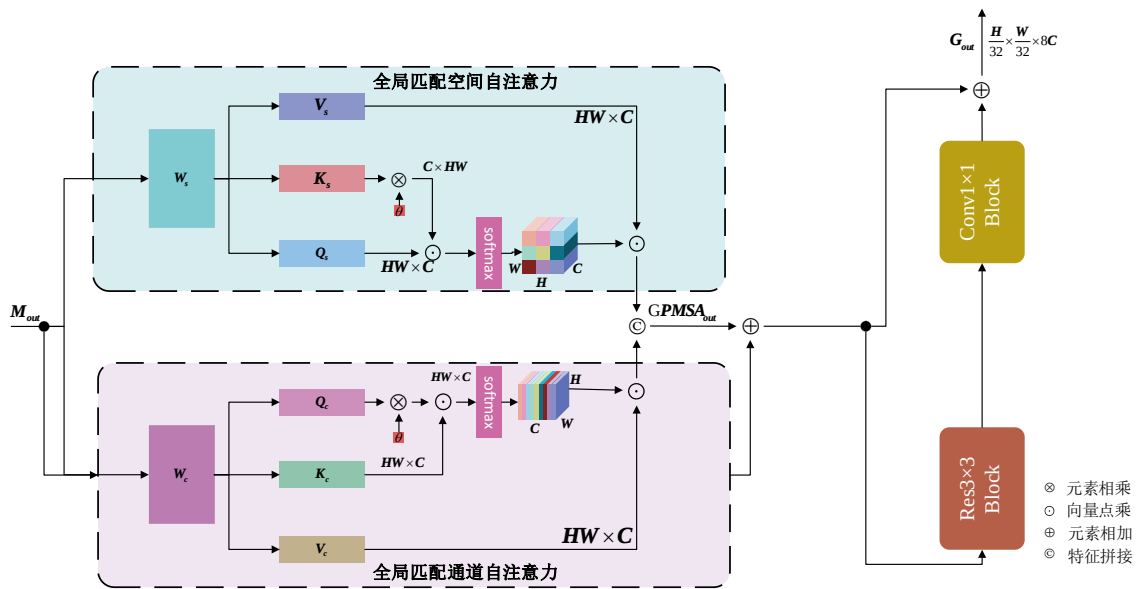


Fig. 3 Global paired self-attention mechanism block

图3 全局匹配自注意力机制模块

2 实验结果与分析

2.1 实验数据集与参数设置

本文实验使用了两个公开数据集:多图谱腹部标记挑战赛(MICCAI 2015)中的多器官分割数据集 Synapse 和公

共心脏 MRI 数据集 ACDC。Synapse 数据集收集了 30 例患者的 3 779 张轴向腹部临床 CT 图像, ACDC 数据集收集了 150 名患者的心脏临床 MRI 图像。实验环境为 Python3.7、Pytorch 1.12 以及 CUDA11.3。输入图像大小设置为 224×224, batch size 设置为 16。实验过程中, 学习率设置为 0.000 1, 使用 Adam 优化函数进行训练, 迭代次数为 150。

实验在 Nvidia Tesla P100 上进行, 显存容量为 16GByte。本文模型训练使用交叉熵损失 (Cross Entropy Loss, CE) 和 Dice 损失 (Dice Loss), 最终损失函数如公式 (12) 所示。

$$Loss = 0.5 \times CE_{Loss} + 0.5 \times Dice_{Loss} \tag{12}$$

本文使用 Mean Dice (DSC) 和 Mean Hausdorff Distance (HD95) 分别评价器官分割精度和边缘精度。其中, DSC 主要用于评估预测分割结果与真实分割结果之间的重叠程度, 数值越高表示整体分割效果越好; HD95 则关注分割边界的最大偏差, 数值越小表示边界预测更加精确。

2.2 实验结果

2.2.1 Synapse 数据集实验结果

在 Synapse 数据集上, 将 MFP Net 与 U-Net^[6]、Att-UNet^[13]、V-Net^[19]、DARR、TransUNet^[17]、Swin-UNet^[10]、UNE-

TR^[20]和 MT-UNet^[21]方法进行对比实验。对比结果如表 1 所示。可以看出, MFP Net 的 DSC 与 HD95 指标分别达到了 77.83% 和 18.28%。相较于 MT-UNet, MFP Net 的分割精度提高了 0.71%。因此, MFP Net 的性能优于 TransUNet、Swin UNet 等方法。同时, 对于贯穿所有轴向切面的细长解剖结构 (如主动脉) 和冠状面中占据较大区域的器官 (如肝脏), MFP Net 的分割精度分别达到了 88.42% 和 94.05%, 达到了很好的效果, 证明多分支融合全局匹配模块能够很好地学习相同解剖结构之间不同尺度下的相似表征。但是, 在分割部分存在重叠的器官时, 如胰脏、胆囊等, MFP Net 尚存在局限性。其原因可能是此类目标较小且分布密集, 同时自注意力机制缺少空间先验假设, 对邻近区域关注度弱, 导致 Transformer 分割困难。

Table 1 Comparison of segmentation accuracy of different networks on the Synapse multi-organ CT dataset

表 1 Synapse 多器官 CT 数据集上不同网络分割精度比较 (%)										
模型名称	Mean Dice	Mean HD95	主动脉	胆囊	左肾	右肾	肝脏	胰腺	脾脏	胃
UNet	74.84	39.70	87.03	67.85	77.77	68.6	90.43	49.78	84.67	72.59
Att-UNet	74.70	36.97	68.92	69.91	79.20	72.71	93.56	51.23	87.19	74.95
V-Net	68.81	—	75.34	51.87	77.10	80.75	87.84	40.05	80.56	56.98
DARR	69.77	—	74.74	53.77	72.31	73.24	94.08	54.18	89.9	45.96
UNETR	76.63	18.59	84.63	56.30	83.60	84.52	92.57	57.74	85.00	68.71
TransUNet	75.23	35.61	85.97	58.77	78.78	71.76	93.08	53.55	84.43	75.61
Swin UNet	76.56	21.55	84.75	60.32	81.35	76.64	91.57	53.31	87.42	77.14
MT-UNet	77.12	26.59	86.73	62.99	80.76	73.29	93.06	58.96	84.37	76.81
MFP Net (Ours)	77.83	18.28	88.42	64.37	82.04	78.26	94.05	57.68	82.23	75.62

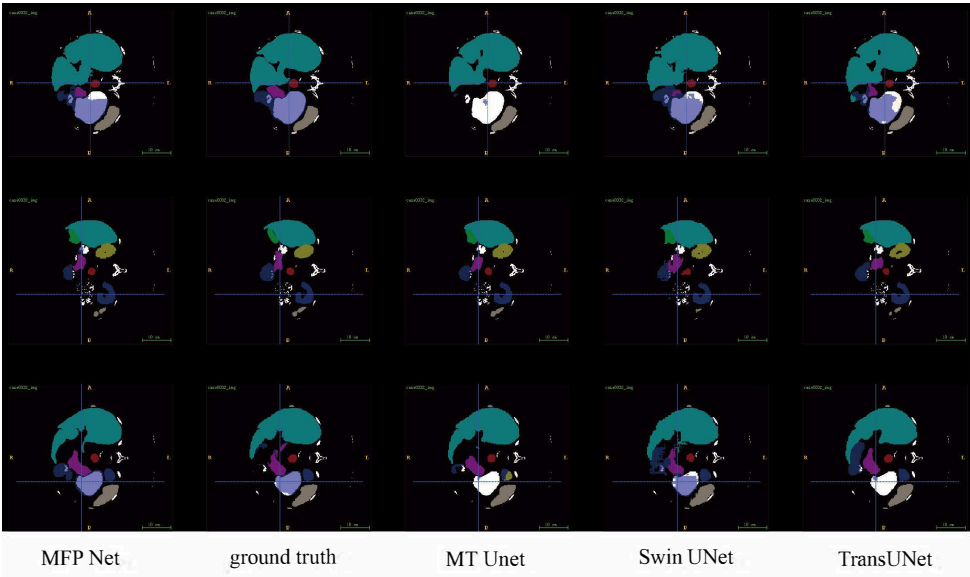


Fig. 4 Segmentation prediction comparison of different models on the Synapse dataset

图 4 不同模型在 Synapse 数据集上分割预测比较

图 4 显示了本文提出的医学影像分割网络 MFP Net 与其他方法在 Synapse 多器官数据集上的部分分割结果。从图 4 中可以看出, 在肾脏分割任务中, MFP Net 网络能够明显地将两者区分, 而 MT-UNet 完全无法区分。在肝脏分割任务中, 得益于本文提出的全局特征融合匹配注意力模块, 相比其他网络, MFP Net 能够更好地将肝脏内部细节与周围组织作出区分。在主动脉分割任务中, MFP Net 分割

结果相较于其他模型也是最优的。但是, 在胰腺以及胆囊分割任务中, 由于分割目标与肝脏、左肾等其它目标距离较近且有重叠, MFP Net 分割结果稍显逊色。

2.2.2 ACDC 数据集实验结果

在 ACDC 数据集上, 将 MFP Net 与 U-Net^[6]、Att-UNet^[13]、TransUNet^[17]、Swin-UNet^[10]和 MT-UNet^[21]方法进行对比实验。由于所有模型在 Mean HD95 指标上差异太小,

对比结果没有参考价值,因此仅选择 Mean Dice 指标作为参考。在 ACDC 数据集上的对比结果如表 2 所示。MFP Net 的 DSC 指标达到了 88.96%,相较于 MT-UNet, MFP Net 的分割精度提高了 0.53%, MFP Net 性能优于 TransUNet、Swin UNet 等方法。在右心室、心肌层分割中, MFP Net 的分割精度分别达到了 86.69% 和 88.19%, 优于所有对比模型。其证实了多分支融合全局匹配模块能够增强模型对于同一解剖结构在不同尺度下相似表征的学习能力。但是,对于左心室这种相对较小的目标,由于缺少空间先验假设以及缺少对邻近区域的关注,该方法仍有进一步提高的空间。

图 5 显示了本文提出的医学影像分割网络 MFP Net 与其他网络在 ACDC 心脏 MRI 数据集上的部分分割结果。

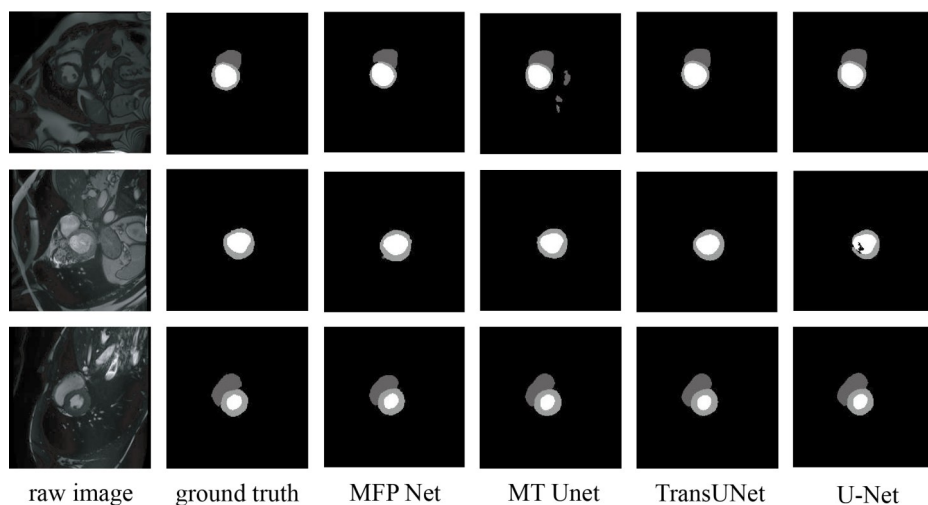


Fig. 5 Segmentation prediction comparison of different models on the ACDC dataset

图5 不同模型在ACDC数据集上分割预测比较

2.3 消融实验

为了验证多分支融合全局匹配注意力模块和注意力门控模块的重要性,本文通过在基础模型中逐步添加各个模块进行消融实验,并对实验结果进行分析。最终的实验结果如表 3 所示。从表 3 可以看出,得益于多分支融合全局匹配注意力模块的加入,提升了网络对不同解剖结构的分割能力,从而加强了网络对不同解剖结构之间存在相似局部特征的理解能力。同时,由于匹配了网络在不同编码器阶段的特征,使网络增强了对不同尺度结构的表征能力。此外,注意力门控单元的加入使得模型能够抑制冗余特征,增强了边界信息,从而改善了融合特征被稀释的问题。

综上所述,在网络结构中引入全局分支融合匹配注意力模块和注意力门控,使 MFP Net 能够在腹部多器官分割问题中更好地学习不同解剖结构之间存在的细微差别,并且加强了网络对不同尺度结构的表征能力,解决了医学图像在不同解剖结构之间存在相似表征的问题。同时,注意力门控的加入使得网络能够及时抑制融合后的冗余特征并增强边界信息,解决了网络因特征融合导致特征被稀释

Table 2 Segmentation accuracy comparison of different networks on the ACDC cardiac MRI dataset

表2 ACDC心脏MRI数据集上不同网络分割精度比较 (%)

模型名称	Mean Dice	右心室	心肌层	左心室
UNet	87.6	84.62	84.52	93.68
Att-UNet	86.9	83.27	84.33	93.53
TransUNet	88.58	85.45	86.12	94.18
Swin-UNet	88.07	85.77	84.42	94.03
MT-UNet	89.43	85.39	87.28	95.62
MFP Net(Ours)	89.96	86.69	88.19	94.98

其中,灰色表示右心室,灰白色表示心肌层,白色表示左心室。从图 5 中第一行分割结果可以看出,相较于 MT-UNet, MFP Net 对左心室的分割效果明显提升;在第二行右心室分割中, MFP Net 相较于 U-Net, 分割结果更加准确。

Table 3 Ablation experiment data of each module in MFP Net

表3 MFP Net各模块消融实验数据 (%)

模型结构	分支融合全局匹配注意力模块	注意力门控	Mean Dice	Mean HD95
基础模型	×	×	74.34	30.94
模型框架+多分支融合全局匹配注意力模块	√	×	76.36	22.58
模型框架+注意力门控	×	√	75.58	25.03
MFP Net	√	√	77.83	18.28

的问题。

3 结语

为了使模型能够学习到相同解剖结构在不同尺度空间中的相似表征,本文设计了一个多分支融合模块。该模块可以将编码器不同下采样阶段的特征进行融合,获取不同阶段同一解剖结构不同的相似表征。然后,设计了一个全局匹配注意力模块,将相似表征进行空间和通道对齐。最后,使用注意力门控模块对冗余特征进行抑制,以防止有用的特征被稀释。基于此,本文模型能够更好地对不同

器官的形状及大小进行区分。实验结果表明,该模型相比之前的方法具有更好的效果。但是,本文模型是基于2.5D方法进行分割的,与3D医学图像分割网络相比,当前的模型无法从空间维度上进行器官特征提取。未来,随着计算机硬件性能的进一步提高,可以更好地利用3D方法对医学图像进行分割,从而进一步优化模型性能。

参考文献:

- [1] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[DB/OL]. <https://arxiv.org/abs/1409.1556>.
- [2] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.
- [4] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4):640-651.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [7] WANG B, YUAN F Q, CHEN Z R, et al. Multi-stage U-Net automatic segmentation of thyroid ultrasound image[J]. Computer Engineering and Applications, 2023, 59(5): 205-212.
王波,袁凤强,陈宗仁,等. 多阶U-Net甲状腺超声图像自动分割方法[J]. 计算机工程与应用, 2023, 59(5): 205-212.
- [8] ZHOU Z, RAHMAN S M M, TAJBAKSH N, et al. UNet++: a nested U-Net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018: 3-11.
- [9] ALOM M Z, HASAN M, YAKOPCIC C, et al. Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation[DB/OL]. <https://arxiv.org/abs/1802.06955>.
- [10] CAO H, WANG Y Y, CHEN J, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation[DB/OL]. <https://arxiv.org/abs/2105.05537>.
- [11] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European Conference on Computer Vision, 2018: 833-851.
- [12] 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.
- [13] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: learning where to look for the pancreas[DB/OL]. <https://arxiv.org/abs/1804.03999>.
- [14] 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, 2017:6000-6010.
- [15] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[DB/OL]. <https://arxiv.org/abs/2010.11929>.
- [16] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 9992-10002.
- [17] CHEN J N, LU Y Y, YU Q H, et al. TransUNet: transformers make strong encoders for medical image segmentation[DB/OL]. <https://arxiv.org/abs/2102.04306>.
- [18] WANG Y, DENG Z J, HU X W, et al. Deep attentional features for prostate segmentation in ultrasound[C]//Medical Image Computing and Computer Assisted Intervention, 2018: 523-530.
- [19] 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, 2016:565-571.
- [20] HATAMIZADEH A, TANG Y, NATH V, et al. UNETR: transformers for 3d medical image segmentation[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision, 2022:1748-1758.
- [21] WANG H Y, XIE S A, LIN L F, et al. Mixed transformer U-Net for medical image segmentation[DB/OL]. <https://arxiv.org/abs/2111.04734>.

(责任编辑:黄健 何佳慧)