

中图法分类号: TP391; TP18 文献标识码: A 文章编号: 1006-8961(2026)01-0320-15

论文引用格式: Peng Y F and Sun W Q. 2026. Medical image segmentation via dual-path encoding and adaptive receptive fields. Journal of Image and Graphics, 31(1):0320-0334(彭晏飞, 孙伟强. 2026. 双路径编码与自适应感受野驱动的医学图像分割. 中国图象图形学报, 31(1):0320-0334)[DOI:10.11834/jig.250128]

双路径编码与自适应感受野驱动的医学图像分割

彭晏飞, 孙伟强*

辽宁工程技术大学电子与信息工程学院, 葫芦岛 125105

摘要: **目的** 受限于局部感受野, 卷积神经网络难以有效建模长程依赖。现有研究尝试将 Transformer 模块引入编码器、解码器或跳跃连接以增强全局信息建模能力, 但此类局部式嵌入仍不足以捕获器官在尺度与形态高度可变情况下所呈现的复杂依赖关系。此外, 传统卷积在训练后趋于静态, 难以适应器官的几何形变, 从而在一定程度上限制了模型对动态形变结构的表征能力。**方法** 针对上述问题, 提出一种端到端的医学图像分割框架, 通过双路径编码与自适应感受野机制的协同设计, 增强模型对全局—局部特征融合能力。具体而言, 首先, 设计了双路径编码结构, 在多个网络层级融合卷积神经网络与 Transformer 特征, 实现局部细节与全局上下文的渐进式融合; 其次, 构建编码器多层次融合机制, 通过跨尺度信息交互整合浅层纹理与深层语义特征, 增强模型对目标结构的多分辨率解析能力; 最后, 提出自适应感受野机制, 基于像素级语义差距动态调整卷积核感知范围, 突破静态卷积在形变组织表征中的瓶颈。**结果** 实验在两个公开数据集上与最新的方法进行比较, 在 Synapse 数据集中, 本文方法较次优模型在 DSC (Dice similarity coefficient) 和 HD95 (95% Hausdorff distance) 评价指标上分别提升 0.54% 和 0.44; 在 ACDC (automated cardiac diagnosis challenge) 数据集上的 DSC 值提高 0.34%; 消融实验进一步验证了双路径编码与自适应感受野机制的协同有效性。**结论** 本文方法通过深度融合卷积神经网络局部感知与 Transformer 全局建模的各自优势, 结合自适应感受野机制, 有效解决了当前医学图像分割模型中全局—局部特征融合不足及卷积核参数静态固化的问题, 实现了 SOTA (state-of-the-art) 级别的分割精度, 为复杂医学图像分割任务提供了新的方案。代码已开源: <https://github.com/Swq308/DPAR-Net>。

关键词: 卷积神经网络 (CNN); Transformer; 双路径编码; 自适应感受野; 多层次融合

Medical image segmentation via dual-path encoding and adaptive receptive fields

Peng Yanfei, Sun Weiqiang*

School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China

Abstract: Objective With the rapid advancement of medical imaging technologies, medical image segmentation has become a critical and extensively researched topic in the field of medical image processing. By enabling precise pixel-level classification and accurate identification of regions of interest—such as organs, structural tissues, and pathological areas—this technique holds substantial value in medical research and clinical practice. Medical image segmentation

收稿日期: 2025-04-23; 修回日期: 2025-05-18; 预印本日期: 2025-05-25

* 通信作者: 孙伟强 lnsqw17@126.com

基金项目: 辽宁省教育厅高等学校基本科研项目 (LJKZ0358, LZKQZ2021152)

Supported by: Basic Research Projects of Higher Education Institutions by the Department of Education of Liaoning Province (LJKZ0358, LZKQZ2021152)

plays a pivotal role in enhancing diagnostic accuracy, supporting personalized treatment planning, and ultimately improving patient outcomes. In recent years, the U-Net architecture, based on convolutional neural network (CNN), has emerged as a foundational model for medical image segmentation. However, due to the intrinsic limitations of CNN's local receptive fields, they struggle to effectively capture long-range dependencies. This shortcoming hinders the accurate segmentation of anatomical structures with diverse morphologies. Although the Transformer architecture was designed to address the limitations of CNN in modeling long-range dependencies, it often struggles to capture fine-grained local details effectively. Consequently, hybrid networks that integrate CNN and Transformers within the U-Net framework have garnered increasing attention. Existing approaches typically integrate Transformer modules into specific components of the convolutional architecture—such as the encoder, decoder, or skip connections—to address this issue. However, these fragmented integration strategies can only establish limited local-global associations at discrete hierarchical levels, making it difficult to model the complex dependencies introduced by anatomical heterogeneity in medical images, ultimately compromising segmentation accuracy. Furthermore, traditional CNN used fixed convolutional kernel parameters once training is complete, rendering them incapable of adapting to geometric deformations of anatomical structures. This limitation notably constrains the capability of the model to represent dynamically deformable targets in medical imaging. **Method** Aiming to address the aforementioned challenges, this study proposes an end-to-end medical image segmentation framework, DPAR-Net, which enhances global-local feature integration through a synergistic dual-path encoding design and an adaptive receptive field mechanism. Specifically, a dual-path encoding architecture is constructed, wherein the CNN path captures fine-grained local textures, while the Transformer path extracts high-level global semantics. A correlation-based fusion mechanism is employed at corresponding hierarchical levels to dynamically adjust feature weights based on local-global semantic associations, replacing traditional fusion paradigms. This strategy maximizes the complementarity between CNN and Transformer representations, enabling a progressive fusion of local detail and global context. Additionally, a multilevel encoder fusion strategy is introduced to integrate shallow texture features with deep semantic features through cross-scale information interaction, thereby enhancing the multi-resolution analytical capability of the model for target structures. Finally, an adaptive receptive field mechanism is proposed to dynamically adjust the perceptive range of the convolutional kernel based on pixel-level semantic discrepancies. This approach overcomes the limitations of static convolutions in modeling deformable anatomical structures. **Result** Experiments conducted on two publicly available datasets—Synapse (abdominal CT) and ACDC (cardiac MRI)—demonstrate that the proposed method achieves a Dice score of 84.57% on the Synapse dataset, along with a substantial reduction in the HD95 metric to 12.23, outperforming existing state-of-the-art approaches. On the ACDC dataset, the method surpasses the classical U-Net by 1.19% and the latest CNN-Transformer hybrid models by 1.2% in average Dice score. Ablation studies further validate the synergistic effectiveness of the dual-path encoding and adaptive receptive field mechanisms. Comparative visualizations also highlight the method's distinctive advantage in segmenting small targets, effectively addressing the issue of feature omission prevalent in existing techniques. **Conclusion** The proposed method effectively overcomes the limitations of inadequate global-local feature integration and the static convolutional kernel parameters in current medical image segmentation models. By deeply integrating the complementary strengths of CNNs for local feature perception and Transformers for global context modeling, augmented with an adaptive receptive field mechanism, this approach notably enhances segmentation accuracy and presents a novel solution for processing complex medical images. The code is available at: <https://github.com/Swq308/DPAR-Net>.

Key words: convolutional neural network (CNN); Transformer; dual-path encoding; adaptive receptive field; multi-level fusion

0 引言

医学图像分割作为医学影像分析领域的核心技术,通过对像素进行精准分类,实现对图像中感兴趣区域(如器官、组织结构及病变区域)的精准识别,在

提升临床诊断准确性、制定个体化治疗方案及改善患者预后方面具有重要作用(廖苗等,2025)。

随着深度学习技术的飞速发展,医学图像分割领域正经历着划时代的范式变革(李赫等,2024)。卷积神经网络(convolutional neural network, CNN)凭借其卓越的层次化特征提取能力,已成为该领域的

基准架构范式。基于U-Net(Ronneberger等, 2015)框架衍生出的诸多改进模型,如U-Net++(Zhou等, 2018)、UNet3+(Huang等, 2020)和Attention U-Net(Oktay等, 2018)、Dense U-Net(Zhou等, 2022)、Eres-UNet++(Li等, 2023)等,已成为该领域的基准模型架构,实现了像素级精度的分割输出。然而,由于CNN倾向于采用小尺寸卷积核进行局部特征编码,导致其在长程语义关联建模方面存在局限(廖苗等, 2025);此外,训练后的卷积核参数静态固定,难以适应医学影像中组织器官的形态异质性,进而制约了模型对感兴趣区域动态变化的表征能力。尽管学界相继提出膨胀卷积、空间金字塔池化等策略,但这些方法在捕获长程依赖关系方面仍存在不足,最终限制了基于卷积神经网络的模型的进一步发展。

随着Transformer架构在自然语言处理领域的成功,其在视觉任务中的迁移应用迅速成为研究焦点(蒋婷和李晓宁, 2024)。视觉Transformer(vision Transformer, ViT)(Dosovitskiy等, 2021)作为首个基于Transformer的视觉特征编码范式,通过图像块嵌入与全局自注意力机制开创了纯序列建模在图像分类中的应用先河。这一突破也推动了包括Swin Transformer(Liu等, 2021)、PVT(pyramid vision Transformer)(Wang等, 2021)、MaxViT(multi-axis ViT)(Tu等, 2022)在内的层次化视觉Transformer架构的发展。这些先进的主干网络被引入医学图像分割领域,催生了如Swin-UNet(Cao等, 2023)、TransDeepLab(Azad等, 2022)、MISSFormer(Huang等, 2023)、TransCeption(Azad等, 2023b)和DAE-Former(Azad等, 2023a)等分割模型,显著提升了分割精度。然而,尽管Transformer在全局语义信息整合方面具有显著优势,但其对局部空间上下文的理解能力相对薄弱,细粒度特征表示能力不足。而局部特征的高精度捕获对异质性器官、组织及病变区域的精确分割至关重要。这一局限性进一步阻碍了Transformer架构在医学图像分割领域的进一步发展。

为充分发挥CNN与Transformer的各自优势,近期研究聚焦于通过三大技术路径实现CNN与Transformer的协同融合:1)串联架构;2)CNN编码-Transformer解码;3)并联架构。TransUNet(Chen等, 2021)作为串联架构的典型代表,通过将CNN的深层特征传递至Transformer模块构建级联式特征增强,显著提升了模型的长程依赖建模能力;ESFPNet

(Chang等, 2023)采用CNN编码器进行层次化渐进编码,并借助Transformer解码器构建全局语义表征,实现了CNN与Transformer的融合;FAT-Net(Wu等, 2022)在编码器中采用并行CNN与Transformer双路径分支,但仅对两条路径的末端特征进行简单融合,未能充分优化局部—全局协同。TransFuse(Zhang等, 2021)、HiFormer(Heidari等, 2023)和ParaTransCNN(Sun等, 2024)通过跨尺度特征金字塔融合及层次化对齐策略机制,实现对局部与全局信息的互补优化,从而显著提升了分割性能。现有研究尽管在突破传统卷积神经网络局部感受野局限方面取得了进展,但普遍忽视了Transformer与CNN提取特征过程中的内在关联。仅对两者特征进行简单集成与叠加,难以实现局部细节信息与全局语义特征的深度融合,进而限制了模型整体性能的提升。

综上所述,针对当前混合CNN-Transformer医学图像分割模型在全局与局部信息融合不充分的核心问题,本文提出一种融合双路径编码与自适应感受野协同优化机制的医学图像分割网络DPAR-Net(dual-path adaptive receptive field network),以实现更有效的多尺度特征建模与上下文信息整合。

本文主要贡献如下:1)构建一种双路径编码器架构,充分融合CNN与Transformer的优势,实现对不同尺度下局部空间上下文特征与全局语义信息的高效捕捉。结合相关性融合模块,以突破传统卷积与Transformer特征融合松散的局限,有效弥合二者在结构差异下的语义断层。此外,提出一种多层次融合机制,实现编码阶段多尺度特征的渐进式融合,以抑制信息损失并增强特征表达的鲁棒性。2)设计了一种自适应感受野机制,根据语义差异动态调整卷积核的感受范围,克服解码过程中静态卷积在捕捉物体几何变化时的局限性,从而提升模型对复杂结构的适应能力。3)在两个公开医学图像分割数据集上开展实验,结果表明,本文方法在主观视觉评估与客观指标评价上均优于当前主流方法,展现出卓越的分割性能与优良的泛化能力。

1 本文方法

1.1 整体架构

DPAR-Net的整体架构如图1所示,区别于传统单路径医学图像分割范式,DPAR-Net引入双路径编

码结构,通过并行化多感受野处理机制协同提取细粒度局部细节特征与全局上下文依赖关系。DPAR-Net的整体结构遵循类似于U-Net的编码器—解码器结构,核心组件包括:基于Transformer与CNN构建的双路径编码器、相关性融合(relevance fusion, RF)机制、多层次融合(multi-level fusion, MLF)机制、自适应感受野(adaptive receptive field, ARF)机制、门控注意力(gated attention, GA)机制以及卷积上采样(conv upsampling, CU)模块。本文提出的相关性融合机制通过学习不同路径特征相关性加权矩阵,对Transformer生成的全局语义信息与CNN提取的局部细节特征进行特征重校准,进而增强编码器特征提取的敏感性。多层次融合机制通过聚合来自编码器的多层次特征,实现编码阶段多尺度特征的深度耦合。自适应感受野机制通过对每个像素的语义差异动态调整卷积感受野,实现解码阶段特征重构过程中的形变感知建模,增强模型在解码阶段的特征恢复能力。

1.2 双路径编码

为充分发挥CNN的局部特征表示优势与Transformer的全局上下文建模能力,本文提出双路径编

码器架构。其中,Transformer路径采用3阶段设计:第1阶段通过图像块嵌入(patch embedding)层将输入图像映射为分辨率为 $H/4 \times W/4 \times C$ 的特征图,随后通过两个Swin Transformer层进行全局长程依赖建模;第2阶段通过图像块合并(patch merging)层对特征图进行下采样,将分辨率降至 $H/8 \times W/8 \times 2C$ 的同时,再利用两个连续的Swin Transformer层进一步提取全局语义信息;第3阶段采用与第2阶段相同的处理原则,将特征图分辨率压缩至 $H/16 \times W/16 \times 4C$,最终构建多尺度特征金字塔。该架构通过4倍、8倍、16倍的渐进式下采样策略,在保持大范围感受野的同时实现多层次化特征提取。其数学表达为

$$\mathbf{F}_T^1 = TL_1(\text{Patch_Embedding}(x)) \in \mathbf{R}^{\frac{H}{4} \times \frac{W}{4} \times C} \quad (1)$$

$$\mathbf{F}_T^2 = TL_2(\text{Patch_Merging}(\mathbf{F}_T^1)) \in \mathbf{R}^{\frac{H}{8} \times \frac{W}{8} \times 2C} \quad (2)$$

$$\mathbf{F}_T^3 = TL_3(\text{Patch_Merging}(\mathbf{F}_T^2)) \in \mathbf{R}^{\frac{H}{16} \times \frac{W}{16} \times 4C} \quad (3)$$

式中, $x \in \mathbf{R}^{H \times W \times C}$ 表示输入图像, $\text{Patch_Embedding}(\cdot)$ 表示图像块嵌入层, $\text{Patch_Merging}(\cdot)$ 表示图像块合并层, $TL_i(\cdot)$ 表示第*i*阶段的Swin Transformer层, $\mathbf{F}_T^i$ 表示Swin Transformer分支第*i*阶段生成的特

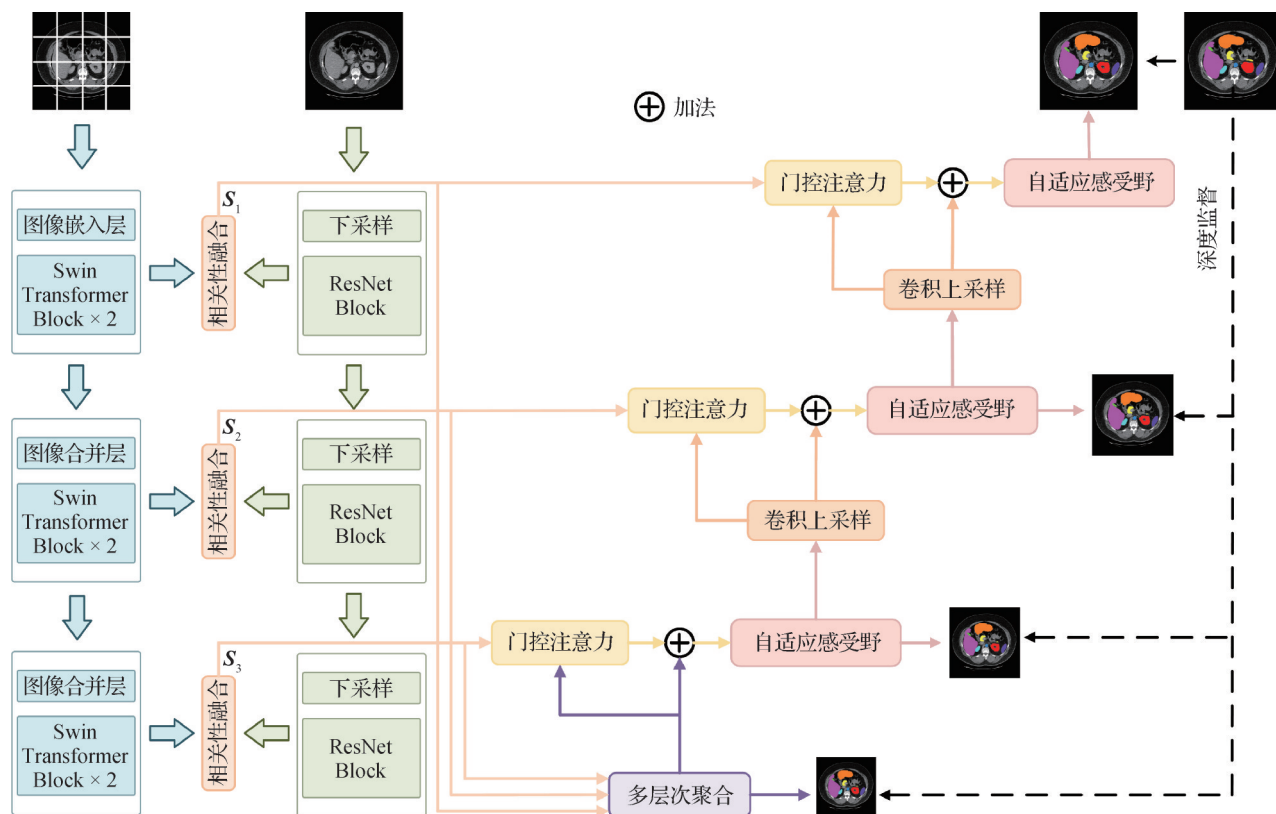


图1 DPAR-Net整体结构

Fig. 1 Overall architecture of DPAR-Net

征, $i \in \{1, 2, 3\}$ 。

在CNN路径的架构设计中,本研究采用预训练的ResNet50 (residual neural network 50) (He 等, 2016)作为骨干网络,旨在精准捕获图像局部纹理及细节信息。为实现CNN路径与Transformer路径特征表征的一致性,仅截取网络末端的3个阶段特征图,通过 1×1 卷积实现特征通道压缩。其数学表示为

$$CBR = \text{ReLU}(BN(f_{1 \times 1}(\cdot))) \quad (4)$$

$$F_c^1 = CBR(RL_1(x)) \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times C} \quad (5)$$

$$F_c^2 = CBR(RL_2(F_c^1)) \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 2C} \quad (6)$$

$$F_c^3 = CBR(RL_3(F_c^2)) \in \mathbb{R}^{\frac{H}{16} \times \frac{W}{16} \times 4C} \quad (7)$$

式中, $f_{k_1 \times k_2}(\cdot)$ 表示卷积核大小为 $k_1 \times k_2$ 的卷积操作, $BN(\cdot)$ 表示批量归一化操作, $\text{ReLU}(\cdot)$ 表示 ReLU (rectified linear unit) 激活函数, $RL_i(\cdot)$ 表示第 i 阶段的 CNN 层, F_c^i 表示 CNN 分支第 i 阶段生成的特征, $i \in \{1, 2, 3\}$ 。

1.3 相关性融合

现有研究虽然探讨了整合Transformer和CNN特征的方法,但未能充分挖掘二者协同工作的潜力。由于CNN与Transformer具备截然不同的特性和计算机制,传统的逐元素相加、相乘以及拼接操作难以

获得最优融合效果。本文提出一种高效融合策略,以最大化编码器的特征表示。如图2所示,首先,对Transformer路径的中间特征重塑形状,以实现与CNN路径特征的空间对齐;随后,执行拼接操作以整合信息,为降低计算开销,本文方法引入深度可分离卷积,先通过 1×1 深度可分离卷积压缩通道维度,再利用 3×3 卷积提取局部细节;接着,采用全局平均池化(global average pooling, GAP)以及 1×1 深度可分离卷积恢复通道数,并通过sigmoid函数计算相关性系数矩阵 M ,随后,将该矩阵与两个原始特征相乘以调整特征表示;最后,将二者相加生成最终分割输出。其具体实现过程为

$$F = \text{Concat}[\text{Reshape}(F_T^i | F_c^i), F_c^i] \quad (8)$$

式中, F 为CNN提取的局部细节特征与Transformer捕获的全局上下文特征经通道拼接后形成的融合表示, $\text{Concat}[\cdot]$ 表示拼接操作, $F_T^i | F_c^i$ 表示将大小为 F_T^i 的特征重塑为 F_c^i 的大小。

$$M = \text{sigmoid}(f_{1 \times 1}(\text{GAP}(f_{3 \times 3}(f_{1 \times 1}(F))))) \quad (9)$$

式中, M 表示相关性系数矩阵, $\text{GAP}(\cdot)$ 表示全局平均池化操作, $\text{sigmoid}(\cdot)$ 表示sigmoid函数。

$$Z = F_T^i \times M + F_c^i \times M \quad (10)$$

式中, Z 代表相关性融合模块的最终输出。

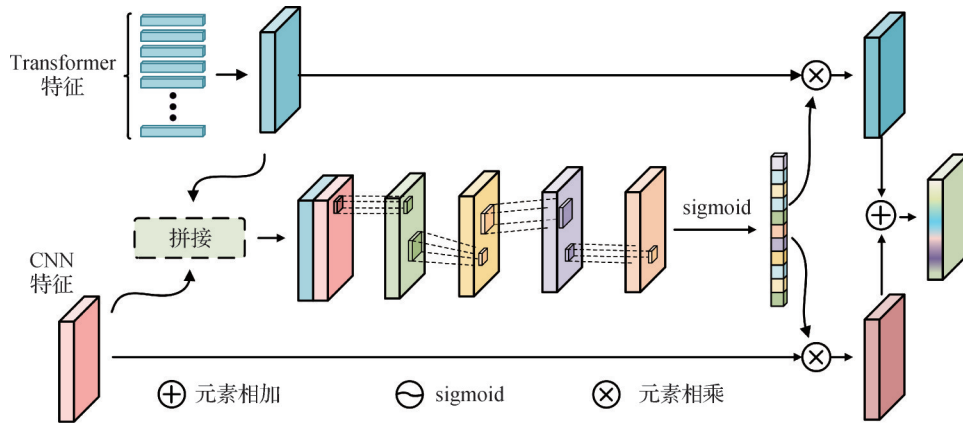


图2 相关性融合

Fig. 2 Relevance fusion

1.4 多层次聚合

本文设计了一种多层次特征聚合模块,以克服随着网络层次加深而引起的细粒度信息损失问题。如图3所示,给定来自编码器的输出特征 S_1 、 S_2 和 S_3 ,首先将 S_2 和 S_3 上采样至与 S_1 相同的尺寸。随后,将3组特征沿通道维度拼接,并通过 1×1 卷

积实现特征压缩与信息整合,形成全局特征图 G 。之后,采用全局平均池化(GAP)提取全局信息,以增强目标区域的完整性。同时,为抑制GAP操作引入的噪声,进一步引入全局最大池化(global max pooling, GMP),通过大小为 k_1 和 k_2 的池化操作捕获局部极值信息,以保留更多前景细节并减小背

景干扰。由此,得到3个池化特征图 G_1 、 G_2 和 G_3 。

随后,对这3幅特征图进行逐元素相加,并通过前馈网络层压缩和恢复通道数。接着利用 sigmoid 计算相关系数矩阵 M_2 ,并将该矩阵与初始特征图 G 相乘,再经过 3×3 卷积操作提取局部细节,得到最终输出。其数学表达为

$$G = \text{Concat}[up(S_3), up(S_2), S_1] \quad (11)$$

$$G_a = f_{1 \times 1}(G) \quad (12)$$

$$G_2 = \text{Sum}(\text{GAP}_{2 \times 2}(G_a) + \text{GMP}_{2 \times 2}(G_a) + \text{GMP}_{4 \times 4}(G_a)) \quad (13)$$

$$M_2 = \text{sigmoid}(\text{MLP}(G_2)) \quad (14)$$

$$\text{out} = f_{3 \times 3}(M_2 \times G) \quad (15)$$

式中, $up(\cdot)$ 表示上采样操作, $\text{GAP}_{k_1 \times k_2}(\cdot)$ 表示池化核为 $k_1 \times k_2$ 的平均池化操作, $\text{GMP}_{k_1 \times k_2}(\cdot)$ 表示池化核为 $k_1 \times k_2$ 的最大池化操作, $\text{MLP}(\cdot)$ 表示前馈网络层, $\text{Sum}(\cdot)$ 表示求和。

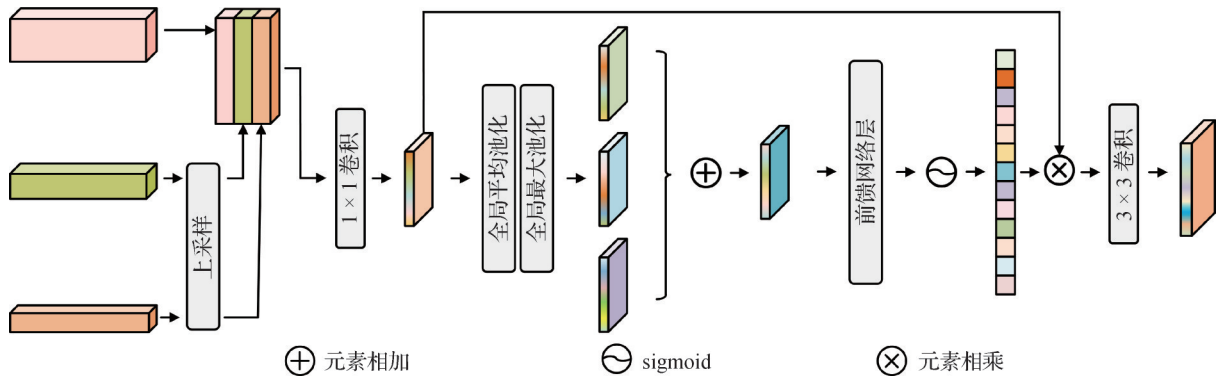


图3 多层次融合

Fig. 3 Multi-level fusion

1.5 自适应感受野

本文提出一种像素级自适应感受野策略,以克服静态卷积在捕捉物体几何形变方面的局限,提升医学图像分割的精度和鲁棒性。如图4所示,自适应感受野模块融合通道注意力块(channel attention block, CAB)、空间注意力块(spatial attention block, SAB)以及自适应感受野卷积,以动态调整感受野大小,从而精准捕获不同尺度的特征信息。对于输入特征,首先通过CAB和SAB进行特征增强,以突出关键通道信息并建模局部空间关系。随后引入自适应感受野机制,根据目标结构的尺度变化动态调整感受野范围,增强网络对不同尺度信息的自适应性。具体而言,自适应感受野卷积首先利用一组预定义的卷积核 $\{k_1, k_2, \dots, k_n\}$ 在不同感受野下提取特征,生成特征图集合 $\{F_1, F_2, \dots, F_n\}$, 其中 k_i 代表对应感受野大小。同时引入全局平均池化(GAP)与全局最大池化(GMP)对每个特征图 F_i 进行池化处理,以分别提取全局信息和高频信息并执行拼接操作。随后,这些池化后的结果经过共享卷积层及 sigmoid 激活函数,生成对应不同感受野的激活图。最终,所有感受野的激活图与池化后的结果逐元素相乘后累

加,并与初始特征进行残差融合,生成最终的输出。其具体计算流程为

$$F_{CS} = \text{SAB}(\text{CAB}(a_i)) \quad (16)$$

$$F_{MA_i} = \text{Concat}(\text{Mp}(\text{MCL}(F_{CS})), \text{Ap}(\text{MCL}(F_{CS}))) \quad (17)$$

$$y = \text{Sum}(\text{sigmoid}(\text{SCL}(F_{MA_i})) \times F_{MA_i}) + F_{CS} \quad (18)$$

式中, F_{CS} 表示空间注意力块的输出特征, $\text{SAB}(\cdot)$ 和 $\text{CAB}(\cdot)$ 分别表示空间注意力和通道注意力, $\text{MCL}(\cdot)$ 表示多尺度卷积层, $\text{Mp}(\cdot)$ 和 $\text{Ap}(\cdot)$ 分别表示最大池化和平均池化, $\text{SCL}(\cdot)$ 表示共享卷积层, F_{MA_i} 表示拼接后的特征, $\text{Sum}(\cdot)$ 表示求和, y 表示模块最终输出。

1.6 门控注意力和卷积上采样

本文提出一种门控注意力机制,利用网络自主学习注意力系数,有效强化相关特征并抑制无关特征,动态调控网络不同阶段的信息流动。如图5所示,首先,对分组卷积输出的特征进行批量归一化处理,以稳定训练过程中的特征分布;随后,通过元素级加法运算实现跨层级特征融合;接着,利用 1×1 卷积进行通道维度压缩,并采用 ReLU 非线性激活函数生成具有空间相关性的中间特征表示经批量归一化后获得单通道空间注意

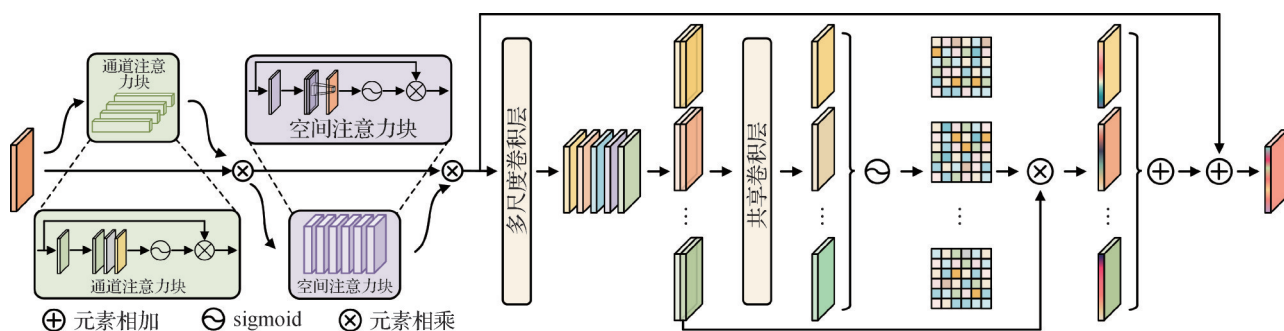


图4 自适应感受野

Fig. 4 Adaptive receptive field

力图;最后,通过 sigmoid 函数将空间注意力值域约束至 $[0,1]$ 区间,形成具有空间选择性的注意力系数矩阵。将该矩阵与初始特征相乘,得到最终结果。其数学表示为

$$A = BN(f_{3 \times 3}(g_i)) + BN(f_{3 \times 3}(S_i)) \quad (19)$$

$$B = \text{sigmoid}(\text{ReLU}(f_{1 \times 1}(A))) \times g_i \quad (20)$$

式中, g_i 表示当前阶段的全局特征, S_i 表示编码器当前阶段特征。

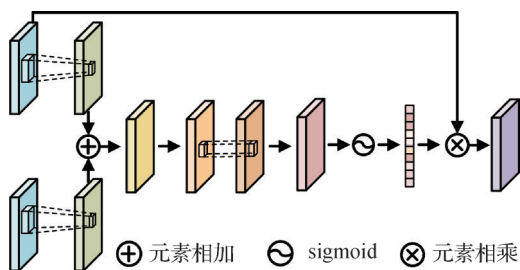


图5 门控注意力

Fig. 5 Gated attention

卷积上采样模块部署于自适应感受野之后,旨在提升特征图的空间解析度并缩减通道维度,其架构细节如图6所示。首先对输入特征执行双线性上采样运算,继而通过 3×3 的深度可分离卷积实施空间特征编码,随后进行批量归一化和非线性激活处理。通过跳跃连接与原始特征进行残差融合后,最终通过 1×1 逐点卷积完成通道重构。其数学表达式为

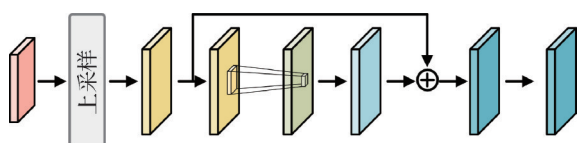


图6 卷积上采样

Fig. 6 Conv upsampling

$$\text{out} = f_{1 \times 1}(\text{ReLU}(\text{BN}(f_{3 \times 3}(\text{up}(x)))) + \text{up}(x)) \quad (21)$$

式中, out 表示模块的输出结果。

2 实验与分析

2.1 实验数据及评价指标

本文采用 Synapse 腹部多器官分割数据集和自动化心脏诊断挑战数据集 ACDC (automated cardiac diagnosis challenge) 对 DPAR-Net 的有效性进行验证。Synapse 数据集包含 30 例腹部 CT (computed tomography) 扫描,共计 3 779 幅轴向临床 CT 图像。为确保研究的严谨性与泛化能力,该数据集随机划分为两个子集,其中 18 例扫描图像用于训练,剩余 12 例用于测试。每个样本由 85~198 幅连续切片构成,并经标准化处理以统一像素尺寸。

ACDC 数据集包含 100 例核磁共振成像 (magnetic resonance imaging, MRI) 扫描样本,数据来源于多位患者,并提供左心室 (left ventricle, LV)、右心室 (right ventricle, RV) 及心肌 (myocardium, MYO) 3 个关键心脏结构的标注。在实验设计中,遵循标准的训练—测试划分策略,从数据集中随机选取 100 例扫描样本用于训练,40 例用于测试。

为客观评估算法的分割性能,实验采用 Dice 相似系数 (Dice similarity coefficient, DSC) 与 95% Hausdorff 距离 (Hausdorff distance at 95% quantile, HD95) 作为评价指标。其中, DSC 取值范围为 0~1, 数值越大,表示分割结果与真实标注的重叠度越高; HD95 值越小,则表明分割结果与真实标注之间的边界误差越低。

2.2 实验细节

为确保实验的公平性,本文所有实验均基于

PyTorch框架实现,并在配备11 GB显存的NVIDIA GeForce RTX 2080Ti GPU上进行训练。针对实验超参数,将输入图像尺寸设定为 224×224 像素,批处理大小设置为4,最大训练epoch设置为400。为充分扩充训练集,采用了包括随机旋转和水平翻转在内的数据增强技术。采用随机梯度下降(stochastic gradient descent, SGD)优化器,为达到最佳分割效果,将初始学习率设置为0.01,动量参数为0.9,权重衰减为 10^{-4} 。

2.3 损失函数

为提升分割性能,本文设计了一种融合Dice损失与二值交叉熵(binary cross entropy, BCE)损失的混合损失函数。其中,Dice损失通过度量预测结果与真实标签之间的空间重叠度,以有效引导模型进行参数优化,其数学定义为

$$L_{\text{Dice}} = \sum_{i=1}^N \left(1 - \frac{2|y_i \cap \hat{y}_i|}{|y_i| + |\hat{y}_i|} \right) \quad (22)$$

式中, N 表示图像中像素总数, y_i 为第 i 个像素的预测值, $\hat{y}_i$ 为第 i 个像素的真实标签值。

此外,本文进一步引入二值交叉熵(BCE)损失,以增强模型对前景与背景像素的判别能力。BCE损失函数的定义为

$$L_{\text{BCE}} = - \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (23)$$

为综合两种损失函数的优势,本文采用Dice与BCE损失的加权混合形式,并结合深度监督以提升分割性能,其数学定义为

$$\text{Loss} = \sum_{i=1}^4 (\alpha \cdot L_{\text{BCE}} + \beta \cdot L_{\text{Dice}}) \cdot p_i \quad (24)$$

式中, α 和 β 为权重系数, $\alpha = 0.5, \beta = 0.5$ 。 p_i 为各阶段分割预测图。

2.4 对比实验及分析

2.4.1 Synapse数据集

为验证DPAR-Net的有效性,将其与近年提出的SOTA(state-of-the-art)算法及一些经典医学图像分割方法进行对比,包括基于CNN的模型(U-Net、AttUNet、AHGNN(Chai等2024))、基于纯Transformer的架构(Swin-UNet、TransDeepLab、MISSFormer、DAEFormer)、混合CNN-Transformer的模型(TransUNet、HiFormer、TransCeption、AFTER-UNet(Yan等,2022)、CASTformer(You等,2022)、MSAANet(Zeng等,

2023)等)以及其他前沿分割方法。

如表1所示,本文提出的方法在多器官分割任务中展现出SOTA级别的分割性能。Dice相似系数达到84.57%的最优水平,95%Hausdorff距离指标显著降至12.23。在左肾及胃分割任务中,较当前最优方法,Dice相似系数分别实现0.69%及0.58%的性能增益。实验结果充分佐证了DPAR-Net在全局上下文建模与局部细节捕获上的协同能力,其构建的渐进式全局—局部融合机制使左肾与胃的识别精度均超越当下最先进模型。图7所示的可视化对比分析进一步表明,本文方法在微小目标分割任务中具备独特优势,能有效克服现有方法存在的特征漏检问题。

2.4.2 ACDC数据集

如表2所示,本研究在ACDC数据集开展的对比实验充分验证了DPAR-Net的优越性。相较于经典的卷积网络架构U-Net,本文方法在平均Dice系数指标上实现1.19%的性能增益;相较于前沿的Transformer架构Swin-UNet、MISSFormer和DAEFormer,平均Dice系数进一步提升1.2%、3.8%和3.35%,充分佐证了本文方法在全局建模与局部细节表征融合方面的优势;与CNN-Transformer融合模型HiFormer、ParaTransCNN和TransCASCADE相比,分别实现1.19%、1.2%和0.34%的性能增益。图8可视化结果表明,相较于现有方法,本文方法在目标区域分割中展现出SOTA级精度,可精确捕获心肌组织的细微形态特征,有力验证了其对复杂器官结构的鲁棒建模能力。

2.5 消融实验

2.5.1 参数消融

在Transformer架构中,网络层数对全局信息捕获至关重要。为实现计算效率与性能的平衡,本研究将Transformer各阶段的层数配置分别设置为[2, 2, 2]、[2, 2, 4]、[2, 4, 4]、[4, 4, 4]。表3所示的实验结果表明,当Transformer各阶段层数配置为[2, 2, 2]时,模型表现出最佳分割性能。这些实验结果共同验证了Transformer分支各阶段层数设置的合理性。

为深入探究自适应感受野机制中不同核尺寸的协同效应,本文在Synapse多器官数据集上展开了对比实验。如表4结果所示,模型效果随核规模扩展呈现先升后降趋势,当核尺寸组合扩展至[1, 3, 5, 7, 9]时,模型取得最优性能指标:平均Dice系数提升

表 1 在 Synapse 数据集上的实验结果比较
Table 1 Comparative experimental results on the Synapse dataset

方法	出版/年份	平均		DSC ↑ /%							
		DSC ↑ /%	HD95 ↓	主动脉	胆囊	左肾	右肾	肝脏	胰腺	脾脏	胃
U-Net	MICCAI/2015	76.85	39.70	89.07	69.72	77.77	68.60	93.43	53.98	86.67	75.58
Att-UNet	MIDL/2018	77.77	36.02	89.55	68.88	77.98	71.11	93.57	58.04	87.30	75.75
TransUNet	CVPR/2021	77.48	31.69	87.23	63.13	81.87	77.02	94.08	55.86	85.08	75.62
Swin-UNet	ECCV/2022	79.13	21.55	85.47	66.53	83.28	79.61	94.29	56.58	90.66	76.60
TransDeepLab	MICCAI/2022	80.16	21.25	86.04	69.16	84.08	79.88	93.53	61.19	89.00	78.40
HiFormer	WACV/2023	80.39	14.70	86.21	65.69	85.23	79.77	94.61	59.53	90.99	81.08
AFTer-UNet	WACV/2022	81.02	—	90.91	64.81	87.90	85.30	92.20	63.54	90.99	72.48
MISSFormer	TMI/2023	81.96	18.20	86.99	68.65	85.21	82.00	94.41	65.67	91.92	80.81
TransCeption	arXiv/2023	82.24	20.89	87.60	71.82	86.23	80.29	95.01	65.27	91.68	80.02
DAE-Former	PRIME/2023	82.43	17.46	88.96	72.30	86.08	80.88	94.98	65.12	91.94	79.19
TransCASCADE	WACV/2023	82.68	17.34	86.63	68.48	87.66	84.56	94.43	65.33	90.79	83.52
MSAANet	ICME/2023	82.85	18.54	89.40	<u>73.20</u>	84.31	78.53	95.10	68.85	91.60	81.78
PAG-TransYnet	arXiv/2024	83.43	15.82	<u>89.67</u>	68.74	86.74	<u>84.88</u>	95.87	68.75	92.01	80.66
MAXFormer	KBS/2023	83.67	15.89	87.73	73.53	87.93	84.68	95.00	66.55	<u>92.47</u>	81.44
ParaTransCNN	arXiv/2024	83.86	15.86	88.12	68.97	<u>87.99</u>	83.84	95.01	69.79	92.71	<u>84.43</u>
AHGNN	MICCAI/2024	<u>84.03</u>	13.26	—	—	—	—	—	—	—	—
MixFormer	TIM/2025	82.64	<u>12.67</u>	87.36	71.53	86.22	83.19	<u>95.23</u>	66.82	89.98	80.77
DPAR-Net	JIG/2026	84.57	12.23	89.04	72.91	88.68	84.45	94.99	<u>69.13</u>	92.35	85.01

注:加粗、下划线字体表示各列最优、次优结果。“↑”表示越高越好,“↓”表示越低越好。“—”表示信息缺失。

至 84.57%, HD95 距离显著降低至 12.23。引入 11 × 11 大核后,出现性能拐点,平均 Dice 系数降低至 84.38%,佐证了过度扩张感受野引发的负优化效应。基于上述分析,本研究最终确定[1,3,5,7,9]的核尺寸组合作为最优配置方案。

2.5.2 结构消融

为验证新提出模块的有效性,在 Synapse 数据集上进行系统性评估,实验结果如表 5 和表 6 所示。

1)双阶段编码验证。在引入双路径编码策略后,模型的 DSC 和 HD95 相较于单 Swin Transformer 分支提升 12.22%(从 71.43% 提高到 83.65%)和 15.06(从 27.69 降低到 12.63);相较于单 ResNet 分支,DSC 和 HD95 分别提升 3.30%(从 80.35% 提高到 83.65%)和 4.10(从 16.73 降低到 12.63)。这些结果表明,双路径编码策略能够有效增强模型的特征表达能力,提升分割精度。

2)相关性加权验证。在引入相关性加权融合后,DSC 较双路径策略提升 0.92%(从 83.65% 提高

到 84.57%),而 HD95 则小幅提升 0.40(从 12.63 降低到 12.23)。这表明,相关性加权融合能够充分整合全局与局部信息,不仅使模型更关注器官的整体结构,还显著强化了边界细节的表达能力。

3)多层次融合验证。通过移除多层次融合组件,模型的平均 DSC 和 HD95 相较于 DPAR-Net 分别下降 1.03% 和 0.9。这一结果验证了多层次融合在多尺度信息融合方面的有效性,表明其能有效解决细节丢失的问题,从而提升模型的分割精度。

4)自适应感受野验证。移除自适应感受野机制更换为两层 3 × 3 卷积后,模型的平均 DSC 和 HD95 相较于 DPAR-Net 分别下降 1.44% 和 2.14。这一变化充分验证了自适应感受野机制的作用,表明其能够有效适应医学图像中不同组织和器官的异质性,大幅增强模型对前景结构的敏感性,从而提升分割精度。

5)门控注意力验证。移除门控注意力并替换为传统的跳跃连接和相加操作后,模型的平均 DSC 相

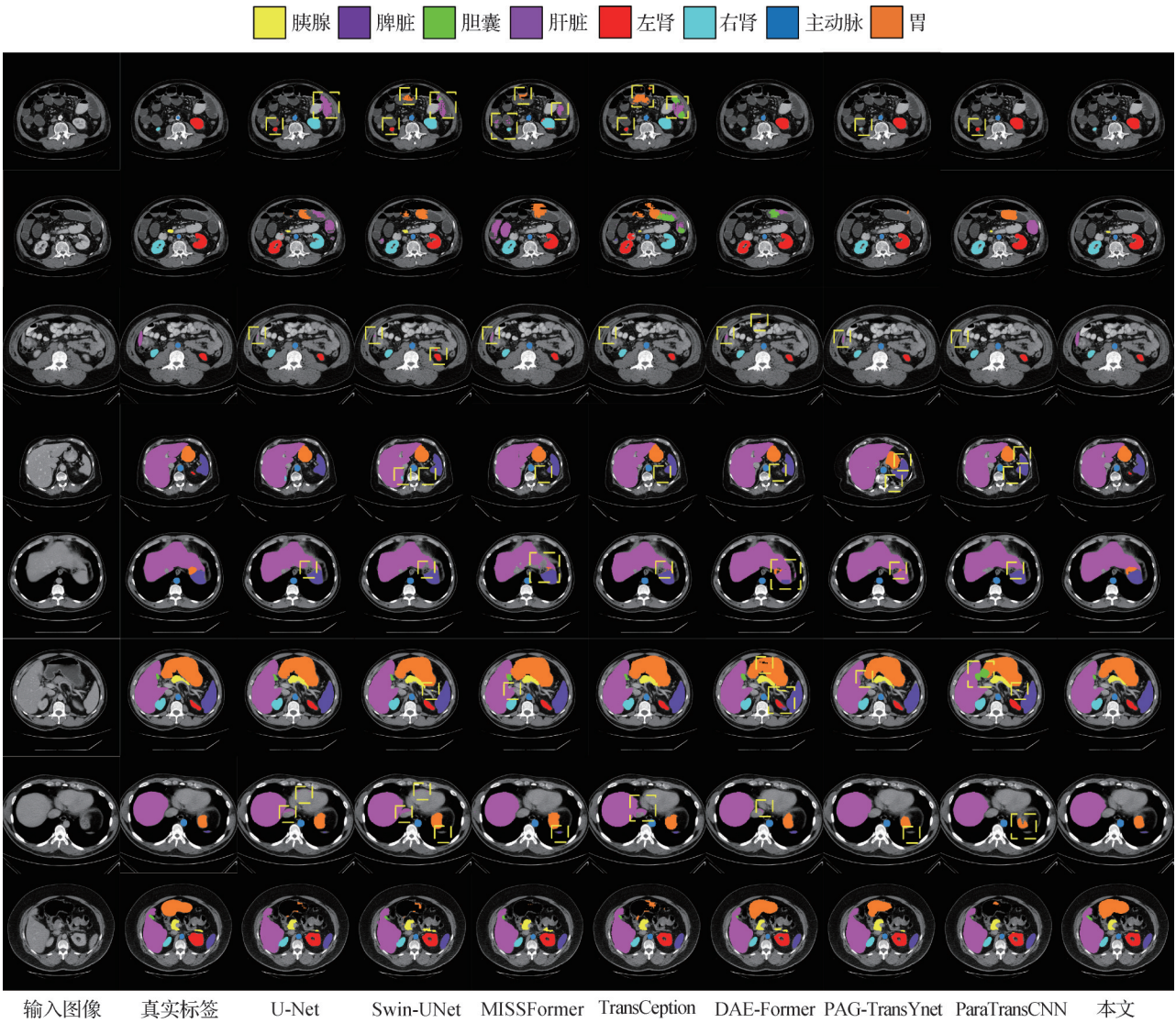


图 7 Synapse 可视化结果

Fig. 7 Synapse visualization results

表 2 在 ACDC 数据集上的实验结果比较
Table 2 Comparative experimental results on the ACDC dataset

方法	平均 DSC ↑	DSC ↑		
		右心室	心肌	左心室
U-Net	90.34	89.43	85.89	95.71
Swin-UNet	90.33	89.87	85.98	95.14
MISSFormer	87.73	87.43	81.82	93.89
DAE-Former	88.18	87.87	82.46	94.21
HiFormer	90.34	89.23	87.36	94.43
ParaTransCNN	90.33	89.44	85.84	95.71
TransCASCADE	91.19	89.41	87.48	96.67
DPAR-Net	91.53	89.67	88.26	96.66

注:加粗字体表示各列最优结果。

较于 DPAR-Net 下降 0. 63%,HD95 则下降 0. 53。这表明,门控注意力机制可有效强化相关特征并抑制无关特征,动态调控网络不同阶段的信息流动。

6)卷积上采样验证。在将卷积上采样替换为普通上采样操作后,模型的平均 DSC 相较于 DPAR-Net 下降 0. 25%,HD95 则小幅提升 0. 06。表明卷积上采样模块可有效弥补普通上采样造成的细节损失问题。

2. 6 复杂度分析

为全面评估各模型性能,本研究从参数量、计算复杂度及分割精度 3 方面开展对比实验。如表 7 所示,传统 CNN 架构 U-Net 的参数量与浮点数分别为 31. 04 M 和 41. 93 G,但其 DSC 指标与边界识别精度显著落后;纯 Transformer 模型(Swin-UNet、MISS-

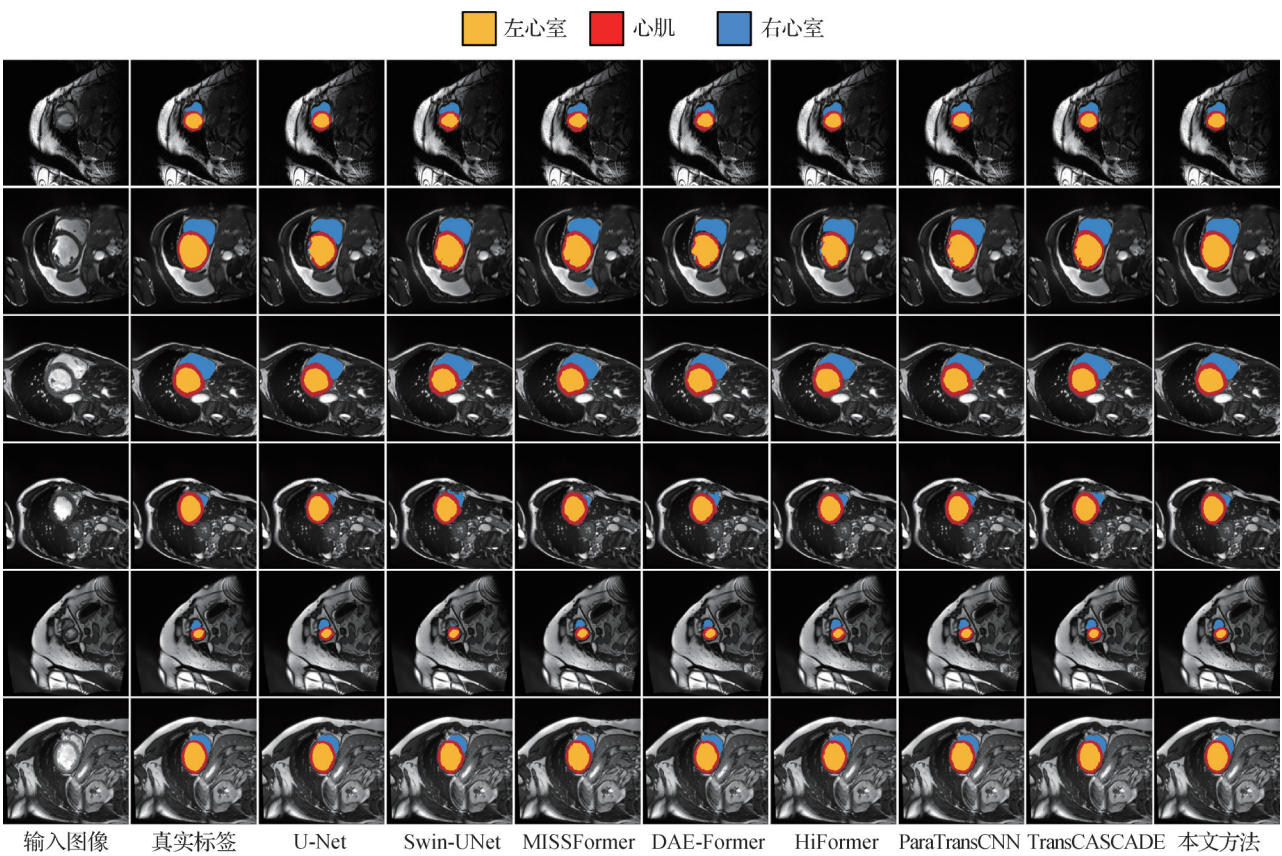


图8 ACDC可视化结果

Fig. 8 ACDC visualization results

表3 在Synapse数据集上Swin Transformer分支层数配置消融实验结果

Table 3 Ablation study results of Swin Transformer branch layer configuration on the Synapse dataset

层数配置	平均 DSC ↑ /%	HD95 ↓	DSC ↑ /%							
			主动脉	胆囊	左肾	右肾	肝脏	胰腺	脾脏	胃
[4,4,4]	83.64	15.18	88.61	73.52	86.35	81.57	94.86	67.34	92.25	84.66
[2,4,4]	83.18	14.23	89.28	67.18	87.85	83.77	95.34	69.10	89.68	83.24
[2,2,4]	83.70	11.21	88.56	70.06	88.52	83.81	95.17	67.99	91.46	84.07
[2,2,2]	84.57	12.23	89.04	72.91	88.68	84.45	94.99	69.13	92.35	85.01

注:加粗字体表示各列最优结果。“↑”表示越高越好,“↓”表示越低越好。

表4 在Synapse数据集上自适应感受野卷积核配置消融实验结果

Table 4 Ablation study results of adaptive receptive field convolution kernel configuration on the Synapse dataset

层数配置	平均 DSC ↑ /%	HD95 ↓	DSC ↑ /%							
			主动脉	胆囊	左肾	右肾	肝脏	胰腺	脾脏	胃
[1,3,5]	82.72	17.17	88.37	67.46	86.87	81.63	95.57	67.48	88.78	85.57
[1,3,5,7]	83.03	13.76	89.29	66.94	87.55	83.85	95.30	68.39	89.03	83.97
[1,3,5,7,9]	84.57	12.23	89.07	72.91	88.68	84.45	94.99	69.13	92.35	85.01
[1,3,5,7,9,11]	84.38	13.99	89.27	70.45	87.96	84.65	95.77	69.65	91.70	85.62

注:加粗字体表示各列最优结果。“↑”表示越高越好,“↓”表示越低越好。

表 5 Synapse数据集上双路径编码消融实验结果

Table 5 Ablation study results of dual-path encoding on Synapse dataset

方法	模块			DSC ↑ /%	HD95 ↓
	Swin	ResNet	相关性加权		
Baseline	√	–	–	71.43	27.69
ResNet	–	√	–	80.35	16.73
Baseline+ResNet	√	√	–	83.65	12.63
DPAR-Net	√	√	√	84.57	12.23

注:加粗字体表示各列最优结果。“√”和“–”表示使用和未使用对应模块。

表 6 在Synapse数据集各模块组合消融实验结果

Table 6 Ablation experiment results of different module combinations on the Synapse dataset

方法	平均DSC ↑ /%	HD95 ↓	DSC ↑ /%							
			主动脉	胆囊	左肾	右肾	肝脏	胰腺	脾脏	胃
w/o MLF	83.54	13.13	89.07	68.33	87.91	83.38	95.08	69.34	90.38	84.88
w/o ARF	83.13	14.37	88.74	70.16	86.15	81.29	94.99	69.03	91.35	83.34
w/o GA	83.94	12.76	88.86	67.04	88.74	85.05	95.35	69.51	91.44	85.58
w/o CU	84.32	12.17	89.23	71.27	88.24	84.15	95.17	70.16	91.26	85.12
DPAR-Net	84.57	12.23	89.04	72.91	88.68	84.45	94.99	69.13	92.35	85.01

注:加粗字体表示各列最优结果。“↑”表示越高越好,“↓”表示越低越好。w/o表示去除该组件。

Former、DAE-Former)在参数量(21.75 M、35.45 M、29.69 M)与浮点数(5.93 G、7.28 G、26.16 G)方面相对占优,但受限于局部建模能力,其分割性能仍有提升空间;融合 CNN 与 Transformer 的混合模型(TransUNet、HiFormer、ParaTransCNN、TransCASCADÉ(Rahman 和 Marculescu, 2023)、PAG-TransYnet(Bougourzi 等, 2024)、MAXFormer(Liang 等, 2023)、Mixformer(Liu 等, 2025))的计算负载较高,分别达到 93.23 M、34.14 M、42.70 M、115.01 M、116.77 M、84.48 M 和 123.35 M 的参数量以及 24.73 G、17.81 G、50.39 G、26.14 G、33.72 G、43.90 G 与 108.32 G 的浮点数,制约了其临床实用性。如图 9 所示,DPAR-Net 以 35.73 M 参数量和 23.93 G 浮点数的适中计算开销,实现 84.57% DSC 值与 12.23 HD95 值的双重突破,验证了本框架在精度与效率平衡方面的优势。

3 结 论

本研究针对医学图像分割中全局与局部特征协同建模不足的核心挑战,提出一种端到端深度神经

表 7 不同方法的参数量与分割性能对比分析表

Table 7 Comparative analysis table of parameters and segmentation performance of different methods

方法	参数量/M	浮点数/G	平均DSC ↑ /%	HD95 ↓
U-Net	31.04	41.93	76.85	39.70
Swin-UNet	21.75	5.93	79.13	21.55
TransUNet	93.23	24.73	77.48	31.69
MISSFormer	35.45	7.28	81.96	18.20
DAE-Former	29.69	26.16	82.43	17.46
HiFormer	34.14	17.81	80.39	14.70
PAG-TransYnet	116.77	33.72	83.43	15.82
TransCASCADÉ	115.01	26.14	82.68	17.34
MAXFormer	84.48	43.90	83.67	15.89
ParaTransCNN	42.70	50.39	83.86	15.86
Mixformer	123.35	108.32	82.64	12.67
DPAR-Net	35.73	23.93	84.57	12.23

网络 DPAR-Net。该框架创新性地构建双路径编码架构,通过并行化局部微尺度特征提取与全局上下

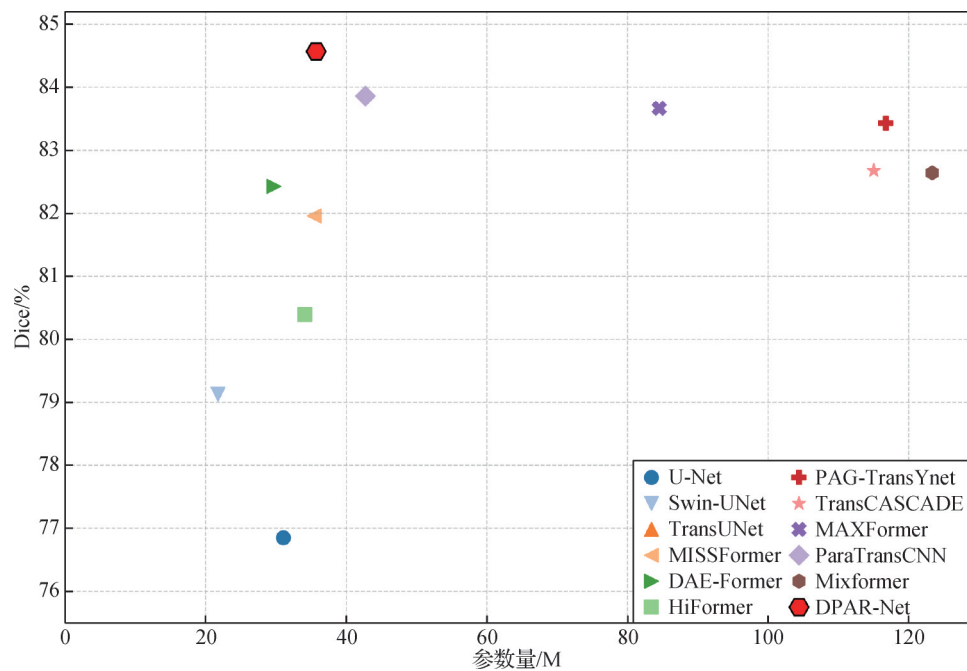


图9 模型性能比较图

Fig. 9 Model performance comparison

文语义建模,有效解决了传统方法中因局部特征与全局语义割裂所导致的性能瓶颈;同时设计了自适应感受野机制,依据医学图像多尺度特性自适应优化卷积核感受野,实现从微小组织到大型脏器的多尺度校准。在多个数据集上的实验表明,本文方法展现出卓越的分割效果和跨域适应性。在Synapse多器官CT分割数据集上取得84.57%的DSC与12.23 HD95指标;在ACDC心脏MRI分割任务中取得91.53%的DSC,实现性能突破。在模型轻量化架构设计中,DPAR-Net通过压缩Transformer模块数量及结合深度可分离卷积的策略,实现参数量级的高效优化,将参数量精简至35.73 M,较同类型模型TransUNet(93.23 M)、PAG-TransYnet(116.77 M)与ParaTransCNN(42.70 M)分别降低61.7%、69.4%与16.3%的存储开销,在精度与效率间获得了良好的平衡。

尽管DPAR-Net在Synapse腹部多器官CT图像分割任务及ACDC心脏MRI图像分割任务中取得了领先的性能,但其在强噪声干扰的超声影像数据中的有效性尚未验证。未来研究重点包括:1)针对超声影像的斑点噪声特性优化模型鲁棒性;2)拓展至高精度边界要求的临床场景,如结肠息肉和皮肤病图像分割任务。

参考文献(References)

Azad R, Arimond R, Aghdam E K, Kazerooni A and Merhof D. 2023a. DAE-Former: dual attention-guided efficient transformer for medical image segmentation//Proceedings of the 6th International Workshop on Predictive Intelligence in Medicine. Vancouver, Canada; Springer: 83-95 [DOI: 10.1007/978-3-031-46005-0_8]

Azad R, Heidari M, Shariatnia M, Aghdam E K, Karimijafarbigloo S, Adeli E, et al. 2022. TransDeepLab: convolution-free transformer-based DeepLab v3+ for medical image segmentation//Proceedings of the 5th International Workshop on Predictive Intelligence in Medicine. Singapore, Singapore: Springer: 91-102 [DOI: 10.1007/978-3-031-16919-9_9]

Azad R, Jia Y W, Aghdam E K, Cohen-Adad J and Merhof D. 2023b. Enhancing medical image segmentation with TransCeption: a multi-scale feature fusion approach. arXiv preprint arXiv: 2301.10847 [DOI: 10.48550/arXiv.2301.10847]

Bougourzi F, Dornaika F, Taleb-Ahmed A and Hoang V T. 2024. Rethinking attention gated with hybrid dual pyramid transformer-CNN for generalized segmentation in medical imaging [EB/OL]. [2025-04-23]. <https://arxiv.org/pdf/2404.18199.pdf>

Cao H, Wang Y Y, Chen J, Jiang D S, Zhang X P, Tian Q, et al. 2023. Swin-UNet: Unet-like pure transformer for medical image segmentation//Proceedings of 2023 European Conference on Computer Vision. Tel Aviv, Israel: Springer: 205-218 [DOI: 10.1007/978-3-031-25066-8_9]

Chai S, Jain R K, Mo S, Liu J, Yang Y, Li Y, et al. 2024. A novel

- adaptive hypergraph neural network for enhancing medical image segmentation//Proceedings of the 27th International Conference Medical Image Computing and Computer Assisted Intervention——MICCAI 2024. Marrakesh, Morocco: Springer: 23-33 [DOI: 10.1007/978-3-031-72114-4_3]
- Chang Q, Ahmad D, Toth J, Bascom R and Higgins W E. 2023. ESF-PNet: efficient deep learning architecture for real-time lesion segmentation in autofluorescence bronchoscopic video//Proceedings Volume 12468, Medical Imaging 2023: Biomedical Applications in Molecular, Structural, and Functional Imaging. San Diego, United States: #1246803 [DOI: 10.1117/12.2647897]
- Chen J N, Lu Y Y, Yu Q H, Luo X D, Adeli E, Wang Y, et al. 2021. TransUNet: transformers make strong encoders for medical image segmentation. arXiv preprint arXiv: 2102.04306 [DOI: 10.48550/arXiv.2102.04306]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. 2021. An image is worth 16×16 words: transformers for image recognition at scale [EB/OL]. [2025-04-23]. <https://arxiv.org/pdf/2102.04306.pdf>
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Heidari M, Kazerouni A, Soltany M, Azad R, Aghdam E K, Cohen-Adad J, et al. 2023. HiFormer: hierarchical multi-scale representations using transformers for medical image segmentation//Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE: 6191-6201 [DOI: 10.1109/WACV56688.2023.00614]
- Huang H M, Lin L F, Tong R F, Hu H J, Zhang Q W, Iwamoto Y, et al. 2020. UNet 3+: a full-scale connected UNet for medical image segmentation//Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE: 1055-1059 [DOI: 10.1109/ICASSP40776.2020.9053405]
- Huang X H, Deng Z F, Li D D, Yuan X G and Fu Y. 2023. MISSFormer: an effective transformer for 2D medical image segmentation. IEEE Transactions on Medical Imaging, 42(5): 1484-1494 [DOI: 10.1109/TMI.2022.3230943]
- Jiang T and Li X N. 2024. Segmentation of abdominal CT and cardiac MR images with multi scale visual attention. Journal of Image and Graphics, 29(1): 268-279 (蒋婷, 李晓宁. 2024. 采用多尺度视觉注意力分割腹部CT和心脏MR图像. 中国图象图形学报, 29(1): 268-279) [DOI: 10.11834/jig.221032]
- Li H, Liu J J and Xiao L. 2024. Dual-encoder global-local cross-attention network for medical image segmentation. Journal of Image and Graphics, 29(11): 3462-3475 (李赫, 刘建军, 肖亮. 2024. 融合交叉注意力与双编码器的医学图像分割. 中国图象图形学报, 29(11): 3462-3475) [DOI: 10.11834/jig.230705]
- Li J, Liu K Y, Hu Y T, Zhang H C, Heidari A A, Chen H L, et al. 2023. Eres-UNet++: liver CT image segmentation based on high-efficiency channel attention and Res-UNet++. Computers in Biology and Medicine, 158: #106501 [DOI: 10.1016/j.combiomed.2022.106501]
- Liang Z W, Zhao K, Liang G, Li S Y, Wu Y F and Zhou Y P. 2023. MAXFormer: enhanced transformer for medical image segmentation with multi-attention and multi-scale features fusion. Knowledge-Based Systems, 280: #110987 [DOI: 10.1016/j.knsys.2023.110987]
- Liao M, Yang R X, Zhao Y Q, Di S H and Yang Z. 2025. Multi-organ segmentation from abdominal CT images based on CE TransNet. Acta Automatica Sinica, 51(6): 1371-1387 (廖苗, 杨睿新, 赵于前, 邸拴虎, 杨振. 2025. 基于CE TransNet的腹部CT图像多器官分割. 自动化学报, 51(6): 1371-1387) [DOI: 10.16383/j.aas.c240489]
- Liu J, Li K Q, Huang C, Dong H, Song Y S and Li R H. 2025. MixFormer: a mixed CNN-transformer backbone for medical image segmentation. IEEE Transactions on Instrumentation and Measurement, 74: #5001220 [DOI: 10.1109/TIM.2024.3497060]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, et al. 2021. Swin transformer: hierarchical vision transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]
- Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, et al. 2018. Attention U-Net: learning where to look for the pancreas. arXiv preprint arXiv: 1804.03999 [DOI: 10.48550/arXiv.1804.03999]
- Rahman M M and Marculescu R. 2023. Medical image segmentation via cascaded attention decoding//Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE: 6211-6220 [DOI: 10.1109/WACV56688.2023.00616]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Sun H K, Xu J and Duan Y P. 2024. ParaTransCNN: parallelized TransCNN encoder for medical image segmentation [EB/OL]. [2025-04-23]. <https://arxiv.org/pdf/2401.15307.pdf>
- Tu Z Z, Talebi H, Zhang H, Yang F, Milanfar P, Bovik A, et al. 2022. MaxViT: multi-axis vision transformer//Proceedings of the 17th European Conference on Computer Vision (ECCV). Tel Aviv, Israel: Springer: 459-479 [DOI: 10.1007/978-3-031-20053-3_27]
- Wang W H, Xie E Z, Li X, Fan D P, Song K T, Liang D, et al. 2021. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 10701-10710 [DOI: 10.1109/ICCV48922.2021.01111]

- tional Conference on Computer Vision (ICCV). Montreal, Canada; IEEE: 548-558 [DOI: 10.1109/ICCV48922.2021.00061]
- Wu H S, Chen S H, Chen G L, Wang W, Lei B Y and Wen Z K. 2022. FAT-Net: feature adaptive transformers for automated skin lesion segmentation. *Medical Image Analysis*, 76: #102327 [DOI: 10.1016/j.media.2021.102327]
- Yan X Y, Tang H, Sun S L, Ma H Y, Kong D Y and Xie X H. 2022. AFTer-UNet: axial fusion transformer UNet for medical image segmentation//*Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, USA; IEEE: 3270-3280 [DOI: 10.1109/WACV51458.2022.00333]
- You C Y, Zhao R H, Liu F L, Dong S Y, Chinchali S, Topcu U, et al. 2022. Class-aware adversarial transformers for medical image segmentation//*Proceedings of the 36th International Conference on Neural Information Processing Systems*. New Orleans, USA; Curran Associates Inc.: 29582-29596
- Zeng H, Shan X X, Feng Y and Wen Y. 2023. MSAANet: multi-scale axial attention network for medical image segmentation//*Proceedings of 2023 IEEE International Conference on Multimedia and Expo (ICME)*. Brisbane, Australia; IEEE: 2291-2296 [DOI: 10.1109/ICME55011.2023.00391]
- Zhang Y D, Liu H Y and Hu Q. 2021. TransFuse: fusing transformers and CNNs for medical image segmentation//*Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Strasbourg, France; Springer: 14-24 [DOI: 10.1007/978-3-030-87193-2_2]
- Zhou Y P, Chang H Y, Lu X L and Lu Y H. 2022. DenseUNet: improved image classification method using standard convolution and dense transposed convolution. *Knowledge-Based Systems*, 254: #109658 [DOI: 10.1016/j.knosys.2022.109658]
- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-Net architecture for medical image segmentation//*Proceedings of the 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018 on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada, Spain; Springer: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]

作者简介

彭晏飞,男,教授,主要研究方向为计算机视觉。

E-mail: pengyf75@126.com

孙伟强,通信作者,男,硕士研究生,主要研究方向为医学图像分割。

E-mail: lnsqw17@126.com