

DOI:10.19651/j.cnki.emt.2417362

融合 CNN 和 Transformer 的三维冠状动脉 CT 图像分割*

潘成龙¹ 刘立程¹ 潘 丹²

(1. 广东工业大学信息工程学院 广州 510006; 2. 广东技术师范大学电子与信息学院 广州 510665)

摘 要: 冠状动脉的分割有利于医生快速诊断心血管疾病, 针对冠状动脉复杂结构和其它血管组织干扰带来的难分割, 造成大量碎片分割的问题, 保证模型对不同形态结构冠脉分割的自适应能力, 提出了一种新的三维冠状动脉分割网络模型 CA-SegNet。融合 CNN 和 Transformer 为骨干网络, 利用其优势和互补性, 充分提取冠状动脉的局部和全局特征。通过提出多尺度特征交互模块, 提取冠脉多尺度特征的同时进行特征通道之间的交互。在解码阶段, 提出注意力加权特征融合模块, 分别从空间和通道的角度对特征进行加权融合, 使模型更加关注冠状动脉区域。实验结果表明, 提出的模型在 Dice 相似系数、Recall、Precision 和 HD95 值分别达到了 81.96%、84.24%、80.11% 和 14.94, 优于当前流行的分割网络模型, 验证了 CA-SegNet 的有效性。

关键词: 冠状动脉分割; 卷积神经网络; Transformer; 特征融合; 注意力机制; 多尺度特征

中图分类号: TN911.73; TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Integration of CNN and Transformer for 3D coronary artery CT image segmentation

Pan Chenglong¹ Liu Licheng¹ Pan Dan²

(1. School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China; 2. School of Electronics and Information, Guangdong Polytechnic Normal University, Guangzhou 510665, China)

Abstract: Segmentation of coronary arteries is crucial for the rapid diagnosis of cardiovascular diseases. Given the challenges posed by the complex structure of coronary arteries and the interference from other vascular tissues, which often result in fragmented segmentation, ensuring the model's ability to adapt to segmenting different morphological structures of the coronary artery, a novel 3D coronary artery segmentation network (CA-SegNet) is proposed. This model incorporates a combination of CNN and Transformer as the encoder and decoder, leveraging their advantages and complementarity to fully extract both global and local features of coronary arteries. By proposing a multi-scale feature interaction module, the model simultaneously extracts multi-scale features of coronary arteries while facilitating feature channel interaction. In the decoding stage, an attention weighted feature fusion module is proposed to weight and fuse features from both spatial and channel perspectives, enabling the model to focus more on the coronary artery regions. Experimental results demonstrate that the proposed model achieves DSC, Recall, Precision, and HD95 values of 81.96%, 84.24%, 80.11% and 14.94 respectively, surpassing current popular segmentation models and validating the effectiveness of CA-SegNet.

Keywords: coronary artery segmentation; CNN; Transformer; feature fusion; attention mechanism; multi-scale feature

0 引 言

心血管疾病 (cardiovascular disease, CVD) 是中国乃至

全世界危害人类健康和导致死亡的主要原因。根据世界卫生组织 WHO 的报告^[1], 心血管疾病每年夺走数百万人的生命, 其中大部分由冠状动脉疾病引起。冠状动脉为心脏

收稿日期: 2024-11-14

* 基金项目: 国家自然科学基金 (61976058)、广东省科技计划项目 (2021A1515012300, 2021B0101220006)、广州市科技计划项目 (202103000034, 202206010007) 资助

供血,其通畅程度与整个心脏的健康密切相关。当胆固醇等脂肪沉积在冠状动脉壁上并转化为斑块时,容易导致冠状动脉狭窄。严重的狭窄会限制血液流动,从而减少心肌供氧。因此,量化冠状动脉狭窄对于冠心病的早期诊断和患者治疗至关重要。冠状动脉计算机断层血管造影(computed coronary tomography angiography, CCTA)作为一种无创诊断方法,因其高分辨率和多方位可视化能力,近年来被广泛应用于冠状动脉疾病的检测与评估。CCTA不仅能够清晰地显示冠状动脉的解剖结构,还能区分钙化和非钙化斑块,帮助临床医生更全面地评估冠状动脉狭窄的程度及斑块特征。然而,目前 CCTA 图像中斑块的位置及严重程度仍需放射科医师手工评估,不仅耗时,而且容易受到医师技术水平的限制,导致误诊和遗漏。此外,面对日益增长的影像数据量,传统手工评估方法的效率已难以满足临床需求。因此,针对 CCTA 图像中冠状动脉的自动分割研究逐步成为重点,为临床诊断提供技术支持。

传统的冠状动脉自动分割算法包括区域生长法^[2-3]、基于冠状动脉中心线的方法^[4]、阈值法^[5-6]、Hessian 矩阵法^[7]等。传统的冠状动脉分割算法在一定程度上能够提取血管结构,但也存在明显的局限性和不足,这些方法通常需要多次复杂的预处理步骤,例如指定固定区域来引导分割过程。另外,由于 CT 图像中血管边界与背景对比度较低,图像中可能存在噪声、血管厚度变化、分叉、交叉等现象,准确得到血管段是很困难的。近年来,深度学习方法^[8-16]在血管分割的效果明显优于非深度学习方法。U-Net^[17]是一种常用的血管分割有监督深度学习方法。受 U-Net 结构的启发,一系列基于 UNet 的模型变体被提出,如 Attention-UNet^[18]、Mobile-UNet^[19]等。一些其他各种架构也被设计出来。Shen 等^[20]首先使用具有注意机制的 3D-FCN 网络对冠状动脉和升主动脉进行分割,然后使用水平集方法对网络的分割结果进行进一步优化。但在 CCTA 图像上,升主动脉不仅显著且易于分割,而且其体积比例远高于冠状动脉。因此,如果只分割冠状动脉,DSC 将显著下降。Zhu 等^[21]提出一种基于时空特征融合结构的 U 形网络,从 CCTA 心脏图像的二维切片中分割冠状动脉。时空特征将多个层次和不同感受野的特征组合在一起,得到更精确的边界。但二维切片只是一例的一张,每次模型只学习一片的特征,片与片之间的信息得不到保留。Wang 等^[22]提出了一种新的冠状动脉分割框架,该框架结合了三维卷积输入层和二维卷积网络。三维输入层利用图像序列中的时间信息,将多幅图像融合成更全面的二维特征图,二维卷积网络完成分割任务。Song 等^[23]首先使用二维分类网络 Densenet 对非冠状动脉切片进行筛选,使训练变得更高效,再将其切块输入到三维分割网络。但其步骤繁琐,需先分类再分割,且用于分割网络的切块较小,网络很难学习到全面的三维冠脉特征。Tian 等^[24]将 V-Net 与级联 U 型网络结合,充分利用图像的上下文信息,通过位置偏移损失提

升冠状动脉分割的连贯性。但网络仅仅学习到冠脉结构的局部信息,冠状动脉的整体轮廓,即全局信息丢失严重。

不同于上述方法,针对冠状动脉复杂结构和其它血管组织干扰带来的难分割,造成大量碎片分割的问题。同时保证模型对不同形态结构冠脉的适应能力,提高分割结果的精度。本研究提出一种融合 CNN 和 Transformer 的三维冠状动脉分割方法。为提升模型的全局和局部特征提取能力,本研究设计了一个 3D Enc 块,在编码阶段构建连续堆叠的 3DSwin-T 块,通过使用常规窗口和移位窗口进行特征采样和窗口信息交流,以获取全局信息。在 3DSwin-T 块旁并连上 3D 卷积块,同时获取特征局部信息。在解码阶段,为了使模型应对不同形态、大小的冠状动脉树具有一定的自适应能力,进一步设计了多尺度特征交互模块(multi-scale feature cross, MFC)。该模块通过使用多个不同扩张率卷积,使得模型能够从不同感受野进行冠状动脉尺度特征信息的提取。此外,通道分割和重组的操作进一步加强了不同尺度信息的融合。最后,为了减少分割碎片的产生,设计了注意力加权特征融合模块(attention weighted feature fusion, AWWF)。该模块同时从通道和空间的维度出发,使模型更加关注图像中冠状动脉区域,减少来自背景中其他器官的噪声信息,保留更多有效信息。

1 本文方法

本文所提方法包括 3 个步骤:首先是数据预处理部分。然后,将预处理后的数据作为模型(coronary artery segmentation network, CA-SegNet)的输入。每个体素块的分割结果都是通过冠状动脉分割模型得到的。最后,通过后处理对分割结果进行细化,主要是将每个体素块分割结果合并成一个完整的冠脉形状。流程图如图 1 所示。

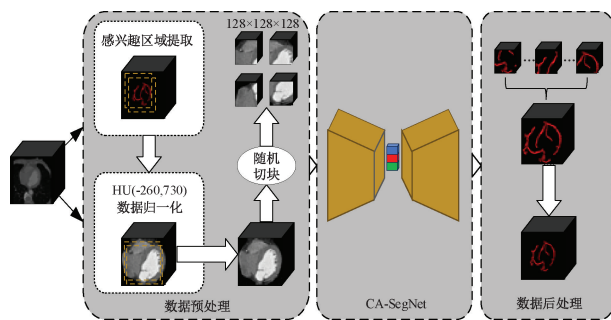


图 1 冠状动脉分割流程图

Fig. 1 Flowchart of coronary artery segmentation

1.1 数据预处理

数据预处理目的是提高血管体素对比度,便于模型特征提取。本文预处理步骤为:

1)为了避免冠脉周围组织的影响,减少计算量,根据冠脉掩模在图像中的位置,对 CT 和标签图像提取感兴趣(region of interest, ROI)区域。

- 2)使图像的体素间距相同,减少不同图像之间的不一致性,对所有的图像重采样到 $[0.5, 0.5, 0.5]$ 。
- 3)将图像的 CT 值限制在 $[-260, 730]$ HU,过滤掉其他器官,使模型更好地聚焦于冠状动脉区域。
- 4)对所有图像进行标准化,去除均值,除以标准差,最后归一化,使所有数据映射到 $[0, 1]$ 的范围内,并将 CT 图像裁剪到为 $128 \times 128 \times 128$ 。

1.2 网络模型

如图 2 所示为基于编码器-解码器架构的 CA-SegNet。编码器使用多个 3D Enc-blk 作为骨干网络,用于同时提取全局特征和局部特征。在编码器和解码器之间放置多个 AFFF 模块代替跳跃连接,用于逐步融合特征,丰富底层特征的语义信息。MFC 模块设计在瓶颈层,增加感受野,提取多尺度特征并实现尺度特征之间的交互。

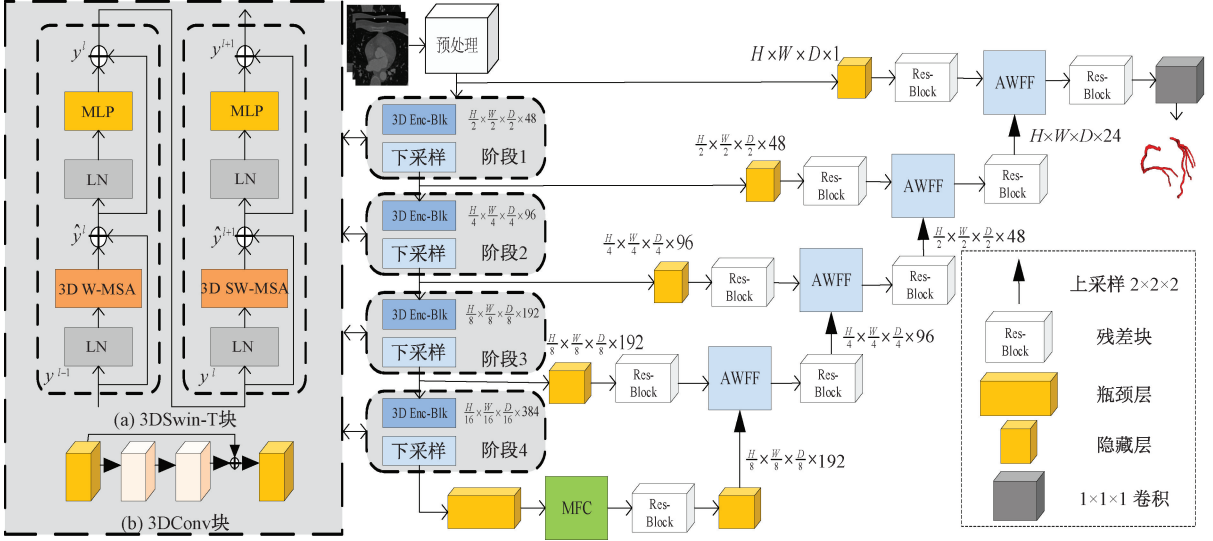


图 2 本文所提模型框架

Fig. 2 The overall framework of the model proposed in this paper

1.3 3D Swin-T 块

对于三维医学图像数据,CA-SegNet 在骨干网络中设计了 3D Enc-blk 作为每个编码块,如图 3 所示。在其中提出了 3D Swin-T 块,该块执行 3D 窗口划分操作并进行注意力的计算,提取图像全局信息。3DSwin-T 块超参数如表 1 所示。

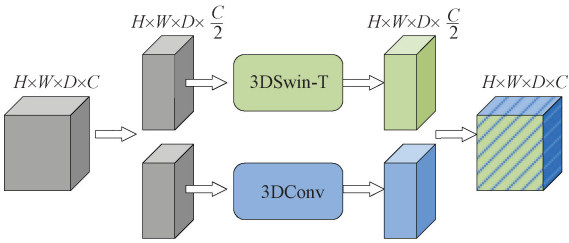


图 3 3DEnc-blk

Fig. 3 3D Encoder block

参考 Swin Transformer^[25]的设计,本文使用了两种类型的窗口划分操作:常规窗口和移位窗口,分别用 3D W-MSA 和 3D SW-MSA 表示。图 4 展示了划分窗口的操作。3D SW-MSA 采用偏移窗口,输入特征图在每个维度上依次移动 $M/2$ 个体素, M 为窗口的大小。假设 3DSwin-T 输入 $8 \times 8 \times 8$ 的特征图,在第 L 层使用 $4 \times 4 \times 4$ ($M=4$) 大小的窗口将其分为 8 个互不重叠的窗口。在第 $L+1$ 层,将特征图往 H, W, D 3 个维度移动 $M/2=2$ 个体

表 1 3D Swin-T 块超参数

Table 1 Hyper-parameter of 3D Swin-T block

参数名称	数量
Head	32
Num heads	(3, 6, 12, 24)
窗口大小	4
MLP ratio	4
Num layers	(2, 2, 6, 2)
下采样因子	(2, 2, 2, 2)

素。如图 4 所示,对于偏移后的 $L+1$ 层,上方 $2 \times 2 \times 4$ 大小长方体块可以融合第 L 层上方两个窗口之间的信息,下方 $2 \times 2 \times 4$ 大小长方体块可以融合第 L 层下方两个窗口之间的信息,对于左边和右边同理。中间 $4 \times 4 \times 4$ 大小的窗口(箭头处)可以融合第 L 层所有 8 个窗口之间的信息。因此,偏移窗口可以解决窗口与窗口间无法信息传递的问题,使网络模型能够学习到全局特征。计算每个窗口中 token 之间的注意力的过程可以用以下公式来描述。

$$Q = X \times W_Q, K = X \times W_K, V = X \times W_V \quad (1)$$

$$attn = \text{Softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \times V \quad (2)$$

其中, X 为单个窗口内的 token 矩阵,每个 W_Q, W_K, W_V

W_V 是具有相同维度的可学习矩阵, d_K 为 K 的维数, $attn$ 为计算得到的 token 之间的相似度。

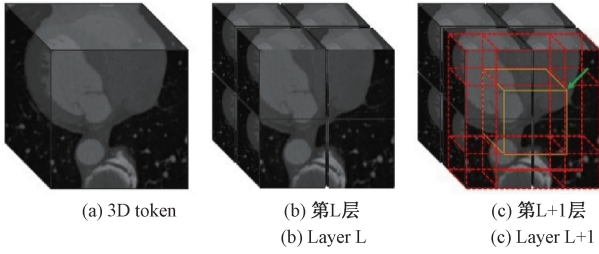


图 4 窗口划分方式

Fig. 4 Window partitioning method

其它 3DSwin-T 的组成包括 MLP、层归一化(LN)和残差连接。图 2(a)描述了每个阶段内的 3DSwin-T 块, 公式为:

$$\hat{y}^l = 3DW - MSA(LN(y^{l-1})) + y^{l-1} \quad (3)$$

$$y^l = MLP(LN(\hat{y}^l)) + \hat{y}^l \quad (4)$$

$$\hat{y}^l += 3DSW - MSA(LN(y^l)) + y^l \quad (5)$$

$$y^{l+1} = MLP(LN(\hat{y}^l + 1)) + \hat{y}^l + 1 \quad (6)$$

其中, 3DW-MSA 和 3DSW-MSA 分别代表常规窗口和偏移窗口机制, $\hat{y}^l$ 和 y^l 分别为 3D(S)W-MSA 和 MLP 在第 L 层的输出。

1.4 3D 卷积块

如图 2(b) 所示, 3D Conv 块由 $3 \times 3 \times 3$ 卷积层、LayerNorm 层、Relu 层的顺序叠加两次, 主要负责学习图像的局部依赖关系, 提取局部特征。该模块的计算过程为:

$$Y_1 = \text{Relu}(\text{LN}(3D\text{Conv}(X))) \quad (7)$$

$$Y_2 = \text{Relu}(\text{LN}(3D\text{Conv}(Y_1))) \quad (8)$$

$$Y = Y_2 + X \quad (9)$$

其中, X 表示 3D Conv 块的输入, Y 表示 3D Conv 块的输出。为了避免该模块带来的额外计算量, 使用深度可分离卷积代替普通卷积。该模块的设计使模型能够更好地拟合图像中的局部细节, 提取局部信息。

1.5 多尺度特征交互模块(MFC)

冠状动脉的尺度信息变化多样, 因此多尺度上下文信息的提取和特征整合至关重要。由于 U-Net 及其变体感受野固定, 导致其在多尺度特征提取上存在不足。针对该问题, Wang 等^[26]提出了一种结合标准卷积和多次扩张卷积的 U-Net 变体, 用于医学图像分割。然而, 该方法在多尺度特征之间缺少信息交互, 且对不同形态和大小的分割目标缺乏自适应性。本文基于此设计了多尺度特征交互模块, 利用不同扩张率的卷积获取多尺度特征, 并通过通道分割和重组增强特征交互效果。

MFC 模块如图 5 所示。MFC 将输入特征 $X \in R^{C \times H \times W \times D}$ 沿着通道维度划分为 4 个不同的头, 生成

$(X_i)_{i=1}^4 \in R^{\frac{C}{4} \times H \times W \times D}$ 。然后, 每个头部分别送入具有不同扩张率的扩张卷积, 得到 $(X'_i)_{i=1}^4 \in R^{\frac{C}{4} \times H \times W \times D}$, 以有效地捕获多尺度的信息。卷积核的扩张率分别为 d_1, d_2, d_3 和 d_4 。

$$X'_i = D\text{Conv}(X_i) \quad (10)$$

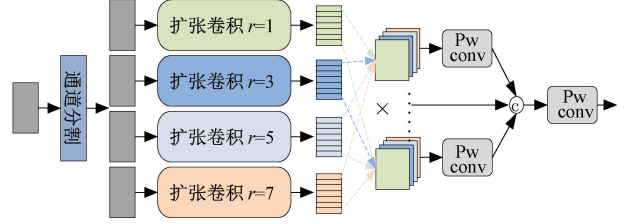


图 5 多尺度特征交互模块

Fig. 5 Multi-scale feature interaction module

其中, X'_i 表示通过对第 i 个头应用扩张卷积获得的特征, $D\text{Conv}$ 表示扩张卷积, 采用的扩张率分别为 $i=1, 3, 5, 7$ 。MFC 通过通道分割和重组来增强特征表示。具体来说, 将 X'_i 分成单独的通道, 对于每个头部, 得到 $(X'_i)_{j=1}^{\frac{C}{4}} \in R^{1 \times H \times W \times D}$ 。接下来, 将每个头部的单个通道交叉重组成 $(h_j)_{j=1}^{\frac{C}{4}} \in R^{4 \times H \times W \times D}$, 从而增强了多尺度特征的多样性。最后, 利用点卷积进行组间和组内信息融合, 得到输出 $Y \in R^{C \times H \times W \times D}$, 实现轻量高效的聚合效果。

$$h_j = \text{Conv}_1(\text{Cat}(X'_1, X'_2, X'_3, X'_4)) \quad (11)$$

$$Y = R(L(\text{Conv}_1(\text{Cat}(h_1, h_2, \dots, h_j)))) \quad (12)$$

其中, Conv_1 表示卷积核大小为 $1 \times 1 \times 1$ 的点卷积, X'_i 表示第 i 个头部的第 j 个通道, h_j 表示第 j 组特征 ($i \in 1, 2, 3, 4; j \in 1, 2, \dots, \frac{C}{4}$)。 R 和 L 分别表示 Relu 激活函数和层归一化。

1.6 注意力加权特征融合模块(AWFF)

大多数现有方法通常采用跳跃连接实现特征融合。跳跃连接通过将上采样输出与对应层编码器特征图进行拼接, 修复下采样过程中丢失的细节信息。这种方式会导致一些噪声的叠加。此外, 不相关特征的拼接可能会导致性能下降, 甚至产生错误的分割碎片。

注意力特征融合 (attention-guided feature fusion, AGFF)^[27] 由通道注意构成, 它计算通道维度上的注意力权重。通道注意力计算出一个通道向量, 用以重新校准特征, 从而提高冠状动脉分割精度。尽管 AGFF 引入了通道注意力来加权特征, 但没有考虑到特征的空间性。

由于冠脉是一个小目标, 分割过程中不能忽略目标的局部细节信息。为了增强目标与背景的可分辨性, 动态地表示信息区域, 本文进一步设计了注意力加权特征融合模块, 充分混合通道注意力和空间注意力。如图 6 所示。

首先将来自编码器的特征 X_1 和同一层的解码器的特征 X_2 相加为 X , 再输入到空间注意力和通道注意力中, 通

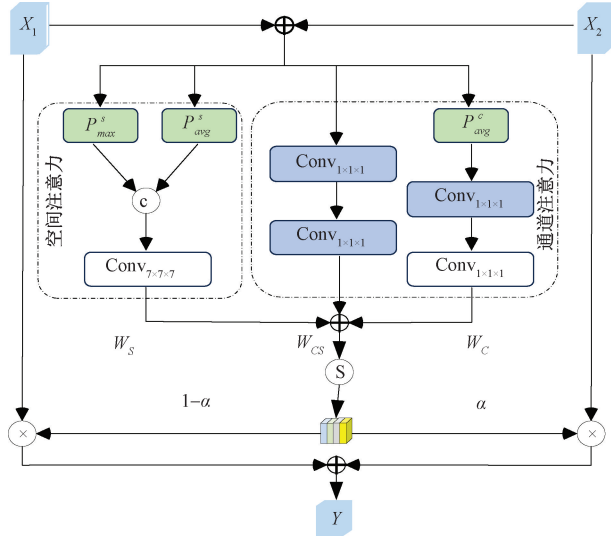


图 6 注意力加权特征融合模块

Fig. 6 Attention weighted feature fusion module

过以下公式计算相应的 W_c 和 W_s 。

$$W_c = \text{Conv}_1(\text{R}(\text{Conv}_1(X))) + \text{Conv}_1(\text{R}(\text{Conv}_1(P_{\text{avg}}^c(X)))) \quad (13)$$

$$W_s = \text{Conv}_7(\text{Cat}(P_{\text{max}}^s(X), P_{\text{avg}}^s(X))) \quad (14)$$

其中, P_{max}^s 、 P_{avg}^s 和 P_{avg}^c 分别代表沿空间维度的全局最大化、全局平均池化和沿通道维度的全局平均池化; Conv_k 表示卷积核大小为 $k \times k \times k$ 的卷积; R 表示 Relu 激活函数; Cat 表示通道维度的拼接操作。为了减少参数数量并限制模型复杂度,第 1 次 $1 \times 1 \times 1$ 卷积将通道维数从 C 降为 $\frac{C}{r}$ (r 为降维比),第 2 次 $1 \times 1 \times 1$ 卷积将通道维数扩展回 C 。在模型中,选择将 r 设置为 4。

根据广播机制,通过加法运算将 W_c 和 W_s 融合在一起,得到 $W_{cs} \in R^{C \times H \times W \times D}$ 。 W_{cs} 和 X 具有相同的特征维度。

$$W_{cs} = W_c + W_s \quad (15)$$

通过学习到的权重 W_{cs} ,使用 Sigmoid 函数获得注意力系数 α ,最后采用加权求和的方法,动态地融合编码器部分的低级特征和相应解码器的高级特征。可表示为:

$$Y = X_1 \times (1 - \alpha) + X_2 \times \alpha \quad (16)$$

$Y \in R^{C \times H \times W \times D}$ 为融合后的特征,AWFF 模块如图 6 所示。图 6 中虚线代表 $1 - \alpha$,与 α 一样,均取值在 $0 \sim 1$ 之间,使得模块可以对输入特征 X_1 和 X_2 进行注意力加权平均。在 CA-SegNet 中,使用了 4 个注意特征融合模块,如图 2 所示。

2 实验结果及分析

2.1 数据集

本实验使用的数据集为广东省人民医院公开的 ImageCAS^[28] 冠状动脉数据集,来自 1 000 例患者。图像尺寸为 $512 \times 512 \times (206 \sim 275)$ 体素,平面分辨率为 $0.29 \sim$

0.43 mm^2 ,体素间距为 $0.25 \sim 0.45 \text{ mm}$ 。每张图像的左右冠状动脉由两名放射科医生独立标记,标注结果经过交叉验证。如果标注不一致,第 3 位放射科医生将重新标记,最终结果由共识决定。数据集的详细信息如表 2 所示。

表 2 数据集
Table 2 Datasets

名称	ImageCAS
图像大小	$512 \times 512 \times (206 \sim 275)$
裁剪大小	$128 \times 128 \times 128$
训练集	800
验证集	100
测试集	100

2.2 实验设置

所有实验均使用深度学习框架 PyTorch 1.16.0 实现,并在 NVIDIA GeForce GTX 3090Ti GPU 上完成。模型参数通过 Adamw 进行优化,初始学习率设置为 1×10^{-4} ,采用余弦退火学习率衰减策略,权重衰减为 1×10^{-4} ,训练迭代次数为 50 轮。

2.3 评价指标

为了定量比较该方法与最先进方法的分割性能,本文采用了多种评估指标。通过计算血管分割的 Dice 相似系数(dice similarity coefficient, DSC)、精确率(Precision)、召回率(Recall)和豪斯多夫距离(hausdorff distance, HD),来比较分割的准确性。Dice 相似系数表示预测正确结果面积的两倍与预测标签和实际标签面积之和的比值。DSC、Precision、Recall 取值范围均为 $0 \sim 1$,值越大,分割的准确率越高。豪斯多夫距离用于描述两组点集之间的相似程度,主要衡量边界的分割精度。本实验使用的是第 95 百分位豪斯多夫距离(HD95),与其他相关研究中常用的方法一致,对异常值的鲁棒性更强。HD95 值越小,表示匹配程度越高。各公式定义为:

$$\text{DSC} = \frac{2TP}{2TP + FP + FN} \quad (17)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (18)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (19)$$

$$d_H\{A, B\} = \max\{\max_{a \in A} \min_{b \in B} d(x, y), \max_{b \in B} \min_{a \in A} d(x, y)\} \quad (20)$$

其中, TP 和 FP 分别表示真阳性和假阳性,分别对应被模型正确分割的血管体素数量和错误分割的背景体素数量。 FN 表示假阴性,对应血管体素被错误分割为背景的数量。 A 和 B 是比较的两个点集, $d(A, B)$ 表示 A 和 B 之间的欧几里得距离。

2.4 实验结果

为了比较 CA-SegNet 与其他 3D 冠状动脉分割模型的性能,本文复现了一些当前主流的医学图像分割模型,包括 3D-Unet^[29], Vnet^[30], Att-Unet^[18], SegresNet^[31], Unetr^[32],Swin-Unetr^[33]和 3DUX-Net^[34]。这些模型同样使用此冠状动脉数据集进行训练和测试,并采用与 CA-SegNet 相同的训练、验证和测试参数设置。

表 3 展示了 CA-SegNet 与其他三维分割模型的比较。在所有 4 个评估指标中,本文提出的 CA-SegNet 比基线 Swin-Unetr 取得了更好的结果。具体来说,与 Swin-Unetr 相比,CA-SegNet 的 DSC 提高了 2.75%,Recall 提高了 2.4%,Precision 提高了 2.92%,HD 降低了 10.02 mm。CA-SegNet 在 DSC、Recall、Precision 和 HD 方面也基本优于其他模型。CA-SegNet 在各种评估指标上的提升,表明其在冠状动脉分割任务中的优势。Recall 的提高表明更多的冠状动脉被正确分割,DSC 的提高表明整体分割结果更好,更接近真实标签。虽然 SegresNet 在 Precision 上优于 CA-SegNet,这表明其错误将背景体素分割为冠状动脉的数量较少,但其 DSC 和 Recall 均低于本文提出的模型,这意味着它在分割结果中遗漏了更多的冠状动脉体素。

表 3 不同模型分割结果对比

Table 3 Comparison of segmentation results of different models				
模型	DSC/%	Recall/%	Precision/%	HD/mm
3D Unet ^[29]	77.81	78.42	77.76	30.28
Vnet ^[30]	76.87	83.90	71.60	36.93
Att-Unet ^[18]	79.15	85.07	74.42	38.08
SegresNet ^[31]	78.39	74.75	82.98	26.82
Unetr ^[32]	77.70	77.61	78.23	22.57
Swin-Unetr ^[33]	79.21	81.84	77.19	24.96
3D UX-Net ^[34]	77.66	82.05	74.13	46.14
CA-SegNet	81.96	84.24	80.11	14.94

2.5 模型参数量与推理时间对比

为了更加准确地分析模型性能,结合上述分割实验,本文比较了 CA-SegNet 与其他三维分割模型的参数量、分割精度以及平均推理时间的数据,如表 4 所示。本文模型参数量为 18.19 M,因为本文模型中多尺度特征交互模块和注意力加权特征融合模块相对轻量化,同时 3DSwin-T 块中(S)W-MSA 窗口大小为 4×4×4,相比于原始窗口大小为 7×7×7 的 Swin-Unetr,参数量大大减少。与 3D UX-Net 相比,参数量减少了 34.82 M,而 Unetr 参数量甚至达到了本文模型的 5 倍多。在分割平均推理时间方面,本文模型的推理速度相较于 3D UX-Net 和 Unetr 更快。但本文模型的在参数量较少的情况下平均分割精度最高,因此本文模型的分割性能整体要优于其他 7 种分

割模型。

表 4 模型参数量与推理时间对比

Table 4 Comparison of model parameters quantity and inference time		
模型	时间/s	参数量/M
3D Unet ^[29]	34.34	7.75
Vnet ^[30]	42.58	23.43
Att-Unet ^[18]	43.63	45.60
SegresNet ^[31]	39.84	18.97
Unetr ^[32]	49.95	93.01
Swin-Unetr ^[33]	43.20	61.98
3DUX-Net ^[34]	57.90	53.01
CA-SegNet	48.27	18.19

2.6 模型对噪声鲁棒性验证

预处理可以减少 CT 图像中的背景噪声,从而提升分割模型的性能。为了验证 CA-SegNet 对背景中噪声信息的鲁棒性,即模型在噪声存在情况下仍能保持稳定性能,进行了对比实验,如表 5 所示。第 1 组实验直接使用未经数据预处理(仅裁剪和归一化)的原始测试图像进行测试,其中图像包含背景中其他器官或无关区域的信息;第 2 组实验则是按照 1.1 节的数据预处理流程处理后进行测试的。两组实验分别在相同的模型架构和参数设置下进行,确保对比的公平性。

表 5 模型对噪声鲁棒性验证

Table 5 Verification of the robustness of the model to noise				
实验组别	DSC/%	Recall/%	Precision/%	HD/mm
①	80.98	84.47	78.13	31.73
②	81.96	84.24	80.11	14.94

实验结果表明,两组测试中 CA-SegNet 均表现出了较高的分割精度。其中第 1 组(未经数据预处理)的主要指标 DSC 略低于第 2 组(经过数据预处理),但差距非常小,仅为 0.98%,且优于现有的分割模型。这表明,尽管背景中存在一定的噪声干扰,模型依然能够准确地分割目标区域,整体性能表现相对稳定,体现出模型对背景噪声具有较好的鲁棒性。但在 HD 指标上,两者的差距较大。这一结果表明,部分背景噪声被模型误判为冠状动脉区域,导致边界偏差增大。同时也表明了数据预处理可以去除部分噪声干扰,通过优化数据预处理流程可以进一步提升性能的潜力。本实验验证了 CA-SegNet 在一定程度上能够抵抗背景噪声的干扰,在主要分割指标上保持较高性能,表现出良好的鲁棒性。

2.7 消融实验

为了证明 CA-SegNet 中各模块相关性能的有效性,进

行了消融研究,结果如表 6 所示。可以看出,与 Baseline 相比,本文在 Baseline 上添加所提出的 MFC 和 AWFF 模块在 4 个评价指标上均有明显的提升。具体来说,添加 MFC 模块后的 Precision 提高了 3.21%,而添加 AWFF 模块后的 Recall 提高了 1.57%。

表 6 消融实验的对比

Table 6 The comparison of ablation experiments						
模型	MFC AWFF		DSC/ %	Recall/ %	Precision/ %	HD/ mm
Baseline			79.97	81.00	79.37	19.26
模型 1	✓		80.98	79.82	82.58	16.30
模型 2		✓	81.12	82.57	80.09	15.62
CASegNet	✓	✓	81.96	84.24	80.11	14.94

消融研究结果表明,单独的 MFC 和 AWFF 模块在提高分割性能方面是有效的。MFC 模块有助于提高精确率,而 AWFF 模块对召回率有较大的影响。通过结合 MFC 和 AWFF 模块,CA-SegNet 在 DSC、Recall、Precision 和 HD 方面都具有良好的性能,突出了两个模块的互补优势。

2.8 分割结果可视化

图 7 所示为 Unetr、Swin-Unetr、3D-UXNet 和 CA-SegNet 从测试集中随机抽取的 4 例分割结果的冠状动脉树最终三维重建可视化对比。可以看出,CA-SegNet 的分割效果优于其他几种方法,对冠脉细节和全局结构都有更好的保留能力。结果表明所提出的 CA-SegNet 比其他方法产生的误报更少,减少了分割碎片的产生,如图 7 圆形框所示。同时,CA-SegNet 使冠状动脉分割更连续,如图 7 矩形框所示。

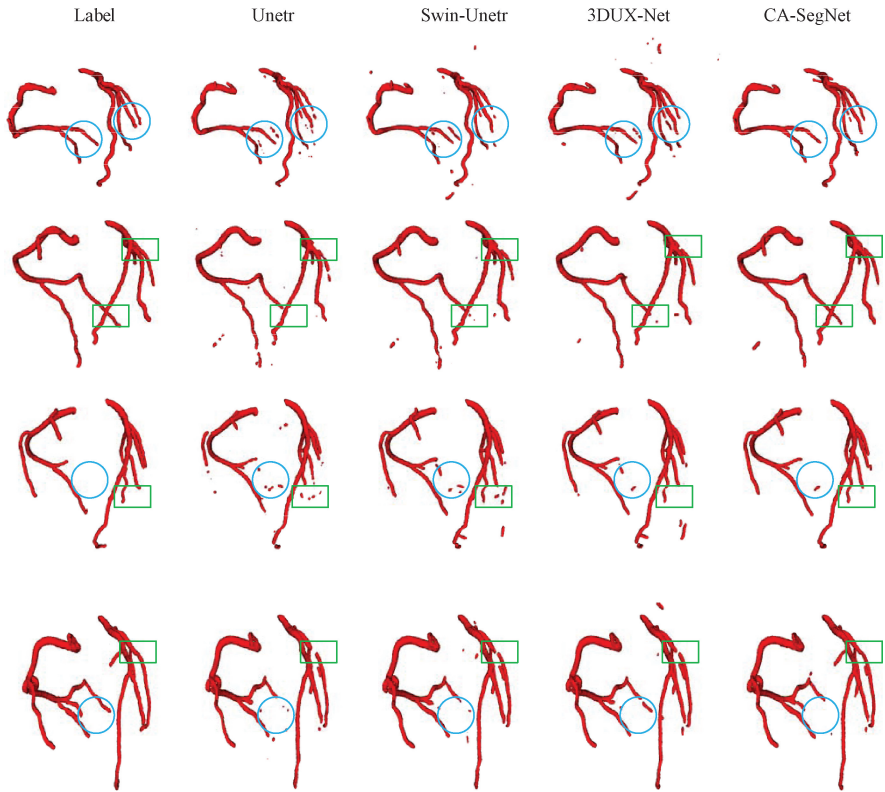


图 7 不同模型分割结果可视化

Fig. 7 Visualization of segmentation results from different models

然而,在某些血管密集的区域,CA-SegNet 表现出轻微过度分割的情况,如图 7 第 1 行的案例,其它的模型也存在类似的问题。这是因为模型误识别了与冠状动脉 CT 值相似的其他结构,这种过度分割现象表明该模型在识别相似结构时仍有改进空间。为了解决这一局限性,计划在后续研究中进行进一步优化分割框架,以更好地区分这些相似的结构。

3 结 论

本研究提出了一种新的三维冠状动脉分割网络模型

CA-SegNet,能有效减少因冠状动脉复杂结构和其它血管组织干扰带来的难分割,造成的大量分割碎片的问题,同时保证模型对不同形态结构冠脉的自适应能力。该模型通过融合 CNN 和 Transformer,利用它们的互补特性,同时提取冠状动脉的局部和全局特征。在多尺度特征交互模块中,采用级联结构提取冠脉多尺度特征,并通过通道分割和重组,以实现特征交互,使模型应对不同形态、大小的冠状动脉树具有一定的自适应能力。在解码阶段,设计了注意力加权特征融合模块,分别从空间和通道的角度对

特征进行加权融合,抑制背景噪声,使模型更加关注冠状动脉区域,减少冠状动脉产生错误分割的问题。在大规模数据集 ImageCAS 上的实验结果表明,本研究提出的 CA-SegNet 优于当前主流的分割网络模型,消融实验验证了所提模块的有效性。

参考文献

- [1] 数据“说”全球心血管疾病负担[J]. 实用心脑血管病杂志, 2023, 31(3): 5.
Data ‘say’ global burden of cardiovascular disease[J]. Practical Journal of Cardio-Cerebro-Pulmonary Vascular Disease, 2023, 31(3): 5.
- [2] KULATHILAKE K, RANATHUNGA L, CONSTANTINE G R, et al. Region growing segmentation method for extracting vessel structures from coronary cine-angiograms[C]. 2015 Moratuwa Engineering Research Conference(MERCon). IEEE, 2015: 142-147.
- [3] WANG SH, LI B N, ZHOU SH J. A segmentation method of coronary angiograms based on multi-scale filtering and region-growing[C]. 2012 International Conference on Biomedical Engineering and Biotechnology. IEEE, 2012: 678-681.
- [4] SHENG X, FAN T, JIN X, et al. Extraction method of coronary artery blood vessel centerline in CT coronary angiography[J]. IEEE Access, 2019, 7: 170690-170702.
- [5] LIAN Y X, WANG Y Y, YU J H, et al. Segmentation of arteriovenous malformations nidus and vessel in digital subtraction angiography images based on an iterative thresholding method[C]. The 8th International Conference on Biomedical Engineering and Informatics(BMEI). IEEE, 2015: 111-115.
- [6] CRUZ-ACEVES I, OLOUMI F, RANGAYYAN R M, et al. Automatic segmentation of coronary arteries using Gabor filters and thresholding based on multiobjective optimization[J]. Biomedical Signal Processing and Control, 2016, 25: 76-85.
- [7] ZHAO J, GONG W K, JIANG SH ZH, et al. Automatic segmentation and reconstruction of coronary arteries based on sphere model and Hessian matrix using CCTA images[C]. Journal of Physics: Conference Series. IOP Publishing, 2019, 1213(4): 042049.
- [8] JIANG W L, LI Y M, JIA Y H, et al. Ori-net: Orientation-guided neural network for automated coronary arteries segmentation[J]. Expert Systems with Applications, 2024, 238: 121905.
- [9] WANG Q J, XU L S, WANG L, et al. Automatic coronary artery segmentation of CCTA images using UNet with a local contextual transformer[J]. Frontiers in Physiology, 2023, 14: 1138257.
- [10] 夏平,何志豪,雷帮军,等. 密集级联卷积与自注意力特征聚合的视网膜血管分割算法[J]. 电子测量与仪器学报, 2024, 38(9): 36-44.
XIA P, HE ZH H, LEI B J, et al. Retinal vascular segmentation algorithm based on full scale dense convolutional u-shaped networks[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(9): 36-44.
- [11] LI Y M, WU Y, HE J J, et al. Automatic coronary artery segmentation and diagnosis of stenosis by deep learning based on computed tomographic coronary angiography[J]. European Radiology, 2022, 32(9): 6037-6045.
- [12] SERRANO A B, OTERO C A, LÓPEZ O D, et al. Coronary artery segmentation based on transfer learning and UNet architecture on computed tomography coronary angiography images[J]. IEEE Access, 2023, DOI: 10.1109/ACCESS.2023.3293090.
- [13] 谷学静, 公海昌, 王静波. 基于深度学习的冠状动脉 OCT 图像易损斑块识别[J]. 国外电子测量技术, 2023, 42(5): 43-49.
GU X J, GONG H CH, WANG J B. Identification of vulnerable plaques in coronary artery OCT images based on deep learning[J]. Foreign Electronic Measurement Technology, 2023, 42(5): 43-49.
- [14] 邸拴虎, 杨文瀚, 廖苗, 等. 基于 RA-Unet 的 CT 图像肝脏肿瘤分割[J]. 仪器仪表学报, 2022, 43(8): 65-72.
DI SH H, YANG W H, LIAO M, et al. Liver tumor segmentation from CT images based on RA-Unet[J]. Chinese Journal of Scientific Instrument, 2022, 43(8): 65-72.
- [15] DING W P, SUN Y, HUANG J S, et al. RCAR-UNet: Retinal vessel segmentation network algorithm via novel rough attention mechanism[J]. Information Sciences, 2024, 657: 120007.
- [16] 许瑶瑶, 单剑锋. 基于密集连接和 Inception 模块的前列腺图像分割[J]. 电子测量技术, 2022, 45(15): 151-157.
XU Y Y, SHAN J F. Prostate image segmentation based on dense connections and Inception module[J]. Electronic Measurement Technology, 2022, 45(15): 151-157.
- [17] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]. Medical Image Computing and Computer-Assisted Intervention-MICCAI the 18th International Conference, 2015, proceedings, part III 18. Springer International Publishing, 2015: 234-241.

- [18] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-net: Learning where to look for the pancreas [J]. ArXiv preprint arXiv:1804.03999, 2018.
- [19] 陈铭, 梅雪, 朱文俊, 等. 一种新型 Mobile-Unet 网络的肺结节图像分割方法[J]. 南京工业大学学报(自然科学版), 2022, 44(1): 76-81, 91.
- CHEN M, MEI X, ZHU W J, et al. A novel pulmonary nodule segmentation method using Mobile-Unet network[J]. Journal of Nanjing Tech University (Natural Science Edition), 2022, 44(1): 76-81, 91.
- [20] SHEN Y, FANG ZH J, GAO Y B, et al. Coronary arteries segmentation based on 3D FCN with attention gate and level set function[J]. IEEE Access, 2019, 7: 42826-42835.
- [21] ZHU H Y, SONG SH N, XU L SH, et al. Segmentation of coronary arteries images using spatio-temporal feature fusion network with combo loss[J]. Cardiovascular Engineering and Technology, 2022: 1-12, DOI: <https://doi.org/10.1007/s13239-021-00588-x>.
- [22] WANG L, LIANG D X, YIN X L, et al. Coronary artery segmentation in angiographic videos utilizing spatial-temporal information [J]. BMC Medical Imaging, 2020, 20: 1-10.
- [23] SONG A L, XU L S, WANG L, et al. Automatic coronary artery segmentation of CCTA images with an efficient feature-fusion-and-rectification 3D-UNet[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(8): 4044-4055.
- [24] TIAN F ZH, GAO Y B, FANG ZH J, et al. Automatic coronary artery segmentation algorithm based on deep learning and digital image processing [J]. Applied Intelligence, 2021, 51(12): 8881-8895.
- [25] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
- [26] WANG SH H, HU S Y, CHEAH E, et al. U-net using stacked dilated convolutions for medical image segmentation[J]. ArXiv preprint arXiv:2004.03466, 2020.
- [27] DONG C X, XU S H, DAI D W, et al. A novel multi-attention, multi-scale 3D deep network for coronary artery segmentation [J]. Medical Image Analysis, 2023, 85: 102745.
- [28] ZENG AN, WU CH B, LIN G S, et al. Imagecas: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images [J]. Computerized Medical Imaging and Graphics, 2023, 109: 102287.
- [29] ÇIÇEK Ö, ABDULKADIR A, LIENKAMP S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]. Medical Image Computing and Computer-Assisted Intervention-MICCAI the 19th International Conference, 2016: 424-432.
- [30] RICH A, STIER N, SEN P, et al. 3DVNet: Multi-view depth prediction and volumetric refinement[C]. 2021 International Conference on 3D Vision (3DV). IEEE, 2021: 700-709.
- [31] MYRONENKO A. 3D MRI brain tumor segmentation using autoencoder regularization [C]. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, 2019: 311-320.
- [32] HATAMIZADEH A, TANG Y CH, NATH V, et al. UNETR: Transformers for 3d medical image segmentation[C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2022: 574-584.
- [33] HATAMIZADEH A, NATH V, TANG Y C, et al. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in mri images [C]. International MICCAI brainlesion workshop. Cham: Springer International Publishing, 2021: 272-284.
- [34] 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[J]. ArXiv preprint arXiv:2209.15076, 2022.

作者简介

潘成龙, 硕士研究生, 主要研究方向为计算机视觉、医学图像处理。

E-mail: pclsolo@163.com

刘立程(通信作者), 副教授, 博士, 主要研究方向为人工智能在图像处理方面的技术研究与应用。

E-mail: celcliu@gdut.edu.cn

潘丹, 副教授, 博士, 高级工程师, 主要研究方向为人工智能、医学图像处理。

E-mail: pandan@gpnu.edu.cn