

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2025)10-3242-13

论文引用格式: Liu J M and Tang Y C. 2025. Two-stage coronary artery segmentation via multidirectional snake convolution and vision Mamba. Journal of Image and Graphics, 30(10):3242-3254(刘建明, 唐煜城. 2025. 采用多方向蛇形卷积和视觉残差 Mamba 的两阶段冠状动脉分割. 中国图象图形学报, 30(10):3242-3254)[DOI:10.11834/jig.240538]

采用多方向蛇形卷积和视觉残差 Mamba 的两阶段冠状动脉分割

刘建明, 唐煜城*

江西师范大学计算机信息工程学院, 南昌 330022

摘要: 目的 基于计算机断层扫描血管造影(computed tomography angiography, CTA)图像的冠状动脉分割具有重要的临床应用价值。冠状动脉具有多分支和细分支的管状结构特点,同时面临前景与背景类别严重不平衡的问题。传统基于卷积神经网络(convolutional neural network, CNN)的冠状动脉分割网络难以建模血管间的长依赖关系,而基于视觉转换器(vision Transformer, ViT)的模型由于复杂度过高,在资源有限的现实环境中难以部署。最新研究表明,以 Mamba 为代表的状态空间模型(state space model)能够在保持线性复杂度的情况下有效建模长依赖关系。**方法** 提出一种采用多方向蛇形卷积和视觉残差 Mamba 的两阶段冠状动脉分割方法——MDSVM-UNet++。采用传统的编码—解码架构:在编码阶段,为了准确捕捉血管细长而曲折的管状结构特征,提出新的多方向蛇形卷积模块,从矢状面、冠状面、水平面 3 个视角进行多视角融合学习,使模型能够更全面地自适应专注于冠状动脉细长局部结构;在解码阶段,为了建模血管切片间的长依赖关系同时保持线性复杂度,设计基于残差视觉 Mamba 的上采样解码器块,解码器块首先使用加法运算执行特征融合,随后将结果输入到残差视觉 Mamba 层中进行长依赖关系建模,最后通过三线性插值操作对特征图进行上采样。为了更准确地分割细分支血管,进一步提出两阶段分割模型:在第 1 阶段,采用 MDSVM-UNet++ 对整幅 CT 图像进行直接分割,结果用于指导原图像的分块;随后,将分块后的数据重新输入 MDSVM-UNet++ 进行第 2 阶段分割,最终合并所有分块的分割结果。在保证分割结果连续性的情况下,进一步减少了分割结果中的假阳性点,同时提高冠脉的连续性和平滑度。**结果** 实验结果表明,提出的两阶段 MDSVM-UNet++ 方法在 IMAGECAS 数据集上的 Dice 相似系数、豪斯多夫距离和平均豪斯多夫距离分别优于基准网络 ImageCAS 5.41%、8.545 6 和 0.809 3。**结论** 本文提出一种采用多方向蛇形卷积和视觉残差 Mamba 的两阶段冠状动脉分割方法:提出的多方向蛇形卷积可以更全面更准确地捕捉血管结构特征;设计的基于残差视觉 Mamba 的解码器模块,在线性复杂度下实现了血管切片间的长依赖关系的建模,最终实现低资源环境下更准确的冠状动脉分割。

关键词: 计算机断层扫描血管造影(CTA);冠状动脉分割;两阶段分割方法;动态蛇形卷积;状态空间模型(SSM)

Two-stage coronary artery segmentation via multidirectional snake convolution and vision Mamba

Liu Jianming, Tang Yucheng*

College of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China

Abstract: Objective Cardiovascular disease (CVD) accounts for approximately half of noncommunicable diseases. Coro-

收稿日期: 2024-09-06;修回日期: 2024-12-13;预印本日期: 2024-12-20

* 通信作者: 唐煜城 879256489@qq.com

基金项目: 国家自然科学基金项目(62266022);江西省自然科学基金项目(20242BAB25110)

Supported by: National Natural Science Foundation of China (62266022); Natural Science Foundation of Jiangxi Province, China (20242BAB25110)

nary artery stenosis is considered a major risk factor for CVD. Computed tomography angiography (CTA) has become one of the widely used noninvasive imaging methods for coronary diagnosis due to its excellent image resolution. Clinically, coronary artery segmentation is crucial for the diagnosis and quantification of coronary artery diseases. Coronary arteries are characterized by their multibranching and tubular structures, with severe imbalance between foreground and background classes. Traditional coronary artery segmentation networks based on convolutional neural networks experience difficulty in modeling long-range dependencies between vessels, while Vision Transformer models are difficult to deploy in resource-constrained real-world environments due to their high complexity. Recent studies have shown that state space models, such as Mamba, can effectively model long-range dependencies while maintaining linear complexity. **Method** Given these reasons, this study applies visual Mamba to coronary artery segmentation for the first time and proposes a two-stage coronary artery segmentation method, called MDSVM-UNet++, based on multidirectional snake convolution (MDSCnv) and visual residual Mamba. MDSVM-UNet++ adopts a traditional encoder-decoder architecture. During the encoding stage, dynamic snake convolution is used to replace traditional convolution to accurately capture the elongated and tortuous tubular structure features of the vessels, introducing deformable offsets into the convolution kernel and employing an iterative strategy to prevent the model from deviating from the target while learning the deformable offsets, ensuring the continuity of attention. In addition, an MDSCnv layer is proposed to extract features from the three dimensions (x, y, z) of a 3D image, retaining attention from multiple directions and further improving segmentation accuracy, allowing the model to focus more on slender and complex tubular structures. During the decoding stage, an up-sampling decoder block based on residual visual Mamba is designed to model long-range dependencies between vascular slices while maintaining linear complexity. This block employs dense spatial pooling techniques to generate fine multi-scale contexts, first performing feature fusion by using addition and then inputting the results into the residual visual Mamba layer for long-range dependency modeling, followed by trilinear interpolation for up-sampling the feature maps. To achieve more accurate segmentation of small branch vessels, a two-stage segmentation model is further proposed. During the first stage, the original CTA images are scaled down to a size of $128 \times 128 \times 64$, and then MDSVM-UNet++ is directly applied to segment the entire CTA image. The results are used to guide the partitioning of the original image into a set of $64 \times 64 \times 64$ voxel blocks, allowing the data to contain more coronary artery information. Subsequently, the segmented data are re-inputted into the MDSVM-UNet++ network for the second stage of segmentation, with all the segmented results merged at the end. This approach reduces false positive points in the segmentation results while ensuring continuity and improving the smoothness of coronary arteries. **Result** In the experimental section, we implemented the model by using the PyTorch framework and trained it on NVIDIA GTX 3090. We selected 750 CTA images from the ImageCAS dataset as the training set, with the remaining 250 CTA images used as validation and test sets. The first-stage segmentation network was trained for 25 epochs by using the Adam optimizer with a learning rate of 0.001. The second-stage segmentation network underwent 50 iterations, with the learning rate decayed by a factor of 0.1 at the 30th and 40th iterations. The Dice similarity coefficient (DSC) was used as the evaluation metric, while the Hausdorff distance (HD) and average HD (AHD) served as auxiliary metrics. The experimental results indicated that the proposed MDSVM-UNet++ method outperformed the latest baseline network, ImageCas, on the ImageCAS dataset, achieving 5.41% improvement in DSC, an increase of 8.545 6 in HD, and an increase of 0.809 3 in AHD. **Conclusion** Given the tubular structural characteristics of coronary arteries, this study applies visual Mamba to coronary artery segmentation for the first time, proposing a two-stage coronary artery segmentation method based on dynamic snake convolution and visual residual Mamba. On the one hand, dynamic snake convolution is utilized to more accurately capture vascular structural features. On the other hand, a decoding module based on visual residual Mamba is designed to model long-range dependencies between vascular slices while maintaining linear complexity, ultimately achieving more accurate coronary artery segmentation in resource-limited environments.

Key words: computed tomography angiography (CTA); coronary artery segmentation; two-stage split method; dynamic snake convolution; state space model(SSM)

0 引言

心血管疾病(cardiovascular disease, CVD)是全球非传染性疾病中的头号杀手之一,已成为当今世界面临的主要健康挑战。在临床上,计算机断层扫描血管造影(computed tomography angiography, CTA)广泛用于冠状动脉疾病的诊断和治疗计划,因为它是无创的,并且能够提供高分辨率的3D成像。影像科医生获得CTA图像后首先定位冠状动脉并划分其边界,然后提取并量化狭窄部分,用于最终的诊断。然而,这种基于手动操作的过程既耗时又容易出错,错误标注的医学图像可能显著影响手动分割的成本和可重复性。更糟糕的是,随着医学图像的数量和种类持续增加,手动分割不仅面临着高昂的成本,还存在可重复性差的弊端,这使得该方式几乎完全不可行(Li等,2018;楼陆飞等,2024)。因此,冠状动脉的自动分割对于冠状动脉疾病的诊断是非常必要的。

近年来,关于血管分割的研究成果显著增加,包括基于传统机器学习的方法(Chen等,2013;尚岩峰等,2013;Zhou等,2010)和基于深度学习的方法(Ronneberger等,2015;Çiçek等,2016;Isensee等,2021;Dong等,2023)。传统机器学习方法以基于规则的技术为主,例如基于区域增长的方法(Chen等,2013)、主动轮廓模型(尚岩峰等,2013)、跟踪算法(Zhou等,2010)等,这些传统方法高度依赖于专家设计的特定领域特征表示,且非常耗时。随着深度学习的快速发展,更多基于深度学习的方法相继提出,并展示出更好的可扩展性和更高的准确性。其中,以基于U-Net架构(Ronneberger等,2015)及其变体的方法最为突出。基于U-Net架构,研究人员开发了多种新方法保留低级细节以精确获取对象边界,例如UNet++(Zhou等,2018)。为了有效捕获跨切片的上下文信息,研究人员将3D卷积神经网络(convolutional neural network, CNN)引入编码器—解码器架构,例如3D U-Net(Çiçek等,2016)、nnUnet(no new net)(Isensee等,2021)、CAS-Net(coronary artery segmentation network)(Dong等,2023)。Zeng等人(2023)提出了结合多尺度块融合和两阶段处理来提取血管细节的方法。此外,基于Transformer

的方法(Hatamizadeh等,2022;Cao等,2022;许聪等,2023;楼陆飞等,2024)在血管分割任务中日益受到关注,因为它们能够捕捉长距离依赖关系并展现出优异的性能。

尽管这些U型结构取得了显著性能,但仍不足以应对冠状动脉分割的挑战(Gharleghi等,2022;Qi等,2023)。1)冠状动脉血管呈细长且多分枝的管状结构,传统卷积没有融合血管几何拓扑结构等领域知识,难以有效提取其管状特征;2)现有U型结构网络无法有效建模血管切片间的长距离依赖关系,而基于Transformer的模型复杂度过高;3)每个阶段的常规跳跃连接通常直接压缩局部信息,这会引入大量不相关的背景噪声,导致冠状动脉与周围组织难以区分。

为了解决上述问题,本文提出一种采用多方向蛇形卷积和残差视觉Mamba的两阶段冠状动脉分割方法。在第1阶段,模型对放缩后的图像进行粗分割,然后通过第1阶段分割结果指导块提取,提取成 $64 \times 64 \times 64$ 大小的体块集,使数据中包含更多的冠脉细节信息,用于作为下一阶段网络的输入数据。在第2阶段,对上一阶段提取得到的块数据进行块分割,最后将分块推理结果融合,得到最终的输出。

本文主要贡献如下:1)将视觉Mamba应用于冠状动脉分割,提出一种采用多方向蛇形卷积和视觉Mamba的冠状动脉分割模型(multi-direction snake vision Mamba UNet++, MDSVM-UNet++):一方面,在编码器部分提出一种新的多方向蛇形卷积模块(multi-direction snake convolution, MDSCConv),从矢状面、冠状面和水平面3个视角进行多视角融合学习,有效捕捉动脉血管管状结构特征;另一方面,在解码器中融入残差视觉Mamba层(residual vision Mamba, RVM),采用密集空间池技术产生更细粒度的多尺度上下文,同时利用Mamba的选择性特征在多尺度特征融合中有效减少同质语义信息,实现了线性复杂度下血管切片间长依赖关系的建模,进一步提高模型的性能。2)提出一种新的逐步细化的两阶段分割方法,在有效解决原始分辨率三维医学图像数据量过大问题的同时,保留了冠状动脉的细节信息,进一步减少了分割结果中的假阳性点,提高冠脉的连续性和平滑度。3)与SOTA(state-of-the-art)方法进行对比,本文方法在更少参数量下取得了更好的性能。

1 相关工作

1.1 冠状动脉分割

血管分割尤其是冠状动脉分割在医学图像分割中起着至关重要的作用。随着深度学习在医学图像分割中的广泛应用,基于深度学习的冠状动脉分割也越来越占据研究主流。Lei 等人(2020)提出一种使用 3D 注意力全卷积网络(fully convolutional network, FCN)(Shelhamer 等, 2017)的方法来自动分割冠状动脉,将注意力机制集成到 FCN 模型中,以突出 CT 冠状动脉成像(computed tomography coronary angiography, CTCA)图像中提取的语义信息特征。尽管将放缩后的原始图像直接输入网络可以获得分割结果,但对调整大小后的图像进行直接分割的插值操作会导致像素级别的信息损失。Huang 等人(2018)将原始 CTCA 图像分割为多个小体块,输入到 3D U-Net 卷积神经网络中进行冠状动脉管腔分割。然而,由于输入的体块较小,网络可能会忽略局部细节,难以学习到全面的特征。Kong 等人(2020)针对冠状动脉的树状结构,提出一种新的树状卷积门控递归单元(tree-structured ConvGRU, TS ConvGRU)来学习冠状动脉的解剖结构。他们将 TS ConvGRU 集成到编码器—解码器结构中,首先使用 3D U-Net 进行粗分割,提取粗分割结果的中心线,然后应用带有 TS ConvGRU 层的 3D U-Net 进行细分割。Wolterink 等人(2019)提出一种使用图卷积神经网络(graph convolutional network, GCN)的方法,预测管状表面网格中顶点的空间位置,自动提取包围整个冠状动脉树的表面网格。在 GCN 中引入网格信息可以改善基线模型的分割重叠和准确性,而无需进行网格上的交互。Dong 等人(2023)提出一种用于冠状动脉分割的新型多注意力、多尺度 3D 深度网络——CAS-Net。该方法设计了一个注意力引导的特征融合模块来融合相邻分层特征,并通过尺度感知特征模块动态调整感受野,以有效提取更具表现力的特征。最后,利用多尺度特征聚合模块来学习更丰富的语义表示,从而进一步细化血管结构的分割。Zeng 等人(2023)提出一个用于在 CTA 图像上进行冠状动脉分割的大规模数据集 ImageCAS,同时提出一种强大的基线方法,该方法结合了多尺度块融合和两阶段处理来提取血管的细节。还有一

类冠状动脉分割方法是在基于深度学习方法的分割基础上进行后处理。Qiu 等人(2023)提出一种保留拓扑结构的全连接冠状动脉提取方法——CorSeg-Rec,将图像分割、中心线重新连接和几何重建有机结合在一起。Wang 等人(2024)则提出一种基于数据驱动策略的创新细化方法,能够有效修复断裂的动脉结构。这些方法均通过在分割结果上进行后处理来提升分割精度。本文提出的方法可以作为这类后处理方法的预分割步骤,结合其后处理策略进一步提高分割效果。

1.2 可变形卷积

Yu 和 Koltun(2016)提出空洞卷积方法,能够系统地聚合多尺度上下文信息,在增加感受野的同时保持特征图尺寸不变,提高语义分割系统的准确性,从而替代上采样和下采样操作,减少信息损失。Dai 等人(2017)指出,CNN 构建模块采用固定的几何结构,这在处理模型几何变换时存在固有局限性。为此,他们提出可变形卷积(deformable convolution)和可变形感兴趣区域池化(deformable ROI pooling),通过在模块中引入额外的偏移量,并从目标任务中学习这些偏移量,无需额外监督即可实现当前位置附近的任意采样。尽管可变形卷积能够适应不同目标的几何变形,但在细长、复杂的管状结构的精确提取上仍面临挑战。Qi 等人(2023)注意到管状结构的特殊性,提出一种动态蛇形卷积方法(dynamic snake convolution, DSConv),能够捕捉管状结构特征,自适应地专注于细长和曲折的局部结构。该方法在卷积核中引入变形偏移量,同时为防止模型在自由学习变形时发生偏移,提出一种迭代策略,为每个待处理目标依次选择下一个待观察目标,确保注意力的连续性,从而提升管状结构分割的准确性。

1.3 状态空间模型

以 Mamba 为代表的状态空间模型(state space model, SSM)已成为传统卷积神经网络(CNN)和视觉 Transformer(vision Transformer, ViT)的强大竞争对手。3D 医学图像的高分辨率会导致基于 ViT 方法的计算负担显著增加,并降低速度性能。为了解决长序列建模的挑战,Gu 和 Dao(2023)提出 Mamba,通过选择机制和硬件感知算法来建模长距离依赖关系,从而提高训练和推理的效率。许多研究已探索了 Mamba 在计算机视觉中的应用。Liu 等人(2024a)提出一种基于 Mamba 的新型模型 Swin-UMamba,其

专为医学图像分割任务而设计,利用了基于 ImageNet 的预训练的优势。Liu 等人(2024b)提出 Vision Mamba,并引入了 ViM 块。ViM 块包含用于数据相关全局视觉上下文建模的双向 SSM 和用于位置感知视觉理解的位置嵌入。Liao 等人(2024)提出一种纯 Mamba 的方法,利用残差视觉 Mamba 层在线性计算复杂度下提取深层语义特征,并对长距离空间依赖性进行建模,从而以更少的参数实现更好的性能。Wu 等人(2024)提出并行视觉 Mamba 块,该块在保持处理通道总数恒定的同时,将其划分为 4 个并行的残差视觉 Mamba,以优化远程空间信息的获取能力并降低计算负载。Zhang 等人(2024)提出 Vision Mamba-UNetV2,引入了视觉状态空间(vision state space, VSS)块来捕获广泛的上下文信息,引入了语义和细节融合来增强低级和高级特征的融合。然而,上述方法主要针对通用医学图像分割问题,未考虑冠状动脉的细长管状特征。

2 本文方法

目前,基于 CTA 图像的冠状动脉分割方法主要

分为直接分割和块分割两类。直接分割方法使用单个神经网络,操作简单。3D 图像直接输入网络可以获得更大的感受野,且输出的概率图无需复杂处理。然而,相较于块分割,直接分割输入的图像通常较大,受限于显存,实践中需要将图像缩小到较低分辨率。块分割方法以三维体块为输入,通常将完整的体数据切分为多个小体块(如 $64 \times 64 \times 64$)输入网络。这导致感受野较小,更易产生假阳性预测。基于直接分割和块分割方法的特点,本文提出一种新的两阶段分割算法,整体框架如图 1 所示。在第 1 阶段,模型进行冠脉的整体结构学习和重建。学习完成后,对所有数据集进行推理,输出结果指导原图像进行块提取,生成第 2 阶段所需的训练集、验证集和测试集。第 2 阶段,使用提取的多块数据进行冠脉血管的特征学习,在较小的邻域内学习更多局部细节,减少和完善血管边缘细节的缺失问题,从而实现更好的分割效果。与现有两阶段动态分割方法(吴春彪,2022)相比,本文方法只将第 1 阶段结果作为下一阶段分割的依据,而不直接采用第 1 阶段的结果,减少了第 1 阶段粗分割不准确导致的假阳性点。

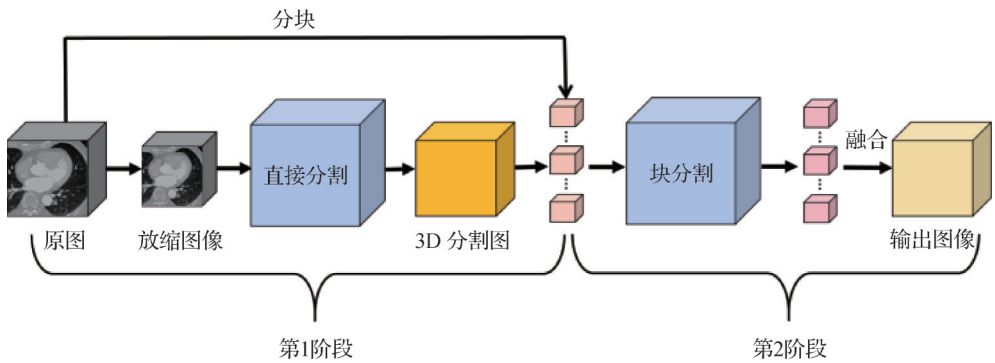


图 1 本文提出的两阶段冠脉分割框架

Fig. 1 The architecture of the proposed two-stage coronary artery segmentation method

2.1 网络框架

本文提出的 MDSVM-UNet++ 网络结构如图 2 所示。给定输入图像,其中 C, H, W 和 D 分别表示 3D 医学图像的通道数、高度、宽度和切片数。MDSVM-UNet++ 共有 5 层结构(编码器、解码器和跳跃连接路径),5 层结构中的通道数设置为 $[16, 32, 64, 128, 256]$ 。MDSVM-UNet++ 在编码器部分采用多方向蛇形卷积替换传统卷积操作,首先经过 4 层 MDSCov 层,最后再经过一个 RVM(residual

vision Mamba)以建模血管切片的长距离依赖关系。每个编码器块特征图中的通道数加倍,而图像分辨率减半。在此之后,采用 3 个 RVM 和 1 个 $3 \times 3 \times 3$ 的卷积层构成解码器,用于特征解码和图像的分辨率恢复。在每个解码器块之后,特征图中的通道数量减半,分辨率加倍。最后,通过一个转置 CNN 解码器块输出与原始图像相同的分辨率的分割图像。跳跃连接路径利用卷积模块(CNN block)形成 UNet++ 架构的密集连接,每一层输出特征和解码器

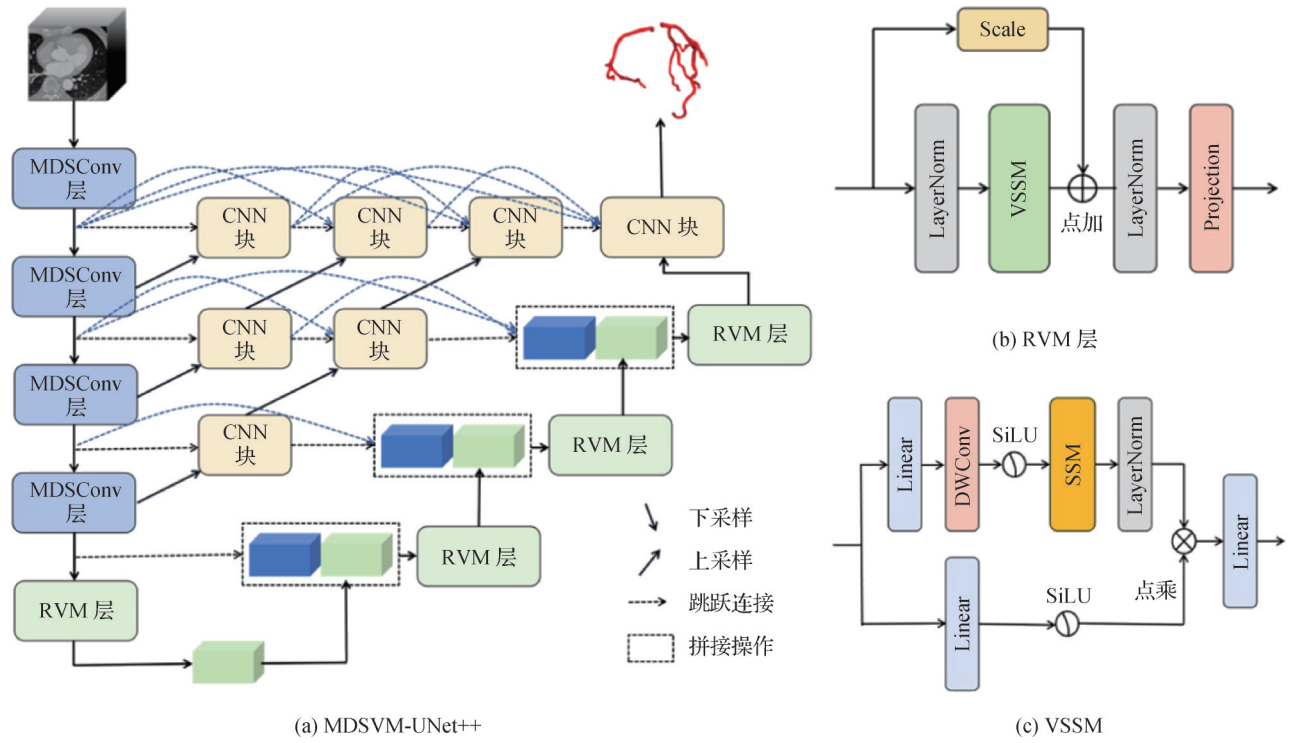


图2 网络框架

Fig. 2 Network framework ((a) MDSVM-UNet++; (b) RVM layer; (c) VSSM)

前一层上采样后特征在通道上融合作为当前 RVM 的输入。

2.2 编码器

2.2.1 多方向蛇形卷积

假设给定标准 3D 卷积核坐标为 $\mathbf{K}$, 其中心坐标为 $\mathbf{K}_i = (x_i, y_i, z_i)$, 一个膨胀指数为 1 的 $3 \times 3 \times 3$ 的卷积核 $\mathbf{K}$ 表示为

$$\mathbf{K} = \{(x-1, y-1, z-1), (x-1, y-1, z), \dots, (x+1, y+1, z+1)\} \quad (1)$$

蛇形卷积 (Qi 等, 2023) 主要是在三维卷积核中引入变形偏移量 Δ , 让模型能够自由学习变形偏移, 同时采用迭代策略, 为每个待处理的目标依次选择待观察的位置, 从而确保注意力的连续性, 并且不会由于大的变形偏移而将感受场扩散得太远。以 x 轴为例, 每个网格在 $\mathbf{K}$ 中的表示为 $\mathbf{K}_{i \pm c} = (x_{i \pm c}, y_{i \pm c}, z_{i \pm c})$, 其中, $c = \{0, 1, 2, 3, 4\}$ 。迭代策略是从中心网格开始, 向两端逐步迭代, 远离中心网格的位置取决于前一网格的位置。 $\mathbf{K}_{i+1}$ 相对于 $\mathbf{K}_i$ 被增大的变形偏移量 Δ 的大小范围为 $[-1, 1]$ 。

迭代策略中, x 轴、 y 轴和 z 轴方向上网格迭代计算为

$$\begin{aligned} k_{i \pm c} &= \begin{cases} (x_{i+c}, y_{i+c}, z_{i+c}) = \left(x_i + c, y_i + \sum_i \Delta y, z_i + \sum_i \Delta z \right) \\ (x_{i-c}, y_{i-c}, z_{i-c}) = \left(x_i - c, y_i + \sum_{i-c}^i \Delta y, z_i + \sum_{i-c}^i \Delta z \right) \end{cases} \quad (2) \\ k_{j \pm c} &= \begin{cases} (x_{j+c}, y_{j+c}, z_{j+c}) = \left(x_j + \sum_j^{j+c} \Delta x, y_j + c, z_j + \sum_j^{j+c} \Delta z \right) \\ (x_{j-c}, y_{j-c}, z_{j-c}) = \left(x_j + \sum_{j-c}^j \Delta x, y_j - c, z_j + \sum_{j-c}^j \Delta z \right) \end{cases} \quad (3) \\ k_{k \pm c} &= \begin{cases} (x_{k+c}, y_{k+c}, z_{k+c}) = \left(x_k + \sum_k^{k+c} \Delta x, y_k + \sum_k^{k+c} \Delta y, z_k + c \right) \\ (x_{k-c}, y_{k-c}, z_{k-c}) = \left(x_k + \sum_{k-c}^k \Delta x, y_k + \sum_{k-c}^k \Delta y, z_k - c \right) \end{cases} \quad (4) \end{aligned}$$

医学图像天然存在矢状面、冠状面和水平面 3 个视角, 三视角特征融合可以更好地观察目标的结构, 提取复杂冠脉结构中隐藏的关键特征。为了分别从矢状面、冠状面和水平面 3 个视角进行多视角融合学习, 本文提出多方向动态蛇形卷积模块 MDSCov, 其结构如图 3 所示。具体地, 对于输入特征 $\mathbf{F}$, 首先经

过普通卷积提取第 l 层的特征图 $f^l(\mathbf{F})$,然后分别从 x 轴、 y 轴和 z 轴提取来自第 l 层的特征图 $f^l(\mathbf{F}_x)$ 、 $f^l(\mathbf{F}_y)$ 和 $f^l(\mathbf{F}_z)$,将得到的4个特征图拼接起来再进行一次卷积操作得到该层最终的特征图 $\mathbf{X}^l$ 。上述过程的计算为

$$\mathbf{X}^l = f_{\text{ReLU}}\left(f_{\text{GN}}\left(\left[f^l(\mathbf{F}), f^l(\mathbf{F}_x), f^l(\mathbf{F}_y), f^l(\mathbf{F}_z)\right]\right)\right) \quad (5)$$

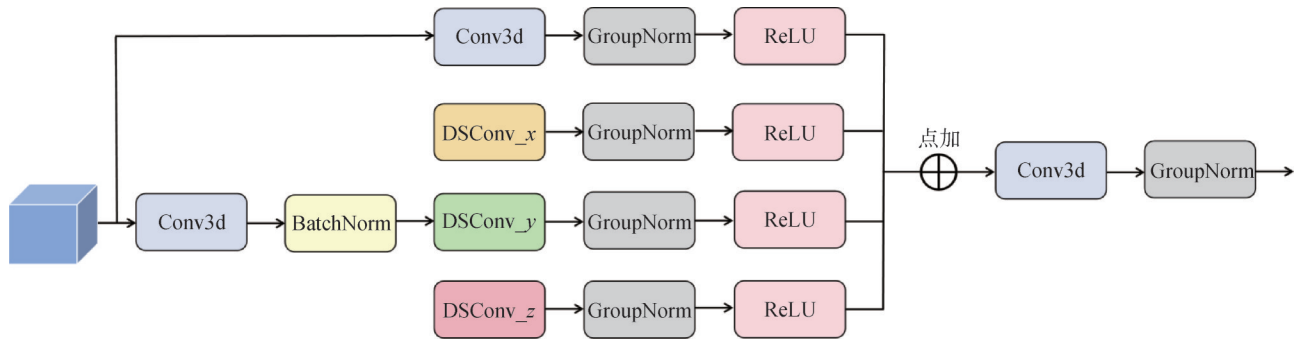


图3 多方向蛇形卷积层

Fig. 3 Multi-direction snake convolution layer

2.2.2 残差视觉 Mamba

给定特征图 $\mathbf{F}^l \in \mathbf{R}^{C \times H \times W \times D}$,其中, C 是通道数, H 是特征图高度, W 是特征图宽度, D 是特征图深度。首先将特征图平移并转置成 (L, C) 的形状,式中, $L = H \times W \times D$ 。随后,利用RVM层来捕获全局信息,之后将特征映射重新调整并转置为 $(C \times 2, H, W, D)$ 。具体而言,RVM利用残差连接和调整因子,在几乎不引入新参数和计算复杂度的情况下,进一步增强了SSM的远程空间建模能力。如图2(b)所示,给定输入深度特征 $\mathbf{x}_{\text{in}}^l \in \mathbf{R}^{L \times C}$,RVM层先进行层归一化(layer norm, LN),然后采用视觉状态空间模块(vision state-space module, VSSM)捕获空间长期依赖性;随后,它在残差连接中利用调整因子 $\mathbf{s} \in \mathbf{R}^C$ 以获得更好的性能;最后对输出特征重组为 (C, H, W, D) 形状得到 $\tilde{\mathbf{Y}}^l$ 。上述过程具体为

$$\tilde{\mathbf{Y}}^l = f_{\text{VSSM}}\left(f_{\text{LN}}(\mathbf{x}_{\text{in}}^l)\right) + \mathbf{s} \times \mathbf{x}_{\text{in}}^l \quad (6)$$

式中,LN表示层归一化,VSSM表示VSS模块。在此之后,RVM层采用另一LayerNorm来归一化 $\tilde{\mathbf{Y}}^l$,并且随后利用投影层来将 $\tilde{\mathbf{Y}}^l$ 转换为更深的特征 $\mathbf{Y}_{\text{out}}^l$ 。上述过程具体为

$$\mathbf{Y}_{\text{out}}^l = f_{\text{Projection}}\left(f_{\text{LN}}(\tilde{\mathbf{Y}}^l)\right) \quad (7)$$

式中,LN表示层归一化,Projection表示线性映射操作。

VSSM将层归一化后的特征 $\mathbf{W}_{\text{in}}^l \in \mathbf{R}^{L \times C}$ 作为输入,并将其引导到两个并行分支中,如图2(c)所示。

式中, $[\]$ 表示拼接操作;GN(group normalization)为组归一化(Wu和He,2018),ReLU(rectified linear unit)为非线性激活函数。三维视角特征融合模块将当前层的3个视角特征融合后传给其后的层,用于提取更加细致的特征信息。

在第1分支中,VSSM使用线性层将特征通道扩展到 $\lambda \times C$,其中, λ 表示预定义的通道扩展因子。随后,经过深度可分离卷积(depthwise separable convolution, DWConv)和SiLU(sigmoid linear unit)激活函数,然后是SSM和LN层,最终得到特征 $\mathbf{W}_1$ 。在第2分支中,首先使用线性层将特征通道扩展到 $\lambda \times C$,再经过一个SiLU激活函数得到特征 $\mathbf{W}_2$ 。最后,使用Hadamard乘积聚合来自两个分支的特征,并将通道号投影回 C 以生成具有与输入 $\mathbf{W}_{\text{in}}$ 相同形状的输出 $\mathbf{W}_{\text{out}}$ 。上述过程具体为

$$\mathbf{W}_1 = f_{\text{LN}}\left(f_{\text{SSM}}\left(f_{\text{SiLU}}\left(f_{\text{DWConv}}\left(f_{\text{Linear}}(\mathbf{W}_{\text{in}})\right)\right)\right)\right) \quad (8)$$

$$\mathbf{W}_2 = f_{\text{SiLU}}\left(f_{\text{Linear}}(\mathbf{W}_{\text{in}})\right) \quad (9)$$

$$\mathbf{W}_{\text{out}} = f_{\text{Linear}}(\mathbf{W}_1 \odot \mathbf{W}_2) \quad (10)$$

式中,Linear表示线性层,DWConv表示深度可分离卷积,SiLU表示激活函数,SSM表示结构化状态空间,LN表示层归一化, $\odot$ 表示Hadamard乘积。

2.3 解码器

MDSVM-UNet++使用上述RVM作为解码模块。解码器由3个RVM层和1个 $3 \times 3 \times 3$ 的卷积层构成,具体地,给定 $\mathbf{F}^l \in \mathbf{R}^{C \times H \times W \times D}$ 来自跳跃连接的输出, $\mathbf{P} \in \mathbf{R}^{C \times H \times W \times D}$ 是先前块得到的输出,解码器块首先使用加法运算执行特征融合,然后输入到RVM层中来解码特征图,随后通过三线性插值操作对特征图进行上采样。最后一个RVM层输出经过卷积

层得到最终输出 P_{out} 。上述过程具体为

$$\tilde{P} = f_{Up}(f_{RVM}(F^l + P)) \quad (11)$$

$$P_{out} = f_{Conv}(\tilde{P}) \quad (12)$$

式中, Up 是上采样操作, $\tilde{P}$ 是解码器每一层的输出, $Conv$ 是卷积操作。

2.4 跳跃连接

本文使用 Zhou 等人(2020)提出的跳跃连接方法, 每个解码器模块都与多个编码器模块进行连接, 这些连接具有不同的深度, 从而能够捕捉不同层次的语义信息。在跳跃连接中使用密集卷积块, 特征图经过密集连接块再被解码器接收。多跳跃连接机制增加了网络的复杂性, 同时提高了网络泛化能力, 减少过拟合的风险。

2.5 损失函数

考虑到冠状动脉分割任务存在严重的类别不平衡问题, 常用的交叉熵损失函数并不适用于这种情况。因此, 本文采用了适合解决前景背景严重不平衡问题的 Dice loss (Li 等, 2020), 以避免模型在分割过程中偏向于背景类别, 从而提高分割模型的性能和准确性。两个阶段的分割网络均采用 Dice loss 作为损失函数, 损失函数计算为

$$Dice = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (13)$$

$$L_{dice} = 1 - Dice \quad (14)$$

式中, A 和 B 分别表示真实标签值集合和 prediction 即预测值集合。

3 实验设计和性能评估

3.1 数据集

对于基于深度学习的冠脉分割方法来说, 具有高质量标注的大型数据集是成功的关键性因素, 但两个常用的数据集 (Kirişli 等, 2013; Schaap 等, 2009) 的数据非常少, 分别仅有 8 幅和 18 幅图像用于训练。此外, 大多数方法仅使用内部数据集进行评估, 它们的图像采集参数、重建技术和标记方法差异很大。因此, 本实验使用 IMAGECAS 数据集 (Zeng 等, 2023) 来评估模型的性能。该数据集是用于研究冠状动脉分割的大规模数据集, 由来自 1 000 名患者在西门子 128 层双源扫描仪捕获的 3D CTA 图像组成。进行高剂量 CTA 重建时, 选择 30%~40% 阶段或

60%~70% 阶段以获取最佳冠状动脉图像, 图像的尺寸为 $512 \times 512 \times (206 \sim 275)$ 体素, 平面分辨率为 $0.29 \sim 0.43 \text{ mm}^2$, 间距为 $0.25 \sim 0.45 \text{ mm}$ 。每幅图像中的左冠状动脉和右冠状动脉由两名放射科医生分别标注, 标注结果进行交叉验证。在标注不一致的情况下, 由第 3 名医生进行标注, 并且最终结果由 3 名医生的共识决定。

3.2 实验环境与配置

本实验都是基于 Python3. 10 和 PyTorch2. 1. 1-cuda11. 8 实现的。所有实验都使用了一个 NVIDIA 3090 GPU, 内存为 24 GB。实验中, 选择 IMAGECAS 数据集中 750 幅 CTA 图像作为训练集, 其余 250 幅 CTA 图像作为验证集和测试集。第 1 阶段的分割网络训练了 25 轮, 采用 Adam 优化器, 学习率为 0. 001, 第 2 阶段的分割网络训练了 50 轮, 并采用 Adam 优化器, 在第 30 和 40 次迭代以衰减率为 0. 1 的比率衰减学习率。

3.3 评估指标

对本文模型的性能评估主要采用 Dice 相似系数 (Dice similarity coefficient, DSC) 作为评价指标。Dice 相似系数主要用于衡量两个集合之间的相似程度, 取值范围在 0 和 1 之间, 并使用豪斯多夫距离 (Hausdorff distance, HD) (Huttenlocher 等, 1993) 和平均豪斯多夫距离 (average Hausdorff distance, AHD) 作为辅助指标。Dice 相似系数值越高说明分割效果越好, HD 和 AHD 越低说明分割效果越好。HD 距离计算为

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (15)$$

$$h(A, B) = \max_{a \in A} \left\{ \min_{b \in B} \|a - b\| \right\} \quad (16)$$

式中, A 表示图像真实像素值, B 表示模型预测图像的像素值, $h(A, B)$ 表示 A 到 B 的豪斯多夫距离, a, b 分别表示图像像素集合 A, B 中的单个像素。

3.4 性能评估

在本节中将所提方法与当前最先进的冠状动脉分割方法在 IMAGECAS 数据集上进行比较, 包括 FCN (Lei 等, 2020)、3D U-Net (Çiçek 等, 2016)、TS ConvGRU (tree-structured ConvGRU) (Kong 等, 2020)、GCN (Wolterink 等, 2019)、3D U-Net++ (Zhou 等, 2018)、DSU-Net (synamic anake convolution UNet) (Qi 等, 2023)、LightM-UNet (Liao 等, 2024)、UltraLight (Wu 等, 2024)、SwinUnet (Cao 等, 2022) 和 ImageCAS

(a large-scale dataset and benchmark for coronary artery segmentation)(Zeng等,2023)。

3.4.1 一阶段分割方法比较

本文使用ImageCAS模型作为基准模型,ImageCAS模型是一个两阶段冠状动脉分割模型,第1阶段使用3D U-Net作为直接分割模型;第2阶段使用3D U-Net++作为块分割模型。在表1中,可看出在第1阶段本文分割网络MDSVM-UNet++的Dice相似系数高于基准模型第1阶段的指标0.026 1,豪斯多夫距离和平均豪斯多夫距离分别优于基准模型第1阶段指标2.003 2和0.203 8。与其他方法进行对比,MDSVM-UNet++在Dice相似系数和平均豪斯多夫距离均取得了最好的性能。模型参数量方面,在解码器中融入残差视觉Mamba层,相对于Transformer,参数量大幅减少,模型参数量相较于DSUNet减少2.3 M。实验结果表明,Dice loss相对于Focal loss(Lin等,2017)在冠状动脉分割的效果更好。

表1 一阶段各种分割方法对比					
Table 1 Comparison of various segmentation methods in one stage					
方法	Loss	Dice	HD	AHD	Param/M
FCN	L_{dice}	0.575 5	42.803 1	1.688 5	20.0
	L_{focal}	0.557 3	37.095 8	1.761 9	
3D U-Net	L_{dice}	0.659 9	31.151 8	1.116 7	25.0
	L_{focal}	0.652 1	30.952 7	1.196 6	
3D U-Net++	L_{dice}	0.671 2	29.517 8	1.019 7	27.5
	L_{focal}	0.671 0	28.871 4	0.984 0	
DSU-Net	L_{dice}	0.675 6	28.368 4	0.943 4	29.0
	L_{focal}	0.664 3	35.462 4	0.971 1	
LightM-UNet	L_{dice}	0.653 0	34.818 1	1.001 7	20.2
	L_{focal}	0.644 6	46.492 7	1.286 4	
UltraLight	L_{dice}	0.574 5	50.959 7	1.443 8	0.238
	L_{focal}	0.567 2	67.151 2	1.347 5	
SwinUnet	L_{dice}	0.672 3	31.806 9	0.927 0	256.3
	L_{focal}	0.664 3	41.958 8	1.236 8	
MDSVM-UNet++ (本文)	L_{dice}	0.686 0	29.148 6	0.912 9	26.7
	L_{focal}	0.681 4	28.241 2	0.926 0	

注:加粗字体表示各列最优结果。

3.4.2 两阶段分割方法比较

两阶段不同网络分割结果如表2所示。相对于只使用一阶段分割方法,使用两阶段分割方法的Dice相似系数提升较大,与基准模型ImageCAS相比,Dice相似系数、豪斯多夫距离和平均豪斯多夫距离分别提升0.054 1、8.545 6和0.809 3。实验结果显示,两阶段分割方法比一阶段直接分割方法性能更优异,这是由于直接分割输入的图像较大,将图像缩小到较小的分辨率后输入到分割网络中容易丢失关键信息,并且对于冠状动脉这种前景和背景类严重不平衡、多分支、细分支的组织来说,单阶段网络尚不能达到很好的分割效果。在直接分割的基础上,使用MDSVM-UNet++进行二次块分割,提高分割结果的精确度。从实验结果上看,本文最终确定的

表2 两阶段各种分割方法对比
Table 2 Comparison of various segmentation methods in two stages

方法	Loss	Dice	HD	AHD	Param/M
3D U-Net	L_{dice}	0.720 1	40.969 3	3.068 6	23.3
	L_{focal}	0.718 3	32.772 4	4.468 7	
TS ConvGRU	L_{dice}	0.687 8	30.335 0	1.433 6	35.2
	L_{focal}	0.676 6	32.450 1	1.586 4	
GCN	L_{dice}	0.706 1	27.871 8	1.243 3	41.5
	L_{focal}	0.698 8	27.997 0	1.435 2	
DSU-Net	L_{dice}	0.779 0	30.152 4	0.909 2	29
	L_{focal}	0.772 0	36.439 0	1.717 1	
LightM-UNet	L_{dice}	0.807 9	28.186 4	0.948 8	5.5
	L_{focal}	0.784 8	28.515 5	1.005 4	
SwinUNet	L_{dice}	0.808 2	29.620 8	1.097 4	256.3
	L_{focal}	0.801 2	28.276 0	0.962 5	
两阶段FCN	L_{dice}	0.795 0	—	—	—
	L_{focal}	—	—	—	
ImageCAS (2024)	L_{dice}	0.782 4	36.388 6	1.711 6	27.6
	L_{focal}	0.774 8	36.174 0	1.7261	
MDSVM-UNet++ (本文)	L_{dice}	0.836 5	27.843 0	0.902 3	26.7
	L_{focal}	0.821 0	27.524 9	0.919 9	

注:加粗字体表示各列最优结果。“—”表示未做实验。

二阶段分割网络比一阶段 Dice 相似系数、豪斯多夫距离和平均豪斯多夫距离分别提升 0.150 5、1.305 6 和 0.010 6。两阶段全卷积神经网络论文未公布代码,且实验数据集不同,仅使用论文数据进行对比。

3.5 消融实验

为全面分析本文方法,以 ImageCAS 为基准模型在 IMAGECAS 数据集上进行消融实验,结果如表 3 所示。可以看出,引入 MDSCov 时 Dice 相似系数、豪斯多夫距离和平均豪斯多夫距离分别提升 0.041 7、8.837 1 和 0.807 0。引入 RVM 时 Dice 相似系数、豪斯多夫距离和平均豪斯多夫距离分别提升 0.041 2、8.780 4 和 0.792 4。在使用单个模块的基础上,同时采用 DSC 和 RVM 模块能够进一步提升性能,Dice 相似系数和平均豪斯多夫距离分别提升 0.054 1 和 0.809 3,表明 MDSCov 和 RVM 模块具有相互促进作用。

表 3 消融实验结果
Table 3 Ablation test results

方法	Dice	HD	AHD
Baseline	0.782 4	36.388 6	1.711 6
Baseline + MDSCov	0.824 1	27.551 5	0.904 6
Baseline + RVM	0.823 6	27.608 2	0.919 2
Baseline + MDSCov + RVM	0.836 5	27.843 0	0.902 3

注:加粗字体表示各列最优结果。

3.6 损失函数对比实验

Dice loss 和 Focal loss 是常用适合解决前景背景严重不平衡问题的两个损失函数,为了对比哪种损失函数针对冠状动脉效果更好,本文采用 Dice loss 和 Focal loss 针对对比算法和本文算法分别进行了实验,如表 1 和表 2 所示。可以看出,针对本文算法,Dice loss 平均效果更好。

3.7 分割结果可视化

将现有其他冠状动脉分割方法和所提 MDSVM-UNet++方法分割的结果进行可视化,结果如图 4(b)——(i)所示。通过对比分割结果的连续性情况,可以看出,所提 MDSVM-UNet++方法有效减少了分割结果中的假阳性点,分割效果比其他方法更好,分割连续性和平滑度也得到进一步提高。

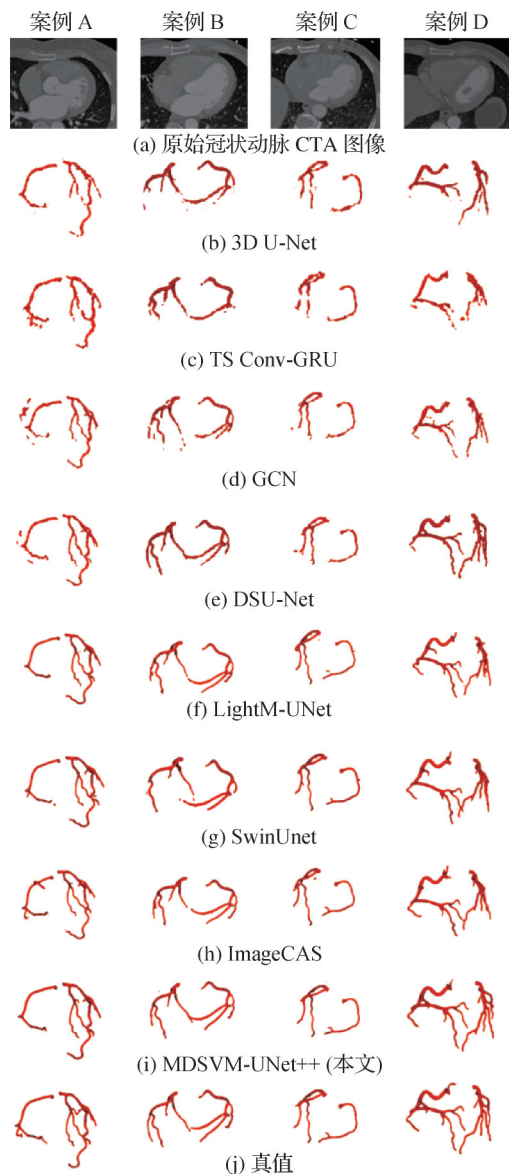


图 4 不同方法的可视化分割结果

Fig. 4 Visualization of segmentation results from different methods((a)original coronary artery CTA images;(b)3D U-Net;(c)TS Conv-GRU;(d)GCN;(e)DSU-Net;(f)LightM-UNet;(g)SwinUnet;(h)ImageCAS;(i)MDSVM-UNet++;(j)ground truth)

4 结论

从冠状动脉计算机断层扫描血管造影(CCTA)中准确提取冠状动脉是计算机辅助诊断冠状动脉疾病(CAD)的先决条件。现有基于 CNN 的冠状动脉分割方法对长距离依赖关系的建模能力有限,很难充分利用图像中的语义信息,而基于 Transformer 的方法复杂度太高。针对以上问题,本文提出一种采

用多方向蛇形卷积和残差视觉 Mamba 的两阶段冠状动脉分割方法。首先,设计了一种多方向动态蛇形卷积模块来自适应地关注细而曲折的血管结构,从而准确捕捉管状结构的特征;其次,提出基于视觉残差 Mamba 的解码器模块,在线性复杂度下实现了血管切片间的长依赖关系的建模;最后,提出一种两阶段逐步细化的分割框架,减少分割结果中的假阳性点,提升了分割冠脉的连续性和平滑度。在最新的公开数据集上进行了对比实验。实验结果表明,本文方法在 Dice 相似系数、豪斯多夫距离和平均豪斯多夫距离等多个指标上均获得了明显提升,在更少参数量情况下,获取了当前最好的效果。然而,本文方法相对于传统基于 CNN 的方法,参数量没有明显提升。未来,将保持相当性能下进一步优化模型参数量,同时将在其他模态的图像中检验所提出方法的性能。

参考文献 (References)

- Kong B, Wang X, Bai J J, Lu Y, Gao F, Cao K L, Xia J, Song Q and Yin Y B. 2020. Learning tree-structured representation for 3D coronary artery segmentation. *Computers in Biology and Medicine*, 80: #101688 [DOI: 10.1016/j.compmedimag.2019.101688]
- Cao H, Wang Y Y, Chen J, Jiang D S, Zhang X P, Tian Q and Wang M N. 2022. Swin-Unet: Unet-like pure transformer for medical image segmentation//*Proceedings of 2022 In European Conference on Computer Vision*. Tel Aviv, Israel: Springer: 205-218 [DOI: 10.1007/978-3-031-25066-8_9]
- Chen Y, Cao Q, Zhuang Z K, Yang Z, Luo L M and Toumoulin C. 2013. 3-D coronary vessel extraction using a novel minimum path based region growing//*Proceedings of the 10th International Conference on Image Analysis and Recognition*. Aveiro, Portugal: Springer: 502-509 [DOI: 10.1007/978-3-642-39094-4_57]
- Çiçek Ö, Abdulkadir A, Lienkamp S S, Brox T and Ronneberger O. 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation//*Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens, Greece: Springer: 424-432 [DOI: 10.1007/978-3-319-46723-8_49]
- Dai J F, Qi H Z, Xiong Y W, Li Y, Zhang G D, Hu H and Wei Y C. 2017. Deformable convolutional networks//*Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy: IEEE: 764-773 [DOI: 10.1109/ICCV.2017.89]
- Dong C X, Xu S H, Dai D W, Zhang Y C, Zhang C Y and Li Z F. 2023. A novel multi-attention, multi-scale 3D deep network for coronary artery segmentation. *Medical Image Analysis*, 85: #102745 [DOI: 10.1016/j.media.2023.102745]
- Gharleghi R, Chen N, Sowmya A and Beier S. 2022. Towards automated coronary artery segmentation: a systematic review. *Computer Methods and Programs in Biomedicine*, 225: #107015 [DOI: 10.1016/j.cmpb.2022.107015]
- Gu A and Dao T. 2023. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. [2024-09-06]. <https://arxiv.org/pdf/2312.00752.pdf>
- Hatamizadeh A, Tang Y C, Nath V, Yang D, Myronenko A, Landman B, Roth H R and Xu D G. 2022. UNETR: Transformers for 3d medical image segmentation//*Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE: 1748-1758 [DOI: 10.1109/WACV51458.2022.00181]
- Huang W M, Huang L, Lin Z P, Huang S, Chi Y L, Zhou J Y, Zhang J M, Tan R S and Zhong L. 2018. Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images//*Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Honolulu, USA: IEEE: 608-611 [DOI: 10.1109/EMBC.2018.8512328]
- Huttenlocher D P, Klanderman G A and Rucklidge W J. 1993. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9): 850-863 [DOI: 10.1109/34.232073]
- Isensee F, Jaeger P F, Kohl S A A, Petersen J and Maier-Hein K H. 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2): 203-211 [DOI: 10.1038/s41592-020-01008-z]
- Kirişli H A, Schaap M, Metz C T, Dharampal A S, Meijboom W B, Papadopoulos S L, Dedic A, Nieman K, de Graaf M A, Meijis M F L, Cramer M J, Broersen A, Cetin S, Eslami A, Flórez-Valencia L, Lor K L, Matuszewski B, Melki I, Mohr B, Öksüz I, Shahzad R, Wang C, Kitslaar P H, Unal G, Katouzian A, Orkisz M, Chen C M, Precioso F, Najman L, Masood S, Ünay D, van Vliet L, Moreno R, Goldenberg R, Vućini E, Krestin G P, Niessen W J and van Walsum T. 2013. Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. *Medical Image Analysis*, 17(8): 859-876 [DOI: 10.1016/j.media.2013.05.007]
- Lei Y, Guo B J, Fu Y B, Wang T H, Liu T, Curran W, Zhang L J and Yang X F. 2020. Automated coronary artery segmentation in coronary computed tomography angiography (CCTA) using deep learning neural networks//*Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*. Houston, USA: SPIE: #34 [DOI: 10.1117/12.2550368]
- Li X Y, Sun X F, Meng Y X, Liang J J, Wu F and Li J W. 2020. Dice loss for data-imbalanced NLP tasks//*Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

- Online: Association for Computational Linguistics: 465-476 [DOI: 10.18653/v1/2020.acl-main.45]
- Li Z Y, Zhang X F, Müller H and Zhang S T. 2018. Large-scale retrieval for medical image analytics: a comprehensive review. *Medical Image Analysis*, 43: 66-84 [DOI: 10.1016/j.media.2017.09.007]
- Liao W B, Zhu Y H, Wang X Y, Pan C W, Wang Y S and Ma L T. 2024. LightM-UNet: Mamba assists in lightweight UNet for medical image segmentation [EB/OL]. [2024-09-06]. <https://arxiv.org/pdf/2403.05246.pdf>
- Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2017. Focal loss for dense object detection//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 2999-3007 [DOI: 10.1109/ICCV.2017.324]
- Liu J R, Yang H, Zhou H Y, Xi Y, Yu L Q, Li C, Liang Y, Shi G M, Yu Y Z, Zhang S T, Zheng H R and Wang S S. 2024a. Swin-UMamba: Mamba-based UNet with ImageNet-based pretraining//*Proceedings of the 27th International Conference on Medical Image Computing and Computer Assisted Intervention*. Marrakesh, Morocco: Springer: 615-625 [DOI: 10.1007/978-3-031-72114-4_59]
- Liu Y, Tian Y J, Zhao Y Z, Yu H T, Xie L X, Wang Y W, Ye Q X, Jiao J B and Liu Y F. 2024b. VMamba: visual state space model. *Advances in Neural Information Processing Systems*, 2024, 37: 103031-103063 [DOI: 10.48550/arXiv.2401.10166]
- Lou L F, Ying J J, Cai K J and Xin Y. 2024. Review of various vessels and airway segmentation in medical imaging. *Journal of Image and Graphics*, 29(9): 2692-2715 (楼陆飞, 应俊杰, 蔡凯俊, 辛宇. 2024. 医学影像多血管和气道分割方法综述. *中国图象图形学报*, 29(9): 2692-2715) [DOI: 10.11834/jig.230240]
- Qi Y L, He Y T, Qi X M, Zhang Y and Yang G Y. 2023. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France: IEEE: 6047-6056 [DOI: 10.1109/ICCV51070.2023.00558]
- Qiu Y H, Li Z H, Wang Y N, Dong P, Wu D J, Yang X N, Hong Q Q and Shen D G. 2023. CorSegRec: a topology-preserving scheme for extracting fully-connected coronary arteries from CT angiography//*Proceedings of the 26th International Conference on Medical Image Computing and Computer Assisted Intervention*. Vancouver, Canada: Springer: 670-680 [DOI: 10.1007/978-3-031-43898-1_64]
- 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]
- Schaap M, Metz C T, van Walsum T, van der Giessen A G, Weustink A C, Mollet N R, Bauer C, Bogunović H, Castro C, Deng X, Dikici E, O'Donnell T, Frenay M, Friman O, Hoyos M H, Kitslaar P H, Krissian K, Kühnel C, Luengo-Oroz M A, Orkisz M, Smedby Ö, Styner M, Szymczak A, Tek H, Wang C L, Warfield S K, Zambal S, Zhang Y, Krestin G P and Niessen W J. 2009. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Medical Image Analysis*, 13(5): 701-714 [DOI: 10.1016/j.media.2009.06.003]
- Shang Y F, Wang H, Wang N, Chen J, Yao L P and Sun K. 2013. Three dimensional vessel extraction model based on tubular characters and active contour model. *Journal of image and graphics*, 18(3): 290-298 (尚岩峰, 汪辉, 汪宁, 陈杰, 姚丽萍, 孙琨. 2013. 管状特性和主动轮廓的 3 维血管自动提. *中国图象图形学报*, 18(3): 290-298) [DOI: 10.11834/jig.20130307]
- Shelhamer E, Long J and Darrell T. 2017. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (4): 640-651 [DOI: 10.1109/TPAMI.2016.2572683]
- Wang Y L, Hsu W C, Shi V, Lin G, Lin C T, Feng X and Bai H. 2024. Car-Dcros: a dataset and benchmark for enhancing cardiovascular artery segmentation through disconnected components repair and open curve snake//*Proceedings of the 27th International Conference on Medical Image Computing and Computer Assisted Intervention*. Marrakesh, Morocco: Springer: 179-189 [DOI: 10.1007/978-3-031-72378-0_17]
- Wolterink J M, Leiner T and Išgum I. 2019. Graph convolutional networks for coronary artery segmentation in cardiac CT angiography//*1st International Workshop on Graph Learning in Medical Imaging*. Shenzhen, China: Springer: 62-69 [DOI: 10.1007/978-3-030-35817-4_8]
- Wu C B. 2022. A study of coronary artery segmentation based on two-stage fully convolutional neural networks. *Computer Science and Application*, 12(4): 828-834 (吴春彪. 2022. 基于两阶段全卷积神经网络的冠状动脉分割研究. *计算机科学与应用*, 12(4): 828-834) [DOI: 10.12677/CSA.2022.124084]
- Wu R, Liu Y, Ning G, Liang P and Chang Q. 2024. UltraLight VM-UNet: parallel vision Mamba significantly reduces parameters for skin lesion segmentation. *Patterns*. [DOI: 10.1016/j.patter.2025.101298]
- Wu Y X and He K M. 2018. Group normalization//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01261-8_1]
- Xu C, Hao H Y, Wang Y, Ma Y H, Yan Q F, Chen B, Ma S D, Wang X G and Zhao Y T. 2023. Vessel segmentation of OCTA images based on latent vector alignment and Swin Transformer. *Journal of Image and Graphics*, 28(9): 2927-2939 (许聪, 郝华颖, 王阳, 马煜辉, 阎岐峰, 陈滨, 马韶东, 王效贵, 赵一天. 2023. 融合隐向量对齐和 Swin Transformer 的 OCTA 血管分割. *中国图象图形学报*, 28(9): 2927-2939) [DOI: 10.11834/jig.220482]

- Yu F and Koltun V. 2016. Multi-scale context aggregation by dilated convolutions//Proceedings of 2016 International Conference on Learning Representations. San Juan, Puerto Rico: OpenReview. net [DOI:10.48550/arXiv.1511.07122]
- Zeng A, Wu C B, Lin G S, Xie W, Hong J, Huang M P, Zhuang J, Bi S S, Pan D, Ullah N, Khan K N, Wang T C, Shi Y Y, Li X M and Xu X W. 2023. ImageCAS: a large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images. Computerized Medical Imaging and Graphics, 109: #102287 [DOI: 10.1016/j.compmedimag.2023.102287]
- Zhang M Y, Yu Y, Jin S, Gu L M, Ling T S and Tao X P. 2024. VM-UNET-V2: rethinking vision Mamba UNet for medical image segmentation//20th International Symposium on Bioinformatics Research and Applications. Kunming, China: Springer: 335-346 [DOI: 10.1007/978-981-97-5128-0_27]
- Zhou S J, Yang J, Wang Y T and Chen W F. 2010. Automatic segmentation of coronary angiograms based on fuzzy inferring and probabilistic tracking. BioMedical Engineering OnLine, 9: #40 [DOI: 10.1186/1475-925X-9-40]
- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-Net architecture for medical image segmentation//4th International Workshop 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]
- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2020. UNet++: redesigning skip connections to exploit multiscale features in image segmentation [EB/OL]. [2024-09-06]. <https://arxiv.org/pdf/1912.05074.pdf>

作者简介

刘建明,男,副教授,主要研究方向为医学图像处理、零样本和小样本学习、细粒度视觉识别。

E-mail:liujianming@jxnu.edu.cn

唐煜城,通信作者,男,硕士,主要研究方向为计算机视觉。

E-mail:879256489@qq.com