Modeling and Simulation 建模与仿真, 2024, 13(2), 1577-1591
Published Online March 2024 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2024.132149

Hans汉斯

基于Transformer和双注意力机制的颈动脉超声造影图像斑块分割方法

王金生*，孙占全

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2023年12月27日；录用日期：2024年3月20日；发布日期：2024年3月27日

摘　要

脑卒中是一项重大的公共卫生挑战，同时也是全球导致死亡人数最多的疾病之一。颈动脉粥样硬化斑块与脑卒中等缺血性疾病密切相关。颈动脉斑块的早发现和早治疗对预防未来缺血性脑卒中疾病的发生具有重要意义。超声造影(CEUS)已经成为常见的诊断颈动脉斑块的成像方式，因此从CEUS图像中准确分割动脉粥样硬化斑块对于预防和治疗缺血性脑卒中至关重要。然而，由于斑块边界模糊和图像噪声强烈等原因，CEUS图像颈动脉斑块自动分割面临巨大挑战。因此，如何提高颈动脉斑块分割性能仍然是迫切需要解决的问题。本文提出了一种创新的医学图像分割框架，称为DATU-Net，该框架将Swin Transformer模块和双注意力机制集成到U形架构中，以实现CEUS图像颈动脉斑块自动分割。DATU-Net采用基于Swin Transformer模块构建的编码器，可以有效地建模远程依赖关系和多尺度上下文信息。为了获得更丰富的特征表示，我们在编码器–解码器的跳跃连接中引入了双级注意力(Dual-Level Attention)模块，以增强图像特定的位置特征和通道特征，从而有效提高了斑块分割性能。此外，我们在解码器中引入了Swin Transformer模块，用于进一步探索上采样过程中的全局上下文信息，同时逐步恢复特征图。我们利用实际的临床数据集对提出的框架性能进行了评估。广泛的实验仿真结果显示，本文提出的方法在Dice系数(0.8548)、交并比(0.7632)、精确度(0.8746)和召回率(0.8863)等方面始终优于其他分割网络。这些实验证明了DATU-Net的有效性，为颈动脉超声造影图像斑块自动分割问题提供了一种可行的解决方案。

关键词

颈动脉斑块，超声造影，医学图像分割，Swin Transformer，双注意力机制

Segmentation of Carotid Plaque in Contrast-Enhanced Ultrasound Image Based on Transformer and Dual Attention Mechanism

Jinsheng Wang*, Zhanquan Sun

*通讯作者。

文章引用: 王金生, 孙占全. 基于 Transformer 和双注意力机制的颈动脉超声造影图像斑块分割方法[J]. 建模与仿真, 2024, 13(2): 1577-1591. DOI: 10.12677/mos.2024.132149

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Dec. 27th, 2023; accepted: Mar. 20th, 2024; published: Mar. 27th, 2024

Abstract

Stroke is a major public health challenge and one of the leading causes of death around the world. Carotid atherosclerotic plaque closely correlates with ischemic diseases such as stroke. Early detection and treatment of carotid plaques are important for preventing future ischemic stroke diseases. Contrast-enhanced ultrasound (CEUS) has emerged as a prevalent imaging modality for the diagnosis of carotid plaques, so accurate segmentation of atherosclerotic plaques from CEUS images is crucial for the prevention and treatment of ischemic stroke. However, automatic segmentation of carotid plaques from CEUS images is tremendously challenging due to blurred plaque boundaries and strong noise in images. Therefore, how to improve the performance of carotid plaque segmentation remains an urgent problem. In this paper, we propose an innovative medical image segmentation framework, called DATU-Net, which aims to integrate the Swin Transformer block and the dual-attention mechanism into a U-shaped architecture for automatic carotid plaque segmentation of CEUS images. DATU-Net employs an encoder built upon the Swin Transformer block, proficient in effectively modelling long-range dependencies and multi-scale contextual information. In order to obtain richer feature representations, we introduce a Dual-Level Attention module in the encoder-decoder skip connection to enhance the image-specific positional and channel features, which effectively improves the performance of plaque segmentation. In addition, we introduce the Swin Transformer block in the decoder for further exploring the global contextual information during the up-sampling process while progressively recovering feature maps. We evaluate the performance of the proposed framework using real clinical datasets. Extensive experimental simulation results consistently show that the method proposed in this paper outperforms other segmentation networks in terms of Dice coefficient (0.8548), intersection over union (0.7632), precision (0.8746) and recall (0.8863). These experiments demonstrate the effectiveness of DATU-Net and provide a viable solution to the problem of automatic plaque segmentation in carotid CEUS images.

Keywords

Carotid Plaque, Contrast-Enhanced Ultrasound, Medical Image Segmentation, Swin Transformer, Dual Attention Mechanism

1. 引言

脑卒中是最常见的脑血管疾病，具有较高的发病率和死亡率，仍然是全世界第二大死因[1] [2]。大多数脑卒中病例是缺血性脑卒中[3]，因此，对于减少中风事件发生问题，精确监测和评估颈动脉斑块在临床上具有重要意义。目前，有多种医学成像方式可用于可视化颈动脉斑块，包括 X 射线、计算机断层扫描(Computed Tomography, CT)、磁共振成像(Magnetic Resonance Imaging, MRI)和常规超声(Ultrasound, US)。在这些成像技术中，常规超声具有无创、便捷、无辐射和价格实惠等特点[4]，因此被广泛用于颈

动脉斑块诊断。超声造影技术(Contrast-enhanced ultrasound, CEUS)是在常规超声的基础上发展的一种新型的非侵入性影像学检查方法，可显著提高对病变组织的检测精准度[5]。在颈动脉斑块治疗过程中，医生需要通过 CEUS 图像来确定斑块的形态和大小[6]。然而，这一过程需要医生手动标注病变区域，人工标注费时费力且标注准确性存在主观性。因此，基于 CEUS 图像自动分割斑块已成为目前医学人工智能的重点研究内容，主要包括传统方法和深度学习方法。

通常来说，基于传统方法的颈动脉斑块分割涉及多个组件的组合，包括图像预处理、感兴趣区域(Region of Interest, RoI)特征提取和斑块分割。大多数传统分割算法的重点在于从图像中提取更具代表性的特征。一些方法仅关注颈动脉血管边界的分割。Sumathi 等[7]尝试使用基于边缘映射的水平集分割方法分割远端血管壁的内中膜厚度(IMT)。Nagaraj 等[8]提出一种基于支持向量机(Support Vector Machine)的感兴趣区域提取和内中膜分割算法。许多其他方法明确地检测超声图像中的斑块。Destrempes 等[9]利用运动场估计并将其整合到贝叶斯模型(Bayesian model)来分割超声图像中的斑块。Mehdi 等[10]提出了一种改进的空间模糊 C 均值和集合聚类方法，用以识别颈动脉超声图像中的斑块。Loizou 等[11]提出了一种基于斑点减少滤波和参数化活动轮廓的方法用于分割颈动脉粥样硬化斑块。许多研究者利用传统算法研究了基于 CEUS 图像的颈动脉斑块分割问题。例如 Hoogi 等[12]利用主动轮廓法分割管腔，并拟合抛物线来估计动脉壁，使斑块分割在单个帧中。Zeynettin 等[13]提出了一种新的分割算法尝试在常规超声和超声造影图像上同时分割颈动脉斑块。尽管这些方法在颈动脉斑块分割领域已经取得了实质性进展，但传统算法仍然存在不可忽视的局限性。由于超声图像质量低，基于超声图像的几何、灰度和纹理特征的方法鲁棒性较差。此外，手动选择的特征通常带有主观性。因此，传统的颈动脉斑块分割算法分割性能一般，分割结果不够准确，而且缺乏足够的鲁棒性。

近年来，深度学习的全面进步推动了医学领域的众多发展。许多基于深度学习的方法致力于颈动脉超声图像斑块分割。由于斑块通常生长在颈动脉血管壁(Carotid Artery Wall, CAW)中，因此也有一些关于颈动脉血管壁分割的研究，这种类型的分割通常用于评估颈动脉内膜－中膜厚度(Intima-Media Thickness, IMT)。Carl 等[14]提出了一种新的深度神经网络用于自动描绘中外膜边界(Media-Adventitia Boundary, MAB)和管腔内膜边界(Lumen-Intima Boundary, LIB)。Zhou 等[15]提出了一种基于动态卷积神经网络和改进的 U-Net 的深度学习分割框架，用于从颈动脉三维超声图像中分割 MAB 和 LIB。然而，这些方法只负责分割颈动脉血管壁，不进一步分割斑块。因此，许多方法明确地分割颈动脉超声图像中的斑块。Mi 等[16]通过设计一个具有三分支的多支特征融合模块的分割算法以实现更好的斑块分割。Xie 等[3]通过集成两级和双解码器卷积 U-Net，用于超声图像中颈动脉血管腔和斑块的分割。Meshram 等[17]提出了一种扩展的 U-Net 架构来分割颈动脉斑块。Pankaj 等[18]将注意力机制与 U-Net 架构结合，用于识别颈内动脉(Internal Carotid Artery, ICA)和颈总动脉(Common Carotid Artery, CCA)图像中的颈动脉斑块。尽管这些基于深度学习的方法减轻了一些手工方法的限制，但现有的颈动脉斑块分割方法仍然面临一些挑战：1) 由于卷积运算中感受野的限制通常会阻碍捕获全局上下文和构建远程依赖关系，影响了分割精度的进一步提高。2) 由于超声造影图像的噪声干扰和低质量，病变区域的边界通常模糊不清，这经常导致边界分割效果不理想。3) 很多分割方法将来自编码器和解码器的特征通过简单的跳跃连接直接结合，忽略了有价值的中间特征，导致低效率的融合。因此，如何更有效地利用颈动脉 CEUS 图像以更准确地分割颈动脉斑块仍然是一个挑战。

在过去的十年中，卷积神经网络(Convolutional Neural Networks, CNN)，特别是全卷积网络(Fully Convolutional Networks, FCN) [19]和 U-Net [20]及其变体在医学图像分割方面表现出色，已被广泛应用于各种医学图像分割任务。U-Net 的提出掀起了医学图像分割的热潮，它采用了编码器－解码器结构，通过跳跃连接将编码阶段和解码阶段的特征进行拼接，用于生物医学图像分割。UNet++ [21]设计了一系列

嵌套的、密集的跳跃连接，旨在减少编码器和解码器之间的语义鸿沟。Attention U-Net [22]提出了一种注意力门(Attention Gate, AG)机制，使模型能够关注不同形状和大小的目标，在突出显著特征的同时抑制不相关的特征。MultiResUNet [23]使用多尺度卷积思想对 U-Net 中的卷积模块进行改进，在多个医学图像数据集上都提升了分割性能。DoubleU-Net [24]通过堆叠两个 U-Net 架构并且采用空洞空间金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)，成为医学图像分割领域的强大基准模型。

作为 Transformer 在计算机视觉领域的首次杰出尝试，ViT [25]通过充分利用预训练的模型，在图像分类方面取得了显著成功。此外，为了获得多尺度特征表示，通过从不同尺度提取信息来提高精度和效率。Swin Transformer [26]使用一种有效的基于移位窗口的方法在局部计算自注意力，在图像识别和密集预测任务(如目标检测和语义分割)中达到了最先进的性能。受到 Transformer 在 CV 取得显著成功的启发，它在医学图像分割领域获得了极大的关注。TransUNet [27]首次尝试将 Transformer 和 U-Net 相结合，利用 Transformer 获取全局上下文信息，结合 U-Net 结构恢复局部信息。TransFuse [28]通过并行的方式将 Transformer 和 CNN 结合在一起，在不丢失浅层特征的前提下提高全局上下文建模效率。Swin-Unet [29]提出了一种基于 Swin Transformer 的 U 形架构的纯 Transformer 模型，用于多器官和心脏分割。DS-TransUNet [30]采用基于 Swin Transformer 的双分支编码器来捕获不同粒度的语义信息，用于学习多尺度特征表示，在多种医学图像分割任务中表现出色。

为解决劲动脉超声造影图像斑块分割领域中分割精度和性能方面的限制，本文提出了一种新颖的医学图像分割框架：DATU-Net，旨在提高颈动脉 CEUS 图像中斑块准确分割效果。DATU-Net 集成了 Swin Transformer、双注意力机制和 U 形架构的优势。编码器分支采用 Swin Transformer 模块进行构建，主要用于捕获远程依赖关系和建模全局上下文信息。同时，通过设计基于 CNN 的辅助分支，充分利用卷积操作的局部特性以增强对细节信息的分割能力。在该框架中，我们还开发了双级注意力(Double-Level Attention, DLA)模块，具有提取图像特定的位置特征和通道特征的能力，旨在优化编码器获得的多尺度特征。此外，在解码器部分引入了 Swin Transformer 模块，以增强网络的解码能力。为了评估本文提出的 DATU-Net 的有效性，我们在超声造影颈动脉斑块数据集上进行了广泛的仿真实验，实验结果充分证明了所提方法的卓越性能，为颈动脉超声造影图像的精准斑块分割提供了一种有效且可靠的解决方案。

2. 方法

2.1. 网络概述

本文提出的 DATU-Net 架构如图 1 所示。给定输入图像的尺寸为 $I \in \mathbb{R}^{H\times W\times 3}$，其中 $H\times W$ 表示输入图像的空间分辨率。DATU-Net 采用典型的编码器－解码器架构，该框架基于 Swin Transformer 模块构造编码器分支，用于捕获全局上下文信息并构建分层特征表示；同时基于 CNN 在视觉任务中所具有的先验知识设计辅助分支对输入图像进行下采样，以有效提取斑块的低级细节特征。DATU-Net 将双级注意力(DLA)模块集成到每层的跳跃连接中，以优化编码器传递的特征，过滤不相关信息，从而更准确地向解码器传递特征并改善图像分割性能。在该模块中，通道注意力模块(Channel Attention Module, CAM) [31]和位置注意力模块(Position Attention Module, PAM) [31]被用于有效提取和利用图像特定的位置特征和通道特征。此外，我们将 Swin Transformer 模块引入到解码器中，以进一步探索上采样过程中的远程上下文信息。最后，DATU-Net 可以准确地获得大小为 $H\times W$ 的像素级分割图。

2.2. Swin Transformer 作为编码器

我们将 Swin Transformer 模块和图像块合并层(Patch Merging)堆叠在一起，构成 DATU-Net 的特征编码路径。该路径主要由包含四个阶段的 Swin Transformer 构成，每个阶段包含一定数量的 Swin Transformer

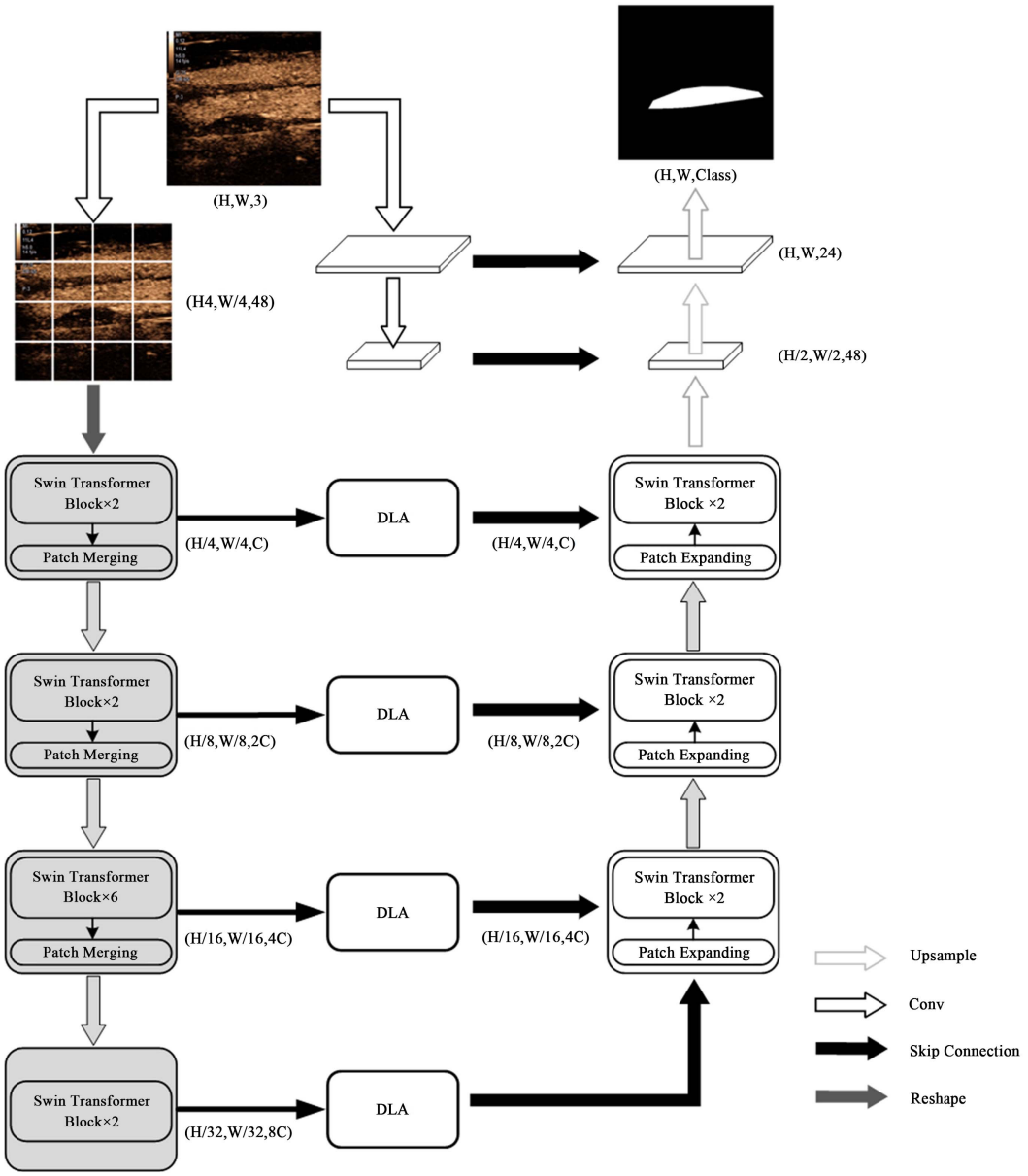

Figure 1. The overall architecture of the proposed DATU-Net
图 1. 提出的 DATU-Net 的整体架构

模块。不同于传统的标准 MSA 模块，为了高效建模，Swin Transformer 模块基于移位窗口的思想构造了更合理的多头自注意力机制。图 2 显示了两个连续的 Swin Transformer 模块，每个模块由 LN (Layer Normalization)层、多头自注意力模块、残差连接和具有 GELU 非线性激活函数的双层 MLP (Multi-layer Perceptron)组成。基于窗口的多头自注意力(Window-based Multi-head Self Attention, W-MSA)模块和基于移位窗口的多头自注意力(Shifted Window-based Multi-head Self Attention, SW-MSA)分别应用于两个连续的 Swin Transformer 模块中。

基于这种窗口划分方法，连续 Swin Transformer 模块可被定义如下：

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

$$z^{l} = MLP\left(LN\left(\hat{z}^{l}\right)\right) + \hat{z}^{l} \tag{2}$$

$$\hat{z}^{l+1} = SW-MSA\left(LN\left(z^{l}\right)\right) + z^{l-1} \tag{3}$$

$$z^{l+1} = MLP\left(LN\left(\hat{z}^{l+1}\right)\right) + \hat{z}^{l+1} \tag{4}$$

其中，z^{l} 和 $\hat{z}^{l}$ 分别表示两个 MSA 模块和 MLP 模块的输出；W-MSA 和 SW-MSA 中自注意力的计算方法可写为如下形式：

$$\text{Attention}\left(Q,K,V\right) = \text{SoftMax}\left(QK^{\text{T}}/\sqrt{d} + B\right)V \tag{5}$$

其中 Q，K，$V \in \mathbb{R}^{M^2\times d}$ 是 query、key 和 value 矩阵。d 表示 query 或 key 的维数，M^2 表示窗口中的图像块数量，$B \in \mathbb{R}^{M^2\times M^2}$ 表示相对位置偏置。

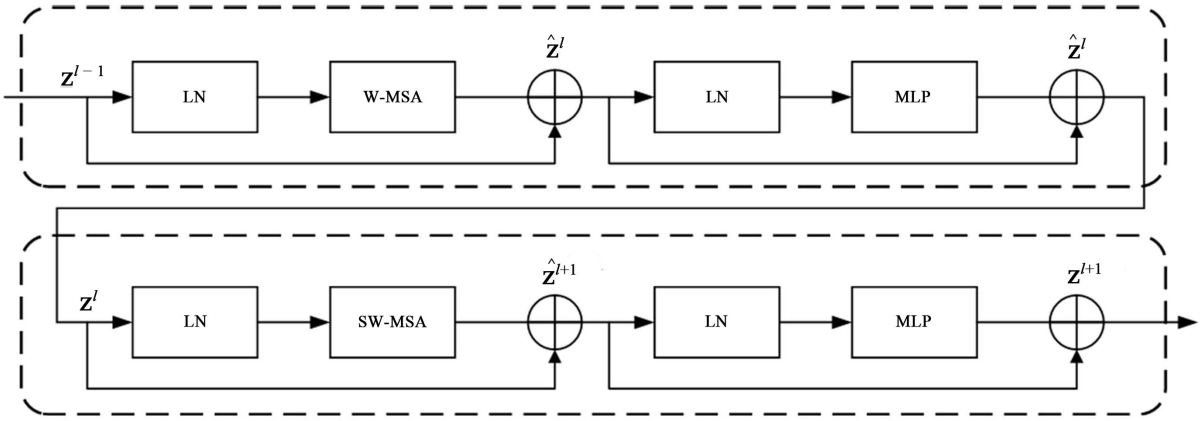

Figure 2. Swin Transformer block
图 2. Swin Transformer 模块

在医学图像被送入到 Swin Transformer 模块之前，它被图像块划分模块(Patch Partition)按照 4 × 4 大小进行分割成一系列不重叠的块。每个图像块都被视为一个“token”，并将通过线性嵌入层(Linear Embedding)将通道数调整到 C 维(C = 96)。随后，这些图像块序列被输入到包含四个阶段的 Swin Transformer 中执行特征表示学习，其中特征维数和分辨率保持不变。为了获得多尺度特征同时产生层级式表示，随着网络的深入，通过块合并(Patch Merging)层来减少图像块的数量。具体而言，块合并层将输入的图像分为 4 部分并将它们沿通道维度拼接(Concatenate)，然后在拼接的特征上应用线性层。这将使图像块数量减少 2 倍，执行 2 倍的分辨率下采样，并将输出特征维数增加 2 倍。因此，四个阶段的输出特征图分辨率分别为 $H/4\times W/4$、$H/8\times W/8$、$H/16\times W/16$ 和 $H/32\times W/32$，通道数分别为 C、2C、4C 和 8C。

2.3. 双级注意力(Dual-Level Attention)模块

如图 3 所示，本文提出的双级注意力(DLA)模块作为 DATU-Net 的关键组件，有效提取和利用了图像特定的位置特征和通道特征，能够分别在空间维度和通道维度上实现特征增强，从而获得更详细和准确的特征集合。DLA 模块同时能够优化跳跃连接中来自编码器的传输特征，过滤掉跳跃连接中的无关信息，方便解码器重建更精确的特征表示。因此，我们将其集成到跳跃连接中以提高模型的分割性能。DLA 由两个主要组件组成：一个具有位置注意力模块(Position Attention Module, PAM)，另一个包含通道注意力模块(Channel Attention Module, CAM)，两者都借鉴了 DANet [31]。

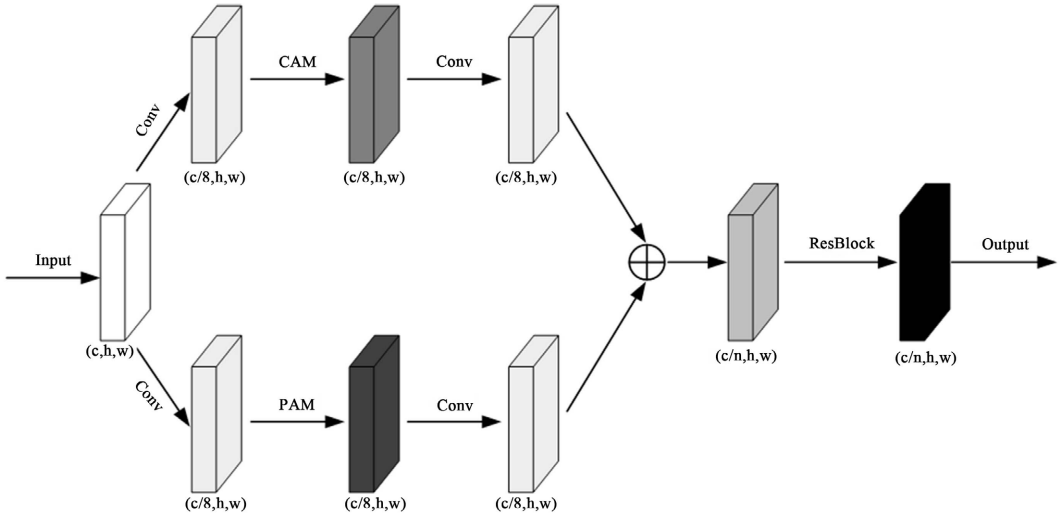

Figure 3. Dual-Level Attention module structure
图 3. 双级注意力模块结构

2.3.1. 位置注意力模块 PAM

如图 4 所示，PAM 捕获特征映射的任意两个位置之间的空间依赖关系，通过所有位置特征的加权和更新特定特征。权重由两个位置之间的特征相似度确定。因此，PAM 在提取有意义的空间特征方面是有效的。位置注意力模块将更广泛的上下文信息编码为局部特征，从而增强了它们的表示能力。

给定 PAM 的初始输入特征图记为 $A\in\mathbb{R}^{C\times H\times W}$，然后 A 将输入到卷积层中，得到三个新的特征图，即 B、C 和 D，每个特征图的尺寸为 $\mathbb{R}^{C\times H\times W}$，接下来将 B 和 C 重塑为 $\mathbb{R}^{C\times N}$，其中 $N=H\times W$ 表示像素数量。之后，在 B 和 C 的转置之间执行矩阵乘法，并应用 softmax 层获得空间注意力图 $S\in\mathbb{R}^{N\times N}$：

$$s_{ji}=\frac{\exp\left(B_i\cdot C_j\right)}{\sum_{i=1}^{N}\exp\left(B_i\cdot C_j\right)} \tag{6}$$

其中，s_{ji} 表示第 i 个位置对第 j 个位置的影响。同时将 D 重塑为 $\mathbb{R}^{C\times N}$，然后与 S 的转置执行矩阵乘法，并将结果重塑为 $\mathbb{R}^{C\times H\times W}$。最后，我们将其乘以参数 α，并于特征图 A 进行元素求和运算，从而得到最终的输出 $E\in\mathbb{R}^{C\times H\times W}$：

$$E_j=\alpha\sum_{i=1}^{N}\left(s_{ji}D_i\right)+A_j \tag{7}$$

其中 α 初始化为 0，并逐渐学习获得更多的权重。由于 E 是所有位置特征和原始特征的加权和，所以它具有全局上下文特征，同时能够根据空间注意力图选择性地聚合上下文。因此，PAM 具有很强的空间特征提取能力，可以有效地提取位置特征，同时保持全局上下文信息。

2.3.2. 通道注意力模块 PAM

通道注意力模块 CAM 结构如图 5 所示，它擅长提取通道特征。与 PAM 不同的是，我们直接将原始特征图 $A\in\mathbb{R}^{C\times H\times W}$ 重塑为 $\mathbb{R}^{C\times N}$，然后将 A 与其转置进行矩阵乘法。随后，我们应用一个 softmax 层，得到通道注意力图 $X\in\mathbb{R}^{C\times C}$：

$$x_{ji}=\frac{\exp\left(A_i\cdot A_j\right)}{\sum_{i=1}^{C}\exp\left(A_i\cdot A_j\right)} \tag{8}$$

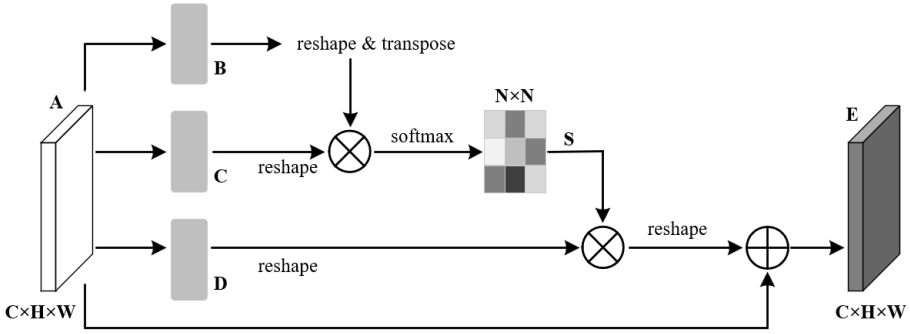

Figure 4. Position attention module structure
图 4. 位置注意力模块结构

这里， x_{ji} 衡量第 i 个通道对第 j 个通道的影响。接下来，对 A 和 X 的转置执行矩阵乘法，并将其将结果重塑为 $\mathbb{R}^{C\times H\times W}$ 。最后，将结果乘以参数 β ，并于 A 进行逐元素求和运算，得到最终的输出 $E\in\mathbb{R}^{C\times H\times W}$ ：

$$E_j=\beta\sum_{i=1}^{N}\left(x_{ji}A_i\right)+A_j \tag{9}$$

其中 β 从 0 开始学习权重。与 PAM 类似，在 CAM 中提取通道特征时，每个通道的最终输出特征都是由所有通道特征和原始特征的加权和生成的，从而赋予了 CAM 模块强大的通道特征提取能力。

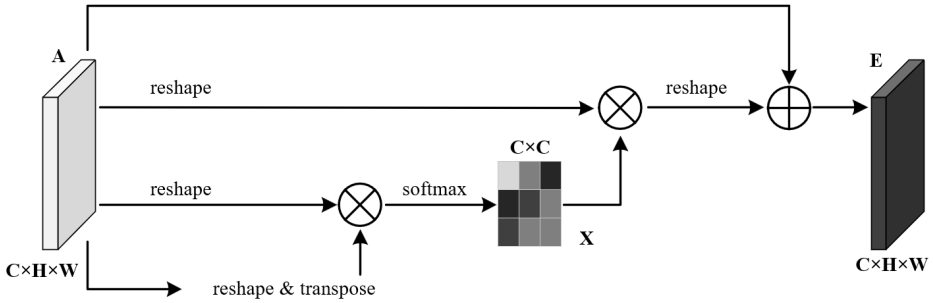

Figure 5. Channel attention module structure
图 5. 通道注意力模块结构

2.3.3. 双级注意力模块 DLA

如图 3 所示，我们给出了双级注意力模块 DLA 的架构示意图。该模块将位置注意力模块(PAM)的强大位置特征提取功能与通道注意力模块(CAM)的通道特征提取优势相结合。DLA 由两部分组成，第一部分以 PAM 为主，第二部分以 CAM 为主。第一部分获取输入特征图，并执行一次卷积，将通道数量缩小为原来的 1/8，得到 f_p 。然后经过 PAM 执行位置特征提取，再进行卷积操作，得到 $\hat{f}_p$ 。具体的计算公式如下所示：

$$f_p=Conv\left(input\right) \tag{10}$$

$$\hat{f}_p=Conv\left(PAM\left(f_p\right)\right) \tag{11}$$

另一部分的组件是相同的，唯一的区别是 PAM 模块被替换为 CAM 模块，其计算公式如下所示：

$$f_c=Conv\left(input\right) \tag{12}$$

$$\hat{f}_c = Conv\left(CAM\left(f_c\right)\right) \tag{13}$$

经过两层注意力模块之后得到输出特征图 $\hat{f}_p$ 和 $\hat{f}_c$，接着进行元素求和，并送入残差模块(ResBlock)[32]传递融合特征，同时恢复通道数，从而得到最终的输出结果。具体计算过程如下所示：

$$output = ResBlock\left(\hat{f}_p + \hat{f}_c\right) \tag{14}$$

这种设计新颖的 DLA 模块很好的集成了 PAM 和 CAM 的优势，以有效增强特征提取，使其成为提高模型整体性能的关键组件。

2.4. 解码器

如图 1 所示，我们基于 Swin Transformer 模块构建了 DATU-Net 的解码器，用于探索上采样过程中的远程依赖关系。受到 Swin-Unet [29]的解码器设计的启发，我们在解码器中使用了图像块扩展层(Patch Expanding)对提取的深度特征图进行上采样。解码器主要由三个阶段组成，每个阶段都包括块扩展层和 Swin Transformer 模块。解码器每一个阶段都会将特征图的分辨率提高 2 倍，并相应地将特征维数降低 2 倍。因此，这三个阶段的输出特征图分辨率分别为 $H/16\times W/16$、$H/8\times W/8$ 和 $H/4\times W/4$，通道数分别为 4C、2C 和 C。同时，也构造了两个常规的解码器模块用于获得分割预测图，每个模块由卷积层、BN 层和 ReLU 激活函数相继组成。经过这两个常规解码器单元执行上采样操作，然后所有的输出特征将被用于获得 $H\times W\times Class$ 的最终分割预测图，其中 $Class$ 表示类别数量。

2.5. 损失函数

为了获得高质量的区域分割和清晰的边界，根据实验我们选择将损失函数定义为混合损失，BCE 损失[33]是二元分类和分割中使用最广泛的损失函数。IoU 最初是为衡量两个集合的相似性而提出的，随后被用作目标检测和分割的标准评估指标。最近，IoU 损失[34]被广泛用作训练损失函数。我们使用的损失函数的计算公式如下：

$$L_{Total} = L_{IoU} + L_{BCE} \tag{15}$$

$$L_{BCE} = -\sum_{(r,c)}\left[G\left(r,c\right)\log\left(S\left(r,c\right)\right)+\left(1-G\left(r,c\right)\right)\log\left(1-S\left(r,c\right)\right)\right] \tag{16}$$

$$L_{IoU} = 1-\frac{\sum_{r=1}^{H}\sum_{c=1}^{W}S\left(r,c\right)G\left(r,c\right)}{\sum_{r=1}^{H}\sum_{c=1}^{W}\left[S\left(r,c\right)+G\left(r,c\right)-S\left(r,c\right)G\left(r,c\right)\right]} \tag{17}$$

其中 $G\left(r,c\right)\in\left\{0,1\right\}$ 表示像素点 $\left(r,c\right)$ 的真值标签，$S\left(r,c\right)$ 表示显著目标的预测概率。

3. 实验

3.1. 数据集

在本文的实验中，我们收集了 146 名颈动脉斑块患者的 CEUS 检测视频，这些视频由上海交通大学医学院附属同仁医院超声科提供，并通过了机构审查委员会的批准。在具有 10 年以上临床经验的临床医生的指导下，对病变区域的像素级标签进行标注，并由多名临床医生进行验证。最终得到的数据集共包含 1200 张 CEUS 图像，并按照 8:1:1 的比例进行划分训练集、验证集和测试集。

3.2. 数据预处理

在实验中，为了增加训练样本的多样性，我们通过数据增强对原始图像进行多种变换。这一举措不

仅有助于提高网络的鲁棒性，降低过拟合的风险，还能有效提高图像分割任务的准确性。在 CEUS 图像尺寸方面，我们进行了裁剪和缩放操作，使其统一为 512*512。同时，我们进行了归一化处理，确保数据在训练过程中保持一致性，提高训练效率。为了更全面地覆盖各种情景，我们还采用了常见的数据增强方法，如图像旋转、图像翻转以及亮度调节等。这些策略的有机结合旨在有效实现数据增强的目标，为模型的仿真训练提供更为丰富的输入信息。

3.3. 实验设置

本文提出的 DATU-Net 的编码器部分采用了 Swin-Tiny [26]作为主干网络，并通过在 ImageNet 数据集上预训练的权重进行参数初始化。在训练过程中，我们设置了 300 个 epoch 的训练轮次，批处理大小为 8，初始学习率为 0.01。为了优化模型的性能，我们选择了 Adam 优化器，其中动量设置为 0.9，权重衰减为 0.0001。DATU-Net 的实现基于 PyTorch 框架，并且所有实验都是在 NVIDIA RTX 3090 GPU 上进行的，以确保计算效率和模型训练的顺利进行。图 6 展示了 DATU-Net 训练 300 个 epoch 的仿真曲线，其中 x 轴表示模型训练的轮次，y 轴表示模型训练的轮次对应的损失值。我们可以观察到训练损失曲线稳步下降，直至趋于平稳，表示模型的训练效果良好，能够正确地捕捉到训练数据中的规律。

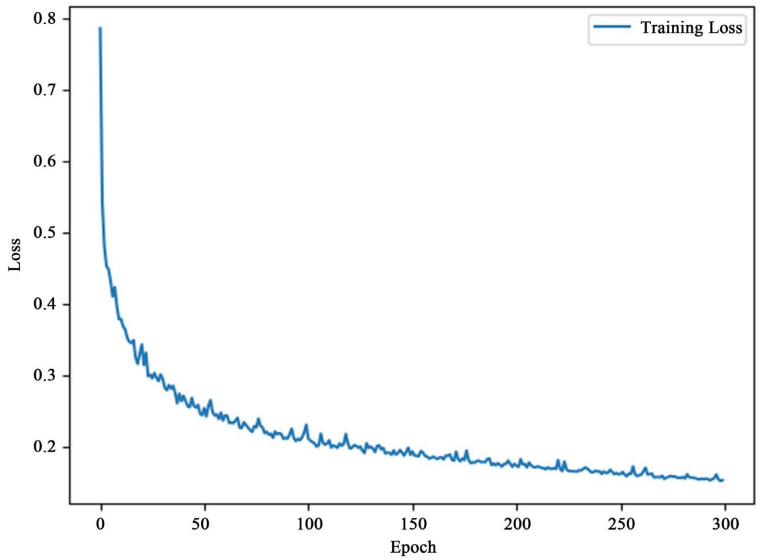

Figure 6. Training loss curve
图 6. 训练损失曲线

3.4. 评价指标

在本文中，我们旨在利用多种评价指标来评估所提出方法的性能，包括 Dice 系数(Dice coefficient score, Dice)、交并比(Intersection over Union, IoU)、精确度(Precision)和召回率(Recall)。这些指标的取值范围在 0 到 1 之间，取值越高表示算法的分割性能越好。Dice 系数评估的是预测像素和真实像素之间的整体相似度。IoU 指标通过计算真实像素与预测像素的交集和并集的比值来衡量分割的准确性。精确度反映了预测样本与真实样本的匹配程度。召回率用于衡量正确识别预测样本的比率。上述指标的公式可定义如下：

$$Dice = \frac{2*TP}{2*TP + FN + FP} \tag{18}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{19}$$

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

$$Recall = \frac{TP}{TP + FN} \tag{21}$$

式中，*TP* (True Positive)表示斑块区域像素的正确分类数；*TN* (True Negative)表示非斑块区域像素的正确分类数；*FP* (False Positive)表示斑块区域像素的错误分类数；*FN* (False Negative)表示非斑块区域像素的错误分类数。

3.5. 对比实验

在本节中，我们针对 CEUS 图像颈动脉斑块分割任务，将提出的 DATU-Net 与多种先进的分割方法(包括 U-Net [20]、UNet++ [21]、Attention U-Net [22]、MultiResUNet [23]、PSPNet [35]、DeepLabV3+ [36]、DenseASPP [37]、TransUNet [12]、Swin-Unet [13])进行了全面比较。定量分析结果如表 1 所示，其中每列中的最佳结果以粗体标出。我们可以发现 DATU-Net 在所有评价指标上都始终优于其他分割模型，具体而言，Dice 和 IoU 分别达到了 85.48%和 76.32%的最佳水平，显著优于 U-Net。与 U-Net 的各种变体相比，如 UNet++和 MultiResUNet，DATU-Net 在 U 形框架的基础上结合了 Swin Transformer 模块，增强了传统编码器－解码器架构的功能性和灵活性。我们观察到 Attention U-Net 通过引入跳跃连接中的注意力门机制相较于 U-Net 有效提升了各项性能指标。相比之下，本文提出的 DLA 模块减少编码器和解码器之间的语义鸿沟方面更为有效，可以有效过滤掉跳跃连接中的无关信息，增强了模型的鲁棒性。PSPNet、DeepLabV3+和 DenseASPP 通过 ASPP 模块获取更大的感受野，涵盖更广泛的上下文信息，进而提高了分割指标，其中 DeepLabV3+和 DenseASPP 分别取得了 79.08%和 80.14%的 Dice 系数。相比之下，DATU-Net 通过在编码器和解码器中引入 Swin Transformer 模块，有效地建模全局上下文信息和远程依赖关系，同时生成多尺度特征表示，提高了医学图像语义分割质量。基于 Transformer 的模型，如 TransUNet 和 Swin-Unet，性能指标明显优于上述模型。其中，Swin-Unet 在 Swin Transformer 模块的指导下取得了第二好的分割成绩，突显了 Swin Transformer 相较于标准 Transformer 的更强大性能。相较于 TransUNet 和 Swin-Unet，DATU-Net 不仅利用 Swin Transformer 模块构造编码器和解码器以有效获取全局依赖和多尺度上下文，还通过将 DLA 模块集成到跳跃连接中，在减少冗余特征的同时提取更有价值的信息，显著提高了整体性能。此外，DATU-Net 中的辅助分支保持了一定的空间细节，有助于获得更准确的斑块分割边界。值得注意的是，DATU-Net 在各项指标上均明显优于排名第二的 Swin-Unet，Dice 和 IoU 分别提高了 1.4%和 2.1%，并且获得了最高的精确度和召回率，分别为 87.46%和 88.63%，显著优于其他竞争对手。

Table 1. Quantitative analysis results of different methods
表 1. 不同方法的定量分析结果

方法	Dice	IoU	Precision	Recall
U-Net	71.49%	64.35%	75.61%	75.86%
U-Net++	72.86%	65.41%	75.93%	76.75%
Attention U-Net	72.92%	65.67%	75.76%	78.25%
MultiResUNet	73.67%	66.75%	78.02%	78.86%
PSPNet	78.60%	69.13%	82.54%	83.40%

续表

DeepLabV3+	79.08%	71.34%	84.42%	85.06%
DenseASPP	80.14%	72.86%	85.19%	86.74%
TransUNet	82.56%	74.40%	86.12%	87.25%
Swin-Unet	83.74%	74.75%	86.54%	87.70%
DATU-Net(ours)	**85.48%**	**76.32%**	**87.46%**	**88.63%**

图 6 显示了本文提出的 DATU-Net 与其他分割模型的定性可视化仿真结果比较。如图 7 所示，我们的模型在分割 CEUS 图像颈动脉斑块任务中表现出色。可以看到，我们的方法始终能够产生最佳的分割结果，这表明 DATU-Net 在整体分割结果和斑块边缘预测方面具有更好的分割能力。总体而言，这些比较结果充分验证了我们提出的 DATU-Net 的有效性和优越性，并且突显了该方法在 CEUS 图像颈动脉斑块自动分割方面的显著优势。

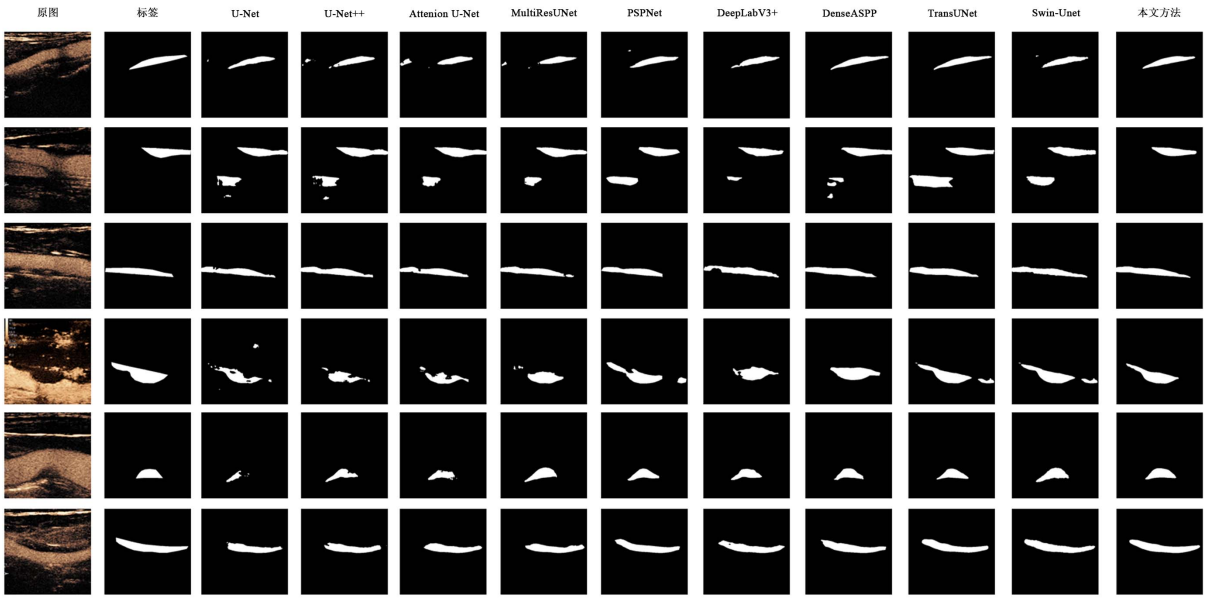

Figure 7. Visualization of plaque segmentation results for different methods
图 7. 不同方法的斑块分割结果可视化

3.6. 消融实验

Table 2. Ablation experimental results
表 2. 消融实验结果

方法	Dice	IoU	Precision	Recall
U-Net	71.49%	64.35%	75.61%	75.86%
U/SE	78.20%	71.39%	82.97%	83.29%
U/SE + SD	81.42%	73.54%	84.04%	86.23%
U/SE + SD + CAB	83.36%	74.62%	86.25%	87.18%
U/SE + SD + CAB + DLA	**85.48%**	**76.32%**	**87.46%**	**88.63%**

为了评估 DATU-Net 中每个组件对分割性能的影响，我们针对 CEUS 图像颈动脉斑块分割任务进行了消融研究，实验结果如表 2 所示，其中每列中的最佳结果以粗体标出。我们将标准的 U-Net 视为基准模型，并基于 U 形结构进行后续实验。具体来说，“U/SE”表示基于 Swin Transformer 编码器的 U 形模型，“U/SD”表示基于 Swin Transformer 解码器的 U 形模型，“U/SE + SD”表示同时基于 Swin Transformer 编码器和解码器的 U 形模型，“U/SE + SD + CAB”表示在“U/SE + SD”的基础上增加了基于 CNN 的辅助分支，“U/SE + SD + CAB”是完整的 DATU-Net 架构。可以看出，在“U/SE”中用基于 Swin Transformer 的编码器替换传统的基于 CNN 的编码器后，Dice 和 IoU 分别显著提高了 6.71%和 7.04%。此外，与“U/SE”相比，在“U/SE + SD”中引入基于 Swin Transformer 的解码器后，Dice 和 IoU 分别提高了 3.22%和 2.15%。通过引入基于 CNN 的辅助分支，“U/SE + SD + CAB”分别将 Dice 和 IoU 提高了 1.94%和 1.08%，这一改进表明，该辅助分支可以有效地增强局部细节，从而提高分割性能。可以看到，“U/SE + SD + CAB + DLA”进一步将 Dice 和 IoU 分别从 83.36%提高到 85.48%和 74.62%提高到 76.32%。如此显著的提升无疑证明了 DLA 模块的巨大贡献，在每个跳转连接层添加 DLA 模块可以有效增强特征提取，为解码器提供更精细的特征，从而减少上采样过程中的特征损失。基于上述消融实验结果，我们认为本文提出方法的所有设计组件都是必要的，因此，完整的 DATU-Net 架构可以实现最卓越的分割性能。

4. 结论

在本文中，我们提出了一种新的医学图像分割方法，即 DATU-Net，该方法融合了 Swin Transformer 模块、双注意力机制和 U 形结构，旨在提高 CEUS 图像颈动脉斑块的分割准确性。我们充分利用 Swin Transformer 模块构建编码器分支，以对全局上下文信息和远程依赖关系进行建模，同时得到多尺度特征表示。基于 CNN 的辅助分支专注于提取细粒度的局部特征，确保获得准确的斑块分割边界。在解码器中引入了 Swin Transformer 模块，以全面建模整个网络的全局上下文信息。我们还提出了一个设计新颖的 DLA 模块，并将其集成到跳跃连接中，有效弥合编码器和解码器之间的语义鸿沟，同时优化输出特征，从而增强了图像分割性能。实验仿真结果表明，DATU-Net 实现了更准确的颈动脉斑块分割，显著提升了分割性能。总体而言，本文提供了一种有效的颈动脉超声造影图像斑块分割方法，有望推动该领域的进展，为临床医生提供更快速的颈动脉斑块识别和相应诊疗支持。

基金项目

上海理工大学医工交叉项目(10-21-302-413)。

参考文献

[1] Organization, W.H. (2019) World Health Statistics Overview 2019: Monitoring Health for the SDGs, Sustainable Development Goals. World Health Organization, Geneva.

[2] Feigin, V.L., Brainin, M., Norrving, B., *et al*. (2022) World Stroke Organization (WSO): Global Stroke Fact Sheet 2022. *International Journal of Stroke*, **17**, 18-29. https://doi.org/10.1177/17474930211065917

[3] Xie, M., Li, Y., Xue, Y., *et al*. (2020) Two-Stage and Dual-Decoder Convolutional U-Net Ensembles for Reliable Vessel and Plaque Segmentation in Carotid Ultrasound Images. *Proceedings of the* 2020 19*th IEEE International Conference on Machine Learning and Applications* (*ICMLA*), Miami, 14-17 December 2020, 1376-1381. https://doi.org/10.1109/ICMLA51294.2020.00214

[4] 张蕴, 敖梦. 超声评价颈动脉易损斑块各病理特征的研究进展[J]. 临床医学进展, 2023, 13(2): 2717-2723.

[5] 孙雪, 李伟, 秦志平, 等. 脑梗死患者颈动脉斑块超声造影特征与梗死特征的相关性[J]. 中国实用神经疾病杂志, 2024, 27(1): 59-63.

[6] Chung, Y.E. and Kim, K.W. (2015) Contrast-Enhanced Ultrasonography: Advance and Current Status in Abdominal Imaging. *Ultrasonography*, **34**, 3-18. https://doi.org/10.14366/usg.14034

[7] Sumathi, K., Mahesh, V. and Ramakrishnan, S. (2014) Analysis of Intima Media Thickness in Ultrasound Carotid Artery Images Using Level Set Segmentation without Re-Initialization. *Proceedings of the* 2014 *International Conference on Informatics*, *Electronics & Vision* (*ICIEV*), Dhaka, 23-24 May 2014, 1-4. https://doi.org/10.1109/ICIEV.2014.7136009

[8] Nagaraj, Y., Hema Sai Teja, A. and Narasimhadhan, A. (2019) Automatic Segmentation of Intima Media Complex in Carotid Ultrasound Images Using Support Vector Machine. *Arabian Journal for Science and Engineering*, **44**, 3489-3496. https://doi.org/10.1007/s13369-018-3549-8

[9] Destrempes, F., Soulez, G., Giroux, M.-F., *et al*. (2009) Segmentation of Plaques in Sequences of Ultrasonic B-Mode Images of Carotid Arteries Based on Motion Estimation and Nakagami Distributions. *Proceedings of the* 2009 *IEEE International Ultrasonics Symposium*, Rome, 20-23 September 2009, 2480-2483. https://doi.org/10.1109/ULTSYM.2009.5441741

[10] Hassan, M., Chaudhry, A., Khan, A., *et al*. (2012) Carotid Artery Image Segmentation Using Modified Spatial Fuzzy C-Means and Ensemble Clustering. *Computer Methods and Programs in Biomedicine*, **108**, 1261-1276. https://doi.org/10.1016/j.cmpb.2012.08.011

[11] Loizou, C.P., Petroudi, S., Pattichis, C.S., *et al*. (2012) Segmentation of Atherosclerotic Carotid Plaque in Ultrasound Video. *Proceedings of the* 2012 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, 28 August-1 September 2012, 53-56. https://doi.org/10.1109/EMBC.2012.6345869

[12] Hoogi, A., Adam, D., Hoffman, A., *et al*. (2011) Carotid Plaque Vulnerability: Quantification of Neovascularization on Contrast-Enhanced Ultrasound with Histopathologic Correlation. *American Journal of Roentgenology*, **196**, 431-436. https://doi.org/10.2214/AJR.10.4522

[13] Carvalho, D.D., Akkus, Z., Van Den Oord, S.C., *et al*. (2014) Lumen Segmentation and Motion Estimation in B-Mode and Contrast-Enhanced Ultrasound Images of the Carotid Artery in Patients with Atherosclerotic Plaque. *IEEE Transactions on Medical Imaging*, **34**, 983-993. https://doi.org/10.1109/TMI.2014.2372784

[14] Azzopardi, C., Camilleri, K.P. and Hicks, Y.A. (2020) Bimodal Automated Carotid Ultrasound Segmentation Using Geometrically Constrained Deep Neural Networks. *IEEE Journal of Biomedical and Health Informatics*, **24**, 1004-1015. https://doi.org/10.1109/JBHI.2020.2965088

[15] Zhou, R., Fenster, A., Xia, Y., *et al*. (2019) Deep Learning-Based Carotid Media-Adventitia and Lumen-Intima Boundary Segmentation from Three-Dimensional Ultrasound Images. *Medical Physics*, **46**, 3180-3193. https://doi.org/10.1002/mp.13581

[16] Mi, S., Bao, Q., Wei, Z., *et al*. (2021) Mbff-Net: Multi-Branch Feature Fusion Network for Carotid Plaque Segmentation in Ultrasound. *Proceedings of the Medical Image Computing and Computer Assisted Intervention-MICCAI* 2021: 24*th International Conference*, Strasbourg, 27 September-1 October 2021, 313-322. https://doi.org/10.1007/978-3-030-87240-3_30

[17] Meshram, N.H., Mitchell, C.C., Wilbrand, S., *et al*. (2020) Deep Learning for Carotid Plaque Segmentation Using a Dilated U-Net Architecture. *Ultrasonic Imaging*, **42**, 221-230. https://doi.org/10.1177/0161734620951216

[18] Jain, P.K., Dubey, A., Saba, L., *et al*. (2022) Attention-Based UNet Deep Learning Model for Plaque Segmentation in Carotid Ultrasound for Stroke Risk Stratification: An Artificial Intelligence Paradigm. *Journal of Cardiovascular Development and Disease*, **9**, Article No. 326. https://doi.org/10.3390/jcdd9100326

[19] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the* 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

[20] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proceedings of the Medical Image Computing and Computer-Assisted Intervention-MICCAI* 2015: 18*th International Conference*, Munich, 5-9 October 2015, 234-241.

[21] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., *et al*. (2018) Unet++: A Nested U-Net Architecture for Medical Image Segmentation. *Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*: 4*th International Workshop*, *DLMIA* 2018, *and* 8*th International Workshop*, *ML-CDS* 2018, *Held in Conjunction with MICCAI* 2018, Granada, 20 September 2018, 3-11.

[22] Oktay, O., Schlemper, J., Folgoc, L.L., *et al*. (2018) Attention U-Net: Learning Where to Look for the Pancreas.

[23] Ibtehaz, N. and Rahman, M.S. (2020) MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. *Neural Networks*, **121**, 74-87. https://doi.org/10.1016/j.neunet.2019.08.025

[24] Jha, D., Riegler, M.A., Johansen, D., *et al*. (2020) Doubleu-Net: A Deep Convolutional Neural Network for Medical Image Segmentation. *Proceedings of the* 2020 *IEEE* 33*rd International Symposium on Computer-Based Medical Systems* (*CBMS*), Rochester, 28-30 July 2020, 558-564. https://doi.org/10.1109/CBMS49503.2020.00111

[25] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al*. (2020) An Image Is Worth 16x16 Words: Transformers for Image

Recognition At Scale.

[26] Liu, Z., Lin, Y., Cao, Y., *et al*. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *The Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 9992-10002. https://doi.org/10.1109/ICCV48922.2021.00986

[27] Chen, J., Lu, Y., Yu, Q., *et al*. (2021) Transunet: Transformers Make Strong Encoders for Medical Image Segmentation.

[28] Zhang, Y., Liu, H. and Hu, Q. (2021) Transfuse: Fusing Transformers and CNNs for Medical Image Segmentation. *Proceedings of the Medical Image Computing and Computer Assisted Intervention-MICCAI* 2021: 24*th International Conference*, Strasbourg, 27 September-1 October 2021, Proceedings, 14-24. https://doi.org/10.1007/978-3-030-87193-2_2

[29] Cao, H., Wang, Y., Chen, J., *et al*. (2021) Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation.

[30] Lin, A., Chen, B., Xu, J., *et al*. (2022) Ds-Transunet: Dual Swin Transformer U-Net for Medical Image Segmentation. *IEEE Transactions on Instrumentation and Measurement*, **71**, 1-15. https://doi.org/10.1109/TIM.2022.3178991

[31] Fu, J., Liu, J., Tian, H., *et al*. (2019) Dual Attention Network for Scene Segmentation. *The Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 3141-3149. https://doi.org/10.1109/CVPR.2019.00326

[32] He, K., Zhang, X., Ren, S., *et al*. (2016) Deep Residual Learning for Image Recognition. *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[33] De Boer, P.-T., Kroese, D.P., Mannor, S., *et al*. (2005) A Tutorial on the Cross-Entropy Method. *Annals of Operations Research*, **134**, 19-67. https://doi.org/10.1007/s10479-005-5724-z

[34] Máttyus, G., Luo, W. and Urtasun, R. (2017) DeepRoadMapper: Extracting Road Topology from Aerial Images. *The Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 3458-3466. https://doi.org/10.1109/ICCV.2017.372

[35] Zhao, H., Shi, J., Qi, X., *et al*. (2017) Pyramid Scene Parsing Network. *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6230-6239. https://doi.org/10.1109/CVPR.2017.660

[36] Chen, L.-C., Zhu, Y., Papandreou, G., *et al*. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *The Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49

[37] Yang, M., Yu, K., Zhang, C., *et al*. (2018) DenseASPP for Semantic Segmentation in Street Scenes. *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 3684-3692. https://doi.org/10.1109/CVPR.2018.00388