

基于MCTUnet的双模态超声图像的 颈动脉斑块分割

何 淳, 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: 2024年12月9日; 录用日期: 2025年1月2日; 发布日期: 2025年1月13日

摘 要

超声检查(US)和对比增强超声(CEUS)是分析病变的空间和时间特征, 以及诊断或预测疾病的重要成像工具。然而, 超声图像通常存在边界模糊和噪声干扰强等特点, 因此, 逐帧评估斑块并描绘病变是一项繁琐且耗时的任务, 这对利用深度学习技术分析超声视频提出了挑战。尽管如此, 现有的超声和对比增强超声图像分割方法中, 能够有效融合这两种不同类型图像的特征信息的仍较为稀缺, 且这些方法在全局上下文信息的提取能力上需优化。为此, 本文提出了一种基于改进Transformer的混合卷积自注意力(MCT)U形结构双分支网络模型, 并在跳跃连接处引入卷积注意力模块。MCT的设计旨在结合卷积和自注意力的优势, 不仅增强了全局上下文信息的捕捉能力, 同时也保留了卷积方法的良好归纳偏置。实验结果表明, 所设计的网络在颈动脉数据集上的表现优于临床医生的诊断结果。

关键词

颈动脉斑块, 分割, 注意力机制, U型结构

Carotid Plaque Segmentation in Dual-Modal Ultrasound Images Based on MCTUnet

Chun He, Zhanquan Sun

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

Received: Dec. 9th, 2024; accepted: Jan. 2nd, 2025; published: Jan. 13th, 2025

Abstract

Ultrasound (US) and contrast-enhanced ultrasound (CEUS) are essential imaging tools for analyzing

the spatial and temporal characteristics of lesions, as well as for diagnosing or predicting diseases. However, ultrasound images are often characterized by blurred boundaries and significant noise interference, making the task of frame-by-frame plaque assessment and lesion depiction labor-intensive and time-consuming. This presents a challenge for using deep learning techniques to analyze ultrasound videos. Despite advancements, there are still few segmentation methods capable of effectively integrating feature information from both types of images, and existing approaches require further optimization in their ability to capture global contextual information. To address this, we propose a U-shaped dual-branch network model based on a Mix Convolutional and Transformer Block (MCT) framework with an improved Transformer, incorporating convolutional attention modules in the skip connections. The MCT framework is designed to leverage the strengths of both convolution and self-attention, enhancing the ability to capture global contextual information while retaining the beneficial inductive biases of convolution. Experimental results demonstrate that our designed network outperforms clinical experts on the carotid artery dataset.

Keywords

Carotid Plaque, Segmentation, Attention Mechanism, U-Shaped Structure

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

颈动脉斑块是常见的心血管疾病, 主要影响颈部的颈动脉, 这些动脉在将血液从心脏输送到大脑的过程中起着至关重要的作用。颈动脉斑块的形成是由于脂肪、胆固醇、钙以及其他物质在血管内壁上的沉积, 这种堆积会增加患者发生心血管事件的风险, 如中风和脑栓塞[1]。中风是成人死亡的第二大原因[2], 并且是严重的神经系统疾病。斑块的狭窄程度和稳定性是影响中风风险的关键因素, 不稳定的斑块容易破裂, 从而可能引发中风。因此, 评估斑块的形态特征, 进而确定狭窄程度和斑块稳定性, 对于预防心血管事件具有重要的临床意义[3]。

超声检查作为一种经济、无创且广泛应用的标准工具, 在颈动脉斑块的诊断中发挥着重要作用。目前, 有多种超声技术可供选择, 每种技术都有其特定的应用和优势: 例如, 标准的 B 超(B-mode 超声) [4], 这是一种常见的超声形式, 通过超声波束生成体内组织的黑白图像图谱, 帮助医生观察斑块在颈动脉中的大小、形状和位置。在 B 超模式下, 斑块通常表现为血管壁上的亮白色区域。彩色多普勒超声则利用多普勒效应测量和显示血流的速度与方向[5], 帮助医生可视化斑块周围的血流阻塞, 从而评估斑块对血液循环的影响。彩色多普勒超声能将血流的彩色映射加入 B 超图像中, 使诊断更加直观。对比增强超声(CEUS) [6]通过注射微泡对比剂, 增强超声波的反射性, 从而提供更清晰的血管和斑块图像[7]。对比增强超声特别适用于揭示斑块的微观特征及血管内皮的状况, 帮助进一步评估斑块破裂的风险。

然而, 在标准超声图像中, 由于伪影、腔内噪声、不规则性和回声斑块的存在, 准确的可视化可能面临挑战。超声对比剂能够更好地描绘颈动脉腔内结构, 而不依赖于血流速度和方向, 但仍存在特定的伪影, 且缺乏组织信息[8], 使得斑块分割变得困难。因此, 颈动脉斑块的分割主要依赖于标准超声图像[9]。

一些研究探索了在标准 B 超图像中进行颈动脉斑块分割的方法[10]。Loizou 等人[9]提出了一种基于灰度标准化、斑点噪声减少滤波和蛇形分割的技术。该研究比较了四种蛇形技术的准确性, 使用彩色多

普勒图像来克服 BMUS 中检测回声透明区域的困难, 并提取初步的蛇形轮廓。Loizou 等人[10]还提出了另一种基于蛇形分割的斑块分割方法, 按不同时间帧对斑块进行分割, 具体取决于手动初始化的第一帧。

对比增强超声(CEUS)为临床医生提供了一种经济、无创的方法, 用于监测组织血流和灌注的动态增强[11]。然而, 手动在每一帧中勾画感兴趣区域——包括病灶及周围微血管——既耗时又费力。临床医生需要仔细观察 CEUS 帧中的超声数据, 以精确定位病灶位置。整合常规超声与对比增强超声图像信息来模拟临床医生的病灶分割方法被认为是一种有前景的解决方案。此前, 在斑块分割方面的挑战有所改善; 特别是 Zheling Meng [11]开发了一种新的 U-net 架构[12], 名为 CEUSegNet, 该模型采用双重自上而下的分支结构和残差连接, 增强了跨模态分割注意力和模态间特征融合。

尽管基于深度学习的颈动脉超声图像分割技术已有进展, 现有研究在分割准确性和斑块定位方面仍存在局限性。受 Transformers 在自然语言处理任务中取得的成功[13]启发, Transformer 的应用已扩展至视觉挑战[14]。本文引入了一种创新的网络架构, 用于自动分割颈动脉斑块, 结合了常规和对比增强超声图像。通过引入 Transformer, 利用交叉注意力机制促进更优的特征融合[15], 并采用常用于图像分割的 U 型网络架构, 实现了完全自动化、无需手动操作的斑块分割, 并显著提高了分割精度。

过去的方法多依赖于标准或对比增强超声, 鲜有方法能有效结合这两种模态。本文提出的方法通过双分支下采样结构有效利用两种模态的信息, 同时改进了原有的 Transformer 模块, 提出了 MCT(混合卷积自注意力), 结合了卷积和自注意力的优势。在保留卷积原有归纳偏置的同时, 利用自注意力来获取全局感受野, 从而增强全局上下文信息的捕捉能力。此外, 通过在跳跃连接处引入 CBAM 模块, 使模型在解码过程中更聚焦于斑块区域, 减少背景干扰, 提高了分割的准确性和鲁棒性。

2. 仿真模型

2.1. 问题定义

本文中主要是针对基于超声和造影两种模态的颈动脉斑块分割, 建立了一种新型的基于双模态 U 型网络架构——MCTU-net, 旨在增强动脉斑块的分割效果, 并克服现有方法的局限性, 同时通过超声(US)和对比增强超声(CEUS)两种不同模态的结合, 提高模型分割精度。本文所提出的模型是以颈动脉超声和造影这两种不同的图像作为输入, 最终输出一个斑块标签图, 即预测颈动脉的斑块位置和大小。

2.2. 仿真模型

本文提出了一种新型的双模态(US 和 CEUS)颈动脉斑块分割模型, 旨在构建一个融合卷积神经网络(CNN)和 Transformer 优势的混合网络。该模型采用双分支输入策略, 将 CEUS 和 US 两种模态分别输入不同的支路。首先, CEUS 支路与 US 支路在通道层面进行拼接, 并通过步长为 2 的 3×3 卷积层进行下采样, 随后输入 ReLU 激活函数和 BatchNorm 操作。输出结果既作为跳跃连接, 也作为下一层的输入。同时, CEUS 支路单独通过 3×3 步长为 2 的卷积层进行下采样, 经过 ReLU 激活和 BatchNorm 处理后, 继续拼接到 US 支路的输入中。此过程重复 3 次, CEUS 和 US 图像经过三层卷积下采样后, 最终双分支的输入再次拼接, 并通过步长为 2 的 3×3 卷积层进行下采样, 以提取局部信息并降低图像尺寸, 从而减少后续自注意力机制的参数复杂度。

接下来, 经过线性投影处理后的输入被重新整形为向量格式, 并输入到一个 12 层的混合卷积自注意力(MCT)模块中, 构成模型的编码器部分。经过 MCT 模块处理后的输出被变形为适合解码器输入的格式。解码器部分由编码器下采样输出的跳跃连接和编码器的最终输出组成。解码器首先对输入进行 3×3 卷积, 再通过转置卷积进行两倍上采样。每层跳跃连接都输入到 CBAM [16]模块中, CBAM 模块的输出与上采样部分进行拼接, 然后再次通过转置卷积进行两倍上采样, 并应用 ReLU 激活函数和 BatchNorm

操作。经过三次这样的上采样操作后, 最后进行一次上采样, 得到最终的分割结果。

本文的模型利用 US 和 CEUS 双模态图像提供了对颈动脉斑块的不同视角信息。由于传统的 US 图像是通过探测超声波在体内的反射来生成图像, 反射信号的强度与组织的密度和弹性等物理特性相关。由于不同类型的组织(如动脉壁、血管腔、软组织等)具有不同的声阻抗, US 图像能够清晰地展示组织的结构和界限, 所以 US 图像强调解剖结构。CEUS 使用超声波与微泡对比剂的组合, 微泡对比剂能够增强血管内血流的反射信号, 从而提供更为清晰的血流信息。对比剂通常具有较强的声学反射信号, 可以突出血管内流动的血液和微血管结构, 尤其是在血流丰富的区域。由于 CEUS 能够增强血流的反射信号, 它能够更清晰地显示血流动态和血管内的灌注情况。对于颈动脉斑块, CEUS 图像能够更好地突出病灶区域, 特别是血流不均匀或被斑块阻碍的区域。通过血流的增强, CEUS 有助于揭示斑块的形态、大小、血供情况, 以及可能的斑块破裂或血栓形成的风险, 从而对诊断具有重要价值。双模态的结合为模型提供了丰富的空间和语义信息, 有助于提高分割的准确性。这种双模态数据输入能够更全面地捕捉斑块特征, 减少因单一模态带来的信息缺失, 使模型在检测和分割病灶区域时更加精准。该网络模型的结构如图 1 所示。

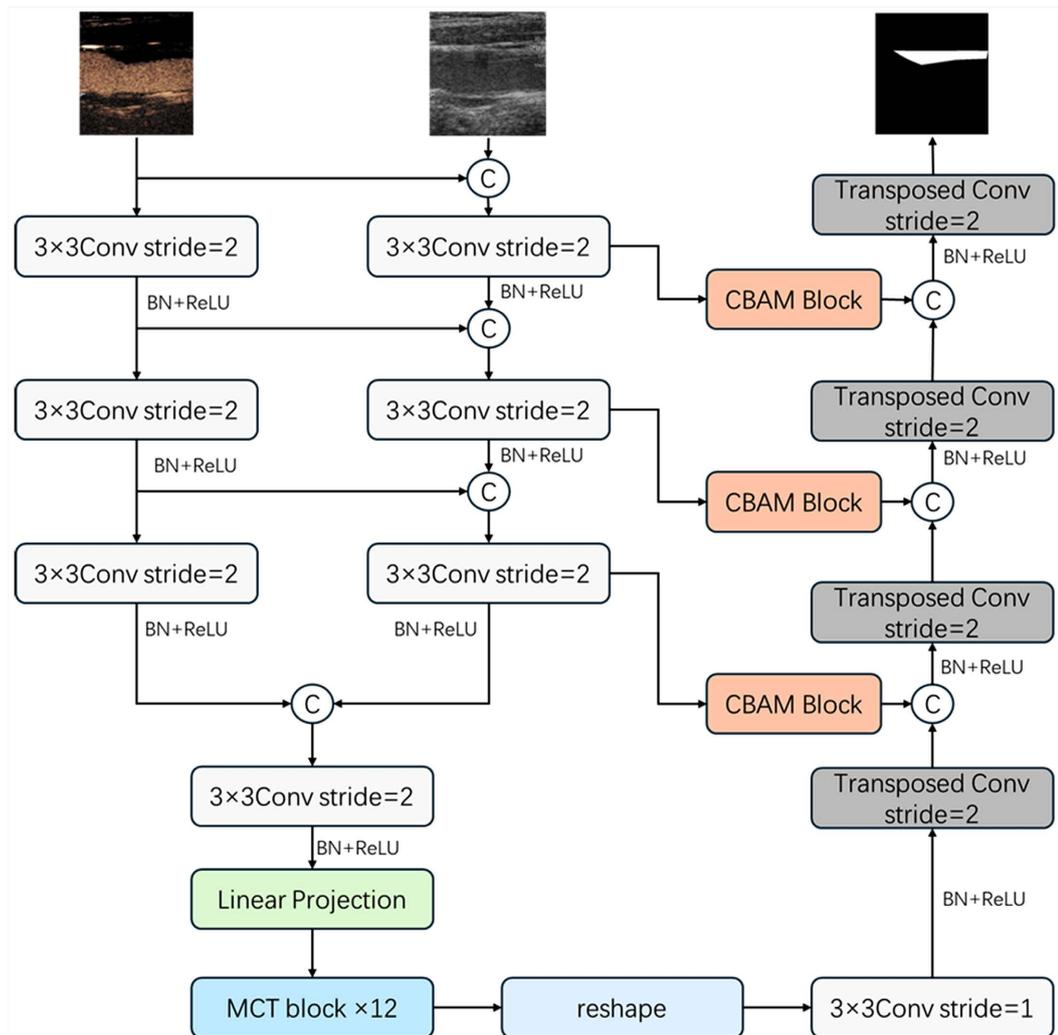


Figure 1. Network architecture diagram

图 1. 网络结构图

2.2.1. 混合卷积注意力(MCT)模块

自 Transformer 技术在自然语言处理领域引入以来, 众多学者开始探索其在图像领域的应用。随着 Vision Transformer 的出现[15], 这类研究进一步加速, 旨在优化 Transformer 技术在图像分析中的整合应用。虽然 Transformer 擅长捕捉图像的全局上下文信息, 并能较好地保留图像细节, 但其缺乏卷积网络所具备的良好归纳偏置, 同时面临计算复杂度过高的问题。为解决这一挑战, 本文对原有的 Transformer 模块进行了改进, 结合 CNN 和 Transformer 的优势, 提出了混合卷积自注意力模块(MCT Block)。MCT 模块由多尺度感知单元(MSPU)、多头自注意力模块(MHSA)以及卷积前馈网络构成, 如图 2 所示。接下来, 本文将详细介绍这三个组成部分。

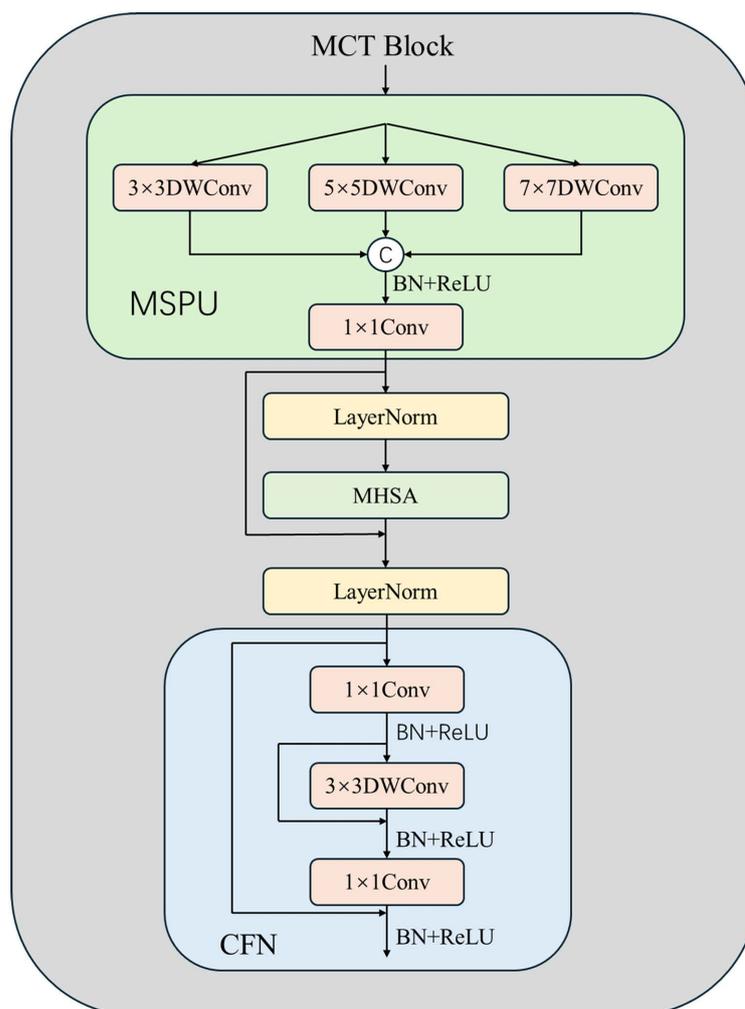


Figure 2. MCT Block
图 2. 混合卷积注意力模块

卷积具有良好的归纳偏置特性, 即权重共享和平移不变性。对于图像任务来说, 旋转和移位是常见的数据增强方法, 这些操作不应影响模型的最终输出结果, 因此在这些任务中实现平移不变性尤为重要[17]。然而, 传统 Transformer 中所使用的绝对位置编码最初是为捕捉 token 顺序而设计的, 这在图像任务中破坏了这种不变性, 因为每个 patch 都被赋予了唯一的位置编码[18]。此外, Transformer 忽略了 patch 内部的局部关系[19]和结构信息[20]。为了解决这些局限性, 本文提出了多尺度感知单元(MSPU), 通过三

种不同大小的卷积核进行分组卷积, 以提取具有不同感受野的信息。输入 $X \in R^{C \times H \times W}$ 分别通过这三种不同大小的卷积进行计算, 并将输出按照通道维度拼接起来得到 $X' \in R^{3C \times H \times W}$, 之后将 X' 进行批量归一化, 并输入激活函数 ReLu 中, 之后再输入 1×1 的点卷积, 降低通道维度得到输出 $Y \in R^{C \times H \times W}$ 。如下公式(1)所示:

$$X' = \text{Concat}(\text{DWConv}(X, \text{size} = 3), \text{DWConv}(X, \text{size} = 5), \text{DWConv}(X, \text{size} = 7)) \quad (1)$$

$$Y = \text{PWConv}(\text{ReLu}(\text{BN}(X'))) \quad (2)$$

之后将 Y 输入本文的多头注意力模块, 输入先经过层归一化, 将 Y 的维度变为向量格式, 并分为 3 部分 $Q, K, V \in R^{\frac{C}{3} \times H \times W}$, 之后将 Q, K, V 输入本文的多头自注意力中得到的输出并与残差连接 Y 进行相加得到最终的输出 Y' , 再将输出 Y' 进行层归一化, 并进行变换维度得到输出 $Z \in R^{C \times H \times W}$ 。如下公式(3)~(5)所示:

$$Q, K, V = \text{split}(\text{LN}(Y)) \quad (3)$$

$$\begin{aligned} Y' &= \text{Attention}(Q, K, V) + Y \\ &= \text{softmax}\left(\frac{Q(K)^T}{\sqrt{d}}\right)V + Y \end{aligned} \quad (4)$$

$$Z = \text{LN}(Y') \quad (5)$$

将输出 Z 输入本文的 CFN 中。本文的 CFN 与 ViT [14]中提出的原始的 FFN 不同的是通过卷积来实现的, 首先将 Z 输入 1×1 的点卷积, 将通道维度变为原来的 4 倍, 并进行批量归一化和激活函数 ReLu 得到 $Z' \in R^{4C \times H \times W}$, 并通过 DW 卷积进行特征提取, 并于残差连接的 Z' 进行相加, 最后再通过 1×1 的点卷积, 将通道维度变为输入的维度 Z'' 。本文使用卷积来替代原线性层, 从而获得了更好的精度表现。如下所示

$$Z' = \text{ReLu}(\text{BN}(\text{PWConv}(Z))) \quad (6)$$

$$Z'' = \text{ReLu}(\text{BN}(\text{PWConv}(\text{ReLu}(\text{BN}(Z' + \text{DWConv}(Z')))))) \quad (7)$$

2.2.2. 模块卷积块注意力模块(CBAM)

CBAM (Convolutional Block Attention Module)是一种旨在提升卷积神经网络性能的注意力机制模块, 由 Sanghyun Woo 等人于 2018 年提出[16]。CBAM 通过引入通道注意力和空间注意力, 增强模型的感知能力, 从而在不显著增加网络复杂度的情况下改善性能。与 Transformer 不同, CBAM 是基于卷积实现注意力机制的。

为了克服传统卷积网络在处理 CEUS 和 US 图像时的局限性, 特别是对不同形状、尺度和方向信息的不足, 本文在跳跃连接的输入上引入了 CBAM 模块。这样可以在解码器的每一层中更好地保留细节信息。CBAM 的作用在于从低层特征中提取关键信息, 避免上采样时将不相关特征带入解码器中。这种方式有效减少了特征冗余, 提升了特征图的表达能力, 从而提高了分割的精度。此外, CBAM 可以更精确地选择要传递的特征, 并弱化不重要的特征, 避免因直接拼接导致的信息丢失或冗余。将注意力机制应用于上采样后的特征图, 使解码器更专注于细粒度信息, 尤其是在边界和细节恢复上表现更出色。CBAM 帮助网络在特征重建过程中减少背景噪声的干扰, 增强上采样效果, 从而使分割结果更加准确和清晰。

CBAM 主要由通道注意力和空间注意力两部分组成, 能够自适应地优化中间特征映射。CBAM 的具体结构如图 3 所示。

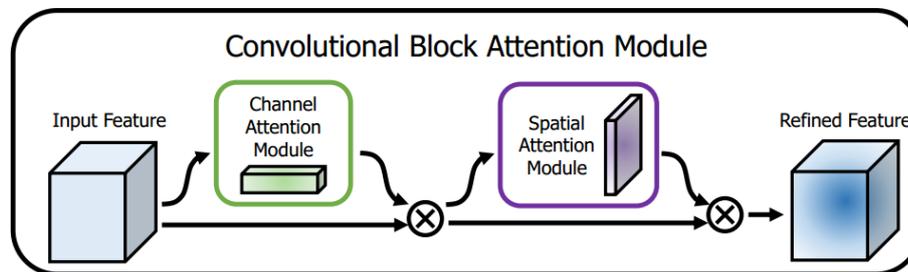


Figure 3. CBAM

图 3. 卷积块注意力模块

以中间特征图 $F \in R^{C \times H \times W}$ 作为输入, CBAM 依次计算得出一维通道注意力图 $M_c \in R^{C \times 1 \times 1}$ 和二维空间注意力图 $M_s \in R^{1 \times H \times W}$, 如图 1 所示。整个注意力过程可以概括为:

$$F' = M_c(F) \otimes F \quad (8)$$

$$F'' = M_s(F') \otimes F' \quad (9)$$

其中 \otimes 表示逐元素的乘法。在乘法过程中, 注意力值相应地被广播(复制): 通道注意力值沿着空间维度被广播, 反之亦然。 F'' 是最终的输出。

通道注意力模块旨在增强每个通道的特征表达。对于输入特征图, 首先对每个通道执行全局最大池化和全局平均池化操作, 以计算每个通道上的最大特征值和平均特征值。这样生成的两个向量分别表示每个通道的全局最大特征和平均特征, 并反映其全局信息。接下来, 将这两个特征向量输入一个共享的全连接层。该全连接层用于学习每个通道的注意力权重, 帮助网络自适应地确定哪些通道对当前任务更为重要。通过结合全局最大特征向量和平均特征向量, 计算得到最终的注意力权重向量。

为了确保注意力权重处于 0 到 1 之间, 应用 Sigmoid 激活函数, 从而生成通道注意力权重。这些权重随后应用于原始特征图的每个通道, 强调对当前任务有贡献的通道并抑制无关的通道。通过将注意力权重与原始特征图的每个通道相乘, 生成加权后的通道特征图, 从而有效提升模型的特征表达能力。通道注意力模块的结构如图 4 所示:

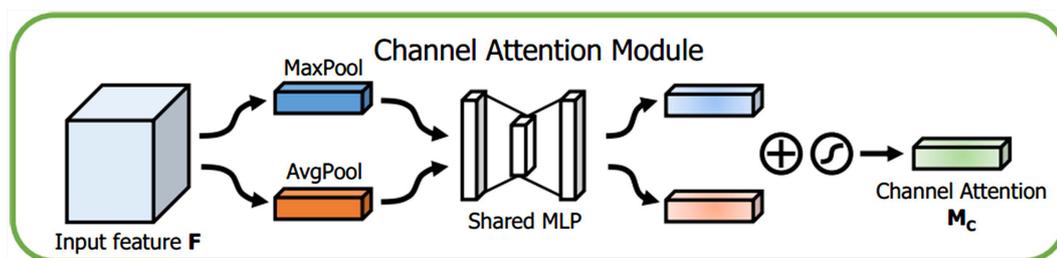


Figure 4. Channel attention module

图 4. 通道注意力模块

空间注意力模块旨在突出图像中不同位置的重要性。对于输入特征图, 首先沿通道维度分别执行最大池化和平均池化操作, 生成不同上下文尺度的特征图。然后, 将这两个池化后的特征图沿通道维度进行拼接, 形成一个包含多尺度上下文信息的特征图。接着, 通过一个卷积层处理该特征图, 以生成空间注意力权重。

与通道注意力模块类似, 空间注意力模块将生成的权重应用 Sigmoid 激活函数, 以将权重限制在 0 到 1 之间。随后, 将这些空间注意力权重应用于原始特征图, 对每个空间位置的特征进行加权, 突出图

像中对任务重要的区域, 同时减少不相关区域的影响。这种方式能够有效增强图像中的显著区域, 提高模型的空间特征表达能力。空间注意力模块的结构如图 5 所示:

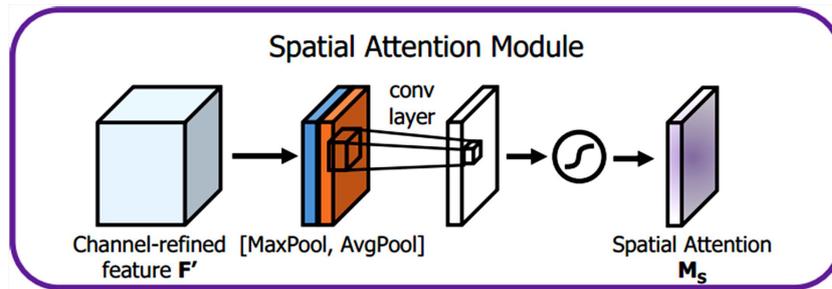


Figure 5. Spatial attention module
图 5. 空间注意力模块

2.3. 损失函数

本文选择 Dice 损失和交叉熵损失作为损失函数。Dice 损失是从 Dice 系数中获得的, 其公式如公式(10)所示。

$$\text{DiceLoss} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (10)$$

交叉熵损失是一种在深度学习中常用的损失函数, 通常用于分类问题。它衡量了模型预测结果与实际结果之间的差距, 是优化模型参数的关键指标之一。其公式如公式(11)所示。

$$L_{(x,y)} = \sum_{i=1}^c x_i \log y_i \quad (11)$$

其中, x_i 表示真实标签的第 i 个元素, y_i 表示模型预测 x 属于第 i 类的概率。当这两个概率分布越接近时, 交叉熵损失越小, 表明模型的预测结果越准确。

本文的损失函数由交叉熵损失函数和 Dice 损失函数加权构成。经过实验, 得出当 λ 为 0.6 的时模型性能最好, 如公式(12)所示

$$\text{Loss} = (1 - \lambda)L_{(x,y)} + \lambda \text{DiceLoss} \quad (12)$$

3. 仿真结果

3.1. 实验环境

本文提出的模型及其他对比模型均在 Linux 平台上通过 1.8 版本 PyTorch 框架实现, 并使用具有 12 G 显存的 NVIDIA 3080 GPU 进行计算。

3.2. 评价指标

为了量化卷积神经网络的性能, 在颈动脉斑块影像数据集上进行了实验, 通过评估指标如 Dice 系数、交并比(IOU)、精确率(Precision)和召回率(Recall)来评估所提出模型的性能。分割结果与标签之间的相似性通过 Dice 相似系数进行计算, 该系数有效地衡量了两者边界之间的重叠程度。交并比是生成的预测图像与图像标签的重叠率, 即它们交集与并集的比值。理想情况下, 重叠率为 1, 即完全重叠。在预测样本中, 召回率为样本图中实际为正的点被最终正确预测的比例。精确率是正确的正类别预测数量与所有正类别预测数量的比率。如公式(13)~(16)所示:

$$\text{Dice} = \frac{2|X+Y|}{|X|+|Y|} \quad (13)$$

$$(14)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (15)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (16)$$

3.3. 仿真结果

3.3.1. 颈动脉斑块数据集仿真实验结果

本文模型实验所选取的数据集是从上海同仁医院收集的颈动脉斑块数据集进行分析, 颈动脉斑块数据集包含 US 和 CEUS 这两种模态的输入, 以及标签 Mask。其中训练集, 验证集, 测试集以 7:2:1 的比例进行划分。所有的结果均是在测试集上得出的指标。

本文在该数据集上对其他先进的模型进行了实验, 并对结果进行了比较。为了确保实验的公平性, 所有模型均设置了统一的学习率和批量大小, 并在相同的运行环境下进行实验。结果如表 1 所示。方法 [12]和[21]-[25]是针对单一模态设计的, 因此, 为了公平性, 本文将 CEUS 和 US 模态按照通道维度进行拼接后输入模型。方法[26]则是针对 CEUS 和 US 模态的分割。从实验结果可以看出, 本文的方法远优于其他方法。

Table 1. Performance of different models on this dataset

表 1. 不同模型在该数据集上的精度

Method	Dice (\uparrow)	Recall (\uparrow)	IOU (\uparrow)	Precision (\uparrow)
TransUnet [21]	83.50%	85.23%	79.62%	87.56%
SwimUnet [22]	76.36%	79.36%	74.80%	82.12%
Unet [12]	75.67%	71.32%	72.38%	80.32%
DeeplabV3 [23]	74.49%	69.45%	71.38%	77.78%
SegNeXt [24]	76.19%	78.34%	73.56%	78.34%
SegViT [25]	77.48%	79.23	74.21%	80.55%
CEUSegNet [26]	82.51%	84.73%	83.91%	86.34%
Our	89.72%	92.84%	87.38%	89.12%

本文的方法通过结合卷积和自注意力机制, 有效提高了全局和局部特征的提取能力。此外, 双分支下采样和 CBAM 模块的引入进一步增强了特征表达和上下文信息的捕捉能力, 显著提升了分割的精度与鲁棒性。这些设计改进在各种衡量指标上都显示了卓越的性能, 均超过其他分割网络, 证明本文提出的网络架构在颈动脉斑块分割任务中的有效性。

3.3.2. 消融实验

为了展示本文提出方法的优越性, 本文对本文所提出的模型进行了一些模块的移除实验。本文在颈动脉斑块的超声和对比增强超声数据集上进行了消融实验, 首先使用 TransUnet 作为基准网络, 然后分别在其基础上加入本文提出的 MCT 模块和 CBAM 模块以及最终设计的网络结构, 并通过 Dice 系数、交并

比、召回率等指标来比较实验性能, 实验结果如表 2 所示。

Table 2. Ablation experiments for different Module
表 2. 针对不同模块的消融实验

Method	Dice (↑)	Recall (↑)	IOU (↑)	Precision (↑)
TransUnet [21]	83.50%	85.23%	79.62%	87.56%
+MCT	86.23%	88.43%	84.95%	88.93%
+CBAM	85.61%	86.71%	85.70%	88.37%
+MCT + CBAM	88.38	89.85%	87.27%	91.42%
All	89.72%	92.84%	87.38%	89.12%

从表 2 所示结果, 可以看出基准网络 TransUnet 将传统的 TransFormer 模块替换成本文提出的 MCT 模块各个指标均有显著的提高。在跳跃连接处加上 CBAM 模块, 模型的性能也均有提高。将 MCT 和 CBAM 这两模块组合使用, 模型的指标也都得到了提高。同时最后使用本文的双分支模型, 可以看到在 Dice 系数, 召回率以及交并比这几个指标上都得到了提高。

同时, 为了验证使用两种模态的有效性, 以及本文模型能够充分提取这两种不同模态的特征并进行良好的信息融合, 本文进一步进行了分别使用单一模态(US 或 CEUS)和使用双模态的实验。由于本文的网络是双输入的, 当使用单一模态(US 或 CEUS)时, 为了确保实验的公平性, 本文在两个输入端分别输入相同的模态。如表 3 所示。

Table 3. Ablation experiments for two modalities of US and CEUS
表 3. 针对超声(US)和对比增强超声(CEUS)两种模态的消融实验

Method	Dice (↑)	Recall (↑)	IOU (↑)	Precision (↑)
US	83.84%	81.79%	83.21%	83.09%
CEUS	87.48%	89.27%	84.95%	88.93%
US + CEUS	89.72%	92.84%	87.38%	89.12%

从表 3 可以看出, 当仅使用一种模态超声(US)时, 模型的表现较差; 而当仅使用对比增强超声(CEUS)时, 模型性能有所提升; 当将两种模态在本文的模型中进行融合时, 能够达到最佳性能。这表明本文的模型能够充分利用来自两种不同模态的信息, 从而提升模型的准确性。

4. 结束语

本文研究了颈动脉斑块分割问题, 这是一项医学人工智能领域的重要任务。为实现高精度且自动化的颈动脉斑块分割, 同时更有效地利用双模态信息(即 US 和 CEUS), 本文提出了一种新型的 U 型网络架构——MCTUnet。该架构基于本文提出的 MCT 模块, 巧妙地融合了卷积和自注意力的各自优势, 在获取全局感受野的同时, 保留了卷积良好的归纳偏置特性。

本文的模型采用双分支结构, 对两种模态分别进行下采样并进行信息交互, 从而充分利用这两种模态的信息, 提高了分割精度。此外, 在跳跃连接处引入 CBAM 模块, 有效减少冗余特征, 增强特征图的表达能力, 进一步提升了分割精度。通过对比实验和消融实验的结果可以明显看出, 本文的模型在各项指标上都优于其他模型, 并且在分割效果上表现最佳, 在 Dice 系数达到了 89.72%, 召回率达到了 92.84%, 交并比达到了 87.38%, 同时精确率达到了 89.12%。本文的模型不同于之前针对单模态的颈动脉斑块分

割, 本文模型同时利用了两种不同模态的信息 CEUS 和 US, 通过信息交互, 从而可以充分利用两种不同模态的优势, 更加精确地提取目标特征, 从而极大地提高了医学上基于颈动脉斑块的分割精度。

基金项目

本文由国防基础研究计划项目(JCKY2019413D001)、上海理工大学医学工程交叉项目(10-21-302-413)资助。

参考文献

- [1] Naghavi, M., Libby, P., Falk, E., Casscells, S.W., Litovsky, S., Rumberger, J., *et al.* (2003) From Vulnerable Plaque to Vulnerable Patient: A Call for New Definitions and Risk Assessment Strategies: Part I. *Circulation*, **108**, 1664-1672. <https://doi.org/10.1161/01.cir.0000087480.94275.97>
- [2] Benjamin, E.J., Muntner, P., Alonso, A., *et al.* (2019) Heart Disease and Stroke Statistics-2019 Update: A Report from the American Heart Association. *Circulation*, **139**, e56-e528.
- [3] He, K., Lian, C., Zhang, B., Zhang, X., Cao, X., Nie, D., *et al.* (2021) HF-UNet: Learning Hierarchically Inter-Task Relevance in Multi-Task U-Net for Accurate Prostate Segmentation in CT Images. *IEEE Transactions on Medical Imaging*, **40**, 2118-2128. <https://doi.org/10.1109/tmi.2021.3072956>
- [4] European Carotid Plaque Study Group (1995) Carotid Artery Plaque Composition-Relationship to Clinical Presentation and Ultra-Sound B-Mode Imaging. *European Journal of Vascular and Endovascular Surgery*, **10**, 23-30.
- [5] Lee, W. (2013) General Principles of Carotid Doppler Ultrasonography. *Ultrasonography*, **33**, 11-17. <https://doi.org/10.14366/usg.13018>
- [6] ten Kate, G.L., van Dijk, A.C., van den Oord, S.C.H., Hussain, B., Verhagen, H.J.M., Sijbrands, E.J.G., *et al.* (2013) Usefulness of Contrast-Enhanced Ultrasound for Detection of Carotid Plaque Ulceration in Patients with Symptomatic Carotid Atherosclerosis. *The American Journal of Cardiology*, **112**, 292-298. <https://doi.org/10.1016/j.amjcard.2013.03.028>
- [7] Feinstein, S.B. (2004) The Powerful Microbubble: From Bench to Bedside, from Intravascular Indicator to Therapeutic Delivery System, and Beyond. *American Journal of Physiology-Heart and Circulatory Physiology*, **287**, H450-H457. <https://doi.org/10.1152/ajpheart.00134.2004>
- [8] ten Kate, G.L., Renaud, G.G.J., Akkus, Z., van den Oord, S.C.H., ten Cate, F.J., Shamdasani, V., *et al.* (2012) Far-Wall Pseudoenhancement during Contrast-Enhanced Ultrasound of the Carotid Arteries: Clinical Description and *in Vitro* Reproduction. *Ultrasound in Medicine & Biology*, **38**, 593-600. <https://doi.org/10.1016/j.ultrasmedbio.2011.12.019>
- [9] Loizou, C.P., Pattichis, C.S., Pantziaris, M. and Nicolaides, A. (2007) An Integrated System for the Segmentation of Atherosclerotic Carotid Plaque. *IEEE Transactions on Information Technology in Biomedicine*, **11**, 661-667. <https://doi.org/10.1109/titb.2006.890019>
- [10] Loizou, C.P., Petroudi, S., Pattichis, C.S., Pantziaris, M., Kasparis, T. and Nicolaides, A. (2012) Segmentation of Atherosclerotic Carotid Plaque in Ultrasound Video. *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>
- [11] Chung, Y.E. and Kim, K.W. (2014) Contrast-Enhanced Ultrasonography: Advance and Current Status in Abdominal Imaging. *Ultrasonography*, **34**, 3-18. <https://doi.org/10.14366/usg.14034>
- [12] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [13] Devlin, J., Chang, M.W., Lee, K., *et al.* (2018) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding.
- [14] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al.* (2020) An Image Is Worth 16 x 16 Words: Transformers for Image Recognition at Scale.
- [15] Chen, C.R., Fan, Q. and Panda, R. (2021) CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 357-366. <https://doi.org/10.1109/iccv48922.2021.00041>
- [16] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. *Computer Vision—ECCV 2018 15th European Conference*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

- [17] Semih Kayhan, O. and van Gemert, J.C. (2020) On Translation Invariance in CNNs: Convolutional Layers Can Exploit Absolute Spatial Location. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 14262-14273. <https://doi.org/10.1109/cvpr42600.2020.01428>
- [18] Chu, X.X., Zhang, B., Tian, Z., Wei, X.L. and Xia, H.X. (2021) Conditional Positional Encodings for Vision Transformers.
- [19] Lowe, D.G. (1999) Object Recognition from Local Scale-Invariant Features. *Proceedings of the 7th IEEE International Conference on Computer Vision*, Kerkyra, 20-27 September 1999, 1150-1157. <https://doi.org/10.1109/iccv.1999.790410>
- [20] Lin, T.-Y., Goyal, P., Girshick, R., He, K.M. and Dollár, P. (2017) Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 318-327.
- [21] Chen, J., Lu, Y., Yu, Q., *et al.* (2021) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation.
- [22] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., *et al.* (2023) Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. *Computer Vision—ECCV 2022 Workshops*, Tel Aviv, 23-27 October 2022, 205-218. https://doi.org/10.1007/978-3-031-25066-8_9
- [23] Yurtkulu, S.C., Şahin, Y.H. and Unal, G. (2019) Semantic Segmentation with Extended DeepLabv3 Architecture. 2019 *27th Signal Processing and Communications Applications Conference (SIU)*, Sivas, 24-26 April 2019, 1-4. <https://doi.org/10.1109/siu.2019.8806244>
- [24] Guo, M.H., Lu, C.Z., Hou, Q., *et al.* (2022) SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation. 2022 *Conference on Neural Information Processing Systems*, New Orleans, 28 November-9 December 2022, 1140-1156.
- [25] Zhang, B., Tian, Z., Tang, Q., *et al.* (2022) SegViT: Semantic Segmentation with Plain Vision Transformers. 2022 *Conference on Neural Information Processing Systems*, New Orleans, 28 November-9 December 2022, 4971-4982.
- [26] Meng, Z., Zhu, Y., Fan, X., Tian, J., Nie, F. and Wang, K. (2022) CEUSegNet: A Cross-Modality Lesion Segmentation Network for Contrast-Enhanced Ultrasound. 2022 *IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, Kolkata, 28-31 March 2022, 1-5. <https://doi.org/10.1109/isbi52829.2022.9761594>