

中图法分类号: TP399; TP18 文献标识码: A 文章编号: 1006-8961(2026)07-2601-13

论文引用格式: Wang X, Deng N, Xie F and Xing W Y. 2026. DMA-UNet: dual-domain multiscale attention analysis network for hemorrhage regions in abdominal ultrasound images. Journal of Image and Graphics, 31(7):2601-2613(王轩, 邓娜, 谢锋, 邢文字. 2026. 面向腹腔超声图像出血区域的双域多尺度注意力分析网络 DMA-UNet. 中国图象图形学报, 31(7):2601-2613)[DOI:10.11834/jig.250510]

面向腹腔超声图像出血区域的双域多尺度 注意力分析网络 DMA-UNet

王轩¹, 邓娜¹, 谢锋², 邢文字^{3*}

1. 上海工程技术大学电子电气工程学院, 上海 201600; 2. 复旦大学生物医学工程与技术创新学院, 上海 200438;
3. 康复大学康复科学与工程学院, 青岛 266113

摘要: 目的 腹腔出血病情凶险, 快速准确诊断至关重要。为了解决超声图像出血分割中存在的低对比度、强噪声和边界模糊等挑战, 提出了一种基于空间域和频域结合的双域多尺度注意力 U 型网络 (dual-domain and multiscale attention U-Net, DMA-UNet) 用于超声图像腹腔出血区域分割。**方法** 核心架构包含 3 个关键创新: 编码器的空洞注意力卷积模块集成并行卷积注意力机制与空洞卷积, 在保证效率的同时增强小目标特征捕获; 中间层的快速傅里叶卷积残差块通过频域变换有效建模全局上下文; 跳跃连接的联合注意力结构实现跨层级特征优化。三者协同构成双域结合的多尺度注意力分析框架。**结果** 实验基于上海长征医院和复旦大学附属华山医院 363 幅腹腔出血超声图像, 通过数据增强对模型进行训练、验证与测试。结果表明, 本文提出的模型 Dice 和交并比 (intersection over union, IoU) 指标在两个中心的测试集中分别达到了 0.879 7、0.796 1 和 0.933 9、0.876 2, 消融实验验证了所设计模块对模型性能提升的有效性。与主流的分割模型相比, 分割性能也具有显著的提升。**结论** 提出了一种双域多尺度注意力分析网络, 通过在多个核心模块中引入注意力机制, 有效融合了空间域和频域特征, 显著增强了 DMA-UNet 全局上下文建模能力、信息融合效率和模型的泛化性能, 为腹腔出血区域分割网络的设计提供了新思路, 具有较高的临床应用价值。

关键词: 腹腔出血; 快速傅里叶卷积 (FFC); 多尺度表征; 联合注意力结构 (JAS); 超声图像分割

DMA-UNet: dual-domain multiscale attention analysis network for hemorrhage regions in abdominal ultrasound images

Wang Xuan¹, Deng Na¹, Xie Feng², Xing Wenyu^{3*}

1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201600, China;
2. College of Biomedical Engineering, Fudan University, Shanghai 200438, China;
3. School of Rehabilitation Science and Engineering, University of Health and Rehabilitation Sciences, Qingdao 266113, China

Abstract: Objective Abdominal hemorrhage is a critical medical condition characterized by a dangerous clinical course, making rapid and accurate diagnosis imperative. This study addresses the challenges in ultrasound image segmentation for abdominal hemorrhage, including low contrast, high noise, and blurred boundaries, which often hinder the precise delineation of bleeding regions. To overcome these limitations, we propose a novel dual-domain and multiscale attention U-Net (DMA-UNet) that integrates spatial- and frequency-domain features for enhanced segmentation of abdominal hemorrhage in ultrasound images. This life-threatening condition typically results from trauma, ruptured aneurysms, or postoperative com-

收稿日期: 2025-10-20; 修回日期: 2025-12-26; 预印本日期: 2026-01-02

* 通信作者: 邢文字 xingwenyu@uhrs.edu.cn

plications. Its abrupt onset and rapid progression can swiftly lead to hemorrhagic shock and multiple organ failure, rendering time a decisive factor in determining patient outcomes. In such critical scenarios, ultrasonography has become an indispensable first-line diagnostic tool because of its portability, absence of radiation exposure, and capability for rapid bedside assessment. Accurately identifying and quantifying the extent of hemorrhage on ultrasound scans directly influence clinical decision-making, such as choosing between conservative management or urgent surgical intervention, and provides crucial prognostic information. However, achieving precise automated segmentation of hemorrhage regions presents multiple inherent challenges. Ultrasound images are inherently characterized by low signal-to-noise ratios, pervasive speckle noise, poor tissue contrast, and various artifact interferences. The heterogeneity of pathological manifestations further compounds this difficulty: hemorrhage areas may present as anechoic, hyperechoic, or mixed echogenicity patterns, often with irregular and poorly defined boundaries that blend with surrounding parenchymal or adipose tissues. This combination of technical limitations and pathological variability makes traditional manual segmentation methods time consuming and operator dependent and subject to substantial interobserver variability. Moreover, complex factors in the clinical examination environment pose additional challenges. Ultrasound examinations are frequently conducted under suboptimal conditions, including patient movement, limited acoustic windows, uneven probe pressure, and intestinal gas interference, all of which considerably degrade image quality and consistency. In emergency and intraoperative settings, the inherent tension between diagnostic speed requirements and accuracy demands creates substantial pressure on medical personnel. Although deep learning techniques, particularly convolutional neural network architectures represented by U-Net and its variants, have achieved remarkable progress in medical image analysis for modalities, such as CT and MRI, their direct application to abdominal ultrasound image segmentation continues to face fundamental challenges. These models are typically designed for images with high contrast and standardized spatial characteristics. When processing ultrasound data, they often exhibit limited receptive fields, insufficient global contextual relationship modeling, and inefficient multiscale feature fusion. This phenomenon results in suboptimal performance when segmenting diffuse, low-contrast hemorrhage regions and frequently leads to under-segmentation of scattered bleeding areas or over-segmentation of anatomically similar soft tissues. **Method** To bridge this technological gap, we propose a sophisticated deep learning architecture called dual-domain multiscale attention network (DMA-Net). This framework is meticulously designed around three synergistic and innovative core modules, each specifically optimized to address the distinct limitations of conventional segmentation models when processing complex ultrasound data. The core architecture incorporates the following key innovative modules. 1) Multiscale atrous attention module in the encoder: This module employs a parallel multibranch architecture that systematically expands the receptive field through atrous convolutional layers with varying dilation rates while adaptively calibrating feature weights via dual channel-spatial attention mechanisms. Each branch integrates contextual information at different scales, thereby enhancing the sensitivity to subtle hemorrhage regions while preserving feature map resolution. A residual attention pathway is specifically designed to retain original texture details. The module dynamically suppresses speckle noise interference through learnable spatial masks, and its hierarchical dilation design captures local vascular structures and global bleeding distribution patterns. This multiscale fusion strategy substantially improves the detection accuracy of early-stage and scattered hemorrhage foci under challenging acoustic conditions. 2) Frequency-domain global modeling module in the bottleneck layer: This module implements a frequency-spatial dual-stream architecture that transforms feature maps into the spectral domain via fast Fourier transform, enabling holistic analysis of hemorrhage distribution patterns. In the frequency pathway, learnable spectral filters suppress ultrasound-specific artifacts while amplifying hemorrhage-related frequency components. The spatial pathway maintains convolutional operations to preserve anatomical details. Through residual connections and frequency-domain attention gates, the module selectively enhances cross-region dependencies that are critical for identifying discontinuous bleeding patterns. This approach effectively bridges the gap between localized feature extraction and global contextual understanding and is particularly valuable for segmenting diffuse hemorrhage regions with irregular boundaries. The integrated design achieves computational efficiency while remarkably improving the model's ability to interpret complex acoustic signatures in abdominal ultrasound imaging. 3) Cross-level joint attention fusion module: a dynamic feature selection mechanism is introduced in the encoder-decoder skip connections. Through multiscale feature cross-computation, it generates adaptive attention weights that intelligently filter semantic features across different levels, sup-

press noise transmission, and promote the optimized fusion of deep semantic information with shallow spatial details for accurate boundary reconstruction. **Result** The experimental framework was established using a clinically heterogeneous dataset comprising 363 intra-abdominal hemorrhage ultrasound images sourced from two independent medical centers to ensure anatomical and acquisition variability. A comprehensive data augmentation pipeline was implemented to enhance model robustness and generalizability; this pipeline incorporates spatial transformations (rotation, scaling, and elastic deformation), intensity variations, and speckle noise simulation. Quantitative evaluation on hold-out test sets from both institutions demonstrated the model's strong segmentation performance and cross-center adaptability. For Center A, the model achieved a Dice similarity coefficient of 0.879 7 and an intersection over union (IoU) of 0.796 1. For Center B, the Dice similarity coefficient and IoU reached 0.933 9 and 0.876 2, respectively, reflecting consistent reliability across different imaging protocols and patient populations. **Conclusion** DMA-UNet introduces an ultrasound-adapted segmentation framework that effectively integrates spatial- and frequency-domain representations through coordinated attention mechanisms. By jointly enhancing global context modeling, multiscale feature fusion, and imaging artifact resilience, the network offers a clinically viable solution for rapid and accurate hemorrhage assessment in critical care. Its strong performance across independent datasets underscores its translational potential for use in bedside decision support systems.

Key words: abdominal hemorrhage; fast Fourier convolution(FFC); multi-scale representation; joint attention structure (JAS); ultrasound image segmentation

0 引言

腹腔作为人体最大的潜在性浆膜腔,其内发生的出血症状是一种危及生命的急症。在腹腔出血的早期诊断中,影像学检查至关重要。Kristensen等人(1971)首次将超声技术应用于钝性创伤患者的腹腔游离液检测,开创了无创诊断的先河。现有研究表明(Soyuncu等,2007),超声诊断腹腔出血的敏感性与特异性分别可达86%和99%,使其成为急诊科创伤评估的重要检查手段。然而,超声诊断的准确性在很大程度上依赖于医师的主观经验。因此,开发自动化算法以实现腹腔出血区域的精准分割,对于减少诊断差异、实现病灶量化评估及辅助临床决策具有重大意义。

尽管超声技术应用广泛,但其成像固有的物理特性为自动化分割带来了严峻挑战(George和Anita,2022)。具体而言,图像中普遍存在的斑点噪声会显著降低信噪比;同时,目标区域与背景之间对比度低、病灶边界模糊且尺度差异大,进一步增加了准确分割的难度。上述因素导致临床诊断仍高度依赖医师经验,存在主观性强、效率有限以及诊断一致性不足等问题。因此,开发能够实现精准自动分割的算法,不仅有助于对出血区域进行定量形态学评估,也可为疾病诊断、治疗规划与疗效监测提供客观依据,具有重要的临床转化价值。

近年来,人工智能技术的突破性进展显著推动了医学图像分割领域的发展(Zhu等,2025),作为深度学习中的核心架构之一,卷积神经网络(convolutional neural network, CNN)凭借其局部感知与权值共享特性,在图像处理任务中取得了广泛成果。以U-Net(Ronneberger等,2015)为代表的U型编解码架构已成为主流解决方案。其衍生模型,包括U-Net++(Zhou等,2018)、U-Net 3+(Huang等,2020)、Attention U-Net(Oktay等,2018)及V-Net(Milletari等,2016)等已成功应用于心脏分割、多器官分割及息肉检测等多种医学影像模态,展现出卓越的临床适用性。这些成果充分验证了CNN在医学图像语义特征学习方面的强大能力。

这些基于CNN的工作极大地促进了超声图像中的病灶分割。然而,此类工作缺少对长程空间依赖关系建模(石军等,2025)。尽管研究者尝试通过空洞卷积、自注意力机制和多尺度金字塔等策略进行改进,但对全局上下文信息的捕获能力仍未取得突破性提升(Ruan等,2024)。

Transformer架构为突破这一瓶颈提供了新思路。ViT(vision Transformer)(Dosovitskiy等,2020)率先将自注意力机制成功引入视觉任务,通过建模图像块(patch)间的全局关系捕获长程依赖。随后,研究者探索将Transformer与U型架构相结合,以兼顾全局语义与局部细节。例如,TransUNet(Chen等,2021)在编码器中融入Transformer模块;MedT

(medical Transformer)(Valanarasu 等, 2021)则针对医学图像设计了门控轴向注意力机制;而Swin-UNet(Cao 等, 2023)基于Swin-T(Swin Transformer)构建了分层的U型架构(Liu 等, 2021),在效率和精度上取得了良好平衡。

尽管基于Transformer的方法在建模全局依赖方面优势明显,并在一系列医学图像分割任务中展现出优越性能,但其在应用于超声图像时仍面临特定挑战:1)缺乏局部归纳偏置导致对超声图像中斑点噪声和微钙化灶的敏感性不足;2)边缘分割精度下降;3)模型训练需要大规模标注数据,在小样本场景下泛化性能显著降低。

为了解决先前研究中存在的局限性,本文提出一种基于多尺度注意力机制的双域协同分析模型——双域多尺度注意力U型网络(dual-domain and multi-scale attention U-Net, DMA-UNet),以实现超声腹腔出血的精准分割。

本文的主要贡献如下:1)设计以空洞注意力卷积(atrous attention convolution, AACnv)模块为核心构建的编码结构,通过通道注意力机制动态学习特征通道的重要性权重。其中,并行卷积块注意力模块(parallel convolutional block attention module, P-CBAM)通过两条路径并行处理平均池化和最大池化,减少了由于共享多层感知机可能造成的信息损失,从而使模型聚焦于图像中的关键区域,改善了在腹腔出血超声图像上边缘分割效果不佳的问题。2)引入快速傅里叶卷积残差块(fast Fourier convolution residual block, FFC-ResBlock)作为编解码器的桥梁(瓶颈层),将输入特征通道分解为局部路径(空间域卷积)和全局路径(频域变换)。通过双域协同机制处理语义信息,从而增强模型的长程依赖捕获能力。3)通过多级扩张卷积结合注意力结构对传统的跳跃连接进行了重新设计,提出了联合注意力结构块(joint attention structure, JAS),旨在解决语义分割中上下文信息不足的问题,从而提高模型的可解释性。

1 方 法

1.1 网络架构

U型网络由于其优越的性能和优雅的架构,在医学图像分割中得到广泛应用。本文提出一种新的

U型编码器—解码器神经网络,用于超声图像中腹腔出血的分割,如图1(a)所示。在腹腔出血超声图像分割任务中,原始图像分辨率通常为 512×512 像素,而出血病灶的尺寸和形态具有高度可变性:少量渗出性出血可能仅占据数十像素的模糊区域,而大量出血则可能形成数百像素的不规则斑块。这种动态变化的尺度特性与病灶边缘的低对比度、纹理异质性共同构成了分割任务的核心挑战。过度的下采样操作可能导致小尺度出血特征丢失,或对大尺度病灶的局部细节建模不足。为了解决这个问题,本文提出了一个基于CNN的编码器—解码器框架,以AACnv模块为核心构建编码器,如图1(b)所示,解码器模块同时接收编码器信息和下层解码器信息,通过双卷积层和修正线性单元(rectified linear unit, ReLU)将融合后的信息进行提炼,最后通过 2×2 卷积进行上采样,从而逐步恢复和放大分辨率。AACnv Block如图1(c)所示,由一个标准卷积块、一个空洞卷积和P-CBAM注意力机制组成,然后在每一级(除第1阶段外)进行一次下采样操作。此外,为了解决边缘小目标分割问题,本文在跳跃连接阶段引入了联合注意力结构块JAS。JAS通过交叉计算联合优化权重,跨维度串联通道和空间注意力,从而增强编码器和解码器特征之间的信息融合,缓解长程依赖丢失,显著提升特征融合的效率和质量。最后,通过FFC模块串联编码器和解码器解决长程依赖捕获难题,从而实现对腹腔出血区域更精确地分割。

1.2 P-CBAM模块

计算机视觉领域的注意力机制研究已取得显著进展,其中通道注意力与空间注意力是两个核心研究方向。在通道注意力方面,代表性工作如SE网络(squeeze-and-excitation network, SENet)(Hu 等, 2018)通过自适应地重标定通道特征响应,有效提升了网络的特征判别能力。其核心思想是通过学习通道间的依赖关系,增强重要特征通道的权重。在空间注意力方面,典型方法如空间变换网络(spatial Transformer network, STN)(Jaderberg 等, 2015)能够对输入数据进行空间变换,自动定位并增强关键区域特征。该方法通过可学习的空间变换参数,实现对目标区域的几何不变性建模。基于上述单维度注意力机制的突破性成果,研究者进一步提出了混合注意力模型。其中,卷积块注意力模块(convolutional

block attention module, CBAM)创新性地将通道注意力与空间注意力以级联方式结合(Woo等,2018),通过双重注意力机制协同建模通道间依赖与空间位置关系,显著提升了复杂场景下的特征表征能力。然而,共享多层感知机无法最优化地提取和放大任一种池化方式带来的独特信息,形成了表征瓶颈,限制了模型的表达容量,从而可能导致优化过程存在冲突和干扰,使得训练效率降低,难以收敛到最优解。因此,为了增强网络对超声图像中噪声的鲁棒性,在

AAConv模块中嵌入一种高效的注意力机制P-CBAM,用于实现模型的鲁棒性和泛化能力。P-CBAM区别于传统的CBAM注意力,通过并行计算通道注意力,独立处理平均池化和最大池化,为全局平均池化(global average pooling, GAP)和全局最大池化(global max pooling, GMP)分别提供独立的多层感知机,避免由于共享多层感知机(如CBAM)可能造成的信息损失,从而提升了模型的整体性能和泛化能力,具体结构如图2所示。

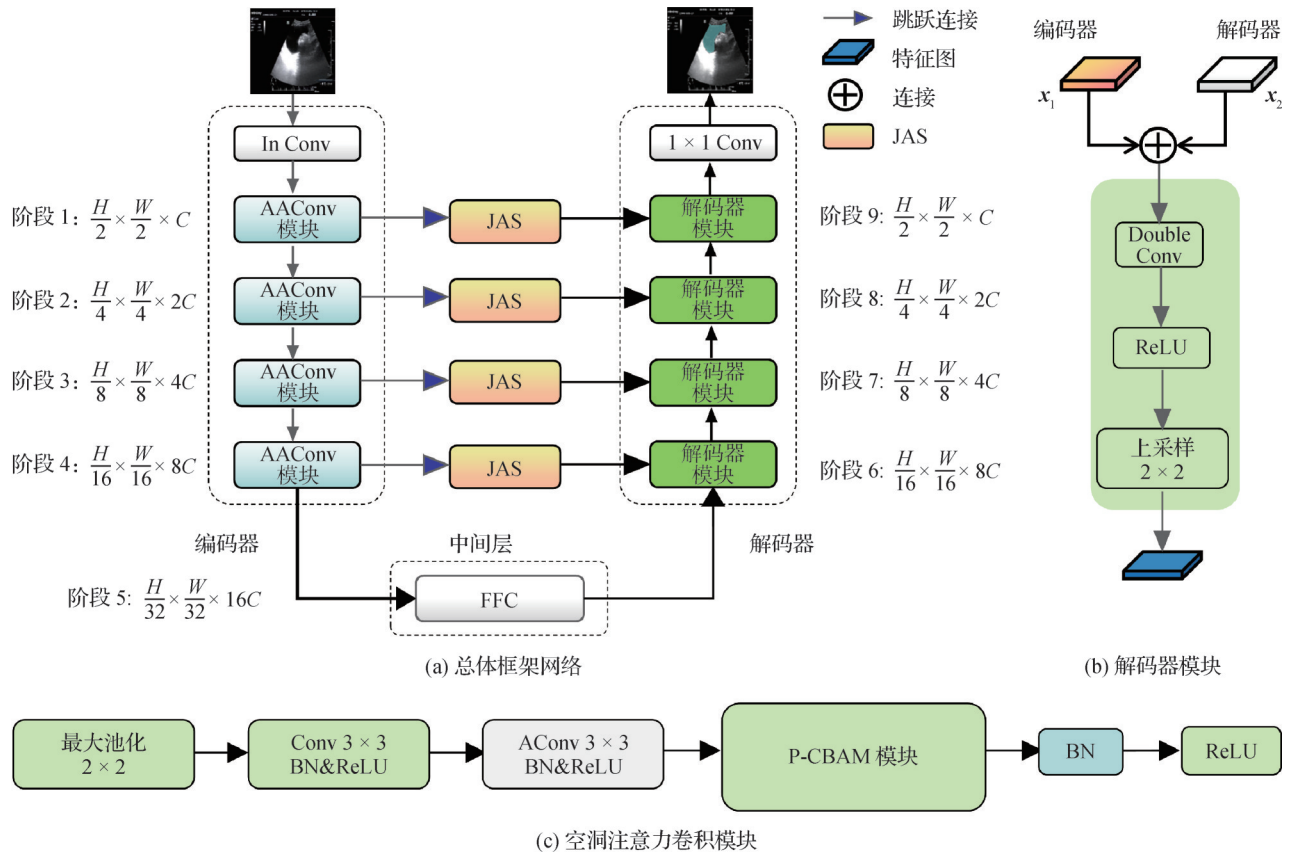


图1 DMA-UNet腹腔出血分割模型

Fig. 1 DMA-UNet abdominal hemorrhage segmentation model

((a)overall framework network; (b)decoder block; (c)atrous attention convolution block)

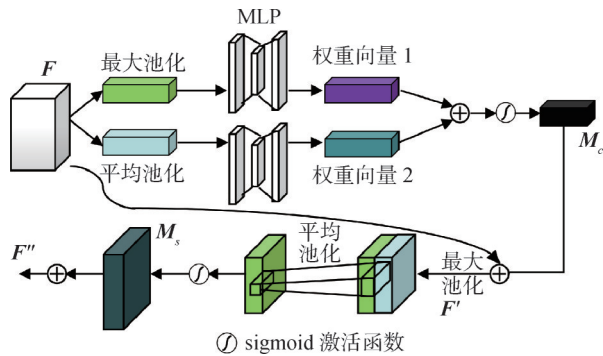


图2 P-CBAM模块示意图

Fig. 2 Schematic diagram of P-CBAM module

对于CNN卷积层输出的特征 $F \in \mathbb{R}^{C \times H \times W}$, P-CBAM模块首先推断一个一维的通道注意力特征向量 $M_c \in \mathbb{R}^{C \times 1 \times 1}$,然后推断一个二维的空间注意力特征图 $M_s \in \mathbb{R}^{1 \times H \times W}$ 。其中,原始特征 F 的通道数、高度和宽度分别为 C, H 和 W 。整个过程可以表示为

$$F' = M_c(F) \otimes F \quad (1)$$

$$F'' = M_s(F') \otimes F' \quad (2)$$

式中, $\otimes$ 表示Hadamard乘法,即逐元素乘法。在逐元素乘积的过程中,通道注意力值将会沿空间维度

传播; F' 为空间注意力模块的输入特征, F'' 为 P-CBAM 模块的输出特征。

1.3 FFC-ResBlock

长距离的像素关系对于捕获图像的整体模式和全局语义信息至关重要(朱智勤 等,2026)。快速傅里叶卷积残差模块(FFC-ResBlock)(Chen 等,2023)作为瓶颈层使 DMA-UNet 能够同时提取全局依赖和局部表示,从而实现对腹腔出血区域更精确的空间分割,如图 3(a)所示。FFC-ResBlock 是一种结合频域和空间域处理的卷积方法,FFC-ResBlock 将输入通道分为局部路径和全局路径。局部路径使用标准卷积处理信息;全局路径则通过傅里叶变换进行捕

捉,这样既保留了局部特征(例如边缘、纹理),又能通过频域变换增强全局特征(例如形状、结构)。

该模型首先在频谱变换块中通过快速傅里叶变换(fast Fourier transform,FFT)捕获全局上下文信息,缓解因下采样导致的高频信息丢失,然后通过压缩激励(squeeze-and-excitation,SE)层,依次执行压缩、激励和重加权操作。SE 层通过学习自动获取每个特征通道的权重,然后根据权重对有利特征进行增强,对不再有利的特征进行抑制。通过使用 SE 层,使模型更加关注到图像中有用的特征,而不是这些符号的特征。最后,再将局部路径和全局路径的输出进行融合。具体步骤如图 3(b)(c)所示。

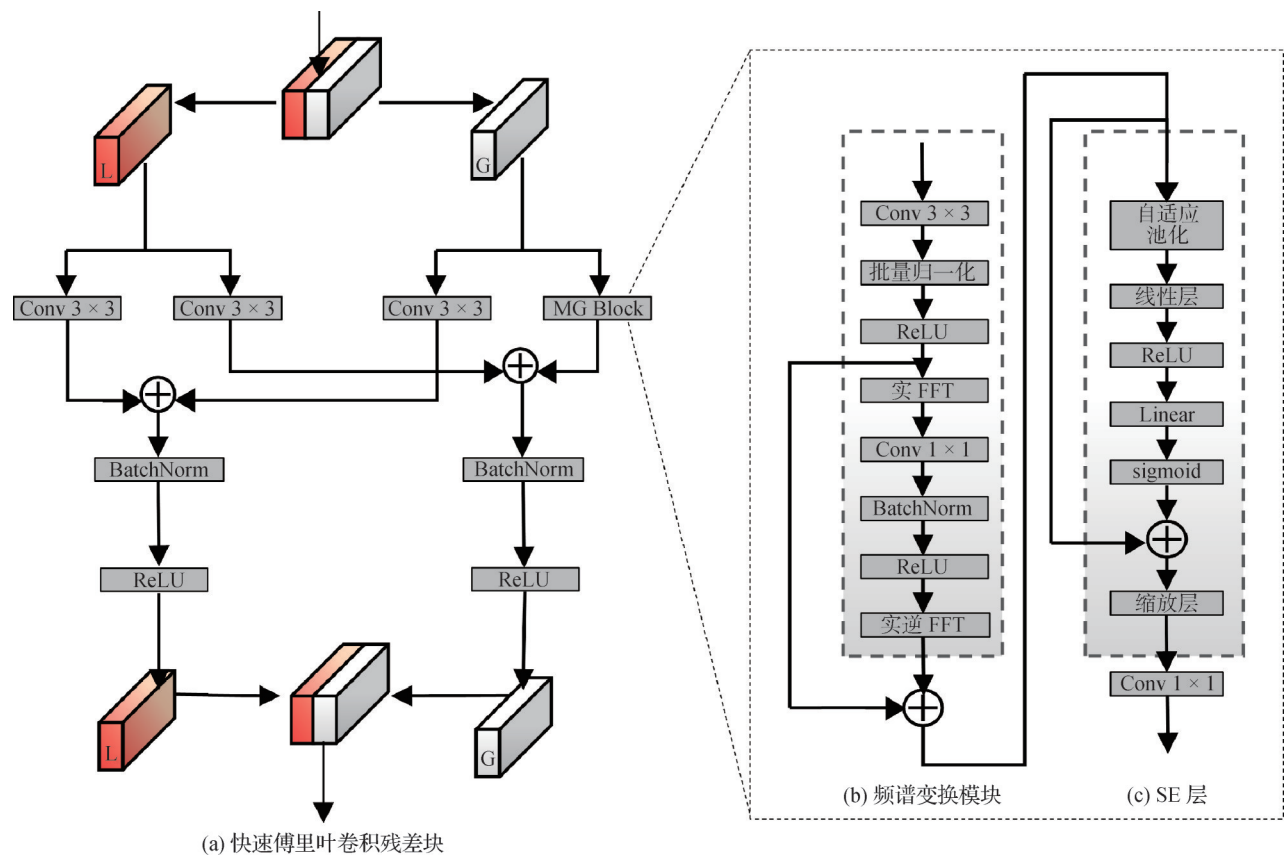


图3 FFC-ResBlock 模块
Fig. 3 FFC-ResBlock module((a)FFC-ResBlock;(b)spectral Transformer block;(c)SE layer)

1.4 JAS 跳跃连接通道

由于超声图像的低信噪比特性,噪声和干扰往往会导致图像质量的显著下降。在图像分割任务中,大多数深度学习模型无法很好地分割小目标区域。此外,腹腔出血的超声图像中往往包含一些视觉上与积液相似的区域,这进一步增加了分割任务的复杂性。JAS 有效解决了这一问题,显著提升了

模型对超声图像中相对位置与空间关系的理解能力(Lei 等,2024)。联合注意力结构特征融合模块采用多级膨胀卷积结合注意力结构,如图 4 所示,其结构类似于空洞空间金字塔池化(atrous spatial pyramid pooling,ASPP)。ASPP 模块是深度学习常用的模块(Chen 等,2018),主要用于图像分割任务,旨在解决语义分割中上下文信息不足的问题。

JAS模块使用多个并行的不同采样率的膨胀卷积层,进一步处理每个采样率在单独分支中提取的特征,并将它们融合生成最终结果。整个过程的计算式可以表示为

$$\mathbf{F}_{\text{jas}} = \text{Cat}(\text{Cat}(\text{DC}_{l=18}, \text{DC}_{l=12})^T \times \text{DC}_{l=6}, \text{SimA}(\text{DC}_{l=3})) \quad (3)$$

式中, DC 表示膨胀卷积, l 表示步长, Cat 表示连接。JAS具有多尺度特征提取和融合能力,可以有效提高网络的泛化性能。与ASPP不同,本文将多尺度膨胀卷积获得的特征图与自注意力融合,然后与无参简单注意力模块(simple attention module, SimAM)增强的特征图连接。SimAM是基于注意力机制的模块,用于计算每个特征图的权重。该模块包含全局平均池化和全连接层两部分。首先,通过全局平均池化计算每个特征图的全局特征向量;随后,全连接层将该向量映射为表示各特征图重要性的注意力权重向量。最终,利用这些权重对输入特征图的相应通道进行加权,生成输出特征图。

SimAM同时考虑了通道和空间信息来计算注意力。它不像CBAM那样将通道和空间注意力分离,也不像SENet那样只关注通道,而是为每一个像素点(每一个通道的每一个位置)计算一个独一无二的权重。遵循SimAM(Yang等, 2021)的设计,为每个神经元(空间和通道维度上的每个点)定义一个能量函数 e^* ,该函数用于衡量该神经元与同一通道内其他神经元的可区分性。能量值越低,表明该神经元越显著。最终的特征增强可以表示为

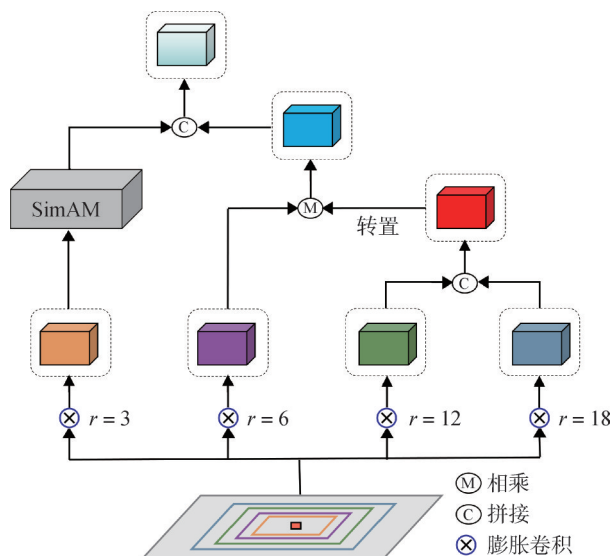


图4 JAS模块

Fig. 4 JAS module

$$\tilde{\mathbf{X}} = f_{\text{sigmoid}}\left(\frac{1}{\mathbf{E}}\right) \odot \mathbf{X} \quad (4)$$

式中, $\mathbf{E}$ 是所有神经元能量值 e_i^* 构成的能量图, f_{sigmoid} 是sigmoid激活函数, $\odot$ 表示逐元素相乘。

1.5 基于频域感知的损失函数

基于频域感知(frequency domain aware, FD)的Dice损失函数是一种结合了传统Dice系数和频域信息的图像分割损失函数(Qu等, 2024),旨在提高模型对图像中不同频率成分的感知能力,特征适用于具有多尺度特征(如不同大小的病变)的医学图像,整个过程的计算式可以表示为

$$\text{Dice} = 2 \cdot |\mathbf{X} \cap \mathbf{Y}| / (|\mathbf{X}| + |\mathbf{Y}|) \quad (5)$$

$$L_{\text{fdice}} = \sum \omega_i \cdot (1 - \text{Dice}_i) \quad (6)$$

式中, $\mathbf{X}$ 是预测分割, $\mathbf{Y}$ 是真实分割, Dice_i 是第 i 个频率分量上计算出的Dice系数, ω_i 是赋予第 i 个频率分量的权重系数。

2 实验

2.1 数据集

实验采用上海长征医院和复旦大学附属华山医院的腹腔出血超声图像数据集,用于模型的训练、验证与测试,所有的分割标签均由有经验的临床医师进行勾勒。本文所使用的腹腔出血超声图像数据集均获得上海长征医院与复旦大学附属华山医院伦理委员会的批准,所有患者均签署了知情同意书。

由于高质量的腹腔出血超声图像数据在临床实践中比较有限,为了在有限数据的基础上提升模型的泛化能力和鲁棒性,在腹腔出血超声图像的预处理阶段,针对211幅上海长征医院原始超声图像,实验中采用了数据增强的方法对原始的腹腔出血超声图像数据集进行处理,其中包括线性变换调整图像灰度分布、模拟不同成像条件下的组织表现以及分别添加高斯噪声(模拟传感器噪声)和椒盐噪声(模拟传输干扰),增强模型对噪声污染的鲁棒性。通过上述方法生成4倍于原始上海长征医院数据的新数据集(共844幅),并按照8:1:1的比例划分为训练集、验证集和测试集,其中源自同一原始数据集中的图像不在训练集、验证集和测试集中交叉出现。此外,为了验证模型的泛化性能,采用了源自复旦大学附属华山医院的152幅腹腔超声图像对训练完成模

型的分割性能进行了验证。

2.2 参数设置

所有模型均在 PyTorch 框架上实现,实验设备为带有 NVIDIA GTX4060 GPU 的工作站。选择 Adam 优化器进行 DMA-UNet 的训练。训练轮次和批次大小分别设置为 100 和 4。初始学习率 $base_{lr}$ 按 0.001 给定,并且按如下方式进行调整,即

$$lr = base_{lr} \times \left(1.0 - \frac{i}{N}\right)^{0.9} \tag{7}$$

式中, lr 为当前学习率, i 表示当前迭代次数, N 表示最大迭代次数。所有超声图像尺寸调整为 512×512 像素。

模型训练过程中,对于 Swin-UNet 和 TransUNet,遵循其原论文的通用实践,使用了在 ImageNet-1K 数据集上的预训练权重来初始化其编码器,这符合此类架构的标准用法,并能体现其最佳性能。所有基于 CNN 的模型(如 U-Net、UNet++、MobileUNet)以及本文提出的 DMA-UNet,为了控制变量,均未使用任何预训练权重,均采用相同的初始化方法和模型训

练超参数。这一设置旨在平衡不同架构的固有特性与对比的公平性。此外,所有模型使用完全相同的训练集、验证集和测试集进行性能评价。

2.3 评价指标

实验采用 Dice 系数、交并比(intersection over union, IoU)、豪斯多夫距离(Hausdorff distance at 95%, HD95)与平均表面距离(mean surface distance, MSD)评估模型预测结果与真值标签的相似程度。

2.4 实验结果与分析

2.4.1 与其他深度学习模型比较

为了评价本文提出的 DMA-UNet 模型在超声图像腹腔出血区域分割的优越性,在两个不同中心的数据集上与其他用于医学图像分割的深度学习方法进行比较,包括基于 CNN 架构的方法(U-Net、UNet++、Attention-UNet、MobileUNet)、基于 Transformer 架构的方法(SwinUNet)和基于混合 CNN-Transformer 架构的方法(TransUNet)。定量评价和显著性分析对比结果如表 1 和表 2 所示。

表 1 不同模型在上海长征医院数据集的分割性能比较

Table 1 Comparison of segmentation performance of different models on the dataset of Shanghai Changzheng Hospital

模型	Dice ↑	IoU ↑	HD95 ↓	MSD ↓	p 值
U-Net	0.813 0±0.06	0.706 6±0.10	6.845 5±1.21	7.616 0±2.58	< 0.001
MobileUNet	0.857 9±0.05	0.773 5±0.08	6.443 6±0.82	5.187 4±1.71	< 0.050
Swin-UNet	0.773 0±0.06	0.650 5±0.13	7.266 1±1.45	9.970 7±3.80	< 0.001
TransUNet	0.814 1±0.07	0.706 9±0.10	6.874 8±1.21	8.648 0±2.11	< 0.001
UNet++	0.845 2±0.05	0.752 0±0.09	6.520 0±0.89	5.800 0±1.96	< 0.010
Attention-UNet	0.831 3±0.04	0.750 3±0.09	6.773 4±1.06	6.234 1±2.08	< 0.010
DMA-UNet(本文)	0.879 7±0.04	0.796 1±0.07	6.298 4±0.53	3.515 4±1.30	—

注:加粗字体表示各列最优结果。“↑”表示值越大越好。“↓”表示值越小越好。“—”表示基线模型。

表 2 不同模型在复旦大学附属华山医院数据集的分割性能比较

Table 2 Comparison of segmentation performance of different models on the dataset of Huashan Hospital, Fudan University

模型	Dice ↑	IoU ↑	HD95 ↓	MSD ↓	p 值
U-Net	0.826 7±0.06	0.766 2±0.10	8.853 7±2.86	7.610 9±2.58	< 0.001
MobileUNet	0.852 1±0.04	0.793 2±0.07	6.127 5±1.08	5.190 6±1.71	< 0.001
Swin-UNet	0.804 8±0.07	0.676 3±0.10	9.490 0±3.24	8.643 4±2.83	< 0.001
TransUNet	0.816 2±0.07	0.712 3±0.10	6.875 5±1.21	8.654 8±2.11	< 0.001
UNet++	0.859 8±0.05	0.753 3±0.09	6.522 1±0.89	5.801 2±1.96	< 0.010
Attention-Unet	0.862 1±0.05	0.799 7±0.08	6.493 6±1.03	5.331 2±1.35	< 0.010
DMA-UNet(本文)	0.933 9±0.01	0.876 2±0.05	3.237 4±0.79	4.035 1±0.66	—

注:加粗字体表示各列最优结果。“↑”表示值越大越好。“↓”表示值越小越好。“—”表示基线模型。

从表1可以观察到,DMA-UNet在上海长征医院数据集上4个评价指标都取得了最好的结果,分别是Dice为0.879 7,HD95为6.298 4,IoU为0.796 1,MSD为3.515 4。相比第2优方法(MobileUNet),Dice和IoU分别提升了0.021 8和0.022 6,在HD95和MSD上则分别降低了0.145 2和1.672。相比于其他方法(U-Net、UNet++、Attention-UNet、Swin-UNet、TransUNet),Dice和IoU平均提升了0.064 4和0.082 8,HD95和MSD平均降低了0.557 9和4.138 3。上述结果表明,本文提出的DMA-UNet模型,通过集成注意力机制和双域协同策略,可以有效地捕获超声图像中腹腔出血的语义和临床层面之间的空间关系,显著提升了模型的分割精度。

在复旦大学附属华山医院数据集上进行实验的结果如表2所示。从表2可以看出,本文提出的DMA-UNet的Dice为0.933 9%,IoU为0.876 2%,HD95为3.237 4,MSD为4.035 1。与其他分割模型相比,DMA-UNet与Swin-UNet的差距最为明显,其中Dice、IoU分别提高了0.129 1和0.199 9,而HD95和MSD分别降低了6.252 6和4.608 3。相比其他分割方法,DMA-UNet的所有指标也都表现出最好的性能,其Dice分别比U-Net、MobileUNet、TransUNet、UNet++和Attention-UNet高出0.107 2、0.081 8、0.117 7、0.074 1、0.071 8。

为了评估性能提升的统计显著性,以Dice为主要指标,对DMA-UNet与所有模型进行了 T 检验统

计分析。结果如表1与表2所示。可以观察到,在两个独立的数据集上,DMA-UNet相对于所有模型在分割性能上存在统计学差异($P < 0.05$)。并且,与部分模型相比(例如与Swin-UNet、U-Net等模型相比),存在更为显著的统计学差异($P < 0.001$)。上述统计分析结果表明了本文提出的DMA-UNet模型在腹腔出血分割任务上的优越性。

2.4.2 对比实验结果可视化

DMA-UNet与先前分割方法在不同的数据集上进行了一系列的可视化对比,如图5所示。从图中可以看出,本文DMA-UNet模型在各种情况下都能够得到与数据标注结果最为接近的分割效果。通过这些图像可以清晰地观察到,在处理边界清晰、周围腔出血超声图像时,多数模型都能够较为准确地识别并分割出病变区域。然而,当遇到如图5第3行所示的边界不规则等干扰因素的复杂情况时,多个模型的分割效果出现了明显的偏差。

2.4.3 消融实验结果分析

为了系统评估各个模块的贡献,在不同的数据集上分别开展了消融实验验证,并对各模块的性能进行了量化分析。实验中,以U-Net作为基准模型,依次引入P-CBAM注意力机制、频域感知(FD)损失函数、空洞卷积(AC)、FFC-ResBlock,以及JAS模块,分别在两个不同测试集上对模型进行了评价,实验结果如表3所示,可视化分析如图6所示。

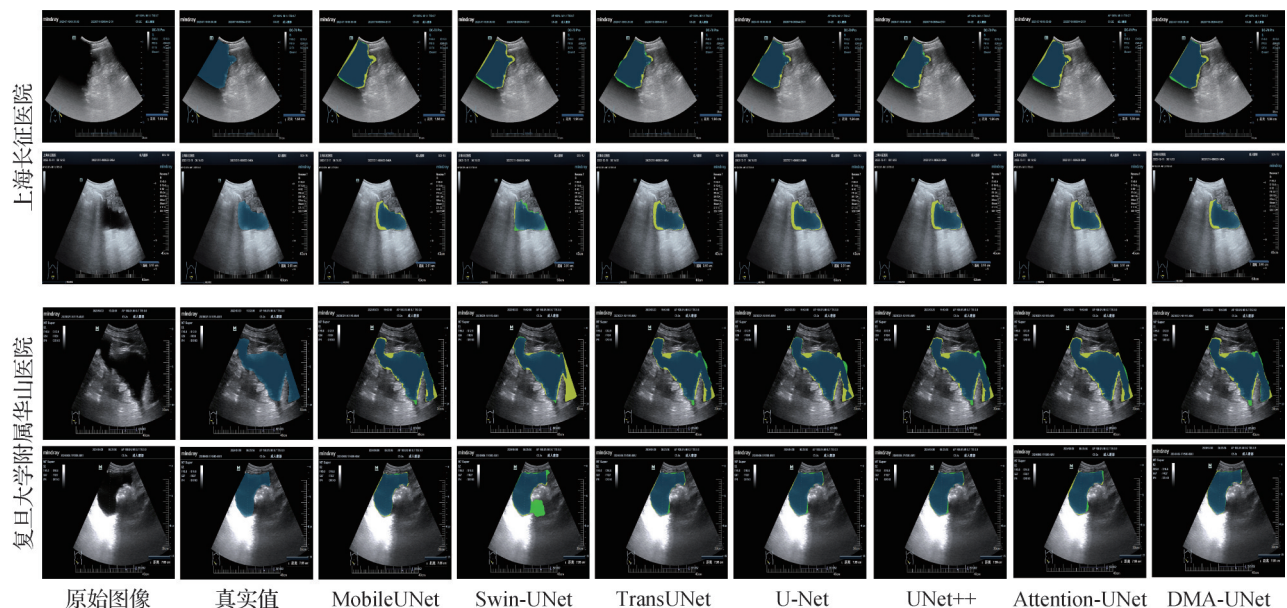


图5 对比实验可视化

Fig. 5 Visualization of comparative experiments

表3 消融实验结果分析
Table 3 Analysis of ablation experiment results

模型	上海长征医院数据集				复旦大学附属华山医院数据集			
	Dice ↑	IoU ↑	HD95 ↓	MSD ↓	Dice ↑	IoU ↑	HD95 ↓	MSD ↓
U-Net(基准)	0.813 0±0.06	0.706 6±0.10	6.845 5±1.21	7.616 0±2.58	0.826 7±0.06	0.766 2±0.10	8.853 7±2.86	7.610 9±2.58
+P-CBAM	0.855 4±0.03	0.763 2±0.05	6.531 9±0.55	6.165 6±2.49	0.903 4±0.04	0.802 2±0.05	7.923 4±1.86	6.650 9±2.24
+FD	0.858 3±0.05	0.765 7±0.08	6.448 0±0.48	5.187 4±1.29	0.913 4±0.03	0.825 2±0.06	7.303 1±1.47	4.765 5±1.64
+AC	0.864 0±0.03	0.773 5±0.06	6.443 6±0.65	4.974 8±2.00	0.928 7±0.02	0.867 7±0.04	7.625 4±1.18	5.036 2±1.12
+FFC-ResBlock	0.877 5±0.04	0.792 7±0.07	6.362 2±0.71	3.558 2±1.41	0.931 6±0.03	0.873 4±0.02	6.493 2±1.24	4.379 1±0.96
+JAS	0.879 7±0.04	0.796 1±0.07	6.298 4±0.21	3.515 4±1.30	0.933 9±0.01	0.876 2±0.01	3.237 4±0.79	4.035 1±0.66

注:加粗字体表示各列最优结果,“↑”表示值越大越好,“↓”表示值越小越好。

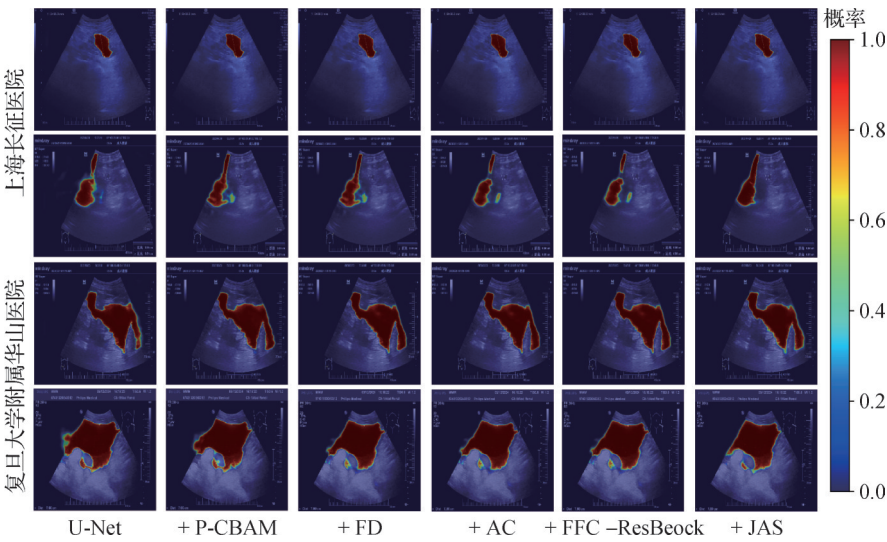


图6 消融实验可视化

Fig. 6 Ablation experiment visualization

实验结果表明,随着各模块的逐步引入,模型在边界刻画能力和信息获取能力方面均有显著提升,腹腔出血区域的分割效果逐步增强。最终,与原始U-Net模型相比,本文提出的模型在两个测试集上的Dice指标分别提高了0.066 7和0.107 2,充分体现了各模块对模型分割性能的积极贡献。

2.4.4 图像特征可视化

为了更加直观地展示FFC-ResBlock模块中核心算子频域卷积(spectral convolution)的功能(章东平等,2025)。对经过空间域卷积、频域卷积以及双域协同处理前后的特征图进行了频谱可视化,如图7所示。从结果中可以发现,原始频谱呈现典型的超声图像频谱特征,中心区域为明亮的黄绿色(高频能量集中区),向外辐射逐渐过渡到蓝色背景。经

过空间域卷积处理后的频谱图中心能量集中区被扩散和拉长,形成沿网格方向分布的明亮十字形线条,背景噪点的视觉强度相较于原始图像呈现出明显减弱。而经过频域卷积处理后的频谱图像,中心区域被深色方块滤除的低频成分取代,背景噪点显著减少,仅保留外围的明亮线条和斑块(增强的高频成分)。这表明频域卷积能够抑制低频信息,去除了均匀组织的冗余信息,减少背景干扰,同时突出了边缘和纹理的频谱能量(如血肿边界、积液分层)。双域协同处理后的频谱图像综合了前述的特点,中心能量被有效抑制,但保留了空间域处理带来的结构性高频线条,同时背景更为纯净。这种量化效果表明,双域处理不仅优化了信噪比,更通过频谱能量的重新分配,显著增强了

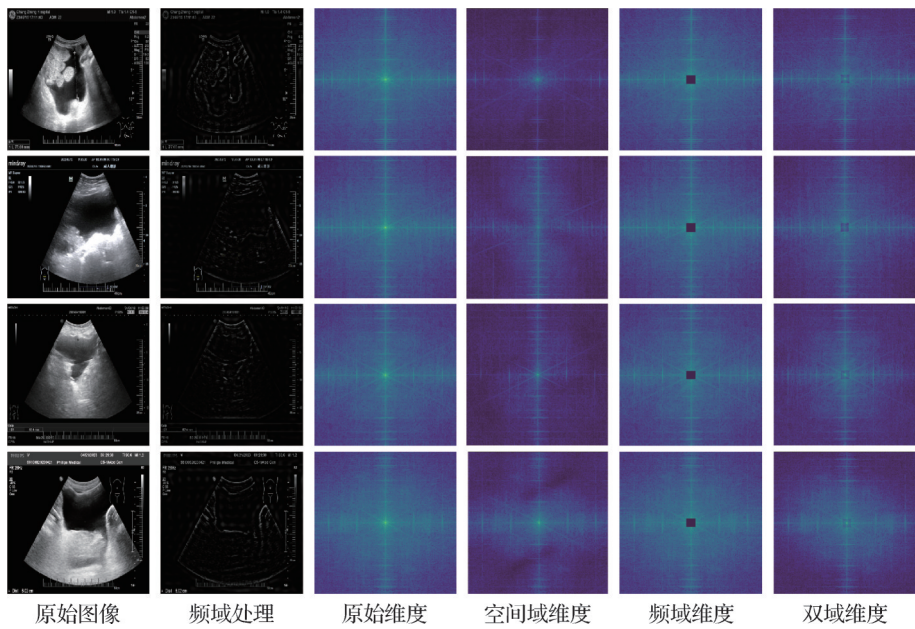


图 7 频域卷积作用结果可视化

Fig. 7 Visualization of the results of convolution in the frequency domain

病灶边界、组织异质性等诊断关键特征的频谱响应灵敏度。

2. 4. 5 模型复杂度分析

模型复杂度分析如表 4 所示。尽管本文 DMA-

UNet 模型的参数量 (19. 27 M) 和 FLOPs (161. 46 G) 并非最低, 例如高于轻量级模型 MobileUNet (参数量 3. 45 M, FLOPs 3. 311 G), 但这种适度的复杂度增加能够带来性能的显著提升。

表 4 模型复杂度分析 (上海长征医院测试集)

Table 4 Analysis of model complexity (test set of Shanghai Changzheng Hospital)

模型	参数量/M	Dice	mIoU	FLOPs/G	推理延迟/ms
U-Net	17.26	0.813 0	0.706 6	160.76	71.46
MobileUNet	3.45	0.857 9	0.773 5	3.311	9.35
Swin-UNet	27.16	0.773 0	0.650 5	5.916	13.52
TransUNet	105.91	0.814 1	0.706 9	129.29	77.18
UNet++	19.83	0.845 2	0.752 0	147.62	63.75
Attention-UNet	18.21	0.831 3	0.750 3	138.47	58.33
DMA-UNet (本文)	19.27	0.879 7	0.796 1	161.46	66.44

DMA-UNet 的 Dice 相比 MobileUNet 提高了约 2. 18%。同时, 推理延迟 (66. 44 ms) 低于 U-Net 和 TransUNet, 与 UNet++、Attention-UNet 较为接近, 但是分割精度同样有显著的提升, 展现了其在效率和精度之间的良好平衡。

综上所述, DMA-UNet 通过优化模型结构, 在保持可接受复杂度的前提下, 实现了最先进的分割性能, 验证了其在实际应用中的有效性和实用性。

2. 4. 6 局限性与不足

上述实验结果表明了 DMA-UNet 的有效性和优越性。然而, 本文模型在面对伪影和边界模糊等因素干扰时, 仍然存在分割错误的情况, 如图 8 所示。具体地, 如图 8 第 1 行和第 3 行所示, 出血区域与周围组织的灰度差异较小 (低对比度), 且出血可能呈现弥散性弱回声, 模型将邻近正常组织误判为出血, 导致本文模型在定位出血区域时出现误判, 进而产生了错误分割。又如图 8 第 2 行所示, 由于腹腔出血

与周围组织的声阻抗差异导致回声渐变,模型未能准确捕捉边界过渡特征,导致边界信息的挖掘出现较大失误。另外,从图8的第4行中可以观察到,由于超声的物理伪影特性未被模型充分学习,尤其在腹腔积液或气体干扰下伪影更显著,进而导致将超声伪影(如混响伪影或声影)误识别为出血点等现象。基于上述失败案例,计划在之后的工作中,进一步探究在伪影和边界模糊影响下的腹腔出血区域的精准分割模型。

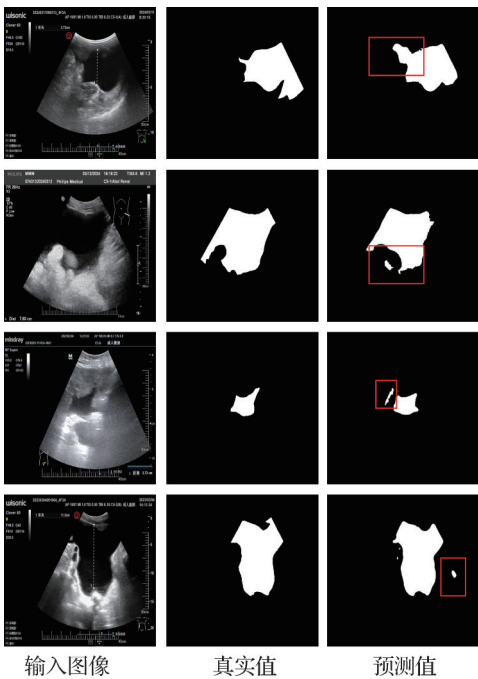


图8 失败案例展示
Fig. 8 Display of failure cases

3 结 论

本文提出的DMA-UNet模型旨在解决腹腔出血超声图像分割这一关键挑战。通过以AAConv Block为核心构建编码器,有效缓解下采样过程中的细节丢失的问题以及增强小目标特征捕获的能力,并采用FFC模块连接编码器与解码器,实现了空间域和频域的双重信息提取和融合,显著提升了特征提取精度,有效减少了因下采样导致的细节损失。在跳跃连接部分引入的JAS联合注意力机制,能够精准聚焦于出血病灶区域,抑制无关背景干扰。这些模块协同工作,使模型在保持较高分割精度的同时,对边界模糊与形态多变的病灶也表现出良好的鲁棒性。

然而,本文仍存在一定局限性。在极端复杂病例(如严重边界模糊或病灶高度不规则)中,其分割精度仍有提升空间。因此未来工作将聚焦于:1)优化模型架构以提高鲁棒性;2)开发轻量化版本以降低计算成本;3)针对复杂边界和多尺度特征进行专项改进;4)探索模型在其他医学图像分割任务中的迁移应用潜力,以验证其泛化性能。

参考文献(References)

Cao H, Wang Y Y, Chen J, Jiang D S, Zhang X P, Tian Q, et al. 2023. Swin-Unet: Unet-like pure transformer for medical image segmentation//Karlinsky L, Michaeli T, Nishino K, eds. Computer Vision—ECCV 2022 Workshops. Tel Aviv, Israel: Springer: 205-218 [DOI: 10.1007/978-3-031-25066-8_9]

Chen J N, Lu Y Y, Yu Q H, Luo X D, Adeli E, Wang Y, et al. 2021. TransUNet: transformers make strong encoders for medical image segmentation [EB/OL]. [2025-10-20]. <http://arxiv.org/pdf/2102.04306.pdf>

Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 801-818 [DOI: 10.1007/978-3-030-01234-2_49]

Chen L J, Qiao C K, Wu M J, Cai L H, Yin C, Yang M K, et al. 2023. Improving the segmentation accuracy of ovarian-tumor ultrasound images using image inpainting. Bioengineering, 10 (2) : #184 [DOI: 10.3390/bioengineering10020184]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. 2020. An image is worth 16 × 16 words; Transformers for image recognition at scale [EB/OL]. [2025-10-20]. <https://arxiv.org/pdf/2010.11929.pdf>

George M and Anita H B. 2022. Analysis of kidney ultrasound images using deep learning and machine learning techniques: a review//Ranganathan G, Bestak R, Palanisamy R, Rocha Á, eds. Pervasive Computing and Social Networking: Proceedings of ICPCSN 2021. Singapore: Springer: 183-199 [DOI: 10.1007/978-981-16-5640-8_15]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Computer Society: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Huang H M, Lin L F, Tong R F, Hu H J, Zhang Q W, Iwamoto Y, et al. 2020. Unet 3+: a full-scale connected UNet for medical image segmentation//Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE: 1055-1059 [DOI: 10.1109/ICASSP40776.2020.9053405]

- Jaderberg M, Simonyan K, Zisserman A and Kavukcuoglu K. 2015. Spatial transformer networks//Proceedings of the 29th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press: 2017-2025 [DOI: 10.5555/2969442.2969465]
- Kristensen J K, Buemann B and Kuehl E. 1971. Ultrasonic scanning in the diagnosis of splenic haematomas. *Acta Chirurgica Scandinavica*, 137(7): 653-657
- Lei H J, Tong G J, Su H Q and Lei B Y. 2024. Weak-supervised attention fusion network for carotid artery vessel wall segmentation//Proceedings of the 27th International Conference on Medical Image Computing and Computer Assisted Intervention. Marrakesh, Morocco: Springer: 349-358 [DOI: 10.1007/978-3-031-72378-0_33]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, et al. 2021. Swin transformer: hierarchical vision transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]
- Milletari F, Navab N and Ahmadi S A. 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation//Proceedings of the 4th International Conference on 3D Vision (3DV). Stanford, USA: IEEE: 565-571 [DOI: 10.1109/3DV.2016.79]
- Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, et al. 2018. Attention U-Net: learning where to look for the pancreas [EB/OL]. [2025-10-20].
<https://arxiv.org/pdf/1804.03999.pdf>
- Qu G, Song M M, Xin G F, Shang Z Q and Sun L M. 2024. Time-convolutional network with joint time-frequency domain loss based on arithmetic optimization algorithm for dynamic response reconstruction. *Engineering Structures*, 321: #119001 [DOI: 10.1016/j.engstruct.2024.119001]
- 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 International Publishing: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Ruan J C, Li J C and Xiang S C. 2024. VM-UNet: vision mamba UNet for medical image segmentation [J/OL]. *ACM Transactions on Multimedia Computing, Communications and Applications*: #2491. [DOI: 10.48550/arXiv.2402.02491]
- Shi J, Wang T T, Zhu Z Q, Zhao M F, Wang B X and An H. 2025. Deep learning-based medical image segmentation methods. *Journal of Image and Graphics*, 30(6): 2161-2186 (石军, 王天同, 朱子琦, 赵敏帆, 王炳勋, 安虹. 2025. 基于深度学习的医学图像分割方法综述. *中国图象图形学报*, 30(6): 2161-2186) [DOI: 10.11834/jig.240467]
- Soyuncu S, Cete Y, Bozan H, Kartal M and Akyol A J. 2007. Accuracy of physical and ultrasonographic examinations by emergency physicians for the early diagnosis of intraabdominal haemorrhage in blunt abdominal trauma. *Injury*, 38 (5): 564-569 [DOI: 10.1016/j.injury.2007.01.010]
- Valanarasu J M J, Oza P, Hacihaliloglu I and Patel V M. 2021. Medical transformer: Gated axial-attention for medical image segmentation//Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2021). Cham, Switzerland: Springer International Publishing: 36-46 [DOI: 10.1007/978-3-030-87193-2_4]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Yang L X, Zhang R Y, Li L D and Xie X H. 2021. SimAM: a simple, parameter-free attention module for convolutional neural networks//Proceedings of the 38th International Conference on Machine Learning. [s.l.]: PMLR: 11863-11874
- Zhang D P, Li Z, Xie Y G, Wang D Y, Tang S L, Bu Y Z, et al. 2025. Boundary cue deep fusion polyp image segmentation network. *Journal of Image and Graphics*, 30(5): 1479-1496 (章东平, 李铮, 谢亚光, 王都洋, 汤斯亮, 卜玉真, 等. 2025. 边界线索深度融合息肉图像分割网络. *中国图象图形学报*, 30(5): 1479-1496) [DOI: 10.11834/jig.240383]
- 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 International Publishing: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]
- Zhu Z Q, Sun M W, Qi G Q, Li Y Y, Yang M J, Cheng J, et al. 2026. Medical image segmentation integrating frequency adaptation and feature transformation. *Journal of Image and Graphics*, 31 (1): 303-319 (朱智勤, 孙梦薇, 齐观秋, 李媛源, 杨梦杰, 程俊, 等. 2026. 融合频率自适应和特征变换的医学图像分割. *中国图象图形学报*, 31(1): 303-319) [DOI: 10.11834/jig.250100]
- Zhu Z Q, Zhang Z M, Qi G Q, Li Y Y, Li Y Z and Mu L. 2025. A dual-branch network for ultrasound image segmentation. *Biomedical Signal Processing and Control*, 103: #107368 [DOI: 10.1016/j.bspc.2024.107368]

作者简介

王轩,男,硕士研究生,主要研究方向为医学图像处理。

E-mail:M320123110@sues.edu.cn

邢文字,通信作者,男,副研究员,主要研究方向为智慧医疗超声。E-mail:xingwenyu@uhrs.edu.cn

邓娜,女,副研究员,主要研究方向为智能控制和图像处理。E-mail:dengna@sues.edu.cn

谢锋,男,博士研究生,主要研究方向为生物医学工程。

E-mail:fxie22@m.fudan.edu.cn