

doi: 10.3969/j.issn.1674-1242.2025.04.005

基于 FA-UNet 的胸腺病变分割

屈泰宇¹, 王兴², 周雷¹

(1. 上海理工大学健康科学与工程学院, 上海 200093;
2. 北京大学第一医院胸外科, 北京 100034)

【摘要】目的 探讨基于 FA-UNet 模型, 在三维医学影像分割中结合全局位置注意力 (Global Position Attention, GPA) 模块和频域感知 (Frequency Aware, FA) 模块提升病变区域分割准确性的效果。方法 该文提出了 FA-UNet 模型, 在原有 UNet 网络架构的基础上, 在桥接层加入了 GPA 模块和 FA 模块。通过这种改进, 模型能够有效地学习三维医学影像中的位置信息, 且不会显著增加计算负担。实验使用三维 CT 胸腺病变数据集, 并通过 Dice 系数等标准指标评估分割效果。结果 与传统 UNet 相比, FA-UNet 在相同算力下获得了显著的性能提升, 病变区域的 Dice 系数达到 86.16%, 且提升效果为增加算力的 3.44 倍。结论 FA-UNet 通过引入 GPA 模块和 FA 模块, 能够在提高分割精度的同时优化计算效率, 是一种有效的三维医学影像分割方法。

【关键词】医学影像分割; 胸腺病变; FA-UNet; 注意力机制; 频域分析

【中图分类号】TP391.41; R445

【文献标志码】A

文章编号: 1674-1242 (2025) 04-0449-08

Segmentation of Thymus Lesions Based on FA-UNet

QU Taiyu¹, WANG Xing², ZHOU Lei¹

(1. School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;

2. Department of Thoracic Surgery, Peking University First Hospital, Beijing 100034, China)

【Abstract】 **Objective** To explore the effectiveness of the FA-UNet model in improving the accuracy of lesion segmentation in three-dimensional medical images by incorporating the global position attention (GPA) block and frequency aware (FA) block. **Methods** This paper proposes the FA-UNet model, which builds upon the original UNet architecture by adding the GPA-block and FA-block in the bridge layer. This modification allows the model to effectively learn positional information in three-dimensional medical images without significantly increasing computational burden. Experiments were conducted using a 3D CT thymus lesions dataset, and segmentation performance was evaluated using metrics such as the Dice coefficient. **Results** Compared to traditional UNet, FA-UNet achieved significant performance improvements with the same computational resources, with the Dice coefficient for lesion segmentation reaching 86.16%, and the performance improvement corresponding to 3.44 times the additional computational cost. **Conclusion** By incorporating the GPA-block and FA-block, FA-UNet can improve segmentation accuracy while optimizing computational efficiency, making it an effective method for 3D medical image segmentation.

收稿日期: 2025-02-10。

基金项目: 国家自然科学基金 (82300122)。

作者简介: 屈泰宇 (2000—), 男, 硕士研究生, 主要从事医学影像处理、计算机视觉方面的研究。邮箱 (E-mail): isqtaiyu@163.com。

【Key words】 Medical Image Segmentation; Thymus Lesions; FA-UNet; Attention Mechanism; Frequency Domain Analysis

0 引言

近年来,随着各类呼吸道传染病流行趋势的增强,胸部CT的普及率逐步提高。医院影像科需要在诊断肺部疾病的同时对胸腺是否存在异常病变进行判断。然而,由于胸腺疾病发病率相对较低,各医院的病例数较少,因此影像科容易忽视对胸腺疾病的判定,从而产生误诊风险^[1-3]。同时,由于胸腺疾病发病率较低,胸腺病变自动分割领域相较于其他疾病领域仍存在较大的进步空间。尽管胸腺疾病发病率较低,但由于我国人口基数庞大,每年仍有数万名新发患者,因此对胸腺病变自动分割算法的探索就显得相当重要。

然而,目前针对胸腺病变自动分割的网络研究较少,其中较为突出的是MG-Net,但由于MG-Net为二维网络,无法有效捕捉切片之间的上下文信息,因此其分割性能并不理想,在本文的数据集上,Dice系数仅为56.60%。这方面研究较少的原因主要有以下两点:一是胸腺疾病的病例数较少,无法得到足够的数据集进行训练;二是当前众多医学影像分割模型都基于卷积神经网络(Convolutional Neural Network, CNN)的UNet^[4]模型,CNN的远程关系建构能力较弱,而医学影像中各个器官的位置关系具有十分重要的定位作用,这就导致这些信息在基于CNN的架构下无法被充分利用。为了获得更多的空间信息,许多研究使用基于Transformer^[5]的注意力信息,在提升效果的同时也极大地增加了算力需求。

因此,本文针对胸腺疾病的数据问题,基于上海交通大学医学院附属上海市第一人民医院胸外科积累的病例开展初步建模和测试,这些病例由该医院与其他多家医院的数据共同构成,均匀地分布在数据集的训练、验证、测试三个部分。这在一定程度上降低了数据量不足对模型可靠性的影响,并确保有足够的数据支持分割模型的训练与效果验证。

在网络方法上,本文提出了频域注意力UNet(Frequency Attention UNet, FA-UNet)模型,并创新性地引入了全局位置注意力(Global Position Atten

tion, GPA)模块和频域感知(Frequency Aware, FA)模块。在UNet编码器和解码器的基础上,本文在桥接层实现了全局信息的提取。三维医学影像主要包含空间位置信息、结构轮廓信息和器官体素特征。传统的CNN模型能够有效提取器官体素特征,但在捕捉空间位置信息和结构轮廓信息方面有局限性。空间位置信息与结构轮廓信息在三维医学影像分析中的重要性不言而喻,在FA-UNet模型中,这些缺失的信息可以通过GPA模块和FA模块进行有效学习,从而提升模型的表现。

在实验方面,本文对比了多种主流网络的性能,包括经典的基于CNN的分割网络UNet和UXNet^[6],以及基于Transformer的网络,如nnFormer^[7]、UNet++^[8]、UNet^[9]、SwinUNet^[10]。此外,本文比较了近年来基于Mamba框架的广泛关注网络LightM-UNet^[11],以及专门用于胸腺分割的MG-Net^[12]。

除此之外,本文还对提出的GPA模块和FA模块进行了消融实验,并与作为基准的UNet进行对比,比较了它们在Dice系数、参数量、浮点运算量及迭代时间等方面的表现,验证了本文所提模块的有效性。

1 实验方法与材料

1.1 FA-UNet 整体架构

图1展示了FA-UNet的整体架构:在保持原始UNet结构不变的前提下,网络进行了6次下采样以提取细节特征,从而在算力与效果之间取得了良好的平衡。网络输入三维CT影像后,通过UNet主干网络进行6次下采样,每次下采样都会在图像的跳跃连接中反映出通道、深度、宽度和高度的变化。在得到最小编码特征图后,UNet结构会对编码后的特征图进行上采样解码,并与编码过程中不同尺度的特征图进行融合,从而得到最终的分割结果。该架构在众多医学影像分割任务中已被验证高效且具有较好的性能。然而,由于缺乏全局信息,该架构的优势尚未得到充分发挥。

在三维医学影像中,良好的分割效果依赖局部特征信息与全局定位信息的有效融合,而这两者与信号

处理中的高频信号和低频信号具有相似性。在信号处理领域,信号从时域转换到频域进行分析与计算是常见的处理方式。在时域中,高频信号与低频信号常常交织在一起,难以分离与分析,而傅里叶变换通过将信号映射到频域,使每个频率上的信号都清晰可见,从而有助于对信号进行细致的分析^[13]。

三维医学影像分割领域同样如此。如果在图像域中直接对整张图像进行分析,可能无法同时获得物体的整体位置信息与细节信息。然而,如果将图像信息转换到频域并从不同的视角进行分析^[14],则能够同时获得细节信息和全局信息。因此,本文提出FA-UNet,其主干网络基于经大量实验证明高效且实用的UNet结构。与传统UNet不同,本文在桥接层引入了GPA模块和FA模块,在极低的算力消耗下获得了丰富的全局位置信息与局部特征信息。这为注意力机制的应用提供了一种新思路:不必在整个编解码器中全局应用昂贵的注意力机制来提取空间信息,仅在桥接层提取空间信息可能更加高效且可靠。此外,可以将

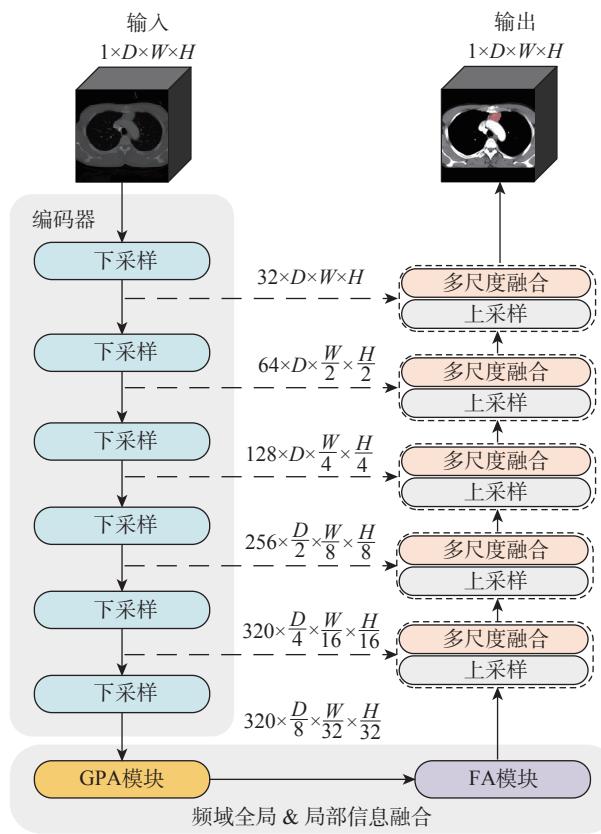


图1 FA-UNet 整体架构
Fig.1 FA-UNet overall architecture

传统的数据处理和变换算法机制引入深度学习中,以便在未曾关注的维度或方向上提取特征信息。

1.2 GPA 模块

本文提出的GPA模块位于桥接层,以提取空间与通道信息,并获取全局位置信息,如图2中的GPA模块所示。在此过程中,空间注意力与通道注意力共享查询(Q)矩阵和键(K)矩阵,但分别为空间和通道赋予不同的值(V)矩阵,从而提取各自的注意力信息。最终,所获得的注意力信息将在通道维度进行融合,并通过卷积操作调整形状,使其与提取前保持一致。GPA模块作为一个即插即用的注意力模块,具有较强的可迁移性。

首先进行特征嵌入(Feature Embedding)操作,将 (C, D, W, H) 的特征整理为 $(C, D \times W \times H)$,即将每个通道视为一个嵌入向量(token),该通道的所有体素作为该token的嵌入维度,令token数量为 C ,维度为 $D \times W \times H$ 。然后进行位置编码,将各特征的位置信息融入token中。

接着分别进行通道注意力和空间注意力的提取。与传统注意力机制相同,本文通过嵌入矩阵获得 Q 矩阵、 K 矩阵和 V 矩阵。与传统注意力机制不同,本文将注意力分为通道注意力和空间注意力,分别计算通道注意力图与空间注意力图,并与原始特征相乘,以关注空间与通道上那些值得注意的位置。因此,本文的 Q 、 K 、 V 矩阵在通道与空间上共享 Q 与 V 的权重,但分配不同的值矩阵 V_s 与 V_c ,得到空间注意力图与通道注意力图,计算公式如下。

$$M_s = \text{SA}(Q, K, V_s) \quad (1)$$

$$M_c = \text{SC}(Q, K, V_c) \quad (2)$$

其中, M_s 、 M_c 分别为空间注意力图与通道注意力图;SA与SC代表空间注意力模块与通道注意力模块。通过计算,得到通道注意力与空间注意力提取后的特征图 X_s 和 X_c ,如下所示。

$$X_s = \text{Softmax} \left(\frac{QK_{\text{proj}}^T}{\sqrt{d}} \right) \times V_s^{\text{proj}} \quad (3)$$

$$X_c = V_c \times \text{Softmax} \left(\frac{\mathbf{Q}^T \mathbf{K}}{\sqrt{d}} \right) \quad (4)$$

其中, Softmax 为激活函数; $\mathbf{K}_{\text{proj}}^T$ 为 \mathbf{K} 矩阵映射的转置; $V_{\text{proj}}^{\text{proj}}$ 为空间 V 矩阵的映射, 映射和转置的目的是满足计算所需的维度要求; d 为嵌入向量的维度。随后, 将经过空间注意力和通道注意力加权后的特征图进行融合, 并将其输入两个连续的 mini UNet 中, 进行多尺度特征整合。接着, 应用卷积核大小为 1 的卷积对通道进行整理, 最终得到 GPA 模块的输入结果, 如下所示。

$$X_a = \text{Conv}_1(\text{Conv}_3(\text{Conv}_3(X_s + S_v))) \quad (5)$$

其中, Conv_1 用于通道整理; Conv_3 用于 mini UNet 多尺度特征整理, 其卷积核大小均为 3; 特征经过 GPA 模块后得到了注意力加权后的特征 X_a , 然后输入 FA 模块, 将 CNN 提取的局部细节信息与 GPA 提取的全局位置信息在频域上进一步提炼。

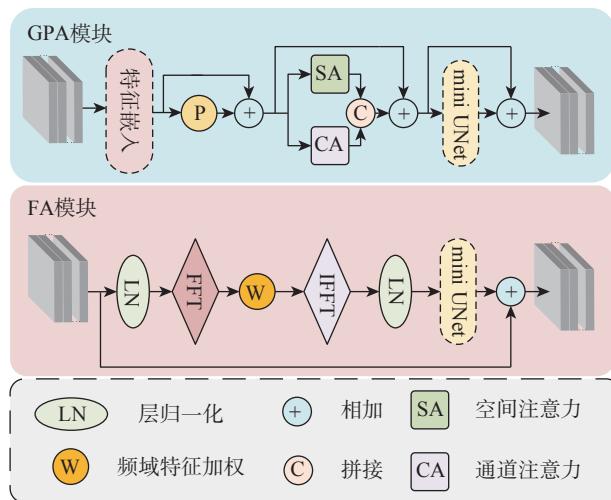


图 2 GPA 模块与 FA 模块

Fig.2 GPA-block and FA-block

1.3 FA 模块

FA 模块在特征图经过 CNN 和 GPA 模块提取局部细节信息与全局位置信息后, 将特征从图像域映射到频域, 进一步提炼更多的高频细节与低频全局信息。如图 2 中的 FA 模块所示, 特征图输入 FA 模块后, 首先经过归一化处理, 接着通过快速傅里叶变换, 将图像域的特征转换到频域。假设输入特征为 F , 形状为 (C, D, W, H) , 三维傅里叶变换的公式如下。

$$F_{\text{fre}} = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} \sum_{z=0}^{D-1} M(x, y, z) e^{-j2\pi(m_x x + m_y y + m_z z)/(HWD)} \quad (6)$$

得到原始特征图 F_{fre} 在频域的特征后, 使用一个可学习的权重矩阵 W 对 F_{fre} 的频域特征进行加权, 以提取最重要的特征信息, 如下所示。

$$F_{\text{fft}}' = F_{\text{fft}} \cdot W \quad (7)$$

此处的乘法并非矩阵乘法, 而是逐像素的权重点乘, 以获得加权后的频域特征信息。接着对 F_{fft}' 采用式(7)中的逆过程将特征从频域映射回图像域, 进行归一化、卷积等操作; 最终与初始特征进行残差连接并求和, 得到最终的输出特征图; 最后进入上采样阶段, 生成分割结果。

1.4 实验数据获取

本文所采用的数据包含 372 例 CT 样本, 均来自前纵隔病变患者。患者年龄范围为 16~83 岁, 均接受了胸部 CT 检查。本文随机使用多种扫描设备, 包括西门子 CT 扫描仪 (Healthineers, 德国) 和飞利浦 CT 扫描仪 (Eindhoven, 荷兰)。扫描参数如下: 螺旋扫描模式, 管电压为 120kVp, 管电流为 300mA, 准直厚度范围为 38.4mm (64mm×0.6mm) 至 80mm (128mm×0.625mm), 重建厚度为 1.0mm, 重建间隔为 0.7mm 或 1.0mm, 旋转速度为 0.6s, 螺距为 1.375。增强 CT 扫描在静脉注射对比剂后进行, 注射速率为 3.5mL/s [1.5mL/kg 体重, 总量为 90~120mL, 使用 300mg/mL 碘海醇 (Omnipaque, GE Healthcare, Waukesha, WI, USA)]。动脉期扫描在注射对比剂后 30s 通过右上肢血管进行。所有 CT 的标注均由胸外科专职医生进行, 医生依据临床经验和影像学特征对病变区域进行精确标注, 确保数据的准确性和可靠性。

在进行数据分析之前, 患者的个人信息已被匿名化。本回顾性研究遵循《赫尔辛基宣言》的原则, 并获得了上海交通大学医学院附属上海市第一人民医院伦理委员会的批准, 所有程序均按照批准的方案进行。此外, 由于本研究具有回顾性, 伦理委员会豁免了每位患者签署书面知情同意书的要求。

1.5 数据预处理

窗宽和窗位是 CT 图像中至关重要的参数, 用于

调节图像的对比度和亮度，以便突出显示特定组织或病变。这对于肿瘤分割任务尤其重要，因为不同组织类型在 CT 图像中的密度范围各异，适当调节窗宽和窗位可以有效提高胸腺异常区域的可见性，减少伪影和噪声，从而提升分割精度。本文通过实验发现，当设定窗位为 65、窗宽为 400 时，相较于其他窗宽、窗位设定，胸腺异常区域的边界更加清晰，更加有助于目标的自动分割。在此窗宽设定下，数据间的差异更加显著，使网络在特征提取过程中能够更有效地分辨不同数据的特征，进而获取更加精准的边界信息。

经过裁剪、重采样和窗宽窗位调整三步预处理后的效果如图 3 所示。图中第一排图像为未经处理的 CT 影像，体素分辨率为 (66, 512, 512)，图像中存在的无关区域（如病人背部）可能干扰自动分割，且胸腺病变区域的边界不明显，难以清晰地分辨。图 3 中第二排图像为预处理后的效果，背部无关区域被有效裁剪，网络的体素分辨率为 (66, 305, 305)。此外，在横切面进行的尺度重采样不仅确保了所有样本具有一致的空间尺度映射值，还有效减少了网络推理过程中的计算量，提高了训练速度。经此处理，胸腺病变区域的边界更加清晰，且能够通过肉眼准确判定。

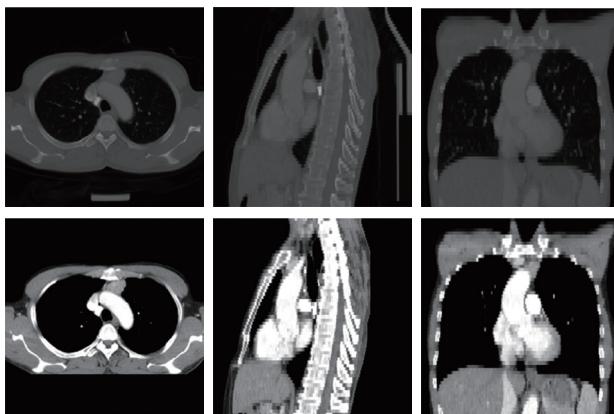


图 3 预处理效果
Fig.3 Preprocessing results

通过上述方法对 372 例胸部 CT 样本进行了预处理，最终将其中的 240 例作为训练集样本，60 例作为验证集样本，72 例作为测试集样本。

2 实验与结果分析

2.1 实验环境与设定

本文采用 Ubuntu 操作系统和 NVIDIA GeForce

RTX 3080 Ti GPU 作为实验平台。为了确保实验结果的一致性和可比性，所有神经网络模型均在相同的超参数配置下进行训练，每个模型的训练周期设定为 500 个 epoch。在模型训练过程中记录两种模型状态：一种是训练结束后的最终模型（Last 模型），另一种是训练过程中性能最优的模型（Best 模型）。所有训练完成后，针对测试集进行评估，选择在测试集上表现最佳的模型作为最终训练结果。此种评估方法能够从每个网络中提取出性能最优的模型，为后续分析和比较奠定坚实的基础。

2.2 评价指标

本文采用了四个关键评价指标来全面评估三维胸腺病变分割模型的性能：Dice 系数、精确率（Precision）、召回率（Recall）和平均对称表面距离（Average Symmetric Surface Distance, ASD）。

Dice 系数是衡量预测分割结果与真实标注之间相似度的指标。通过比较两者的重叠部分与各自体积的比例进行计算，值越接近 1，表示预测结果与真实标注的一致性越好。

精确率是指模型所预测病变区域中实际为病变区域的比例，反映了模型预测结果的准确性。

召回率衡量了模型在所有实际为病变区域的样本中预测正确的比例，反映了模型的覆盖能力，即模型能够识别出多少实际的病变区域。

平均对称表面距离是衡量预测分割表面与真实标注表面之间平均距离的指标，通过计算预测表面与真实表面之间的最小距离得出，值越小表示预测分割的表面精度越高。

通过以上四个指标的综合评估，从多个角度评价分割模型的性能，包括其准确性、精确性、覆盖度和表面精度，从而确保模型在实际应用中能够达到最佳性能。

2.3 分割网络对比实验

本文的分割对比实验涵盖了丰富的对比网络，包括经典的三维医学影像分割奠基之作 UNet、基于 CNN 的大卷积核分割网络 UXNet、大量基于 Transformer 或混合结构的网络（如 UNet++、UNet、SwinUNet、nnFormer），以及自 2024 年以来备受关注的基于 Mamba 的 LightM-UNet 和专门

针对胸腺病变分割设计的二维分割网络 MG-Net。相关实验如表 1 所示。

表 1 分割网络对比实验
Tab.1 A comparative experiment on segmentation networks

网络	Dice 系数 /%	精确率 /%	召回率 /%	ASD/mm
FA-UNet	86.16±4.71	86.59±6.09	88.05±4.69	0.46±0.14
UNet	82.82±7.42	84.85±6.28	85.55±7.56	3.87±10.19
UXNet	79.81±9.52	80.13±8.85	84.85±7.39	5.35±12.08
LightM-UNet	78.53±8.34	82.85±8.01	78.99±9.61	1.31±1.19
UNetr++	72.36±9.28	73.86±8.39	80.12±5.37	6.43±4.12
UNetr	66.01±11.72	67.41±12.87	77.21±11.06	16.56±13.06
SwinUNet	63.62±11.01	66.78±9.32	70.13±12.56	17.26±12.77
nnFormer	59.06±10.7	68.72±13.60	59.69±5.73	12.27±7.32
MG-Net	56.60±14.06	67.42±15.09	58.56±13.36	5.41±3.69

本文提出的 FA-UNet 在 Dice 系数、精确率、召回率和 ASD 四个指标上均取得了最佳值，并且相较于其他网络表现出显著优势。此外，FA-UNet 在 95% 置信度下的表现也极为出色，均获得最佳结果。除精确率的 95% 置信度未在 5 以下外，其他指标均低于 5，表明结果具有高度可信度和较强稳定性。

通过与其他网络的对比可见，基于 CNN 的网络（如 UNet 和 UXNet）依然表现较好，而基于 Transformer 或混合结构的网络在性能上远逊色于 FA-UNet 和基于 CNN 的网络。与此同时，基于 Mamba 的 LightM-UNet 虽然进行了轻量化处理，

但训练时的时间消耗仍然远超其他网络，且其效果差强人意。

图 4 展示了四个表现最优的网络在某一测试样本中的分割结果与 GT 标注情况，分别呈现了三个水平面的分割效果与三维分割示意。从图中可见，FA-UNet 的分割边界相较于其他网络更加清晰、准确，成功分割了样本的上凸部分，并且没有发生下侧的误分割现象，最终取得了最佳的 Dice 系数值。

值得注意的是，MG-Net 是一个二维分割模型。在本实验中，三维医学体积沿着 z 轴逐片处理，每个包含胸腺瘤的二维切片作为独立的训练样本，并使用对应的地面实况掩码进行训练。模型在每个二维平面

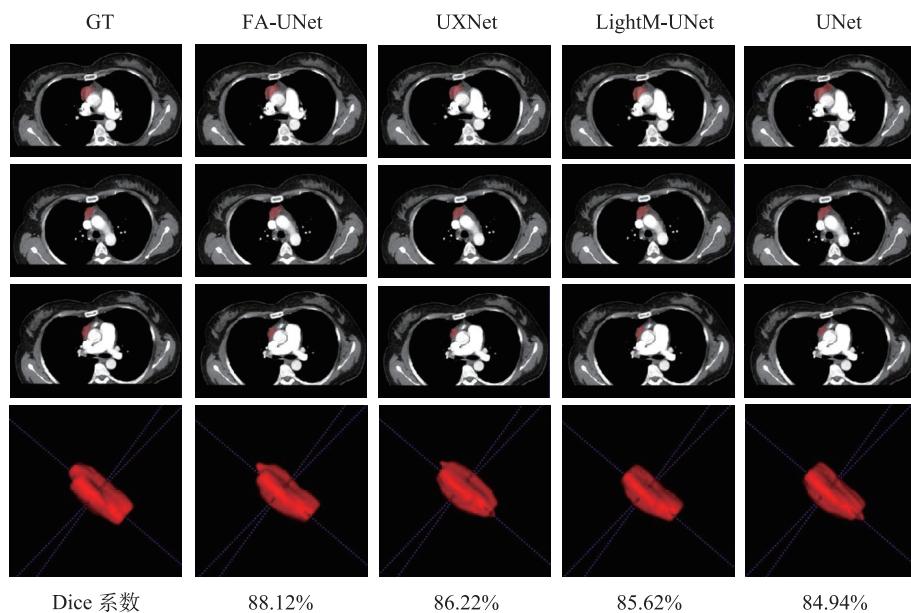


图 4 分割效果示意
Fig.4 Segmentation results illustration

内学习分割胸腺瘤。在推理过程中，三维体积同样沿 z 轴切片，并由训练好的模型对每个切片进行单独处理。最终得到的二维预测结果被重新堆叠成三维体积，得出最终的分割结果。这种方法保留了MG-Net在二维特征提取方面的有效性，同时使其能够处理三维医学影像数据。然而，MG-Net无法捕捉切片之间的上下文信息，导致分割结果之间存在不连续性，进而使其分割性能显著低于其他三维分割模型。

表2 消融实验结果
Tab.2 Ablation study results

桥接层模块	Dice系数/%	Flops/亿次	参数量/百万个	推理时间/ms	时间增长率/%	Dice系数增长率/%	增长比例
GPA	85.04	1.26	11.68	122.30	0.63	2.69	4.27
FA	84.21	0.01	0.10	72.54	0.53	1.68	3.17
GPA+FA	86.16	1.25	11.79	133.92	1.17	4.03	3.44

对Dice系数进行观察，无论是GPA模块、FA模块，还是两者兼具，均显著提升了最终的Dice系数，且在两者兼具的情况下，取得了最佳的Dice效果。此外，三者的使用在极小的时间增长率下，获得了极大的Dice系数增长率，其增长比例均大于1，这意味着这两个模块的增加具有较高的性价比，并且它们的增长比例超过3，表明它们在提高效率方面的表现极为突出。

值得注意的是，FA模块在浮点算力需求和参数量方面表现出极小的消耗，表明此模块能够在非常低的算力消耗下，将胸腺病变分割任务中的图像域特征信息转换到频域进行提取和整理，并能够取得非常优秀的效果。

3 结论

本文提出的FA-UNet模型通过引入GPA模块和FA模块，显著提升了三维胸腺病变分割的性能。传统的UNet模型虽然能够有效提取器官的体素信息，但缺乏对全局位置信息和局部细节信息的充分建模，限制了其在复杂医学影像中的分割效果。为了解决这一问题，FA-UNet在编码器和解码器的桥接部分引入了GPA模块与FA模块，从而在低计算消耗下实现了对图像全局位置信息和局部细节信息的双重捕捉。

相较于传统UNet网络，FA-UNet网络在扩展深度的同时，融入了具有高效性与特征形状不变

2.4 消融实验

对本文所提的GPA模块与FA模块进行消融实验，探究它们在推理过程中的浮点算力(Flops)需求、参数量、推理时间，以及加入模块后相对于整个网络的时间增长率和Dice系数增长率，最终得到它们的Dice系数增长率/时间增长率的增长比例，也就是模块使用的效率评估指标：增长比例。消融实验结果如表2所示。

性的即插即用型GPA模块和FA模块。这一设计使各类网络均可通过添加此类模块显著提升性能。与对比实验中的其他网络相比，FA-UNet的网络结构更加明晰，效果更优。此结果表明，在医学影像分割任务面临数据稀缺的情况下，过度依赖基于Transformer的主干网络尽管可以大幅增加训练与推理时长，但也未必能实现性能的等比例提升。相反，轻巧易用的模块或许是更有效的选择。

虽然FA-UNet在胸腺分割中表现优异，但其对复杂病变区域的细微边界处理仍有改进空间。实践中尚未对其他病变分割进行实验以验证FA-UNet的泛用性，因此未来可结合Mamba等新兴技术，进一步增强模型对多尺度特征的捕获能力，并探索更高效的轻量化设计，以降低计算成本。同时，未来可开展针对其他疾病病变区域的实验研究，以深入探究FA-UNet是否具备广泛的适用性与通用性。

参考文献

- [1] Multidisciplinary Committee of Oncology, Chinese Physician Association. Chinese guideline for clinical diagnosis and treatment of thymic epithelial tumors (2021 edition)[J]. *Chinese Journal of Oncology*, 2021, 43(4): 395-404.
- [2] ENGELS E A. Epidemiology of thymoma and associated malignancies[J]. *J Thorac Oncol*, 2010, 5(10 Suppl 4): S260-S265.
- [3] FANG W, FU J, SHEN Y, et al. Management of thymic tumors-consensus based on the Chinese Alliance for Research in Thymomas Multi-in-

- stitutional retrospective studies[J]. *J Thorac Dis*, 2016, 8(4): 641-645.
- [4] Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation[C]// International conference on medical image computing and computer-assisted intervention. Cham: Springer International Publishing, 2016: 424-432.
- [5] VASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st Conference on Neural Information Processing Systems(NIPS 2017). Long Beach, CA: Neur IPS Foundation, 2017: 5998-6008.
- [6] PATEL K, KOVALYOV A, PANAHY I. UX-Net: Filter-and-process-based improved U-Net for real-time time-domain audio separation[C]// Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island: IEEE, 2023: 1-5.
- [7] ZHOU H Y, GUO J S, ZHANG Y H, et al. nnFormer: Volumetric medical image segmentation via a 3D Transformer[J]. *IEEE Trans Image Process*, 2023, 32: 4036-4045.
- [8] SHAKER A M, MAAZ M, RASHEED H, et al. UNETR++: Delving into efficient and accurate 3D medical image segmentation[J]. *IEEE Trans Med Imaging*, 2024, 43(9): 3377-3390.
- [9] HATAMIZADEH A, TANG Y C, NATH V, et al. Unetr: Transformers for 3D medical image segmentation[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision(2022). Vancouver: IEEE 2022: 574-584.
- [10] CAO H, WANG Y Y, CHEN J, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation[C]// KARLINSKY L, MICHAELEI T, NISHINO K, et al. Computer Vision-ECCV 2022 Workshops. Lecture Notes in Computer Science, vol 13803. Cham: Springer Nature Switzerland, 2023: 205-218.
- [11] LIAO W, ZHU Y, WANG X, et al. Light M-UNet: Mamba assists in Lightweight UNet for medical image segmentation[J]. *arXiv*, 2024 Mar 11. arXiv:2403.05246.
- [12] LI J Y, SUN W F, KAREN M, et al. MG-Net: Multi-level global-aware network for thymoma segmentation[J]. *Comput Biol Med*, 2023, 155: 106635.
- [13] OPPENHEIM A V, SCHAFER R W. Discrete-time signal processing[M]. Edinburgh: Pearson Education Limited, 1999.
- [14] CHI L, JIANG B, MU Y. Fast fourier convolution[J]. *Neur IPS*, 2020, 33: 4479-4488.
- [15] CHEN Y, QUAN L, LONG C, et al. Brain tumor segmentation based on the U-NET++ network with efficientnet encoder[J]. *Technol Health Care*, 2024, 1878-7401.
- [16] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces[J]. *arXiv*, 2024 May 31. arXiv:2312.00752.
- [17] HUANG J, RIZK N P, TRAVIS W D, et al. Comparison of patterns of relapse in thymic carcinoma and thymoma[J]. *J Thorac Cardiovasc Surg*, 2009, 138(1): 26-31.
- [18] TSAO M S, NICHOLSON A G, MALESZEWSKI J J, et al. Introduction to 2021 WHO classification of thoracic tumors[J]. *J Thorac Oncol*, 2022, 17(1): e1-e4.