Journal of Image and Signal Processing 图像与信号处理, 2025, 14(4), 443-456
Published Online October 2025 in Hans. https://www.hanspub.org/journal/jisp
https://doi.org/10.12677/jisp.2025.144041

Hans汉斯

HFF-UNet：基于混合特征融合的轻量化医学图像分割网络

蒋林烽[1]，卢洪轩[1]，张耀严[2]，徐爱茹[1]，余子怡[1]，姚兴兴[1*]

[1]武汉工程大学数理学院，湖北 武汉
[2]武汉工程大学光电信息与能源工程学院，湖北 武汉

收稿日期：2025年8月29日；录用日期：2025年9月19日；发布日期：2025年9月30日

摘 要

医学图像分割在疾病诊断和治疗规划中发挥着关键作用，但现有方法如U-Net及其变体仍面临全局上下文建模不足、特征融合效率低和模型复杂度高等挑战。为此，本文提出一种轻量化医学图像分割网络HFF-UNet (Hybrid Feature Fusion U-Net)，通过三个创新模块显著提升分割性能：首先，设计高效多尺度注意力模块以增强多尺度特征表达能力；其次，提出金字塔池化激励模块改进跳跃连接，缓解编码器与解码器间的语义差距；最后，开发混合特征融合模块优化特征融合过程，提升细节恢复能力。在Glas和CVC-ClinicDB公开数据集上的实验结果表明，与原始U-Net算法相比，Dice系数分别提升1.80%和2.05%，IoU分别提升2.06%和1.89%，参数量分别降低92.20%，82.45%，计算量分别降低90.89%和73.48%，并且显著优于现有轻量化模型。消融实验进一步验证了各模块的有效性。本研究为复杂医学图像分割提供了一种精度与效率兼顾的解决方案，具有重要的临床应用价值。

关键词

医学图像分割，轻量化网络，注意力机制，特征融合，U-Net

HFF-UNet: A Lightweight Hybrid Feature Fusion Network for Medical Image Segmentation

Linfeng Jiang[1], Hongxuan Lu[1], Yaoyan Zhang[2], Airu Xu[1], Ziyi Yu[1], Xingxing Yao[1*]

[1]School of Mathematics and Physics, Wuhan Institute of Technology, Wuhan Hubei
[2]School of Optical Information and Energy Engineering, Wuhan Institute of Technology, Wuhan Hubei

Received: August 29, 2025; accepted: September 19, 2025; published: September 30, 2025

*通讯作者。

文章引用: 蒋林烽, 卢洪轩, 张耀严, 徐爱茹, 余子怡, 姚兴兴. HFF-UNet: 基于混合特征融合的轻量化医学图像分割网络[J]. 图像与信号处理, 2025, 14(4): 443-456. DOI: 10.12677/jisp.2025.144041

Abstract

Medical image segmentation plays a pivotal role in disease diagnosis and treatment planning. However, existing methods such as U-Net and its variants still face challenges including insufficient global context modeling, inefficient feature fusion, and high model complexity. To address these issues, this paper proposes a lightweight medical image segmentation network, HFF-UNet (Hybrid Feature Fusion U-Net), which significantly improves segmentation performance through three innovative modules: 1) an efficient multi-scale attention module to enhance multi-scale feature representation; 2) a Pyramid Pooling Excitation Module to refine skip connections and mitigate the semantic gap between the encoder and decoder; and 3) a Hybrid Feature Fusion Block to optimize the feature fusion process and improve detail recovery. Experimental results on the publicly available GlaS and CVC-ClinicDB datasets demonstrate that, compared to the original U-Net algorithm, the proposed method achieves improvements in Dice coefficients of 1.80% and 2.05%, and in IoU of 2.06% and 1.89%, respectively, while reducing the number of parameters by 92.20% and 82.45%, and computational costs by 90.89% and 73.48%. Moreover, HFF-UNet outperforms existing lightweight models significantly. Ablation studies further validate the effectiveness of each module. This study provides a balanced solution for accuracy and efficiency in complex medical image segmentation, offering substantial clinical application value.

Keywords

Medical Image Segmentation, Lightweight Network, Attention Mechanism, Feature Fusion, U-Net

1. 引言

医学图像分割是医学影像分析的核心技术，在疾病诊断、治疗规划和手术导航中发挥着关键作用。近年来，深度学习技术的突破显著提升了分割性能。2015 年 Ronneberger 等人[1]提出的一种基于卷积神经网络(Convolutional Neural Network, CNN)的 U-Net 架构凭借其编码器–解码器架构和跳跃连接设计，有效提升了分割的精度与准确性，成为医学图像分割的经典模型。然而，U-Net 的局部卷积操作限制了提取全局上下文信息的能力，编码器解码器的特征融合存在语义差距。后续改进方法，如 ResUNet [2]和 UNet++ [3]，分别通过残差连接和嵌套结构提高了特征提取能力，缓解了冗余特征问题。TransUNet [4]通过结合 Transformer 和 U-Net，利用自注意力机制增强了全局特征建模能力，提升了分割精度。然而，TransUNet 面临跳跃连接冗余信息和高计算复杂度的问题。SwinUNet [5]结合了 Swin Transformer [6]和 U-Net 结构，通过滑动窗口机制提升了全局上下文建模能力，但对小目标的分割效果有限。

尽管 U-Net 及其变体在医学图像分割领域取得了显著进展，但现有方法仍存在局限性。本文针对现有医学图像分割方法在参数量过大、编码器特征提取能力有限、跳跃连接冗余、编码器和解码器特征融合的语义差距等方面存在的不足[7]，提出了一种权衡模型参数量大小与分割效果的轻量化医学图像分割网络结构 HFF-UNet，在两个公开数据集上取得了良好的分割效果。本文主要工作如下：

1) 在编码器解码器中引入注意力机制，采用高效多尺度注意力(Efficient Multi-Scale Attention, EMA)模块[8]，在保持较低的计算复杂度和参数量的同时，高效捕获通道和空间维度的交互信息。

2) 为缓解编码器与解码器的语义差距，在跳跃连接部分设计基于注意力的特征传递模块(Pyramid Pooling-based Excitation Module, PPEM)。对来自编码器的特征图进行筛选，动态调整特征权重。

3) 在解码器端与跳跃连接传递的特征图进行融合时，以传统的通道拼接为基础，提出一种新的以跳跃连接传递特征图引导的混合特征融合模块(Hybrid Feature Fusion Block, HFFB)，利用传递而来的空间、语义信息指导解码器上采样逐步恢复分辨率，有效提升对目标细节信息的建模。

4) 在 Glas 和 CVC-ClinicDB 两个公开数据集上开展实验。实验结果表明本文提出的方法能够权衡参数量与分割精度，具有不错的分割性能。并通过消融实验验证了提出模块的有效性。

2. 相关工作

2.1. U-Net 网络

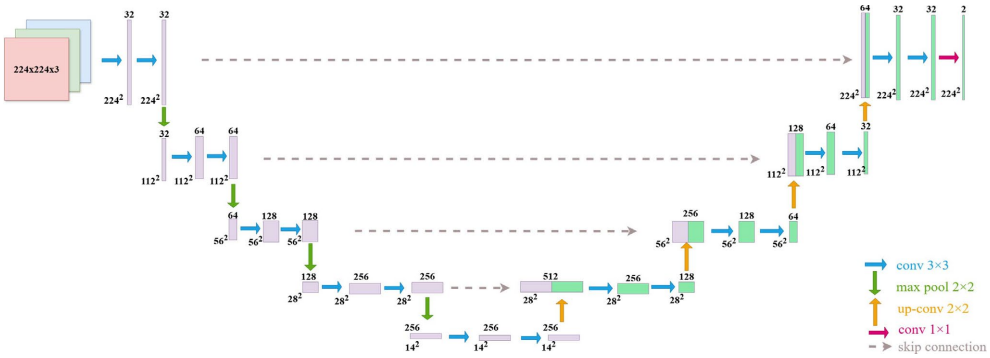

Figure 1. U-Net network architecture
图 1. Unet 网络结构

U-Net 自提出以来，就广泛应用于生物医学图像分割，整体模型呈 U 形结构，如图 1 所示。其独特的对称编码器–解码器架构和跳跃连接设计，使得网络能够有效地融合细粒度特征与高级语义特征[1]。然而，U-Net 的局部卷积操作限制了提取全局上下文信息的能力。为了改进这些局限，后续基于 U-Net 的改进方法，如 ResUNet [2]和 MultiRes U-Net，引入残差连接提高了特征提取能力，缓解了冗余特征问题。UNet++进一步通过嵌套结构减少了编码器和解码器之间的语义差距，在不同层次上实现更有效的特征融合，显著提升了小目标分割精度，但增加了计算复杂度[3]。

近年来，为开发轻量化的网络模型，减少参数量和计算复杂度。Valanarasu 等人[9]在 U-Net 的网络基础上，提出 Tokenized MLP 块和轴向移位操作，结合卷积和多层感知机的优点，设计了一种高效、快速且轻量级的分割网络 UNeXt。其中两个 MLP 块之间利用了深度可分离卷积(Depthwise Separable Convolution)编码位置信息。深度可分离卷积最早于 2017 年由 Google 提出[10]，在 MobileNets[11]网络中得到广泛应用。Dinh 等人[12]在 2023 年提出的 ULite 基于深度可分离卷积原理设计，进一步减少参数量，展现了轻量级 CNN 在医学图像分割领域的潜力。此外，Ruan 等人[13]提出了 Malunet，是一种用于皮肤病变分割的多注意力轻量化 U-Net 结构，该研究验证了将注意力机制与轻量化架构相结合的有效性。

2.2. 注意力机制

Wang 等人[14]率先将自注意力机制引入计算机视觉任务中，提出了 Non-Local 模块，通过非局部操作捕获长距离依赖关系。注意力机制通过加权增强关键特征，在医学图像分割中得到广泛应用。在跳跃连接方面，Attention U-Net [15]在跳跃连接中引入注意力门控，利用解码器端的上采样特征作为门控信号，对编码器输出的特征进行重新校准和筛选，在腹部胰腺 CT 图像分割任务中获得优异结果。Schlemper 等

人[16]则设计了一种注意力门控网络，通过注意力门模块选择性地关注目标区域并过滤无关区域，在超声波扫描平面检测数据集上表现突出。这些工作主要聚焦于利用解码器信息对跳跃连接传递的编码器特征进行加权选择，但融合方式相对直接，且较少考虑特征图内部的多尺度上下文信息。在解码器特征利用方面，2021 年，Petit 等人[17]提出 U-Transformer，采用基于自注意力和交叉注意力的 Transformer 结构，构建 U 形网络，克服了 U-Net 在建模长距离上下文交互和空间依赖方面的不足。此外，Sinha 和 Dolz [18]提出了一种多尺度自引导注意力机制，旨在解码过程中自适应地整合多尺度特征。Ruan 等人[13]同样在皮肤病变分割的轻量化网络 Malunet 中集成了多注意力机制，验证了在轻量化架构中嵌入注意力模块(包括空间和通道注意力)以提升特征表示的有效性。

2.3. Transformer 在医学图像分割中的应用

近年来，Dosovitskiy 等人在 2020 年提出的视觉 Transformer (Vision Transformer, ViT)通过将全局自注意力机制直接应用于全尺寸图像。随着 Transformer [19]在计算机视觉领域的成功，许多研究者将 Transformer 引入医学图像分割任务。Chen 等人提出的 TransUNet 是首个基于 Transformer 的医学图像分割框架，基于 ViT 设计，处理图像特征序列以建模全局依赖，弥补 CNN 局部感受野的不足。Valanarasu 等人[20]提出门控轴向注意力模型 MedT，并采用局部全局训练策略，同时考虑医学图像中的细节信息和远程依赖关系。Liu 等人在 2021 年提出的基于滑动窗口自注意力机制的 Swin transformer，引入层次化特征表示和移位窗口自注意力机制，解决 ViT 在图像分割任务中的限制。随后，Cao 等人受到 Swin Transformer 的启发，首次提出纯 Transformer 的 U 形架构 Swin-Unet，用 Swin Transformer 模块替代 U-Net 中的卷积块。这些方法确实能够建立起有效的长程依赖关系，并且取得了不错的效果。然而，仍然面临着冗余计算和过高算力消耗。为解决这些挑战，近期研究致力于将注意力机制与 Transformer 架构相融合。例如，Sun 等人[21]提出的 DA-TransUNet，通过将空间与通道双重注意力机制与 Transformer U-Net 架构相结合，聚焦空间和通道特征重要性。尽管这些 Transformer 方法在全局建模上表现出色，但其计算开销通常较大，且对跳跃连接和解码器融合策略的轻量化与精细化设计关注相对较少。

3. 本文方法

3.1. 整体架构

图 2 展示了本文的整体架构。该网络基于经典的编码器解码器架构，整体架构设计呈 U 形结构。将编解码器中普通的卷积替换为深度可分离卷积。编解码器加入轻量化的 EMA 模块，高效地捕获通道和空间维度的交互信息，增强模型特征提取能力。在跳跃连接部分，PPEM 模块可以对来自编码器的特征图进行筛选，动态调整特征权重，抑制无关信息，缓解冗余特征问题。解码器部分加入的 HFFB，该模块利用经过 PPEM 筛选后传递而来的空间、语义信息指导解码器上采样逐步恢复分辨率，能够减小解码器、编码器之间存在的语义差距。

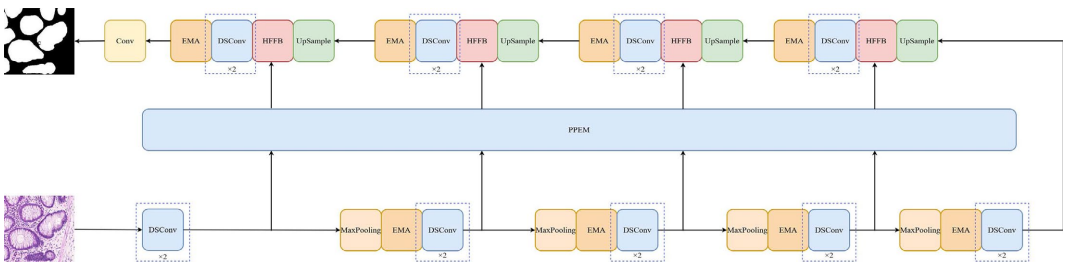

Figure 2. Overall architecture of our method
图 2. 本文整体架构

3.2. EMA 注意力模块

高效多尺度注意力模块[8]是一种新型的注意力机制，用于增强卷积神经网络中的特征表示能力。该模块通过结合特征分组、并行子网络和跨空间学习，在不进行通道维度降维的情况下，高效地捕获通道和空间维度的交互信息，提升模型性能，同时保持较低的计算复杂度和参数量。具体结构如图 3 所示。

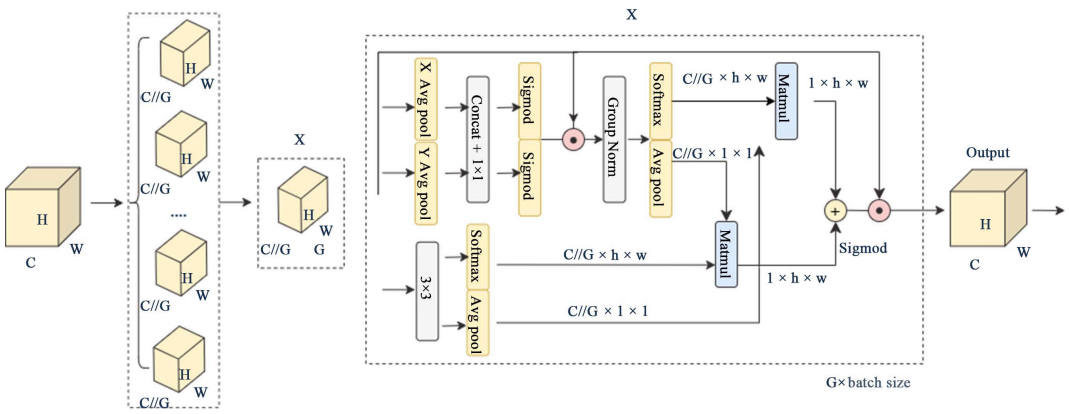

Figure 3. Structure of the EMA
图 3. EMA 模块结构

给定输入特征图 $X \in R^{C\times H\times W}$，其中 C 为通道数，H 和 W 分别为空间高度和宽度，图 3 中，G 表示将输入特征图 X 沿通道维度划分成 G 个子特征组，表示为：

$$X=\left[X_0,X_1,\cdots,X_{G-1}\right] \quad X_i \in R^{C/G\times H\times W} \tag{1}$$

通过将 G 个子特征组重塑到批次维度，此操作避免了通道降维，可以增强对特征的提取能力，同时降低了计算复杂度。EMA 模块包含两个并行分支：1 × 1 分支和 3 × 3 分支，分别处理不同尺度的特征信息。其中 1 × 1 分支，通过两个一维全局平均池化操作捕获空间方向的通道信息，生成水平和垂直方向的特征描述符，由公式(2)表示

$$z_c^H(h)=\frac{1}{W}\sum_{i=0}^{W-1}x_c(h,i),\quad z_c^W(w)=\frac{1}{H}\sum_{j=0}^{H-1}x_c(j,w) \tag{2}$$

其中，$x_c(h,i)$ 和 $x_c(j,w)$ 表示第 c 个通道在空间位置 (h,i) 和 (j,w) 的像素值。两个向量进行拼接后通过 1 × 1 卷积处理，生成通道注意力权重，随后通过 Sigmoid 函数激活，得到通道注意力图：

$$\omega_c=\sigma\left(\mathrm{Conv}_{1\times1}\left[z_c^H,z_c^W\right]\right) \tag{3}$$

其中 σ 为 Sigmoid 函数，$Conv_{1\times1}$ 表示 1 × 1 卷积。3 × 3 分支通过 3 × 3 卷积核捕获多尺度空间信息，扩大感受野，增强局部跨通道交互。

EMA 采用跨空间学习策略，通过矩阵点积操作捕获像素级别的空间关系，有效地增强了特征的聚合。对 1 × 1 分支的输出 $Y_{1\times1}$ 进行二维全局平均池化，生成全局特征描述符：

$$z_c=\frac{1}{H\times W}\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}y_{1\times1,c}(i,j) \tag{4}$$

然后将 $Y_{1\times1}$ 重塑为 $R^{C/G}$，3 × 3 分支的输出 $Y_{3\times3}$ 重塑为 $R^{C/G\times H\times W}$，通过矩阵点积生成第一个空间注意力图。同样对 3 × 3 分支的输出进行二维全局平均池化，将 1 × 1 分支重塑后生成第二个注意力图，最后融

合两个空间注意力图，通过 Sigmoid 函数生成最终特征图，具体公式如下：

$$Y_{EMA}=\sigma\left(\text{Softmax}\left(Y_{1\times1}\cdot Y_{3\times3}\right)+\text{Softmax}\left(Y_{3\times3}\cdot Y_{1\times1}\right)\right)\cdot X, Y_{EMA}\in R^{C\times H\times W} \tag{5}$$

3.3. 跳跃连接改进

受 Attention U-Net 的启发，设计了一种新型基于挤压激励(Squeeze and Excitation, SE)模块[22]与金字塔池化(Pyramid Pooling) [23]改进的跳跃连接特征传递模块 PPEM。如图 4 所示。

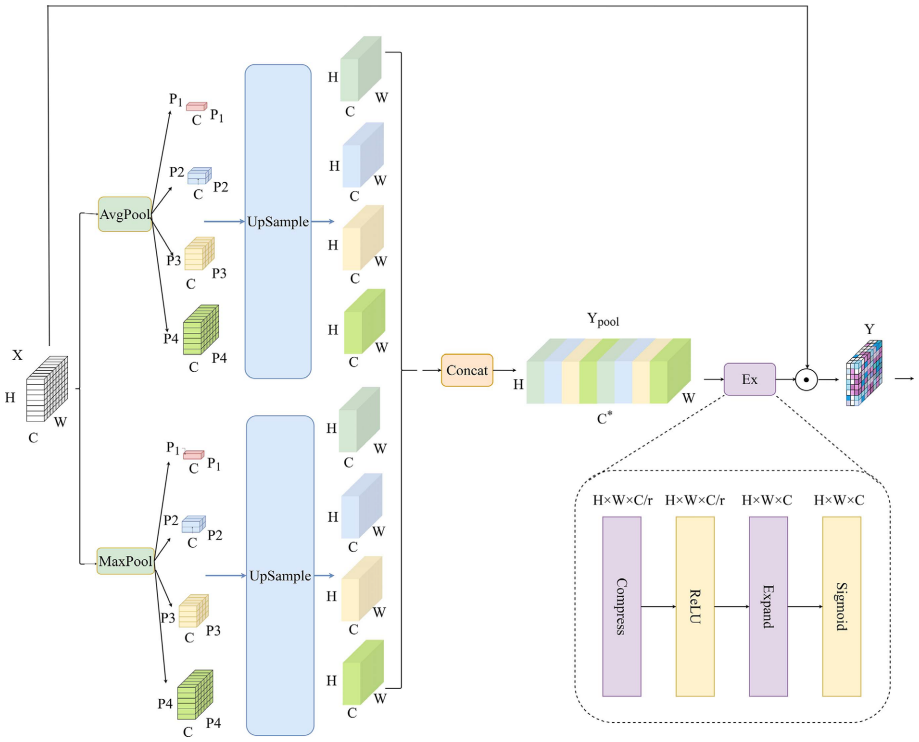

Figure 4. Structure of the PPEM
图 4. PPEM 模块结构

该模块由金字塔池化和通道注意力两个核心组件构成，首先通过金字塔池化包括平均池化和最大池化提取空间上不同尺度的特征，捕获全局和局部上下文信息。通道注意力部分的实现通过降维和升维的 1 × 1 卷积操作，动态调整各通道的重要性。以实现同时在空间和通道上完成特征筛选。

给定输入特征图 $X\in R^{C\times H\times W}$，金字塔池化模块通过一组预定义的池化尺度 $\{P_1,P_2,P_3,\cdots,P_s\}$ 对输入特征图进行池化操作。对于每个池化尺度 $P\in\{P_1,P_2,P_3,\cdots,P_s\}$，特征图被池化为 $R^{C\times P\times P}$，然后通过双线性插值上采样恢复到原始尺寸 $H\times W$，之后将所有池化结果沿着通道维度拼接。表示为：

$$F_{\text{avg}}^{P_i}(X)=AdaptiveAvgPool2d(P_i)(X),\quad F_{\max}^{P_i}(X)=AdaptiveMaxPool2d(P_i)(X) \tag{6}$$

$$Y_{\text{avg}}^{P_i}=Bilinear\left(F_{\text{avg}}^{P_i}(X)[H,W]\right),\quad Y_{\max}^{P_i}=Bilinear\left(F_{\max}^{P_i}(X)[H,W]\right) \tag{7}$$

$$Y_{pool}=Concat\left(\left[Y_{\text{avg}}^{P_1},Y_{\max}^{P_1},\cdots,Y_{\text{avg}}^{P_s},Y_{\max}^{P_s}\right]\right) \tag{8}$$

输出特征图 $Y_{pool}\in R^{(C\cdot2S)\times H\times W}$，$S$ 表示使用的池化尺度数量，$F_{\text{avg}}^{P_i}(X)$ 和 $F_{\max}^{P_i}(X)$ 表示输入特征图使用池化尺度 P_i，经过平均池化和最大池化得到的特征图，$Bilinear$ 表示双线性插值上采样。$Y_{\text{avg}}^{P_i}$，$Y_{\max}^{P_i}$ 分别

表示上采样回原图大小后的空间特征图，Y_{pool} 是经过拼接的结果。通道注意力权重 w_c 的生成过程对应着图 4 所示的激励(Excitation, Ex)过程，首先将线性变换矩阵 W_1 作用到 Y_{pool} 上降低通道维度，以减少参数量。再通过 W_2 将通道维度变换为初始特征图 X 的通道数，最终输出为输入特征图与通道权重的哈达玛积，可以由公式(9)表示：

$$Y = X \odot \sigma\left(\left\{W_2\delta\left(W_1 Y_{pool}\right)\right\}\right) \tag{9}$$

其中，σ 表示 Sigmoid 函数，δ 表示 ReLU 激活函数，$W_1 \in R^{2S\cdot C\times\frac{C}{r}\times H\times W}$，$W_2 \in R^{\frac{C}{r}\times C\times H\times W}$，$r$ 表示降维比率，$\odot$ 表示哈达玛积。最终输出 $Y \in R^{C\times H\times W}$，与原始特征图一致。本文除此以外还提出另两种基于 SE 模块的优化方案，在第 4 节实验部分开展实验进一步讨论。

3.4. 特征融合模块

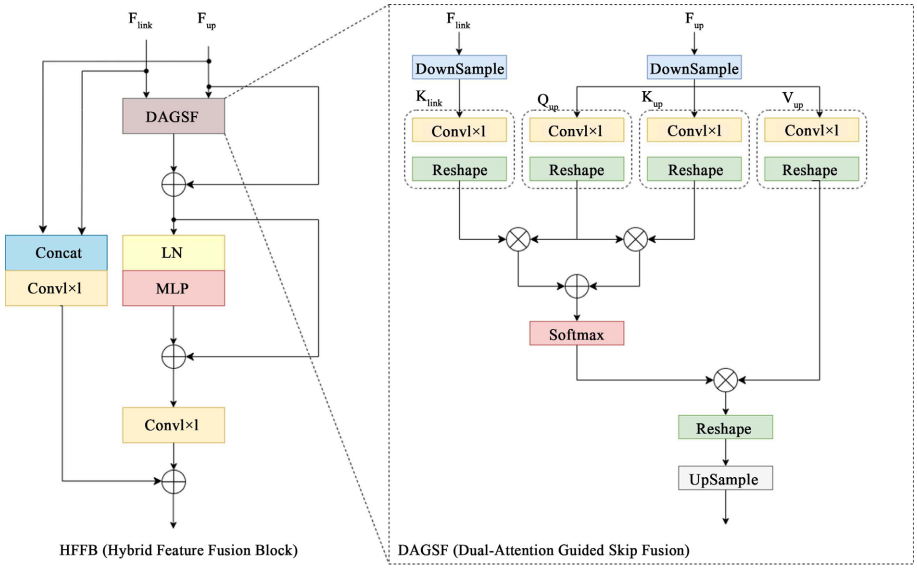

Figure 5. Structure of the HFFB
图 5. HFFB 模块结构

在医学图像分割中，分割场景复杂多样。使用传统 U-Net 的简单通道拼接在面对复杂场景时往往得不到较好的分割效果，基于 Transformer 对建模长程依赖具有优势的认识，本文提出一种以跳跃连接特征图引导的混合特征融合模块 HFFB (Hybrid Feature Fusion Block)旨在改善编码器与解码器的语义不匹配，提高模型的鲁棒性。该模块中主要包含了跳跃连接引导双注意力特征融合 DAGSF (Dual-Attention Guided Skip Fusion)部分、包含跳跃连接与上采样特征图信息的捷径分支以及 Transformer 标准设计的层归一化(Layer Norm, LN)与多层感知机(Multilayer Perceptron, MLP)，整体模块结构如图 5 所示。

DAGSF 模块通过结合自注意力和跨注意力机制，利用跳跃连接传递的特征图作为先验知识，融合编码器与解码器特征，实现对语义差距的减小。具体地，给定特征图 $F_{link} \in R^{B\times C\times H\times W}$、$F_{\text{up}} \in R^{B\times C\times H\times W}$，分别表示跳跃连接传递而来的特征图与上采样的特征图。先进行下采样率为 r 的平均池化，得到下采样后的特征图 F'_{link}，F'_{up}，使用 1×1 卷积生成查询向量 Q，键向量 K_{link}、K_{up} 以及值向量 V，之后进行 Reshape 操作进行多头拆分，上述过程的计算公式表示为：

$$F'_{link} = AvgPool_r\left(F_{link}\right) \in R^{B\times C\times\frac{H}{r}\times\frac{W}{r}},\quad F'_{up} = AvgPool_r\left(F_{up}\right) \in R^{B\times C\times\frac{H}{r}\times\frac{W}{r}} \tag{10}$$

$$Q = Conv_Q\left(F'_{up}\right), K_{up} = Conv_{kv}\left(F'_{up}\right), K_{link} = Conv_{kv}\left(F'_{link}\right), V = Conv_v\left(F'_{up}\right) \in R^{B\times C\times \frac{H}{r}\times \frac{W}{r}}$$
$$Q = G(Q), K_{up} = G\left(K_{up}\right), K_{link} = G\left(K_{link}\right), V = G\left(V\right) \in R^{B\times N_h\times N_{ds}\times D_h}$$

其中 $G(\cdot)$ 表示 Reshape 操作， N_h 是注意力头数， $N_{ds} = H\times W/r\times r$ ， $D_h = C/N_h$ 。

之后，为了让 Q 向量能同时与上采样和跳跃连接特征交互。先把 K_{link} 、 K_{up} 双分支融合得到 K_{fusion} 。随后使用矩阵乘法计算 Q 与融合后的 K_{fusion} 的缩放点积相似度并进行 Softmax 归一化，得到注意力分数后对值向量 V 施加权重，进一步得到下采样后的注意力输出。最后将注意力输出经过重塑、上采样操作，恢复原始特征图大小。

$$K_{fusion} = K_{up} + K_{link} \in R^{B\times N_h\times N_{ds}\times D_h} \tag{11}$$

$$O_{ds} = \text{softmax}\left(\frac{Q\cdot K_{fusion}}{\sqrt{D_h}}\right)\cdot V \in R^{B\times N_h\times N_{ds}\times D_h}, \quad O = U\left(G\left(O_{ds}\right)\right) \in R^{B\times C\times H\times W} \tag{12}$$

其中， $G(\cdot)$ 表示 Reshape 操作， $U(\cdot)$ 表示上采样操作。遵循标准 Transformer 设计，通过残差连接将输入的上采样特征图添加到注意力输出，保留原始信息。之后应用层归一化以稳定和归一化特征，输入到 MLP 中，增加非线性并进一步转换特征。MLP 的结构包含两个线性层和一个 ReLU 激活函数。第二个残差连接从 LN 之前连接到 MLP 之后，经卷积后与包含跳跃连接与上采样特征的捷径分支相加，得到整个 HFFB 模块的输出，具体流程如图 5 的 DAGSF 模块所示。

4. 实验与结果分析

4.1. 数据集

本实验采用 Glas [24] (gland segmentation)数据集和 CVC-ClinicDB [25] (Colorectal Cancer-Clinic Dataset)数据集。Glas 是 MICCAI2015 腺体分割挑战赛的公开数据集，该数据集包含 165 张图像，其中 85 张用于训练，80 张用于测试。每张图像都标注了腺体的边界，以便进行分割任务。CVC-ClinicDB 是一个公开可用的内窥镜图像数据集，该数据集包含 31 段结肠镜检查视频，从中提取了 612 张高质量的图像，分辨率为 348 × 288 像素。

4.2. 评估指标

为了评估模型的性能，使用 Dice 相似系数(DSC)和交并比(IoU)进行度量。Dice 和 IoU 都用于评估标签与预测之间的相似程度。Dice 和 IoU 的数学公式表示如下：

$$\text{DSC} = \frac{2TP}{2TP + FP + FN}, \quad \text{IoU} = \frac{TP}{TP + FP + FN} \tag{13}$$

其中， TP 为真正类(被正确预测的前景像素个数)， FP 为假正类(被错误预测为前景的像素个数)， FN 为假负类(被错误预测为背景的像素个数)。对于 Glas 数据集，使用 5 折交叉验证进行实验，实验结果报告五折的均值和标准差。对于 CVC-Clinic 数据集，由于官方并没有划分数据集，将数据集划分成 459 张训练集、153 张验证集，报告单折的实验结果。

4.3. 实验设置

本实验使用 Python 3.11.8 和 Pytorch 2.2.1，CUDA 版本 12.1，硬件使用单个 12 GB 显存 4070 GPU。对于 Glas 数据集，设置 batch size 大小为 4，调整图像分辨率为 224 × 224。对于 CVC-Clinic 数据集设置 batch size 大小为 8，所有实验使用 Adam 优化器进行训练，其中初始学习率设置为 0.001，采用余弦退火

法进行学习率衰减。采用早停策略进行模型训练，设置最大训练轮次 500，若在连续训练 40 轮后 Dice 系数没有提升，则结束训练。为避免模型训练过拟合，采用随机旋转、随机翻转策略进行数据增强。损失函数采用混合交叉熵损失和 Dice 损失函数，如下：

$$\text{Loss} = \left(\text{CE}_{loss} + \text{Dice}_{loss}\right) \tag{14}$$

其中 CE_{loss} 表示交叉熵损失， Dice_{loss} 表示 Dice 损失。

4.4. 对比实验

本文在 GlaS 和 CVC-ClinicDB 两个医学图像分割基准数据集上进行了系统评测。如表 1 和表 2 所示，我们选取了七种代表性方法进行对比，包括四类经典模型(U-Net, U-Net++, AttU-Net, MultiResUnet)和三种先进轻量化模型(UNeXt, MedT, ULite)。

HFF-UNet 在 GlaS 和 CVC-ClinicDB 数据集上均实现了更优的性能–效率平衡。完整版(1.347 M 参数)在 GlaS 上达 90.15%DSC 和 82.77%IoU，较 U-Net 提升 1.80%/2.06%，同时参数量减少 92.20%，计算量降低 90.89%。在 CVC-Clinic 上提升 2.05%DSC/1.89%IoU，参数量降 82.45%，计算量降 73.48%。参数量/计算量较 Unet++、AttU-Net 等均降低 80%以上。相比更轻量的 UNeXt (0.439GFLOPs)，HFF-UNet (2.802GFLOPs)在 GlaS 上换取 3.38%DSC 和 5.21%IoU 提升。压缩版 HFF-UNetS (0.766M 参数，DSC87.92%)以更少参数较 ULite (0.878M)提升 2.29%DSC 和 3.39%IoU。

Table 1. Comparison of our method with other methods on the Glas dataset
表 1. 本文方法在 Glas 数据集与其他方法的对比

模型	DSC/%	IoU/%	参数量/M	FLOPs/G
Unet	88.35 ± 0.26	80.71 ± 0.24	17.263	30.768
Unet++	89.30 ± 0.64	81.52 ± 1.04	9.16	26.723
AttU_Net	89.39 ± 1.07	81.81 ± 1.66	34.879	51.015
MultiResUnet	88.99 ± 0.60	81.10 ± 1.02	7.238	14.281
UNeXt	86.67 ± 1.00	77.56 ± 1.41	1.472	**0.439**
MedT	78.55 ± 1.84	66.66 ± 2.39	1.371	1.950
HFF-UNet (OURS)	**90.15 ± 0.41**	**82.77 ± 0.64**	**1.347**	2.802
ULite	85.63 ± 1.27	76.11 ± 1.79	0.878	**0.580**
HFF-UNetS (OURS)	**87.92 ± 0.84**	**79.49 ± 1.21**	**0.766**	1.707

Table 2. Comparison of our method with other methods on the CVC-ClinicDB dataset
表 2. 本文方法在 CVC-ClinicDB 数据集与其他方法的对比

模型	DSC/%	IoU/%	参数量/M	FLOPs/G
Unet	90.06	83.96	7.675	10.564
Unet++	91.39	85.14	9.163	26.723
AttU_Net	91.44	85.25	34.879	51.015
MultiResUnet	89.5	82.52	7.238	14.281
UNeXt	85.36	77.17	1.472	**0.439**

续表

MedT	81.72	71.69	1.371	1.950
HFF-UNet (OURS)	**92.11**	**85.85**	**1.347**	2.802
ULite	86.08	77.86	0.878	**0.580**
HFF-UNetS (OURS)	**90.72**	**83.62**	**0.766**	1.707

分割结果可视化分析如图 6 所示，结果表明：在 GlaS 数据集上，HFF-UNet 对腺体复杂边界和多样结构的分割效果显著优于 U-Net 和 UNeXt，边界清晰度更高。在 CVC-ClinicDB 数据集上，HFF-UNet 能有效捕捉低对比度息肉区域细节，边缘平滑性优于 AttU-Net 和 MultiResUnet。以 GlaS 数据集第二行图像为例，其他模型未能完全分割左下角两个腺体，而 HFF-UNet 分割完整。再如 CVC-ClinicDB 数据集最后一行，其他模型在左下角区域出现误分割，HFF-UNet 结果准确。这些定性观察结果与定量分析一致，验证了 HFF-UNet 在复杂场景下的鲁棒分割性能。

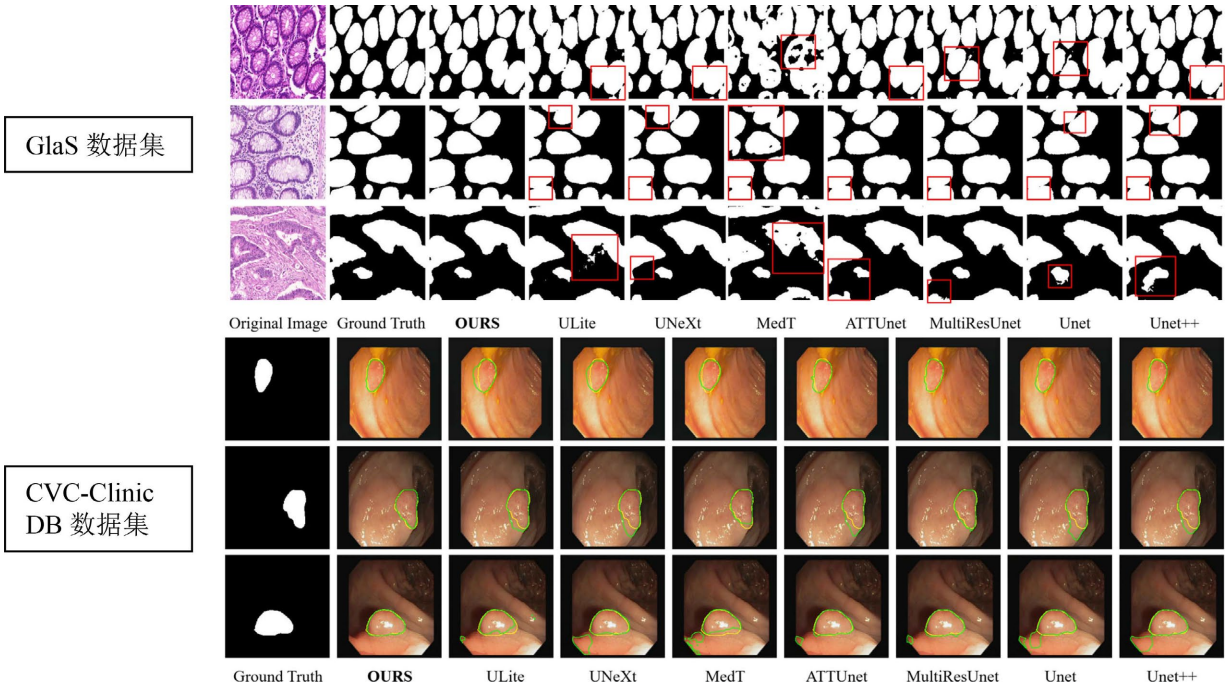

Figure 6. Comparison of segmentation results of different models on the CVC-ClinicDB dataset
图 6. 不同模型分别在 GlaS 和 CVC-ClinicDB 数据集上的分割结果对比

为了进一步研究 PPEM 模块在特征筛选与特征传递的优越性，本文基于 SE 模块提出两种改进版本，针对标准 SE 模块中全局平均池化(Global Average Pooling, GAP)的固有局限：过度压缩导致细粒度信息丢失及尺度敏感性不足。本研究引入多尺度特征增强策略：采用 GAP 与全局最大池化(GMP)并行的双路径池化机制，以增强多尺度上下文特征表达能力。其中，V1 采用加性融合，V2 采用通道拼接融合。

将两种改进方法应用于 UNet 跳跃连接并与 SE 对比，实验结果如表 3 所示，实验表明其 DSC/IoU 均优于 SE，验证了多尺度特征增强的有效性，并为 PPEM 设计提供了依据。进一步地，PPEM 在 GlaS 数据集上与先进注意力机制(ECA [26], SE, CBAM [27], CA [28])对比显示，在少量增加参数量的前提下，DSC/IoU 均有显著提升。其中，PPEM 较 SE 提升 1.17%DSC 和 1.68%IoU。

Table 3. Comparison of PPEM with other different attention mechanisms on the GlaS dataset
表 3. PPEM 在 GlaS 数据集上与其他不同注意力机制的对比

模型	DSC/%	IoU/%	参数量/M	FLOPs/G
V1	87.49 ± 1.83	78.86 ± 2.56	0.5256	1.2487
V2	88.19 ± 0.77	79.90 ± 1.08	0.7001	1.2489
ECA	87.92 ± 0.71	79.47 ± 0.95	**0.5147**	**1.2487**
SE	87.13 ± 1.20	78.38 ± 1.72	0.5256	1.2487
CBAM	88.14 ± 0.39	80.05 ± 0.58	0.5259	1.2553
CA	87.51 ± 1.03	78.86 ± 1.55	0.5272	1.2527
PPEM (OURS)	**88.3 ± 0.27**	**80.06 ± 0.38**	0.5418	1.3734

同时本文采用了 Grad-CAM [29]作为不同注意力机制的可视化方法，可视化了表 3 中的注意机制方法，如图 7 所示。图 7 中每行展示一种方法的三重对比图(无注意力首层、有注意力首层、末层注意力图)。结果显示，本文提出的 PPEM 能更精准捕捉边缘和小目标信息，有效优化编码器–解码器特征传递，减少冗余信息干扰。

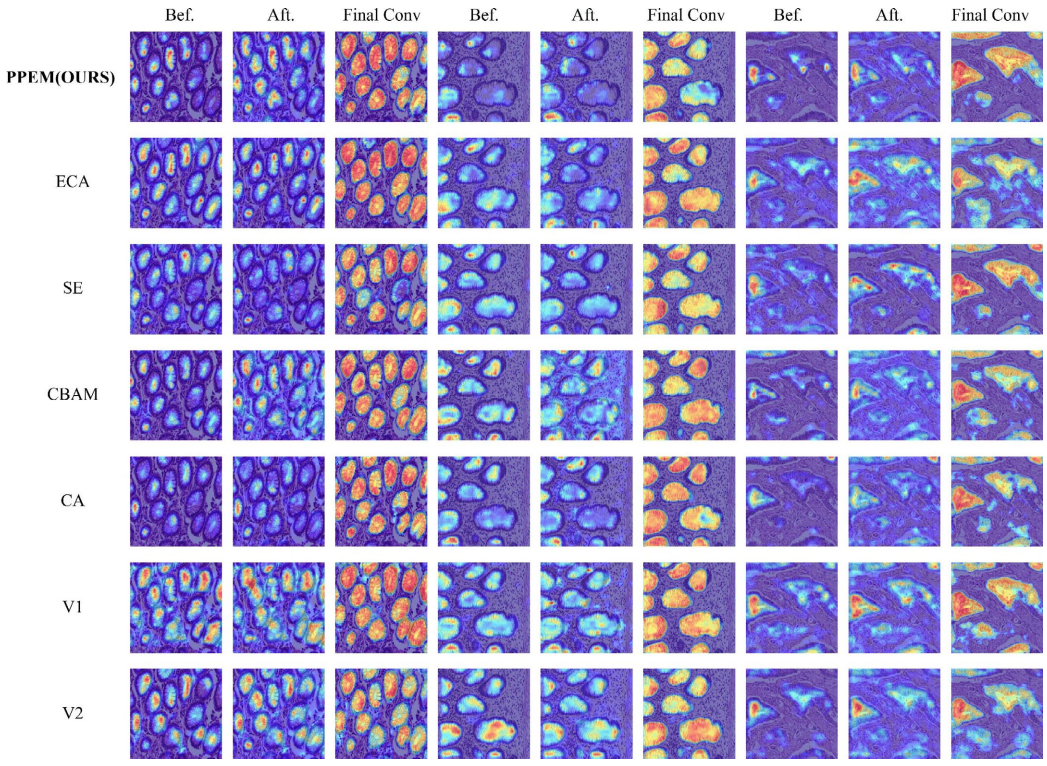

Figure 7. Grad-CAM visualization of attention maps
图 7. Grad-CAM 可视化注意力图

4.5. 消融实验

为验证各模块的贡献，我们在 GlaS 和 CVC-ClinicDB 数据集上进行了系统的消融实验，实验结果如表 4 和表 5 所示。

Table 4. Ablation study results on the Glas dataset
表 4. 在 Glas 数据集上消融实验的结果

	EMA	PPEM	HFFB	Dice/%	IoU/%
1				84.84	74.80
2	√			87.18	78.35
3		√		86.31	77.19
4			√	88.11	79.78
5	√	√		88.3	80.06
6	√	√	√	**90.15**	**82.77**

Table 5. Ablation study results on the CVC-ClinicDB dataset
表 5. 在 CVC-ClinicDB 数据集上消融实验的结果

	EMA	PPEM	HFFB	Dice/%	IoU/%
1				89.36	82.12
2	√			91.21	84.60
3		√		90.06	83.17
4			√	90.48	83.52
5	√	√		91.62	85.27
6	√	√	√	**92.11**	**85.85**

Table 6. Module computational cost analysis
表 6. 模块计算成本分析

模块	参数量/M	参数量占比	FLOPs/G	FLOPs 占比
EMA	0.01	0.75%	0.20	7.14%
PPEM	0.02	1.49%	0.12	4.29%
HFFB	0.80	59.70%	1.43	51.07%
Baseline	0.51	38.06%	1.05	37.50%
总计	1.34	100.00%	2.80	100%

在 GlaS 数据集上，EMA 模块提升 DSC 至 87.18% (+2.34%)，IoU 至 78.35% (+3.55%)，证明其增强特征表达的有效性；在 CVC 上 EMA 也提升 DSC1.85%，IoU2.48%，表明普适性。PPEM 单独提升 DSC1.47%，与 EMA 组合达 DSC88.3%，IoU80.06%，有效缓解语义差距。仅 HFFB 在 GlaS 提升 DSC 至 88.11% (+3.27%)，在 CVC 提升 IoU 至 84.60% (+2.48%)，验证跳跃引导融合。完整模型在两个数据集均达最优(GlaS:DSC90.15%, IoU82.77%; CVC:DSC92.11%, IoU85.85%)，较双模块组合提升 1.85%DSC，证明三模块互补性：EMA 增强特征提取，PPEM 优化特征传递，HFFB 改善融合质量。

从表 6 可以看到设计的三个模块的计算成本的分布，EMA 和 PPEM 以极低的计算开销显著提升了特征提取与传递效率，充分体现了其设计的轻量高效特性。HFFB 作为负责高分辨率特征融合与细节恢复的核心模块，其设计的引导式融合操作需处理更大尺寸的数据，导致计算成本显著高于其他模块，虽然消耗主要计算资源，但是从消融实验的结果可以发现它对最终精度提升的显著贡献，证明该计算资源分

配的合理性。三者协同工作，实现了性能与效率的优异平衡。

5. 结论

针对现有网络在全局建模、特征融合和轻量化上的不足，本文提出了一种轻量级医学图像分割网络HFF-UNet。其核心模块(EMA、PPEM、HFFB)有效提升特征提取与融合能力，在 Glas 和 CVC-ClinicDB 数据集上取得高精度与低复杂度的平衡，为精准医疗提供技术支持。

基金项目

武汉工程大学第十八期学生校长基金项目(XZJJ2024043)。

参考文献

[1] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., *et al*., Eds., *Medical Image Computing and Computer-Assisted Intervention*, Springer International Publishing, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[2] Diakogiannis, F.I., Waldner, F., Caccetta, P. and Wu, C. (2020) ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **162**, 94-114. https://doi.org/10.1016/j.isprsjprs.2020.01.013

[3] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N. and Liang, J. (2020) UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, **39**, 1856-1867. https://doi.org/10.1109/tmi.2019.2959609

[4] Chen, J., Lu, Y., Yu, Q., *et al*. (2021) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. https://doi.org/10.48550/arXiv.2102.04306

[5] 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

[6] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., *et al*. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 10012-10022. https://doi.org/10.1109/iccv48922.2021.00986

[7] Wang, H., Cao, P., Wang, J., *et al*. (2022) UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-Wise Perspective with Transformer. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 2441-2449.

[8] Ouyang, D., He, S., Zhang, G., *et al*. (2023) Efficient Multi-Scale Attention Module with Cross-Spatial Learning. *Proceedings of ICASSP* 2023-*IEEE International Conference on Acoustics*, *Speech and Signal Processing*, Rhodes Island, 4-10 June 2023, 1-5.

[9] Valanarasu, J.M.J. and Patel, V.M. (2022) UNeXt: MLP-Based Rapid Medical Image Segmentation Network. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Singapore, 18-22 September 2022, 23-33.

[10] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 1251-1258. https://doi.org/10.1109/cvpr.2017.195

[11] Howard, A.G., Zhu, M., Chen, B., *et al*. (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. https://doi.org/10.48550/arXiv.1704.04861

[12] Dinh, B., Nguyen, T., Tran, T. and Pham, V. (2023) 1M Parameters Are Enough? A Lightweight CNN-Based Model for Medical Image Segmentation. 2023 *Asia Pacific Signal and Information Processing Association Annual Summit and Conference* (*APSIPA ASC*), Taipei, 31 October-3 November 2023, 1279-1284. https://doi.org/10.1109/apsipaasc58517.2023.10317244

[13] Ruan, J., Xiang, S., Xie, M., *et al*. (2022) MALUNet: A Multi-Attention and Lightweight UNet for Skin Lesion Segmentation. 2022 *IEEE International Conference on Bioinformatics and Biomedicine* (*BIBM*), Las Vegas, 6-8 December 2022, 1150-1156.

[14] Wang, X., Girshick, R., Gupta, A. and He, K. (2018) Non-Local Neural Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7794-7803. https://doi.org/10.1109/cvpr.2018.00813

[15] Schlemper, J., Oktay, O., Schaap, M., *et al*. (2019) Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images. *Medical Image Analysis*, **53**, 197-207.

[16] Schlemper, J., Oktay, O., Chen, L., *et al*. (2018) Attention-Gated Networks for Improving Ultrasound Scan Plane Detection. https://doi.org/10.48550/arXiv.1804.05338

[17] Petit, O., Thome, N., Rambour, C., Themyr, L., Collins, T. and Soler, L. (2021) U-Net Transformer: Self and Cross Attention for Medical Image Segmentation. *Machine Learning in Medical Imaging*: 12*th International Workshop*, *MLMI* 2021, Strasbourg, 27 September 2021, 267-276. https://doi.org/10.1007/978-3-030-87589-3_28

[18] Sinha, A. and Dolz, J. (2021) Multi-Scale Self-Guided Attention for Medical Image Segmentation. *IEEE Journal of Biomedical and Health Informatics*, **25**, 121-130. https://doi.org/10.1109/jbhi.2020.2986926

[19] Vaswani, A., Shazeer, N., Parmar, N., *et al*. (2017) Attention Is All You Need. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010. https://doi.org/10.5555/3295222.3295349

[20] Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I. and Patel, V.M. (2021) Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention*, *MICCAI* 2021, Strasbourg, 27 September-1 October 2021, 36-46. https://doi.org/10.1007/978-3-030-87193-2_4

[21] Sun, G., Pan, Y., Kong, W., Xu, Z., Ma, J., Racharak, T., *et al*. (2024) DA-Transunet: Integrating Spatial and Channel Dual Attention with Transformer U-Net for Medical Image Segmentation. *Frontiers in Bioengineering and Biotechnology*, **12**, Article ID: 1398237. https://doi.org/10.3389/fbioe.2024.1398237

[22] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. 2018 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/cvpr.2018.00745

[23] Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J. (2017) Pyramid Scene Parsing Network. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 2881-2890. https://doi.org/10.1109/cvpr.2017.660

[24] Sirinukunwattana, K., Pluim, J.P.W., Chen, H., *et al*. (2017) Gland Segmentation in Colon Histology Images: The GlaS Challenge Contest. *Medical Image Analysis*, **35**, 489-502. https://doi.org/10.1016/j.media.2016.07.004

[25] Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C. and Vilariño, F. (2015) WM-DOVA Maps for Accurate Polyp Highlighting in Colonoscopy: Validation vs. Saliency Maps from Physicians. *Computerized Medical Imaging and Graphics*, **43**, 99-111. https://doi.org/10.1016/j.compmedimag.2015.02.007

[26] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q. (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 11534-11542. https://doi.org/10.1109/cvpr42600.2020.01155

[27] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[28] Hou, Q., Zhou, D. and Feng, J. (2021) Coordinate Attention for Efficient Mobile Network Design. *Proceedings of the IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, 19-25 June 2021, 13713-13722.

[29] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 618-626. https://doi.org/10.1109/iccv.2017.74