基于小波的低频和高频融合网络用于生物医学图像的全监督和半监督语义分割

摘要

随着深度神经网络 (DNNs) 的发展，生物医学图像的全监督和半监督语义分割得到了推进。然而，到目前为止，DNN 模型通常被设计为仅支持这两种学习方案中的一种，支持全监督和半监督分割的统一模型仍然存在局限。此外，很少有全监督模型关注图像固有的低频 (LF) 和高频 (HF) 信息来提高性能。基于一致性的半监督模型中的摄动通常是人为设计的。它们可能会引入不利于训练的负面学习偏差。本文首先复现了原作者所提出的小波的 LF 和 HF 融合模型 XNet 模型及其改进版本 XNetv2 模型，同时基于该模型尝试设计了 RGHPA 模块从多个角度提取特征信息以增强特征提取和泛化能力，该模型支持全监督和半监督语义分割。该模型强调提取 LF 和 HF 信息进行一致性训练，以减轻人为扰动造成的学习偏差。在 2 个公开的 2D 数据集和 2 个私有的 2D 数据集上进行了复现对比实验，验证了该模型的有效性。

关键词： 低频; 高频; 全监督；半监督; 语义分割

1 引言

医学图像获取和解释是基于医学成像诊断的基础。近年来，图像获取的速度和分辨率得到了极大的改善，但是在传统医学诊断中，专家判断通常被认为是决定性因素。在医学诊断领域中虽然得益于深度学习和人工智能的迅速发展，但是仍然严重依赖于医生的判断，然而这种依赖存在很多挑战和不确定性，包括耗时的过程以及由于医疗工作的要求高和重复性导致的疲劳和相关错误。虽然深度学习方法在特定条件下展示了有希望的诊断结果，但它们仍然受到数据库限制、泛化和准确性问题的显著制约。从复杂的生物医学图像中精准提取有意义的区域，如器官、病变组织等，面临巨大挑战。传统的图像分割方法难以适应生物医学图像的高度复杂性与多样性。与此同时，深度学习技术凭借强大的自动特征学习能力在图像分割领域崭露头角，但它极度依赖大规模的精确标注数据，这在生物医学领域成为瓶颈，因为专业的医学图像标注不仅要求深厚的医学知识，耗费高昂的人力、时间成本，还易出现标注不一致等问题。在此背景下，探索高效的全监督与半监督训练策略，结合新兴的基于小波的低频和高频融合网络技术，为生物医学图像语义分割开辟新途径显得尤为迫切。医学图像分割主要通过评估图像间所给定特征因素的相似度，然后将图像分为不同的区域，其中的目标区域主要包括组织、器官和细胞，这种区域分割方法主要依赖于图像的局部空间特征因素，例如颜色，形状等特征。基于边界的分割方法主要使用梯度信息来确定目标的边界，例如 Fast Marching 算法和 Watershed 算法可以快速准确地分割医学图像 [27]。

全监督学习模式在过往研究中，当配备充足且精准标注的生物医学图像样本时，展现出卓越的分割精度，能够针对特定的医学图像任务，如视网膜血管分割、肺部结节分割等，训练出高敏感度与特异度的模型。这得益于全监督训练让模型充分学习标注样本中的细微特征差异，构建起精准的判别模式。另一方面，获取大量标注良好的生物医学图像在现实中困难重重，与之形成鲜明对比的是，未标注的生物医学图像数据资源极为丰富。半监督训练策略恰好能桥接这一鸿沟，利用少量标注样本引导模型对大量未标注样本进行学习挖掘，其核心原理如基于一致性正则化确保模型对未标注样本变换的稳定预测，以及伪标签技术赋予未标注样本可靠的预测标签以迭代优化模型。此外，小波分析在图像处理领域久负盛名，它能够将图像分解为低频和高频分量，低频分量保留图像整体轮廓等大致信息，高频分量蕴含丰富细节纹理，通过设计巧妙的融合网络整合两者优势，有望克服生物医学图像中不同尺度结构、模糊边界等分割难点，前期小波理论在图像去噪、增强等方面的成功应用也为其在图像分割的深入研究提供了可行性支撑。[63, 64]

语义分割是生物医学图像分析中的一项基本任务，其目标是为每个像素分配一个类 label。基于卷积神经网络（convolutional neural networks, cnn）的生物医学图像语义分割方法已经取得了显著的成功 [10, 22, 39, 44]。最近，transformers 和 CNNs 的结合已经流行起来 [3, 20, 51, 58, 58]。transformers 可以捕获远程依赖关系 [2, 5, 59]，以补偿 CNNs 有限的接受域。然而，现有的大多数方法都侧重于模型架构，以更好地提取特征 [19, 38, 66]。很少有方法探索图像的内在 LF 和 HF 信息，这些信息可能对分割有用 [26, 56]。对于生物医学图像的语义分割，监督方法需要大规模的标记图像，这是昂贵且耗时的。为了缓解这一问题，研究人员提出了使用少量标记图像和大量未标记图像进行学习的半监督方法 [4, 28, 47]。常见的解决方案包括对抗性训练 [25, 45]、伪标记 [23, 30, 61] 和一致性正则化 [1, 50]。一致性正则化是目前表现最好的方法 [53, 54]，它扰动输入图像、中间特征或输出预测，允许模型从扰动中学习一致性 [53, 55, 57]。然而，目前的扰动策略是人为设计的，如旋转 [55]、噪声添加 [57]、距离映射 [54] 和 dropout [57] 等。它们可能会引入负的学习偏差，比如分割噪声图像相当于学习了一个额外的去噪任务。此外，完全监督和半监督语义分割被视为两个不同的研究领域。同时达到最先进水平的统一模型仍然有限。为了解决上述问题，作者提出了一种基于小波的 LF 与 HF 融合模型 XNet。XNet 可以同时实现基于 LF 和 HF 信息融合的全监督学习，以及基于 LF 和 HF 输出一致性的半监督学习。原作者使用小波变换生成 LF 和 HF 图像，并将其输入 XNet。XNet 融合它们的 LF 和 HF 信息，然后生成双分支分割预测。对于监督学习，分割预测吸收原始图像的完整 LF 和 HF 信息。对于半监督学习，双输出对 LF 和 HF 信息的关注不同，导致一致性差异。这些差异被用于对未标记图像的训练。对于语义分割问题，HF 信息一般表示图像细节，而 LF 信息往往是抽象语义。提取和融合不同频率信息的策略可以帮助模型更好地关注 LF 语义和 HF 细节，从而提高性能。XNet 模型使用小波变换生成 LF 和 HF 图像，用于基于一致性差异的半监督学习。这些一致性差异源于对 LF 和 HF 信息关注的不同，从而缓解了人为设计造成的学习偏差。

2 相关工作

2.1 多模态医学图像分割

对医学图像处理而言，不管是特征提取还是图像融合，一个不可避免的步骤是对肿瘤区域进行分割，然后对该区域进行有针对性地处理，由于医学图像具有特异性，单模态的融合分割方法通常不适用于多模态医学图像的融合分割 [42]。当前多模态医学图像融合分割方法主要利用融合信息进行辅助分割。双分支分割网络 [9] 利用不同模态的有效特征以提高模型对病灶区域的分割能力，全卷积神经网络 [13] 则学习模态之间更复杂的特征表示以进行精确的多模态图像分割。文献 [12] 设计了更复杂的网络模型来学习和组合不同模态的多尺度上下文信息以进行病灶分割。上述分割网络复杂度高并且忽略了多个分割任务之间的相关性，因此，MFCNs [37] 为每个模态训练一个网络以提取每个模态的深层特征并进行融合。多任务分割网络 [62] 将病灶分割任务分解为三个不同但相关的子任务，从而提高了模型的分割性能。而 TSCN [49] 则进一步结合多任务和多视图技术，根据病灶的层次结构将多类分割任务分解为若干简单的分割任务，以对病灶区域逐个进行分割。为了更有效地利用多模态信息，文献 [60] 利用相互学习策略提取不同模态高级表示之间的共性，而交叉编码器 [65] 可学习原始图像之间的互补属性，迫使网络学习多模态图像之间的互补信息。类似地，也有研究将空间和通道维度注意力机制用于提取病灶特征以进行病灶分割 [17]。

2.2 UNet 算法及其变体

近年来，UNet [39] 算法因其在医学图像处理方面具有优异性能而被广泛采用。作为一种深度学习架构，它由对称编码器和解码器组成。这种结构能够有效捕获图像中的局部和全局特征，以及拥有强大的上下文信息传播能力。但是 UNet [39] 的卷积结构决定了其只能获取局部感受野的信息，对于全局上下文信息的理解能力相对较弱，并且 UNet [39] 在解码器阶段进行上采样操作，可能导致分割结果的边界模糊，无法得到清晰的分割边界，在处理大尺寸图像时效果较差，因此在乳腺肿瘤分割任务中这种非刚性组织的成像上效果更差。

受 UNet [39] 及 Transformer [5] 的启发，TransUNet [8] 融合了二者的优势成为医学图像分割的一个强有力的选择。其中，Transformer 通过编码来自卷积神经网络（CNN）中被标记化的特征图的图像块，然后将其编码后的图像块输入序列来提取全局上下文信息。同时，解码器会对编码器提取的特征进行上采样处理，然后与高分辨率的卷积神经网络（CNN）输出的特征图进行融合，从而能够实现精准的目标定位。MedNeXt [40] 是一个大型卷积核分割网络，作为纯 ConvNeXt 3D [52] 编码器和解码器架构，用于医学图像分割。它结合了 ConvNeXt [33] 上采样和下采样块与跳跃连接，以保持跨尺度的语义丰富性。UNeXt [48] 结合了 UNet [39] 和 MLP [46]，开发了一种轻量级模型，在减少参数和计算量的同时实现了卓越的性能。跳跃连接将每一对下采样层和上采样层连接起来，使空间信息能够直接传播到更深的层，从而产生更准确的分割结果。在 UNet [39] 网络中，跳跃连接用于融合从解码器输出的高级语义特征图与从编码器获得的相应低级的特征图。然而，在 UNet [39] 中有一个问题是在普通的跳跃连接中如何融合不同语义上的特征。作为 UNet [39] 的增强版本，UNet++ [62] 通过开发具有嵌套和密集连接组件的跳跃连接架构来解决此问题，然而它无法在多尺度上充分探索信息，导致特征捕获不完整。一种用于医学成像的新型注意力门（AG）模型 Attention UNet [38] 可

以自动学习聚焦于不同形状和大小的目标结构，使用 AG 训练的模型隐式学习抑制输入图像中的不相关区域，同时突出显示对特定任务有用的显著特征。

在医学图像分割任务中由于高精度的要求因此监督学习是最受欢迎的方法。图像语义分割的目标是对图像中的像素进行分类，主要使用包含编码器和解码器的全卷积网络 (FCN) 架构 [34]。编码器模块主要负责用来提取图像中的特征，同时解码器模块主要负责用来将编码器模块提取到的图像特征恢复到与原始图像大小一致，并且实现输出最终的分割掩码。常见的架构包括 FCN、UNet [39]、DeepLab [10] 等。常规图像的分割通常质量较高，噪声较少，拍摄条件相对稳定且物体边界相对清晰易于区分，在特征提取方面常规图像中可以依赖图像的低级特征（如颜色、纹理）进行分割，并且对分割的精度要求相对较低，能够容忍一定误差，但是在医学图像中常包含较多噪声，图像质量可能受到设备和成像条件的影响导致组织和器官的边界模糊，不易准确分辨，同时低级特征往往不足，需要依赖高级特征（如形状、上下文信息）进行准确分割，目标通常为人体内部的组织和器官，形态相对固定，但结构复杂，并且对分割的精度要求极高，细小的误差可能影响诊断和治疗结果，因此在医学图像中检测和识别人体内部的组织和器官目标具有非常大的挑战性。另外缺乏图像细节信息会使得仅仅依赖语义特征不足以准确稳定地描绘目标边界。为了应对这些挑战，医学图像分割常常采用更为复杂的网络结构和先进的算法，UNet [39] 网络因为其加入了跳跃连接，因此能够将低分辨率特征和高分辨率特征实现有效融合，成为了医学图像分割任务中较为理想的解决方案。目前，UNet [39] 网络已成为许多医学图像分割任务的基准架构，同时由于其具有对下游任务的灵活性，促进了各种领域内的许多重大改进。

2.3 生物医学图像的全监督语义分割

随着深度学习的兴起，CNNs 在语义分割中得到了广泛的应用 [15]，如 FCN [34]、DeepLab v3+ [10] 等。对于生物医学图像，高效的编解码器架构实现了卓越的性能 [22]，如 UNet [39]、UNet++ [66]、UNet 3+ [21] 等。此外，研究人员将这种架构扩展到 3D，以满足体分割的需求。[36] 提出了一个 3D 全 CNN VNet。[67] 将 UNet 扩展到 3D。ConResNet 提出了片间上下文残差学习。SwinUNet [7]、TransBTS [51]、UN- ETR [3] 等，这些模型既捕获了远程依赖关系，也捕获了局部特征，以提高性能。

2.4 生物医学图像的半监督语义分割

为了缓解标记图像的不足，生物医学图像的半监督语义分割已经成为一个关键的方法 [24, 32, 33]。目前占主导地位的策略包括对抗性训练、伪标记 [23, 30, 61] 和一致性正则化 [1, 50]。对抗训练使用生成式对抗网络 [18] 来不断提高生成分割预测的生成器和判断预测真实性的判别器的性能。伪标记利用高置信度预测来提高模型性能。基于一致性正则化的方法具有更好的性能 [35, 53, 54]。它们通过加强不同预测之间的一致性来利用未标记的图像。DTC [35] 提出了一种双任务一致性网络，用于预测分割映射和几何感知水平集表示。TCSMv2 [55] 利用转换一致性，允许网络为不同的扰动输入生成一致的预测。[53] 提出了一种不确定性校正金字塔一致性 (URPC) 策略。

2.5 基于小波的深度神经网络语义分割

基于小波的深度神经网络语义分割。基于小波变换强大的频率和空间表示能力，深度神经网络与小波变换相结合，并探索了一些用于语义分割的方法 [16,31]。常用的策略包括使用小波变换作为预处理或后处理 [26,56]，用小波变换替换 cnn 的某些层 (如上采样和下采样) [?,16]。然而，它们中的大多数只适用于特定的分割对象，这限制了它们的泛化和应用。[6] 提出了一种基于小波变换增强的对称 CNN (Aerial LaneNet)，用于航路图像的车道标记语义分割。CWNN [14] 使用小波约束池化层取代常规池化，用于合成孔径雷达图像分割。WaveSNet [29] 在下采样期间使用小波变换提取图像细节，在上采样期间使用逆变换恢复细节。相反，作者使用小波变换生成 LF 和 HF 图像作为双支路输入提取 LF 和 HF 特征。

3 本文方法

3.1 本文方法概述

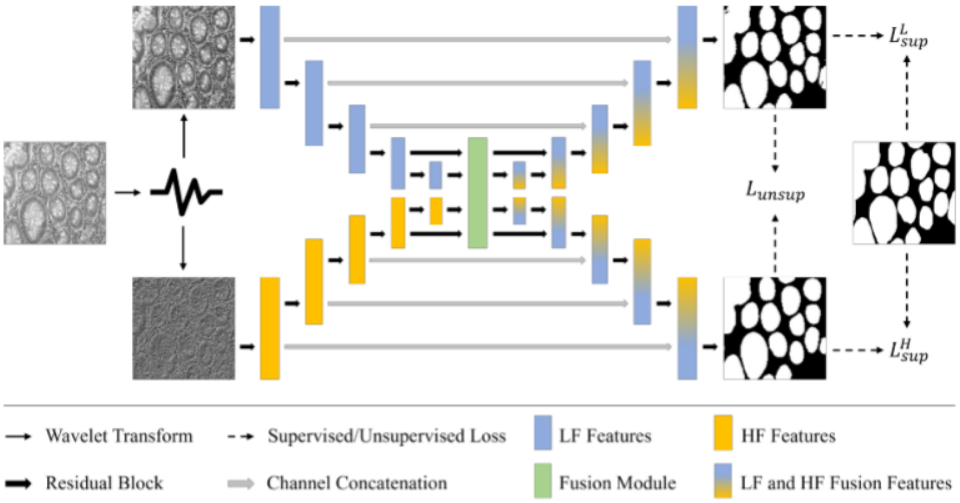

图 1. 提出的 XNet 模型概述。蓝色和橙色分别代表 LF 和 HF 编码器，绿色代表融合模块。混合颜色表示双分支解码器。XNet 通过最小化 L_{unsup} 学习未标注的图像，并通过最小化 L^L_{sup}，L^H_{sup} 和 L_{unsup} 学习标注的图像。[63]

融合模块将 LF 语义和 HF 细节特征融合，生成融合特征。双分支解码器使用融合特征生成分割预测。

如图 1所示，训练过程通过对原始图像进行小波变换，提取相应的 LF 和 HF 图像，并将其输入到 LF 和 HF 编码器中以生成 LF 和 HF 特征。这些特征在融合模块中融合后输入到解码器，以生成 LF 和 HF 分支的分割预测。对于有监督训练，通过最小化标注图像的监督损失和双输出一致性损失来优化模型。对于半监督训练，通过最小化标注图像的监督损失和未标注图像的双输出一致性损失来优化模型。因此，无论是完全监督还是半监督训练，总损失函数 L_{total} 定义如下：

$$L_{\text{total}} = L_{\text{sup}} + \lambda L_{\text{unsup}} \tag{1}$$

其中 L_{sup} 为监督损失，L_{unsup} 为无监督损失（即双输出一致性损失），λ 为控制 L_{sup} 和 L_{unsup} 之间平衡的权重。具体来说，监督损失 L_{sup} 由 LF 监督损失 L_{sup}^{L} 和 HF 监督损失 L_{sup}^{H} 组成：

$$L_{\text{sup}} = L_{\text{sup}}^{L}(p_i^L, y_i) + L_{\text{sup}}^{H}(p_i^H, y_i) \tag{2}$$

其中 p_i^L 和 p_i^H 分别表示第 i 张图像的 LF 和 HF 分割预测，y_i 表示第 i 张图像的真实值。无监督损失 L_{unsup} 通过交叉伪监督（CPS）损失实现：

$$L_{\text{unsup}} = L_{\text{unsup}}^{L}(p_i^L, \hat{p}_i^H) + L_{\text{unsup}}^{H}(p_i^H, \hat{p}_i^L) \tag{3}$$

其中 L_{unsup}^{L} 和 L_{unsup}^{H} 分别表示 LF 和 HF 的无监督损失，$\hat{p}_i^L$ 和 $\hat{p}_i^H$ 分别表示由 p_i^H 和 p_i^L 生成的 LF 和 HF 伪标签。

在本研究中，$L_{\text{sup}}^{L}(\cdot)$、$L_{\text{sup}}^{H}(\cdot)$、$L_{\text{unsup}}^{L}(\cdot)$ 和 $L_{\text{unsup}}^{H}(\cdot)$ 均采用 Dice 损失。原文在训练阶段选择性能更好的分支作为推理过程中的最终输出。

3.2 小波变换

二维（2D）或三维（3D）图像本质上是离散的非平稳信号，包含不同的频率范围和空间位置信息。小波变换可以在分解这些信号的同时有效地保留这些信息。

具体来说，以二维图像为例。作者使用小波变换将原始图像分解为 LF 分量和 HF 分量，包括水平 HF、垂直 HF 和对角 HF 分量（LL、HL、LH 和 HH）。这些分量分别保存原始图像的 LF 和 HF 信息。作者用 LF 分量表示 LF 图像 L，用不同方向 HF 分量的和表示 HF 图像 H。L 和 H 定义为：

$$L = LL \tag{4}$$

$$H = HL + LH + HH \tag{5}$$

L 和 H 在图 2 中显示。可以看出，H 强调细节信息，而 L 关注语义信息。

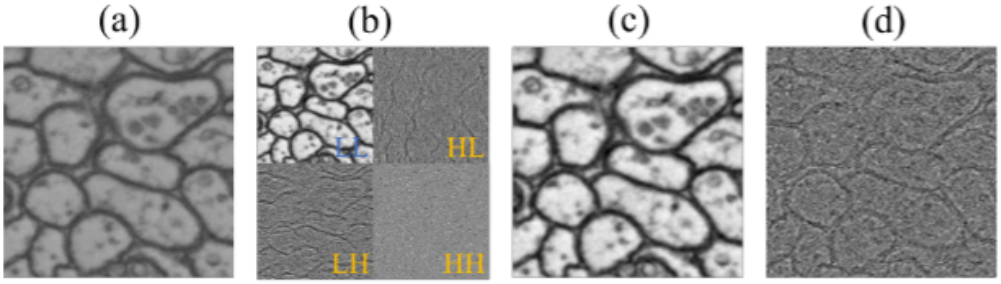

图 2. 以 CREMI 数据集为例，展示 LF 和 HF 结果。(a) 原始图像。(b) 小波变换结果。(c) LF 图像。(d) HF 图像。[63]

与其他方法（如傅里叶变换）相比，小波变换是一种高效生成 L 和 H 的方法。以 L 作为输入，XNet 能够更关注 LF 语义特征，因为 L 含有较少的噪声和细节信息。而 H 相比之下噪声更多但物体边界更清晰，有助于模型更关注 HF 细节。此外，在半监督训练中，L 和 H 的一致性差异源于图像中内在的 LF 和 HF 信息，这有助于缓解人工扰动带来的学习偏差。

3.3 LF 和 HF 融合模块

LF 和 HF 融合模块的架构如图 11 所示。通过使用 LF 和 HF 特征作为输入，融合模块通过 3×3 卷积进行相同尺寸处理、上采样或下采样，随后将其通道连接起来。然后，这些通道连接后的特征被输入到 1×1 卷积中，以生成 LF 和 HF 的融合特征。

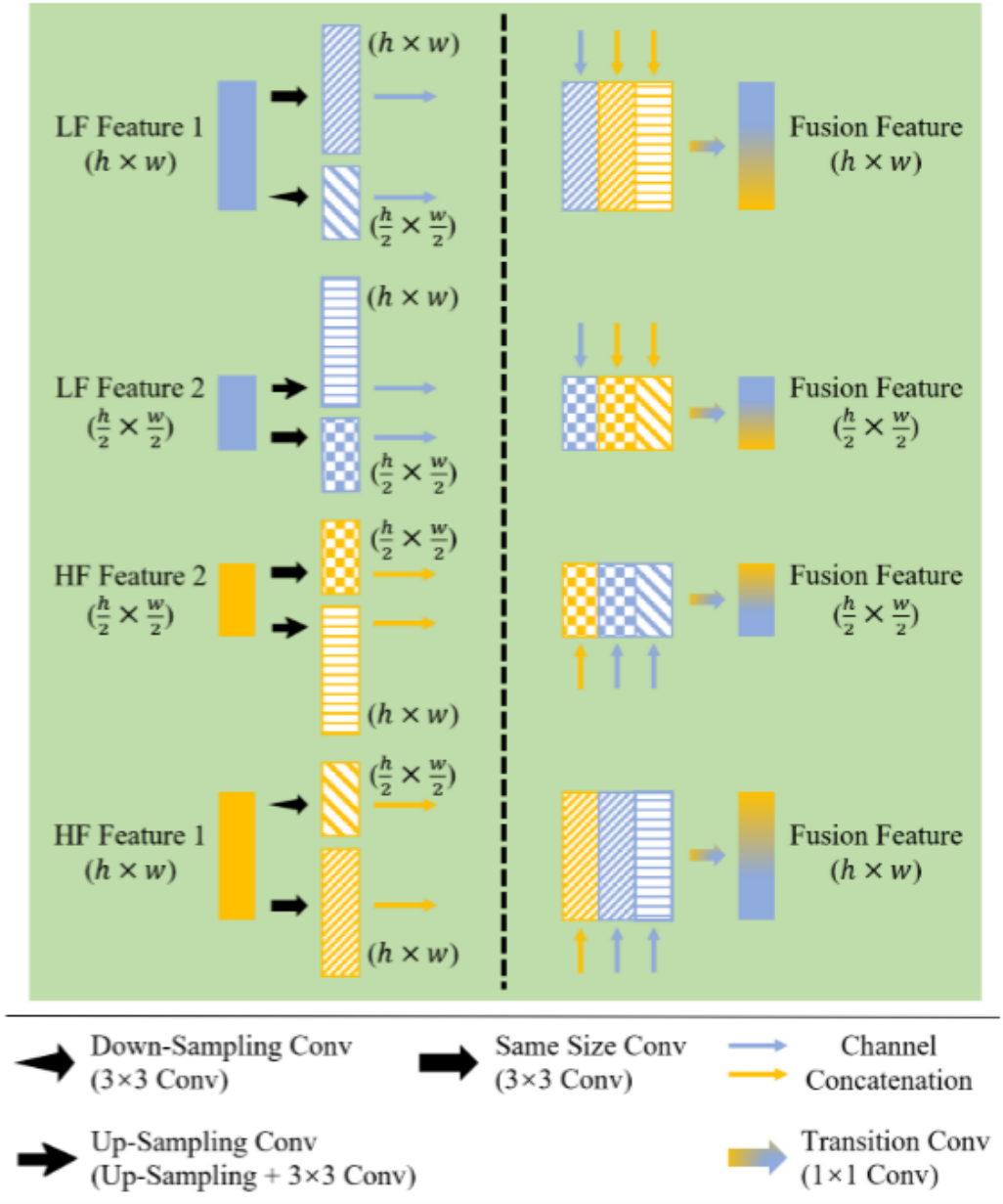

图 3. LF 和 HF 融合模块的架构。相同尺寸卷积表示输入和输出特征尺寸相同。下采样卷积将特征尺寸减半，上采样卷积将特征尺寸加倍。通道连接用于连接输入和输出特征。

融合模块可以将 LF 和 HF 特征结合为完整特征。若不使用融合模块，每个分支可能会缺乏语义或细节特征，这会对分割结果产生不利影响。

3.4 全监督与半监督的可行性

对于生物医学图像，作者假设原始图像 I 包含 LF 特征 F_L、HF 特征 F_H、LF 附加噪声 N_L 和 HF 附加噪声 N_H。因此，I 定义为：

$$I = F_L + F_H + N_L + N_H \tag{6}$$

相关研究表明，生物医学图像中的噪声通常是加性的 [63, 64]。对于语义分割问题，准确的分割需要 LF 语义特征（如形状、颜色等）和 HF 细节特征（如边缘、纹理等）。小波变换 W 可以将原始图像 I 分解为 LF 图像 L 和 HF 图像 H：

$$L, H = W(I) \tag{7}$$

$$L = F_L + N_L \tag{8}$$

$$H = F_H + N_H \tag{9}$$

LF 和 HF 编码器 E_L 和 E_H 分别从 L 和 H 中提取特征 F_L 和 F_H：

$$F_L = E_L(L) \tag{10}$$

$$F_H = E_H(H) \tag{11}$$

融合模块 M 将 F_L 和 F_H 融合为完整特征 F_M：

$$F_M = M(F_L, F_H) \tag{12}$$

对于监督学习，解码完整特征可以获得分割预测。对于半监督学习，由于 LF 和 HF 特征信息在解码分支中的不同，导致双分支解码器之间的预测存在 LF 和 HF 细节差异。这些差异可以用于基于一致性正则化的半监督训练。

LF 和 HF 分支的分割预测定义为：

$$P_L, P_H = D(F_M) \tag{13}$$

其中，P_L 和 P_H 分别表示 LF 和 HF 分割预测，D 表示双分支解码器。

3.5 Residual Group multi-axis Hadamard Product Attention 模块

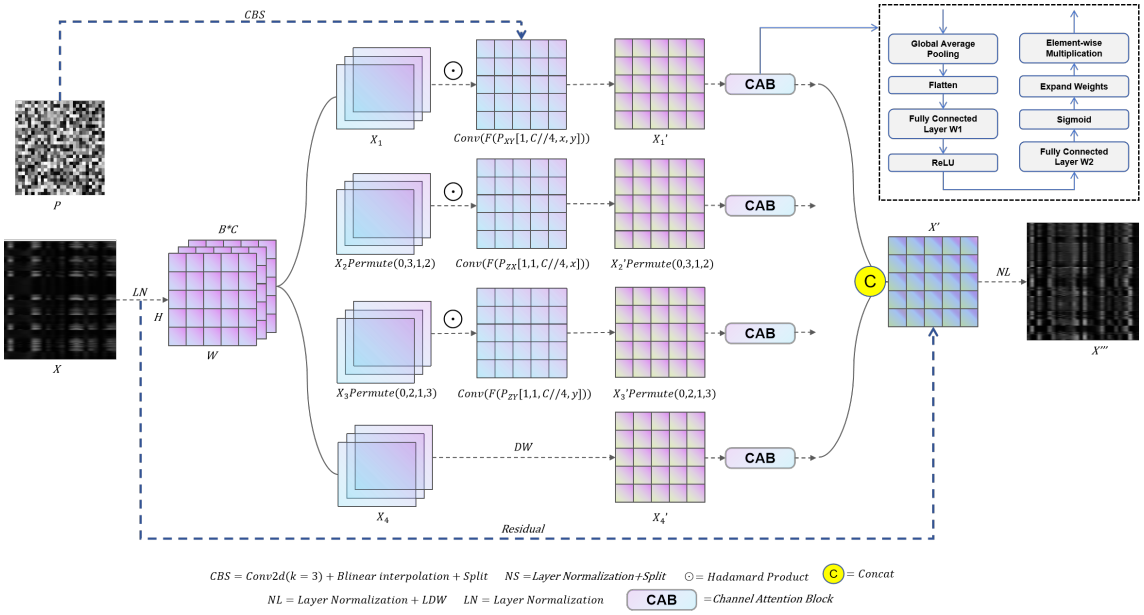

图 4. RGHPA 模块

受到 MHSA 中多头模式的启发，EGE-UNet 基于 HPA 引入了 GHPA 模块 [41] 且表现较好，在本文中使用该模块通过对不同轴上的输入特征进行分组和执行 HPA 操作，从多个角度提取特征信息，具有线性复杂度，从而减小了模型大小，同时将各个角度提取的特征经过 CAB 模块 (Channel Attention Block) 计算通道注意力权重值，再根据权重值重新校准特征图以提取每个通道的全局信息，能够更好地捕捉重要的肿瘤区域特征，抑制不重要或噪声特征，最后结合残差连接操作来增强特征提取和泛化能力，结构如图 4 所示，具体实现流程如下：假设输入张量为 $x \in \mathbb{R}^{B\times C\times H\times W}$，其中 B 是批量大小，H 和 W 分别是输入张量的高度和宽度，C 是通道数。首先，将输入张量 x 进行层归一化，以得到 x_{norm}。随后，根据通道数将 x_{norm} 分成四部分：x_1, x_2, x_3, x_4，每个部分的维度是 $B \times \frac{C}{4} \times H \times W$。每个部分都执行卷积和 GELU 激活函数的组合，以实现线性映射和非线性激活：

假设输入张量为 $x \in \mathbb{R}^{B\times C\times H\times W}$，其中 B 是批量大小，H 和 W 分别是输入张量的高度和宽度，C 是通道数。首先，将输入张量 x 进行层归一化，以得到 x_{norm}。随后，根据通道数将 x_{norm} 分成四部分：x_1, x_2, x_3, x_4，每个部分的维度是 $B \times \frac{C}{4} \times H \times W$。每个部分都执行卷积和 GELU 激活函数的组合，以实现线性映射和非线性激活：

1. 对于 x_1：

$$x_1' = x_1 \odot \text{Conv2d}_x y(params_{xy}) \tag{14}$$

其中，$params_{xy} \in \mathbb{R}^{1\times \frac{C}{4}\times x\times y}$，$\odot$ 表示逐元素相乘，$\text{Conv2d}_x y(params_{xy})$ 表示参数共享的二维卷积操作，具有参数 $params_{xy}$。该操作实现了对 x_1 中每个位置的通道的线性转换。该操作类似于在每个位置的像素上应用不同的权重。

2. 对于 x_2：

$$x_2' = x_2 \odot \text{Conv1d}_z x(params_{zx}) \tag{15}$$

其中，$params_x \in \mathbb{R}^{1\times1\times\frac{C}{4}\times x}$，$\mathrm{Conv1d}_z x(params_{zx})$ 表示参数共享的一维卷积操作，具有参数 $params_{zx}$。该操作实现了对 x_2 中每个位置的通道的线性转换，类似于在水平方向上对每一列像素应用不同的权重。

3. 对于 x_3：

$$x_3' = x_3 \odot \mathrm{Conv1d}_z y(params_{zy}) \tag{16}$$

其中，$params_y \in \mathbb{R}^{1\times1\times\frac{C}{4}\times y}$，$\mathrm{Conv1d}_z y(params_{zy})$ 表示参数共享的一维卷积操作，具有参数 $params_{zy}$。该操作实现了对 x_3 每个位置的通道的线性转换，类似于在垂直方向上对每一行像素应用不同的权重。

4. 对于 x_4：

$$x_4' = DW(x_4) \tag{17}$$

其中，DW 表示深度可分离卷积，x_4 被输送到深度可分离卷积层以进行特征提取和通道维度的转换。输入通过 1×1 卷积进行通道维度的转换操作，然后通过 $GELU$ 激活函数进行非线性转换，随后使用 3×3 的深度可分离卷积操作进行特征提取。

然后，将 x_1', x_2', x_3', x_4' 分别经过 CAB 模块（Channel Attention Block），对通道注意力机制进行建模，从而捕捉输入特征的显著性信息。CAB 模块操作流程如下：

假设其大小为 $B\times C\times H\times W$，首先，经过平均池化层进行全局平均池化，得到大小为 $B\times C\times1\times1$ 的全局特征：

$$Z_l = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{l,i,j} \tag{18}$$

然后将 Z_l 展平为大小为 $B\times C$ 的张量，即 $Z_l \to Z \in \mathbb{R}^{B\times C}$。接着进行通道权重的计算：经过第一个全连接层。将展平后的维度从 C 降到 C/r：

$$Z_l' = W_l^e Z_l \tag{19}$$

其中，$W_l^e \in \mathbb{R}^{C\times\frac{C}{r}}$，$r$ 是缩减系数。

ReLU 激活。对全连接层的输出 Z_l' 进行 $ReLU$ 激活：

$$Z_l'' = \mathrm{ReLU}(Z_l') \tag{20}$$

经过第二个全连接层，将激活后的特征 Z_l'' 输入到第二个全连接层，维度从 $\frac{C}{r}$ 恢复到 C：

$$S_l = W_l^z Z_l'' \tag{21}$$

其中，$W_l^z \in \mathbb{R}^{\frac{C}{r}\times C}$。

Sigmoid 激活。对全连接层的输出 S_l 进行 $Sigmoid$ 激活，得到其每个通道的权重范围在 $(0,1)$ 之间：

$$S_l = \sigma(S_l) \tag{22}$$

特征重新校准。将通道权重 S_l 重新调整为 $B\times C\times1\times1$ 的形状，然后扩展为 $B\times C\times H\times W$ 的张量，最后与输入特征相乘，实现重新权重。得到输出特征 x_4'。

$$x_4' = x_4 \odot S_l \tag{23}$$

同理可得各个角度的特征图 x_1', x_2', x_3', x_4'。CAB 模块（Channel Attention Block）能够有效地计算每个通道的重要性权重，并根据这些权重重新校准特征图，从而提高模型在乳腺肿瘤分割任务中的表现。

最后，利用残差连接将 x_4' 与原输入 x_4 相加以得到最终的特征图：

$$x_4'' = x_4' + x_4 \tag{24}$$

最后一层采用深度分离卷积，第一层卷积层使用 3×3 的卷积核，第二个卷积层使用 1×1 的卷积核，与 DW 相反。最终的卷积操作用于提取更深层次的特征：

$$x''' = \mathrm{LDW}(\mathrm{LayerNorm}(x'')) \tag{25}$$

表 1. RGPA Module Implementation Pseudo - code

RGPA Pseudo - code:
Input: X, the feature map with shape [B, C, H, W]
Output: Out, the feature map with shape [B, C, H, W]
Params: a, the hyperparameter and default by 8 in this paper
b, the hyperparameter and default by 8 in this paper
P_{xy}, the randomly initialized tensor with shape [1, C/4, a, b]
P_{zx}, the randomly initialized tensor with shape [1, 1, C/4, a]
P_{zy}, the randomly initialized tensor with shape [1, 1, C/4, b]
Operator: DW, Depthwise Separable Convolution
LN, LayerNorm
BI, Bilinear interpolation
CAB, Channel Attention Block
x1, x2, x3, x4 = torch.chunk(LN(X), 4, dim = 1)
x1 = x1 * DW(BI(P_{xy})), DW(x4)
x1 = CAB(x1)
x2 = (x2.permute(0, 3, 1, 2) * DW(BI(P_{zx}))).permute(0, 2, 3, 1)
x2 = CAB(x2)
x3 = (x3.permute(0, 2, 1, 3) * DW(BI(P_{zy}))).permute(0, 2, 1, 3)
x3 = CAB(x3)
x4 = DW(x4)
x4 = CAB(x4)
Out = DW(LN(torch.cat([x1, x2, x3, x4], dim = 1)))
Out = Out + X # Residual connection

表 2. CAB Module Implementation Pseudo - code

CAB Pseudo - code:
Input: x, the input tensor with shape [B, C, H, W]
Output: x * y, the attention - weighted tensor
Params: channel, the number of input channels
reduction, the channel reduction ratio (default = 16)
Operator: avg_pool, Adaptive Average Pooling
fc, Fully Connected layers
ReLU, Rectified Linear Unit
Sigmoid, Sigmoid Activation
get b, c, _ from x.size()
y = avg_pool(x).view(b, c)
y = fc(y).view(b, c, 1, 1)
return x * y.expand_as(x)

4 实验

4.1 数据集

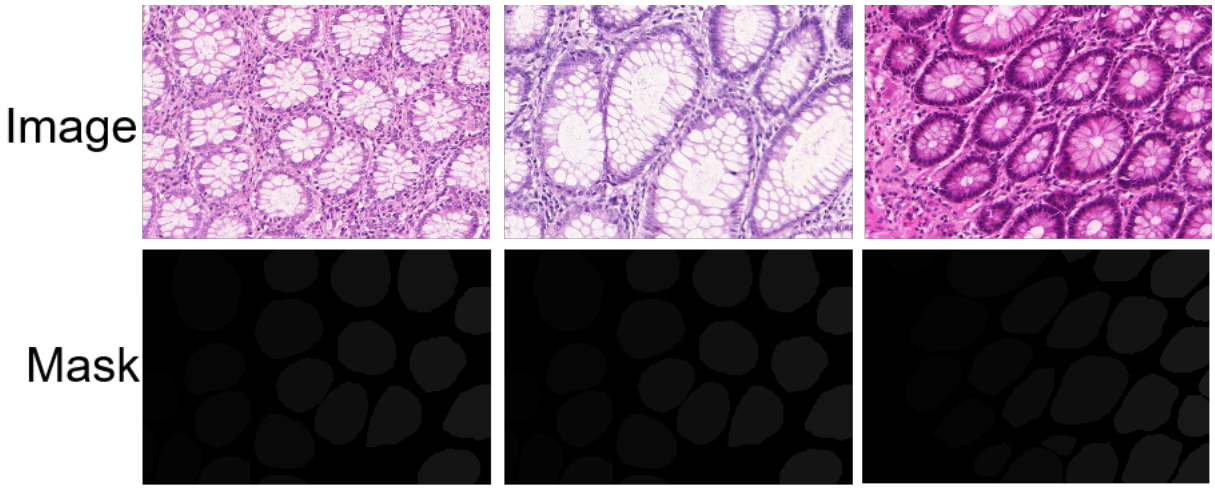

图 5. GlaS 数据集

GlaS 数据集是用于结直肠腺体组织的分割任务，可视化示例如图 5 所示，由挑战赛 MICCAI 2015 Gland Segmentation Challenge 发布。共包括 165 张图像，这些图像来源于 16 个 HE 染色的组织学切片，这些切片代表了 T3 或 T4 阶段的结直肠腺癌，每个切片来自不同的患者，并且切片在不同的时间在实验室中处理，这些组织学切片被数字化为整个切片图像（WSIs），使用 Zeiss MIRAX MIDI 滑片扫描仪完成，像素分辨率为 0.465 微米，图像尺寸为 775×522，训练图像和测试图像的数量分别为 85（37 个良性，48 个恶性）和 80（37 个良性，43 个恶性）。[43]

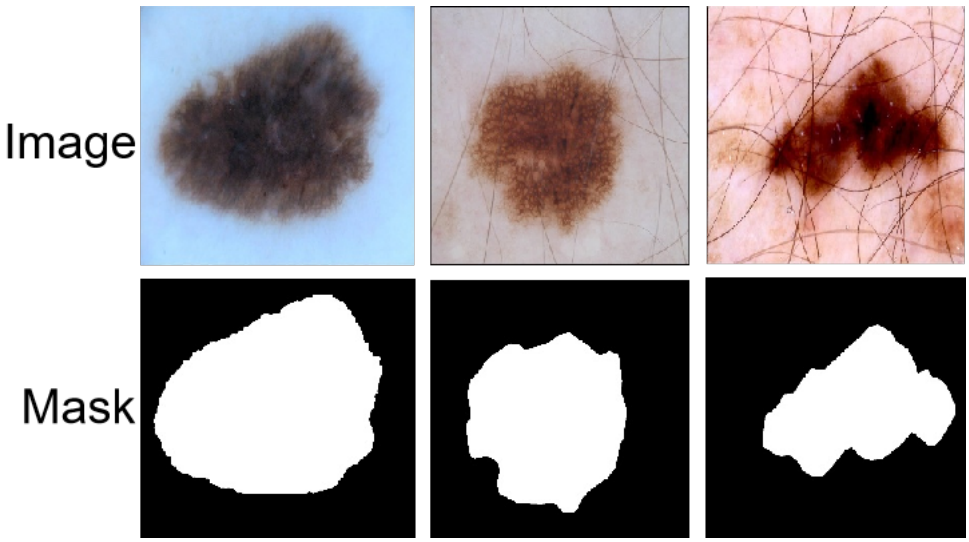

图 6. ISIC2017 数据集

ISIC2017 数据集是用于皮肤病变分割任务，可视化示例如图 6 所示，由 International Skin Imaging Collaboration (ISIC) 发布的一个大规模的皮肤镜图像数据集共 2750 张，该数据集的目标是帮助参与者开发图像分析工具，以便从皮肤镜图像自动诊断黑色素瘤，图像尺寸为 256×256，该数据集包含 2000 张训练图像，150 张验证图像和 600 张测试图像。[11]

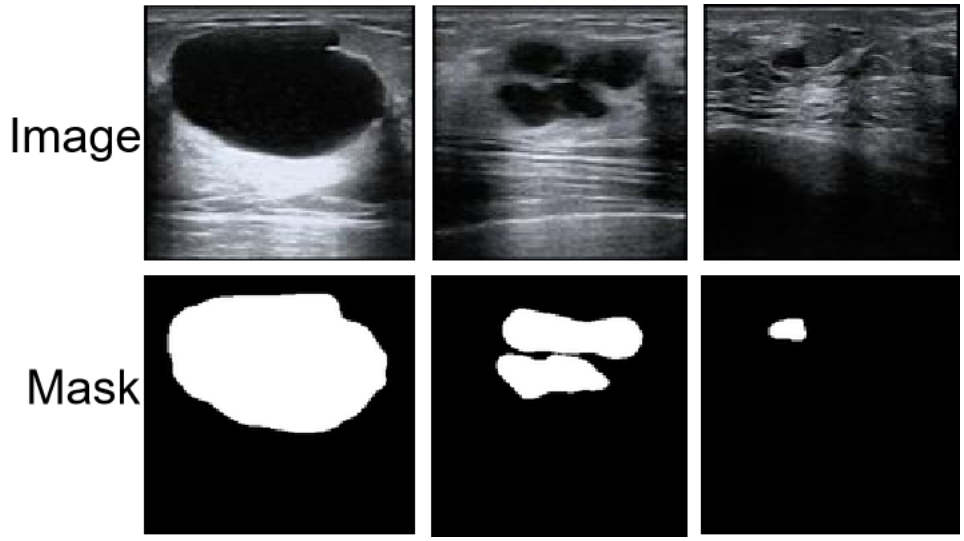

图 7. SZU-BCH-BUS2215 数据集

SZU-BCH-BUS2215 数据集用于乳腺结节超声图像分割任务，可视化示例如图 7 所示，由深圳大学赖志辉教授团队提供，该数据集的主要目的是促进乳腺结节的自动识别和分类算法的研究和开发，图像尺寸为 256×256，包含 2000 张训练图像，215 张测试图像。

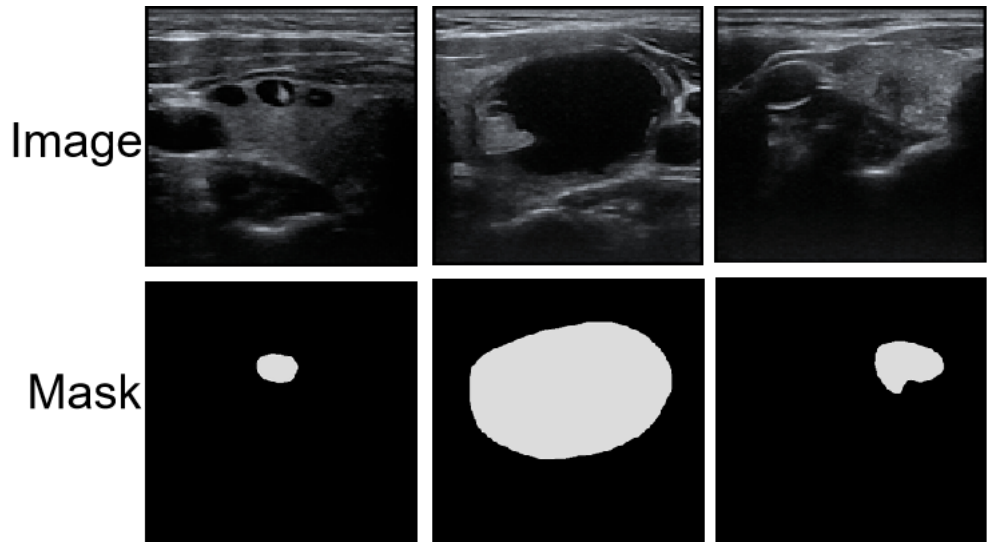

图 8. SZU-BCH-TUS221 数据集

SZU-BCH-TUS221 数据集用于甲状腺结节超声图像分割任务，可视化示例如图 8 所示，由深圳大学赖志辉教授团队提供，该数据集的主要目的是促进甲状腺结节的自动识别和分类算法的研究和开发，图像尺寸为 256×256，包含 2000 张训练图像，215 张测试图像。

本次复现在以上四个数据集上均进行了实验，分别利用模型的全监督和半监督训练策略进行实验。

4.2 数据集 LF 和 HF 特征可视化

观察图 9-图 12 可以看出，在甲状腺肿瘤图像和乳腺肿瘤图像中高频信息有许多噪声，并且包含除病变区域外的组织特征，这对于模型提取特征过程加大了难度。

本次实验数据集划分设置如下：

半监督训练：80% 无标签训练集 / 20% 有标签训练集 / 验证集

全监督训练：100% 有标签训练集 / 验证集

(1) GlaS 数据集：

半监督训练：无标签训练集 124 张 / 有标签训练集 31 张 / 验证集 10 张

全监督训练：有标签训练集 155 张 / 验证集 10 张

(2) ISIC2017 数据集：

半监督训练：无标签训练集 1980 张 / 有标签训练集 495 张 / 验证集 275 张

全监督训练：有标签训练集 2475 张 / 验证集 275 张

(3) SZU-BCH-BUS2215 数据集：

半监督训练：无标签训练集 1600 张 / 有标签训练集 400 张 / 验证集 215 张

全监督训练：有标签训练集 2000 张 / 验证集 215 张

(4) SZU-BCH-TUS2215 数据集：

半监督训练：无标签训练集 1600 张 / 有标签训练集 400 张 / 验证集 215 张

全监督训练：有标签训练集 2000 张 / 验证集 215 张

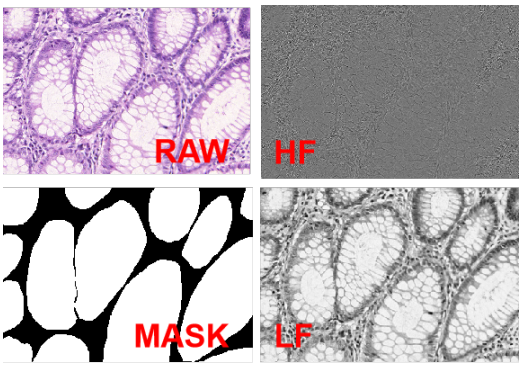

图 9. 以 GlaS 为例，可视化 LF 和 HF 结果

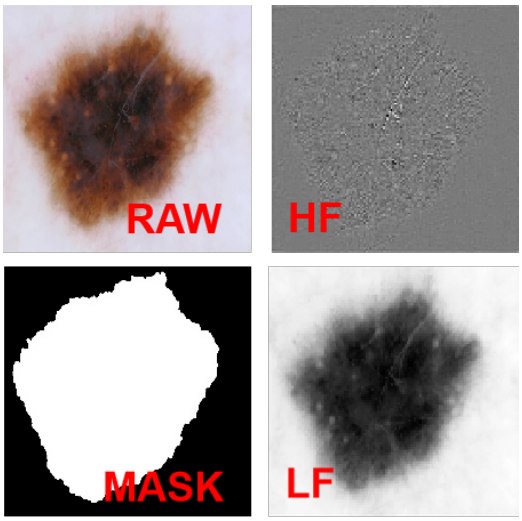

图 10. 以 ISIC2017 为例，可视化 LF 和 HF 结果

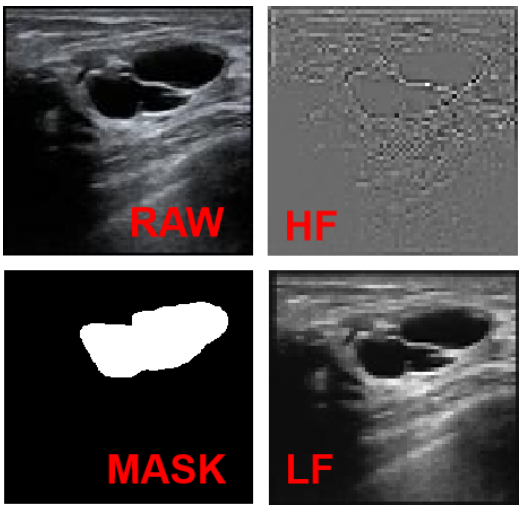

图 11. 以 SZU-BCH-BUS2215 为例，可视化 LF 和 HF 结果

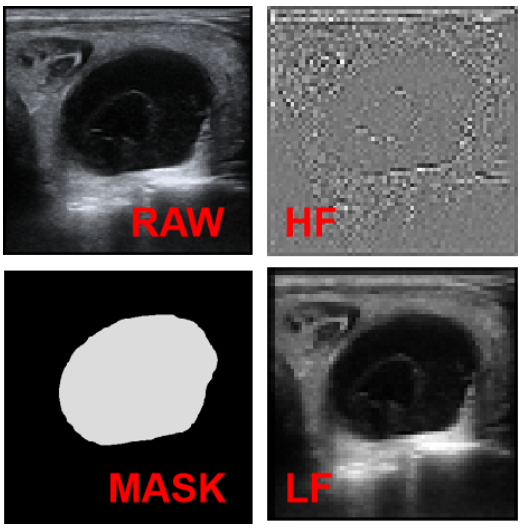

图 12. 以 SZU-BCH-TUS2215 为例，可视化 LF 和 HF 结果

4.3 评价

本文使用 Jaccard 指数 (JI)、Dice 系数 (DC)、第 95 百分位 Hausdorff 距离 (95HD) 和平均表面距离 (ASD) 作为性能指标来评估分割结果。JI 和 DC 强调像素精度，而 95HD 和 ASD 强调边界精度。这些指标被广泛用于生物医学图像分割的基准性能。

4.4 实现细节

本文使用 PyTorch 实现该模型。所有模型的训练和推理原论文作者都在 4 台 NVIDIA GeForce RTX3090 上进行，本次复现在 2 台 NVIDIA GeForce RTX4090D 上进行。本文使用带动量的 SGD 来训练模型，动量设置为 0.9，权值衰减设置为 0.00005。epoch 的个数设置为 200。学习率每 50 个 epoch 衰减 0.5。无监督损失函数的权值 λ 线性增加 epoch：

$$\lambda = \lambda_{max} * \frac{epoch}{max\_epoch}$$

4.5 实验结果

表 3. XNet 原论文结果

Method	Datasets	JI↑	DC↑	95HD↓	ASD↓
Fully - Supervised (100%)	GlaS	84.77	91.76	7.87	1.55
	ISIC2017	73.94	85.02	9.81	4.14
Semi - Supervised (20%+80%)	GlaS	80.89	89.44	9.86	2.07
	ISIC2017	71.17	83.16	11.46	4.73

表 4. XNet 复现结果

Method	Datasets	JI↑	DC↑	95HD↓	ASD↓
Fully - Supervised (100%)	GlaS	84.84	91.80	5.82	1.45
	ISIC2017	79.79	88.76	7.26	2.88
	SZU-BCH-BUS2215	74.96	85.69	7.23	2.13
	SZU-BCH-TUS2215	72.08	83.77	7.65	2.94
Semi - Supervised (20%+80%)	GlaS	79.82	88.78	7.95	1.86
	ISIC2017	66.13	79.61	11.94	4.97
	SZU-BCH-BUS2215	72.24	83.88	8.41	2.46
	SZU-BCH-TUS2215	67.29	80.44	10.38	4.11

表 5. XNetv2 原论文结果

Method	Datasets	JI↑	DC↑	95HD↓	ASD↓
Fully - Supervised (100%)	GlaS	84.03	91.32	9.12	1.79
	ISIC2017	76.04	86.39	3.86	9.78
Semi - Supervised (20%+80%)	GlaS	83.17	90.81	8.54	1.75
	ISIC2017	74.07	85.11	3.97	9.95

表 6. XNetv2 复现结果

Method	Datasets	JI↑	DC↑	95HD↓	ASD↓
Fully - Supervised (100%)	GlaS	87.64	93.41	4.01	1.06
	ISIC2017	80.00	88.89	6.98	2.67
	SZU-BCH-BUS2215	75.53	86.06	7.76	2.05
	SZU-BCH-TUS2215	73.86	84.96	8.06	2.76
Semi - Supervised (20%+80%)	GlaS	84.92	91.85	5.75	1.30
	ISIC2017	79.58	88.63	8.14	2.86
	SZU-BCH-BUS2215	76.35	86.59	6.94	1.87
	SZU-BCH-TUS2215	74.45	85.35	7.02	2.47

复现结果对比如表 3-表 6 所示，在 XNet 及 XNetv2 模型上分别进行了复现对比实验，使用了原论文中的两个公共 2D 数据集，使用了私有的两个 2D 数据集对模型进行评价，根据结果可以看到复现结果与原论文中的结果基本相同，微小差异主要是因为数据集的划分等因素。在两个私有数据集中可以看到，在甲状腺和乳腺超声图像数据集中，缺少高频信息且充满噪声信息，但模型任然能够较为准确的对病变区域进行分割，说明该模型具有较强的泛化能力。

分割结果可视化对比图如图 13-图 16 所示，由于 XNet 强调 HF 信息，当图像几乎没有 HF 信息时，XNet 的性能会受到负面影响，与直接使用小波变换生成的 LF 和 HF 图像作为

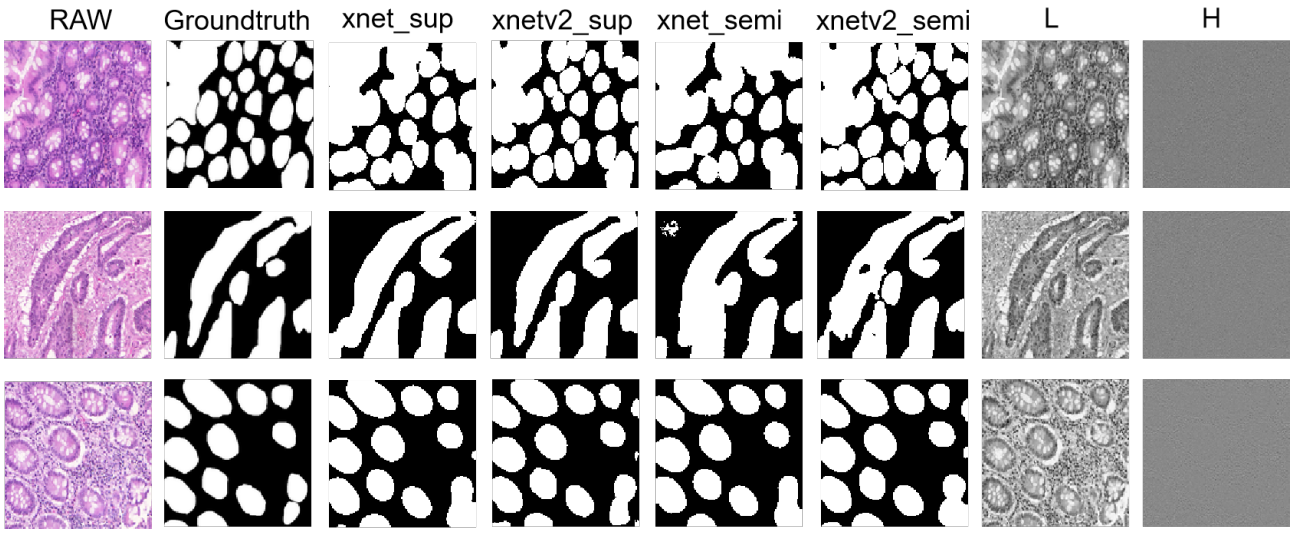

图 13. GlaS 数据集分割结果可视化对比

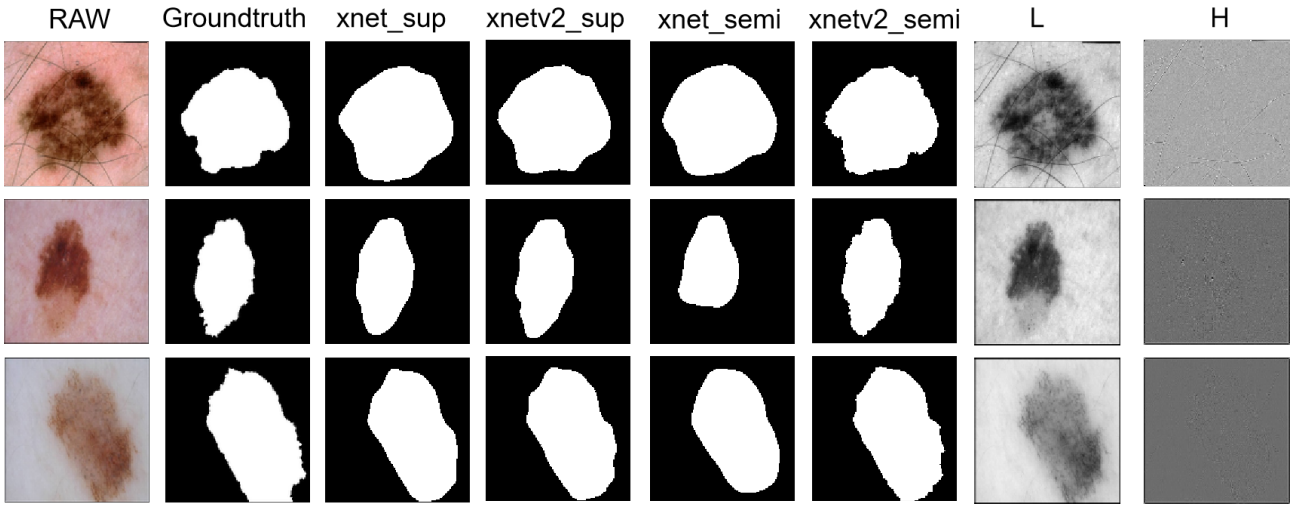

图 14. ISIC2017 数据集分割结果可视化对比

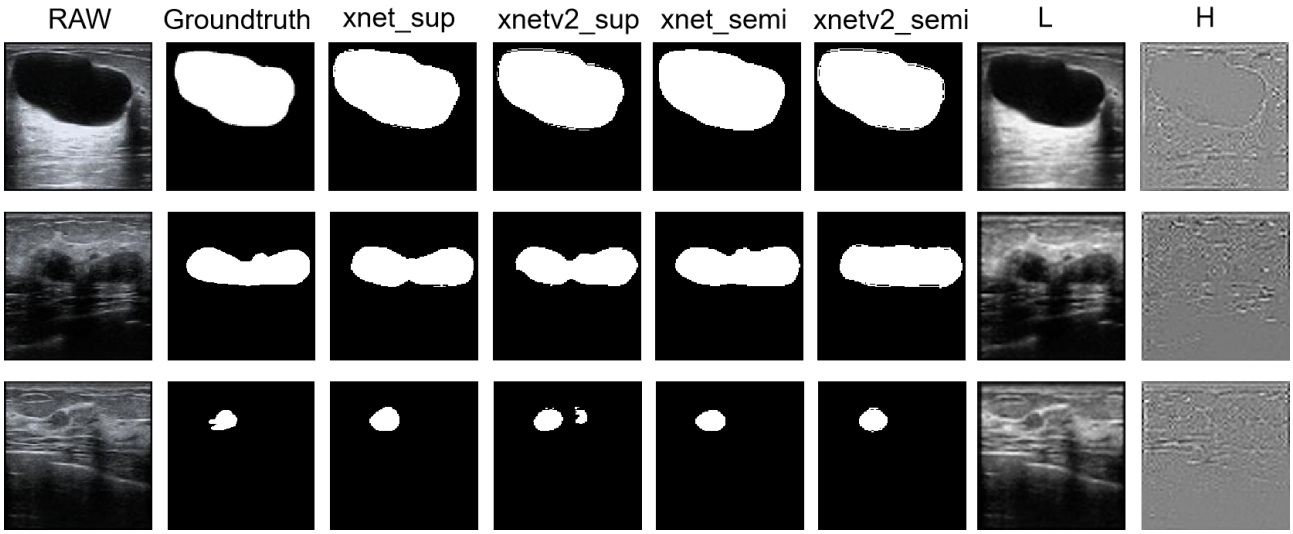

图 15. SZU-BCH-BUS2215 数据集分割结果可视化对比

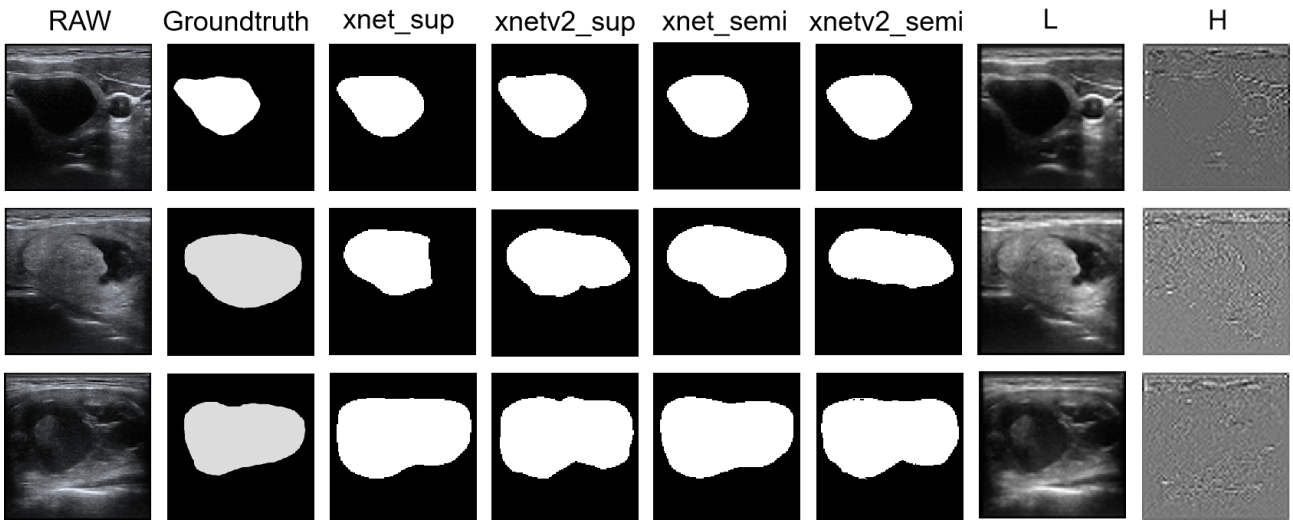

图 16. SZU-BCH-TUS2215 数据集分割结果可视化对比

表 7. RGHPA 模块消融实验

Method	Model	JI ↑	DC ↑	95HD ↓	ASD ↓
Fully - Supervised (100%)	XNetv2+RGHPA	87.81	94.79	3.86	1.02
	XNetv2	87.64	93.41	4.01	1.06
	XNet+RGHPA	83.90	91.24	6.59	1.51
	XNet	84.84	91.80	5.82	1.45
Semi - Supervised (20%+80%)	XNetv2+RGHPA	84.82	93.79	4.86	1.15
	XNetv2	84.92	91.85	5.75	1.30
	XNet+RGHPA	77.67	87.43	7.87	1.91
	XNet	79.82	88.78	7.95	1.86

输入不同，XNet v2 对 LF 和 HF 图像进行图像级互补融合。将融合结果与原始图像一起输入到三个不同的网络 (主网络、LF 网络和 HF 网络) 中，生成用于一致性学习的分割预测。但是在甲状腺和乳腺超声图像中缺少大量高频信息的同时充满许多噪声，降低了模型分割性能。

为了增加模型的通道注意力机制，通过对不同轴上的输入特征进行分组和执行 HPA 操作，从多个角度提取特征信息，能够更好地捕捉重要的肿瘤区域特征，抑制不重要或噪声特征，最后结合残差连接操作来增强特征提取和泛化能力，修改后的模型结构如图 17 和图 18 所示。

RGHPA 模块的消融实验结果如表 7 所示，观察到 RGHPA 模块虽然可以更有效地传播梯度，并有助于防止过拟合，但是从指标结果来看分割效果有所减弱，主要原因可能是由于模型对于单通道的高频和低频信息分别进行特征提取，而 RGHPA 模块主要计算了不同通道间的注意力，因此在该网络架构中无法充分利用通道信息注意力策略来增强捕捉重要的病变区域特征并抑制噪声特征。

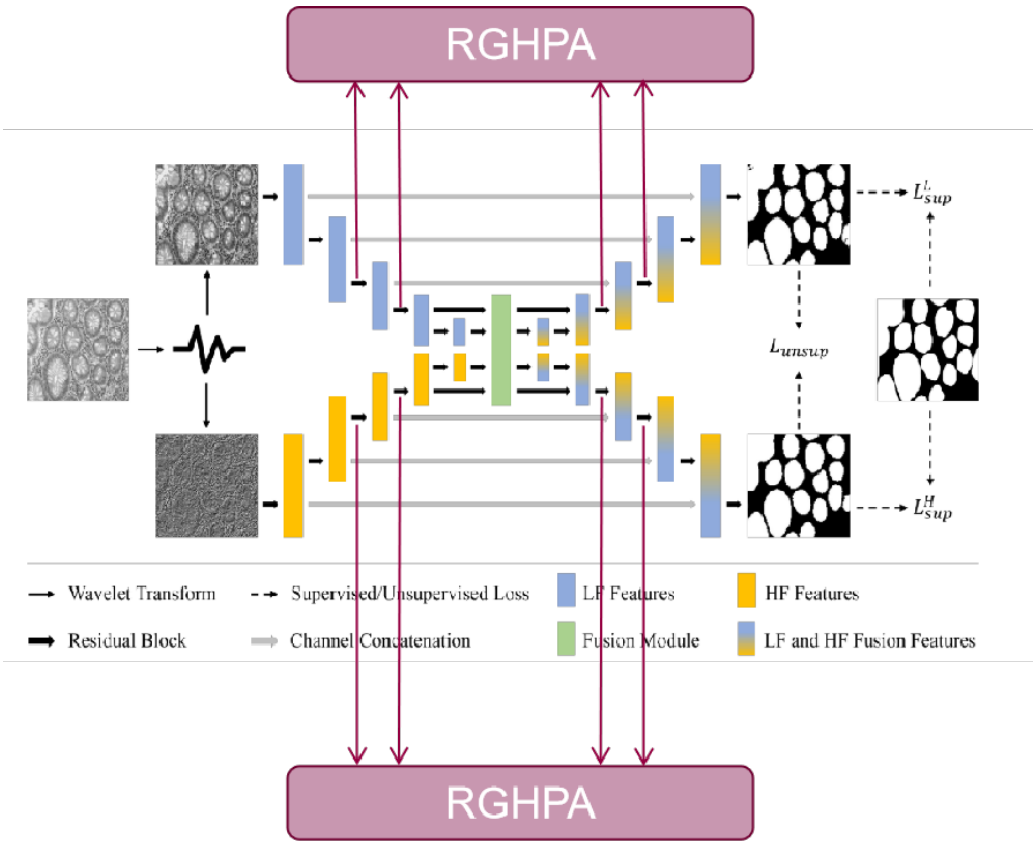

图 17. XNet 加入 RGHPA 模块

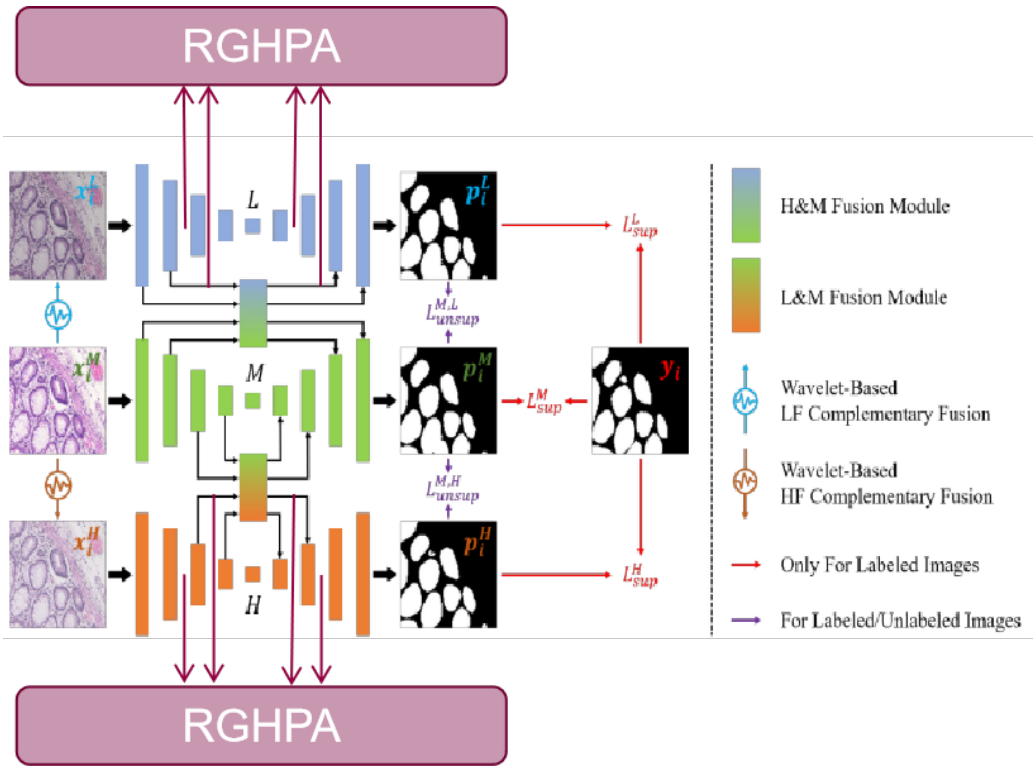

图 18. XNetv2 加入 RGHPA 模块

5 总结与展望

本文复现了基于小波的低频（LF）与高频（HF）融合模型 XNet 及 XNetv2，旨在解决生物医学图像语义分割中的全监督和半监督学习问题。该研究针对现有模型的局限性，如多数模型仅支持单一学习方案、全监督模型对图像固有信息利用不足、半监督模型扰动策略存在学习偏差等展开深入探索。通过小波变换生成 LF 和 HF 图像，并将其输入 XNet。该网络利用融合模块将 LF 和 HF 特征融合，再通过双分支解码器生成 LF 和 HF 分支的分割预测。对于全监督训练，通过最小化标注图像的监督损失和双输出一致性损失来优化模型；半监督训练则通过最小化标注图像的监督损失和未标注图像的双输出一致性损失进行优化。

本文提出的 RGHPA 模块，通过对不同轴上的输入特征进行分组和执行 HPA 操作，从多个角度提取特征信息，并结合残差连接操作增强特征提取和泛化能力。小波变换能够将原始图像分解为 LF 和 HF 分量，使模型更好地关注 LF 语义和 HF 细节，缓解人为设计扰动带来的学习偏差。

在复现过程中，分别在 GlaS、ISIC2017、SZU-BCH-BUS2215 和 SZU-BCH-TUS2215 四个数据集上进行实验。结果表明，复现结果与原论文基本相同，在甲状腺和乳腺超声图像等私有数据集上，模型仍能较为准确地对病变区域进行分割，展现出较强的泛化能力。同时，对 RGHPA 模块的消融实验显示，该模块虽能有效传播梯度、防止过拟合，但在当前网络架构下，分割效果有所减弱，可能是由于无法充分利用通道信息注意力策略。本文复现了 XNet 模型，在生物医学图像的全监督和半监督语义分割任务中取得了一定成果，图像本质上是离散的非平稳信号，而小波变换可以有效地分析它们，但 RGHPA 模块在该架构下的应用仍需进一步优化，未来可探索更适配的注意力策略，以提升模型能够更好的利用小波变换在生物医学图像分割中的性能。

参考文献

[1] Chartsias Agisilaos, Joyce Thomas, Papanastasiou Giorgos, Semple Scott, Williams Michelle, Newby David, Dharmakumar Rohan, and Tsaftaris Sotirios A. Factorised spatial representation learning: Application in semi-supervised myocardial segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 490–498*, 2018.

[2] Dosovitskiy Alexey, Beyer Lucas, Kolesnikov Alexander, Weissenborn Dirk, Zhai Xiaohua, Unterthiner Thomas, Dehghani Mostafa, Minderer Matthias, Heigold Georg, and Gelly Sylvain. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[3] Hatamizadeh Ali, Tang Yucheng, Nath Vishwesh, Yang Dong, Myronenko Andriy, Landman Bennett, Roth Holger R, and Xu Daguang. Unetr: Transformers for 3d medical image segmentation. *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 574–584*, 2022.

[4] Tarvainen Antti and Valpola Harri. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems, 30*, 2017.

[5] Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N, Kaiser Łukasz, and Polosukhin Illia. Attention is all you need. *Advances in neural information processing systems, 30*, 2017.

[6] Seyed Majid Azimi, Peter Fischer, Marco Korner, and Peter Reinartz. Aerial lanenet: Lane-marking semantic segmentation in aerial imagery using wavelet-enhanced cost-sensitive symmetric fully convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(5):2920–2938, 2018.

[7] Hu Cao, Yueyue Wang, Joy Chen, et al. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv preprint arXiv:2105.05537*, 2021.

[8] Jie Chen, Yutong Lu, Qihang Yu, et al. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[9] L. Chen, Y. Wu, A. M. DSouza, A. Z. Abidin, A. Wismüller, and C. Xu. Mri tumor segmentation with densely connected 3d cnn. In E. D. Angelini and B. A. Landman, editors, *Medical Imaging 2018: Image Processing*, volume 10574, page 105741F. International Society for Optics and Photonics. SPIE, 2018.

[10] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[11] Noel CF Codella, David Gutman, M Emre Celebi, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 168–172. IEEE, 2018.

[12] J. Dolz, C. Desrosiers, and I. Ben Ayed. Ivd-net: Intervertebral disc localization and segmentation in mri with a multi-modal unet. In *Computational Methods and Clinical Applications for Spine Imaging*, pages 130–143, 2019.

[13] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, and I. Ben Ayed. Hyperdense-net: A hyper-densely connected cnn for multimodal image segmentation. *IEEE Transactions on Medical Imaging*, 38(5):1116–1126, 2019.

[14] Yiping Duan, Fang Liu, Licheng Jiao, et al. Sar image segmentation based on convolutional-wavelet neural network and markov random field. *Pattern Recognition*, 64:255–267, 2017.

[15] Mingyuan Fan, Shenqi Lai, Junshi Huang, et al. Rethinking bisenet for real-time semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9716–9725, 2021.

[16] Feng Gao, Xiao Wang, Yunhao Gao, et al. Sea ice change detection in sar images based on convolutional-wavelet neural networks. *IEEE Geoscience and Remote Sensing Letters*, 16(8):1240–1244, 2019.

[17] Y. Gao, X. Fu, Y. Chen, C. Guo, and J. Wu. Post-pandemic healthcare for covid-19 vaccine: Tissue-aware diagnosis of cervical lymphadenopathy via multi-modal ultrasound semantic segmentation. *Applied Soft Computing*, 133:109947, 2023.

[18] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. Generative adversarial nets. *Advances in Neural Information Processing systems*, 27, 2014.

[19] Zhao Hengshuang, Shi Jianping, Qi Xiaojuan, Wang Xiaogang, and Jia Jiaya. Pyramid scene parsing network. *In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2881–2890*, 2017.

[20] Cao Hu, Wang Yueyue, Chen Joy, Jiang Dongsheng, Zhang Xiaopeng, Tian Qi, and Wang Manning. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv preprint arXiv:2105.05537*, 2021.

[21] Huimin Huang, Lanfen Lin, Ruofeng Tong, et al. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059. IEEE, 2020.

[22] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021.

[23] Yuan Jianlong, Liu Yifan, Shen Chunhua, Wang Zhibin, and Li Hao. A simple baseline for semi-supervised semantic segmentation with strong data augmentation. *In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 8229–8238*, 2021.

[24] Qiangguo Jin, Hui Cui, Changming Sun, et al. Semi-supervised histological image segmentation via hierarchical consistency enforcement. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 3–13. Springer, 2022.

[25] Springenberg Jost Tobias. Unsupervised and semi-supervised learning with categorical generative adversarial networks. *arXiv preprint arXiv:1511.06390*, 2015.

[26] Upadhyay Kamini, Agrawal Monika, and Vashist Praveen. Wavelet based fine-to-coarse retinal blood vessel extraction using u-net model. *In 2020 International Conference on Signal Processing and Communications (SPCOM), pages 1–5*, 2020.

[27] Thomas Martin Lehmann, Claudia Gonner, and Klaus Spitzer. Survey: Interpolation methods in medical image processing. *IEEE transactions on medical imaging*, 18(11):1049–1075, 1999.

[28] Yu Lequan, Wang Shujun, Li Xiaomeng, Fu Chi-Wing, and Heng Pheng-Ann. Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 605–613*, 2019.

[29] Qiufu Li and Linlin Shen. Wavesnet: Wavelet integrated deep networks for image segmentation. In *Pattern Recognition and Computer Vision: 5th Chinese Conference, PRCV 2022, Shenzhen, China, November 4–7, 2022, Proceedings, Part IV*, pages 325–337. Springer, 2022.

[30] Yang Lihe, Zhuo Wei, Qi Lei, Shi Yinghuan, and Gao Yang. St++: Make self-training work better for semi-supervised semantic segmentation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4268–4277*, 2022.

[31] Pengju Liu, Hongzhi Zhang, Wei Lian, et al. Multi-level wavelet convolutional neural networks. *IEEE Access*, 7:74973–74985, 2019.

[32] Xiaofeng Liu, Fangxu Xing, Nadya Shusharina, et al. Act: Semi-supervised domain-adaptive medical image segmentation with asymmetric co-training. *arXiv preprint arXiv:2206.02288*, 2022.

[33] Zhuang Liu, Hanzi Mao, Chunyu Wu, et al. A convnet for the 2020s. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022.

[34] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[35] Xiangde Luo, Jieneng Chen, Tao Song, et al. Semi-supervised medical image segmentation through dual-task consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8801–8809, 2021.

[36] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.

[37] D. Nie, L. Wang, Y. Gao, and D. Shen. Fully convolutional networks for multi-modality isointense infant brain image segmentation. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1342–1345, 2016.

[38] Oktay Ozan, Schlemper Jo, Folgoc Loic Le, Lee Matthew, Heinrich Mattias, Misawa Kazunari, Mori Kensaku, McDonagh Steven, Hammerla Nils Y, and Kainz Bernhard. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[39] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[40] Soumyabrata Roy, Gregor Koehler, Christoph Ulrich, et al. Mednext: transformer-driven scaling of convnets for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 405–415. Springer Nature Switzerland, 2023.

[41] J. Ruan, M. Xie, J. Gao, et al. Ege-unet: an efficient group enhanced unet for skin lesion segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 481–490. Springer Nature Switzerland, 2023.

[42] B. Shareef, M. Xian, and A. Vakanski. Stan: Small tumor-aware network for breast ultrasound image segmentation. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, 2020.

[43] Korsuk Sirinukunwattana, Josien PW Pluim, Hao Chen, et al. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35:489–502, 2017.

[44] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5693–5703, 2019.

[45] Miyato Takeru, Maeda Shin-ichi, Koyama Masanori, and Ishii Shin. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(8):1979–1993*, 2018.

[46] I. O. Tolstikhin, N. Houlsby, A. Kolesnikov, et al. Mlp-mixer: An all-mlp architecture for vision. *Advances in Neural Information Processing Systems*, 34:24261–24272, 2021.

[47] Vu Tuan-Hung, Jain Himalaya, Bucher Maxime, Cord Matthieu, and Perez Patrick. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2517–2526*, 2019.

[48] J. M. J. Valanarasu and V. M. Patel. Unext: Mlp-based rapid medical image segmentation network. *arXiv preprint arXiv:2203.04967*, 2022.

[49] G. Wang, W. Li, S. Ourselin, and T. Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 178–190, 2018.

[50] Cui Wenhui, Liu Yanlin, Li Yuxing, Guo Menghao, Li Yiming, Li Xiuli, Wang Tianle, Zeng Xiangzhu, and Ye Chuyang. Semi-supervised brain lesion segmentation with an adapted mean teacher model. *In International Conference on Information Processing in Medical Imaging, pages 554–565*, 2019.

[51] Wang Wenxuan, Chen Chen, Ding Meng, Yu Hong, Zha Sen, and Li Jiangyun. Transbts: Multimodal brain tumor segmentation using transformer. *In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 109–119*, 2021.

[52] Sung Woo, Sourav Debnath, Ruibin Hu, et al. Convnext v2: Co-designing and scaling convnets with masked autoencoders. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16133–16142, 2023.

[53] Luo Xiangde, Wang Guotai, Liao Wenjun, Chen Jieneng, Song Tao, Chen Yinan, Zhang Shichuan, Metaxas Dimitris N, and Zhang Shaoting. Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. *Medical Image Analysis, 80:102517*, 2022.

[54] Luo Xiangde, Chen Jieneng, Song Tao, and Wang Guotai. Semi-supervised medical image segmentation through dual-task consistency. *In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 8801–8809*, 2021.

[55] Li Xiaomeng, Yu Lequan, Chen Hao, Fu Chi-Wing, Xing Lei, and Heng Pheng-Ann. Transformation-consistent self-ensembling model for semisupervised medical image segmentation. *IEEE Transactions on Neural Networks and Learning Systems, 32(2):523–534*, 2020.

[56] Yin Xin and Xu Xiaoyang. A method for improving accuracy of deeplabv3+ semantic segmentation model based on wavelet transform. *In Communications, Signal Processing, and Systems: Proceedings of the 10th International Conference on Communications, Signal Processing, and Systems, Vol. 2, pages 315–320*, 2022.

[57] Ouali Yassine, Hudelot Celine, and Tami Myriam. Semi-supervised semantic segmentation with cross-consistency training. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12674–12684*, 2020.

[58] Xie Yutong, Zhang Jianpeng, Shen Chunhua, and Xia Yong. Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. *In International conference on medical image computing and computer-assisted intervention, pages 171–180*, 2021.

[59] Liu Ze, Lin Yutong, Cao Yue, Hu Han, Wei Yixuan, Zhang Zheng, Lin Stephen, and Guo Baining. Swin transformer: Hierarchical vision transformer using shifted windows. *In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 10012–10022*, 2021.

[60] Y. Zhang, J. Yang, J. Tian, Z. Shi, C. Zhong, Y. Zhang, and Z. He. Modality-aware mutual learning for multi-modal medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention –MICCAI 2021*, pages 589–599, 2021.

[61] Feng Zhengyang, Zhou Qianyu, Gu Qiqi, Tan Xin, Cheng Guangliang, Lu Xuequan, Shi Jianping, and Ma Lizhuang. Dmt: Dynamic mutual training for semi-supervised learning. *Pattern Recognition, page 108777*, 2022.

[62] C. Zhou, C. Ding, Z. Lu, X. Wang, and D. Tao. One-pass multi-task convolutional neural networks for efficient brain tumor segmentation. In *Medical Image Computing and Computer Assisted Intervention –MICCAI 2018*, pages 637–645, 2018.

[63] Yanfeng Zhou, Jiaxing Huang, Chenlong Wang, Le Song, and Ge Yang. Xnet: Wavelet-based low and high frequency fusion networks for fully-and semi-supervised semantic segmentation of biomedical images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21085–21096, 2023.

[64] Yanfeng Zhou, Lingrui Li, Zichen Wang, Guole Liu, Ziwen Liu, and Ge Yang. Xnet v2: Fewer limitations, better results and greater universality. *arXiv preprint arXiv:2409.00947*, 2024.

[65] Y. Zhu, X. Wang, L. Chen, and R. Nie. Cefusion: Multi-modal medical image fusion via cross encoder. *IET Image Processing*, 16(12):3177–3189, 2022.

[66] Zhou Zongwei, Siddiquee Md Mahfuzur Rahman, Tajbakhsh Nima, and Liang Jianming. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging, 39(6):1856–1867*, 2019.

[67] Ozgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, et al. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.