

DA-SwinUnet: 基于方向注意力模块与 SwinUnet的股骨头坏死分割模型

刁瑞新, 王恒澳, 王 琨, 计晓斐*

青岛大学计算机科学技术学院, 山东 青岛

收稿日期: 2026年3月3日; 录用日期: 2026年4月8日; 发布日期: 2026年4月21日

摘 要

股骨头坏死是临床常见的骨科疾病, 早期诊断与精准病灶定位对有效治疗至关重要。针对当前影像诊断主观性强、缺乏客观性的问题, 以及现有分割方法未能充分利用图像解剖一致性的问题, 本研究提出了一种融合定向注意力机制的DA-SwinUnet模型。该模型在Swin-Unet架构基础上引入定向注意力机制以增强方向性特征提取, 并通过加权跳跃连接优化网络性能, 有效降低了特征融合阶段的计算冗余, 提升了对不同大小病灶的分割精度; 同时整合方向连通向量概念以进一步强化解剖一致性约束。在股骨头坏死临床MRI数据集上的实验结果显示, DA-SwinUnet模型的Dice系数达到87.3, IoU达到78.7, 性能优于传统CNN模型及其他基于Transformer的分割方法。研究表明, DA-SwinUnet能够在保持解剖一致性的前提下实现股骨头坏死病灶的精确分割, 显著提升分割性能, 为疾病的早期诊断和临床治疗提供了有力支持, 具有广阔的前景。

关键词

股骨头坏死, 定向注意力机制, 图像分割

DA-SwinUnet: A Segmentation Model for Avascular Necrosis of the Femoral Head Based on Directional Attention Module and SwinUnet

Ruixin Diao, Heng'ao Wang, Kun Wang, Xiaofei Ji*

College of Computer Science & Technology, Qingdao University, Qingdao Shandong

Received: March 3, 2026; accepted: April 8, 2026; published: April 21, 2026

*通讯作者。

文章引用: 刁瑞新, 王恒澳, 王琨, 计晓斐. DA-SwinUnet: 基于方向注意力模块与 SwinUnet 的股骨头坏死分割模型[J]. 软件工程与应用, 2026, 15(2): 168-180. DOI: 10.12677/sea.2026.152017

Abstract

Osteonecrosis of the femoral head (ONFH) is a common orthopedic condition, where early diagnosis and precise lesion localization are crucial for effective treatment. To address the issues of subjectivity in current image-based diagnosis and the failure of existing segmentation methods to fully leverage the anatomical consistency of medical images, this study proposes a DA-SwinUnet model enhanced with a directional attention mechanism. Built upon the Swin-Unet Transformer architecture, this model integrates a directional attention mechanism to enhance directional feature extraction and employs weighted skip connections to optimize network performance, reducing computational complexity while improving segmentation accuracy for lesions of varying sizes. The concept of a Directional Connectivity Vector is also incorporated to further enforce anatomical consistency. Experimental results on a clinical MRI dataset for ONFH demonstrate that the DA-SwinUnet model achieves a Dice coefficient of 87.3 and an IoU of 78.7, outperforming traditional CNN models and other Transformer-based segmentation methods. This study indicates that the DA-SwinUnet can achieve precise segmentation of ONFH lesions while maintaining anatomical consistency, significantly enhancing segmentation performance. It provides robust support for the early diagnosis and clinical treatment of the disease and holds promising potential for application.

Keywords

Osteonecrosis of the Femoral Head, Directional Attention Mechanism, Image Segmentation

Copyright © 2026 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

股骨头坏死(Osteonecrosis of the femoral head, ONFH)是一种因血供中断导致的进行性骨组织坏死, 常见于 30 至 50 岁人群。若不及时干预, 病情进展可导致股骨头塌陷, 引发严重的髋关节功能障碍, 晚期往往需要进行髋关节置换手术。流行病学研究表明, 我国每年新发病例超过 15 万, 且由于早期症状隐匿, 多数患者在确诊时已处于中晚期[1][2]。因此, 实现早期病变的检出与精确定位对于指定有效的临床干预策略起到了至关重要的作用。

目前, ONFH 的临床诊断主要依赖医生对 MRI 影像的主观判读, 这就导致诊断结果易受个人经验和诊断水平差异的影响[3]。尤其对于体积微小、边界模糊或呈现非典型方向性分布的病灶, 误诊风险显著增高。研究显示, 部分病例的误诊率可超过 20% [4], 这严重阻碍了辅助诊断系统的发展。

近年来, 深度学习技术在医学图像分析领域取得了显著进展[5]。基于卷积神经网络的图像分割模型, 如 U-Net 及其变体, 已广泛应用于器官和病灶的自动识别, 并在多个公开医学影像数据集上展现出优异性能[6]。亦有研究尝试将卷积神经网络应用于 ONFH 的自动分级, 例如 Wang 等人设计的分级模型, 验证了深度学习方法在 ONFH 识别中的可行性与优越性[7]。然而, 分类模型仅提供疾病严重程度的粗略估计, 无法实现精确的病灶定位与定量分析, 也难以满足个性化治疗方案制定的需求。

相比之下, 病灶分割能够精确勾勒病变区域, 是评估 ONFH 进展、实现自动化诊断和术前规划的关键[8]。然而, ONFH 病灶通常形态不规则、体积较小且具有明显的方向性分布特征。传统卷积神经网络架构受限于局部感受野, 难以有效捕捉病灶的方向性结构及其空间连通性, 导致分割边界模糊、连续性

欠佳。为应对上述挑战,方向连通性建模被引入医学图像分割领域。该方法通过建模像素间的方向关联,显著提升了模型对病灶区域空间结构的捕捉能力。DconnNet 提出的方向连通向量(Directional Connectivity Vector, DCV),通过构建 8 通道的方向连通性掩码,将方向信息显式地融入潜在特征空间,有效增强了模型识别微小结构并保持解剖一致性的能力[9]。

然而,现有的方向连通性方法通常应用于卷积神经网络架构,尚未拓展至基于 Transformer 的模型。Swin-Unet 作为一种层级式 Transformer 分割网络,凭借其强大的全局建模能力和多尺度上下文信息捕获能力,在医学图像分割任务中表现出优越性能[10]。

本研究提出了一种创新的 DA-SwinUnet 模型,首次将方向连通性概念融入 Swin-Unet 架构,构建了一个具有方向感知能力的 Transformer 分割网络。该模型引入了方向注意力机制,以增强方向性信息的提取,使网络能够更精准地捕捉 ONFH 病灶的方向性及空间结构特征。此外,针对医学数据集规模通常较小、Transformer 大参数模型易过拟合的问题,模型采用了加权跳跃连接策略[11],在避免由于通道拼接带来参数膨胀的同时,提升了模型在小样本医学数据集上的分割性能。

2. 相关工作

2.1. ONFH 图像分析中的深度学习研究

在股骨头坏死(osteonecrosis of the femoral head, ONFH)的影像分析领域,现有文献的核心任务主要围绕病变检出与分级评估展开。早期工作多采用二维卷积神经网络(2D CNN)实现简单的阳性/阴性判别或基于 Steinberg 分期的分级预测。例如,Wang 等提出多路径 CNN,并在小样本 MRI 数据上取得了超过 80% 的分级准确率[7];Shen 等将残差网络与迁移学习结合,用于评估 ONFH 的严重程度[12]。上述方法验证了深度学习用于 ONFH 诊断与评估的可行性,但其输出通常仅为整体标签,难以提供像素级病灶定位及体量化信息,因此在个性化治疗所需的精细评估方面仍存在不足。

相较于分级任务,针对 ONFH 病灶分割的研究仍较为有限。Ruckli 等提出从常规髋关节 MRI 中对坏死区域进行定量的深度学习方法,病灶分割 Dice 系数达到[13];Uemura 等基于三维 MRI 影像构建用于 ONFH 体积评估的深度学习模型,其病灶分割 Dice 系数达到 0.89,平均表面距离为 0.76 mm [14]。尽管这些研究为 ONFH 自动分割提供了有价值的探索,但在小病灶敏感性不足、结构信息建模能力有限等方面仍面临挑战。近年来,部分研究尝试将 Transformer 编码器引入 U-Net 框架,利用全局自注意力缓解病灶形态多样导致的特征匹配困难。然而,由于缺乏显式的空间连通性约束,分割结果在方向一致性方面仍可能出现不连续的问题[15]。总体而言,如何在保持全局上下文感知的同时兼顾方向连续性与结构一致性,仍是 ONFH 精细分割模型亟待解决的关键问题。

2.2. 通用医学图像分割网络

医学图像分割方法整体上可归纳为两类主流框架:基于卷积神经网络(CNN)的分割框架与基于 Transformer 的分割框架。

2.2.1. 卷积神经网络(CNN)

自 U-Net [16]提出以来,带有跳跃连接的编码器-解码器(encoder-decoder)结构已成为器官与病灶分割任务的经典结构。随后,SegNet [17]、引入空洞空间金字塔池化(ASPP)的 DeepLab [18]以及 HRNet [19]等代表性模型,分别通过池化索引引导的上采样、空洞卷积的多尺度上下文聚合与多分辨率并行特征学习等策略,进一步增强了多尺度特征表征与细节恢复能力。然而,卷积算子的局部感受野使其难以共同捕获大体积与小体积且方向性明显的病灶[20],易出现边界断裂与拓扑错误。

2.2.2. Transform

TransUnet [21]是将 ViT (Vision Transformer) [22]融合进 U-Net 框架的模型; UNETR [23]与 Swin UNETR [15]则采用分层式窗口注意力机制,使 2D 与 3D 分割任务能够受益于全局上下文建模能力。然而,这些基于 Transformer 的网络多是在通道或 patch 级别隐式编码像素间关系[24],缺乏对方向连通性与显式拓扑结构的建模。因此,在需要解剖结构连续性与拓扑一致性的精细分割场景(例如 ONFH 中小范围坏死区域)中,分割结果仍可能出现断裂或假阳性扩散等现象。

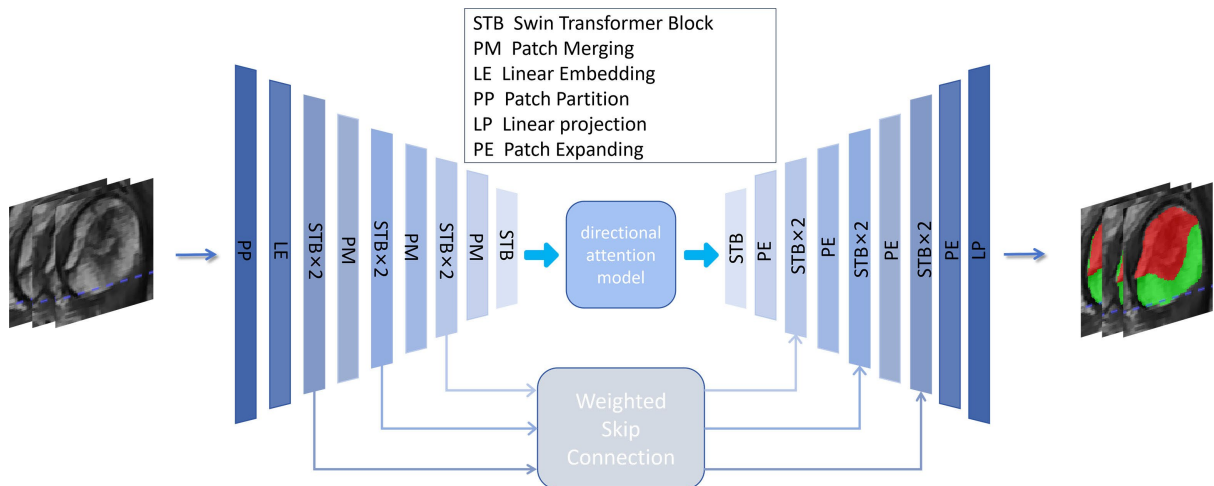


Figure 1. Overall architecture of the DA-SwinUnet model

图 1. DA-SwinUnet 模型的整体架构

3. 方法论

3.1. 任务定义与模型总览

本研究旨在实现 MRI 图像中股骨头坏死(osteonecrosis of the femoral head, ONFH)病灶的自动分割。该任务以单张二维 MRI 切片作为输入,输出与输入尺寸一致的多类别分割掩膜。不同于一般语义分割任务,ONFH 病灶分割除需达到像素级分割精度外,还必须尽可能保持病灶区域的空间连通性与解剖一致性。这是因为 ONFH 病灶的空间分布与股骨头内部骨小梁结构相关,病灶形态具有一定方向性与连续性,对分割结果的结构一致性有一定的要求。因此,分割结果的结构连续性对后续定量评估与临床决策具有重要意义。

为应对上述挑战,本文提出一种新型 DA-SwinUnet 网络结构,通过引入方向注意力机制与加权跳跃连接,提升对小尺度病灶及边界模糊区域的建模能力。如图 1 所示,整体框架由编码器、方向注意力模块(Directional Attention Module, DAM)以及解码器三部分组成。

网络主干采用对称的 U 形结构。编码器基于 Swin Transformer 提取多尺度特征表示,在兼顾局部细节的同时捕获全局上下文信息。具体而言,输入图像首先被划分为若干不重叠的 patch,并经线性映射嵌入为高维特征向量;随后通过多层 Swin Transformer 模块进行特征编码,利用窗口多头自注意力(window-based multi-head self-attention, W-MSA)与移位窗口注意力(shifted window multi-head self-attention, SW-MSA)实现局部建模与跨窗口信息交互。在各阶段之间,patch merging 模块完成空间下采样与通道扩展,从而形成分层抽象的特征表达。解码器结构与编码器结构对称,通过 patch expanding 模块逐级恢复空间分辨率,并融合跨层特征以细化病灶边界,最终将输出映射回原始图像尺度生成分割掩膜。

在编码器与解码器之间的瓶颈位置,本文引入方向注意力模块(DAM),以显式建模像素间的方向连

通关系。该模块沿八个主方向(水平、垂直及两组对角方向)构建方向掩膜,引导网络学习与解剖连续性一致的结构特征。进一步地,方向引导的特征图在监督学习约束下提取全局方向权重,并将其用于通道级特征融合,从而增强模型对小病灶方向连通性的捕获能力,提高边界感知与结构完整性。

此外,DA-SwinUnet 在跨层特征融合过程中引入加权跳跃连接。不同于传统的直接拼接方式,加权跳跃连接为不同层级特征分配自适应权重,使网络能够突出更具判别力的表示并抑制冗余信息。该设计一方面有助于降低不必要的融合开销,另一方面可提升多尺度病灶的分割鲁棒性,并在小样本医学影像场景下缓解 Transformer 架构潜在的过拟合风险。

3.2. 编码器

编码器采用 Swin Transformer 从 MRI 图像中提取多尺度特征。Swin Transformer 基于 ViT (Vision Transformer)架构构建,并引入窗口多头自注意力机制(window-based multi-head self-attention, W-MSA)及移位窗口策略(shifted window multi-head self-attention, SW-MSA)。与 ViT 的全局自注意力不同,W-MSA 将输入特征划分为若干互不重叠的局部窗口,并仅在窗口内计算自注意力,从而显著降低计算复杂度。为弥补窗口间交互不足的问题,SW-MSA 在相邻层之间对窗口划分进行平移,使得不同窗口区域能够在层间发生信息交换,在保持计算效率的同时获取更丰富的跨区域上下文信息。典型 Swin Transformer 块的结构如图 2 所示。

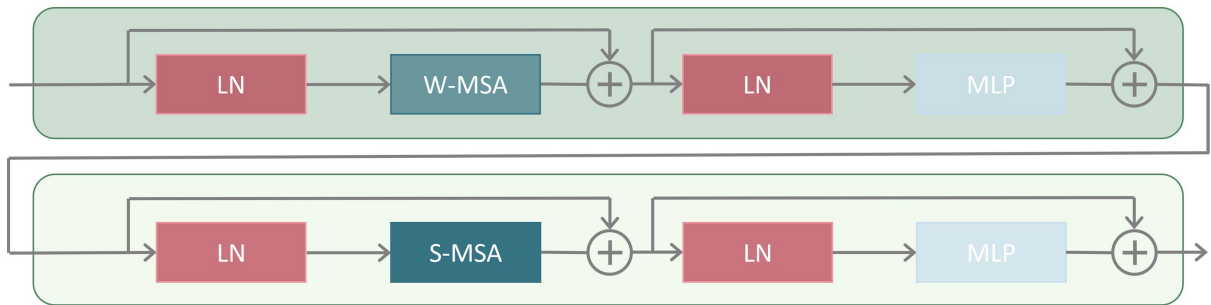


Figure 2. Structure of the Swin Transformer block. It consists of two sub-layers, W-MSA and SW-MSA, and includes LN, MLP, and residual connections to achieve efficient feature modeling

图 2. Swin Transformer 块的结构。它由两个子层 W-MSA 和 SW-MSA 组成,并包含 LN、MLP 和残差连接,以实现高效的特征建模

3.3. 方向注意力模块

Swin Transformer 基于自注意力机制,能够在图像内部同时建模局部与全局的信息交互,使得所提取的特征图不仅包含丰富的语义表征,还保留了较为细致的空间结构信息,并在特征表达中隐式编码了像素之间的相对位置关系。从几何意义上看,任意两像素之间的相对位置可表示为一个方向向量,因此整幅特征图中蕴含了大量可用于刻画结构走向与连通性的方向性特征。

然而,这类方向性信息往往被复杂的高维特征表示所“淹没”,难以被网络直接显式利用。为充分挖掘并强化方向性特征对分割的贡献,本文提出将传统二值掩膜(M_{gt})转换为方向掩膜(D_{gt}),并以其作为监督信号。通过该转换,模型在学习前景类别概率的同时,还能够学习像素间的空间连通性与方向关联,从而在分割过程中更好地保持局部结构完整性与整体一致性。

3.3.1. 方向掩膜标签的构建

为使模型能够准确感知病灶区域的方向性结构,本文提出将传统二值掩膜(M_{gt})转换为方向掩膜标签(D_{gt}),并以此作为监督信号。如图 3 所示,原本单通道的标签图被扩展为八通道方向掩膜,其中每个通

道对应一种特定方向(上、下、左、右、左上、左下、右上、右下), 用于显式表征像素在相应方向上的连通关系与结构一致性。

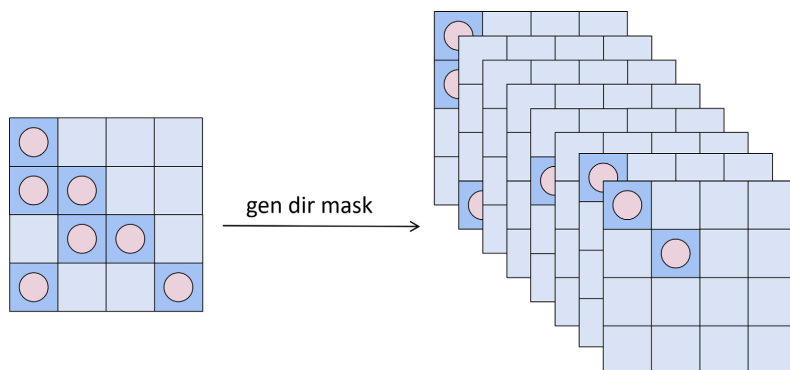


Figure 3. Schematic diagram of orientation mask generation. The left side shows the input binary mask, and the right side shows the generated eight-channel orientation mask, with each channel representing information from a different direction
图 3. 方向掩膜生成的示意图。左侧显示输入的二进制掩膜, 右侧显示生成的八通道方向掩膜, 每个通道表示不同方向的信息

例如, 在“上”方向通道中, 某一位置取值为 1 表示: 在二值掩膜中, 该位置像素及其正上方相邻像素的标签均为 1, 即两者在该方向上存在连通关系。方向掩膜标签 D_{gt} 的构建过程如下:

$$D_{gt}^{(0)}(x, y) = \begin{cases} 1, & \text{if } M_{gt}(x, y) = M_{gt}(x, y-1) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

这里, $D_{gt}(i, j)$ 表示位置 (i, j) 处的方向标签, $M_{gt}(i, j)$ 表示位置 (i, j) 处的二值掩膜标签。该公式通过在对应方向通道中将满足连通条件的像素对标记为 1, 从而构建方向掩膜, 实现对像素间方向连通关系的显式编码。

3.3.2. 方向特征引导与方向加权机制

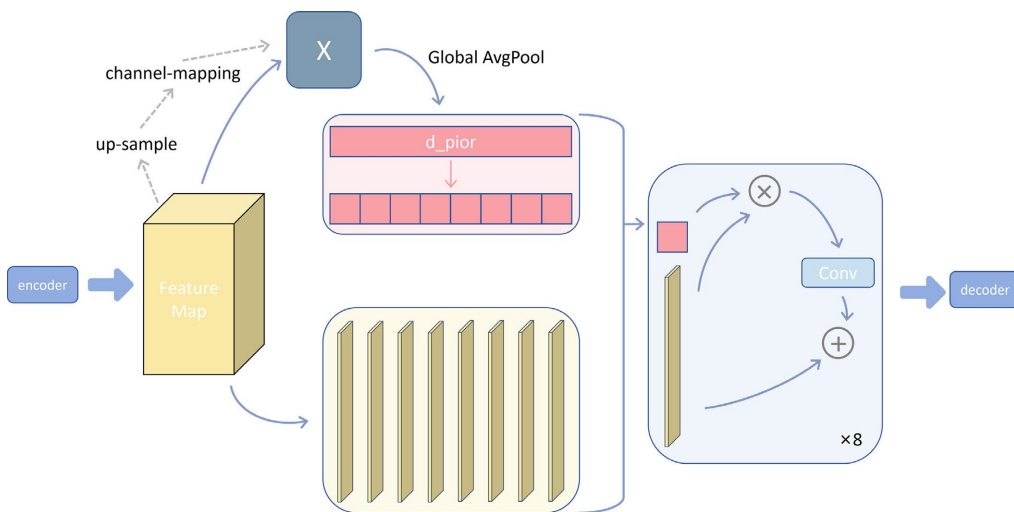


Figure 4. Schematic diagram of directional attention module (DAM)

图 4. 方向注意力模块(DAM)的示意图

为引导网络准确感知方向性特征, 本文基于编码器提取的特征图 F 构建了方向引导特征图 F_{dgr} 。具体而言, 首先对特征图 F 进行通道映射(channel mapping), 然后将其上采样至与方向掩膜相同的空间分辨

率, 从而确保其尺寸与原始标签图一致。得到的方向引导特征图 F_{dgr} 不仅作为辅助监督信号来增强方向性学习, 还用于后续方向权重的提取与建模。

随后, 我们对方向引导特征图 F_{dgr} 应用全局平均池化(global average pooling), 得到一个八维的方向权重向量, 该向量表示图像在各个方向上的整体注意力。这一方向权重向量将在后续模块中用于引导主特征图的加权融合, 从而增强不同方向上的结构信息表示。

3.3.3. 方向特征图加权融合

在获得方向权重后, 我们将编码器输出的特征图沿通道维度划分为八个子组, 每个子组表示对应特定方向的特征子集。然后, 将方向权重与各子特征图相乘, 从而增强对应方向的响应并抑制其他方向的响应。这些加权特征图随后通过卷积层进行尺度调整, 并添加回原始特征图。

最终, 所有八个加权特征图被拼接在一起, 生成一个统一的增强特征图, 该图整合了显式的方向性结构信息。作为方向注意力模块的输出, 这一增强特征图显著提升了模型捕捉结构信息的能力, 特别是在边缘检测、纹理识别以及局部形状变化等任务中。

整体过程如图 4 所示, 其中编码器生成特征图, 随后这些特征图被上采样并处理以生成方向特征图, 这些特征图再经过加权和融合, 最终形成输出。

3.4. 解码器

解码器的设计基于传统的 U-Net 架构, 简洁而有效。它通过上采样和跳跃连接恢复图像的空间分辨率, 同时将编码器提取的高分辨率特征与当前特征图进行融合。这一过程有助于保留编码器提取的详细信息。

3.5. 加权跳跃连接

在分割任务中, 跳跃连接对于将低分辨率特征从编码器传递到解码器至关重要, 从而恢复图像的细节信息[25]。然而, 传统的跳跃连接(如拼接或加和)并未考虑不同层特征的重要性差异, 这可能导致冗余或丢失重要特征。特别是在我们的任务中, 病灶仅占整个目标区域的一小部分, 这一问题尤为明显。为了解决这个问题, 我们引入了一种加权跳跃连接机制。该机制使用一个可学习的权重 α 来融合编码器和解码器的特征, 融合特征计算公式为:

$$F_{\text{fusion}} = \alpha F_{\text{encoder}} + (1 - \alpha) F_{\text{decoder}} \quad (2)$$

其中, α 是由网络学习得到的, 表示编码器和解码器特征在融合中的贡献。为了确保 α 保持在[0, 1]范围内, 我们使用 Sigmoid 激活函数:

$$\alpha = \text{sigmoid}(W_{\alpha} \cdot F_{\text{encoder}} + b_{\alpha}) \quad (3)$$

这一设计允许网络根据任务的需求自适应地调整特征的贡献。对于小病灶, 网络增加低层特征的权重, 以更好地捕捉细节[26]。相反, 对于较大的目标区域(如股骨头), 网络减少对低层特征的依赖, 更侧重于高层特征, 从而增强对整体形状和结构的捕捉。

与传统的拼接操作不同, 后者可能导致内存溢出和过拟合问题, 加权跳跃连接引入了可学习的参数来融合特征, 避免了不必要的通道扩展, 从而显著减少了模型在融合阶段的参数量。具体而言, 在传统的 U-Net 架构中, 通道维度的直接拼接会导致特征图通道数翻倍(例如由 C 膨胀至 $2C$)。为了进行特征对齐并输入至下一层, 通常必须紧跟额外的卷积层(如 1×1 或 3×3 卷积)进行通道降维, 这会引入高达 $O(C^2)$ 数量级的庞大参数开销。相比之下, 本文提出的加权跳跃连接通过逐元素加权求和(Element-wise Addition)的方式进行融合, 使得融合后的特征图通道数始终维持在 C 。此外, 计算权重 α 所需的线性映

射参数仅为 $O(C)$ 级别。因此,该机制从根本上避免了拼接操作带来的通道膨胀以及随后的降维卷积开销,在降低计算复杂度与内存消耗的同时,实现了多尺度特征的高效聚合。

3.6. 损失函数

设计了两类损失函数:连通性损失和像素级分类损失。连通性损失用于确保分割结果的空间一致性[27],而像素级分类损失则主要用于优化模型的类别预测准确性[28]。

3.6.1. 连通性损失(Connectivity Loss)

连通性损失的目标是优化图像中像素的空间连通性,确保模型输出的分割结果没有出现不连续或断裂的区域。具体而言,我们通过计算模型生成的方向性连通性图与扩展后的真实标签方向性连通性图之间的差异来评估模型的空间一致性。

在实现中,我们首先使用模型输出的特征图 c_{map} 计算每个像素在多个方向上的连通性信息。然后,通过比较预测的方向性连通性图与真实标签的方向性连通性图来计算连通性损失。此时, c_{map} 包含每个像素在不同方向(如上、下、左、右以及四个对角方向)上的连通性概率。连通性损失计算公式为:

$$L_{\text{conn}} = \text{BCELoss}(\sigma(c_{\text{map}}), c_{\text{true}}) \quad (4)$$

其中, c_{map} 是模型输出的连通性图,经过 Sigmoid 激活函数处理后,值域在 $[0, 1]$ 之间,表示每个像素在不同方向上的连通性; c_{true} 是真实标签的方向性连通性图,表示经过扩展后的真实标签连通性图。

3.6.2. 投票机制

为了生成最终的分割图,我们结合模型输出的连通性图使用投票机制。具体而言,我们计算每个像素在多个方向上的投票,综合来自所有方向的连通性信息,以生成预测的分割图。

投票过程包括计算每个像素与其在上、下、左、右及对角方向上的邻近像素之间的连通性,为每个方向生成一个投票。这些投票通过平移操作进行计算,评估每个像素在不同方向上基于邻近像素位置的连通性。所有方向的结果随后被加权并组合,以做出每个像素的最终连通性决策。最终的预测分割图通过选择每个像素在所有方向上的最大投票来生成。

3.6.3. 像素级分类损失

像素级分类损失结合了交叉熵损失(Cross-Entropy Loss)和 Dice 损失(Dice Loss),以优化模型的分割性能。交叉熵损失侧重于通过比较预测标签和真实标签来提高像素分类的准确性。Dice 损失则特别适用于小物体或不平衡物体的分割任务,因为它衡量预测区域和真实区域之间的重叠度,鼓励更好的边界对齐。两者的结合不仅提高了像素级的预测精度,还提升了整体分割质量。

4. 实验设计与结果分析

4.1. 数据集

本研究使用的数据集来自青岛大学附属医院,包括 43 例股骨头坏死(ONFH)患者的 MRI 图像。该数据集涵盖了年龄在 24 至 60 岁之间的不同患者,涉及 ONFH 的不同阶段,如图 5 所示。

所有患者均经临床确诊为 ONFH,但病灶分布于股骨头不同区域,这保证了数据集的多样性与较强的代表性。MRI 图像采用 T1 或 T2 加权冠状位序列进行采集。为确保扫描流程和参数设置的一致性,所有影像数据均由具有丰富经验的放射科医师完成采集。

病灶区域由两名高级职称放射科医师使用 ITK-SNAP 工具进行人工标注,并由第三名放射科医师对每例标注结果进行复核,以保证标注的一致性与高质量。

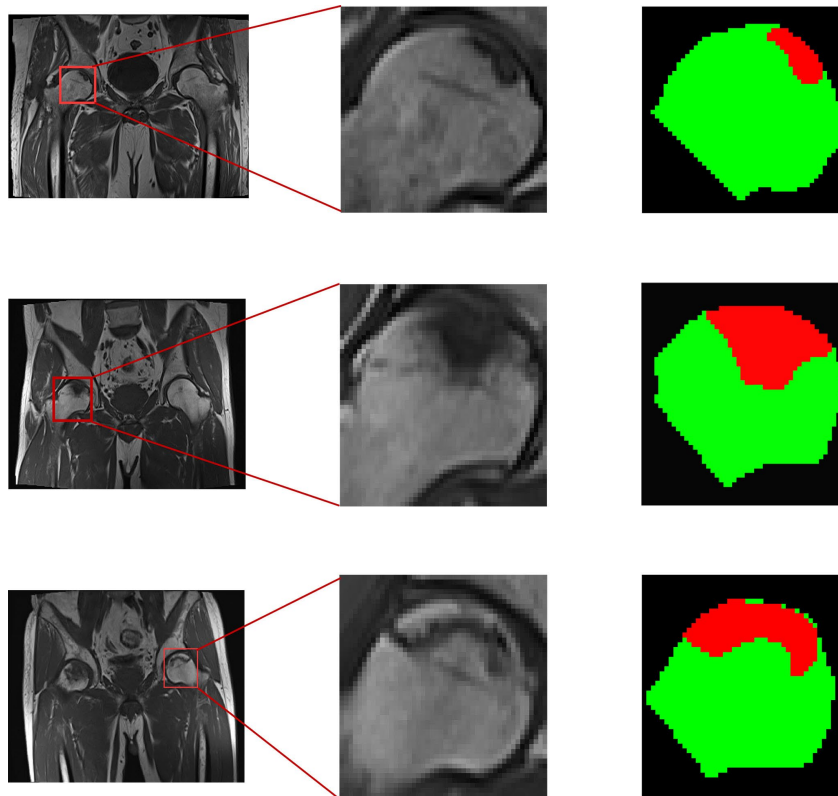


Figure 5. MRI images of femoral head necrosis (ONFH) of different severities and their corresponding ground truth labels. The left column shows the original MRI images, the middle column shows the magnified areas, and the right column shows the manually labeled results, where green represents normal femoral head and red highlights the necrotic areas

图 5. 不同严重程度股骨头坏死(ONFH) MRI 图像及对应的真实标签。左列显示原始 MRI 图像, 中间列显示放大的区域, 右列展示人工标注结果, 其中绿色表示正常股骨头, 红色突出显示坏死区域

模型在正式训练前首先需要对数据进行了预处理, 包括去噪及数据增强等操作, 这样不仅能提升图像质量, 提高模型的泛化能力, 还可以增强模型在小病灶分割场景下的鲁棒性数据集的多样性以及高质量的人工标注, 使其在 ONFH 自动分割与辅助诊断研究中具有较高的应用价值和临床意义。

4.2. 实验细节

在本研究中, 我们基于 PyTorch 通过矩阵平移操作为每幅图像生成对应的方向连通性掩膜。编码器采用在 ImageNet 上预训练的 Swin Transformer, 用于初始化模型的主干网络。不同实验配置下的超参数略有差异, 为保证对比的合理性, 本实验中所有输入图像的尺寸统一调整为 224×224 像素。

训练过程中采用随机梯度下降(Stochastic Gradient Descent, SGD)优化器, 初始学习率设为 0.01, 动量为 0.9, 权重衰减系数为 0.0001。学习率随迭代次数增加按照多项式衰减策略进行更新。批量大小(batch size)设为 8, 最大训练轮数 epoch 设置为 100。

损失函数由交叉熵损失(Cross-Entropy Loss, CE Loss)、Dice 损失以及连通性损失三部分组成。其中, 连通性损失通过构建水平和垂直方向的平移矩阵, 对模型预测的八方向连通性图进行对齐与监督。总体损失采用多分支监督形式, 将各项损失进行加权融合以共同优化模型。

整个框架基于 PyTorch 2.5.0 和 Python 3.10 实现。训练过程中记录损失值、学习率变化及预测结果, 并进行可视化分析。所有实验均在 NVIDIA GeForce RTX 3060 Ti GPU 上完成, 同时固定 Python、NumPy 及 PyTorch 的随机种子, 以保证实验结果的可复现性。

4.2.1. 对比试验

为全面评估所提出的 DA-SwinUnet 模型性能, 本文在自建的股骨头坏死(ONFH) MRI 数据集上, 与多种主流语义分割模型进行了系统对比实验。对比模型包括 FPN、DeepLabv3+、U-Net、U-Net++、PAN [29]、TransUnet 以及 Swin-Unet。所有模型均在相同的实验设置下进行训练与测试, 以保证结果的公平性与可比性。

评估指标包括交并比(Intersection over Union, IoU)、Dice 系数、准确率(Accuracy, Acc)、灵敏度(Sensitivity, SEN)、特异度(Specificity, SPE)以及阳性预测值(Positive Predictive Value, PPV)。具体实验结果详见表 1。

Table 1. Performance comparison of different segmentation models, with the highest value bolded

表 1. 不同分割模型的性能比较, 最高值已加粗

	IoU (%)	DICE (%)	Acc (%)	SEN (%)	SPE (%)	PPV (%)
fpn	75.6	85.5	92.8	84.4	94.2	86.7
deeplab	75.2	85.2	93.1	83.1	94.1	88.3
unet	75.8	85.7	93.1	85.1	94.4	86.3
Unet++	76.0	85.8	93.2	84.6	94.3	87.3
pan	75.3	85.2	92.9	83.5	94.1	87.7
TransUnet	76.3	86.1	93.2	85.2	94.5	87.1
Swin-Unet	76.8	86.4	93.1	85.8	94.4	87.2
DA-SwinUnet	78.7	87.3	93.3	89.1	95.9	89.2

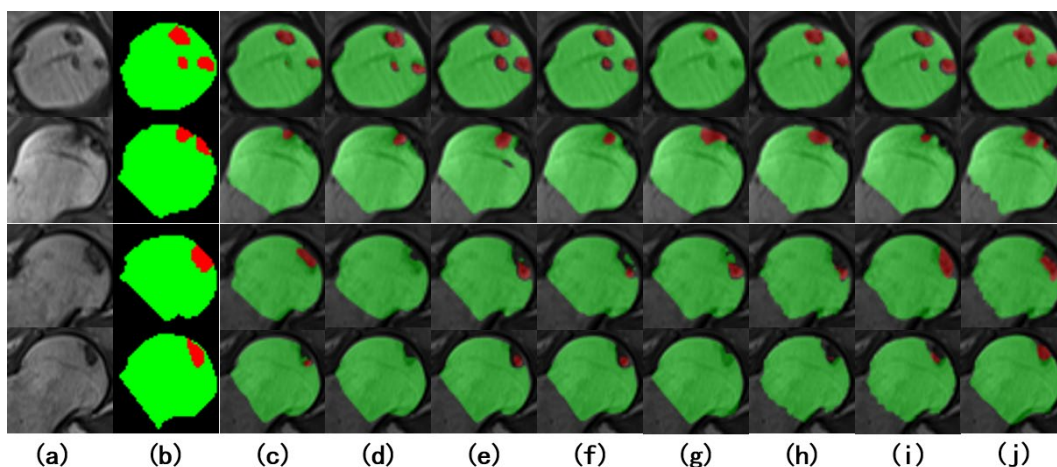


Figure 6. Visual comparison of segmentation results of different models in femoral head necrosis. (a) Input image; (b) Ground truth label; (c) U-Net; (d) U-Net++; (e) FPN; (f) PAN; (g) DeepLabV3; (h) TransUnet; (i) SwinUnet; (j) DA-SwinUnet

图 6. 不同模型在股骨头坏死分割结果的可视化对比。(a) 输入图像; (b) 真实标签; (c) U-Net; (d) U-Net++; (e) FPN; (f) PAN; (g) DeepLabV3; (h) TransUnet; (i) SwinUnet; (j) DA-SwinUnet

实验结果表明, 各模型之间的性能差异显著。基于卷积神经网络(CNN)的方法, 如 U-Net 和 DeepLabv3+, 表现相似(IoU 范围为 75.2%到 76.0%), 这表明传统的 CNN 架构在该任务上的性能已经接近饱和。相比之下, 结合 Transformer 架构的模型, 如 Swin-Unet, 展现了更强的竞争力, 达到了 76.84%的 IoU 和 86.48%的 Dice 系数, 显著优于所有 CNN 模型。这突显了基于 Transformer 的模型在捕捉长程

依赖和全局上下文信息方面的优势。

为了更直观地展示各模型的分割效果，图 6 给出了不同模型在股骨头坏死测试集上的可视化对比结果。值得注意的是，DA-SwinUnet 模型在所有对比模型中表现最佳，IoU 和 Dice 分别达到了 78.70%和 87.30%，创下了新的性能基准。与强基线模型 Swin-Unet 相比，DA-SwinUnet 在 IoU 上取得了近 1.86%的显著提升。从图 6 的可视化结果可以看出，这一显著的性能提升归因于 DA-SwinUnet 模型中创新的方向注意力模块(DAM)和加权跳跃连接(WSC)模块的协同作用。DAM 显式地建模了目标的结构化方向性信息，显著提升了边界分割精度，而 WSC 模块自适应地融合多尺度特征，优化了不同层级信息的利用。这些实验结果充分验证了 DA-SwinUnet 模型的有效性与先进性，为复杂医学图像分割任务提供了一种稳健的解决方案。

4.2.2. 消融实验

为了验证所提出的方向注意力模块(DAM)和加权跳跃连接(WSC)的有效性，本文基于 Swin-Unet 模型进行了消融实验。通过逐步引入 WSC 和 DAM，定量评估了每个模块对模型性能的贡献。所有实验均在相同设置下进行，实验结果详见表 2。

Table 2. Ablation experiments were conducted on the proposed weighted skip connections and directional attention modules
表 2. 对所提的加权跳跃连接和方向注意力模块进行了消融实验

	IoU (%)	DICE (%)	Acc (%)	SEN (%)	SPE (%)	PPV (%)
Baseline	76.8	86.1	93.1	85.8	94.4	87.2
Baseline + WSC	77.7	86.6	95.1	88.7	95.9	84.7
Baseline + DAM	77.2	86.3	94.9	87.4	95.9	85.2
Baseline + WSC + DAM	78.7	87.3	93.3	89.1	95.9	85.7

首先，当仅引入 WSC 模块时，IoU 从 76.84%增加到 77.7%，提高了 0.86%，而 Dice 系数从 86.1%提升到 86.6%。这表明 WSC 通过自适应地融合来自编码器的低层细节特征和来自解码器的高层语义特征，增强了模型在复杂场景下，尤其是既有小病灶又有大目标的情况下的分割能力。

接下来，引入 DAM 模块使 IoU 提升了 0.36% (从 76.84%提升至 77.20%)，Dice 系数提高了 0.3%(从 86.1%提升至 86.3%)。这表明 DAM 通过显式建模目标的方向性信息，有效提高了边界分割的精度，并增强了模型捕捉几何结构的能力。

当同时引入 WSC 和 DAM 模块时，DA-Swin-Unet 模型达到了 78.70%的 IoU 和 87.30%的 Dice 系数，分别较 Swin-Unet 基线模型提高了 1.86%和 1.2%。这一提升显著超过了每个模块单独增益的总和，表明 DAM 和 WSC 之间具有强大的协同效应。具体而言，DAM 增强了对边界区域的方向感知，而 WSC 优化了多尺度特征的融合，进一步提升了整体分割性能。

5. 结论与讨论

在本研究中，我们提出了基于 Swin-Unet 的 DA-SwinUnet 模型，结合方向注意力模块(DAM)和加权跳跃连接(WSC)来解决股骨头坏死(ONFH) MRI 图像的分割问题。实验结果表明，DA-SwinUnet 在各类传统 CNN 和 Transformer 基础模型中表现优异，相较于强基线模型 Swin-Unet，IoU 提升 1.86%，Dice 提升了 1.2%。消融实验进一步验证了 WSC 和 DAM 的有效性，每个模块都显著贡献于性能的提升。WSC 模块改善了编码器和解码器之间多尺度特征的融合，能够有效处理既有小病灶又有大病灶的复杂场景。DAM 模块通过显式建模方向性信息，提升了模型对边界精度和几何结构的捕捉能力。

尽管取得了良好的实验结果,本研究仍存在一些局限性,包括数据集规模相对较小、缺乏 3D 空间信息的考虑以及相较于传统 CNN 模型的较高计算成本。未来的工作将着重于将 DA-SwinUnet 扩展到 3D MRI 数据卷积,探索多模态数据融合,和研究半监督学习技术,以增强模型的泛化能力并减少对标注数据的依赖。总体而言,DA-SwinUnet 模型通过其创新模块,为 ONFH 病灶分割与临床诊断提供了强有力的工具,展示了其在医学图像分析中的潜力。

参考文献

- [1] Zhao, D., Yu, M., Hu, K., Wang, W., Yang, L., Wang, B., *et al.* (2015) Prevalence of Nontraumatic Osteonecrosis of the Femoral Head and Its Associated Risk Factors in the Chinese Population: Results from a Nationally Representative Survey. *Chinese Medical Journal*, **128**, 2843-2850. <https://doi.org/10.4103/0366-6999.168017>
- [2] Mont, M.A. and Hungerford, D.S. (1995) Non-Traumatic Avascular Necrosis of the Femoral Head. *The Journal of Bone & Joint Surgery*, **77**, 459-474. <https://doi.org/10.2106/00004623-199503000-00018>
- [3] Alkhatatbeh, T., Alkhatatbeh, A., Guo, Q., Chen, J., Song, J., Qin, X., *et al.* (2025) Interpretable Machine Learning and Radiomics in Hip MRI Diagnostics: Comparing ONFH and OA Predictions to Experts. *Frontiers in Immunology*, **16**, Article ID: 1532248. <https://doi.org/10.3389/fimmu.2025.1532248>
- [4] Zhang, Y., Cao, X., Li, X., Chen, J., Zhao, Y., Tian, Z., *et al.* (2018) Accuracy of MRI Diagnosis of Early Osteonecrosis of the Femoral Head: A Meta-Analysis and Systematic Review. *Journal of Orthopaedic Surgery and Research*, **13**, Article No. 167. <https://doi.org/10.1186/s13018-018-0836-8>
- [5] Li, M., Jiang, Y., Zhang, Y. and Zhu, H. (2023) Medical Image Analysis Using Deep Learning Algorithms. *Frontiers in Public Health*, **11**, Article ID: 1273253. <https://doi.org/10.3389/fpubh.2023.1273253>
- [6] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., *et al.* (2017) A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, **42**, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- [7] Wang, P., Liu, X., Xu, J., Li, T., Sun, W., Li, Z., *et al.* (2021) Deep Learning for Diagnosing Osteonecrosis of the Femoral Head Based on Magnetic Resonance Imaging. *Computer Methods and Programs in Biomedicine*, **208**, Article ID: 106229. <https://doi.org/10.1016/j.cmpb.2021.106229>
- [8] Gao, S., Zhu, H., Wen, M., He, W., Wu, Y., Li, Z., *et al.* (2024) Prediction of Femoral Head Collapse in Osteonecrosis Using Deep Learning Segmentation and Radiomics Texture Analysis of MRI. *BMC Medical Informatics and Decision Making*, **24**, Article No. 320. <https://doi.org/10.1186/s12911-024-02722-w>
- [9] Yang, Z. and Farsiu, S. (2023) Directional Connectivity-Based Segmentation of Medical Images. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 18-22 June 2023, 11525-11535. <https://doi.org/10.1109/cvpr52729.2023.01109>
- [10] 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. In: Karlinsky, L., *et al.*, Eds., *Computer Vision—ECCV 2022 Workshops*, Springer, 205-218. https://doi.org/10.1007/978-3-031-25066-8_9
- [11] Shen, Y., Guo, P., Wu, J., Huang, Q., Le, N., Zhou, J., *et al.* (2023) MoViT: Memorizing Vision Transformers for Medical Image Analysis. In: Cao, X.H., *et al.*, Eds., *Machine Learning in Medical Imaging*, Springer, 205-213. https://doi.org/10.1007/978-3-031-45676-3_21
- [12] Shen, X., He, Z., Shi, Y., Liu, T., Yang, Y., Luo, J., *et al.* (2024) Development and Validation of an Automated Classification System for Osteonecrosis of the Femoral Head Using Deep Learning Approach: A Multicenter Study. *The Journal of Arthroplasty*, **39**, 379-386.e2. <https://doi.org/10.1016/j.arth.2023.08.018>
- [13] Ruckli, A.C., Nanavati, A.K., Meier, M.K., Lerch, T.D., Steppacher, S.D., Vuilleumier, S., *et al.* (2023) A Deep Learning Method for Quantification of Femoral Head Necrosis Based on Routine Hip MRI for Improved Surgical Decision Making. *Journal of Personalized Medicine*, **13**, Article No. 153. <https://doi.org/10.3390/jpm13010153>
- [14] Uemura, K., Takashima, K., Otake, Y., Li, G., Mae, H., Okada, S., *et al.* (2025) Development of a Deep Learning Model for the Volumetric Assessment of Osteonecrosis of the Femoral Head on Three-Dimensional Magnetic Resonance Imaging. *The Journal of Arthroplasty*, **40**, S160-S166.e1. <https://doi.org/10.1016/j.arth.2025.05.126>
- [15] Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R. and Xu, D. (2022) Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: Crimi, A. and Bakas, S., Eds., *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer International Publishing, 272-284. https://doi.org/10.1007/978-3-031-08999-2_22
- [16] 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—MICCAI 2015*, Springer

- International Publishing, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [17] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. <https://doi.org/10.1109/tpami.2016.2644615>
- [18] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2018) Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected Crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/tpami.2017.2699184>
- [19] Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., *et al.* (2019) High-Resolution Representations for Labeling Pixels and Regions.
- [20] Ma, X., Chen, M., Zhang, J., Song, L., Du, F. and Yu, Z. (2025) PARF-Net: Integrating Pixel-Wise Adaptive Receptive Fields into Hybrid Transformer-CNN Network for Medical Image Segmentation.
- [21] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., *et al.* (2021) Transunet: Transformers Make Strong Encoders for Medical Image Segmentation.
- [22] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., *et al.* (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.
- [23] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R. and Xu, D. (2021) Unetr: Transformers for 3D Medical Image Segmentation. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 5-9 January 2021, 574-584.
- [24] Ou, Y., Yuan, Y., Huang, X., Wong, S.T.C., Volpi, J., Wang, J.Z., *et al.* (2022) Patcher: Patch Transformers with Mixture of Experts for Precise Medical Image Segmentation. In: Wang, L.W., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022*, Springer, 475-484. https://doi.org/10.1007/978-3-031-16443-9_46
- [25] Li, Z., Lyu, H. and Wang, J. (2023) FusionU-Net: U-Net with Enhanced Skip Connection for Pathology Image Segmentation. *Asian Conference on Machine Learning*, PMLR, İstanbul, 11-14 November 2023, 694-706.
- [26] Wang, H., Cao, P., Yang, J. and Zaiane, O. (2023) MCA-UNet: Multi-Scale Cross Co-Attentional U-Net for Automatic Medical Image Segmentation. *Health Information Science and Systems*, **11**, Article No. 10. <https://doi.org/10.1007/s13755-022-00209-4>
- [27] Jian, M., Wu, R., Xu, W., Zhi, H., Tao, C., Chen, H., *et al.* (2024) Vasuconnet: An Enhanced Connectivity Network for Vascular Segmentation. *Medical & Biological Engineering & Computing*, **62**, 3543-3554. <https://doi.org/10.1007/s11517-024-03150-8>
- [28] Azad, R., Heidary, M., Yilmaz, K., *et al.* (2023) Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook.
- [29] Lin, T.-Y., Doll'ar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2016) Feature Pyramid Networks for Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 2117-2125.