

# 基于 DH-Swin Unet 的医学图像分割算法<sup>①</sup>



王艺妮, 时宏伟

(四川大学 计算机学院, 成都 610065)

通信作者: 时宏伟, E-mail: shihw001@126.com

**摘要:** 骨关节疾病自古以来是人类最高发的疾病之一, 随着老龄化的不断加快, 这类疾病日趋广泛, 关节外科医师面临着巨大挑战. 对人体关节的图像分割方法研究可以帮助医生进行临床诊断和治疗, 然而, 由于存在噪声、模糊、对比度低等问题, 医学图像的特征提取比普通图像更具挑战性, 而且目前大多数分割模型在编码器和解码器之间都采用了普通的跳跃连接, 没有注重解决跳跃连接过程中的信息间隙和损失问题. 为解决这些问题, 提出一种基于 DH-Swin Unet 的医学图像分割算法, 该模型在 Swin-Unet 模型的基础上, 在跳跃连接中引入密集连接的 Swin Transformer 块, 并加入混合注意力机制, 来强化网络的特征信息传递. 通过在某三甲医院提供的真实临床数据对所提方法的性能进行评价, 结果表明, 所提出的方法取得了 *DSC* 为 86.79%、*HD* 为 32.05 mm 的分割结果, 在关节疾病的临床诊断中具有一定的实用价值.

**关键词:** U-Net; 跳跃连接; 医学图像分割; Swin Transformer; 注意力机制

引用格式: 王艺妮, 时宏伟. 基于 DH-Swin Unet 的医学图像分割算法. 计算机系统应用, 2024, 33(3): 206-212. <http://www.c-s-a.org.cn/1003-3254/9440.html>

## Medical Image Segmentation Method Based on DH-Swin Unet

WANG Yi-Ni, SHI Hong-Wei

(College of Computer Science, Sichuan University, Chengdu 610065, China)

**Abstract:** Bone and joint diseases have been one of the most prevalent diseases in human history. With the acceleration of aging, these diseases have become increasingly widespread, posing great challenges to orthopedic surgeons. Research on image segmentation for human joints can assist doctors in clinical diagnosis and treatment. However, due to the presence of noise, blurring, and low contrast, medical image feature extraction is more difficult than ordinary images. In addition, most segmentation models only use simple skip connections between the encoder and decoder, without addressing the issues of information gaps and losses during the skip connection process. To this end, this study proposes the DH-Swin Unet algorithm for medical image segmentation. On the basis of the Swin-Unet model, the densely connected Swin Transformer block and hybrid attention mechanism are introduced into skip connection to enhance the feature information transmission. The real clinical data provided by a hospital ranking top three are used to evaluate the performance of the proposed method. The results show that the *DSC* and *HD* of the model reach 86.79% and 32.05 mm respectively, and the model has certain practical value in the clinical diagnosis of joint diseases.

**Key words:** U-Net; skip connection; medical image segmentation; Swin Transformer; attention mechanism

骨关节疾病自古以来是人类最高发的疾病类别之一, 常见的骨关节疾病包括关节炎、颈椎炎、类风湿

关节炎等. 20 世纪以来, 由于人口老龄化的不断加快, 这类疾病日趋广泛, 给关节外科医师带来了巨大的挑

① 收稿时间: 2023-09-20; 修改时间: 2023-10-20; 采用时间: 2023-11-03; csa 在线出版时间: 2024-01-09

CNKI 网络首发时间: 2024-01-10

战. 而医学影像是疾病诊断、筛选、评估和确定治疗方案的重要工具, 图像分割技术在临床诊断和治疗中有着重要作用, 能够显著提高诊断效率和准确性. 早期的医学图像分割方法往往依赖于边缘检测、模板匹配技术、统计形状模型等. 尽管这些方法取得了一定的成效, 但特征提取算法的复杂性、局限性、稳定性等因素限制了图像分割技术的进一步发展<sup>[1]</sup>. 近年来, 在医学图像分割研究中, 深度学习的快速发展使得图像分割算法成为一种有效的工具, 它能够用来提取并分割医学图像中的骨关节结构, 帮助外科医生更有效地进行临床诊断和治疗.

卷积神经网络 (convolutional neural network, CNN) 是深度学习领域的重要成果, 自卷积神经网络在手写字体识别领域获得巨大突破后<sup>[2]</sup>, 逐渐开始应用在图像处理领域. CNN 在医疗影像分析中发挥着重要作用, 用于特征学习的 CNN 对图像噪声、模糊、对比度等不敏感, 因此在医学图像处理中提供了很好的分割结果. 它可以用于肿瘤检测、疾病诊断、器官分割等任务, 帮助医生更准确地进行诊断和治疗. Long 等人<sup>[3]</sup> 2015 年提出了全卷积网络 (fully convolutional network, FCN) 结构. FCN 可以对图像进行像素级的分类, 从而解决了语义级别的图像分割问题.

2015 年, Ronneberger 等人<sup>[4]</sup>提出了 U-Net 网络结构, U-Net 是一种基于 FCN 的语义分割网络, 适用于医学图像分割. 但是 U-Net 网络只能关注到局部性, 不能很好地建立显式的长距离依赖, 也没有完全捕获全局上下文关系. Transformer 结构开始用于视觉领域之后, 基于 U-Net 和 Transformer 结合的相关研究逐渐兴起. 由于 CNN 的平移不变性和捕捉长期依赖能力的不足, 限制了 U-Net 模型的性能, 而 Transformer 擅长捕捉序列之间的长期依赖关系. 于是, 研究者们尝试将 Transformer 结构融入到以 CNN 为主体的 U-Net 中, 提出了 TransUNet 网络<sup>[5]</sup>. TransUNet 是第 1 个将 Transformer 结构应用于医学图像分割的研究, 它大大地提高了 U-Net 的分割性能. 但是 TransUNet 仍然采用 CNN 和 Transformer 混合编码的结构, 解码器也是基于 CNN 的上采样, 这种混合编码的结构没有完全发挥出 Transformer 的优势. 此前, MSRA 提出了 Swin Transformer<sup>[6]</sup>作为新的视觉 Transformer backbone, 更适合用于医学图像分割, 基于 Swin Transformer 的 Swin-Unet 应运而生<sup>[7]</sup>.

Swin-Unet 采用纯 Transformer 结构的 U-shaped 网络, 大量实验表明, Swin-Unet 在多个器官和心脏数据集上具有良好的分割精度和鲁棒泛化能力.

注意力机制 (attention mechanism) 最初在机器翻译领域<sup>[8]</sup>取得成功, 随后被广泛应用于图像处理<sup>[9]</sup>、语音识别、自然语言处理等方面. 注意力机制可以从众多信息中选择出对当前任务目标更关键的信息, 同时抑制那些不重要的信息. 目前, 注意力机制广泛应用于在图像处理中, 在文献<sup>[10]</sup>中, Hu 等人提出了 SENet, 他们在通道维度上对特征层进行特征提取, 从而提高了通道特征信息的表达能力. Fu 等人<sup>[11]</sup>提出双重注意力网络, 他们通过自注意力机制自适应地融合局部语义特征, 更有效地捕获空间和通道维度上的全局依赖关系. 大量实验研究表明, 注意力机制能够促进图像分割模型更准确地关注关键信息, 提高分割结果的准确性和有效性, 提升模型的效率和可解释性.

目前图像分割也广泛应用于软骨、半月板等人体关节的临床研究, 基于卷积神经网络的图像分割方法可以通过学习关节区域的特征来实现准确的分割. 刘云鹏等人<sup>[12]</sup>提出一种基于图像块和全卷积神经网络相结合的自动分割算法, 用于对肩关节 MRI 图像进行分割. Liao 等人<sup>[13]</sup>提出了一种 CNN+Bi-RNN 方法来确定椎骨的位置, 改善了对特征信息的获取和处理, 提高了分割结果的准确性. Zeng 等人<sup>[14]</sup>提出了一种深度监督的 3D U-Net, 用于近端股骨的图像分割. Nie 等人<sup>[15]</sup>使用 3D FCN 对颅颌面进行分割. 宋平等<sup>[16]</sup>构建了 PointRend\_Unet 神经网络, 该方法用于对膝关节 CT 图像进行三维重建与分割, 有效提高了算法的准确性和分割速度, 提高了人工全膝关节置换术前规划的工作效率.

但是上述这些方法都没有关注编码器和解码器中间的连接通路, 也没有对信息接收提供有效的跳跃连接. 因此, 在先前工作的基础上, 我们基于 Swin-Unet 模型, 提出了一种新的图像分割算法——DH-Swin Unet, 我们在跳跃连接中引入密集连接的 Swin Transformer 块, 从而可以更好地使用不同尺度下的移位窗口方案, 增强网络的多尺度特征表示能力并改善长距离相关性. 另外我们在跳跃路径中引入混合注意力机制, 来增强特征描述. 我们构建了一个关节术后感染数据集, 其中包括髌骨数据和膝盖数据. 通过在我们的数据集上进

行多任务分割实验, 结果表明, 所提出的方法在不同数据集上取得了更好的分割结果.

### 1 研究方法

#### 1.1 网络架构

图 1 展示了我们提出的医学图像分割模型的结构. 我们基于 Swin-Unet, 提出了一种由编码器、解码器和

跳跃连接组成的 U 型对称网络模型——DH-Swin Unet, 跳跃连接采用由密集 Swin Transformer 块 (DSTB) 和混合注意力模块 (hybrid attention block) 组成的跳跃连接路径, 从而加强了编码器和解码器之间的信息传递有效性. 来自编码器的特征信息在进入解码器之前先经过密集连接的 Swin Transformer 块, 再经过混合注意力模块, 最后输出到解码层.

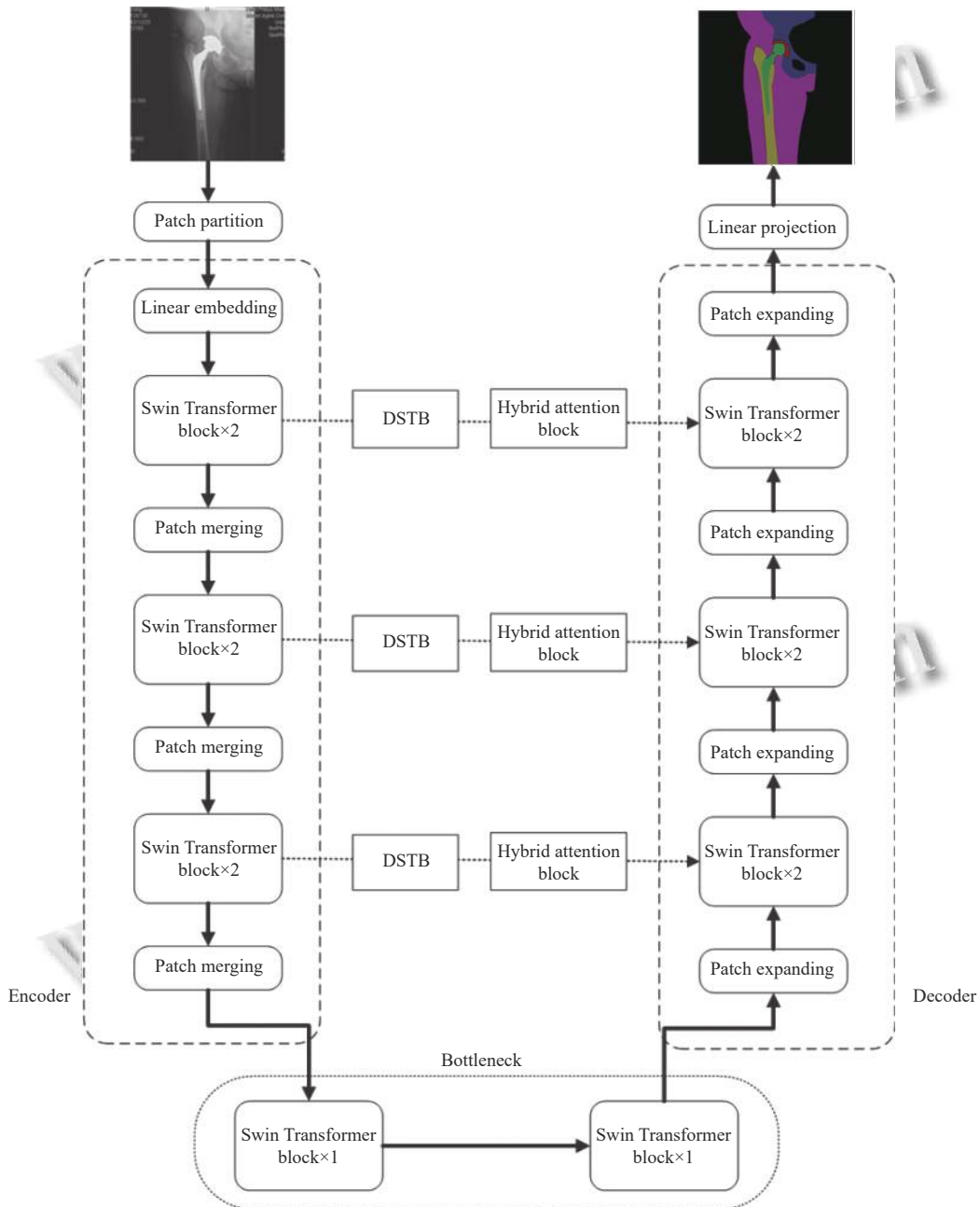


图 1 DH-Swin Unet 模型结构

#### 1.2 跳跃路径

U-Net 的大多数变体都使用了简单的跳跃连接, 但

是这种跳跃连接不利于长距离的依赖, 会带来语义差距大、信息损失、细粒度细节丢失等问题<sup>[17]</sup>. U-Net<sup>[4]</sup>



中,来自编码器的特征与解码器直接相连,这就要求它们具有相同的比例.然而,Zhou等人<sup>[18]</sup>提出即使这些特征图处于相同的比例,它们在语义上并不相似,且缺少理论支持这种融合是最优策略.因此,他们将注意力放在重新设置编码器和解码器之间的跳跃连接上,采用密集的卷积块替换了这些简单的连接,使得编码器特征的语义信息更丰富,并在合并之前使其语义级别更接近等待的解码器.这些密集跳跃连接也在文献<sup>[19]</sup>的Mask-RCNN分割中被证明了有效性.同样,密集U-Net++<sup>[20]</sup>使用半密集U-Net<sup>[21]</sup>,包含密集跳跃连接和跳跃路径,每个图层上的密集块使用之前所有图层的聚合要素,得到更优的结果.

基于以上研究的启发,我们采用由密集Swin Transformer块和混合注意力模块组成的跳跃连接路径,来增强模型的显式远程依赖,促进特征信息在编码器和解码器之间的有效传递.

### 1.3 密集Swin Transformer模块

受密集连接的启发,我们在跳跃连接路径中采用了密集连接Swin Transformer块,每个DSTB由多个Swin Transformer组成,具体结构如图2和图3所示.每个Transformer将特征信息传播给下一个Swin Transformer.Swin Transformer块是基于平移窗口(shifted windows)构造的,每个Swin Transformer块由LayerNorm(LN)层、多头自注意模块、剩余连接和具有GELU非线性的两层MLP组成.在两个连续的Transformer模块中分别采用了基于窗口的多头自注意(W-MSA)模块和位移的基于窗口的多头自注意(SW-MSA)模块<sup>[6]</sup>.这个机制表示为:

$$\hat{z}^l = W-MSA(LN(z^{l-1})) + z^{l-1} \quad (1)$$

$$z^l = MLP(LN(\hat{z}^l)) + \hat{z}^l \quad (2)$$

$$\hat{z}^{l+1} = SW-MSA(LN(z^l)) + z^l \quad (3)$$

$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \quad (4)$$

其中, $z^l$ 和 $\hat{z}^l$ 分别表示第 $l$ 块的SW-MSA模块和MLP的输出.每个窗口的自注意力公式为:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d}} + B\right)V \quad (5)$$

其中, $Q, K, V \in \mathbb{R}^{M^2 \times d}$ 分别表示查询、键和值矩阵, $M^2$ 表示窗口中的patch数量, $d$ 表示 $Q$ 或 $K$ 的维度, $B \in \mathbb{R}^{M^2 \times M^2}$ 为相对位置偏置.

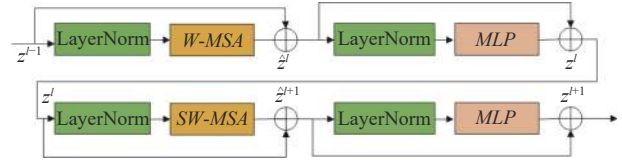


图2 Swin Transformer结构

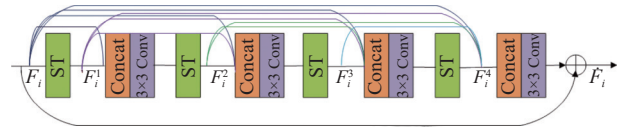


图3 Dense Swin Transformer block结构

图3中 $F_i^l$ 为每个Swin Transformer(ST)的输出,公式为:

$$F_i^l = P_{ST}(P_{FF}(F_i, F_i^1, \dots, F_i^{l-1})) \quad (6)$$

其中, $P_{ST}$ 和 $P_{FF}$ 分别表示Swin Transformer操作和特征融合操作.

### 1.4 混合注意力机制

在本文中,我们采用混合注意力机制,顺序捕获多尺度编码器特征的通道和空间依赖性来减小编码器和解码器特征之间的语义差距.

#### 1.4.1 通道注意力模块

通道注意力(CA)模块关注图像的关键特征,如图4所示,输入特征图的通道信息是一个 $H \times W \times C$ 的特征 $F_C$ , $F_C$ 首先经过平均池化和最大池化操作进行聚合.再通过多层感知机(MLP)对池化后的特征进行处理,结果分别应用于两个通道,再通过Sigmoid函数得到通道的注意力结果图.这个机制可表示为<sup>[22]</sup>:

$$\begin{aligned} M_{channel}(F_C) &= \sigma(MLP(Avg(F_C)) + MLP(Max(F_C))) \\ &= \sigma(W_1(W_0(F_C^{Avg})) + W_1(W_0(F_C^{Max}))) \end{aligned} \quad (7)$$

其中, $\sigma$ 表示Sigmoid激活函数, $W_0 \in \mathbb{R}^{\frac{C}{r} \times C}$ 和 $W_1 \in \mathbb{R}^{C \times \frac{C}{r}}$ 表示MLP的共享权重, $r$ 表示压缩比, $F_C^{Avg}$ 和 $F_C^{Max}$ 分别表示 $F_C$ 通过平均池化和最大池化后的特征.

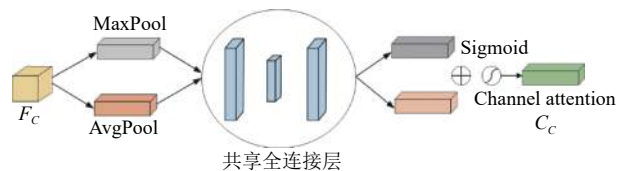


图4 通道注意力模块

#### 1.4.2 空间注意力模块

空间注意力(SA)模块的目的是找到特征图中的

关键位置关系. 如图 5 所示, 与通道注意力相似, 给定一个  $H \times W \times C$  的输入特征  $F'_C$ , 通过两个池化来处理输入特征, 结果拼接成一个特征图, 再经过卷积操作, 最后经过 Sigmoid 激活函数得到特征图, 这个机制可表示为<sup>[22]</sup>:

$$\begin{aligned} M_{\text{spatial}}(F'_C) &= \sigma(\text{conv}_{3 \times 3}([\text{AvgPool}(F'_C), \text{MaxPool}(F'_C)])) \\ &= \sigma(\text{conv}_{3 \times 3}([F'_C{}^{\text{Avg}}, F'_C{}^{\text{Max}}])) \end{aligned} \quad (8)$$

其中,  $\sigma$  表示 Sigmoid 激活函数,  $\text{conv}_{3 \times 3}$  表示  $3 \times 3$  卷积,  $F'_C{}^{\text{Avg}}$  和  $F'_C{}^{\text{Max}}$  分别表示为  $F'_C$  通过平均池化和最大池化后的特征.

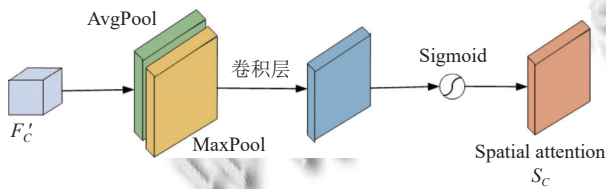


图 5 空间注意力模块

### 1.4.3 混合注意力模块

通过在跳跃路径中引入混合注意力模块, 来增强模型的特征描述. 如图 6 所示, 混合注意力模块首先对空间分支和上下文分支的输出特征进行连接, 接着对融合的特征进行  $3 \times 3$  卷积、批量归一化和 ReLU 单元处理, 再将空间分支的输出通过空间注意力模块进行处理, 得到的结果与之前融合的特征相乘. 再将上下文分支的输出应用于通道注意力模块, 得到的注意力图乘以之前的融合特征, 结果添加到融合特征, 最后对获得的特征进行卷积和上采样操作得到最后的特征图.

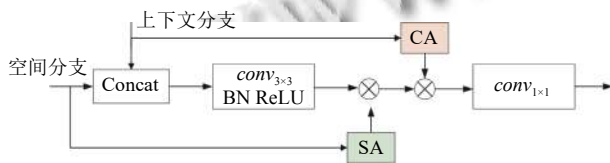


图 6 混合注意力模块

## 2 实验

### 2.1 数据集和评价指标

本文的实验数据来自某三甲医院的临床医学图像数据, 其中包括髌骨图像数据 757 例, 膝盖图像数据 241 例, 分别实现了多分类医学图像分割任务. 为避免

样本过少可能出现的过拟合问题, 我们通过随机旋转、缩放、几何平移等方法对数据集进行了预处理, 最终得到 4 500 张髌骨图像, 1 700 张膝盖图像作为数据集. 我们用各类别的平均 Dice 相似系数 ( $DSC$ ) 和平均 Hausdorff 距离 ( $HD$ ) 来评估我们的方法.  $DSC$  用来度量两个集合的相似度, 侧重于度量分割的内部填充的准确性.  $HD$  主要用来度量边界的分割准确度. 它们的定义如下:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (9)$$

$$HD(A, B) = \max(h(A, B), h(B, A)) \quad (10)$$

$$h(A, B) = \max_{a \in A} \{\min_{b \in B} \|a - b\|\} \quad (11)$$

$$h(B, A) = \max_{b \in B} \{\min_{a \in A} \|b - a\|\} \quad (12)$$

其中,  $A$  代表医学影像的标签图,  $B$  代表算法分割的预测图. 式 (10) 为双向  $HD$ , 式 (11) 和式 (12) 分别为从  $A$  到  $B$  和从  $B$  到  $A$  的单向  $HD$ .

### 2.2 实验细节

基于 Python 3.7, PyTorch 1.6.0, Nvidia 1080Ti GPU 完成实验. 训练时超参数使用如下相同设置: batch\_size 为 4, 初始学习率为 0.01, 优化器为随机梯度下降 (SGD) 混合使用交叉熵损失函数 CrossEntropyLoss() 和 Dice Loss() 两个损失函数.

在相同的实验条件下, 将我们的模型分割结果与经典的分割网络 U-Net, U-Net++, TransUNet, Swin-Unet 等进行对比分析, 实验结果如表 1, 表 2 所示.

表 1 不同方法在髌骨数据集上的分割结果

| 方法             | $DSC$ (%)    |              |              |              |              |              | $HD$ (mm)    |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                | 平均           | 髌臼杯          | 股骨柄          | 股骨           | 髌臼           | 股骨周围组织及肌肉    |              |
| U-Net          | 75.30        | 77.99        | 73.95        | 75.17        | 74.62        | 74.77        | 39.81        |
| U-Net++        | 84.69        | <b>84.88</b> | 84.93        | 84.68        | 84.50        | 84.45        | 36.34        |
| TransUNet      | 86.23        | 78.38        | 90.61        | 85.86        | <b>88.85</b> | 87.48        | 64.68        |
| Swin-Unet      | 84.16        | 73.53        | 89.59        | 82.92        | 86.49        | 88.27        | 34.56        |
| CNN+Bi-RNN     | 84.00        | 77.74        | 89.01        | 84.71        | 81.05        | 87.49        | 38.76        |
| PointRend_Unet | 84.71        | 80.50        | 87.48        | 79.43        | 86.80        | 89.34        | 32.87        |
| Ours           | <b>86.79</b> | 78.64        | <b>91.15</b> | <b>85.96</b> | 88.60        | <b>89.59</b> | <b>32.05</b> |

从表 1 中可以看出, 在髌骨数据集上, 我们的模型得到的  $DSC$  和  $HD$  参数分别为 86.79% 和 32.05 mm, 取得了较为准确的分割结果. 与经典的医学图像分割网络 U-Net, U-Net++ 相比, 我们的模型分割结果在  $DSC$  和  $HD$  上得到了较大的提升. 与 TransUNet 和 Swin-

Unet 相比, 我们的  $DSC$  分别提高了 0.56%, 2.63%,  $HD$  分别下降了 32.63 mm, 2.51 mm. 与其他关节图像分割方法 CNN+Bi-RNN、PointRend\_Unet 相比, 我们的分割结果得到的  $DSC$  分别提高了 2.79% 和 2.09%, 说明我们的模型加强了编码器和解码器之间的信息传递, 减少了信息损失, 提高了算法的分割精度. 我们还在膝盖数据集上评估了我们的方法. 所有实验都是在相同的设置下进行的, 以便进行公平的比较. 表 2 总结了实验结果, 在膝盖数据集上, 得到的  $DSC$  和  $HD$  参数分别为 88.17% 和 29.57 mm, 也取得了较好的分割结果. 这表明了我们提出的模型具有通用性和鲁棒性.

表 2 不同方法在膝盖数据集上的分割结果

| 方法             | $DSC$ (%)    |              |              |              |              | $HD$ (mm)    |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                | 平均           | 股骨周围组织及肌肉    | 关节囊及周围组织及肌肉  | 骨头与股骨头之间的区域  | 骨头与股骨假体之间的区域 |              |
| U-Net          | 80.54        | 82.15        | 81.15        | 80.29        | 79.71        | 79.42        |
| U-Net++        | 85.92        | 85.10        | 86.23        | 86.42        | 85.77        | 86.09        |
| TransUNet      | 86.55        | 78.08        | <b>91.60</b> | 86.04        | <b>88.15</b> | 88.90        |
| Swin-Unet      | 84.43        | 88.17        | 84.87        | 87.60        | 73.82        | 87.68        |
| CNN+Bi-RNN     | 84.57        | 83.62        | 88.10        | 85.40        | 79.17        | 86.56        |
| PointRend_Unet | 81.45        | 82.33        | 80.87        | 76.15        | 78.30        | 89.60        |
| Ours           | <b>88.17</b> | <b>90.29</b> | 88.96        | <b>89.91</b> | 81.57        | <b>29.57</b> |

通过对我们提出的模型与 Swin-Unet 网络的分割结果进行可视化对比分析, 结果如图 7. 可以看出我们的模型得到的分割结果边界更加清晰和光滑, 而且对不同类别的分割结果更加准确. 说明本文提出的算法提高了信息传递的有效性.

### 2.3 消融实验

通过分别移除跳跃路径上的密集 Swin Transformer 模块和混合注意力模块, 来验证它们的有效性, 并在我们的数据集上完成了消融实验, 结果如表 3、表 4 所示. 我们将 Swin-Unet 作为基准实验模型, 实验表明, 当算法中仅有 DSTB 模块时, 取得了  $DSC$  为 86.59%,  $HD$  为 32.32 mm 的结果, 较好地提升了分割的准确性; 仅有混合注意力模块时,  $DSC$  为 84.41%, 小幅度地提高了分割结果的准确性. 当同时加入 DSTB 模块和混合注意力模块时, 在髌骨数据集上取得了  $DSC$  为 86.79%、 $HD$  为 32.05 mm 的分割结果, 在膝盖数据集上取得了  $DSC$  为 88.17%、 $HD$  为 29.57 mm 的分割结果, 均有显著提升. 因此说明二者对于算法性能的提升都不可缺少, 也证明了我们提出的算法的有效性.

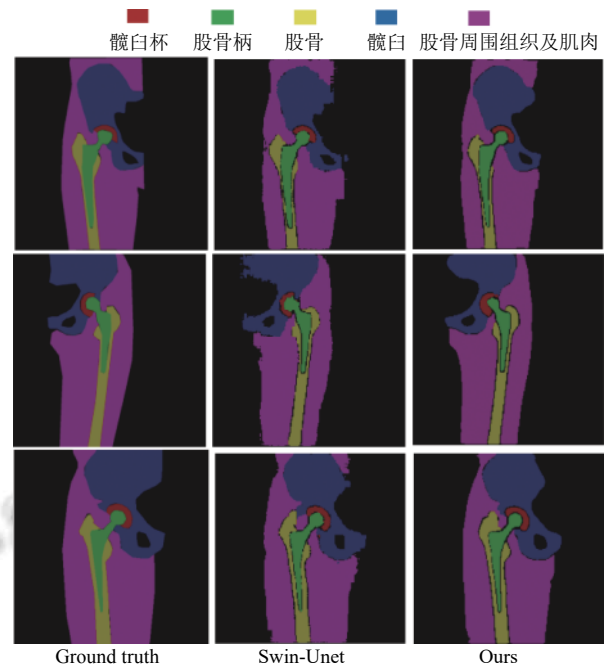


图 7 不同方法的分割结果图

表 3 在髌骨数据集上的消融实验

| 方法           | $DSC$ (%) | $HD$ (mm) |
|--------------|-----------|-----------|
| Swin-Unet    | 84.16     | 34.56     |
| DSTB模块       | 86.59     | 32.34     |
| 混合注意力模块      | 84.41     | 39.06     |
| DSTB+混合注意力模块 | 86.79     | 32.05     |

表 4 在膝盖数据集上的消融实验

| 方法           | $DSC$ (%) | $HD$ (mm) |
|--------------|-----------|-----------|
| Swin-Unet    | 84.43     | 30.92     |
| DSTB模块       | 86.82     | 35.39     |
| 混合注意力模块      | 85.41     | 36.42     |
| DSTB+混合注意力模块 | 88.17     | 29.57     |

### 3 结束语

在本文中, 我们提出了一种基于 Swin-Unet 的图像分割方法——DH-Swin Unet, 用于关节图像分割. 我们基于 Swin-Unet 网络, 重新设计改进其跳跃连接路径, 我们在跳跃连接中引入密集连接的 Swin Transformer 块和混合注意力模块, 来增强特征描述和信息传递. 通过在多种医学图像数据上进行实验, 结果表明我们提出的模型具有更好的密集连接效率, 更稳定的信息传递效果. 其中在髌骨数据集上得到了  $DSC$  为 86.79%、 $HD$  为 32.05 mm 的分割结果, 说明本文所提出的算法的有效性.



## 参考文献

- 1 Wang RS, Lei T, Cui RX, *et al.* Medical image segmentation using deep learning: A survey. *IET Image Processing*, 2022, 16(5): 1243–1267. [doi: [10.1049/ipr2.12419](https://doi.org/10.1049/ipr2.12419)]
- 2 LeCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324. [doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791)]
- 3 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 3431–3440.
- 4 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 234–241.
- 5 Chen JN, Lu YY, Yu QH, *et al.* TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv:2102.04306*, 2021.
- 6 Liu Z, Lin YT, Cao Y, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021. 10012–10022.
- 7 Cao H, Wang YY, Chen J, *et al.* Swin-Unet: Unet-like pure Transformer for medical image segmentation. *Proceedings of the 2022 European Conference on Computer Vision*. Tel Aviv: Springer, 2022. 205–218.
- 8 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *Proceedings of the 3rd International Conference on Learning Representations*. San Diego, 2015.
- 9 Hou RB, Chang H, Ma BP, *et al.* Cross attention network for few-shot classification. *arXiv:1910.07677v1*, 2019.
- 10 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 11 Fu J, Liu J, Tian HJ, *et al.* Dual attention network for scene segmentation. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 3146–3154.
- 12 刘云鹏, 蔡文立, 洪国斌, 等. 应用图像块和全卷积神经网络的肩关节 MRI 自动分割. *中国图象图形学报*, 2018, 23(10): 1558–1570. [doi: [10.11834/jig.180044](https://doi.org/10.11834/jig.180044)]
- 13 Liao HF, Mesfin A, Luo JB. Joint vertebrae identification and localization in spinal CT images by combining short- and long-range contextual information. *IEEE Transactions on Medical Imaging*, 2018, 37(5): 1266–1275. [doi: [10.1109/TMI.2018.2798293](https://doi.org/10.1109/TMI.2018.2798293)]
- 14 Zeng GD, Yang X, Li J, *et al.* 3D U-Net with multi-level deep supervision: Fully automatic segmentation of proximal femur in 3D MR images. *Proceedings of the 8th International Workshop on Machine Learning in Medical Imaging*. Quebec City: Springer, 2017. 274–282.
- 15 Nie D, Wang L, Trullo R, *et al.* Segmentation of craniomaxillofacial bony structures from MRI with a 3D deep-learning based cascade framework. *Proceedings of the 8th International Workshop on Machine Learning in Medical Imaging*. Quebec City: Springer, 2017. 266–273.
- 16 宋平, 范哲奇, 智信, 等. 基于深度学习的膝关节 CT 图像自动分割准确性验证研究. *中国修复重建外科杂志*, 2022, 36(5): 534–539.
- 17 Chaurasia A, Culurciello E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. *Proceedings of the 2017 IEEE Visual Communications and Image Processing*. St. Petersburg: IEEE, 2017. 1–4.
- 18 Zhou ZW, Siddiquee MR, Tajbakhsh N, *et al.* UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 2020, 39(6): 1856–1867. [doi: [10.1109/TMI.2019.2959609](https://doi.org/10.1109/TMI.2019.2959609)]
- 19 Zhao HS, Qi XJ, Shen XY, *et al.* ICNet for real-time semantic segmentation on high-resolution images. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 405–420.
- 20 Chen FJ, Ding Y, Wu ZX, *et al.* An improved framework called Du++ applied to brain tumor segmentation. *Proceedings of the 15th International Computer Conference on Wavelet Active Media Technology and Information Processing*. Chengdu: IEEE, 2018. 85–88.
- 21 Wu ZX, Chen FJ, Wu DY. A novel framework called HDU for segmentation of brain tumor. *Proceedings of the 15th International Computer Conference on Wavelet Active Media Technology and Information Processing*. Chengdu: IEEE, 2018. 81–84.
- 22 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 3–19.

(校对责编: 牛欣悦)