计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(1):141−147 [doi: 10.15888/j.cnki.csa.009374]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

IUINet: 基于 Shift 的双流映射 3D 医学分割模型①

朱庚鑫, 程远志, 刘　豪

(青岛科技大学 信息科学技术学院, 青岛 266061)
通信作者: 程远志, E-mail: yzchengqust2007@163.com

摘　要: 为了提高特征融合, 我们设计了动态全连接层 (DyFC), 该方法重新定义了权重和偏置, 使用基向量来代表新的权重和偏置, 基向量的系数是根据每一个输入特征进行学习得到的, 权重和偏置不再是共享的, 而是特有的, 这对于每一个特征的表达更具有专向性. 在本文中, 我们提出了一种双流映射结构模型 IUINet. IUINet 是通过 3DShift 操作、空间可分离卷积的组合来实现医学图像分割任务, 同时保持精度和效率之间的平衡. 所提出来的 IUINet 遵循编码器-解码器结构, 其中编码器一部分包含 Shift 操作、逐点 Conv1×1 操作, 另一部分包含空间可分离卷积操作. IUINet 运用了多尺度输入以及多尺度特征映射层, 提高反向传播速度, 降低反向传播的平均距离. 提高模型的精确度, 增加模型泛化能力, 减少过拟合.

关键词: Shift; 基向量; 动态全连接层; 医学图像分割

引用格式: 朱庚鑫,程远志,刘豪.IUINet: 基于 Shift 的双流映射 3D 医学分割模型.计算机系统应用,2024,33(1):141–147. http://www.c-s-a.org.cn/1003-3254/9374.html

IUINet: Two-flow Mapping 3D Medical Segmentation Model Based on Shift

ZHU Geng-Xin, CHENG Yuan-Zhi, LIU Hao

(College of Information Science and Technology, Qingdao University of Science & Technology, Qingdao 266061, China)

Abstract: This study designs the dynamic fully connected layer (DyFC) to enhance the feature fusion, which redefines the weights and biases by adopting base vectors to represent the new weights and biases. The coefficients of the base vectors are learned based on each input feature, and the weights and biases are no longer shared but unique, which provides more directional expressiveness for each feature. In this study, a dual-stream mapping architecture model IUINet is proposed. IUINet combines the 3DShift operation and spatial separable convolution to achieve medical image segmentation tasks and maintain a balance between accuracy and efficiency. The proposed IUINet follows an encoder-decoder structure, where the encoder consists of two parts. One part includes the Shift operation and pointwise Conv1×1 operation, and the other part incorporates spatial separable convolution operation. IUINet utilizes multi-scale inputs and multi-scale feature mapping layers to improve the backpropagation speed and reduce the average backpropagation distance. Finally, this enhances the model accuracy, improves generalization ability, and reduces overfitting.

Key words: Shift; base vector; dynamic fully connected layer (DyFC); medical image segmentation

1　引言

骨干神经网络特征提取能力是计算机视觉领域的基础. 卷积神经网络自 AlexNet[1]取得巨大的进展以来, 卷积神经网络在骨干网络这一领域占据主导地位已经 10 多年的时间, 其强大而复杂的映射能力, 已经被证实在处理各种视觉任务上是有用模型. 卷积神经网络在

① 基金项目: 国家自然科学基金 (61702135, 61806107)
收稿时间: 2023-07-20; 修改时间: 2023-08-29; 采用时间: 2023-08-31; csa 在线出版时间: 2023-11-28
CNKI 网络首发时间: 2023-11-30

医学图像分析领域成为主流, 特别是编码器-解码器结构网络, 在过去的几年里受到了广泛的关注. U-Net 网络凭借着跳跃连接方式将编码器提取的低级特征连接到解码器, 以此帮助网络模型获得在下采样过程中模型丢失的上下文信息, 展现了十分优越的性能. 许多 U-Net 变体也随之产生, 它们通过改进跳跃连接方式或者在跳跃连接之前进行增强编码器特征等方案成功的改进了基线的 U-Net.

尽管 U-Net 及其变体在各种医学图像分割任务上取得了良好的性能, 但是仍然存在性能缺陷. 首先, 卷积更加重视局部信息, 无法获取全局特征之间的长期依赖关系, 这是由于卷积是通过卷积核逐渐获取局部特征, 而不是一次性提取全局特征. 其次, 普通的特征融合没有建模全局多尺度上下文信息的能力, 尽管很多 U-Net 变体通过卷积层或者残差层来提高这个能力, 并取得了一定的效果, 但是在减小语义差距上仍然还在努力.

在本文中, 我们研究了一种新的 3D 医学图像分割模型 IUINet. 普通的全连接层中的权重对于所有特征来说是共享的, 将全连接层的权重变为每个特征独有的, 这有助于每个特征的表达. 在视频识别任务中 Shift 的作用起到了显著的效果, 将 Shift 应用到 3D 医学 CT 图像上, 通过切片维度 (H) 和空间维度 (W, D) 上的有效位移, 可以有效地提高效益以及精确度, 并且相对于传统的卷积来说, 使用 Shift 可以使参数和计算量减少, 解决了 3D 医学 CT 图像需要大量硬件资源的挑战. 空间可分离主要是处理图像的空间维度, 通过将卷积核划分为两个小的卷积核可以有效地减少计算复杂度, 使得网络模型的运行速度更快. 我们将上述两个方案作为 IUINet 的编码器, IUINet 是双流架构, 必然会带来巨大的参数量和计算量, 使用上述两种方案作为 IUINet 模型架构的编码器有效地降低了参数量和计算量, 并且整个模型的性能表现仍然可观.

综上所述, 在目前的研究中全连接层仍是共享参数以及模型运行所需要的计算量和参数量大等问题, 为了解决这些问题, 本文贡献如下.

(1) 为了应对复杂的特征分布, 增加模型泛化性以及鲁棒性, 我们设计了动态全连接层, 该方法在特征筛选上提供了一个新视角, 从而更好地获取特征上的上下文互信息.

(2) 我们设计了一个新的双流映射结构模型 IUINet, 在参数量和计算量微增加的情况下, 在 CT 医学图像器官分割任务上取得了可观的性能.

2 相关工作

在本节, 我们主要从医学图像模型主要方法、Shift 操作以及动态权重这 3 个方面进行了相关工作的综述.

2.1 医学图像模型

医学图像分析是近几年深度视觉领域的热点, 随之而来的就是一些医学图像处理模型, 其中 U-Net 被广泛关注. U-Net 在医学图像分割中具有开创性和广泛的应用, 在各种医学图像分割任务中取得了先进的结果. 然后, 开发了许多 U-Net 的变体, 如 Unet++、nnUnet、UNet3+和 3DUnet, 它们主要基于编码器-解码器范式. 这些方法以 CNN 为主干网络提取图像特征, 并结合一些精细的技巧如: 跳跃连接、多尺度表示、特征交互、嵌套结构等新的网络结构或机制, 进行特征增强, 提高模型表现能力. 上述模型, 具有不错的性能, 但是在处理 3D 医学图像分割任务上具有大量的计算量和模型参数. 本文提出的 IUINet 可以有效地降低网络的计算量以及模型参数量, 并且模型仍然可以保证性能的稳定, 甚至可以超过最新方法.

2.2 Shift 操作

在 2017 年 Shift 操作已经是深度学习中应用的技术, 其思想是通过对输入数据进行像素级的位移操作来实现特征提取和信息交互. 在相关研究中, 出现了一些与 Shift 操作相关的工作.

ShiftNet[2]方法使用了 Shift 操作, 以提高模型的性能和计算效率. ShiftNet 使用一种无参数、无 flop 的 Shift 操作来代替传统的空间卷积操作, 以降低模型的计算复杂度和提高计算效率, 并可以在分类等任务上以较少的参数取得较强的性能. 但是在移动操作的时候参数是固定的, 缺乏了灵活性.

Shift-invariant neural network[3]方法通过使用自适应多相采样, 使得卷积神经网络真正实现位移不变性.

Shift 操作在自然图像上的应用已经不是很新鲜的事情了, 但是在医学 CT 影像上仍然比较新颖. 文献[4]中提出的 Shift 与 VIT 的结合用来代替注意力机制的方法, 并表示注意力机制可能不是 VIT 成功的重要因素. 他们在论文中使用的是一种部分移位操作, 遵循的是 TSM 中提出的方法[5]. 在通道维度, 只有一部分的通道进行移位操作, 其他的通道保持不变. 上述方法中,

Shift 操作中的参数是固定的, 无法适应数据集的变化或者处理一些复杂的数据分布. 我们的方法遵循的是文献[6]中提到的可学习的 3DShift 操作, 在这个方法中操作系数是可以学习的, 可以有效地适应数据集的变化或者处理一些复杂的数据分布从而获得更好的性能.

2.3 动态权重

当前深度学习两大领域计算机视觉领域和 NLP 领域中都有在使用动态卷积核. 在计算机领域中, 通过线性层直接生成卷积核, 但是卷积核具有大量的参数并且在硬件上线性层的效率很低. 在 NLP 领域, 一些作品也是通过线性层直接生成卷积核的. 在 DyNet[7]中解决了上述所描述的问题, 通过预测线性组合固定卷积核的系数, 降低了卷积核之间的相关性解决了这个问题, 并且 CNN 在硬件上的速度也得到了提升, 这个技术已经在华为得到发展. 尽管动态卷积在减少冗余计算量上表现卓越的乘积, 但是卷积网络仍然是依赖于局部特征的提取, 总会丢失一定的全局特征信息. 在本文中, 我们提出的方法通过基向量注意力重新定义权重和偏置, 对于每个特征来说都有独有的权重以及偏置.

3 方法

在本节中, 首先概述了我们提出的模型结构. 然后给出了所提出分层的具体细节.

我们提出的模型的整体流程如图 1 所示, 遵循了 U-Net 的设计思想, 由编码层、跳跃层、解码层、映射输出层. 该模型结构中的编码器分为两个骨干网络作为特征提取. 其主要新颖之处: 其一, 在于骨干 A 运用了 shift 的思想进行分层特征提取, 骨干 B 运用了空间可分离的思想进行分层特征提取, 然后将两个骨干网络获取的分层特征进行融合. 其二, 特征融合、残差、输出层, 不再是传统的方法, 而是使用我们设计的动态全连接层进行操作. Conv 模块和 SAMC 模块如图 2 所示.

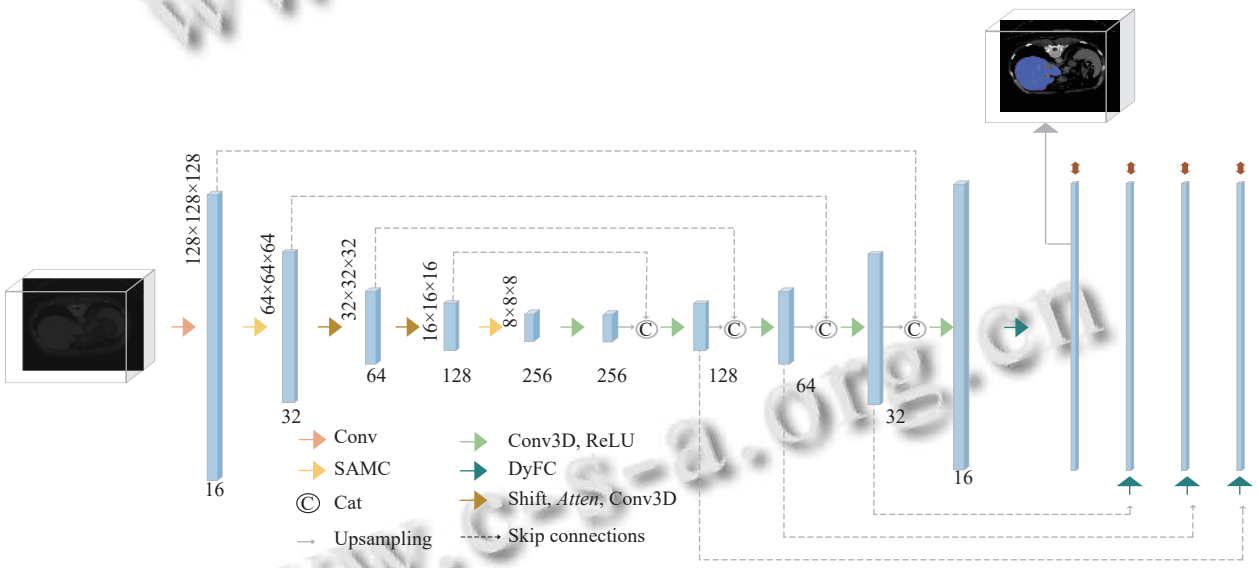

图 1 本文模型结构

具体来说, 给定输入图像 $I\in(B, C, H, W, D)$, H、W、D、C、B 分别代表图像的高度、宽度、深度、通道数和批量. 对于骨干网络 A 而言, 将图片放入网络的 5 个阶段中生成分层特征, 5 个阶段的特征图分辨率分别为原始分辨率的{1/1, 1/2, 1/4, 1/8, 1/16}, 这些特征通过与骨干网络 B 进行特征融合, 然后进入跳跃连接, 为分割任务提供低级到高级的特征. 同时对于骨干网络 B 而言, 将图片放入网络的 5 个阶段中生成分层特征, 5 个阶段的特征图分辨率分别为原始分辨率的{1/1, 1/2, 1/4, 1/8, 1/16}. 将这两个网络获得的分层特征进行融合, 通过特征选择捕获多尺度特征以及局部特征, 获得新的分层特征, 将 5 个阶段的层次特征输入到 Decoder 模块, 再到 Mapping 层, 预测分割结果. 接下来, 我们详细阐述了提出的网络设计方案.

3.1 动态全连接层

近年来, 网络深度不断增加, 梯度爆炸, 梯度消失等问题也随之出现. 直到残差连接被广泛地使用, 卷积融合是进行残差处理的常用方法之一. 由于卷积更专注于局部特征, 对于全局特征的捕获能力仍然存在不足, DyFC 可以有效地解决这个问题. 如图 3 所示.

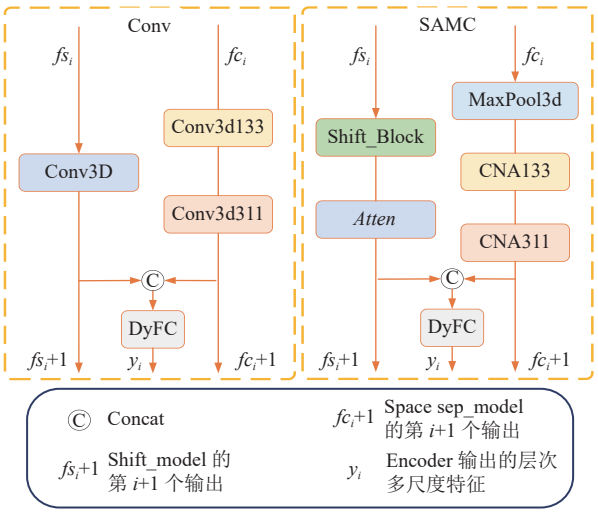

图 2 Conv 模块和 SAMC 模块

动态全连接层作为残差连接的新方法. 相比较卷积融合操作, 在完成残差操作的同时, 还增强了融合过程中全局特征的捕获能力.

Fully connected layer: 在具有三维图像的传统全连接网络中, 网络的输入是一个四维张量 $X\in\{C, H, W, D\}$, C 为输入通道, H、W、D 分别为高度、宽度和深度. 可以将三维图像全连接运算定义为:

$$Y = X@Weight + bias \tag{1}$$

其中, $Weight$ 是一个 $(N\times M)$ 的矩阵, $bias$ 是一个 $(M\times 1)$ 的矩阵. 这样的全连接层特征的权重和偏置是共享的, 不具备特殊性, 泛化能力还需要继续努力.

我们提出的动态全连接层与之前的全连接工作有所不同, 即我们的模型可以学习每个特征的权重, 在进行特征融合的时候可以保留更多的上下文联系, 已达到获取更多的全局信息. 这使得在网络参数不过多增加的情况下, 可以更好地表达每个特征.

在先前的全连接工作之后, 我们考虑了注意力机制的方法, 允许我们通过注意力机制来对全连接参数进行优化. 我们动态全连接层的参数可以定义为:

$$Y = (X@Weight_s + bias_s)@Atten \tag{2}$$

其中, $Weight_s = \{(w_1, w_2, \cdots, w_i)|i\in k\}$, w_i 是 $Weight_s$ 的第 i 个基向量, 原始的 $Weight$ 是一个 $(N\times M)$ 的矩阵, 我们设计的全连接层中 $Weight_s$ 是一个由 i 个 $(N\times M)$ 的基向量构成的. $bias_s = \{(b_1, b_2, \cdots, b_i)|i\in k\}$, b_i 是一个 $bias_s$ 的第 i 个基向量, 原始的 $bias$ 是一个 $(M\times 1)$ 的矩阵, 我们设计的全连接层中 $bias_s$ 是一个由 i 个 $(M\times 1)$ 的基向量构成, 这样的权重和偏置可以增加模型的学习能力. 根据向量的线性组合, $Weight$ 与 $bias$ 由基向量可以表示为:

$$Weight = Weight_s@Atten \tag{3}$$

$$bias = bias_s@Atten \tag{4}$$

其中, $Atten = \{(a_1, a_2, \cdots, a_i)|i\in k\}$, 这里 $Atten$ 是一个可学习的参数, 用来表示 $Weight_s$ 基向量中每个基向量的系数. 通过注意力机制来学习到 $Atten$, 可以表示为:

$$Atten = attention(xlr) \tag{5}$$

其中, xlr 是通过用原图像的均值和方差, 计算得到的. xlr 可以表示为:

$$xlr = concat(mean - std, mean + std) \tag{6}$$

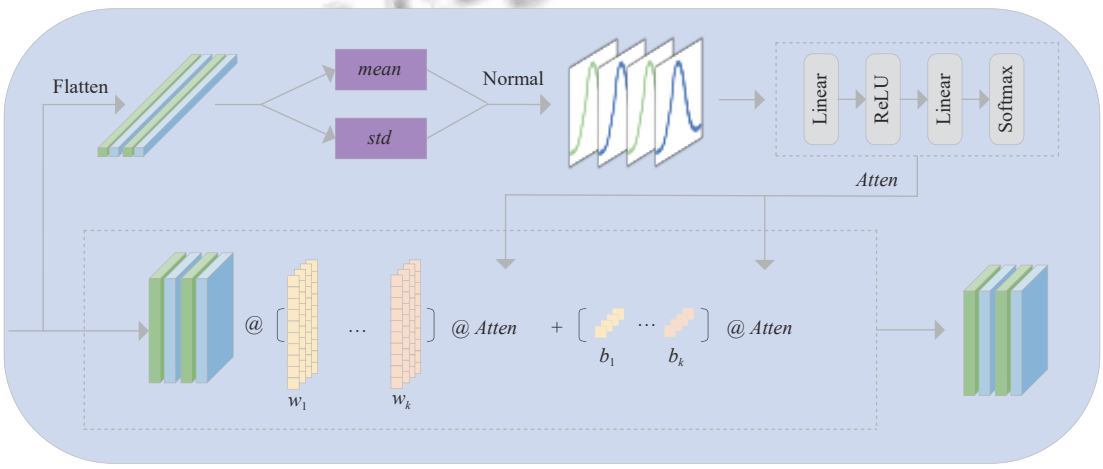

图 3 DyFC 模块

我们设计的 DyFC 可以用在多个场景: 作为残差连接的新方式、作为输出头. 在我们设计的 IUINet 中在这两个方法上的改进都进行了实现. 作为残差连接时, 以往的工作是通过直接进行 Add、卷积融合等方式实现的, DyFC 可以作为一个全新的连接方式, 不仅可以在通道维度进行操作, 还可以指定维度进行操作. 原始的全连接层特征的权重和偏置是共享的, 不具备特殊性, 通过基向量重新设计权重和偏置可以有效地解决共享的问题, 增加权重和偏置的特殊性, 这种基向量的选择可以方便地提取、表示图像的特征, 选择合适的基向量, 可以将高维度特征空间中的冗余特征去除, 从而提高模型的泛化能力和精度.

3.2 Shift_Encoder: Model Architecture Design

该模型整体架构反映了 ResNet 架构中的残差块的设计. 每一个 Shift_Block 包含 Shift 操作以及逐点 Conv1×1 操作, 用来促进跨通道的信息交换. 这里的残差块的设计是使用了我们的动态全连接层 (DyFC) 进行特征融合的, 这样可以更好地融合信息, 使得信息具有更多的全局特征.

我们的 Shift_Block 具体来说, 该块是由 3 个顺序堆叠的组件组成: 卷积、Shift 和动态全连接层 (DyFC). Shift 运算在 CNN 中已经得到了很好的研究. 它有很多方向, 例如主动位移[8]、稀疏位移[9]和部分位移操作[5]等. 在这项研究中, 我们遵循的是可学习的 3D 位移操作[6], 如图 4 所示. 给定一个输入张量, 在 H, W, D 方向, 分别沿着通道维度进行移动. 在这里使用了传统离散位移操作的连续形式, 允许位移操作直接通过反向传播来优化位移参数. 这种可学习 3D 移动操作能够使整体架构以高效的方式学习一个联合的 3D 位移核, 该核以有效的方式聚合激活中的判别特征, 并且可以学习切片维度 (H) 位移原语的能力, 在联动切片 (H) 和空间 (W, D) 上下文上共同移动, 这使得网络能够以更少的总体参数有效地参考重要的切片 (H) 和空间 (W, D) 上下文跨度.

3.3 空间可分离网络: Model Architecture Design

该模型架构主要是运用空间可分离方法, 将传统的卷积替换为参数量更少的两个小核卷积. 3D 医学 CT 图像在训练时, 参数大, 计算量大是其面临的重大挑战之一, 将空间可分离卷积应用到分析 3D 医学 CT 图像上可以有效地改善这个问题. 空间可分离在参数方面, 由于使用两个更小的卷积核, 所以空间可分离卷积的参数量比传统卷积要少很多; 在计算量方面, 计算量比传统卷积也要小很多.

我们这个 IUINet 模型的设计采用了双流模型的思想. 在编码器阶段, 我们使用 Shift 网络和空间可分离网络进行特征提取, 这两个模型都在效率上有着一定的优势, 通过动态全连接层将两个网络的特征进行融合作为对方的信息补丁, 然后进入解码器阶段, 对解码器输出的特征进行映射, 这使得模型在具有精度不丢失甚至有所提升的情况下, 整体参数和计算并不会明显提升.

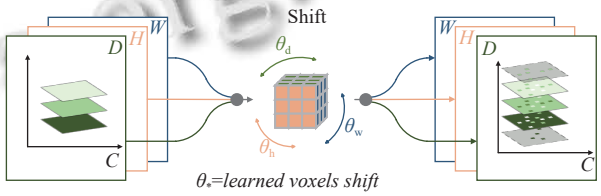

图 4 Shift 模块

4 实验分析

4.1 数据集描述

本文使用的数据集来自 Synapse 的多器官腹部分割数据集. 该数据集是从正在进行的结直肠癌化疗试验和回顾性腹壁疝病例研究中随机选取了 50 例腹部 CT 扫描, 该过程是在机构审查委员会 (IRB) 的监督下进行的. 这 50 个扫描是在门静脉造影期间捕获的, 具有不同的体积尺寸 (512×512×85–512×512×198) 和视野范围 (约 280×280×280 mm^3–500×500×650 mm^3). 平面分辨率从 0.54×0.54 mm^2 到 0.98×0.98 mm^2 不等, 切片厚度范围为 2.5–5.0 mm. 我们先使用 Training-Testing 进行预训练然后利用 TransUNet[10]中提供的数据分割在 18 个训练样本上训练我们的模型, 并在 12 个验证案例上进行评估.

4.2 实验详情

本文的方法均使用 Torch 实现的, 所有的实验均在一台配置为 Intel(R) Core(TM) i9-10900K CPU @ 3.70 GHz 处理器、24 GB NVIDIA GeForce RTX 3090, 操作系统为 Linux 系统上进行的.

模型训练和超参数设置: 为了防止过拟合, 减小模型参数, 加快模型训练速度, 我们将图像和标签的大小调整为 128×128×128, 批次大小为 1, Epoch 设置为 2 000, 但当模型在验证集中的损失值不降低的轮次大

于等于 100 的时候结束训练. 使用 Adam 优化器, 学习率初始设置为 0.001, 训练过程中最优 DSC 超过 20 次训练后不变化, 学习率就减小为 0.2 倍.

在本文中, 为了定量评估方法的准确性, 本文使用了 Dice similarity coefficient (DSC) 作为评判标准.

4.3 模型复杂度

我们模型按照输入尺寸 (1, 1, 128, 128, 128) 得到计算量为 153.8G, 参数量为 49.0M, 本文使用了 ptflops 中的 get_model_complexity_info 方法在 NVIDIA GeForce RTX 3090 对分割方法进行了计算量和参数的计算.

4.4 对比实验

为了证明我们提出的方法的有效性, 我们将在 Synapse 数据集上与其他 14 种方法进行比较得到定量结果. 表 1 中提供了在 Synapse 数据集中的 3 个器官上我们方法与其他 14 种方法的分割统计结果, 在肝脏、脾脏、胃上的平均 DSC 结果我们提出的方法为第 1 名, 3 个器官的 DSC 结果分别为 0.9546、0.9135、0.9063. 肝脏分割结果与第 1 名相差 0.0138 个精度, 但是这明显高于基线方法. 胃部的分割精度达到了 0.9063 的高度, 比第 1 名高出 0.038 个精度. 3 个器官的平均得分为 0.9248.

表 1 实验结果对比

Model	AVG	Liver	Spleen	Stomach
VNet[11]	0.7513	0.8784	0.8056	0.5698
U-Net[12]	0.8523	0.9343	0.8667	0.7558
Att-UNet[13]	0.8554	0.9357	0.8730	0.7575
DARR[14]	0.7665	0.9408	0.8990	0.4596
R50 U-Net[10]	0.8389	0.9335	0.8441	0.7392
MISSFormer[15]	0.8904	0.9441	0.9192	0.8081
R50 Att-UNet[10]	0.8523	0.9356	0.8719	0.7495
TransUNet[10]	0.8493	0.9408	0.8508	0.7562
Swin-Unet[16]	0.8718	0.9429	0.9066	0.7660
TransClaw U-Net[17]	0.8519	0.9428	0.8774	0.7355
LeVit-UNet-384s[18]	0.8491	0.9311	0.8886	0.7276
WAD[19]	0.8756	0.9440	0.8892	0.7935
nnFormer[20]	0.9139	0.9684	0.9051	0.8683
UNETR++$_{\text{Learnable Weight}}$[21]	0.9244	0.9624	0.9541	0.8566
IUINet (Our)	**0.9248**	0.9546	0.9135	**0.9063**

图 5 是 IUINet 和 UNETR++$_{\text{Learnable Weight}}$ 的胃部对比实验图, 其中红色细线是真实标签, 蓝色区域是预测结果, 如黄色正方形所在区域显示, 本文的方法是有效的. 这些结果表明, 我们方法的性能比第一名的方法具有一定的优势, 尤其是分割胃部的能力.

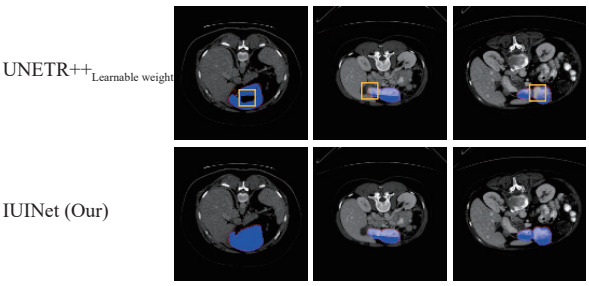

图 5 对比实验图

4.5 消融实验

本节对我们提出的 DyFC 以及 Shift 进行消融实验, 在表 2 中我们展示了基线效果、DyFC 作为输出头和不作为输出头的对比结果以及使用 Shift 和不使用 Shift 的对比结果.

表 2 基线实验结果对照

Shift+DyFC	Params (M)	FLOPs (G)	Liver	Spleen	Stomach
√	49.0	153.8	0.9546	0.9135	0.9063
×	48.63	153.89	0.9410	0.8847	0.8187

在表 3 中展示了使用 DyFC 作为输出头和不使用 DyFC 作为输出头的结果. 在实验中, 通过比较两种模型的参数量 (Params)、计算量 (FLOPs)、DSC 等指标, 评估了 DyFC 对模型性能的影响. 表 3 中的√代表使用 DyFC 作为输出头, ×代表不使用 DyFC 作为输出头. 结果表明使用 DyFC 作为输出头比不使用的性能要好. 而且参数量仅增加了 0.37, 计算量并没有提升反而降低了 0.09. 综上所示, DyFC 对模型是具有积极作用的.

表 3 DyFC 消融实验对照

DyFC	Params (M)	FLOPs (G)	Liver	Spleen	Stomach
√	49.0	153.8	0.9546	0.9135	0.9063
×	48.63	153.89	0.9422	0.8966	0.8745

在 Shift 模块消融实验中, 针对使用和不使用两种情况对模型性能进行了评估. 表 4 中展示了实验结果, 实验表明 Shift 模块不影响模型的参数以及计算量, 但是在精度上具有略微的提升. 3 个器官分别提升了 0.0024、0.0067 和 0.0118. 在各项指标上都具有更好的性能. 因此使用 Shift 模块可以有效地提升模型的性能.

表 4 Shift 消融实验对照

Shift	Params (M)	FLOPs (G)	Liver	Spleen	Stomach
√	49.0	153.8	0.9546	0.9135	0.9063
×	49.0	153.8	0.9522	0.9068	0.8945

5 结论与展望

我们提出了一个用于 3D 医学分割的双流映射结

构的模型, 命名为 IUINet. 在模型中我们有效地使用了 DyFC 和可学习的 Shift 模块, 通过注意力机制来应对复杂的特征分布, 增加模型的泛化性以及鲁棒性. 其中 DyFC 改变了传统的全连接层的权重和偏置的设置, 使它们变得更加灵活. 与现有的方法相比, 我们的 IUINet 在 Synapse 数据集上具有良好的分割性能, 同时相比传统的 3D 网络也减小了参数量和计算量.

参考文献

1 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2012. 1097–1105.

2 Wu BC, Wan A, Yue XY, *et al*. Shift: A zero flop, zero parameter alternative to spatial convolutions. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 9127–9135.

3 Chaman A, Dokmanić I. Truly shift-invariant convolutional neural networks. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021. 3772–3782.

4 Wang GT, Zhao YC, Tang CX, *et al*. When shift operation meets vision transformer: An extremely simple alternative to attention mechanism. Proceedings of the 36th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2022.

5 Lin J, Gan C, Han, S. TSM: Temporal shift module for efficient video understanding. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 7082–7092.

6 Fan LX, Buch S, Wang GZ, *et al*. RubiksNet: Learnable 3D-shift for efficient video action recognition. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 505–521.

7 Zhang YK, Zhang J, Wang Q, *et al*. DyNet: Dynamic convolution for accelerating convolutional neural networks. arXiv:2004.10694, 2020.

8 Jeon Y, Kim J. Constructing fast network through deconstruction of convolution. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 5955–5965.

9 Chen WJ, Xie D, Zhang Y, *et al*. All you need is a few shifts: Designing efficient convolutional neural networks for image classification. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7234–7243.

10 Chen JN, Lu YY, Yu QH, *et al*. TransUNet: Transformers make strong encoders for medical image segmentation. arXiv:2102.04306, 2021.

11 Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. Proceedings of the 4th International Conference on 3D vision (3DV). Stanford: IEEE, 2016. 565–571.

12 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.

13 Oktay O, Schlemper J, Le Folgoc L, *et al*. Attention U-Net: Learning where to look for the pancreas. arXiv:1804.03999, 2018.

14 Fu SH, Lu YY, Wang Y, *et al*. Domain adaptive relational reasoning for 3D multi-organ segmentation. Proceedings of the 23rd International Conference on Medical Image Computing and Computer-assisted Intervention. Lima: Springer, 2020. 656–666.

15 Huang XH, Deng ZF, Li DD, *et al*. MISSFormer: An effective medical image segmentation transformer. arXiv: 2109.07162, 2021.

16 Cao H, Wang YY, Chen J, *et al*. Swin-Unet: Unet-like pure transformer for medical image segmentation. Proceeding of the 2022 European Conference on Computer Vision. Tel Aviv: Springer, 2022. 205–218.

17 Yao C, Hu MH, Li QL, *et al*. TransClaw U-Net: Claw U-Net with Transformers for medical image segmentation. Proceedings of the 5th International Conference on Information Communication and Signal Processing (ICICSP). Shenzhen: IEEE, 2022. 280–284.

18 Xu GP, Wu XR, Zhang X, *et al*. LeViT-UNet: Make faster encoders with transformer for medical image segmentation. arXiv:2107.08623, 2021.

19 Li YJ, Cai WT, Gao Y, *et al*. More than encoder: Introducing transformer decoder to upsample. Proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Las Vegas: IEEE, 2022. 1597–1602.

20 Zhou HY, Guo JS, Zhang YH, *et al*. nnFormer: Interleaved transformer for volumetric segmentation. arXiv:2109.03201, 2021.

21 Kunhimon S, Shaker A, Naseer M, *et al*. Learnable weight initialization for volumetric medical image segmentation. arXiv:2306.09320, 2023.

(校对责编: 牛欣悦)