

动态感受野的图像分割神经网络模型

刘洋^{1,2)}, 舒利明³⁾, 孔雨秋^{1,4)*}, 杨婉肖²⁾, 李英平²⁾, 孔程玉²⁾

¹⁾ (大连理工大学创新创业学院 大连 116024)

²⁾ (大连理工大学电子信息与电气工程学部 大连 116024)

³⁾ (大连理工大学机械工程学院 大连 116024)

⁴⁾ (大连民族大学信息与通信工程学院 大连 116600)

(yqkong@dlut.edu.cn)

摘 要: 针对基于 U-Net 的分割方法在医学图像分割领域存在的下采样阶段信息丢失, 以及因固定尺寸卷积核带来的局部多尺度语义信息提取不足的问题, 提出一种动态感受野的神经网络模型. 首先通过构建特征递进级联模块, 获取编码器多尺度局部语义特征并将其赋值给解码器, 提升模型解码阶段图像语义信息修复的效果; 然后设计局部视野偏移模块, 增强固定尺寸卷积核提取视野内上下文语义信息的能力. 在 ISIC2018 和 BUSI 癌症图像分割数据集上的实验结果表明, 所提模型的 IoU 指标达到 83.92 ± 0.26 和 70.45 ± 1.70 , Dice 系数达到 91.09 ± 0.23 和 83.39 ± 1.15 , 比现有的医学图像分割方法更优.

关键词: 医学图像分割; 动态感受野; 特征递进级联; 局部视野偏移

中图分类号: TP391.41

DOI: 10.3724/SP.J.1089.2024.19979

Dynamically Receptive Field Networks for Image Segmentation

Liu Yang^{1,2)}, Shu Liming³⁾, Kong Yuqiu^{1,4)*}, Yang Wanxiao²⁾, Li Yingping²⁾, and Kong Chengyu²⁾

¹⁾ (School of Innovation and Entrepreneurship, Dalian University of Technology, Dalian 116024)

²⁾ (Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024)

³⁾ (School of Mechanical Engineering, Dalian University of Technology, Dalian 116024)

⁴⁾ (College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600)

Abstract: To address the issues of information loss during down-sampling and the insufficient extraction of the multi-scale local semantic information caused by the fixed-size convolutional kernels when applying a U-Net-based method to realize medical image segmentation, a dynamically receptive field network (DRFN) is proposed. Firstly, the network enhances the reducibility of the semantic information during decoding process by introducing the feature progressively cascaded integration (FPCI) module to extract multi-scale local semantic features from the encoder and assigns these features to the decoder. On this basis, the local vision adjustment (LVA) module is designed to strengthen the ability of the fixed-size convolution kernel to obtain contextual semantic information in a large receptive field. The experimental results on the ISIC2018 and BUSI cancer image segmentation datasets show that the IoU score of the proposed method reaches 83.92 ± 0.26 and 70.45 ± 1.70 while the Dice score reaches 91.09 ± 0.23 and 83.39 ± 1.15 , which outperforms the state-of-the-art medical image segmentation methods.

收稿日期: 2022-09-22; 修回日期: 2023-03-02. 基金项目: 国家自然科学基金(62006037, 61972064, 62172073); 国家重点研发计划(2021YFB3301904); 中央高校基本科研业务费(DUT22JC06, DUT22RC(3)064). 刘洋(1993—), 男, 博士, 讲师, 硕士生导师, CCF 会员, 主要研究方向为计算机视觉、人工智能等; 舒利明(1990—), 男, 博士, 教授, 博士生导师, 主要研究方向为人体数字孪生、医疗增强机器人等; 孔雨秋(1992—), 女, 博士, 讲师, 硕士生导师, 论文通信作者, 主要研究方向为计算机视觉、深度学习等; 杨婉肖(1998—), 女, 硕士研究生, 主要研究方向为计算机视觉; 李英平(1999—), 男, 硕士研究生, 主要研究方向为计算机视觉; 孔程玉(1996—), 男, 硕士研究生, 主要研究方向为计算机视觉.

Key words: medical image segmentation; dynamically receptive field; feature progressively cascaded integration; local vision adjustment

医学图像分割是计算机辅助医疗中一项重要的基础性工作, 分割结果的准确性对病灶辅助分析、医学 3D 重建等工作具有重要的影响^[1]. 然而, 医学应用场景的特殊性给医学图像分割的研究带来了挑战: (1) 医疗场景下的数据集往往规模不大, 因此医学图像分割模型在设计时需要避免模型过于复杂带来的过拟合问题; (2) 由于医学分析对时间效率要求较高, 因此医学图像分割模型的推理效率也是模型设计时需要考虑的重要因素之一. 医学图像分割是计算机辅助医疗领域的重要研究方向之一.

在医学图像分割任务中, 早期研究者利用病灶与正常组织之间视觉上的差异性来实现医学图像的分割^[2]. 随着深度学习技术的普及, 其在医学图像分割领域也取得了许多积极的成果^[3]. 尽管部分深度学习方法在医学图像分割领域取得了较好的分割精度, 但其往往结构较为复杂, 推理时间较长, 影响了该方法在医学辅助分析领域的应用. 当前, 能够较好地兼顾效率和精度的典型方法之一是基于 U-Net^[4]的医学图像分割模型. U-Net 通过一个自编码和解码的过程实现分割边界的确定, 具有网络层数少、效率高的优点; 在此基础上, 研究者也提出了很多改进模型^[5]. 然而, 基于 U-Net 的图像分割方法尚存在 2 方面的问题: (1) 在编码阶段, 基于 U-Net 的方法通过小尺度卷积层和下采样层的堆叠逐渐缩小特征图的空间尺寸, 进而实现关键信息的提取; 然而, 无论卷积还是下采样操作都只关注局部信息, 都是利用一种小尺寸的内核提取输入特征图的局部信息, 若要进行大范围图像语义信息的提取, 则需要通过级联多个小尺寸的卷积和下采样操作. 由于医学图像分割往往需要提取大范围的语义信, 因此很多基于 U-Net 的神经网络方法需要较深的编码器. 然而, 在编码阶段级联较多的卷积和下采样操作将带来模型参数的增加, 导致模型的训练和推理效率受到影响. 由于在 U-Net 设计中, 每次卷积操作之后都会将特征图数量翻倍^[4], 因此当使用很多卷积操作时, 模型的训练和推理效率将受到很大影响. 此外, 更多的下采样操作会导致图像空间语义信息丢失得更多, 极大地影响医学图像分割的准确性^[6]. (2) U-Net 的解码阶段结构与编码阶段相似, 仅将下采样操作改为上采样操作. 当前主流的上采样方法主要包

括反卷积和反池化, 这也是提取局部信息的操作; 然而, U-Net 上采样的过程涉及图像空间语义信息的修复, 仅使用单一尺度局部语义信息很难获得理想的图像分割结果. 因此, 设计一个能提取多尺度局部信息的结构, 将极大地提升基于 U-Net 的图像分割方法的效果.

针对基于 U-Net 的方法存在的问题, 本文在 U-Net 结构的基础上, 提出一种面向医学图像分割的动态感受野的网络(dynamically receptive field networks, DRFN)模型, 既能够保留 U-Net 高效性的特点, 也改进了 U-Net 存在的单一尺度局部语义信息提取不足的问题. 在编码的初始阶段, 设计一种特征递进级联(feature progressively cascaded integration, FPCI)模块获取不同感受野大小下图像的语义信息, 并将其传递到解码器上, 突破了 U-Net 在编码阶段因卷积核尺寸固定带来的仅能提取局部信息的限制; 还设计一种局部视野偏移(local vision adjustment, LVA)模块, 丰富网络在编码阶段通过卷积操作提取的图像语义信息. 在 2 个主流肿瘤检测数据集上对 DRFN 模型进行实验, 结果表明, 该模型在使用较少参数的情况下, 分割准确性更高.

1 相关工作

研究者从不同角度对医学图像分割准确性和高效性进行了探索, 早期主要从病灶与正常组织之间的差异性出发, 设计了许多传统方法实现对医学图像的分割, 如基于阈值的图像分割方法^[7-8]、基于边缘的图像分割方法^[9-10]、基于区域的图像分割方法^[11-12]和引入图论的方法^[13-14]. 随着深度学习技术的快速发展, 很多学者将深度学习技术应用到医学图像分割领域, 取得了显著成效.

在利用深度学习技术解决医学图像分割问题的过程中, 研究者从 3 类基本深度神经网络模型出发, 包括基于全卷积网络(fully convolutional networks, FCN)的方法、基于 U-Net 的方法, 以及基于生成对抗网络(generative adversarial network, GAN)的方法. Long 等^[15]通过保持卷积神经网络输出和输入尺寸一致来构建 FCN, 并通过像素级别的分类实现图像的分割. 为了克服 FCN 未能有效地利用全局上下文语义信息的问题, Liu 等^[16]提出

ParseNet, 利用每层的平均特征将全局上下文信息添加到 FCN 中. Roth 等^[17]和 Christ 等^[18]采用级联的方式训练 3D FCN, 进一步提升医学图像语义分割的精度和鲁棒性. 对于部分复杂医学图像, 研究者也开展了针对性的研究. 由于头颈器官的锥形束 CT 图像的软组织对比度低, FCN 方法通常会发生边界泄漏问题. 因此为了缓解该问题, Liu 等^[19]提出一种采用双金字塔网络合成磁共振成像 (magnetic resonance imaging, MRI) 辅助 CT 图像分割的方法; Nie 等^[20]将多流技术应用于 FCN, 实现婴儿脑组织图像的分割 (白质、灰质和脑脊液).

Ronneberger 等^[4]在 FCN 基础上提出 U-Net 架构, 是医学图像分割的经典架构之一. U-Net 包含一个编码路径和一个解码路径, 并且在等分辨率层之间使用跳跃连接, 它的 U 型结构解决了 FCN 无法获得上下文信息和位置信息的缺点. Milletari 等^[21]引入了残差操作, 提出 V-Net, 并且使用 Dice 损失函数代替二分类交叉熵损失, 解决了类别不平衡问题; Oktay 等^[22]将注意力机制用于 U-Net, 提出 Attention U-Net 模型, 在等分辨率的编码器和解码器中间拼接一种新型的注意力门控模块. 为了减少编码器和解码器的特征图之间的语义差距, Zhou 等^[23]提出 U-Net++ 模型, 在编码器和解码器之间嵌入密集的跳跃连接; Huang 等^[24]在 U-Net 和 U-Net++ 基础上继续改进, 提出 UNet 3+, 采用全尺寸跳跃连接和深度监督最大限度地利用全尺度的特征图; Chen 等^[25]首次将 Transformer 和 U-Net 结合, 提出 TransUNet, 进一步提升医学图像分割的性能. 然而, 利用密集的跳跃连接和 Transformer

极大地增加了 U-Net 的参数量, 不利于网络在轻量化设备上的应用^[26].

随着 GAN^[27]在多个图像分割任务上大获成功, 研究人员将其用于医学图像分割领域. Xue 等^[28]提出一个用于图像分割任务的端到端对抗网络架构 SegAN, 利用一个多尺度 L_1 损失函数来克服分割结果和真实标签不同尺度的特征差异; Zhao 等^[29]提出 Deep-sup GAN 架构, 将原始的 MRI 图像和生成的 CT 图像结合来完成分割任务; Dai 等^[30]提出结构校正对抗网络 (structure correcting adversarial network, SCAN), 使用逐像素损失函数进行预训练, 解决了 GAN 的不稳定性问题; Dong 等^[31]深度结合 U-Net 和 FCN, 将前者作为生成器、后者作为判别器进行对抗学习; Huo 等^[32]提出 SSNet (splenomegaly segmentation network), 将 FCN 和条件 GAN (conditional GAN, CGAN) 结合, 提高系统对器官大小和形状分割的鲁棒性.

2 DRFN 模型

2.1 整体结构

DRFN 模型在 U-Net 的基础上, 通过构建 FPCI 模块和 LVA 模块提升卷积操作获取局部语义信息的能力; 同时, 将更多的信息在编码-解码的过程中保留下来, 不仅充分利用 U-Net 高效的优点, 还在仅使用少量参数的情况下提升了网络分割效果.

DRFN 模型的整体结构如图 1 所示, 分为编码阶段和解码阶段. 与 U-Net 一致, 为了保证解码阶段的效果, 每层的编码器输出与解码器输入都通

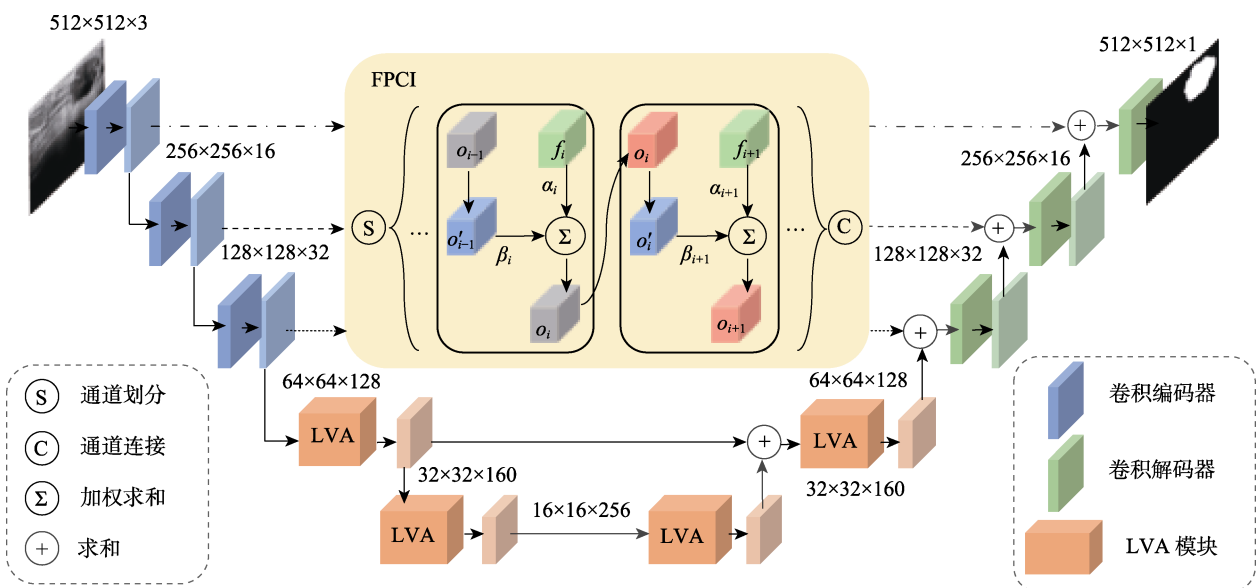


图 1 DRFN 模型示意图

过求和的方式连接. 其中, 编码阶段和解码阶段结构相似, 都包含 5 层网络模块. (1) 编码阶段的前 3 层中, 每层利用卷积、批归一化、ReLU 激活函数和池化操作构成一个卷积编码器, 每个编码器的为了提升 DRFN 模型的效率, 从上到下通道数分别为 32, 64, 128, 160 和 256, 比现有大多数基于 U-Net 的模型^[5]的参数量更小. 输出分为 2 个走向: 一是作为下一层编码器的输入; 二是作为 FPCI 模块的输入, 实现不同尺度局部信息的提取, 进而将输出结果与对应的解码器输入相融合. (2) 编码阶段的最后 2 层和解码阶段的前 2 层则由 LVA 模块构成, 通过对卷积核局部视野的调整实现对图像语义信息的提取. (3) 解码阶段的最后 3 层由与编码阶段前 3 层结构相呼应的解码器组成, 为了降低模型的参数量, 解码阶段利用双线性插值实现上采样, 在保留特征图语义信息的同时进一步提升模型的效率.

2.2 FPCI 模块

为了提升模型对医学图像中语义信息的提取能力, 更好地实现图像分割的目标. 本文设计 FPCI 模块来提取不同尺度感受野下特征图的语义信息, 将提取的语义信息传递给解码模块, 实现语义信息的保留.

对于从编码阶段前 3 层编码器输出的特征图, 首先使用一个 1×1 的卷积核将输入特征 $f_{enc} \in \mathbb{R}^{c \times h \times w}$ 映射到一个新的特征空间中并改变其通道数, 将新的特征表示为 $f_{local} \in \mathbb{R}^{\hat{c} \times h \times w}$; 其中, $\hat{c}$ 表示预设的映射后特征的通道数. 在此基础上, 将 f_{local} 沿通道维度分割为 m 个子部分, 分别记作 $f_1, f_2, \dots, f_m$, 即 $f_1, f_2, \dots, f_m = \text{Conv}_{in}(f_{enc})$.

至此, 新特征空间中的特征 $f_{local} \in \mathbb{R}^{\hat{c} \times h \times w}$ 已被分解为 m 个子部分. 为了克服卷积操作中因卷积核大小固定带来的感受野大小受限的问题, FPCI 模块利用多尺度卷积单元(multi-scale convolutional unit, MCU)将 m 个子特征级联聚合, 输出 m 个聚合特征, 记作 $o_1, o_2, \dots, o_m$. 此后, 沿通道维度连接所输出的聚合特征, 并通过一个 1×1 的卷积核将连接后的聚合特征重组, 重新映射到输出空间, 得到级联聚合模块的最终输出 f_{skip} , 这一过程可以被描述为

$$f_{skip} = \text{Conv}_{out}(\text{Concat}(o_1, o_2, \dots, o_m)).$$

本文设计的 FPCI 的结构如图 2 所示. 给定第 i

组分解后的特征 f_i , 第 $i-1$ 级的输出特征 o_{i-1} 将经过一层 3×3 的卷积得到 o'_{i-1} , 并与输入特征 f_i 一起按比例加权求和得到本级的输出特征 o_i , 公式为

$$o_i = \sigma(\alpha_i) \cdot f_i + \sigma(\beta_i) \cdot \text{Conv}_i(o_{i-1}).$$

其中, 为了让各级运算中来自前级的累积特征和来自本级的输入特征的加权比例更具有适应性, 引入一组可学习的参数 α_i 和 β_i 作为各级运算中的权重; $\sigma(\cdot)$ 表示 Sigmoid 运算, 保证限制加权重为 (0,1).

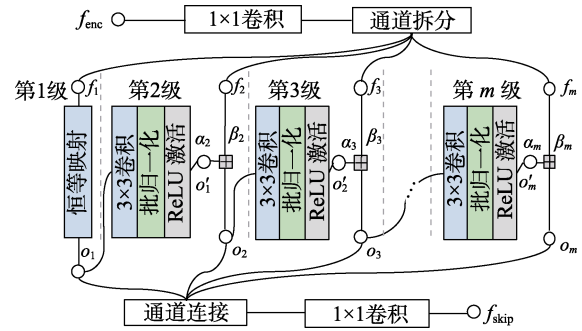


图 2 FPCI 模块结构

考虑到第 1 级运算时没有前级输出结果, 故将第 1 级的输入特征进行恒等映射后直接作为第 1 级的输出, 即 $o_1 = f_1$. 在此基础上, 即可驱动 MCU 逐级运算每级的输出特征. 由于每级的运算都有来自前级运算结果的参与, 在卷积层的作用下, o'_{i-1} 提供的感受野将逐级递增. 因此, 在输出的聚合特征 $o_1, o_2, \dots, o_m$ 中, 其所对应的感受野也将从 o_1 的只包括每个元素本身向其邻域不断扩张, 经过 MCU 运算并连接重组的输出特征 f_{skip} 能够拥有不同尺度邻域范围内的上下文信息, 进一步强化了多层次视觉特征的融合能力. 随后, 将 MCU 的输出与解码阶段对应层的解码器的输入相加, 将编码阶段图像多尺度的视觉语义信息传递到解码阶段, 克服基于 U-Net 的方法存在的解码阶段语义信息不足的问题.

FPCI 模块在丰富模型提取的上下文视觉语义信息时, 比传统的卷积模块使用的参数更少. 对于输入特征 $f_{enc} \in \mathbb{R}^{c \times h \times w}$, FPCI 模块先将其解耦为 m 份通道数为 $\frac{c}{m}$ 的子特征以进行级联聚合; 级联聚合的第 1 阶段为一个恒等映射, 无需参数, 后续的 $m-1$ 个聚合阶段中, 每阶段的参数量均为 $\frac{c}{m} \times \frac{c}{m} \times 3 \times 3 + 2$. 其中, $\frac{c}{m} \times \frac{c}{m} \times 3 \times 3$ 表示卷积参数

量, $\frac{c}{m} \times \frac{c}{m} \times 2$ 表示自适应加权重 α 和 β 所带来的参数量. 因此, FPCI 模块中总的参数量为 $\left(\frac{c}{m} \times \frac{c}{m} \times 3 \times 3 + 2\right) \cdot (m-1)$, 对于高通道特征 f_{enc} , 与一层普通 3×3 卷积相比, FPCI 以其约 $\frac{1}{m}$ 的参数量获得了更大的感受野, 以及更丰富的上下文语义信息.

2.3 LVA 模块

在编码器的最后 2 层, 构建 LVA 模块动态地调整卷积神经网络卷积核感受野在不同通道特征图上对应的区域, 增强模型对局部语义信息提取的能力^[33].

给定输入特征, 首先利用 3×3 的卷积核, 以 2 为步长压缩输入特征的特征图尺寸, 同时将特征图的通道数变为 T , 将这一特征表示为 $X \in \mathbb{R}^{\tilde{h} \times \tilde{w} \times T}$. 与 FPCI 模块采用的通道分割方案类似, LVA 模块将 T 个通道分成 n 份, 则 $X = [x_1, x_2, \dots, x_i, \dots, x_n]$, 再通过对 x_i 进行不同尺度的处理丰富模型获得的局部语义信息. 为了充分考虑每个局部卷积提取的感受野内图像的上下文语义信息, LVA 模块利用横向平移、竖向平移和旋转改变 x_i 上特征图的分布. 以横向平移为例, 对 X 包含的每个特征图集合横向平移不同的长度, 可用 $S_w(x_i, w_i)$ 表示, 其中, w_i 表示横向平移的距离, 则经过横向平移后 $X' = S_w(X, w)$; 同理, 竖向平移和旋转可分别表示为 $S_h(x_i, h_i)$ 和 $R(x_i, \theta_i)$, 其中, h_i 和 θ_i 分别表示竖向平移的长度和旋转的角度. 对于平移或旋转后的特征图, 为了进一步挖掘不同平移长度和旋转角度下的语义信息, 本文在通道上对特征图进行重组, 利用一个基本多重感知机嵌入层(basic multilayer perceptron embedding layer, BME)来实现这一目标, 表示为 $BME(\cdot) = \text{GeLU}(\text{MLP}(\cdot))$, 其中, $\text{MLP}(\cdot)$ 为通道上的多层感知机(multi-layer perceptron, MLP), $\text{GeLU}(\cdot)$ 为激活函数. 为了保证模型的效率, 利用深度可分离卷积(depth wise convolutional layer, DWConv)进一步处理局部信息重组后的特征. 除了参数量少, DWConv 还能有效地保持特征的位置信息^[26]. 在 LVA 模块的最后, 利用层归一化处理不同平移距离和旋转角度的特征图, 保持整个特征图信息的一致性. 本文设计的 LVA 模块的结构如图 3 所示.

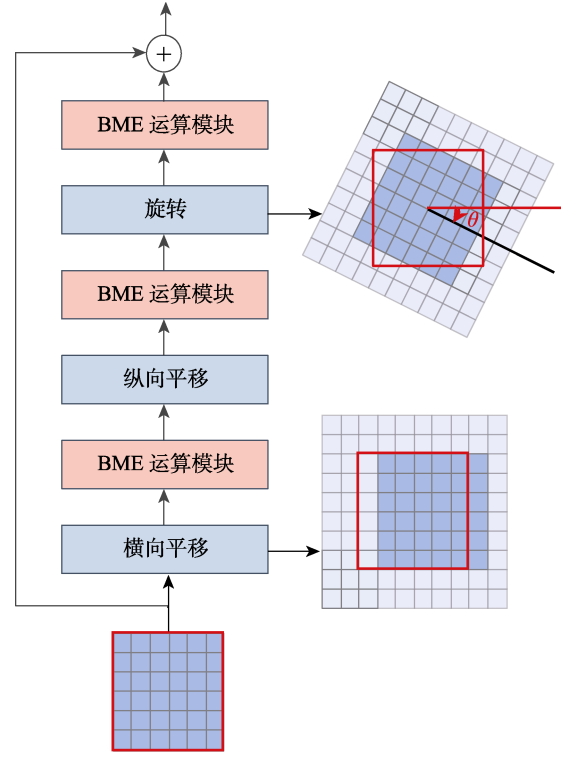


图3 LVA 模块结构

LVA 模块对输入的操作步骤如下:

Step1. 横向的位移和处理, 公式为

$$X_{\text{shift}}^W = \text{DWConv}\left(\text{BME}\left(S_w(X, w)\right)\right).$$

Step2. 竖向的位移及特征处理, 公式为

$$X_{\text{shift}}^H = \text{DWConv}\left(\text{BME}\left(S_h(X_{\text{shift}}^W, h)\right)\right).$$

Step3. 进行旋转处理, 公式为

$$X_{\text{rotate}} = \text{DWConv}\left(\text{BME}\left(R(X_{\text{shift}}^H, \theta)\right)\right).$$

Step4. 将输入和 Step1~Step3 的操作结果进行加和并做层归一化处理, 公式为

$$X_{\text{out}} = f\left(\text{LayerNorm}\left(X + X_{\text{rotate}}\right)\right);$$

其中, $f(\cdot)$ 表示一个线性映射.

2.4 损失函数

本文模型的损失函数为 $L = \lambda \cdot L_{\text{BCE}} + L_{\text{Dice}}$. 其中, λ 表示参数, 与现有大多数基于 U-Net 框架的方法一致, 可令其值为 0.5^[26]; 二值交叉熵损失 L_{BCE} 和 Dice 损失 L_{Dice} 分别为

$$L_{\text{BCE}} = -\sum_i^N y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log(1 - \hat{y}_i),$$

$$L_{\text{Dice}} = 1 - \frac{2 \cdot \sum_i^N \hat{y}_i \cdot y_i}{\sum_i^N \hat{y}_i^2 + \sum_i^N y_i^2};$$

$\hat{y}_i$ 表示模型推理结果, y_i 表示图像分割结果的真值.

3 实验与结果分析

3.1 数据集及评价指标

在皮肤癌病灶分割数据集(2018)(简称 ISIC 2018 数据集)^[34]和乳腺癌病灶分割数据集^[35](简称 BUSI 数据集)这 2 个主流癌症数据集上, 评估 DRFN 模型的分割性能。

3.1.1 ISIC 2018 数据集

ISIC 2018 数据集是一个开放的黑色素瘤检测的皮肤病变数据集, 由国际皮肤成像合作组织(international skin imaging collaboration, ISIC)于 2018 年提供的世界皮肤病 AI 诊断模型挑战赛专用数据集。该数据集包含皮肤病学图像和相应的皮肤癌病变区域分割图; 由 2 594 幅图像组成, 与现有主流医学图像分割方法评价方式^[36]一样, 每次随机抽取 2 074 幅图像进行训练, 520 幅图像用于测试。鉴于 ISIC 2018 数据集图像尺寸差异较大, 训练前需将图像的分辨率统一调整为 512×512 像素。

3.1.2 BUSI 数据集

BUSI 数据集即乳腺超声图像(breast ultrasound images)数据集, 是一个乳腺癌病灶分割数据集。该数据集是一个开放数据集, 其中共有来自 600 名女性患者(25~75 岁)的 780 幅乳腺超声图像, 包括 133 幅无肿块的正常图像、437 幅癌症肿块的恶性图像和 210 幅良性肿块的图像。与文献[26]中的设置一样, 本文使用 647 幅良性和恶性图像组成的集合来验证 DRFN 模型, 随机选择 80%作为训练集, 剩余 20%作为测试集进行实验。为了便于在 BUSI 数据集上进行评估, 训练前需将图像的分辨率统一调整为 256×256 像素。

3.1.3 评价指标

为了评估 DRFN 模型的有效性, 以交并比(intersection over union, IoU)和 Dice 系数作为指标, 定量分析所提图像分割方法的性能。从数据集和模型推理结果, 可以得到测试集上的预测结果(predict result, PR)与真实标签(ground truth, GT), 并计算得到真阳性(true positive, TP)、真阴性(true negative, TN)、假阳性(false positive, FP)和假阴性(false negative, FN)指标。IoU 指预测样本和实际样本的交并比, 计算公式为

$$\text{IoU} = \frac{\text{PR} \cap \text{GT}}{\text{PR} \cup \text{GT}} = \frac{\text{TP}}{\text{FP} + \text{TP} + \text{FN}}.$$

Dice 系数是一种评估相似度的函数, 用于计算预测结果和真实结果之间的相似度, 其计算结果与常用的 F_1 -score 评价指标一致。Dice 系数 D 的

$$\text{计算公式为 } D = \frac{2|\text{PR} \cap \text{GT}|}{|\text{PR}| + |\text{GT}|} = \frac{2\text{TP}}{\text{FP} + 2\text{TP} + \text{FN}}.$$

为了更全面地评估模型的效果, 本文还对模型的效率进行评估, 以模型的大小和测试阶段推理单幅图像所需的时间作为对模型效率的评估指标。

3.2 实验设置

DRFN 模型基于 Pytorch 开发框架实现, 所有实验均在 NVIDIA GeForce RTX 3090 显卡上进行。使用学习率为 1×10^{-4} 和权值衰减为 1×10^{-4} 的 Adam 优化器训练模型; 还使用余弦退火算法调整学习率, 其中, 最小学习率为 1×10^{-5} , 批处理大小都设置为 8, 迭代 300 次后终止训练。随机选择 LVA 模块的值, 利用平移和旋转参数的轻微扰动提升模型对边缘信息的提取能力; 选择范围不宜过大, 避免由于引入噪声引起效果的下降^[33]。在 FPCI 模块中, 将通道划分为 4 份; 在 LVA 模块中, 将通道划分为 5 份。此外, 在 LVA 模块中, 不同通道先在水平方向位移 $[-2, -1, 0, 1, 2]$, 接着在垂直方向上位移 $[-2, -1, 0, 1, 2]$, 在此基础上, 旋转的角度为 $[-10^\circ, -5^\circ, 0^\circ, 5^\circ, 10^\circ]$ 。为了保证对实验结果评估的客观性, 本文在每个数据集上对训练集和测试集进行 4 次随机拆分, 并记录实验结果的均值和标准差来综合评价 DRFN 模型的性能。

3.3 对比实验

将 DRFN 模型与当前广泛使用的医学图像分割方法进行了比较, 包括基于 U-Net 及其改进模型的 U-Net^[4], U-Net++^[23], ResU-Net^[37]和 UNeXt^[26], 基于全卷积结构的 MedT 模型^[38], 以及基于 U-Net 和 Transformer 结合的 TransU-Net 模型^[25]。

3.3.1 ISIC 2018 数据集

由于 ISIC 2018 数据集由彩色图像构成, 因此其中包含的语义信息较为丰富; 同时, 皮肤癌区域和周围正常皮肤的对比较为鲜明, 丰富的语义信息有利于图像的精准分割。尽管皮肤癌图像边界清晰, 但边界细微形状往往没有固定规律, 这也给图像的分割带来挑战。Dice 指标一般倾向于衡量模型的平均性能, 而 IoU 指标倾向于衡量模型最坏的表现。因此, 一个好的模型不仅要在容易识别的样本上取得良好表现, 也要克服困难样本带来的划分挑战。DRFN 模型通过构建 FPCI 模块和 LVA 模块, 提升卷积操作局部视野下对图像上下文语义信息的捕捉能力, 克服基于 U-Net 的方法在降采样过程中带来的信息丢失。表 1 所示为 7 种模型在 2 个数据集上的分割效果对比。可以看出, 所有模型

在 ISIC 2018 数据集上的 Dice 值都比较高; DRFN 模型不仅在 Dice 指标上取得了最优的效果, 而且显著地提升了 IoU 指标; 而 MedT 模型和 TransU-Net 模

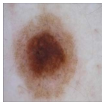
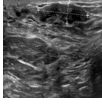
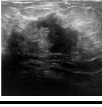
型则对困难样本考虑不足. 实验结果表明, DRFN 模型能够较好地保留复杂对象的局部语义, 其分割效果可视化对比如表 2 所示.

表 1 7 种模型在 2 个数据集上分割效果对比

模型	参数量	推断时间/ms	ISIC 2018 数据集		BUSI 数据集	
			IoU	Dice 系数	IoU	Dice 系数
U-Net ^[4]	31.13M	223	74.55±0.98	84.03±0.93	63.85±1.06	76.35±0.94
U-Net++ ^[23]	9.16M	173	75.12±0.81	84.96±0.84	64.33±0.87	77.54±0.86
ResU-Net ^[37]	62.74M	333	75.62±1.05	85.60±0.82	64.89±0.91	78.25±0.86
MedT ^[38]	1.60M	751	79.54±0.51	87.35±0.42	63.89±0.74	76.93±0.33
TransU-Net ^[25]	105.32M	246	80.51±0.85	88.91±0.79	66.92±0.87	79.30±0.61
UNeXt ^[26]	1.47M	25	81.70±1.24	89.70±0.98	66.95±1.10	79.37±0.75
DRFN	1.67M	33	83.92±0.26	91.09±0.23	70.45±1.70	83.39±1.15

注: 粗体表示最优值.

表 2 不同模型在 2 个数据集上的分割效果可视化示意图

数据集	输入图像	GT	U-Net ^[4]	U-Net++ ^[23]	ResU-Net ^[37]	MedT ^[38]	TransU-Net ^[25]	UNeXt ^[26]	DRFN
ISIC 2018									
									
BUSI									
									

3.3.2 BUSI 数据集

BUSI 数据集由分辨率较低的灰度图构成. 与 ISIC 2018 数据集相比, BUSI 数据集包含的语义信息没有那么丰富; 此外, 乳腺结构相对复杂, 乳腺癌区域和周围正常区域区分不明显, 边界模糊. 与 ISIC 2018 数据集相比, BUSI 数据集存在训练数据少、图片分辨率低、结构复杂, 且只有灰度图等挑战. 因此, 无论 IoU 指标和 Dice 指标, 对 BUSI 数据集的识别均比 ISIC 2018 数据集上的低. DRFN 模型通过构建 FPCI 模块提取不同尺度的局部语义信息, 并将编码块与解码块直接连接, 提升基于 U-Net 的重构能力. FPCI 模块和 LVA 模块能够充分利用图像的低级特征和高级语义信息, 显著地提高了基于 U-Net 的方法在边界分割能力. 从表 1 可以看出, DRFN 在 IoU 指标和 Dice 指标上均获得了显著的提升. 从表 2 可以看出, DRFN 模型在医学图像分割上表现更好, 分割边界更加清晰.

TransU-Net 模型参数多, 推理速度受限, 而 DRFN 模型则是一个兼具轻量化和分割能力的模型. 从表 1 可以看出, DRFN 模型的大小小于 TransU-Net 模型大小的 2%, 推理速度也明显快于现有大部分对比方法; 与基于 U-Net 的改进方法相比, DRFN 模型在有效地降低模型复杂度的同时, 提升了模型的分割效果, 有利于推动深度学习技术在计算机辅助医疗图像分割领域的应用.

3.4 消融实验

为了验证本文提出的 FPCI 模块和 LVA 模块捕获图像语义信息的能力及其对图像分割的作用, 在 2 个数据集上进行消融实验. 以 U-Net 基本网络架构为基准, 按照本文设置的通道层数和特征图尺寸进行调整, 验证 FPCI 模块和 LVA 模块对分割效果的影响. LVA 模块通过调整局部感受野不同通道感受到的视野范围, 提升卷积提取图像局部语义信息的能力, 如表 3 所示. 可以看出, LVA 模块

能够显著地提升 U-Net 的表现; FPCI 模块通过通道递进级联实现对固定尺寸卷积核感受野的扩大, 能有效地减少 U-Net 编码解码过程中语义信息的丢失. 在 BUSI 数据集上, DRFN 模型对 IoU 提升效果较好. 实验结果表明, 在 LVA 和 FPCI 模块的共同作用下, DRFN 模型不仅能够提升模型分割效果, 而且对于复杂困难样本的提升效果较好.

表 3 DRFN 模型各模块对分割效果影响对比

模型	ISIC 2018 数据集		BUSI 数据集	
	IoU	Dice 系数	IoU	Dice 系数
U-Net	74.32±1.04	83.98±0.89	63.60±1.15	75.95±0.95
U-Net+LVA	83.26±0.58	90.71±0.34	69.07±0.53	81.11±0.30
U-Net+FPCI	82.92±0.72	90.34±0.69	70.10±0.75	81.95±0.74
U-Net+LVA+FPCI	83.92±0.26	91.09±0.23	70.45±1.70	83.39±1.15

4 结 语

针对基于 U-Net 的方法存在的下采样阶段信息丢失和因固定尺寸卷积核带来的多尺度局部语义信息提取不足的问题, 本文提出一种 DRFN 模型. 首先构建 FPCI 模块, 获取编码器多尺度局部语义特征并将其赋值给解码器, 克服下采样阶段信息的丢失; 然后构建 LVA 模块, 通过对特征图进行平移和旋转操作增强固定尺寸卷积核, 提取视野上下文语义信息的能力. 实验结果表明, DRFN 模型在主流癌症图像分割上取得了显著提升. 由于医学图像种类纷繁, 获取难度大, 未来将不断扩大对不同场景下医疗图像分割问题的研究.

参考文献(References):

- [1] Sharma N, Aggarwal L M. Automated medical image segmentation techniques[J]. *Journal of Medical Physics*, 2010, 35(1): 3-14
- [2] Grau V, Mewes A U J, Alcaniz M, *et al.* Improved watershed transform for medical image segmentation using prior information[J]. *IEEE Transactions on Medical Imaging*, 2004, 23(4): 447-458
- [3] Hesamian M H, Jia W J, He X J, *et al.* Deep learning techniques for medical image segmentation: achievements and challenges[J]. *Journal of Digital Imaging*, 2019, 32(4): 582-596
- [4] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C] // *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Heidelberg: Springer, 2015: 234-241
- [5] Siddique N, Paheding S, Elkin C P, *et al.* U-net and its variants for medical image segmentation: a review of theory and applications[J]. *IEEE Access*, 2021, 9: 82031-82057
- [6] Çiçek Ö, Abdulkadir A, Lienkamp S S, *et al.* 3D U-Net: learning dense volumetric segmentation from sparse annotation[C] // *Proceedings of the 19th International Conference on Medical Image Computing and Computer-assisted Intervention*. Heidelberg: Springer, 2016: 424-432
- [7] Yen J C, Chang F J, Chang S. A new criterion for automatic multilevel thresholding[J]. *IEEE Transactions on Image Processing*, 1995, 4(3): 370-378
- [8] Lee C, Huh S, Ketter T A, *et al.* Unsupervised connectivity-based thresholding segmentation of midsagittal brain MR images[J]. *Computers in Biology and Medicine*, 1998, 28(3): 309-338
- [9] Gong M G, Liang Y, Shi J, *et al.* Fuzzy C-means clustering with local information and kernel metric for image segmentation[J]. *IEEE Transactions on Image Processing*, 2013, 22(2): 573-584
- [10] Nalwa V S, Binford T O. On detecting edges[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, PAMI-8(6): 699-714
- [11] Wan S Y, Higgins W E. Symmetric region growing[J]. *IEEE Transactions on Image Processing*, 2003, 12(9): 1007-1015
- [12] Manousakas I N, Undrill P E, Cameron G G, *et al.* Split-and-merge segmentation of magnetic resonance medical images: performance evaluation and extension to three dimensions[J]. *Computers and Biomedical Research*, 1998, 31(6): 393-412
- [13] Rother C, Kolmogorov V, Blake A. "GrabCut": interactive foreground extraction using iterated graph cuts[J]. *ACM Transactions on Graphics*, 2004, 23(3): 309-314
- [14] Ünver H M, Ayan E. Skin lesion segmentation in dermoscopic images with combination of YOLO and GrabCut algorithm[J]. *Diagnostics*, 2019, 9(3): Article No.72
- [15] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2015: 3431-3440
- [16] Liu W, Rabinovich A, Berg A C. ParseNet: looking wider to see better[OL]. [2022-09-22]. <https://arxiv.org/abs/1506.04579>
- [17] Roth H R, Oda H, Zhou X R, *et al.* An application of cascaded 3D fully convolutional networks for medical image segmentation[J]. *Computerized Medical Imaging and Graphics*, 2018, 66: 90-99
- [18] Christ P F, Elshaer M E A, Ettlinger F, *et al.* Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields[C] // *Proceedings of the 19th International Conference on Medical Image Computing and Computer-assisted Intervention*. Heidelberg: Springer, 2016: 415-423
- [19] Liu Y Z, Lei Y, Fu Y B, *et al.* Head and neck multi-organ auto-segmentation on CT images aided by synthetic MRI[J]. *Medical Physics*, 2020, 47(9): 4294-4302
- [20] Nie D, Wang L, Gao Y Z, *et al.* Fully convolutional networks for multi-modality isointense infant brain image segmentation[C] // *Proceedings of the 13th International Symposium on Biomedical Imaging*. Los Alamitos: IEEE Computer Society Press, 2016: 1342-1345
- [21] Milletari F, Navab N, Ahmadi S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C] // *Proceedings of the 4th International Conference on*

- 3D Vision. Los Alamitos: IEEE Computer Society Press, 2016: 565-571
- [22] Oktay O, Schlemper J, Le Folgoc L, *et al.* Attention U-Net: learning where to look for the pancreas[OL]. [2022-09-22]. <https://arxiv.org/abs/1804.03999>
- [23] Zhou Z W, Siddiquee M M R, Tajbakhsh N, *et al.* U-Net++: redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867
- [24] Huang H M, Lin L F, Tong R F, *et al.* UNet 3+: a full-scale connected UNet for medical image segmentation[C] //Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Los Alamitos: IEEE Computer Society Press, 2020: 1055-1059
- [25] Chen J N, Lu Y Y, Yu Q H, *et al.* TransUNet: transformers make strong encoders for medical image segmentation[OL]. [2022-09-22]. <https://arxiv.org/abs/2102.04306>
- [26] Valanarasu J M J, Patel V M. UNeXt: MLP-based rapid medical image segmentation network[C] //Proceedings of the 25th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2022: 23-33
- [27] Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144
- [28] Xue Y, Xu T, Zhang H, *et al.* SegAN: adversarial network with multi-scale L_1 loss for medical image segmentation[J]. Neuroinformatics, 2018, 16(3): 383-392
- [29] Zhao M Y, Wang L, Chen J W, *et al.* Craniomaxillofacial bony structures segmentation from MRI with deep-supervision adversarial learning[C] //Proceedings of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2018: 720-727
- [30] Dai W, Dong N Q, Wang Z Y, *et al.* Scan: structure correcting adversarial network for organ segmentation in chest X-Rays[C] //Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Heidelberg: Springer, 2018: 263-273
- [31] Dong X, Lei Y, Wang T H, *et al.* Automatic multiorgan segmentation in thorax CT images using U-net-GAN[J]. Medical Physics, 2019, 46(5): 2157-2168
- [32] Huo Y K, Xu Z B, Bao S X, *et al.* Splenomegaly segmentation using global convolutional kernels and conditional generative adversarial networks[C] //Proceedings of the Medical Imaging: Image Processing. Bellingham: Society of Photo-Optical Instrumentation Engineers, 2018, 10574: Article No.1057409
- [33] Cao H, Wang Y Y, Chen J, *et al.* Swin-UNet: U-Net-like pure transformer for medical image segmentation[C] //Proceedings of the International Conference on Computer Vision. Heidelberg: Springer, 2023: 205-218
- [34] Codella N, Rotemberg V, Tschandl P, *et al.* Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC)[OL]. [2022-09-22]. <https://arxiv.org/abs/1902.03368>
- [35] Al-Dhabyani W, Gomaa M, Khaled H, *et al.* Dataset of breast ultrasound images[J]. Data in Brief, 2020, 28: Article No. 104863
- [36] Azad R, Bozorgpour I A, Asadi-Aghbolaghi M, *et al.* Deep frequency re-calibration u-net for medical image segmentation[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 3267-3276
- [37] Zhang Z X, Liu Q J, Wang Y H. Road extraction by deep residual U-Net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753
- [38] Valanarasu J M J, Oza P, Hacihaliloglu I, *et al.* Medical transformer: gated axial-attention for medical image segmentation[C] //Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2021: 36-46