

# 基于2.5D网络和尺度注意力感知的肝脏与肿瘤分割

李家健, 黄国恒

广东工业大学计算机学院, 广东 广州

收稿日期: 2021年12月25日; 录用日期: 2022年1月21日; 发布日期: 2022年1月28日

## 摘要

深度学习技术已被广泛应用在肝脏与肿瘤的分割任务中。但是, 现有的3D网络模型大都忽略了医学图像横断面的像素距离过大的问题, 直接使用3D卷积操作难以学习到准确的三维空间信息。此外, 肿瘤形状大小高度可变的特点使得分割肿瘤更具挑战性。针对第一个问题, 本文提出改进后的2.5D ResNet34对肝脏和肿瘤特征进行编码, 提高模型对三维空间信息的建模能力。同时, 利用DSC损失函数来提高模型对整体结构的分割能力。针对第二个问题, 本文提出尺度注意力感知模块, 通过建模不同尺度特征下的局部和全局三维空间信息, 以有效地整合低级上下文信息和高级区域语义信息, 从而实现精准的肝脏和肿瘤分割。本文所提出的方法在Liver Tumor Segmentation (LiTS)数据集上测试了肝脏与肿瘤的分割性能, 其中肝脏分割的DSC为96.4%, 肿瘤分割的DSC为72.3%, 并与近三年的模型相比, 本文提出的方法在肝脏和肿瘤分割中表现最好。

## 关键词

2.5D网络, 尺度注意力感知, 语义分割, CT图像

## A 2.5D Network and Scale Attention Awareness for Liver and Tumor Segmentation

Jiajian Li, Guoheng Huang

School of Computer, Guangdong University of Technology, Guangzhou Guangdong

Received: Dec. 25<sup>th</sup>, 2021; accepted: Jan. 21<sup>st</sup>, 2022; published: Jan. 28<sup>th</sup>, 2022

## Abstract

Deep learning technology has been widely used in the segmentation task of liver and tumor. How-

ever, the existing 3D network models mostly ignore the problem of large pixel distances across medical images, and it is difficult to learn accurate three-dimensional spatial information directly using 3D convolution operations. In addition, the highly variable shape and size of tumors make segmenting tumors more challenging. Aiming at the first problem, this paper proposes an improved 2.5D ResNet34 to encode liver and tumor features, and improve the modeling ability to model three-dimensional spatial information. At the same time, the DSC loss function is utilized to improve the modeling ability to segment the overall structure. In response to the second problem, this paper proposes a scale attention awareness module, which can effectively integrate low-level context information and high-level regional semantic information by modeling local and global three-dimensional spatial information under different scale features, thereby achieving accurate segmentation of liver and tumor. The proposed method tested the performance on the Liver Tumor Segmentation (LiTS) dataset. The DSC of liver segmentation was 96.4%, and the DSC of tumor segmentation was 72.3%. Moreover, compared with the methods in the past three years, the proposed method performs best in liver and tumor segmentation.

## Keywords

2.5D Network, Scale Attention Awareness, Semantic Segmentation, CT Image

Copyright © 2022 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

肝脏是构成人体生命所需的重要器官, 肝癌是世界上最常见的癌症之一。根据世界卫生组织发布的最新数据显示, 每年都会有大量的病人因为肝癌疾病而死亡[1]。在医学上, 筛查肝脏疾病的主要手段是通过计算机断层扫描(Computed Tomography, CT)形成具有多个连续的层面图像, 然后由放射科医生手工或半手工地来勾画出肝脏或病变区域, 以作进一步的诊断及治疗计划[2]。但是, 手工或半手工地来勾画出肝脏或病变区域是非常耗时的, 并且带有强烈的主观意志, 容易出错。因此, 肝脏与肿瘤的自动分割技术成为当前一个热点研究课题。

在 CT 图像中, 肝脏与周围器官及组织之间的对比度非常低, 从 CT 中自动、鲁棒地分割肝脏与肿瘤是一项非常具有挑战性的任务。而与肝脏分割相比, 肿瘤具有形状和大小高度可变的特点, 因此, 肿瘤分割被认为是一项更具挑战性的任务。

近年来, 深度学习技术已被广泛应用在医学图像的分割任务中, 越来越多的研究者在肝脏与肿瘤分割任务中使用了深度学习方法[3] [4] [5]。目前, 最先进的肝脏与肿瘤分割方法大多是基于全卷积神经网络(Fully Convolutional Networks, FCN) [6]。FCN 的核心主要是基于编码 - 解码的结构, 对图像进行像素级分类, 其可以输入任意大小的图像并产生相同大小的输出。基于 FCN, Ronneberger 等人[7]提出一个完全对称的 U 形网络, 称为 UNet, 其主要创新是每一次上采样都与对应的特征提取部分进行融合, 以弥补更多的低级语义信息, 并且只需用少量的数据进行训练就能获得比较精确的结果。但是, 该网络的缺点是, 下采样操作会丢失大量的语义信息。为解决这个问题, Liu 等人[8]在 UNet 的基础上, 设计了适用于三维医学图像分割的网络, 称为 VNet, 引入了 ResNet 的残差连接思想[9], 从编码器提取肝脏特征, 解码器产生全分辨率输出。该方法虽然提高了分割精度, 但由于三维卷积的参数量大, 需要更长的训练时间。为此, Li 等人[10]提出 H-DenseUNet 来缓解这个问题, 其思路是混合 2D DenseUNet 与 3D DenseUNet

的特征来加速 3D DenseUNet 的收敛, 同时在 UNet 中嵌入密集连接块, 能进一步提升肝脏与肿瘤的分割精度。进一步地, Han 等人[11]提出以多个 2D 信息堆叠成 3D 信息进行肝脏与肿瘤的分割, 该方法基于 UNet 的跳跃连接和 ResNet 的残差连接, 然后通过输入多个连续的单张切片来提供 3D 上下文信息。但是, 这种方法得到的 3D 上下文信息是有限的。此外, 为了分割多尺度的肿瘤特征, Jin 等人[12]提出 RA-UNet 对肝脏和肿瘤进行分割, 其以 3D UNet 为基准网络, 在网络的深层引入了注意力机制来实现肝脏与肿瘤的分割, 虽然能在一定程度上提升分割精度, 但是由于深层的特征越来越抽象, 语义信息非常少, 因此提升是非常有限。相似地, Chen 等人[13]提出 Deeplabv3+模型。该模型是在空洞卷积和空间金字塔(ASPP)的基础上[14], 将深度可分卷积应用在 ASPP 和解码模块中, 以此来学习更多的尺度特征。该方法在肿瘤的分割效果有进一步的提升, 但在肝脏分割中, 提升不明显。为此, Tang 等人[15]提出两阶段的分割模型, 首先用 Res2Net50 [16]作为基准粗分割出肝脏区域, 再通过粗分割出来的肝脏区域裁剪输入, 使用 E2Net 进行肝脏和肿瘤的精分割。在精分割阶段, 通过距离变换得到边缘增强监督信息, 从而提高分割精度。

总的来说, 以上方法大多都是基于 UNet 及 ResNet 的残差连接来实现肝脏与肿瘤的分割, 证明了 ResNet 具有较强的特征提取能力。受此启示, 本文直接采用 ResNet 作为特征提取网络。但是, 上述方法中都忽视了三维医学图像的像素距离信息。如果 CT 图像的横断面的像素距离过大, 直接采用 3D 卷积操作反而会带来很多干扰信息[17], 如图 1 所示, 相邻横断面(xy)之间的像素距离约为 6 毫米。此外, 上述方法难以拟合形状大小高度可变的肿瘤中的多尺度特征。针对第一个问题, 本文提出改进的 2.5D ResNet34 在网络浅层学习单张切片的特征, 在网络的深层学习整个图像的空间上下文信息, 提高模型对三维空间信息的建模能力。针对第二个问题, 本文提出尺度注意力感知模块, 通过建模不同尺度特征下的局部和全局三维空间信息, 以有效地整合低级上下文信息和高级区域语义信息, 从而实现精准的肝脏和肿瘤分割。

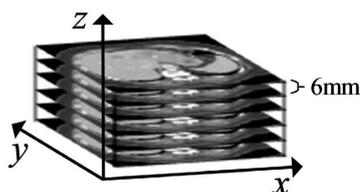


Figure 1. Example of pixel distance among cross-sections of 3D medical images  
图 1. 3D 医学图像横断面间像素距离示例图

## 2. 肝脏与肿瘤分割算法介绍

### 2.1. 整体结构

本文提出一种基于 2.5D 网络和尺度注意力感知模块的肝脏与肿瘤分割模型, 该模型由三部分组成: 1) 融合单张切片特征和相邻切片特征的 2.5D ResNet34 作为编码器; 2) 尺度注意力感知模块; 3) 多层解码单元组成的解码器, 如图 2 所示。其中, 与常用的 ResNet34 不同的是, 本文在 ResNet34 的前两层改用 2.5D 卷积操作, 在后三层采用正常的 3D 卷积操作, 以缓解三维医学图像普遍存在的像素距离过大的问题。经过编码部分对输入图像进行编码后, 产生四个不同尺度的特征向量  $F_s$ ,  $s = \{1, 2, 3, 4\}$ 。然后, 尺度注意力感知模块(SAA)用于捕获  $F_s$  的注意力特征, 并通过将高层级与底层级的注意力特征融合, 形成具有尺度感知的注意力特征  $F_s^A$ 。SAA 不仅可以缓解由于下采样操作导致肿瘤边界难以精准分割的问题, 还能解决由于肿瘤形状高度可变而导致的误分割问题。解码器旨在融合不同层级的  $F_s^A$ , 最终输出与输入图像大小相同的分割概率图。

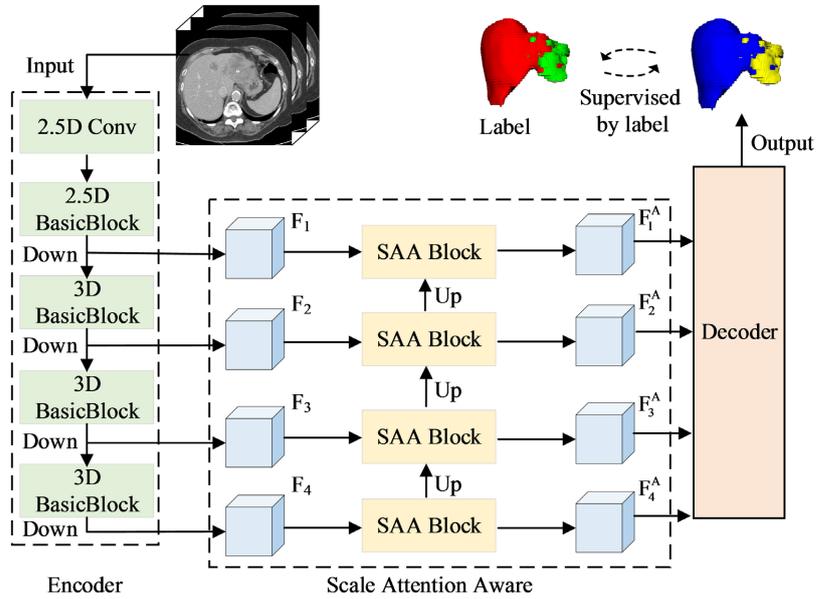


Figure 2. Overall structure of model  
图 2. 模型整体结构

## 2.2. 编码器

编码器使用 ResNet34 作为骨干网络，它具有较好的特征提取能力且参数量适中，非常适用于 3D 图像的语义分割。然而，直接将 3D 图像输入到 3DResNet34 中进行训练是不合理的。现有的神经网络无法学习到像素间的距离信息，如果像素距离过大，直接执行 3D 卷积操作反而难以学习三维的空间信息。

为缓解这一问题，提出改进后的 2.5D ResNet34 作为编码器，使编码器能在相邻切片像素距离过大的条件下，仍然能同时有效地学习单张切片特征和相邻切片特征。编码器的具体细节如图 3 所示，为有效提取单张切片的特征，将原  $3 \times 3 \times 3$  卷积核替换为  $1 \times 3 \times 3$  卷积核，称为 2.5D 方法。相应地，padding 设为 (0, 1, 1)。一般地，网络的浅层使用 2.5D 方法，网络的深层使用 3D 方法能更好地平衡模型对单张切片特征和相邻切片特征的学习能力。因此，提议的方法只在前两层使用 2.5D 方法，其他层使用 3D 方法。

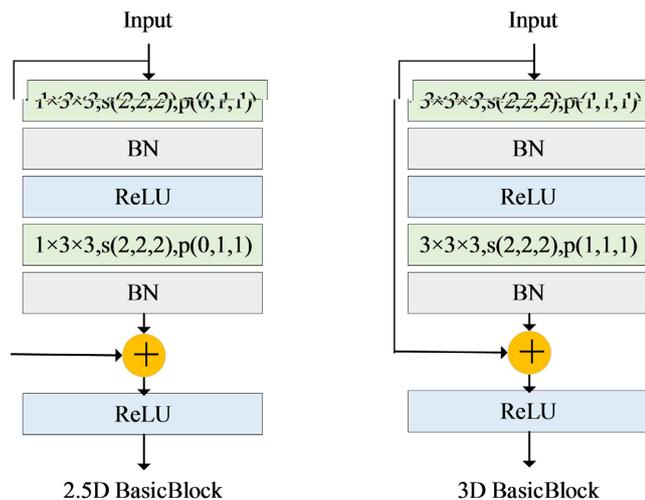


Figure 3. 2.5D basic block and 3D basic block structure diagram  
图 3. 2.5D 基础模块和 3D 基础模块结构图

### 2.3. 尺度注意力感知模块

为了学习有效上下文信息, 区域语义信息和边界信息在内的从粗到细的特征, 本文提出了尺度注意力感知模块(SAA)。SAA 的目标是探索不同尺度特征下的局部和全局三维空间信息。构造 SAA 的灵感来自 RFB [18], 其旨在通过卷积和扩展卷积的无缝组合来获得全局信息。与 RFB 不同的是, SAA 不仅引入了注意力机制, 而且还根据不同的尺度建立了不同的分支, 以进一步促进视觉感受野。具体地, SAA 形成具有不同尺度的多分支注意力特征池。然后, 嵌入多尺度特征融合策略, 以有效地整合低级上下文信息和高级区域语义信息。

图 4 展示了尺度注意力感知模块的所有细节。假设编码器生成的多尺度特征为  $F_s$ ,  $s = \{1, 2, 3, 4\}$ , 我们将不同尺度特征下的视觉感知过程表述如下:

$$F_s^T = \begin{cases} \text{Conv}_3(F_s + \text{Up}(F_{s+1}^A)), & s < 4 \\ \text{Conv}_3(F_s), & s = 4 \end{cases} \quad (1)$$

$$C_r = \sigma(\text{Conv}_3(F_s^T, r)) \cdot F_s^T \quad (2)$$

$$F_s^A = \text{Conv}_1(C_1, C_3, C_5, C_7) \quad (3)$$

其中,  $F_s^T$  表示获得的中间特征,  $\text{Up}(\cdot)$  表示上采样操作,  $\sigma(\cdot)$  表示 softmax 激活操作,  $C_r$  表示单尺度的注意力特征,  $F_s^A$  表示第  $s$  个尺度注意力特征。  $r = \{1, 3, 5, 7\}$ 。

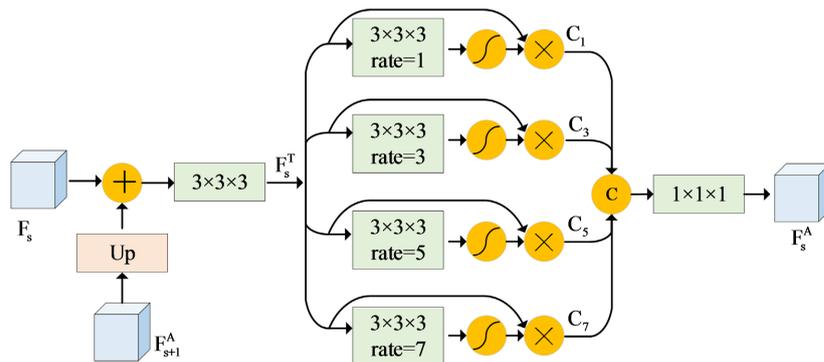


Figure 4. Structure diagram of scale attention awareness module  
图 4. 尺度注意力感知模块结构图

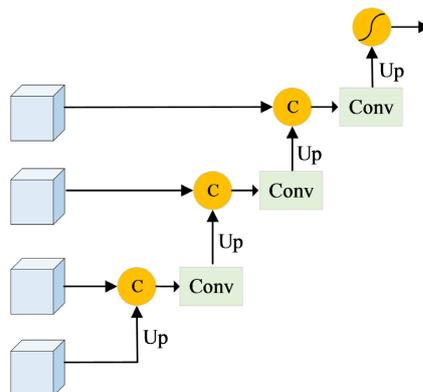


Figure 5. Diagram of the decoder module structure  
图 5. 解码模块结构图

## 2.4. 解码模块

解码模块主要由四层解码单元组成, 如图 5 所示。每一层解码单元都是对高层级的尺度注意力感知特征进行上采样, 得到与上一层级特征大小一致的新特征, 然后将这两个大小一致的特征进行拼接融合, 以弥补由于下采样操作中丢失的语义信息, 从而增强对边缘细节的分割。最后, 在最后一层解码单元中输出的特征向量被输入到 softmax 激活函数中, 将特征向量映射成概率值, 对应图像中每一个像素点, 并选择概率最大的类别作为预测结果。

## 2.5. 损失函数

交叉熵损失函数是目前语义分割任务中使用最广泛的损失函数。但是, 在肝脏与肿瘤分割任务中, 将面临着类别不平衡问题, 直接使用交叉熵损失函数会使模型倾向拟合类别数量多的样本。目前, 加权的交叉熵能有效的缓解这类问题, 其主要思想是对每个类别设置一个权重系数, 对数量少的类别加大惩罚, 公式如下所示:

$$L_{WCE} = -\frac{1}{\sum_{c=1}^3 \omega_c N_{y=c}} \sum_{i=1}^N \sum_{c=1}^3 \omega_i^c y_i \log p_i^c \quad (4)$$

其中,  $N$  表示图像的所有像素点的数量,  $c$  表示类别(背景, 肝脏和肿瘤),  $\omega$  表示权重,  $y_i$  表示第  $i$  个像素点的真实值,  $p_i^c$  表示第  $i$  个像素点被预测为类别  $c$  的概率值。

对于语义分割任务, 加权交叉熵只是在像素级计算损失值, 忽视了整体结构的损失值。因此, 引入了 Dice 相似性系数(Dice similarity coefficient, DSC)作为辅助损失函数。DSC 用于度量两个集合的相似性, 是评估分割精度最常用的度量指标, 公式如下所示:

$$DSC = \frac{2yp}{y+p} \quad (5)$$

其中  $y$  表示真实值,  $p$  表示预测概率值。DSC 损失函数可以定义为:

$$L_{DSC} = \sum_{c=1}^3 (1 - DSC) \quad (6)$$

最后, 我们使用联合损失函数用于肝脏与肿瘤分割, 能够缓解类别不平衡的问题的同时, 强调整体结构的损失值。公式如下所示:

$$L = L_{WCE} + \alpha L_{DSC} \quad (7)$$

其中,  $\alpha$  为 DSC 损失函数的调整权重系数。

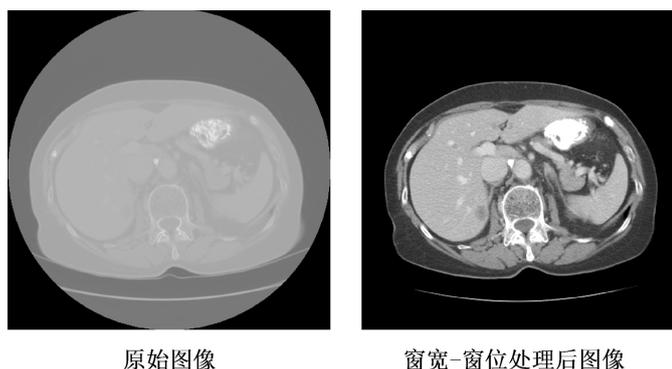
## 3. 实验与结果分析

### 3.1. 数据集及预处理

数据集是从国际医学图像计算和计算机辅助干预协会(Medical Image Computing and Computer Assisted Intervention Society, MICCAI)在 2017 年发起的肝脏肿瘤分割竞赛中获得的(Liver Tumor Segmentation, LiTS) [19]。LiTS 包括来自 6 个不同医院的 200 个对比增强的 3D 腹部 CT 扫描图像, 其中训练样本包含 130 个病例, 测试样本包含 70 个病例。为了进行有监督的训练, MICCAI 提供了由有经验的放射科医生对肝脏和肿瘤进行手工勾画的掩码图及注释。数据集在图像质量、空间分辨率和视觉方面存在显著差异, 平面内分辨率固定为  $512 \times 512$  像素, 像素距离为  $0.6 \text{ mm} \times 0.6 \text{ mm} \sim 1.0 \text{ mm} \times 1.0 \text{ mm}$ 。切片间距离为  $0.45 \sim 6.0 \text{ mm}$ , 不同病例的切片数量为  $42 \sim 1026$  张不等。

与自然图像相比, CT 扫描图像中存储的像数值单位为霍斯菲耳德单位(Hounsfield unit, Hu), 像素值范围在 $[-3000, 3000]$ 之间。在本实验中, 通过分析 CT 图像和肝脏区域的窗宽和窗位值得到 $[-200, 250]$ 内为最合适的强度值。所以, 把大于 250 Hu 的像素设为 250, 小于 $-200$  Hu 的像素设为 $-200$ , 从而更清晰观察到需要分割的肝脏区域[20]。图 6 展示了原始 CT 图和经过窗宽和窗位调整后的对比效果, 由图中可以观察到, 经过窗宽和窗位调整后, 能更清晰地观察到肝脏区域。

在数据集划分方面, 本文把 130 个病例按照 8:2 的比例随机划分为训练集和验证集, 剩下的 70 例作为测试集。此外, 为防止模型过拟合, 本文通过随机裁剪的方式对原图进行裁剪, 裁剪后的大小为  $32 \times 256 \times 256$ , 同时进行上下和左右翻转来达到数据扩充的目的。



**Figure 6.** Comparison before and after of CT image processing  
**图 6.** CT 图像处理前后对比图

### 3.2. 评价指标

为了评估模型在肝脏与肿瘤分割的有效性, 本文采用以下三种评价指标。

1) Dice 相似度系数(Dice similarity coefficient, DSC)表示预测结果  $p$  与真实结果  $y$  之间的重叠率, 取值范围 0~1, 0 表示预测结果与真实结果发生严重偏离, 1 表示预测结果与真实结果完全重合。其公式如公式 5 所示。

2) VOE (volumetric overlap error)表示错误率, 单位为%, 0%表示完全分割, 100%表示预测结果与真实结果完全不重叠。其公式为:

$$VOE = 1 - \frac{yp}{y + p - yp} \quad (8)$$

3) ASD (average symmetric surface distance)表示预测结果与真实结果之间对称位置的平均表面距离, 单位是 mm, 其中 0 mm 表示完美分割。其公式为:

$$ASD = \frac{1}{S(y) + S(p)} \times \left( \sum_{s_y \in S(y)} d(s_y, S(p)) + \sum_{s_p \in S(p)} d(s_p, S(y)) \right) \quad (9)$$

其中,  $S(x)$  代表集合  $x$  中的表面体素,  $d(v, S(x))$  表示任意体素  $v$  到  $S(x)$  的最短距离。

### 3.3. 实验参数设置

在参数设置上, 初始学习率为 0.001, 同时利用权重衰减为  $1e-5$  的自适应矩估计优化器来实现梯度下降。此外, 总训练轮次为 200 epoch, 每 50 个 epoch 将学习率降低 10%, 并在每个 epoch 之后在验证集上评估性能。本文所有的实验都是在 NVIDIA GTX 3090 GPU 上进行的。

### 3.4. 消融实验

为验证本文提出的不同组件的有效性, 本实验采用相同的参数和数据集, 通过逐一添加组件进行对比, 肝脏与肿瘤分割在测试集上的实验结果分别如表 1 和表 2 所示。

**Table 1.** Comparison of the results of different components in the liver segmentation

**表 1.** 不同组件在肝脏分割的结果对比

| 模型                                | DSC (%)     | VOE (%)    | ASD (mm)   |
|-----------------------------------|-------------|------------|------------|
| 2D ResNet34 + LWCE                | 92.2        | 13.7       | 6.6        |
| 3D ResNet34 + LWCE                | 94.1        | 10.8       | 5.2        |
| 3D ResNet34 + LWCE + LDSC         | 94.9        | 9.5        | 2.9        |
| 2.5D ResNet34 + LWCE + LDSC       | 95.7        | 8.3        | 1.2        |
| 2.5D ResNet34 + LWCE + LDSC + SAA | <b>96.4</b> | <b>7.3</b> | <b>1.1</b> |

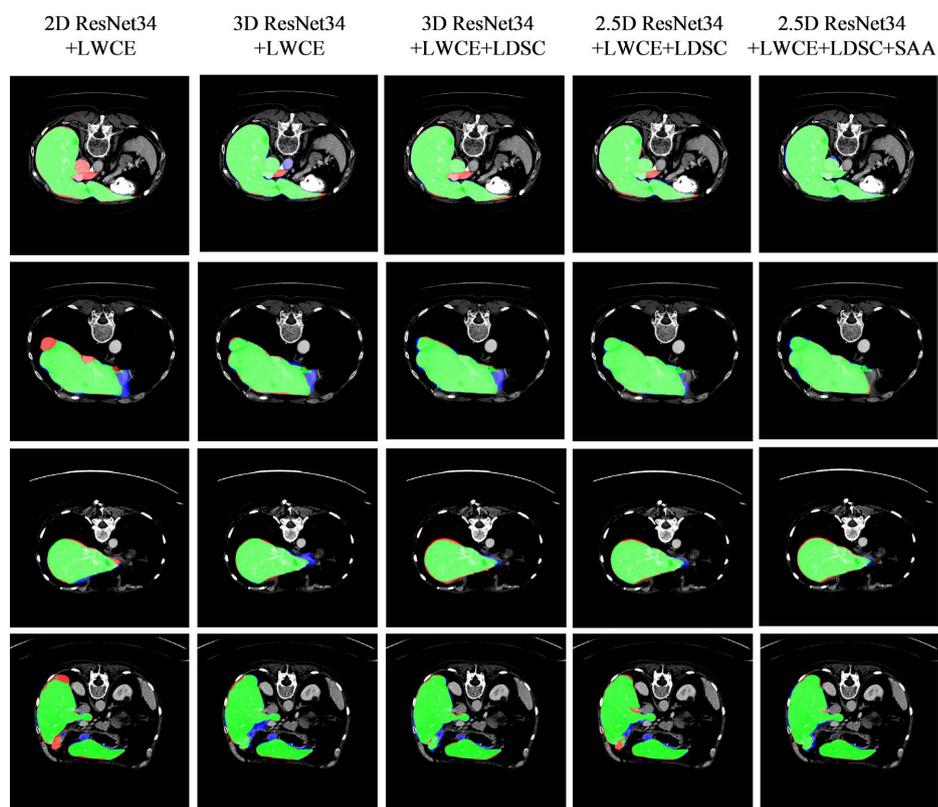
**Table 2.** Comparison of the results of different components in the tumor segmentation

**表 2.** 不同组件在肿瘤分割的结果对比

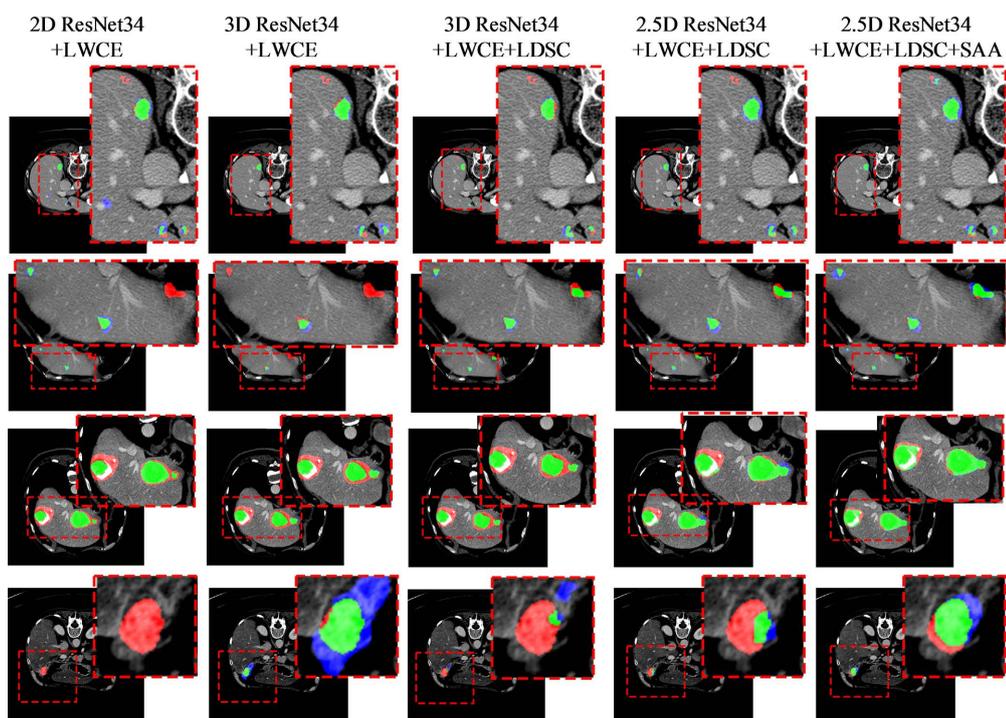
| 模型                                | DSC (%)     | VOE (%)     | ASD (mm)   |
|-----------------------------------|-------------|-------------|------------|
| 2D ResNet34 + LWCE                | 62.1        | 55.4        | 3.6        |
| 3D ResNet34 + LWCE                | 68.4        | 37.9        | 3.1        |
| 3D ResNet34 + LWCE + LWCE         | 70.2        | 34.2        | 2.2        |
| 2.5D ResNet34 + LWCE + LWCE       | 71.8        | 34.6        | 1.5        |
| 2.5D ResNet34 + LWCE + LWCE + SAA | <b>72.3</b> | <b>33.2</b> | <b>1.3</b> |

由表 1 和表 2 数据可知, 相比 2D ResNet34, 3D ResNet34 由于学习到三维的空间信息, 因此有助于肝脏和肿瘤的分割, 肝脏的 DSC 提升 1.9%, 肿瘤的 DSC 提升 63%。进一步地, 通过将加权交叉熵损失函数(LWCE)和 DSC 损失函数(LDSC)结合形成混合损失函数, 在肝脏和肿瘤分割的 DSC 上能得到进一步的提升, 证明这种组合对于肝脏和肿瘤分割任务能发挥更好的效果。2.5D ResNet34 方法在优化了三维空间信息后, 肝脏分割的 DSC 能提高至 95.7%, ASD 进一步优化至 1.2 mm。虽然 3D ResNet34 同样具有三维空间信息, 但由于图像横截面的像素距离大, 过多的使用三维空间信息反而会带来差的效果, 2.5D ResNet34 方法同样被证明在肿瘤分割中有效, 相比 3D ResNet34, DSC 提升 1.6%。最后, 通过嵌入尺度注意力感知模块(SAA), 在肝脏分割中, DSC 被进一步提升至 96.4%, 在肿瘤分割中, DSC 被进一步提升至 72.3%, 证明了本文提出的方法在肝脏和肿瘤分割中的有效性。

图 7 和图 8 分别展示了各种组件对肝脏和肿瘤的推理结果, 其中, 绿色区域表示正确分割肝脏/肿瘤, 红色区域表示该区域是肝脏/肿瘤, 但模型预测为背景, 蓝色区域表示该区域为背景, 但模型预测为肝脏/肿瘤。由图可知, 2D ResNet34 由于没有利用三维的空间信息, 无论是肝脏分割或者肿瘤分割, 都有较多的漏分割的区域。相反, 由于医学图像在横断面的像素距离较大, 直接使用 3D 卷积操作导致出现较多的误分割区域。通过使用 2.5D ResNet34 能在一定程度上缓解这个问题。最后结合尺度感知注意力模块, 本文最终的方法在肝脏分割中的效果达到最好, 其漏分割和误分割的区域非常少, 这也是这个方法 DSC 最高, VOE 和 ASD 最低的直接证明。



**Figure 7.** The inference results of liver among each of methods  
**图 7.** 各种方法对肝脏的推理结果



**Figure 8.** The inference results of tumor among each of methods  
**图 8.** 各种方法对肿瘤的推理结果

### 3.5. 对比实验

为进一步验证提出的方法的有效性, 本文在 LiTS 的 70 例测试集上, 比较了过去三年提出的肝脏与肿瘤分割方法以及一些主流的分割算法的 DSC 值。

**UNet 2D:** 该模型是最先应用在医学图像的 2D 分割方法。其特点主要是在解码阶段, 融合编码阶段的特征来弥补丢失的特性信息, 被证明在医学图像分割中非常有效。

**UNet 3D:** 由于 UNet 2D 不能充分学习三维的空间信息, 因此衍生出 3D 版本。相比 UNet 2D, UNet 3D 能通过 3D 卷积操作来学习三维的空间信息, 实现更准确的分割。

**VNet:** 该模型主要是引入了残差连接弥补下采样的特征丢失语义信息的问题, 并采用卷积替代上采样和平均池化, 以保留更过细节信息。

**DeeplabV3 + 3D:** 在空洞卷积和空间金字塔模块(ASPP)的基础上, 将深度可分卷积应用在 ASPP 和解码模块中, 提高网络的训练速度。

**RA-UNet:** 该模型是一个三阶段分割模型, 首先用 RA-UNet 2D 对肝脏进行定位, 然后用 RA-UNet 3D 对定位的肝脏进行精细分割, 最后用 RA-UNet 3D 从肝脏的感兴趣区域中分割肿瘤。

**H-DenseUNet:** 该模型主要是在 UNet 的基础上加入了密集连接块来获取浅层的特征, 同时通过融合 2D DenseUNet 和 3D DenseUNet 的特征进行精细的分割。

上述方法在 LiTS 的 70 例测试集的分割结果如表 3 所示。

**Table 3.** Comparison of the results of different components in the liver and tumor segmentation

**表 3.** 不同方法在肝脏与肿瘤分割的结果对比

| 模型             | 肝脏 DSC (%)  | 肿瘤 DSC (%)  |
|----------------|-------------|-------------|
| UNet 2D        | 91.1        | 49.8        |
| UNet 3D        | 93.2        | 52.5        |
| VNet 3D        | 94.9        | 62.4        |
| DeeplabV3 + 3D | 95.7        | 66.6        |
| RA-UNet 3D     | 96.1        | 59.5        |
| H-DenseUNet 3D | 96.1        | 72.2        |
| Ours           | <b>96.4</b> | <b>72.3</b> |

通过对比实验得出, 本文提出的方法在肝脏和肿瘤分割中性能表现良好。其中, RA-UNet 是通过使用两个模型来进行由粗到细的肝脏与肿瘤分割, 而本文提出的是端到端的单模型分割方法。虽然多模型融合能有一定的 DSC 提升, 但模型更复杂, 且需要更多的推理时间。与 RA-UNet 相比, 本文提出的方法在肝脏分割中, DSC 提高了 0.3%。在肿瘤分割中, DSC 达到 72.3%。同样地, H-DenseUNet 的结果也是通过多个模型融合得到的, 但本文的方法结果仍然比 H-DenseUNet 的更优, 这显示了本文的方法在临床实践中具有更高效的应用潜能。

## 4. 结论

在肝脏与肿瘤分割任务中, 针对三维医学图像的横断面间像素距离大的问题, 本文基于 2D 和 3D

ResNet34, 提出利用改进后的 2.5D ResNet34 对肝脏和肿瘤特征进行编码, 提高模型对三维空间信息的建模能力。同时, 利用 DSC 损失函数来提高模型对整体结构的分割能力。此外, 针对肿瘤形状大小高度可变的问题, 本文设计了尺度注意力感知模块来建模不同尺度特征下的局部和全局三维空间信息, 以有效地整合低级上下文信息和高级区域语义信息。同时, 在与其他先进的方法对比中, 证明了本文提出的方法的有效性。在未来的工作中, 将注重研究小目标的语义分割算法, 以进一步提高肿瘤的分割性能。

## 参考文献

- [1] Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C. and Parkin, D.M. (2010) Estimates of Worldwide Burden of Cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*, **127**, 2893-2917.
- [2] Müller, S.A., Maier-Hein, L., Mehrabi, A., Pianka, F., Rietdorf, U., Wolf, I. and Schmied, B.M. (2007) Creation and Establishment of a Respiratory Liver Motion Simulator for Liver Interventions. *Medical Physics*, **34**, 4605-4608.
- [3] 马金林, 邓媛媛, 马自萍. 肝脏肿瘤 CT 图像深度学习分割方法综述[J]. 中国图象图形学报, 2020, 25(10): 2024-2046.
- [4] Dou, Q., Chen, H., Jin, Y.M., Yu, L.Q., Qin, J. and Heng, P.A. (2016) 3D Deeply Supervised Network for Automatic Liver Segmentation from CT Volumes. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Athens, 17-21 October 2016, 149-157. [https://doi.org/10.1007/978-3-319-46723-8\\_18](https://doi.org/10.1007/978-3-319-46723-8_18)
- [5] Bi, L., Kim, J., Kumar, A. and Feng, D. (2017) Automatic Liver Lesion Detection Using Cascaded Deep Residual Networks. <https://arxiv.org/abs/1704.02703>
- [6] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 640-651. <https://doi.org/10.1109/TPAMI.2016.2572683>
- [7] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, 5-9 October 2015, 234-241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [8] Milletari, F., Navab, N. and Ahmadi, S.A. (2016) V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 *Fourth International Conference on 3D Vision (3DV)*, Stanford, 25-28 October 2016, 565-571. <https://doi.org/10.1109/3DV.2016.79>
- [9] He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [10] Li, X., Hao, C., Qi, X., Qi, D., Fu, C.W. and Pheng-Ann, H. (2017) H-DenseUNet: Hybrid Densely Connected UNet for Liver and Liver Tumor Segmentation from CT Volumes. *IEEE Transactions on Medical Imaging*, **37**, 2663-2674.
- [11] Han, X. (2017) Automatic Liver Lesion Segmentation Using a Deep Convolutional Neural Network Method. <https://arxiv.org/abs/1704.07239>
- [12] Jin, Q., Meng, Z., Sun, C., Wei, L. and Su, R. (2018) RA-UNet: A Hybrid Deep Attention-Aware Network to Extract Liver and Tumor in CT Scans. <https://arxiv.org/abs/1811.01328>
- [13] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *European Conference on Computer Vision*, Munich, 8-14 September 2018, 833-851. [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49)
- [14] Chen, L.C., Papandreou, G., Schroff, F. and Adam, H. (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. <https://arxiv.org/abs/1706.05587>
- [15] Tang, Y.B., Tang, Y.X., Zhu, Y.Y., Xiao, J. and Summers, R.M. (2020) E<sup>2</sup>Net: An Edge Enhanced Network for Accurate Liver and Tumor Segmentation on CT Scans. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Lima, 4-8 October 2020, 512-522. [https://doi.org/10.1007/978-3-030-59719-1\\_50](https://doi.org/10.1007/978-3-030-59719-1_50)
- [16] Gao, S., Cheng, M.M., Zhao, K., Zhang, X.Y. and Torr, P. (2019) Res2Net: A New Multi-Scale Backbone Architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 652-662.
- [17] Heimann, T., Ginneken, B.V., Styner, M.A., Arzhaeva, Y. and Wolf, I. (2009) Comparison and Evaluation of Methods for Liver Segmentation from CT Datasets. *IEEE Transactions on Medical Imaging*, **28**, 1251-1265. <https://doi.org/10.1109/TMI.2009.2013851>
- [18] Liu, S.T., Huang, D. and Wang, Y.H. (2018) Receptive Field Block Net for Accurate and Fast Object Detection. *European Conference on Computer Vision*, Munich, 8-14 September 2018, 409-419.

[https://doi.org/10.1007/978-3-030-01252-6\\_24](https://doi.org/10.1007/978-3-030-01252-6_24)

- [19] Bilic, P., Christ, P.F., Vorontsov, E., Chlebus, G. and Wu, J. (2019) The Liver Tumor Segmentation Benchmark (LiTS). <https://arxiv.org/abs/1901.04056v1>
- [20] Liu, T., Liu, J., Ma, Y., He, J., Han, J., Ding, X., *et al.* (2020) Spatial Feature Fusion Convolutional Network for Liver and Liver Tumor Segmentation from CT Images. *Medical Physics*, **48**, 264-272.