

# 基于改进 UNETR++ 的肝脏分割<sup>①</sup>

马力<sup>1</sup>, 王骏<sup>2</sup>, 梁羨和<sup>2</sup>, 郝金华<sup>2</sup>

<sup>1</sup>(中山仰视科技有限公司, 中山 528400)

<sup>2</sup>(中山火炬开发区人民医院 放射科, 中山 528400)

通信作者: 王骏, E-mail: ejun1101@163.com



**摘要:** 肝脏 MRI 影像的脂肪定量标准化过程中常需要对肝脏感兴趣区域进行手工采样, 但手工采样策略耗时且结果多变. 基于深度学习方法的全肝分割与手工勾勒的感兴趣区域在进行脂肪定量分析时, 变异性误差和不确定性程度更低, 性能更优越. 在进行全肝分割任务时, 为了提升分割性能, 本文在 UNETR++ 模型的基础上, 进行改进. 该方法融合卷积神经网络和 Transformer 结构各自的优点, 增加卷积结构分支用于补足局部特征, 同时引入门控注意力机制, 抑制不相关的背景信息, 使模型更为突出分割区域的显著特征. 相比于 UNETR++ 及其他分割模型, 改进的方法具有更优的 *DCS* 及 *HD<sub>95</sub>* 指标.

**关键词:** 全肝分割; 卷积神经网络; 门控注意力; UNETR++

引用格式: 马力, 王骏, 梁羨和, 郝金华. 基于改进 UNETR++ 的肝脏分割. 计算机系统应用, 2024, 33(2): 246–252. <http://www.c-s-a.org.cn/1003-3254/9414.html>

## Liver Segmentation Based on Improved UNETR++

MA Li<sup>1</sup>, WANG Jun<sup>2</sup>, LIANG Xian-He<sup>2</sup>, HAO Jin-Hua<sup>2</sup>

<sup>1</sup>(Zhongshan Yangshi Technology Co. Ltd., Zhongshan 528400, China)

<sup>2</sup>(Radiology Department, Zhongshan Torch Development Zone People's Hospital, Zhongshan 528400, China)

**Abstract:** In the process of fat quantification standardization in liver MRI images, it is often necessary to manually sample the liver area of interest, but the manual sampling strategy is time-consuming and the results are variable. Compared with manually sketched regions of interest, the whole liver segmentation based on deep learning method has lower variability error and uncertainty, and better performance in fat quantitative analysis. To improve the segmentation performance during the whole liver segmentation task, this study makes improvements based on the UNETR++ model. This method combines the advantages of a convolutional neural network and Transformer structure and adds convolutional structure branches to supplement local features. Meanwhile, it introduces a gated attention mechanism to suppress irrelevant background information to make the model more prominent features of the segmented region. The improved method has better *DCS* and *HD<sub>95</sub>* indexes than UNETR++ and other segmentation models.

**Key words:** whole liver segmentation; convolutional neural network (CNN); gated attention; UNETR++

## 1 引言

### 1.1 背景

目前, 测量脂肪分数需要人工选择肝实质中的感

兴趣区域 (region of interest, ROI). 这一策略非常耗时, 因为需要多个 ROI 才能获得可重复性的结果. 人工全肝分割的诊断性能与不同的人工 ROI 采样方法结果显

① 基金项目: 中山市科技计划 (2020B1077)

收稿时间: 2023-07-05; 修改时间: 2023-08-24; 采用时间: 2023-10-16; csa 在线出版时间: 2023-12-26

CNKI 网络首发时间: 2023-12-28

示出质子密度脂肪分数有很强的相关性<sup>[1,2]</sup>。在可重复性条件下,人工 ROI 和人工全肝分割获得了相似的观察者间一致性,但是这两种方法都导致了影像组学测量的差异。目前 Martí-Aguado 等人<sup>[3]</sup>提出的基于卷积神经网络 (convolutional neural network, CNN) 的全肝自动分割方案,获得与手动分割相似的诊断准确性,且自动分割具有稳定结果。文献<sup>[3]</sup>前瞻性研究证实了基于深度学习的全肝脏分割算法,在测量 MRI 质子密度脂肪分数时,诊断准确性与手工采样 ROI 进行评估得到的结果相似。深度学习方法与手工勾勒的 ROI 估计值相比,变异性误差和不确定性程度更低,性能更优越。准确的 MRI 自动全肝分割方法可以被扩展并无缝整合到临床实践工作流程中,能简化肝脏脂肪定量分析流程。

## 1.2 相关工作

肝脏 MRI 影像的脂肪定量标准化过程中,传统方法常需要对肝脏感兴趣区域进行手工采样,通过对多个手工勾勒的感兴趣区域(通常 3-4 个)进行脂肪定量分析求取均值。但手工采样感兴趣区域时需要排除肝脏中血管部分,防止血管部分对结果计算产生影响,这种策略耗时且因采样的区域不同而导致结果多变。目前大部分的肝脏分割相关任务主要集中在肝脏及其病变分割上,陆雪松等人<sup>[4]</sup>针对腹部 MR 中呈现出浸润现象、与相邻器官灰度值相似、边界较为模糊的问题,提出了一种基于深度卷积编解码的迭代网络结构。它将前次分割输出的概率图引入到网络浅层,与来自编码器的低水平特征图融合,弥补下采样时丢失的细节信息,最终得到 93% 的 Dice 系数。Amin 等人<sup>[5]</sup>提出了一种由两个主要步骤组成的方法:首先,采用局部拉普拉斯滤波提高图像质量;其次,提出语义分割模型,将特征提取到预训练好的 VGG16 模型中,传递到 U 型网络中,所提出的分割框架能精确地分割受感染的肝脏区域。该方法在 3D-IRCADB 和 LiTS17 两个数据集上进行了性能测试,达到 98% 的 Dice 系数。对于肝脏中剔除血管的分割研究较少, Martí-Aguado 等人<sup>[3]</sup>利用 CNN 进行全肝自动分割方案,该方案验证了基于深度学习的全肝分割与手工勾勒的感兴趣区域在进行脂肪定量分析时,获得与手动分割相似的诊断准确性,且自动分割具有稳定结果。王志明<sup>[6]</sup>构建了剔除血管的肝脏私有数据集,并设计并实现一种基于 U-Net<sup>+</sup> 网络<sup>[7]</sup>,用于肝脏 MR 图像的分割,该方法 Dice 系数还

到 93.8%。

上述分割任务通常采用与 U-Net 类似的编码器-解码器架构,其中编码器学习图像的分层低维表示,解码器将学习到的表示映射到体素进行分割。基于 CNN 的方法在编码器和解码器中分别使用卷积和反卷积,由于局限的局部感受野而难以获得更为准确的分割结果。随着 Transformer 结构<sup>[8]</sup>被用到计算机视觉领域,这种本质上是全局操作的神经网络,展现了具有竞争力的性能。

本文主要探索 CNN 及 Transformer 混合架构,以结合卷积和自注意力各自的优点来实现更好的分割。

## 1.3 本文贡献

本文对基于 Transformer 为主的 UNETR++ 网络进行改进,用于肝脏分割,后续用来替换肝脏脂肪定量分析中手工选取 ROI 步骤。创新点如下:(1) 本文模型是一种 CNN 及 Transformer 相混合的编码器、解码器架构,在 UNETR++ 基础上引入 CNN 结构分支用于捕获更多的局部特征,用于提升模型的分割准确性。(2) 引入改进的注意力门控模块,用于融合 CNN 结构及 Transformer 结构提取的特征,使模型能隐式学习,抑制输入特征中的无关区域的同时,使分割能按注意力权重聚焦到目标区域。

## 2 方法

CNN 结构的成功依赖于其两个固有的归纳偏置,即平移不变性和局部相关性,而视觉 Transformer 结构通常缺少这种特性,导致通常需要大量数据才能超越 CNN 的表现, CNN 在小数据集上的表现通常比纯 Transformer 结构要好。但 CNN 感受野有限导致很难捕获全局信息,而 Transformer 可以捕获长距离依赖关系,本文尝试将 CNN 和 Transformer 结合,使得网络结构能够继承 CNN 和 Transformer 各自的优点,并且最大程度保留全局和局部特征。

### 2.1 网络整体结构

本文提出的改进后 UNETR++<sup>[9]</sup> 分割网络结构如图 1 所示,模型采用分层设计,是一种 CNN-Transformer 相混合的编码器、解码器架构。编码器有两条分支,一条基于 Transformer 结构,另一条基于 CNN 结构。编码器 Transformer 结构分支有 4 个阶段,其第 1 阶段包括 patch 嵌入 (patch embedding), 其将体积块输入划分为 3D patches, 然后是新颖的高效配对注意力 (efficient

paired-attention, EPA) 块. 在 patch 嵌入中, 将每个 3D 输入  $x \in R^{H \times W \times D}$  转换为非重叠 patch  $x_u \in R^{N \times (P_1, P_2, P_3)}$ , 其中  $(P_1, P_2, P_3)$  是每个 patch 的分辨率, 然后将 patch

投影到  $C$  通道维度, 生成尺寸为  $H/P_1 \times W/P_2 \times D/P_3 \times C$  的特征图. 对于剩余的每个编码器阶段, 使用非重叠卷积的下采样层将分辨率降低两倍. 然后输入至 EPA 块.

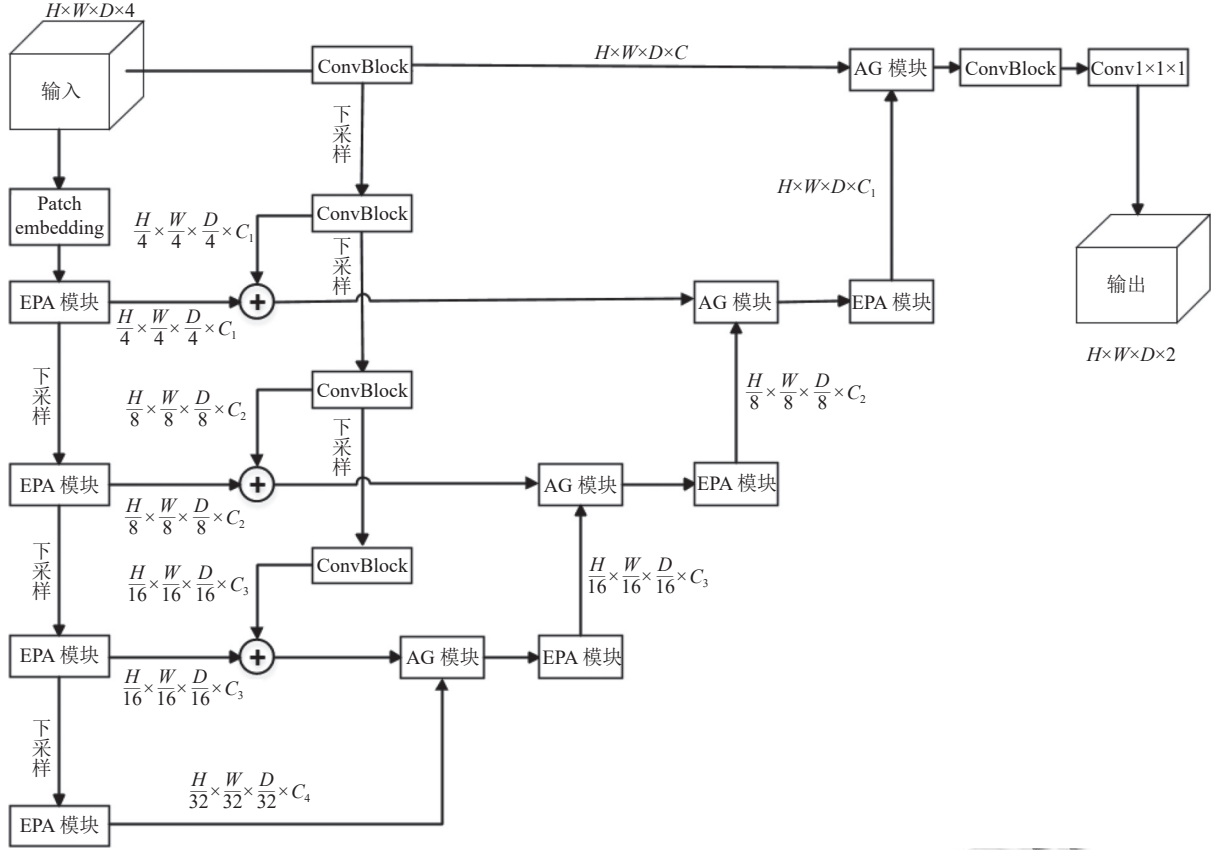


图 1 改进的 UNETR++ 架构

与编码器类似, 解码器也包括 4 个阶段, 其中每个解码器阶段都包含注意力门控 (attention gate, AG) 模块<sup>[10]</sup>, 在 AG 模块中集成了上采样操作, 通过使用反卷积将解码器部分的 EPA 块生成的特征图的分辨率提高两倍. 编码器层级通过跳跃连接与 CNN 分支结构生成的特征图进行相加操作, 得到融合后的特征. 再经过跳跃连接输入至同解码器层级的 EPA 模块捕获长距离依赖关系. 对输出的特征输入至 AG 模块, 通过注意力门控机制使图像分割时能按注意力权重聚焦到目标区域. 每两个解码器之间的通道数量减少两倍, 因此, 最后一个解码器的输出与卷积特征图融合以恢复空间信息并增强特征表示. 最后将生成的输出, 经  $3 \times 3 \times 3$  和  $1 \times 1 \times 1$  卷积块以生成体素方面的最终分割掩码预测.

接下来对 EPA 模块和 AG 模块进行详细介绍.

### 2.2 EPA 模块

EPA 模块是 UNETR++ 基于最近的 UNETR 框架<sup>[11]</sup>

进行的改进, 通过在两个分支中应用空间注意力和通道注意力, 有效地捕捉了丰富的相互依存的空间和通道特征. 如图 2 所示, 给出 EPA 模块的结构图, 每个 EPA 块包括两个注意模块, 通过使用共享 keys-queries 方案, 对空间和通道维度的信息进行编码来有效地学习空间通道特征表示. patch 嵌入生成的特征图被直接送到 EPA 块中, 随后是 3 个编码器层.  $Q$  和  $K$  线性层的权重在两个注意力模块之间共享, 每个注意力模块使用不同的  $V$  层. 两个注意力模块计算如下:

$$\hat{X}_s = SA(Q_{shared}, K_{shared}, V_{spatial}) \quad (1)$$

$$\hat{X}_c = CA(Q_{shared}, K_{shared}, V_{channel}) \quad (2)$$

其中,  $\hat{X}_s$  和  $\hat{X}_c$  分别表示空间和通道注意图. SA 是空间注意力模块, CA 是通道注意力模块.  $Q_{shared}$ 、 $K_{shared}$ 、 $V_{spatial}$  和  $V_{channel}$  分别是共享 query、共享 key、空间 value 层、通道 value 层的矩阵.

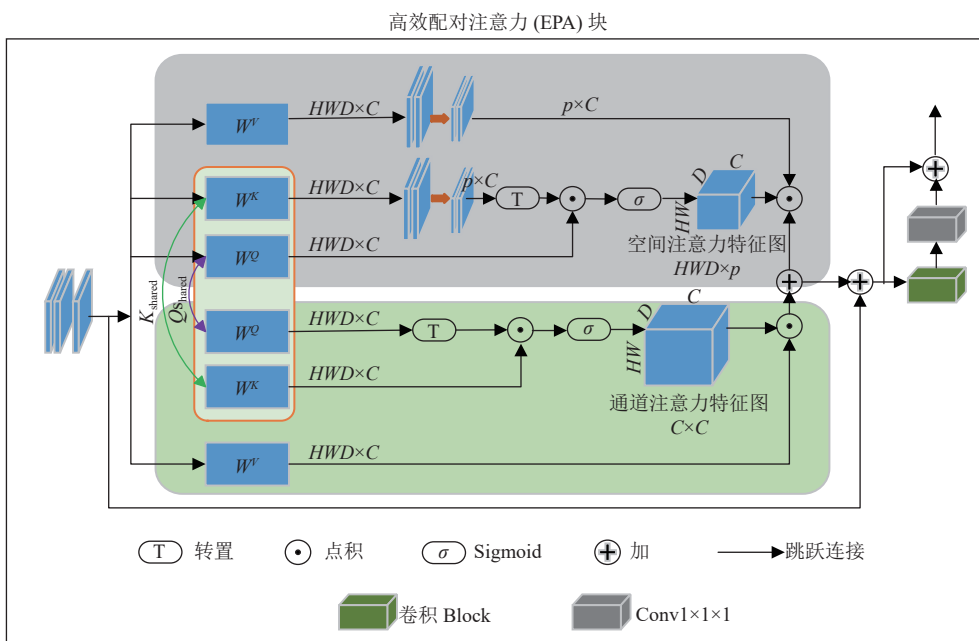


图2 EPA 模块

空间注意力: 该部分通过将复杂度从  $O(n^2)$  降低到  $O(np)$  来有效地学习空间信息, 其中  $n$  是 token 的数量,  $p$  是投影向量的维度,  $p \ll n$ . 给定  $HWD \times C$  形状的归一化张量  $X$ , 使用 3 个线性层计算  $Q_{shared}$ 、 $K_{shared}$  和  $V_{spatial}$  投影, 得出  $Q_{shared} = W^Q V$ ,  $K_{shared} = W^K V$ ,  $V_{spatial} = W^V X$ , 维度为  $HWD \times C$ , 其中  $W^Q$ 、 $W^K$  和  $W^V$  分别是  $Q_{shared}$ 、 $K_{shared}$  和  $V_{spatial}$  的投影权重. 然后, 执行 3 个步骤: 首先, 将  $K_{shared}$  和  $V_{spatial}$  层从  $HWD \times C$  投影到形状为  $p \times C$  的低维矩阵中. 其次, 通过将  $Q_{shared}$  层乘以投影  $K_{shared}$  的转置来计算空间注意力图, 然后使用 *Softmax* 来计算每个特征与其余空间特征之间的相似度. 第三, 将这些相似性乘以投影的  $V_{spatial}$  层, 以生成  $HWD \times C$  形状的最终空间注意力图. 空间注意力的定义如下:

$$\hat{X}_p = \text{Softmax} \left( \frac{Q_{shared} K_{shared}^T}{\sqrt{d}} \right) \cdot \tilde{V}_{spatial} \quad (3)$$

通道注意力: 该部分通过在通道 value 层和通道注意力图之间的通道维度中应用点积运算来捕获特征通道之间的相互依赖关系. 使用与空间注意力模块相同的  $Q_{shared}$  和  $K_{shared}$ , 计算通道的 value 层, 并使用线性层学习互补特征, 得到  $V_{channel} = W^V X$ , 维度为  $HWD \times C$ , 其中  $W^V$  是  $V_{channel}$  的投影权重. 通道注意力定义如下:

$$\hat{X}_p = \text{Softmax} \left( \frac{Q_{shared} K_{shared}^T}{\sqrt{d}} \right) \cdot \tilde{V}_{spatial} \quad (4)$$

其中,  $V_{channel}$ 、 $Q_{shared}$ 、 $K_{shared}$  分别表示通道 value 层、

共享 query 和共享 key,  $d$  是每个向量的大小. 最后, 执行融合, 并通过卷积块变换两个注意力模块的输出, 以获得丰富的特征表示. EPA 模块的最终输出如下所示:

$$\hat{X} = \text{Conv}_1(\text{Conv}_3(\hat{X}_s + \hat{X}_c)) \quad (5)$$

### 2.3 Attention gate 模块

Attention gate (AG) 是一种注意力门控模块, 最早应用于医学图像分割任务之中, 该模块能够自动学习聚焦不同形状和大小的目标结构. 使用 AG 训练的模型能隐式学习, 抑制输入图像中的无关区域, 同时突出对特定任务有用的显著特征.

本文改进的 AG 模块结构如图 3 所示, 其中  $x_e$  为编码器分层特征,  $x_d$  为解码器分层特征, 经 EPA 模块生成的特征  $x_d$  输入 AG 后, 先进行两倍上采样操作, 将特征图除通道外的尺寸扩大至  $x_e$  相同尺寸  $x'_d$ .  $x_e$  及  $x'_d$  通过  $1 \times 1 \times 1$  卷积来增加线性操作, 然后将生成的特征图在通道维度进行拼接 (Concat) 操作, 采用动态 ReLU (dynamic ReLU, DY-ReLU)<sup>[12]</sup> 根据输入动态地调整对应的分段激活函数, 来增强网络的表达能力. 随后再通过  $1 \times 1 \times 1$  卷积进行通道维度的变换, 使之同  $x'_d$  的通道维度保持一致, 通过 Sigmoid 激活函数处理后得到注意力权重  $\alpha$ , 其元素值范围为 (0, 1), 将  $\alpha$  与  $x'_d$  逐元素相乘, 使图像分割时能按注意力权重聚焦到目标区域, 最终得到 AG 模块的输出特征  $\hat{x}_d$ .

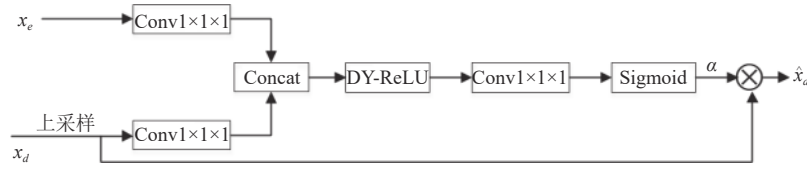


图3 改进的AG模块

## 2.4 损失函数

损失函数基于常用 soft Dice 损失<sup>[13]</sup>和交叉熵损失的总和,以同时利用两个互补损失函数的优势.其定义为:

$$\mathcal{L}(Y, P) = 1 - \sum_{i=1}^I \left( \frac{2 \times \sum_{v=1}^V Y_{v,i} \cdot P_{v,i}}{\sum_{v=1}^V Y_{v,i}^2 + \sum_{v=1}^V P_{v,i}^2} + \sum_{v=1}^V Y_{v,i} \log P_{v,i} \right) \quad (6)$$

其中,  $I$ 表示类别数量,  $V$ 表示体素数量,  $Y_{v,i}$ 和  $P_{v,i}$ 分别表示类  $i$  在体素  $v$  处的真实值和输出概率.

## 3 实验分析

### 3.1 数据集

本研究通过广东中山火炬开发区医院伦理委员会批准.从病历中收集患者的临床、影像学和组织病理学资料,共筛选出 95 例 MRI 数据进行研究.每个患者采用 1.5T MR 扫描仪进行上腹部 MR 平扫,使用 T1 VIBE Dixon 序列,取仰卧位进行检查.扫描参数:回波时间 (TE)=2.39 ms, 重复时间 (TR)=7.03 ms, 反转角度为 10, 带宽= $\pm 490$  kHz, 层厚 3.0 mm, 切片数: 72. 于检查前对受试者进行呼吸训练,以确保受试者每次吸气末屏气幅度尽可能一致.一次屏气扫描完成后即可同时获得 4 幅图像,同相位,反相位,脂肪相及水相.将数据随机分为训练集 (70 例)、验证集 (15 例) 和测试集 (10 例).

标注的制定:目前公开的相关腹部器官分割数据集中,肝脏标签图像并未对其中的血管组织进行剔除,本研究将构建分割数据集.数据集由 3 名放射科医师共同制定,其中由一名有 10 年腹部影像经验的放射科医师应用 ITK-SNAP 软件在 T1 横断面勾画肝脏区域并剔除其中较大的血管组织;另一名放射科医师(在腹部影像方面有 10 年经验)检查和调整标签,如果对某些区域的标注没有一致意见,则由第 3 名医师(在肝脏成像方面有 15 年的经验)进行检查和调整,并做出最终决定,横断面切片上的肝脏分割(剔除血管)标签如图 4 所示.

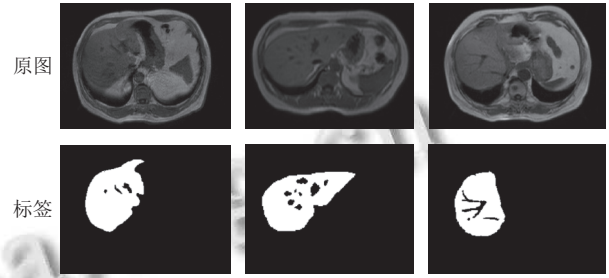


图4 肝脏分割(剔除血管)标签示例

### 3.2 实验配置

实验环境配置如表 1 所示.训练过程中使用了 Adam 作为优化器,初始学习率为 0.001,每 50 个 epoch 衰减到原始值的一半.将总训练时间点设置为 500 个,批次大小为 4 个,并保留训练期间在验证集中表现最好的模型用于预测.

表 1 实验环境

名称	配置
操作系统	Ubuntu 16.04.1 LTS
编程语言	Python 3.7.8
AI框架	PyTorch 1.9.0
CPU	Intel(R) Core(TM) i7-7820X CPU @ 3.60 GHz
GPU	NVIDIA GeForce GTX 1080 Ti×2 (11 GB×2)
内存	32 GB

### 3.3 评估指标

在医学图像分割的评估指标中, Dice 相似性系数 (Dice similarity coefficient, DSC) 常用来评估分割性能,其又称为重叠指数,表示两个物体相交的面积点总面积的比值, DSC 评价指标公式如下:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (7)$$

其中,  $X \in [0, 1]$  和  $Y \in [0, 1]$  分别为预测和人工标注的分割掩码, 0 表示分割结果与真实标签无相交, 1 代表分割结果与真实标签完全重合.

95% 豪斯多夫距离 (95% Hausdorff distance,  $HD_{95}$ ) 表示预测分割区域边界与真实区域边界之间的距离,其值越小代表预测边界分割误差越小.定义如下:

$$HD_{95}(X, Y) = \max\{d_{XY}, d_{YX}\} \quad (8)$$

其中,  $d_{XY}$  为预测体素与真实值之间的最大第 95 百分位距离,  $d_{YX}$  为真实值与预测体素之间最大的第 95 百分位距离.

### 3.4 对比实验

本文将提出的算法与 3D U-Net、3D V-Net<sup>[14]</sup>、UNETR、UNETR++模型进行了对比实验,结果如表 2 所示.

表 2 模型比较

模型	DSC (%)↑	HD <sub>95</sub> (mm)↓
3D U-Net	93.52	6.34
3D V-Net	93.89	6.01
UNETR	94.25	5.47
UNETR++	95.17	4.63
本文改进的UNETR++ (w/o AG)	95.59	4.12
本文改进的UNETR++	95.73	3.95

所有对比的实验方法均为 3D 分割网络. 在训练轮次达到 500 个 epoch 后, 3D U-Net、3D V-Net、UNETR、

UNETR++的平均 DSC 系数分别为 93.52%、93.89%、94.25%、95.17%, HD<sub>95</sub> 分别为 6.34、6.01、5.47. 改进的模型 DSC 系数及 HD<sub>95</sub> 指标, 均比其他模型更优. 对 AG 模块的验证, 从表 2 中可得到, 加入 AG 模块后, 改进的模型 DSC 系数提高了 0.14%, HD<sub>95</sub> 减少 0.17. 根据实验结果, 本文提出的方法与其他方法相比在全局信息的提取和局部细节保留上更具有优势, 在目前最先进的 UNETR++的基础上其分割效果 DSC 提升了 0.56%, HD<sub>95</sub> 减小 0.68, 证实了本文提出模型在肝脏分割任务上的有效性.

不同模型预测结果如图 5 所示, 相比本文改进模型, 3D U-Net、3D V-Net、UNETR 及 UNETR++等模型肝脏中血管部分有的未分割出, 且分割细节不精确, 本文模型能最大程度地排除掉血管组织.

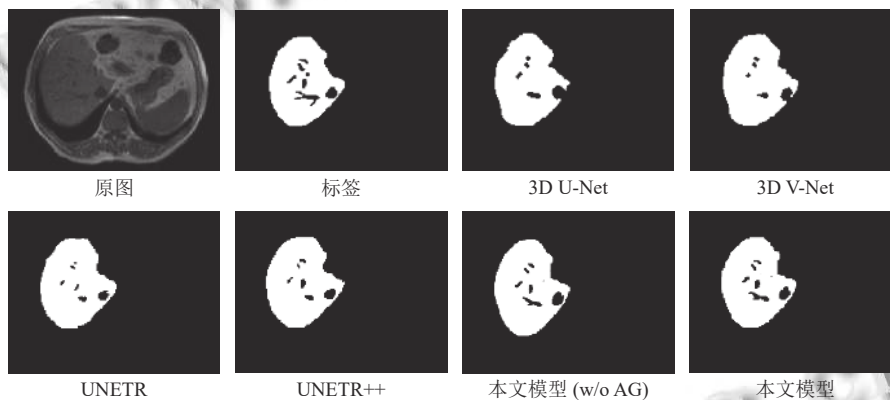


图 5 不同模型的分割结果对比

## 4 结论与展望

在本文中, 我们基于 UNETR++模型提出了一种改进的结构, 用于自动分割肝脏, 该模型的性能优于目前最先进的 UNETR++网络, 证明了我们的方法在肝脏分割任务上的有效性. 模型不足之处主要集中在提高了分割精度的同时, 增加了模型的参数和计算量. 同时, 由于构建数据集时对肝脏中的血管组织的剔除标注较为困难, 实验构建出的数据集较小, 后续研究需要进一步扩充数据集大小. 肝脏的准确分割是后续定量分析的重要前提, 后续可以利用深度学习方法进行自动全肝区域提取来替换手工 ROI 选取肝脏区域, 极大地帮助简化肝脏脂肪定量分析操作流程. 在未来的研究中, 我们计划通过在更大的数据样本中使用本文的解决方案, 并将自动分割算法扩展到其他器官, 如胰腺等其他腹腔组织.

## 参考文献

- Procter AJ, Sun JY, Malcolm PN, *et al.* Measuring liver fat fraction with complex-based chemical shift MRI: The effect of simplified sampling protocols on accuracy. *BMC Medical Imaging*, 2019, 19(1): 14. [doi: 10.1186/s12880-019-0311-y]
- Song JJ, Yu XL, Song WL, *et al.* MRI-based radiomics models developed with features of the whole liver and right liver lobe: Assessment of hepatic inflammatory activity in chronic hepatic disease. *Journal of Magnetic Resonance Imaging*, 2020, 52(6): 1668–1678. [doi: 10.1002/jmri.27197]
- Martí-Aguado D, Jiménez-Pastor A, Alberich-Bayarri Á, *et al.* Automated whole-liver MRI segmentation to assess steatosis and iron quantification in chronic liver disease. *Radiology*, 2022, 302(2): 345–354. [doi: 10.1148/radiol.2021211027]
- 陆雪松, 闫书豪. 基于迭代卷积神经网络的肝脏 MRI 图像

- 分割. 中南民族大学学报(自然科学版), 2022, 41(3): 319–325.
- 5 Amin J, Anjum MA, Sharif M, *et al.* Visual geometry group based on U-shaped model for liver/liver tumor segmentation. *IEEE Latin America Transactions*, 2023, 21(4): 557–564. [doi: [10.1109/TLA.2023.10128927](https://doi.org/10.1109/TLA.2023.10128927)]
- 6 王志明. 基于 U-Net<sup>+</sup>的全身脂肪及肝脏 MR 图像分割算法研究[硕士学位论文]. 武汉: 华中师范大学, 2021.
- 7 Çiçek Ö, Abdulkadir A, Lienkamp SS, *et al.* 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *Proceedings of the 19th International Conference on Medical Image Computing and Computer-assisted Intervention*. Athens: Springer, 2016. 424–432. [doi: [10.1007/978-3-319-46723-8\\_49](https://doi.org/10.1007/978-3-319-46723-8_49)]
- 8 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 9 Shaker A, Maaz M, Rasheed H, *et al.* UNETR++: Delving into efficient and accurate 3D medical image segmentation. *arXiv:2212.04497*, 2022.
- 10 Deng WJ, Shi Q, Li J. Attention-gate-based encoder–decoder network for automatic building extraction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 2611–2620. [doi: [10.1109/JSTARS.2021.3058097](https://doi.org/10.1109/JSTARS.2021.3058097)]
- 11 Hatamizadeh A, Tang YC, Nath V, *et al.* UNETR: Transformers for 3D medical image segmentation. *Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa: IEEE, 2022. 1748–1758.
- 12 Chen YP, Dai XY, Liu MC, *et al.* Dynamic ReLU. *Proceedings of the 16th European Conference on Computer Vision*. Glasgow: Springer, 2020. 351–367.
- 13 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.
- 14 Lei Y, Wang TH, Wang B, *et al.* Ultrasound prostate segmentation based on 3D V-Net with deep supervision. *Proceedings of the 2019 Medical Imaging Conference on Ultrasonic Imaging and Tomography*. San Diego: SPIE, 2019. 198–204. [doi: [10.1117/12.2512558](https://doi.org/10.1117/12.2512558)]

(校对责编: 孙君艳)