

分类号 TP391.1

密级 \_\_\_\_\_

UDC \_\_\_\_\_

编号 10741



## 硕士学位论文

论文题目 基于混合卷积神经网络的 CT 肝脏及肝脏  
肿瘤自动分割研究

研究生姓名: 刘祥强

指导教师姓名、职称: 丁晓阳 教授

学科、专业名称: 管理科学与工程

研究方向: 信息管理与信息系统

提交日期: 2021年5月31日

## 独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 刘祥强 签字日期： 2021.05.15

导师签名： 王刚 签字日期： 2021.5.15

## 关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意”/“不同意”）以下事项：

- 1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；
- 2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 刘祥强 签字日期： 2021.05.15

导师签名： 王刚 签字日期： 2021.5.15

# **Automatic segmentation of CT liver and liver tumor based on hybrid convolutional neural network**

**Candidate: Liu Xiangqiang**

**Supervisor: Ding Xiaoyang**

## 摘要

利用机器学习以及图像处理技术的方法从计算机断层扫描(Computed Tomography, CT)图像中寻找肝脏及肝脏中的肿瘤对于计算机辅助诊断治疗具有重要的意义。近年来,由于计算机性能的提升,深度学习成为医学图像分割的主流。肝脏 CT 图像是一个三维数据,二维全卷积神经网络(Fully Convolutional Networks, FCN)虽然在语义分割(Semantic Segmentation)任务中表现优秀,但是缺少了利用三维数据中的空间上下文信息这个过程。同时,三维全卷积神经网络虽然将 Z 轴上的信息与二维平面上的信息进行了综合优化,但由于其参数量巨大,网络性能往往受限于不能采取一定的深度甚至会过度考虑 Z 轴的信息。

基于以上各自维度的卷积神经网络的不足,本文提出了一种端到端的 2.5D 卷积神经网络,改进的工作主要由以下两点构成:(1)在二维全卷积神经网络中加入三维卷积部分同时提取空间特征,模型可以同时接受二维和三维图像的输入,并且二维卷积层和三维卷积层可以一起参与训练。(2)为了有效的融合二维和三维信息,本文提出了 AFM(Attention Fusion Module)模块用于将三维信息生成一个注意力掩模与二维特征图相乘,在训练中使模型得到一个重点关注区域。将模型对 MICCAI 2017 肝脏肿瘤分割挑战赛的测试集进行预测,通过消融实验表明,本文提出的同时提取二维和三维特征并且将它们融合的这种方式使得该混合全卷积神经网络(Fuse multi-dimensional Network, FMD-Net)模型相对于纯 2D 卷积神经网络和纯 3D 全卷积神经网络在测试结果指标上得到了明显提升,将 2D Unet 和 3D Unet 在测试集上的肝脏肿瘤预测结果的 Dice per case 值分别从 0.61 和 0.55 提高到了 0.662,与此同时将指标 Dice global 的值分别从 0.774 和 0.729 提高到了 0.803,并且将 2D Unet 的在测试集上的肝脏肿瘤预测结果的 Precision 这个指标从 0.193 提高了到了 0.253,这个指标对于从 CT 扫描中精准识别到肝脏肿瘤具有重要意义。

**关键词:** 三维特征 融合 卷积神经网络 端到端

## Abstract

The search for liver and tumors in liver from computed tomography (CT) images using machine learning and image processing techniques is important for computer-aided diagnosis and treatment. In recent years, deep learning has become the mainstream of medical image segmentation due to the improvement of computer performance. The CT image of liver is a three-dimensional data, and although the two-dimensional Fully Convolutional Networks (FCN) performs well in the Semantic Segmentation task, it lacks the process of utilizing the spatial contextual information in the three-dimensional data. Meanwhile, although 3D fully convolutional neural networks integrate and optimize the information on the Z-axis with that on the 2D plane, the network performance is often limited by the fact that it cannot take a certain depth or even over-consider the information on the Z-axis due to its huge number of parameters.

Based on the above deficiencies of the respective dimensional convolutional neural networks, this paper proposes an end-to-end 2.5D convolutional neural network with improvements consisting of the following two main components: (1) By adding a 3D convolutional part to the 2D full convolutional neural network to extract spatial features simultaneously, the model can accept inputs from both 2D and 3D images, and the 2D convolutional and 3D convolutional layers can participate in training together. (2) In order to effectively fuse 2D and 3D information,

the AFM (Attention Fuse Module) module is proposed in this paper for generating an attention mask from 3D information and multiplying it with the 2D feature map to make the model get a focused attention region during training. The model is predicted on the test set of the MICCAI 2017 Liver Tumor Segmentation Challenge, and the ablation experiments show that the proposed approach of extracting both 2D and 3D features and fusing them makes the hybrid Fuse multi-dimensional network (FMD-Net) model more efficient than the pure 2D convolutional network and pure 3D convolutional network. The Dice per case value for liver tumor prediction on the test set of 2D Unet and 3D Unet is improved from 0.61 and 0.55 to 0.662 respectively, while the Dice global value is improved from 0.774 and 0.774 respectively. The Dice global value was increased from 0.774 and 0.729 to 0.803, and the Precision of 2D Unet's liver tumor prediction in the test set was increased from 0.193 to 0.253, which is important for the accurate identification of liver tumors from CT scans.

**Keywords:** 3D features; Fusion; Convolutional neural network; End-to-end

# 目 录

<b>1 绪论</b> .....	1
1.1 研究背景及研究意义.....	1
1.2 肝脏及肝脏肿瘤分割方法研究现状.....	2
1.2.1 基于传统的机器学习和图像特征的方法.....	2
1.2.2 基于深度学习的方法.....	5
1.3 主要创新点和贡献.....	7
<b>2 全卷积神经网络</b> .....	8
2.1 神经网络.....	8
2.2 全卷积神经网络.....	9
2.2.1 卷积核.....	9
2.2.2 非线性修正单元.....	11
2.2.3 池化层.....	13
2.2.4 批归一化.....	15
2.2.5 上采样.....	16
2.3 本章总结.....	17
<b>3 基于 FMD-Net 模型的肝脏及肝脏肿瘤 CT 图像分割</b> .....	17
3.1 算法流程.....	17
3.2 模型结构.....	19
3.2.1 端到端.....	19
3.2.2 同时提取图像中的三维信息和二维信息.....	20
3.3 AFM 模块.....	21
3.3.1 attention 机制.....	22
3.3.2 降维卷积.....	23
3.4 本章总结.....	24
<b>4 实验设置与结果分析</b> .....	25
4.1 数据处理.....	25

4.1.1 数据集介绍.....	25
4.1.2 HU 值调整.....	25
4.1.3 归一化.....	27
4.1.4 数据增强.....	27
4.2 实验评价指标.....	28
4.2.1 Dice.....	28
4.2.2 Precision.....	29
4.3 实验设置.....	30
4.4 实验结果.....	31
4.4.1 消融实验.....	31
4.4.2 方法对比.....	32
<b>5 总结与展望</b> .....	<b>35</b>
5.1 总结.....	35
5.2 展望.....	36
<b>参考文献</b> .....	<b>38</b>
<b>致谢</b> .....	<b>42</b>

# 1 绪论

本章主要介绍了本文研究背景、研究意义和当今已经存在的从 CT 扫描中预测肝脏及肝脏肿瘤的传统方法和现代方法,同时说明从医学图像中全自动的分割肝脏以及肝脏肿瘤依然是医学界的一个难点;其次,阐述了在医学图像分割中已经存在的一些二维和三维的深度学习方法,分别说明了单纯从二维空间或三维空间提取特征的不足;最后,给出本文基于以往工作的不足所提出的创新点与改进。

## 1.1 研究背景及研究意义

肝癌是世界上七大癌症类型之一,也是第四大癌症致死类型<sup>[1]</sup>。并且每年的患癌人数以及肝癌发病人数正在逐年上升<sup>[2]</sup>。临床上勾画出 CT 扫描中癌症的位置以及轮廓需要经验丰富的医师和放射科医生,然而,以逐个切片的方式对具有数百个切片的体积的图像进行标注是非常繁琐耗时且非常昂贵<sup>[3]</sup>。同时医生在此类疾病的诊断和治疗时,经常会出现漏诊、误诊的情况,因此亟需要一种全自动的精准的方法来辅助医生诊断治疗,可以与医生的诊断结果进行交叉验证,降低误诊率,提高精准性,帮医生减轻工作负担。与此同时,一个可靠的肝脏肿瘤分割系统能够协助医生准确评估原发或继发肿瘤的发展情况,快速安排治疗时间。计算机辅助诊断(Computer-aided diagnosis, CAD)就是非常适合这个问题的工具,它不仅仅可以量化图像指标,同时可以勾画出图像感兴趣区域(Area of Interest, ROI)<sup>[4]</sup>,对后续的治疗、诊断工作具有巨大的辅助作用。本文提出的模型也是通过计算机全自动的从 CT 扫描中勾画出肝脏以及肝脏肿瘤区域。

从 CT 扫描中使用计算机识别出肝脏以及肝脏肿瘤一直是比较困难的研究课题,主要这个课题一直存在以下几个难点。第一,因为 CT 扫描本身属于灰度图像,它的对比度相对自然图像较低且肝脏与周围脾脏、心脏等其他器官纹理、灰度值接近如图 1.1 所示,肝脏肿瘤在 CT 扫描中与空气阴影、水的纹理特征也十分接近,不同病例间肝脏肿瘤的特征变化较大,例如某一个病人的肝脏肿瘤灰度值大小与其他病人的肝脏、脾脏接近。第二,肝脏肿瘤在形状、大小、数量和位置上的异质性,很难用一个普通的机器学习算法去总结他们之间的共同特征。第

三，像素级标注的数据不足，远远不如自然图像数据来的丰富，这个问题广泛存在我们的医学数据当中，不仅是医生繁忙，标注代价大，而且病人的数据往往涉及隐私从而难以获取。因此，利用计算机从 CT 扫描中分割出肝脏以及肝脏肿瘤一直是医学图像处理中的难点和重点。现阶段的大多数的肝脏肿瘤分割都比较依赖一个好的肝脏分割结果。因此首先模型能有一个精准的肝脏分割结果对于后续的肝脏肿瘤分割至关重要。肝脏的分割方法主要分为传统的机器学习方法和现在流行的深度学习方法。

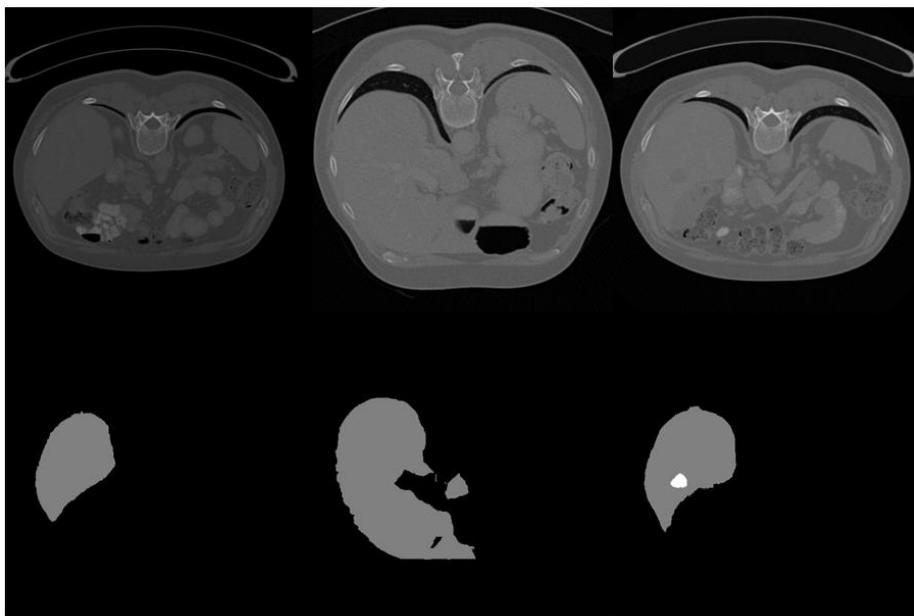


图 1.1 CT 图像的对比度显著低于自然图像

## 1.2 肝脏及肝脏肿瘤分割方法研究现状

### 1.2.1 基于传统的机器学习和图像特征的方法

传统的研究方法，肝脏分割通常使用的方法有区域生长法<sup>[5][6]</sup>、灰度阈值法<sup>[20]</sup>、图割算法、边缘检测法、基于局部特征的聚类算法等。阈值分割方法，是一种将基于感兴趣区域的灰度值分析的自适应阈值与基于模型的形态学处理相结合的混合方法。是使用一个人工设定好的阈值，将高于阈值的像素点认定为前景，将低于阈值的像素点认定为背景，通过这种方法，虽然很快把感兴趣区域提取出来，过程简单快速但是存在着众多缺点，首先这种方法不能保留肝脏中的背景色

区域，抗干扰能力很差且容易出现预测结果不连续的现象，其次，不同的 CT 扫描机器成像的效果是不一样的，对某一台机器生成的 CT 使用的阈值可能不适用于另一个机器的 CT 扫描结果，所以这种方法所依赖的人工设定的阈值并不可靠，同时也是因为如此，他需要人工参与，所以它并不是一个全自动的方法。

区域生长法是一种基于相似区域的分割算法。其基本思想是选择一个出发点，从某个像素点出发后按照一定准则将具有相似特征的点的连接起来，当生长满足终止条件时算法迭代结束<sup>[7]</sup>。Chen<sup>[8]</sup>等人提出了一种基于区域生长方法从 CT 图像中分割肝脏区域。为了得到更快、更准确的结果，肝脏分割方法总是结合先验知识，重点研究肝脏的位置、大小、强度分布和片间的拓扑关系。他们充分利用一个数据集中相邻 CT 切片之间的拓扑连续性，形成片与片之间区域生长的准则。考虑到肝区在中心坐标和强度分布上的连续性且肝区中心坐标在各片之间变化不大，可作为下一张片的部位引导；上一张切片肝区强度的均值和标准差值可较好地估计当前片上的肝区强度分布。此外，采用一些形态学操作如连通域和孔洞填充以确保快速而准确找到 CT 扫描中的肝脏区域。这种方法可以为感兴趣区域提供很好地边界信息和分割结果，但是其缺点也同样明显，区域生长法是一种迭代的算法，训练和测试时的计算资源和测试时间花费代价都比较大。其次，此算法的结果很依赖种子点的选择和生长的停止条件，这也造成了该算法可能在某一个 CT 上的分割结果表现良好，而同样的超参数设置不太适用另一组数据，因为他们在文中默认种子点的位置是在 CT 中心位置的，但是面对海量数据，有很多病人的 CT 中心位置并不在肝脏区域，这可能造成后续的肝脏分割失败。最重要的是利用此方法在 CT 扫描中分割肝脏的话，因为目标区域存在大量噪声和灰度不一的现象，导致分割结果产生大量且大小不一的孔洞，然而，这样的孔洞填充本身就是另一个难点，因此该方法的肝脏分割结果精确性十分有限。

全自动的从 CT 扫描中分割肝脏及其肿瘤的典型的无监督方法有模糊 C 均值聚类以及 K 均值聚类。Goryawala M<sup>[9]</sup>等人创建了一种将多种方法融合的方式，利用基于 k-means 的分割算法和专门用于 CT 数据集的局部轮廓算法。并且开发一个使用多个步骤的算法，在分割结果上有不错的表现。k-means 算法的种子点的选择也是一项重要的任务，因为随机选择聚类的种子点，可能生成效果十分差的掩模。使用同样的算法对每幅图像进行 3 次分割，其中每幅图像切片的 3 次运

行结果的交集即为最终分割结果,以达到更好的效果。该方法的缺点也是只适用特定的数据集,不同数据集的肝脏跨度是不同的,在大多数情况下,肝脏显示出的是三角形器官的正常解剖结构,其大部分体积在身体右侧,在右肾上方一点,然而也有一些情况,肝脏横跨整个腹腔,只是在膈肌下方。Zhao<sup>[10]</sup>等人提出了一种新的肝脏分割算法。首先,用阈值法去除初始图像中的肋骨和棘突,采用模糊 C-means 聚类算法和形态重建滤波对初始肝脏 CT 图像进行分割。然后,根据初始图像的分割结果,采用 BP 算法和已有金标准训练一个神经网络来处理预测结果。先用算法对相邻的 CT 扫描的切片进行分割,最后,用训练好的神经网络对生成的肝脏边缘的轮廓进行再一次处理。

边缘检测方法主要利用梯度寻找感兴趣区域的,但是这种方法对噪声比较敏感,容易出现分割不连续的边界,随后的区域填充同样面对孔洞的问题,不连续且变化不规则的边界,目前现存的算法很难完美的连接这些断点。在其他基于机器学习的肝脏分割方法中,C.L.Kuo 等人<sup>[11]</sup>在 CT 图像中提取肝脏肿瘤的纹理特征,随后将其用于医学临床中生存分析的预测。李登旺等人<sup>[45]</sup>发现分割肝脏使用显式模型容易受到适用条件的限制,分割进程不能得到有效的控制,因此限制了算法整体性能的提高,针对这个问题,他们将分割 CT 肝脏任务转化为隐式函数最优值搜索,根据图像纹理特征、先验形状信息有效地控制分割进程,用迭代搜索的策略得到最终的肝脏预测结果。

还有各类基于像素点特性的图割(Graph Cut)算法,图割是另一种基于图形的分割技术,它主要方法包括随机设置的种子点和成本函数。种子点指定了前景和背景体素,而图割分割法则是在图像中寻找将前景和背景体素分开的最佳切割边。更准确的说是将图像数据表示为一个图,图上有两组特殊的节点,称为终端,代表前景和背景种子体素,切边是图中完全将前景与背景种子分开的一组边,也就是说,如果去掉属于切边的边,任何前景和背景体素之间都没有路径,那将保留这条切边最终一步步迭代达到分割感兴趣区域的目的,有许多边缘集表现出这种特性。切口选择标准是有一定的根据的,如果在没有任何成本函数的情况下,它只是将具有最小数量的边缘的进行保留,如果存在成本函数,将对边缘进行加权,选择边缘集具有最低的整体权重的切口进行保留。Wu 等人在 2016 年提出了基于超体像素的图割方法,取得了传统方法中的在 CT 上分割肝脏的最佳表现

[12]。不同于当时的二维图割技术，它提出的是一种基于超体素的图割技术在 CT 三维图像上自动划定肝脏区域。首先，它是根据最大强度投影(Maximum Intensity Projection, MIP) 确定腹部区域以提取感兴趣的肝脏体积 (Volume of Interest, VOI) 和阈值化方法，然后通过基于直方图的自适应阈值法和形态学操作，从腹部区域提取患者特异性肝 VOI。肝脏 VOI 的超体素是使用简单的线性迭代聚类方法生成的。图切的前景、背景种子是在拥有最大面积的肝脏切片上生成的，并且图切算法已应用于 VOI 超体素。实验结果表明，所提出的方法能够准确地检测出肝脏，并且极大地减少了处理时间，尤其是在处理患病的肝脏病例时。但是分割结果表明在一些细而长的边界处表现效果不好，主要是由于图像此处的强度的相似性以及图形切割的收缩偏差最小化。

综上所述，以上方法主要是基于空间正则化和图像局部特征的方法，他们在肝脏的分割效果上最精准的模型 Dice 值仅仅为 92.7，肿瘤分割效果更是远不尽如人意，远远达不到临床的使用需求。更重要的是，他们的方法很多是半自动，例如阈值和种子点的选择。一些方法过程的确是全自动的，但他们更多利用的是数据集的统一性和肝脏区域的先验知识，这种方式的局限性很高，因为很多数据并不是像它们训练时使用的数据分布一样。

## 1.2.2 基于深度学习的方法

1998 年 Lecun<sup>[13]</sup>等人提出的全卷积神经网络因为其强大的特征提取和高维数据处理能力，已经被应用于各类语义分割任务，其中包括医学图像处理<sup>[14][15]</sup>。卷积神经网络的主要特点是无需人工挑选特征和鲁棒性，它通过卷积核、激活函数并且不断加深卷积层数来将卷积到的特征映射到高维空间，最终通过损失函数确定各个卷积核的参数，使模型在某一任务上的效果达到最优，所以卷积神经网络也被称为黑盒子。J Long<sup>[16]</sup>等人在卷积神经网络中加入了上采样的过程提出了全卷积神经网络，这是种端到端、像素到像素的训练方式，它去掉了用来输出分类结果的全连接层，通过各种上采样方式如双线性插值、反置卷积等方式将特征图恢复成原始尺寸，创造了用于分割感兴趣区域的卷积神经网络，这种方法立马成为复杂语义分割任务中的主流工具，随后出现的各类基于卷积神经网络的分割模型都是对他进行的改造。

2015 年 Ronneberger<sup>[17]</sup>等人面对医学数据相对较少的问题,提出 U-Net 网络,通过添加跳层连接 (Skip Connection) 等连接方式,将编码器中的信息与上采样的信息进行相加或者堆叠,弥补了在下采样中的不断损失的信息,它能有效地恢复目标对象的细微细节。在各类医学图像分割中取得了很好地效果,并且成为各种语义分割任务中的基础模型架构。它可以在拥有多个背景的图片上生成细节恢复特别好的前景对象,这种连接方式也是实例分割模型成功的基础,它甚至可以很好地预测被遮挡的物体。随后, Xiao Han<sup>[18]</sup> 等人就将 FCN 网络应用于肝脏以及肝脏肿瘤分割当中,这里为了获取一定的空间上下文信息,采用了同时输入 5 张切片的方式,这种空间信息的获取的方式并不是通过卷积的方式获得,与三维卷积相比对于信息的提取的并不充分。Zhou<sup>[19]</sup>等人提出了 UNet++网络,这是一种新的分割模型架构,它基于的是密集的、嵌套的跳层连接。这种架构背后的基本假设是当网络中解码器中的每一个卷积层接受来自对应的编码器特征之前,先逐渐丰富来自下采样阶段中的特征信息,这样可以更有效地捕获细粒度信息,综合考虑不同分辨率的特征图,当来自模型下采样和上采样过程中的特征图在语义上相似时,这种学习任务对于模型来说更加容易处理。这与 U-Net 中常用的普通跳层连接不同,不同于只能同一层特征图进行堆叠或者相加,这种连接直接建立了编码器中的高分辨率的特征图与解码器不同阶段卷积层之间的联系。通过不同的方法对比与消融实验表明这种提出的架构是有效的且与 U-Net 网络相比它的精准率和泛化性得到了显著提升。

面对医学影像大多是三维数据的特点, Milletari<sup>[20]</sup>等人提出了 U-Net 模型的扩展版 V-Net 模型,它是一种基于全卷积神经网络和体积的三维医学图像分割方法。模型接受的数据类型不再是一张张切片而是整个三维医学图像中的一块区域,模型的卷积部分采用的是三维卷积的方式,使网络在训练过程中可以充分考虑了三维数据中的 Z 轴信息,在模型卷积运算过程中使用的卷积核大小为 5 且步长为 2 的移动方式,随着特征块在不同阶段的下采样过程中下降,其特征块大小逐渐降低但是感受野逐渐增大。模型输出与输入同样大小,输出二分类预测结果,如果输入概率大于 0.5 的像素点则属于前景像素,否则就是背景。此外作者还提到在面对医学图像数据的时候不仅仅需要的是一个三维的预测结果,可能还需要肿瘤的体积以及具体的位置,这对于临床治疗和计算机辅助诊断是十分有意义的。

这篇论文中的另一个亮点是引入了 Dice 损失函数，有效的解决了大多医学图像分割任务存在的类别不平衡问题。PF Christ<sup>[21]</sup>等人提出了级联式的全卷积神经网络框架，首先使用第一个 U-Net 得到粗略的肝脏区域，然后使用第二个 U-Net 在肝脏区域中分割肝脏肿瘤，这也是今后大多数分割肝脏肿瘤工作的流程。

### 1.3 主要创新点和贡献

已经存在的分割模型大多是从纯二维或者纯三维的角度对数据进行特征的选取。但是在现实中，放射科医生在观察 CT 扫描时，并不是根据某一张切片进行最后的诊断的，而是在发现可疑区域时同时浏览连续几张或者更多的切片对某一层的切片的诊断作参考。因此，二维卷积神经网络在面对三维影像数据时，难以考虑某一层切面和相邻切片之间的关系<sup>[22]</sup>。三维全卷积神经网络虽然可以在训练中考虑到了数据本身 Z 轴上的信息，但由于三维卷积神经网络本身参数量巨大，复杂度极高，需要更多标注的训练数据和计算资源<sup>[23]</sup>，所以在某些医疗影像的分割任务中，三维分割模型在一些主要指标上还不如二维分割模型。X. Li<sup>[24]</sup>等人提出的 H-DenseNet 模型，在训练过程中分为多个阶段，将上一个阶段 2D FCN 得到的特征信息通过他们文章中提出的特征混合模块加入 3D FCN 进行训练，获得了十分理想效果。可见，不同维度之间的信息相互补充和融合对于模型性能的提升是有帮助的。针对以往工作中提出的模型的限制和优点，本文提出一种端到端的混合全卷积神经网络 FMD-Net，同时使用多维卷积提取 CT 扫描的特征，并使用 AFM (Attention Fusion Module) 模块融合多维度信息，本文的贡献主要有以下两点：

(1) 数据输入形式同时有图像切片和块，并且使用 2D 和 3D 卷积核在下采样阶段同时提取平面信息和空间信息，整个流程都是端到端的，不需要分开训练。

(2) 使用 AFM 模块将每一层 3D 卷积得到的特征使用注意力机制的方式与 2D 卷积得到的特征进行融合，帮助二维卷积得到空间信息。

## 2 全卷积神经网络

随着计算机硬件资源的逐渐提升和专门处理图像和视频的显卡的出现，深度学习已经成为了人工智能领域的主要工具，例如自然语言处理（Natural Language Processing, NLP）、计算机视觉（Computer Vision, CV）、持续学习（Continual Learning）、迁移学习（Transfer Learning）等等，其中全卷积神经网络更是将深度学习热度推向高潮，尽管后来出现了图神经网络（Graph Neural Network）等新的工具，但是目前全卷积神经网络依然是工业界和学术界主流工具。全卷积神经网络作为本文方法中所使用的主要工具，我们将对它做一定介绍方便接下来介绍我们的方法以及对于全卷积神经网络的改进。全卷积神经网络的发展主要经历了以下几个发展阶段：神经网络（Artificial Neural Network, ANN）、卷积神经网络（CNN）、全卷积神经网络（FCN）。

### 2.1 神经网络

人工神经网络是一种模拟人脑的神经网络以期能够实现类人工智能的机器学习技术。人脑中的神经网络是一个非常复杂的组织。成人的大脑中估计有 1000 亿个神经元之多。神经网络中每个神经元类似真实的神经元处理各种形式的输入信息，之后通过激活函数这个单一的输出单元与其他神经元相连接，达到了使模型拥有非线性拟合的能力，可以模拟现实生活中的各种数据分布。神经网络使用一种称为反向传播的技术来达到拟合的目的，以降低损失函数的损失为目标来更新每个神经元的权重。简单的人工神经网络如图 2.1 所示，它主要由三个层级的神经元组成：输入层、隐藏层、输出层。在设置神经网络的超参数时，隐藏层之外的神经元节点的个数通常是不变的，输入层节点个数通常与输入数据中拥有的属性或者特征值个数一致，输出层节点个数通常与输入数据所拥有的类别一致，隐藏层的节点个数则是不固定的，往往根据经验和任务选择。训练一个神经网络就是让它的权重的值调整到一个固定的值使得整个模型的性能最好，因此连接（图 2.1 中蓝色的箭头）是模型中最重要的东西，每一个连接上都被赋予了一个权重，在训练完成后是固定不变的。

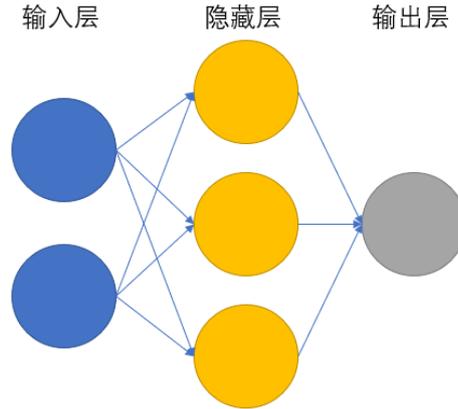


图 2.1 简单的神经网络示意图。

## 2.2 全卷积神经网络

与神经网络相比，卷积神经网络相当于先对输入数据做了一个预处理，全连接层相当于神经网络中的隐藏层中的神经元，是将高维抽象的数据做一个低维映射，输出图像中对应类别的概率。一张自然图片在计算机当中相当于一个三维矩阵，有着长宽高三个维度，高度代表着图像或者特征图拥有的通道数(Channel)，传统的神经网络无法面对这海量的计算规模，通过加入卷积、池化等操作，达到了让神经网络处理图像这种数据的目的。下面，将通过重点介绍卷积核、池化层、上采样等模块来说明卷积神经网络的主要结构。

### 2.2.1 卷积核

卷积核(Convolution Kernel)也称滤波器(Filters)的加入是卷积神经网络与传统神经网络的重要区别。卷积核是卷积神经网络的重要组成单位，卷积神经网络的主要参数都在成千上万的卷积核中，卷积核就是模型训练过程中主要的优化对象。卷积层是卷积神经网络的核心构件，它是模型训练过程中计算发生的主要地方。它主要由以下几个部分组成分别是输入图像、滤波器和特征图。我们假设输入将一张彩色 RGB 图像，它是由三维的像素矩阵组成的，这意味着输入图像将有三个维度包括高度、宽度和深度，对应图像中的 RGB。还有一个特征检测器，也称为卷积核或滤波器，它将在图像矩阵上按照设定好的步长进行移动，与对应的数据进行乘积的方式，这个过程被称为卷积。卷积层是卷积神经网络的核心构

件，它是模型训练过程中计算发生的主要地方。它主要由以下几个部分组成分别是输入图像、滤波器和特征图。我们假设输入将一张彩色 RGB 图像，它是由三维的像素矩阵组成的，这意味着输入图像将有三个维度包括高度、宽度和深度，对应图像中的 RGB。还有一个特征检测器，也称为卷积核或滤波器，它将在图像矩阵上按照我们设定好的步长进行移动，与对应的数据进行乘积的方式，这个过程被称为卷积。

在二维卷积神经网络中滤波器是一个二维的权重阵列，在训练过程中权重是不断变化的。虽然他们的大小可以不尽相同，但滤波器的大小大家通常是选择一个  $3 \times 3$  矩阵，这也决定了每个卷积层中感受野的大小。然后将滤波器应用于图像的一个区域，并计算输入像素和滤波器之间的点积，然后，这个点积被输入到一个输出阵列中。之后，滤波器按照设定好的步长进行移动，重复这个过程，直到卷积核扫描完整个图像。输入图像在经过滤波器的一系列点积之后最终输出的阵列被称为特征图、激活图或卷积特征。卷积的示意图如图 2.2 所示。

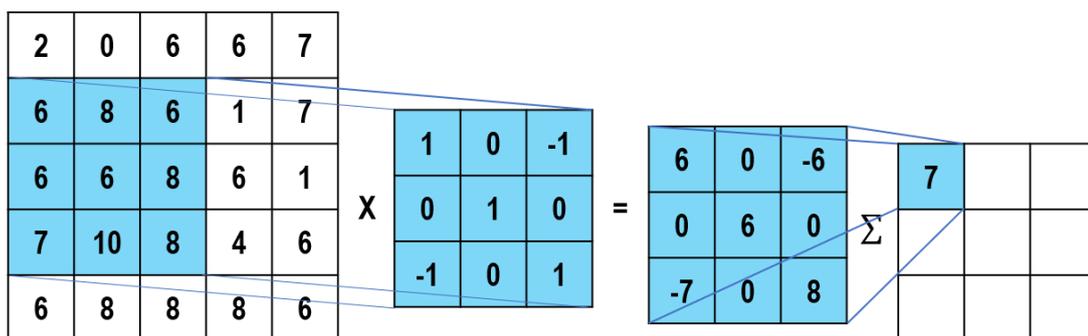


图 2.2 卷积过程的示意图

需要注意的是，当特征检测器在图像上移动时，它的权重是固定的，这也就是所谓的参数共享。有些参数，比如权值，在训练过程中会通过反向传播和梯度下降的过程进行调整。但是，在神经网络开始训练之前，有三个影响输出体积大小的超参数需要设置。

- (1) 滤波器的数量会影响输出的深度。例如，三个不同的滤波器将产生三个不同的特征图，形成三个通道。在我们的实验中，无论是三维卷积还是二维卷积部分所设置的卷积层的输出通道数都是上一层的两倍，上采样部分卷积层输

出的通道数与之相反。

(2) 步长 (Stride) 是滤波器在输入矩阵上移动的距离或者像素点的个数。理论上来说为步长为 2 或更大的情况相对比较很少, 跨距越大, 输出越小, 但是丢失的信息可能也越多, 因此在我们的模型中, 卷积采取的步长为 1。

(3) 零填充通常在滤波器不适合输入图像, 而我们需要制定输出特征矩阵的大小时使用。就是将所有落在输入矩阵之外的元素都设置为零, 从而产生一个更大或同等大小的特征矩阵, 有三种类型的填充方式。

- Valid padding。这也就是所谓的无填充。在这种情况下, 如果尺寸不对齐, 最后一个卷积会被放弃。
- Same padding。它通过矩阵卷积公式, 提前计算好需要补零的个数, 保证输入的图像和输出的特征图大小一致。
- Full padding。这种类型的填充方式通常通过在输入的边框上加零来增加输出的大小。

在本文实验中我们采取所有的填充方式都是 same padding 的操作。在有了以上几个关键的超参数设置之后, 就可以指定卷积方式输出特定大小的特征矩阵, 卷积核输出大小的公式 2-1 如下:

$$N = (W - F + 2P) / S + 1 \quad (2-1)$$

其中 N 表示输出之后的特征图的大小, W 代表输入特征图的大小, F 代表卷积核也就是滤波器的大小, 在本文实验中我们设定的是 3。P 和 S 分别代表采用零填充的值和步长大小。

## 2.2.2 非线性修正单元

在经历每次卷积操作之后都会将输出的特征矩阵用一个激活函数 (Activation Function) 进行变换, 从而使模型具有非线性拟合的能力。在卷积层中, 主要采用了卷积的方式来产生特征矩阵, 滤波器中的每个参数就是与像素点求点积的权重, 但是这个操作显然是一个线性的过程。在现实生活中, 大多数需要分类的数据不一定是线性可分的, 所以为了解决这个问题, 可以引入具有非线性映射能力的激活函数, 来拟合现实中非线性函数。因此如果没有激活函数存在, 那么多层感知机类型的神经网络只能进行线性映射, 因为就算有再多的神经元节点,

其整个模型跟单层感知机的映射意义是一样的，只能拟合线性的函数。激活函数又被称为“转移函数”。如果激活函数的输出范围是有限的，那么它可以被称为“压制函数”，许多激活函数是非线性的。激活函数的选择对神经网络的能力和性能有很大的影响，在模型的不同部分或者针对不同的任务可以使用不同的激活函数。但是从技术上讲，网络中每个节点的经过的激活函数都是采取统一的除了最后一层输出结果根据自己的需要选择不同的激活函数。激活函数另一个优点是它可以去除冗余信息让神经网络具有稀疏性，它通过构建稀疏矩阵来最大限度的保留信息中的特征，将大多数冗余的信息用 0 来表示。主要的激活函数有以下几种：Sigmoid 函数、Tanh 函数、ReLU 函数。

#### (1) Sigmoid 函数

Sigmoid 激活函数<sup>[25]</sup>也叫 logistic 函数，它与逻辑回归分类算法中使用的函数相同。该函数取任意实数作为输入，输出值在 0 到 1 的范围内，输入越大（越是正值），输出值越接近 1，而输入越小（越是负值），输出越接近于 0。Sigmoid 激活函数如公式 2-2 所示：

$$f(x) = \frac{1}{1 + e^x} \quad (2-2)$$

因为 Sigmoid 激活函数的导数还是其本身，因此计算起来非常方便且节省时间。在神经网络训练过程当中，Sigmoid 函数可以很容易的抑制误差尤其是相差比较大的误差，但是它也存在几点不足。首先，当输入的数据距离坐标原点较远时，函数的梯度就会逼近极限出现非常小的状况甚至几乎为零。在卷积神经网络利用 BP 算法进行优化时，通过链式法则使每一个权重都微分产生梯度，但是经过加权过后的值经过了 Sigmoid 函数之后，这个权重上的梯度就几乎很小了，就会出现我们常说的梯度消失现象。另外 Sigmoid 函数另一个很大的缺陷就是他的输出值永远是恒大于零的，这样会导致模型收敛可能出现过慢的状况，因为我们希望后续输入滤波器的信号是 0 均值的，如果输入信号都为正，那个所得到的梯度也都是正的，在反向传播更新过程当中也都是往一个方向更新的，最终导致收敛变慢。

#### (2) tanh 函数

tanh 激活函数<sup>[26]</sup>即双曲正切（Hyperbolic Tangent）函数，相当于 Sigmoid 幅度增大的版本，将输入信号压缩到-1 到 1 之间，解决了 Sigmoid 函数输出值只能为正的缺点。tanh 函数的导数取值范围在 0 至 1 之间，更加优于 Sigmoid 激活函

数取值范围 0 到 1/4 之间，在一定程度上，减轻了梯度消失和梯度弥散的问题。 $\tanh$  函数的如公式 2-3 所示。

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2-3)$$

$\tanh$  函数虽然一定程度上解决了 Sigmoid 函数的一些局限性，但是依然存在着梯度消失和幂运算的问题，幂运算是指数运算，会显著的降低计算速度。在输入信号的值距离原点很大或是很小的时候，输出的信号值也几乎是平滑的，梯度很小。

### (3) ReLu 函数

ReLU<sup>[27]</sup>(Rectified Linear Activation Unit)被认为是深度学习革命的几个里程碑之一。它很简单，却真的比一些常用的其他激活函数如 Sigmoid 或  $\tanh$  函数效果要好的多。首先，与传统的激活函数相比，ReLU 函数能够加快深度神经网络的训练速度，因为 ReLU 的导数对于正输入是 1。由于是常数，深度神经网络在训练阶段不需要花费额外的时间来计算误差项。当随着网络层数增加时，ReLU 函数不会引发消失梯度或者弥散问题。这是因为这个函数没有渐变的上下限。因此，最早的层（第一个隐藏层）能够接收来自最后几层的误差，以调整层间的所有权重。相比之下，像 Sigmoid 这样的传统激活函数被限制在 0 和 1 之间，因此对于第一个隐藏层来说，误差变得很小。这种情况将导致神经网络的训练效果不佳。

ReLU 的输出是零和输入值之间的最大值。当输入值为负时，输出等于零，当输入值为正时，输出等于输入值。因此，ReLU 函数遵循的公式如 2-4 所示。

$$f(x) = \max(0, x) \quad (2-4)$$

但是 ReLU 函数并不是一点缺点也没有，虽然它增加了网络的稀疏性，但是同时也是因为强制性的稀疏处理会让他屏蔽掉很多特征，例如属于在原点左边的信号，导致模型无法学习或者错过很多有效的特征。但是综合以上的激活函数的优点以及缺点后，我们最终还是选择 ReLU 函数作为整个网络卷积层当中的激活函数。

## 2.2.3 池化层

与之前的卷积核和激活函数类似，每一个卷积层当中也都有一个池化层，最后一层除外。使用池化层主要是因为它有两个特点，首先，池化层可以用于减少

输出特征图的大小,因此,它可以减少模型中的参数量和训练时的计算量。另外,卷积层输出的特征图中的一片区域中的相邻像素点具有特征值接近、属性类似的特点<sup>[28]</sup>,这说明卷积层输出的特征图中许多信息都是相近、冗余的。池化层按照设定好的准则挑选特征,减小特征图的大小。因此它的另一个特点就是对卷积层生成的特征图中的特征进行再挑主要特征。所以它的操作是在上一步总结的特征上进行的,而不是卷积层生成的固定位置的特征。这使得该模型在应对输入图像中特征位置的变化时更加稳健,因为它变向的增加了不断下采样过程中的感受野。池化层有多种池化方式,它包括最大值池化平均值池化。

### (1) 平均值池化

平均值池化 (Average Pooling) 计算的是利用指定大小的过滤器按照设定好的步长在特征图上移动,将所覆盖的特征图区域中所存在的元素加起来求和再平均得到平均值。因此,最大值池化给出了特征图中某一特定区域中最突出的特征,而平均值池化则给出了一个滤波器覆盖范围中存在的特征的平均值。平均值池化的示意图如图 2.4 所示。

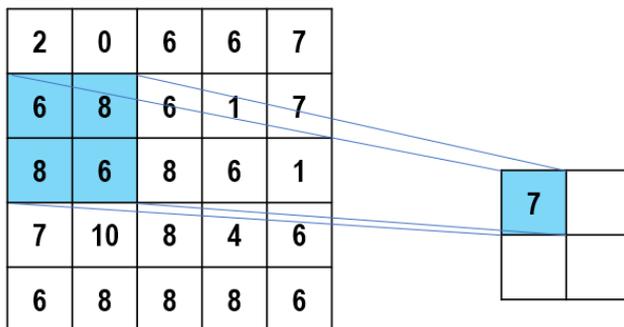


图 2.4 平均值池化示意图

### (2) 最大值池化

最大值池化 (Max Pooling) 是一种比较流行的池化操作,从过滤器覆盖的特征图区域中选择最大的元素。因此,最大值池化层后的输出是将包含前一个特征图中最突出的特征。虽然,最大值池化在操作过程当中导致了一些信息的丢失,但是我们通过跳层连接的方式尽可能的弥补了这个过程当中的损失的精度。在本文的实验中我们在下采样模块就是采用最大值池化的方式。最大值池化的示意图如图 2.5 表示。池化操作的过滤器的大小比特征图的大小要小很多,具体来说,它几乎总是  $2 \times 2$  像素,应用的步幅为 2 像素。输出的特征图大小是输入特征图的一半。

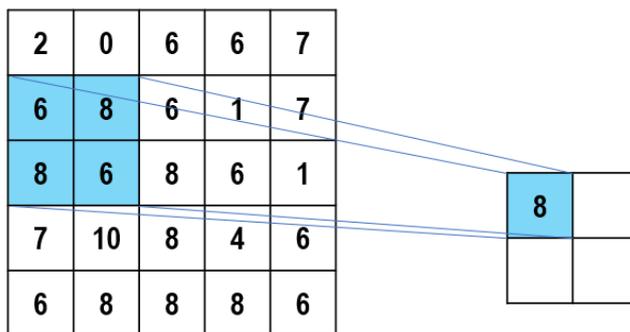


图 2.5 最大值池化示意图

## 2.2.4 批归一化

批归一化 (Batch Normalization) 层也叫 BN 层, Loffe 等人<sup>[29]</sup>将其用于深度卷积神经网络中, 至今, 批归一化已经成为各种流行的深度卷积神经网络的标配。归一化是一种数据预处理工具, 用于将输入的数值数据带到一个统一的尺度, 而不扭曲其形状。一般来说, 当将数据输入到机器学习或深度学习算法时, 我们倾向于将数值统一到同一个空间。之所以要进行归一化, 部分原因是为了确保我们的模型能够适当地进行泛化, 从而防止过拟合现象的产生。添加归一化层就是通过深度神经网络中增加额外的 BN 层来使神经网络更快、更稳定的过程了, BN 层对来自上一层的输入进行标准化和归一化操作。Batch 的含义是一个典型的神经网络是使用收集到的一组输入数据进行训练的, 称为批处理。同样, 批处理归一化中的归一化过程也是以批为单位进行的, 而不是作为单一的输入。具体而言, 我们使用 BN 层具体有以下几点目的:

- 对模型初始的权重不再敏感, 以往在设置深度学习的模型的初始权重通常采样符合 0 均值某方差的高斯分布, 对高斯分布的方差设置十分重要, 但是有了批归一化存在后它对与同一个输出节点相连的权重进行了缩放, 其标准差也会放缩同样的倍数。
- BN 层可以防止梯度爆炸或弥散, 因为 BN 层抑制了梯度消失, 可以让大部分的激活函数能够远离其饱和区域。另外也正是因为有了 BN 层的存在, 深度学习模型可以不需要太依赖 Dropout 层就达到减少过拟合的目的。可以将批归一化的函数定义为如公式 2-5。

$$z^n = \left( \frac{z - m_z}{s_z} \right) \quad (2-5)$$

公式中的 $m_z$ 代表着每个神经元的输出，而 $s_z$ 代表着每个上一层神经元输出的标准差。已经有很多工作和论文证明将 BN 层放置在 ReLu 前，因为 ReLu 激活函数的输出非负，不能近似为高斯分布，放在 ReLu 后相当于直接对每层的输入进行归一化。注意当 Batchsize 取 1 时 BN 层就失去了它的意义，所以为了能使用 BN 层，在肝脏肿瘤分割过程中我们将大小不一的 patch 图像统缩放到统一大小。

### 2.2.5 上采样

正是由于上采样(Upsample)的存在，才实现了 CNN 网络到 FCN 网络的跨越。通过提取卷积特征和池化降低分辨率后，对于图像分割任务来说，下一步就是将低分辨率的图像上调回输入图像的原始分辨率。池化就是将一片区域数值转换为单一数值，而上采样则相反，将单一数值转换为一片区域的数值。这个采用扩大图像尺寸，实现图像由小分辨率到大分辨率的映射的操作，叫做上采样。上采样有 3 种常见的方法：最近邻插值 (Nearest Neighbor Interpolation)、双线性插值 (Bilinear Interpolation)、转置卷积(Transposed Convolution)。

转置卷积并不叫反卷积，但从大家叫错它的名字可以看出它是与卷积相反的操作。它们与传统插值的操作不同，而是学习一组可用于重建原始输入分辨率的权重，也就是说它反置卷积是有参数的并且可以学习的，它们可以在训练过程中与下采样中的卷积层联合训练。反卷积的输入与输出尺寸的关系如公式 2-6：

$$L = s(x - 1) - 2p + k \quad (2-6)$$

在公式 2-6 中，L 代表输出的二维特征图的尺寸，注意长与宽是一样的。x 是输入二维特征图的尺寸，p 代表在输入特征图外填充的大小，s 代表反置卷积滤波器的移动的步长。

插值指利用已知的点来“猜”未知的点，图像邻域插值常用在修改图像尺寸的过程，由旧的图像矩阵中的点计算新图像矩阵中的点并插入，常用的插值方法有最近邻插值、双线性插值等等，在本文实验中模型的上采样采用的是双线性插值的方法。与转置卷积的卷积核权重是可通过模型训练不断学习相比，双线性插值的上采样方法显得不那么智能，因为插值的方法相当于人工特征工程，在人工预

定的方式下利用上采样的方式,但是本文的模型上采样方式依然选择双线性插值的方式,主要原因有以下几点。首先,也正是因为没有参数需要训练,插值的操作可以帮助我们节省大量的计算资源和训练时间。另外,在上采样操作的每一层中都有与下采样卷积层一样的卷积核、BN 层、激活函数,完全可以将放大之后的特征再进行映射、学习,达到转置卷积中的智能的效果,因为其实过多的卷积有点像重复思考的过程,造成模型的冗余,因此在双线性插值也能很好地恢复图像分辨率的情况下,不需要使用转置卷积。

## 2.3 本章总结

作为本文方法使用的主要工具,重点介绍了全卷积神经网络的主要结构,通过介绍这些组成部分的作用和超参数设置,方便接下来一章来体现对已有工作的改进。在介绍全卷积神经网络的组成部分时,首先说明了全卷积神经网络是由简单的神经网络发展而来,依旧沿用了 BP 算法来优化网络。其次说明了卷积神经网络的基础单位卷积层分别由负责提取特征的卷积核,负责减少计算量和增大感受野的池化层,使模型具有非线性拟合能力的激活函数,并且能增强模型鲁棒性和泛化能力。最后介绍了全卷积神经网络能实现图像分割的重要操作上采样,以及我们所使用的上采样方式是双线性插值。

# 3 基于 FMD-Net 模型的肝脏及肝脏肿瘤 CT 图像分割

## 3.1 算法流程

肝脏以及肝脏肿瘤的分割流程主要分为以下两个部分:首先使用本文提出的模型用来训练只预测肝脏结果。然后,将第一阶段得到的粗略的肝脏结果作为 ROI 区域提取出来用同样的模型进行第二阶段的训练,同时预测肝脏以及肝脏肿瘤。使用 ROI 区域训练的好处是因为医学图像数据存在着严重类别不平衡的状况,尤其是在预测肝脏肿瘤时这种现象更为严重,所以使用 ROI 区域训练的好处是在一定程度上排除了不相关的背景信息,减少了类别不平衡,如图 3.1 所示:

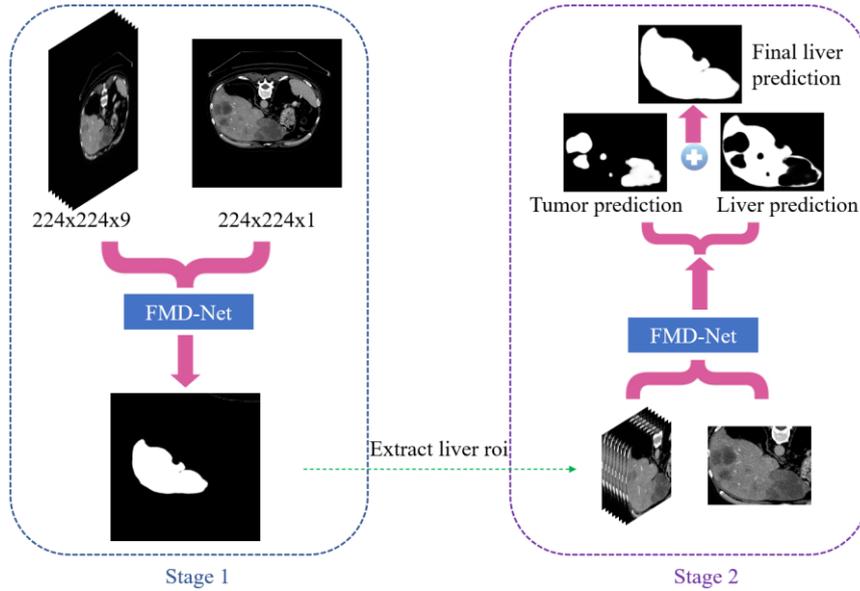


图 3.1 阶段 1 的粗略的肝脏分割阶段以及阶段 2 所示的肝脏及其肿瘤精准分割阶段

从图 3-1 中可以看出阶段 1 网络同时输入二维数据和三维数据，在阶段 2 中输入网络的数据同样如此。值得一提的是在阶段 1 的 FMD-Net 网络是进行一个二分类分割任务，只是单一的输出肝脏结果，因此阶段 1 也称为肝脏的粗略分割阶段。与之相对应的，在阶段 2 中，FMD-Net 网络进行的是一个三分类的分割任务，这样做的好处是让模型在学习肝脏肿瘤的知识同时学习肝脏知识，既充分利用了数据，让模型尽可能的学习更多的知识从而更加鲁棒，又可以在面对肿瘤与肝脏灰度值纹理特征接近时加以区分。在最终输出结果的时候让模型同时产生非肿瘤肝脏区域、肝脏肿瘤、以及背景，将该阶段的肝脏肿瘤结果作为最终结果提交，将肝脏肿瘤结果与非肿瘤肝脏区域进行叠加作为整个算法流程最终肝脏预测结果。

在算法的最后一个阶段分别采用两种数据后处理方式分别处理肝脏和肝脏肿瘤分割结果。模型对于 CT 图像中肝脏的预测结果实验中使用的是三维最大连通域的后处理方式，使用这个方式是因为我们充分考虑考虑到一个先验信息就是肝脏是一个连续单一器官，且在胸部中它的体积是最大，因此将模型对于肝脏的输出结果在三维空间中进行连通域标记，计算出各个连通域的体积并且保留最大的那个作为实验最后的肝脏预测结果，这样做可以有效去除模型预测结果中的过分割现象。对于 CT 图像中肝脏肿瘤的预测结果采用的是一个三维条件随机场

[30](3D Fully Connected Conditional Random Fields, 3D CRF)的数据后处理方式以实现更结构化的预测。三维条件随机场能够对任意大型体素邻域进行建模且计算方式十分高效,使其非常适合处理 3D 多模态医学图<sup>[31]</sup>,因此也适合本文模型所采用的肝脏增强 CT 扫描数据。对于模型输入图像  $I$  和模型输出结果  $z$ , CRF 模型中的吉布斯能量可以由以下公式 3-1 表述:

$$E(z) = \sum_i \Psi_u(z_i) + \sum_{i,j,i \neq j} \psi_p(z_i, z_j) \quad (3-1)$$

其中  $i$  代表原图像和模型预测结果的像素点,对于每个像素点  $i$  具有类别标签还有对应的观测值  $z_i$ ,这样每个像素点作为节点,像素与像素间的关系作为边,即构成了一个条件随机场。而且我们通过观测变量  $z_i$  来推测像素  $i$  对应的类别标签。其中的一元势函数  $\sum_i \Psi_u(z_i)$  即来自于全卷积神经网络 FMD-Net 的输出。而二元势函数的公式 3-2 如下:

$$\psi_p(z_i, z_j) = \mu(z_i, z_j)k(f_i, f_j) \quad (3-2)$$

二元势函数就是描述像素点与像素点之间的关系,鼓励相似像素分配相同的标签,而相差较大的像素分配不同标签,而这个“距离”的定义与颜色值和实际相对距离有关。

## 3.2 模型结构

本文提出的 FMD-Net (Fuse Multi-dimensional Network) 模型是一种端到端混合全卷积神经网络,这是与以往的一些 2.5D 模型不一样的地方,并不需要将二维模型和三维模型分开来训练,模型最终的预测结果同时受二维编码器和三维编码器部分影响。且在数据的输入形式上与任何模型的输入有所不同,并非单一的输入二维数据或者三维数据,而是同时输入。所以, FMD-Net 模型最大的特点就是端到端的同时提取多维信息和融合多维信息。

### 3.2.1 端到端

相对于端到端的网络有每一个卷积核的对于网络贡献分配在训练过程无法约束同时他们的贡献度也无法得知、网络可解释性的低的缺点,端到端网络的优

点其实是更加明显。首先，端到端的网络将多步骤、多模块的问题使用一步就可以解决，这个方法在工业界应用当中十分重要，为了将开发出的深度学习算法嵌入到设备当中，我们需要在保持模型鲁棒性的前提下尽可能的将模型进行有效的压缩，因此将多模型和多步骤融合到一块对于减少模型的参数量是十分可观的。X. Li 等人提出的 H-DenseNet 模型，在他们的模型中他们也使用的是 2.5D 的模型，不过他们的 2.5D 模型虽然可以做到全自动但不是端到端的，而是先由一个 2D 模型生成二维的分割结果，第二步将第一步得到的二维分割结果通过一个信息融合模块加入 3D DenseUnet 网络，整个过程分为了两步。端到端的训练过程可以将信息综合考虑而并不是分开单独训练相互之间缺少辅助。与单独训练二维分割网络和三维分割网络不同，在 FMD-Net 网络中，当最终的输出结果与金标准之间产生损失后进行反向传播更新网络权重时，二维编码器和三维编码器包括融合他们的 AFM 模块的参数都是同时更新的，不会单独的考虑某一维信息对于最终结果的影响，将这个综合考虑的步骤也加入整个训练的过程当中，更加符合现实生活当中医生阅片的习惯，这是 FMD-Net 的另一个好处。

### 3.2.2 同时提取图像中的三维信息和二维信息

FMD-Net 网络接受的数据输入形式同时有三维和二维图像，这种方式与之前的任何模型和方法都不同。大多数的二维分割和三维分割网络只接受二维图像和三维图像块的输入，只有一些二维分割网络它通过同时输入三张或者五张切片，它们称这种方法也为 2.5D 的网络，但是我们通过大量的实验表明，排除了随机性之后，连续输入多张和只输入一张对于最后模型的预测结果几乎是没什么影响的，这是因为这种获取空间的信息是有限的，他们仅仅使用点积的方式将不同的切片信息进行组合。在 FMD-Net 网络中，我们将接受的三维图像块使用传统三维卷积神经网络的编码器部分来提取特征，相对于点积的方式，使用三维卷积核提取特征增加了模型提取特征的丰富度也可以说是相对于二维卷积核，三维卷积核的非线性拟合能力更强。输入的三维数据的高度是 9，为什么选择这个高度将在接下来的 AFM 模块当中解释。因此三维数据输入的形式是  $N \times 1 \times 9 \times 224 \times 224$  的五维数据，二维数据输入的形式为  $N \times 1 \times 224 \times 224$  的四维数据，其中 N 为输入数据的 Batchsize，三维输入图像中的中间一张也就是第五张对应着输入的

二维数据，他们的长和宽都是一样的。在 3D Encoder 部分，我们取消了三维最大池化的操作，因为三维图像块高度有限，与 AFM 模块中的降维卷积操作相似，所以使用连续卷积的方式来降低图像块的高度用来增加模型在 Z 轴上感受野。在经历了四次三维卷积之后图像块的高度已经变为了 1，此时再使用压缩（squeeze）的操作来将五维数据降维成四维，在与相对应的二维信息进行融合时，与之前的 AFM 模块不同的是此时的 AFM 模块取消了连续卷积的操作。整个模型的模块设置和流程如图 3.2 所示：

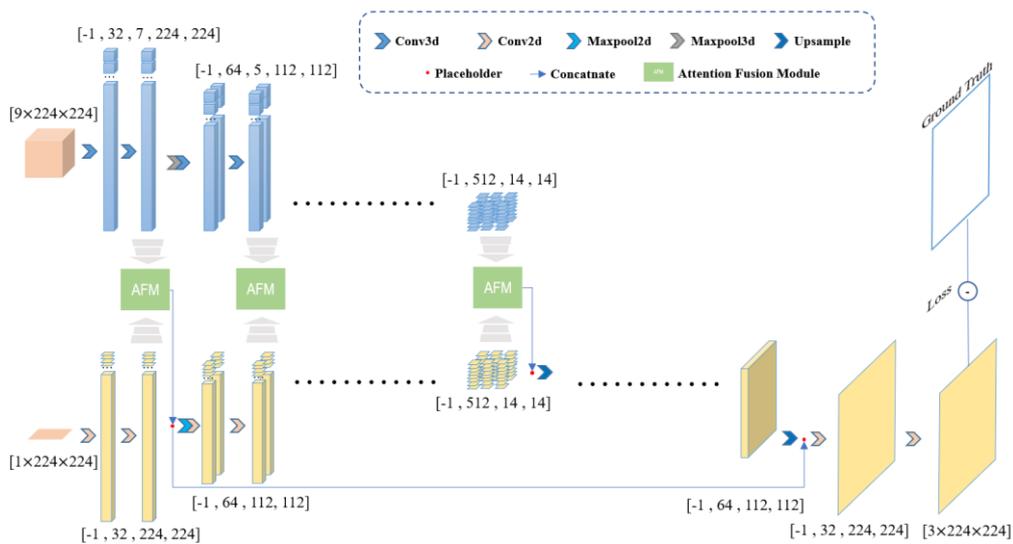


图 3.2 FMD-Net 示意图。

综上所述，FMD-Net 网络结构的思想主要是利用三维信息帮助二维模型进行分割。二维特征和三维特征使用 AFM (Attention Fuse Module) 模块进行融合，与 U-Net 网络类似，在上采样过程中也采用了跳层连接的措施来弥补下采样过程中的信息损失，与之不同的是将 AFM 模块产生的特征图与相对应上采样层进行堆叠。在阶段 1 网络只产生粗略的肝脏预测结果，在阶段 2 网络同时预测背景、肝脏及肝脏肿瘤。

### 3.3 AFM 模块

AFM 模块的主要作用就是将三维信息和二维信息进行融合，AFM 模块通过两种方式实现这个目的，一个就是降维，另一个工具就是注意力机制。我们通过消融实验和可视化的方式可以明显的看出 AFM 模块对于多维信息融合上面的帮

助。

### 3.3.1 attention 机制

FMD-Net 模型输出的二维结果，所以需要有一个机制使三维卷积的得到的信息能够帮助到二维卷积部分，Chen LC 第一次将注意力机制(Attention mechanism)<sup>[32]</sup>用在计算机视觉领域，在本文的多维信息融合中也可以采用这种机制。当人类的视野中存在一系列不同的物体时，我们必须将注意力集中在某一个物体上，我们视觉感知系统内的注意力机制会使用一组复杂的滤镜来制造模糊效果(类似于数码摄影中的“虚化”)，从而使感兴趣的物体成为焦点，而周围的物体则被淡化或模糊。注意力机制在人类认知中起着至关重要的作用。因此，首先在计算神经科学中，后来在机器学习中，有许多试图复制它的尝试。最近，基于注意力的模型在自然语言处理中取得了巨大的成功，随后在计算机视觉的深度学习模型中也有一些成功的应用。K. Xu 等人搭建了一个基于注意力的神经网络被用来生成图像的标题。图像首先由一个标准的卷积神经网络处理，然后用加入注意力机制的 Recurrent Neural Network (RNN)运行几个步骤来生成文本，RNN 的功能得到了增强，它可以决定下一步要关注图像的哪一部分。

空间和通道这两个词可能是计算机视觉中被提及最多的术语，尤其是涉及到卷积层的时候。这个张量是由 3 维度的指标来表征的： $h$  代表每个特征图的高度； $w$  代表每个特征图的宽度； $c$  代表通道总数、特征图总数或张量的深度。这就是为什么我们通常将输入的维度表示为  $(c \times h \times w)$ 。特征图是堆叠在一起的。 $c$  值由该层中使用的卷积滤波器总数决定。空间注意力代表了特征图上的注意力机制或注意力掩模，或者说是张量的单个切片。例如，感兴趣的对象是一只鸟，因此空间注意力将生成一个掩模，它将增强定义该鸟的特征。通过这样使用 Spatial Attention 来细化特征图，我们会增强后续卷积层的输入，从而提高模型的性能。在 AFM 模块中我们采用的就是 Spatial attention 的方式，我们将三维空间的特征通过卷积和激活函数的映射，我们得到一个注意力掩模，然后用这个掩模与二维特征图进行相乘，得到了加权之后的特征图。

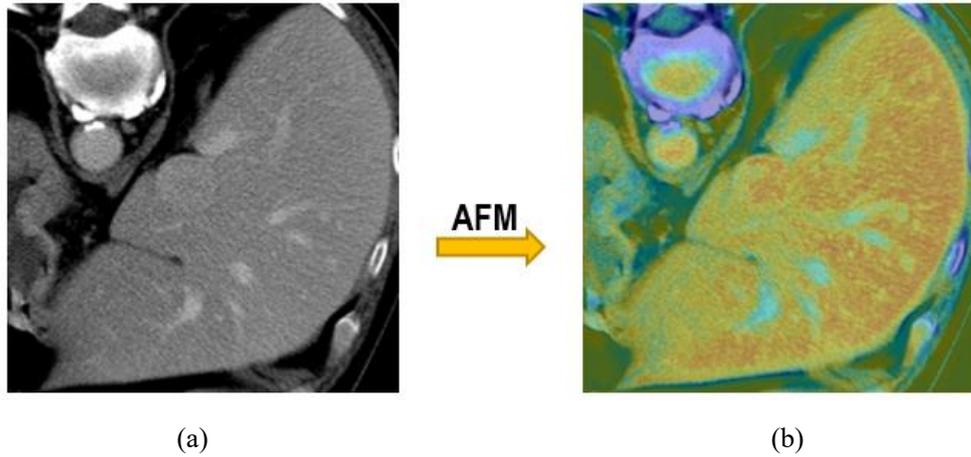


图 3.3 如图所示，a 为输入模型的二维原图，b 为得到了关注肝脏区域的热力图。

### 3.3.2 降维卷积

为了融合三维信息，首先需要将三维数据卷积成一维，这个卷积过程并没有池化层，而是完全通过连续的三维卷积，每一次三维卷积，三维特征图高度都会减少 2，这个过程类似于传统下采样过程中的池化层，由于肿瘤十分小，所有在不断下采样过程中在一些层数比较深的卷积层肿瘤的信息可能已经丢失了，使用这种连续卷积的方式可以有效减少信息的丢失。但是连续卷积的计算量并不小，这也进一步解释了输入的三维数据的高度为什么不能太高。这个模块是本实验流程中的重要环节，它是将三维特征图与二维特征图融合的关键，在模块中产生的权重图是二维的，所以在产生的权重图与二维特征图进行相乘。

值得一提的是，三维数据的高度之所以选择 9，因为在训练时，输入的二维数据对应着三维数据的中间一张，同时，在数据的首尾部分是没有肝脏区域的，在预测时，首尾的一部分数据不参与测试，模型默认预测结果为全零，输入的三维数据高度取得越高，默认预测结果为零的图片越多，因此高度不能取太高。另外通过训练集的金标准，计算出个体肿瘤的平均切片张数为 16，考虑在一些数据中存在着大肿瘤它们可能由上百张切片组成，但是在医生对某一张肿瘤进行标注时，并不会考虑相邻甚远的切片信息，最终为了方便计算和减少计算量，并且通过实验对比，将输入模型的三维数据高度选择为 9。在得到三维特征图之后，为了利用其信息对二维卷积部分提供帮助，只用连续卷积的方式来提取关键信息

和将特征块高度压缩为 1。不同于传统下采样中的卷积、池化等方式，在 AFM 模块中，虽然同样采取卷积的方式来压缩特征块高度，但是在连续卷积的过程中取消了池化的过程且保持了长和宽的大小不变，目的是为了产生与二维特征图同样大小的注意力掩模，之后与二维特征图进行相乘。

综上所述，当模型可以同时做到提取三维信息和二维信息的时候，如何融合和利用的他们十分重要，在本文中提出了使用 attention 和连续卷积的方式将三维特征信息进行降维和二维信息进行融合，从融合后的效果也可以看出这种方式是十分有意义的。整个 AFM 模块的具体操作和流程如图 3.4 所示。

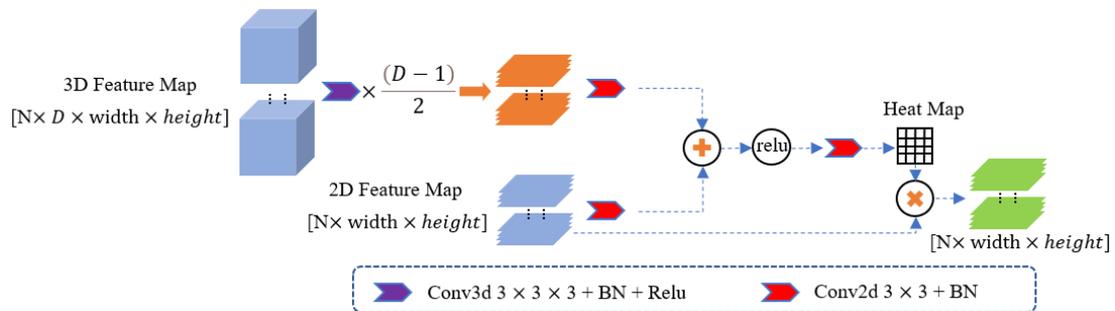


图 3.4 AFM 模块示意图

### 3.4 本章总结

本章介绍了我们使用模型进行从 CT 图像中分割肝脏以及肝脏肿瘤的流程。主要分为两个阶段，第一个阶段粗略的肝脏分割，第二个阶段利用第一阶段的结果再进行精准的肝脏以及肝脏肿瘤分割，值得注意的是第二个阶段的肝脏预测结果是排除肝脏肿瘤区域的，所以最终的肝脏预测结果要加上模型对于肿瘤的预测结果。随后详细介绍了本文使用的模型 FMD-Net 网络，它的最大特点就是同时接受二维数据和三维数据的输入，虽然是 2.5D 模型但是它的输入输出都是端到端的。最后，介绍了模型中的关键部分 AFM 模块，它的作用是将模型提取到的二维特征和三维特征进行融合并且使用跳层连接操作与上采样信息进行堆叠实现信息恢复。

## 4 实验设置与结果分析

### 4.1 数据处理

#### 4.1.1 数据集介绍

为了保证实验结果的精准性，模型在 MICCAI 2017 肝脏肿瘤分割挑战赛（Liver Tumor Segmentation Challenge, LiTS）的数据集上面进行训练和测试，并将测试的结果提交给比赛官方进行评估。该数据由 131 个病人的增强 CT 扫描组成的训练数据和 70 个病人的增强 CT 扫描组成的测试集。

LiTS 的图像数据来自多个医学中心和研究中心。研究对象中包括了各种肝癌类型患者，例如存在原发性肿瘤疾病（例如 HCC）以及继发性肝肿瘤和源自结直肠癌转移过来的肿瘤类型，还有乳腺癌和肺癌的转移类型患者。该数据集采用了不同强度的对比度增强，如高或低密度对比度。数据集中的三维医学图像数据充分包含了治疗前和治疗后腹部 CT 扫描的混合图像且图像数据是使用不同的 CT 扫描机器和采集协议采集的<sup>[43]</sup>。图像数据在分辨率和图像质量方面也非常多样化，图像分辨率的大小范围为轴方向 0.56mm 到 1.0mm，纵轴方向 0.45mm 到 6.0mm。另外数据集中每个病人三维数据的纵轴方向上的切片张数从 42 到 1026 张不等。一些图像包含成像伪影、噪声（如金属学人造物），这些都存在于真实的临床数据中，所以这些现象在该数据集中也存在增加了数据的丰富性。在训练集和测试集的肝脏体积呈现正态分布，并与已知分布相似<sup>[44]</sup>。肿瘤的数量在 0 和 75 之间变化。与训练集中的 131 个病人数据相比，测试集中 70 个病人的数据中肝脏肿瘤发生率更高。肿瘤的大小在 38mm 和 349mm 之间变化。综上所述，该数据集的数据多样且完整，提供了数据可能产生的各种随机性例如不同的机器产生的图片、治疗前后的病人肝脏状况、肿瘤类型等以应对可能需要各类临床需求。

#### 4.1.2 HU 值调整

该数据集中的所有数据切片大小都是 512×512，高度范围从 42 到 1026 不等，所有训练集 131 个病人的切片方向都是横向，但是在测试集 70 个病人中有

一个特殊的病人的数据样式，他的不同的点在于它的切片方向是纵向的，因此作为与训练集中截然不同的样本，模型对于其的预测结果可以很好地检验模型的提取特征和学习能力。HU 值(Hounsfield Units, HU)是计算机断层扫描中普遍使用的一种无维度单位，以标准化和方便的形式表达 CT 值。HU 值是由测量的衰减系数 1 的线性变换得到的。这种变换是基于任意分配的空气和纯水的密度，例如，蒸馏水在标准温度和压力下的辐射密度(STP)=0 HU，空气为-1000HU，以此使用公式对其进行变换方便表示。一般来说，CT 扫描使用 12 比特图像，能够存储-1024 至 3071 之间的值，由于人类眼睛可以感知的灰度范围是有限的，所以通常会将 HU 值范围调整到一个较窄的部分来突出显示感兴趣的组织或者器官，这个过程叫做“窗宽”。每一个像素的 HU 值反映了给定位置的组织成像密度，它给每一个像素点分配了灰度强度，数字越高越亮从而成为了数字图像。每一个组织器官拥有属于自己范围的 HU 值，这与组织的特性、结构、密度等各类因素有关，例如，由于脂肪的密度比水小，所以它的 HU 值的范围是-30 到-70 之间，在 CT 图像当中就可以表现为脂肪的颜色是比水深。

在该实验中所用到的 LiTS 数据集是增强 CT 扫描数据，它的各个主要组织器官的 CT 值范围如表 1 所示。为了更好地凸显肝脏与周围组织，增强对比度，我们将 HU 值的范围设定正在[-50,250]，这样这的取值范围既尽可能的保留了原始图像中的信息，也去除了绝大部分的不相关组织的信息，例如水和空气等造成的阴影。

表 4.1 在腹部 CT 中主要组织的 CT 值范围分布

器官、组织	CT 值
肝	50-70
脾脏	35-60
胰腺	30-55
肾脏	25-50
骨组织	>400

### 4.1.3 归一化

在深度卷积神经网络训练中，大多数模型的损失计算，需要假定数据的所有特征都是零均值并且具有同一阶方差的。这样在计算损失时，才能将所有特征属性统一处理。在机器学习或者深度学习中，大多数模型在计算损失的时候，一般假定我们的数据的所有特征都是零均值并且具有同一阶方差的。这样在计算损失时，才能将所有特征属性统一处理。然后在现实中数据的分布可能不是那么理想，因此我们通过归一化可以将特征属性统一到同一量纲，一定程度上可以提高模型的精度。另外，模型使用的是梯度下降的优化算法，梯度下降法的本质上是找到损失函数的最小值（通常描绘成等高线图）。当数据被归一化后，这个等高线图是对称的，梯度下降发现它更简单地达到平均值，同时模型的收敛速度比数据未归一化时要快，因为未归一化时它需要更多的步骤才能找到达到最小值的方法。每次在进行参数优化的时候会找到产生梯度最大的方向进行迭代。在更新过程中，可能会出现更新过程不是一直朝向最小极值点更新的，而是呈现非直线情况，这样会使迭代很慢。使用了归一化之后，这个模型更新的过程就变成了在近似圆形空间，不断向最小极值点迭代的过程，从而过程会加快很多。当处理图像时，通常将图像数据归一化到[0,255]，以确保像素值的范围在 0 和 1 之间。本文中所采用的数据归一化的方法是最大最小值归一化（Max-min Normalization），也称为离差标准化，通过公式 4-1：

$$x = (x - min)/(max - min) \quad (4-1)$$

将每个像素点的像素值  $x$  线性的缩放到[0,1]范围，通过之前的 HU 值调整后，其中最小值为样本数据中的最小值等于-100，最大值为样本数据中的最大值等于 250。

### 4.1.4 数据增强

在训练深度学习模型时，常常会遇到数据不足的情况，尤其是医学图像数据这个情况更加明显。由于其标注代价大、难以获取，相对于自然图像对标注人员的要求也更高，需要经验丰富的放射科医师，因此各类任务的医学图像数据往往是小样本数据，包括使用的 LiTS 2017 数据集，他的训练集的病人样本数也只有

131 个。全监督的深度学习模型的预测精度十分依赖训练过程中数据的数量和多样性,为了在复杂任务上获得高性能而训练的深度学习模型一般都有大量的神经元。随着隐藏神经元数量的增加,可训练参数的数量也会增加。在最先进的计算机视觉模型中,如 ResNet (60M) [33]和 Inception-V3 (24M) [34]的参数数量在千万级。它们有大量的可调整参数,因此,它们在训练阶段需要大量的数据来学习大量参数的值。否则就会出现模型泛化能力差、过拟合等现象的发生。为了解决这种数据不够或者样本不够丰富的问题,一般采用的方法是在现有数据上应用不同的变换来合成新数据,这种方法也称为数据增强 (Data Augmentation)。在本文中所使用的数据增强的方法主要由图像旋转和图像翻转构成,可以对图片进行左右和前后的翻转。为了防止样本冗余和重复,本文实验没有使用组合的方式对他们增强,例如,一个上下反转的图片等同于图片的 180 度旋转,然后再执行水平翻转。在旋转这个数据增强方式中主要采用左右旋转 90 度的方式。值得注意的是,所有的数据增强的操作方式都是随机的,在输入数据的过程之前产生一个 0 到 5 的随机数,每两个数字对应一种变换方式,如果产生的随机数是 0 或 1,那么数据输入采用原图的方式进行。如果产生的数字是 2 或 3,那么会采用图像翻转的方式,翻转的角度是固定而且是翻转方向也是随机的,所有对于输入 CT 图像的数据增强方式都会应用到金标准上面,最终达到数据增广的目的。这里的数据增强利用到的一个先验知识就是肿瘤的方向、位置是没有规律的,通过这种方式可以有效地、高质量地增加可用样本的数量,增强模型的泛化能力和精准性。

## 4.2 实验评价指标

本文所有的数据均来自与 LiTS 这个比赛,所有试验中所有的实验指标均来自比赛官方所使用的标准,以及本文给出的相关指标的结果皆来自于比赛官网的测评,所以结果公正可靠。下面,将针对本文所使用的到的指标进行详细说明。

### 4.2.1 Dice

Dice (Dice Similarity Coefficient, DSC)<sup>[35]</sup>最初是设计用来衡量目标检测任务、配准任务、图像分割任务的预测结果与金标准之间的相似性。与平衡 F 分数 (Balanced F Score) 类似,同时考虑模型的准确率(Precision)和召回率 (Recall)。

在该实验中，模型的分割结果是一个二分类的图像，所以就是用该指标衡量每个像素点的分类准确度。Dice 指标的计算方式如公式 4-2 表示：

$$Dice(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (4-2)$$

其中 X 属于模型预测结果的集合，Y 属于输入 CT 图像对应的金标准的集合，从公式中我们可以看出 Dice 的值得范围是 0 到 1 的，当 Dice 值接近于 0 时表示我们模型的预测结果比较差，相反当 Dice 值接近 1 时表示模型预测的结果与金标准高度相似。为了更加公平和全面的衡量分割结果，Dice 评价指标具体又分为了 Dice per case (DC) 和 Dice global (DG)，Dice per case 就是 Dice 系数的均值，它通过计算每个肿瘤与金标准的相似程度再除以肿瘤的个数得到的值。这种计算方式的侧重点在于计算分割结果中假阳性的情况，与分割肝脏类似，大的肿瘤相对于小肿瘤来说总是比较容易分割的，所以模型将小肿瘤分割的好的话可以显著提高 Dice per case 这个指标。因为肿瘤体积差异非常大，用平均 Dice 系数也就是 Dice per case 进行评价可能会不太客观，所以就引入了 Dice global，通常看一些论文的话，算法的 Dice global 可能会比 Dice per case 高十个百分点，当然这特指肝脏肿瘤，如果是肝脏这类体积本身就很大的器官，两种评价指标几乎是一样的，Dice global 的计算方式是把测试集样本 Dice 系数的分子的累加除以测试集样本 Dice 系数分母的累加。

#### 4.2.2 Precision

精准率表示被模型预测正确的正例占模型预测为正例的样本的比例。首先要了解两个概念就是 True Positives(TP)表示被正确地划分为正例的个数，即实际为正例且被分类器划分为正例的实例数，False Positives(FP)表示被错误地划分为正例的个数，即实际为负例但被分类器划分为正例的实例数，因此，精准率的计算公式与 4-3 所示：

$$precision = \frac{TP}{TP + FP} \quad (4-3)$$

这个指标对于评价模型关于肝脏肿瘤的分割结果是十分有意义的，如之前提到了实验所用的数据中存在着严重类别不平衡的现象，在 512 长宽的一张 CT 图像切片中，肿瘤的大小可能只占到几个像素点，且大多数的 CT 图像切片中是没

有肿瘤的。因此模型如果只预测负例是没有意义的，在实验中，肝脏肿瘤作为模型要预测的正例，而医生手动勾画的肝脏肿瘤结果为所有正例的样本，通过计算模型预测为正确的肿瘤像素点个数所占的比例，可以充分评价模型的精准率。在计算这个指标上时，当模型对于肝脏肿瘤的预测结果与金标准有一半以上的重合区域的时候，认为该肿瘤已经被检测到；相反，当肿瘤分割结果与医生分割结果的重合区域小于一半，则表示模型对肝脏肿瘤识别失败，这样的计算方式更加严苛但是对于临床意义是十分重大的，因为在临床中可以准确、快速的帮助医生找到肿瘤或者帮助医生查漏补缺。

### 4.3 实验设置

该实验所使用的显卡为 NVIDIA Tesla V100 32G，使用的编程语言是 Python，搭建卷积神经网络采用的深度学习框架是由脸书人工智能学院推出的 PyTorch(<https://pytorch.org/>)，之所以采用它是因为 PyTorch 是一个动态的框架<sup>[34]</sup>，这就和 python 的逻辑是一样的，要对变量做任何操作都是灵活的，不用像 Tensorflow 一样先要构建一个计算图，构建好了之后，这样一个计算图是不能够变的了，然后再传入不同的数据进去，进行计算进而带来一个问题，就是固定了计算的流程，势必带来了不灵活性。数据预处理、数据加载方式已经在 4.1 中介绍过，下面介绍在训练和测试阶段所使用的实验设置。本文将训练过程分为两个阶段，粗略的肝脏分割阶段和精准的肝脏以及肝脏肿瘤分割阶段。在第一个阶段，为了使用更大的 Batchsize，因为增大 Batchsize 不仅通过并行化提高内存利用率，还可以因为单次 epoch 的迭代次数减少，提高运行速度，不仅如此，适当的增加 Batchsize，梯度下降方向准确度增加，训练震动的幅度减小，因此，在将数据输入模型之前，直接使用 Pytorch 中的 Transform 工具包统一的将二维和三维图像 Resize 到 224 大小。这个阶段模型采用的是 Dice 损失函数，它与 Dice 评价指标类似，相对于普通的交叉熵损失函数，它可以充分考虑到类别不平衡的情况。所使用的优化算法是 Adam<sup>[36]</sup> (Adaptive Moment Estimation)算法，Adam 实际上就是将 Momentum 和 RMSprop 集合在一起，把一阶动量和二阶动量都使用起来了，模型所采用的学习率大小为固定在  $10e-5$ ，网络的最后一层输出通道为 2，将每个像素点的肝脏预测和背景预测作为二分类任务，至此，第一阶段肝脏分割阶段

结束。

第二阶段的肝脏以及肝脏肿瘤精准分割阶段,正如本文第三章方法 1 中所提到将第一阶段的肝脏分割结果作为肝脏 ROI 区域然后当成第二阶段模型的输入。此时每个人病人图像数据的大小是不一样的,PyTorch 可以接受不同大小数据的输入,但是 Batchsize 只能取 1,这样对于模型的训练和收敛是不利的,因此统一将不同大小的图像在输入前 Resize 到 224 大小。这个阶段使用的不再是 Dice 损失函数,而是使用带权重的交叉熵损失函数,因为通过实验发现由于肝脏肿瘤类别不平衡的问题要严重的多,尽管采取 ROI 区域和有几率的选择带有肿瘤的数据进行训练,依然会出现模型大概率预测为负例的情况,因此我们使用带权重的交叉熵损失函数如公式 4-4 表示:

$$loss(p, q) = - \sum_x (1 - \omega)(p(x) \log q(x) + \omega(1 - p(x)) \log(1 - q(x))) \quad (4-4)$$

其中  $x$  属于图像中的每个像素点,  $p$  是我们的预测结果,  $q$  是医生手动勾画的金标准,可以看到当  $p$  与  $q$  的值越接近时,整个交叉熵的值越小。这里的  $\omega$  就是加权交叉熵中的权重,它可以有效的增加正例的损失对于整个模型的影响。在肝脏肿瘤分割当中,模型对于每个像素点做的是 3 分类预测,它分别输出背景、肝脏、肝脏肿瘤三个结果,它们所拥有的权重分别是 0.78、1.05、8.07。并在学习率方面不再采取肝脏分割阶段中的固定学习率的方式,而是采用逐渐下降的方式,初始学习率设置为  $10e-3$ ,学习率衰减的方式使用等间隔的方法,每轮的最新的的学习率为上一轮的学习率乘以 0.1,训练一共设置 50 轮,每轮迭代 10 万次。

## 4.4 实验结果

### 4.4.1 消融实验

本文的提出的 FMD-Net 是一种端到端的 2.5D 模型,它所使用的 Backbone 是 U-Net 网络,因此将使用消融实验来证明本文提出的这种三维和二维信息融合的方式比单纯的二维和三维图像分割相对来说更加有效,主要包括了与 2D U-Net 和 3D U-Net 网络的实验指标对比。结果如表 4.2 所示,加粗的数值表示在这个评价指标上的最佳表现:

表 4.2 FMD-Net 在 LiTS 2017 挑战赛上的肝脏及肝脏肿瘤分割表现

模型	liver segmentation		tumor segmentation		
	Dice per case	Dice global	Dice per case	Dice global	precision
2D U-Net	0.957	0.959	0.61	0.774	0.193
3D U-Net <sup>[37]</sup>	\	0.951	0.55	0.729	\
<b>FMD-Net(ours)</b>	<b>0.959</b>	<b>0.960</b>	<b>0.662</b>	<b>0.803</b>	<b>0.253</b>

其中 3D U-Net 的实验结果引用自 Wang<sup>[42]</sup>等人在 MICCAI 2019 上所展示结果。从表 2 中可以看出, 相对于 2D U-Net 模型和 3D U-Net, FMD-Net 无论是在 CT 图像中分割肝脏还是肝脏肿瘤的表现都更加优越, 由于肝脏的对比度和形状相对于肝脏肿瘤较为明显且类别不平衡问题相对不那么突出, 所以肝脏的 Dice per case 和 Dice global 指标提升相对较少, 分别从 Dice global 的 0.959 提升到 0.960 和 Dice per case 的 0.957 提升到 0.959, 与 3D U-Net 模型相比较, Dice global 从 0.951 提升到了 0.960。由于肿瘤分割难度较大, FMD-Net 网络对于 2D U-Net 和 3D U-Net 网络的提升更加明显。在 Dice per case 和 Dice global 这两个指标上面 FMD-Net 分别将 2D U-Net 网络的 0.61、0.774 和 3D U-Net 网络的 0.55、0.729 分别提高到了 0.662 和 0.803, 在评价算法是否能够准确找到单体肿瘤指标 Precision 方面, 本文提出的 FMD-Net 网络相对于 2D U-Net 网络也是提升明显, 将 0.193 提升到了 0.253。

#### 4.4.2 方法对比

FMD-Net 模型在 MICCAI 2017 肝脏肿瘤分割挑战赛数据集上肝脏预测结果的主要指标 Dice Global 为 0.96, Dice per case 为 0.959; 模型对于肝脏肿瘤分割结果的主要评价指标 Dice Global 为 0.803, Dice per case 为 0.662, 准确率为 0.253。为了客观评价 FMD-Net 的网络的性能, 将其与既往的一些其他流行的二维和三维分割方法的比较。Liu 等<sup>[38]</sup>将深层监督机制(Deep supervision)应用于肝脏及肝脏肿瘤分割, 这种机制本来被开发用于自然图像中的边缘检测, 他们使用这种方法将模型的性能进行了较大的提升。Kaluva 等人<sup>[40]</sup>将传统的二维卷积方式采用了新的 Densely 连接的网络结构从 CT 图像中分割肝脏及肝脏肿瘤, 达到了不错的效果。由于 ResNet 在图像分类任务上取得了重大提升, Bi 等人<sup>[39]</sup>将这种残差

连接的方式 (Shortcut Connection) 分应用于分割网络当中并且用来分割 CT 图像中的肝脏及肝脏肿瘤。Li 等人<sup>[24]</sup>在实验环节中给出了 3D DenseUnet 网络在 CT 图像上分割肝脏及肝脏肿瘤的表现。与各个流行的二维及三维分割方法的性能展示如表 4.3 所示:

表 4.3 SFF-Net 模型与 MICCAI 2017 肝脏肿瘤挑战赛其他模型的对比

模型	liver segmentation		tumor segmentation		
	Dice per case	Dice global	Dice per case	Dice global	precision
2D SFF-Net <sup>[38]</sup>	0.937	0.955	0.592	0.746	<b>0.369</b>
2D ResNet <sup>[39]</sup>	\	0.959	0.50	\	\
2D DenseNet <sup>[40]</sup>	0.912	0.923	0.492	0.625	0.117
3D DenseUNet <sup>[24]</sup>	0.936	0.929	0.594	0.788	\
3D I3D <sup>[41]</sup>	0.957	0.960	0.624	0.776	
Blacknean	<b>0.961</b>	0.964	0.563	0.730	0.314
Summer	0.941	0.945	0.631	0.786	0.130
H-Densenet <sup>[24]</sup>	0.961	<b>0.965</b>	<b>0.722</b>	<b>0.824</b>	\
<b>FMD-Net(ours)</b>	0.959	0.960	0.662	0.803	0.253

综上所述,通过消融实验证明了这种使用三维信息辅助二维网络分割的思路以及提出的 AFM 多维信息融合模块对于整个模型性能提升的帮助。其中在 Precision 这个指标上面, Liu 等人在论文中提到了深层监督机制对于找到肝脏肿瘤的重要性,因此,在这个指标上面他们模型的性能是高于 FMD-Net 模型的。FMD-Net 对 LiTS 测试集的一些分割结果展示如图 4.1。其中图中上一面一行表示输入的用作测试的原始 CT 图像。第二行表示模型对于输入 CT 图像做出的肝脏及肝脏肿瘤分割结果,绿色是肝脏的预测结果,黄色是肿瘤的预测结果。由于官方给出的测试集是没有金标准的,所以此处无法提供相应的金标准作为对比。

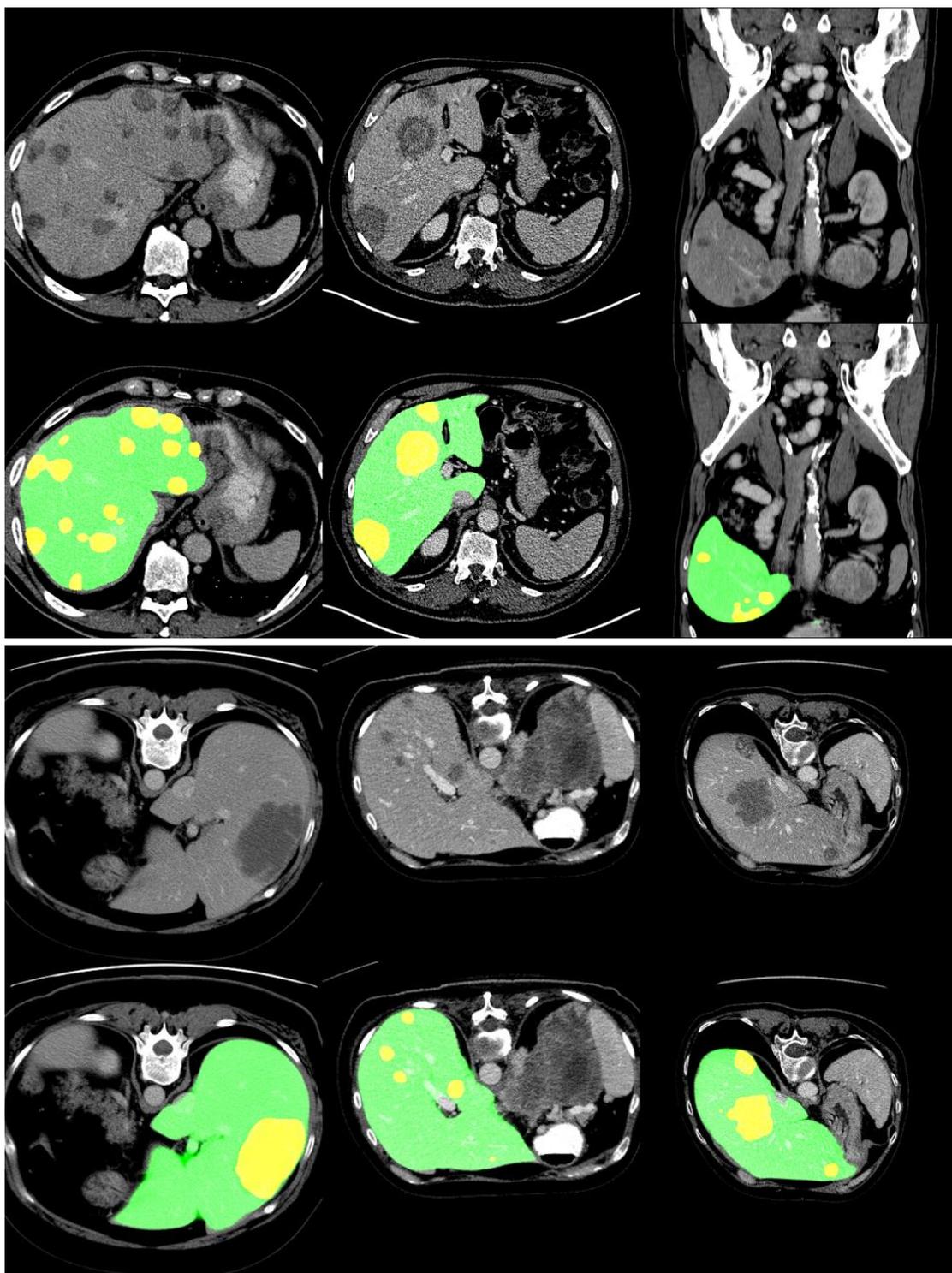


图 4.1 模型对于 LiTS 2017 测试集上的测试图像的预测结果

## 5 总结与展望

### 5.1 总结

本文研究的主要内容是通过深度学习的算法来从 CT 图像中全自动的分割肝脏及肝脏肿瘤区域来达到使用计算机辅助医生诊断治疗的目的。基于已有的二维和三维分割模型以及肝脏和肝脏肿瘤分割流程，本文所采用的基础模型为 2D U-Net 网络，在此基础上，为了利用空间上下文信息输入三维的数据块同时利用三维卷积核、连续卷积的方式来提取三维信息，之后为了利用三维信息帮助 2D U-Net 网络分割 CT 图像，又提出了利用 attention 机制和连续卷积的 AFM 模块来融合三维和二维信息，通过对 AFM 模块可视化可以发现该模块对于让模型专注于肝脏区域具有重要的促进作用。综上，本文提出的算法综合了传统的二维分割网络和三维分割网络的优势，解决了不同维度的分割模型在各自维度上的局限性以及全自动的从 CT 图像中分割肝脏和肝脏肿瘤的一些问题。

首先，本文中所有的模型和方法流程都是为了利用计算机从 CT 图像当中自动的分割出肝脏及肝脏肿瘤，面对实验对象是三维的医疗图像这个特点，传统的二维卷积神经网络天生的面临的一个很大的局限性就是无法充分利用数据特性和信息的特点。神经网络模仿的是人的思维模型，恰巧医生在勾画肝脏或者肝脏肿瘤时不可能只考虑一张切片上的 x、y 轴信息，同时会考虑三维医学数据上的 z 轴信息，这个“同时”很关键，因此 FMD-Net 不像其他的 2.5D 模型是分开训练二维或者三维的分割模型，我们的模型因为同时存在二维和三维卷积核的特征，可以同时接受二维和三维数据，充分利用了医学影像数据是三维数据的特点。

另外，面对三维医疗影像数据，并不是使用三维全卷积神经网络的模型就可以一劳永逸，可以从实验结果发现单从 CT 图像中分割肝脏及肝脏肿瘤这个任务而言，大多数的三维分割模型在一些主要评价指标上是不如二维的分割模型的。究其原因，三维卷积神经网络面对的主要问题是三维的卷积神经网络因为显存的限制，导致其感受野有限，通常只能专注于细节和局部特征。由于其超大量计算参数的特质，它在训练和测试时占用的显存是巨大的，这往往导致三维的网络不能采取更深的下采样因而限制了其感受野。在不考虑计算和显存性能的情况下，

三维分割网络因为可以结合图像层间信息，能够保证隔层图像 Mask 之间的一个变化连续性，效果可能会比二维的分割网络更好，但是三维卷积神经网络面临的另一个问题是超参数的设置，例如选取多少张连续相邻的切片才是合适但是对网络性能的影响非常大。此外，二维分割网络更加专注于平面上的细节分割，三维分割网络更加注重预测结果在空间上与目标的相似度。针对以上的三维分割网络的局限性，在不占用大量计算资源的前提下，使用了 AFM 模块融合了三维与二维的信息，既保证了在平面上分割的细节，又充分利用了三维信息辅助二维分割网络。

最后，本文提出的整个模型和方法都是端到端的，在之前的一些方法中已经有人意识到三维分割和二维分割网络的各自局限性。也有工作充分证明了将上一个阶段的二维预测结果放入三维分割模型中会取得更好的效果，但是这并不是一个端到端的分割模型。端到端的重要意义在于可以将二维信息和三维信息综合在一起对模型预测结果产生影响达到优化模型的目的。另一个重要的影响就是在以后将算法嵌入前段设备时需要小的模型，相对于多个模型，端到端的模型明显拥有明显的更少的参数量。

## 5.2 展望

本文提出的 FMD-Net 网络目的是为了作为计算机辅助诊断治疗的手段，虽然在 CT 图像中分割肝脏和肝脏肿瘤达到了不错的效果，但是将其投入到临床应用当中还有很多需要提升的地方，一个是模型的预测精度尤其是肝脏肿瘤这项任务的预测上，还远远达不到代替医生手工勾画的要求，另外在临床上还存在着医疗事故责任划分的问题，需要相关部门进一步完善相关法律。我们需要做的就是在未来进一步完善模型的精度。展望未来，针对全自动从 CT 图像中分割肝脏及肝脏肿瘤这个任务，主要有以下几点值得改进：

(1) 从 CT 图像识别肝脏肿瘤的意义大于肝脏，但是能精准的分割肝脏是分割肝脏肿瘤的前提。但是在含有病灶的 CT 影像数据中，大多数的病灶的体积是相当小的，与分割肝脏不同，在不断的下采样过程中，由于肝脏肿瘤的体积太小，很有可能已经丢失了肝脏肿瘤信息。针对这个问题可以采用深层监督机制，在不断的下采样过程中，使用侧输出的方式来约束每层卷积块的内容可以变向的

恢复下采样过程中的信息损失，在本文模型中可以采取的侧输出地方有很多，例如在三维卷积部分可以使用三维的侧输出，相对于从二维卷积下采样部分使用深层监督，三维的深层监督内容更加丰富。

(2) 模型本身的性能是一定的，但是一个好的图像预处理、训练技巧、测试方式对于实验结果影响也很大。本文中的图像预处理方式只是简单的调整了 HU 值，但是并不是每个病人的数据都适合这个 HU 值，因此可以通过增加模态的方式例如经过直方图均衡过后的 CT 图像，它的肿瘤信息可能更加明显。在训练中，我们只是简单的采取图像的旋转和翻转功能，但其实图像增强的手段远远不止这些，尤其是在肝脏肿瘤分割任务当中，一些论文当中采用的镜像变换、随机采取 patch 的方式十分适合这个任务，有效的达到了数据增广的目的。在测试过程当中，将模型输出一次的结果作为最终结果，只用了一些传统的后处理方式，但是如果可以使用训练过程当中数据增加的方式，每种数据变化输出一次结果，最后使用一个合适的准则来计算多个结果的重叠区域，这样的实验结果也能得到很大的提升，这也是模型可以提高的地方。

## 参考文献

- [1] A. Forner, JM. Llovet, J. Bruix, Hepatocellular carcinoma [J]. The Lancet, 2012, Vol. 379: 1245-55.
- [2] Patricia C. Valery et al. Projections of primary liver cancer to 2030 in 30 countries worldwide[J]. Hepatology, 2018, 67(2) : 600-611.
- [3] A M T C , A C H C . Potential Roles for Spectroscopic Coherent Raman Imaging for Histopathology and Biomedicine - ScienceDirect[J]. Neurophotonics and Biomedical Spectroscopy, 2019:547-570.
- [4] Tang Y, Yan K, Xiao J, et al. One Click Lesion RECIST Measurement and Segmentation on CT Scans[M]. 2020.
- [5] Gonzalez R C, Woods R E. Digital Image Processing[M]. 电子工业出版社, 2007.
- [6] 彭丰平, 鲍苏苏, 曾碧卿. 基于自适应区域生长算法的肝脏分割[J]. 计算机工程与应用, 2010 (33) :202-204.
- [7] 罗文村. 基于阈值法与区域生长法综合集成的图像分割法[J]. 现代计算机 (专业版), 2001 (05) :42-45.
- [8] Chen Y, Wang Z, Zhao W, et al. Liver Segmentation from CT Images Based on Region Growing Method[C]// International Conference on Bioinformatics & Biomedical Engineering. IEEE, 2009.
- [9] Goryawala M, Gulec S, Bhatt R, et al. A Low Interaction Automatic 3D Liver Segmentation Method Using Computed Tomography for Selective Internal Radiation Therapy[J]. Biomed Research International, 2014, 2014(8): 12.
- [10] Zhao Y, Zan Y, Wang X, et al. Fuzzy C-means clustering-based multilayer perceptron neural network for liver CT images automatic segmentation[C]// 2010 Chinese Control and Decision Conference. IEEE, 2010.
- [11] Kuo C L, Cheng S C, Lin C L, et al. Texture-based treatment prediction by automatic liver tumor segmentation on computed tomography[C]// International Conference on Computer. IEEE, 2017.
- [12] Weiwei W, Zhuhuang Z, Shuicai W, et al. Automatic Liver Segmentation on

- Volumetric CT Images Using Supervoxel-Based Graph Cuts[J]. *Computational and Mathematical Methods in Medicine*, 2016, (2016-4-5), 2016, 2016:9093721.
- [13]Lecun Y, Bottou L. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11):2278-2324.
- [14]Z. Xu, R. P. Burke, C. P. Lee, R. B. Baucom, B. K. Poulouse, R. G. Abramson, B. A. Landman. Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning [J]. *Medical image analysis*, 2015, 24(1), 18-27.
- [15]Rezaei M, Yang H, Meinel C. Deep Learning for Medical Image Analysis[J]. 2017.
- [16]Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 39(4):640-651.
- [17]Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. Springer, Cham, 2015.
- [18]Han X. Automatic Liver Lesion Segmentation Using A Deep Convolutional Neural Network Method. 2017.
- [19]Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation[J]. 2018.
- [20]F Milletari, Navab N, Ahmadi S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation[C]// 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016.
- [21]Christ P F, Elshaer M, Ettliger F, et al. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2016.
- [22]Zhang J, Xie Y, Zhang P, et al. Light-Weight Hybrid Convolutional Network for Liver Tumor Segmentation[C]// Twenty-Eighth International Joint Conference on Artificial Intelligence {IJCAI-19. 2019.
- [23]Tang Y, Tang Y, Zhu Y, et al. E<sup>2</sup>Net: An Edge Enhanced Network for Accurate Liver and Tumor Segmentation on CT Scans[J]. 2020.

- [24] Li Xiaomeng et al. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes[J]. IEEE transactions on medical imaging, 2018, 37(12): 2663-2674.
- [25] Han J, Moraga C. The Influence of the Sigmoid Function Parameters on the Speed of Backpropagation Learning[C]// International Workshop on Artificial Neural Networks: from Natural to Artificial Neural Computation. Springer-Verlag, 1995.
- [26] Hahnloser R, Sarpeshkar R, Mahowald M A, et al. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit[J]. Nature, 2000, 405(6789):947-951.
- [27] Y Jing, Mao Q, Liu D, et al. Method to Improve the Performance of Restricted Boltzmann Machines[J]. International Symposium on Neural Networks, 2018.
- [28] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks[J]. In Computer Vision—ECCV 2014, Springer, 2014,8689: 818–833.
- [29] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. JMLR.org, 2015.
- [30] Krhenbühl, Philipp, Koltun V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials[J]. 2012.
- [31] Kamnitsas K, Ledig C, Newcombe V F J, et al. Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation[J]. Medical Image Analysis, 2016, 36:61.
- [32] Chen L C, Yang Y, Wang J, et al. Attention to Scale: Scale-aware Semantic Image Segmentation[J]. 2015.
- [33] Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.
- [34] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision[J]. IEEE, 2016:2818-2826.
- [35] Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945;26(3):297–302.
- [36] Kingma D, Ba J. Adam: A Method for Stochastic Optimization[J]. Computer

- Science, 2014.
- [37] M. Bellver, K. K. Maninis, J. P. Tuset, X. G. Nieto, J. Torres, L. V. Gool. Detection-aided liver lesion segmentation using deep learning [J]. NIPS, 2017.
- [38] Liu T, Liu J, Ma Y, et al. Spatial feature fusion convolutional network for liver and liver tumor segmentation from CT images[J]. Medical Physics, 2020.
- [39] Bi L, Kim J, Kumar A, et al. Automatic Liver Lesion Detection using Cascaded Deep Residual Networks[J]. 2017.
- [40] Kaluva K C, Khened M, Kori A, et al. 2D-Densely Connected Convolution Neural Networks for automatic Liver and Tumor Segmentation[J]. 2018.
- [41] Carreira J, Zisserman A. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset[J]. IEEE, 2017.
- [42] Wang X, Han S, Chen Y, et al. Volumetric Attention for 3D Medical Image Segmentation and Detection[J]. 2020.
- [43] Bilic P, Christ P F, Vorontsov E, et al. The Liver Tumor Segmentation Benchmark (LiTS)[J]. 2019..
- [44] Kubota K, Makuuchi M, Kusaka K, et al. Measurement of liver volume and hepatic functional reserve as a guide to decision-making in resectional surgery for hepatic tumors[J]. Hepatology, 1997, 26(5):1176-81.
- [45] 李登旺, 王杰, 陈进琥, 等. 基于最优值搜索框架的 CT 序列图像肝脏分割[J]. 计算机学报, 2016, 039(007):1477-1489.

## 致谢

少年不管，流光如箭，因循不觉韶光换。青丝白发转瞬间，相比于刚走出校门的懵懂少年，鬓角上几根白发见证了硕士三年在兰州财经大学经历的充实生活。良师、益友是我这三年来最大的收获，借此我想向他们表示衷心的感谢。

我要感谢培养我的导师丁晓阳教授。丁老师事务繁忙，尤其是在抗击疫情期间挑起全校师生的安全重担，在工作岗位和家之间来回奔波、不解衣带。即便如此丁老师还是会在百忙之中关心我的学术、生活和个人的未来发展，丁老师对于我的教育从来都是循循善诱和一语中的，学生打心里感谢老师的栽培。

师傅领进门，修行在个人，但是何江萍副教授对于我们的学习指导从来都是面面俱到，从一手教会我们入门到后面对我们科研素质和思维方式的培养，真正的体现了什么是授人以渔。不止在学习，在生活上何老师也是一个优秀的人，他艰苦、自律、负责。他教导我们永远要试着接触更优秀的人，以后的生活教会我这不仅是一种生活方式，更是一种态度，不做井底之蛙，不要一叶障目。

感谢韩金仓韩教授，韩老师平时对于我们的关心和指导从来都不是溢于表面的。实验室的搭建，离不开韩老师的心血倾注。韩老师还经常给我们分享当前的相关领域的会议和实用的科研工具，让我们收益颇丰。

感谢李强教授，李老师是一个无论是科研和实践都十分卓越的人。李老师对我科研的帮助特别大，正是李老师的大力支持，我才能利用如此优越的计算机资源使我的实验顺利进行。

在此，我也要感谢给了我新知识的各位老师，正是你们的默默无闻的辛勤付出，我们兰州财经大学的信工学院才会有越来越广阔的天地。

感谢我的益友们。我的好兄弟、班长何多魁，他平时为我们班级任劳任怨，对兄弟们尽心尽力。我的情感导师边柯橙，乐于助人的董亚兰，认真努力的吴瑶还有其他同学们，可爱的师弟刘伟和黄建民还有其他师弟师妹们，一路上给予我重大帮助的学长刘天宇和师姐姜威威，还有帮助我一开始学会编程的高洪萍师姐，因为有他们我的研究生生涯才会多姿多彩。

感谢所有评审我论文的老师，感谢你们抽出宝贵的时间对我的论文进行指导，谢谢你们的批评与指正。

## 攻读硕士期间参与科研与研究成果

### 攻读硕士期间参加的科研项目：

- 1、参与国家自然科学基金项目“三维 CT 图像中肝脏肿瘤的计算机辅助诊断方法研究”（项目批准号：61661024）。

### 攻读硕士期间研究成果：

- [1]刘祥强,李强,何多魁,王文瑞,黄建民.基于自监督学习的卷积神经网络在 CT 图像中的肝脏自动分割研究[J].科学技术创新,2021(03):78-79.
- [2]姜威威,刘祥强,韩金仓.基于深度协同训练的肝脏 CT 图像自动分割方法[J].电子设计工程,2020,28(14):175-179.
- [3]王堃,刘祥强.基于 MSCA-UNet 模型的肝脏 CT 图像分割.移动通信,2020(8):00018-00021.