

引文格式:

贺宝春, 贾富仓. 基于组合 U-Net 网络的 CT 图像头颈放疗危及器官自动分割 [J]. 集成技术, 2020, 9(2): 17-24.

He BC, Jia FC. Head and neck CT segmentation based on a combined U-Net model [J]. Journal of Integration Technology, 2020, 9(2): 17-24.

基于组合 U-Net 网络的 CT 图像头颈放疗危及器官自动分割

贺宝春^{1,2} 贾富仓^{1,2}

¹(中国科学院深圳先进技术研究院 深圳 518055)

²(中国科学院大学深圳先进技术学院 深圳 518055)

摘要 CT 图像头颈分割面临着以下难点: CT 图像的低对比度导致边界不清, 图像扫描间距过大导致冠状面和矢状面图像分辨率低, 头颈中待分割的 22 个器官对于神经网络构建建模的需求不同, 且由于存在极小器官造成了类间不平衡。为解决上述问题, 该文提出一种 U-Net 组合模型——由 3 种 U-Net 模型组成, 分别是 2D U-Net 模型、3D U-Net 模型及 3D-small U-Net 模型。其中, 2D U-Net 模型用于厚层图像的分割, 3D U-Net 模型利用三维空间信息, 3D-small U-Net 模型用于分割最小的两个器官以解决类不平衡问题。该方法在 MICCAI 2019 StructSeg 头颈放疗危及器官分割任务中取得了第 2 名的成绩, 平均 DSC 系数为 80.66%, 95% 豪斯道夫距离为 2.96 mm。

关键词 头颈; 危及器官; 图像分割; nnU-Net; 组合模型

中图分类号 TG 156 文献标志码 A doi: 10.12146/j.issn.2095-3135.20191216001

Head and Neck CT Segmentation Based on a Combined U-Net Model

HE Baochun^{1,2} JIA Fucang^{1,2}

¹(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

²(Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China)

Abstract Head and neck (HaN) segmentation in CT image is difficult due to low contrast and large slice thickness that resulted in useless information in coronal and sagittal plane for some organs. In addition, complex and small organs have different requirements on neural network modeling. To achieve an accurate segmentation of 22 HaN organs, we combined three U-Net models. The first model was a 2D model that is advantage for dealing with thick slice images. The second model was a 3D model using a cropped input to

收稿日期: 2019-12-16 修回日期: 2020-02-10

基金项目: 国家自然科学基金项目 (U1613221)

作者简介: 贺宝春, 博士研究生, 研究方向为医学图像处理; 贾富仓 (通讯作者), 博士, 研究员, 研究方向为计算机辅助医疗干预, E-mail: fc.jia@siat.ac.cn.

cover most organs with original resolution in the transverse plane. The third model was a 3D-small U-Net model that focuses on the segmentation of two small organs together and uses a small region of interest (ROI) computed from the bounding box of 2D model segmentation. All the three models were trained using nnU-Net method. The final trained model was submitted through docker image to the StructSeg challenge. The leaderboard showed that the proposed method achieved the second place among all methods on ten unseen testing cases with an average Dice value of 80.66% and 95% Hausdorff distance of 2.96 mm.

Keywords head and neck; organ at risk; image segmentation; nnU-Net; model combination

1 引 言

放疗是癌症治疗的重要手段之一，术前针对肿瘤及普通器官放射剂量计划是手术的关键步骤。放疗计划的目的在于确保肿瘤接受足够剂量的辐射，并防止放疗时高危及器官正常细胞受到过多损伤^[1-2]。危及器官对辐射非常敏感，如视神经和视交叉不能接受太多放疗，否则病人容易失明。而放疗计划的关键一步就是高危及器官的边界识别。因此，高危及器官的自动分割有助于减少医生放疗计划的工作量，无论是从时间还是经济角度都能够减少放疗的整体代价。此外，CT头颈高危及器官的分割存在对比度低、器官边界不清、图像扫描间距过大导致的冠状面和矢状面分辨率模糊等问题，其中最主要的是过小的器官视交叉和脑垂体带来的类不平衡问题。

头颈分割方法主要包括传统的多图谱方法^[3]、基于形状模型的方法^[4]及近些年基于深度学习的方法^[5-10]。其中，传统基于配准的多图谱方法对于微小器官分割精度很低。基于形状模型方法的缺点是局部结构对比度低导致模型难以收敛到结构边界。深度学习方法能够非线性刻画多个器官的局部和全局特征，分割精度更高，但针对头颈 CT 图像的分割，深度学习方法主要的挑战是极小器官带来的类不平衡问题。Gao 等^[7]和 Zhu 等^[11]提出使用焦点损失代价函数解决类不平衡问题，并在网络中减少池化次数，降低下采样对微

小器官分割的副作用。Gao 等^[7]提出的 FocusNet 使用额外的网络去单独训练微小器官，以解决类不平衡问题。

目前，绝大多数医学图像深度学习分割方法都是基于全卷积网络 (Fully Convolutional Network, FCN)^[12]或 U-Net 网络结构^[13]，根据分割图像模态、器官或肿瘤以及图像维度等特点，在网络的宽度、深度、损失函数、预处理方法以及注意力机制等方面进行个性化设计，以获得最优分割效果。对于尺寸较大的三维图像分割来说，内存是制约分割精度的关键因素。在输入到三维网络前，需要对三维图像进行裁剪或下采样，这导致输入信息必然有所损失。为解决上述问题，已有的方法通过并行或串行方式提取上下文信息和高分辨率信息。如 Kamnitsas 等^[14]提出一种两路并行 FCN 方法解决脑肿瘤分割问题，其中一路使用下采样但具有较多的上下文信息，另一路则使用原始分辨率的局部感兴趣区域作为输入，两路特征最后通过裁剪形式进行信息融合共同决策。Zhou 等^[15]提出一种串行的固定点胰腺分割模型，即以一个固定点提取感兴趣区域 (Region of Interest, ROI) 进行分割，得到的分割结果再进行更精确的 ROI 提取，如此迭代直到收敛。Han^[16]和 Li 等^[17]则都使用了粗分割结果提取肝脏的感兴趣区域。Isensee 等^[18]提出的 nnU-Net (no-new U-Net) 方法包含 3 种模型：2D U-Net 模型、3D U-Net 模型和级联的 3D U-Net

模型。其中, 级联的 3D U-Net 模型使用了下采样图像的粗分割结果作为高分辨率输入网络的一个通道输入, 以此实现上下文信息与高分辨率信息的兼顾。该方法最终在 MICCAI 2018 Medical Decathlon Segmentation 竞赛的 13 项任务中取得了最佳分割精度, 表明得到良好训练的 U-Net 结构即可获得有竞争力的结果。

本文方法的提出建立在上述已有方法的研究成果之上, 主要基于 nnU-Net 网络的训练策略, 针对头颈 CT 图像的特点, 提出新的图像裁剪和预处理方法, 并使用独立的网络训练最小的两个器官以解决类不平衡问题。

2 方 法

本文提出一种组合 U-Net 模型, 具体如图 1 所示。其中, 2D U-Net 模型以原始横断面切片作为输入, 3D U-Net 模型则将图像裁剪作为初始图像, 3D-small U-Net 模型则通过 2D U-Net 模型分割结果计算最小器官的中心点, 并以此提取最小器官的 ROI。在预测阶段, 根据器官在模型中的表现对比, 确定最终的组合策略。最终分割结果的不同器官来自于不同的分割模型, 如图 1 所

示, 标签值为 4、5、6、15、16、19 的器官来自于 2D 模型, 标签值为 8 和 11 的器官的分割结果来自于 3D-small 模型, 其他器官的分割结果来自于 3D 模型。选取该组合策略的具体原因参见结果部分中对不同模型、不同器官的分割结果的分析。

2.1 图像预处理

图像采用 z-score 归一化方法, 不进行任何重采样操作。其中, 3D U-Net 模型的裁剪操作定义为: 以 $(\frac{1}{4}S_x, \frac{1}{4}S_y, 0)$ 起始点, 裁剪尺寸为 $(\frac{1}{2}S_x, \frac{1}{2}S_y, S_z)$ 的区域。这里用 (S_x, S_y, S_z) 表示图像的原始尺寸, 其目的是去除不含分割器官对象的信息区域, 如左右胳膊、身体前方空气及后部扫描床, 从而缩小图像尺寸, 最终在训练 3D 模型时能够在不进行下采样的同时覆盖大部分器官区域, 具体如图 2 所示。

3D-small 模型主要针对最小的两个器官: 视神经交叉及脑垂体, 主要原因是这两个器官的整体体积非常小, 很容易出现类不平衡问题。此外, 这两个神经在空间上彼此距离最近, 因此将两者同时训练。本文在裁剪时, 首先计算二者在

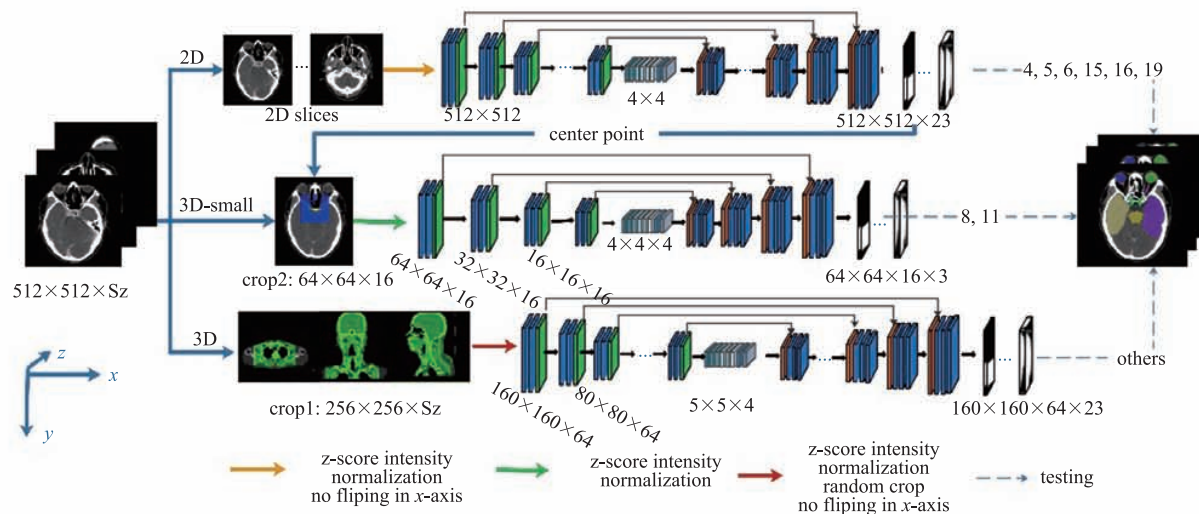


图 1 本文提出的组合 U-Net 模型算法流程

Fig. 1 Illustration of the proposed combined U-Net model

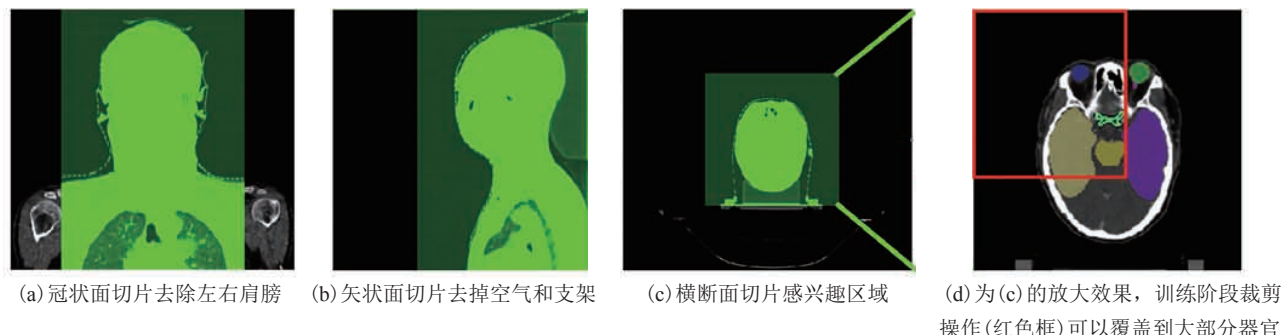


图2 3D模型的感兴趣区域(绿色部分)

Fig. 2 The region of interest in 3D-model (Green)

二维模型上的分割结果,接着计算分割结果的中心位置,并以此中心位置裁剪大小为(64, 64, 16)的区域。该尺寸是二者器官大小的3倍,这样包括一定的背景信息有助于二者的定位识别(见图3),同时不容易因分割结果导致裁剪出的区域只含部分器官。

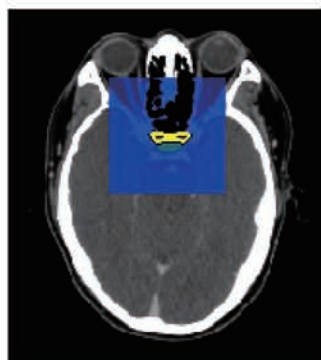


图3 3D-small模型的裁剪图示

Fig. 3 The illustration of cropping ROI in 3D-small model

2.2 网络结构

本文采用的3种模型皆为U-Net架构,图1所示的架构给出了网络的初始输入尺寸,根据最小特征图的尺寸可以推断出网络深度及池化次数。根据内存实际情况,3个网络的初始特征图设置为30,最大三维网络的特征图个数为320,二维网络的最大特征图个数为480,每个卷积层后面加入dropout和batch norm操作,激活函数采用LRelu,池化采用最大值采样,上采样采用线性插值法。整体结构及参数参考于Isensee等^[18]提出

的nnU-Net。

2.3 网络训练

本文网络采用pytorch实现。训练时,2D、3D-small、3D模型分别采用的batch size为11、2、2,优化器为adam,初始学习率设为 3×10^{-4} ,采用pytorch中的平均滑动指数更新学习率(学习率低于 10^{-6} 时模型训练自动停止),训练迭代次数约150 000次,训练消耗显存约为12 Gb。训练2D、3D-small、3D模型分别需要30 h、4 h、60 h。训练的损失函数为Dice与交叉熵之和,具体训练方法参见Isensee等^[18]提出的nnU-Net中2.3小节。

3 实验

数据来源自MICCAI2019 StructSeg 2019 (<https://structseg2019.grand-challenge.org/>)分割竞赛,数据的尺寸为 $512 \times 512 \times (315 \sim 456)$,扫描层间距为 $(0.977 \sim 1.184) \text{ mm} \times (0.977 \sim 1.184) \text{ mm} \times 3 \text{ mm}$ 。其中,训练数据为50套,测试数据集为10套。此外,所有参赛者均不可访问测试数据集,而是通过将训练好的分割模型以docker镜像方式提交给竞赛主办方,由主办方测试并给出分割结果,且分割每套图像的时间不超过2 min,因此测试结果具有公平性和说服力。

本文算法数据增强采用 nnU-Net 数据增强方式, 包括弹性形变、任意旋转、镜像翻转、Gamma 变换、比例变换。从图 4 可以看出, 大

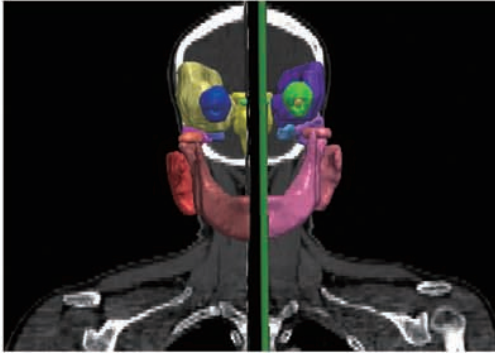


图 4 待分割器官关于矢状面对称图示

Fig. 4 The organs are sagittal plane symmetrical

部分器官相对于矢状面是对称的, 因此本文在训练 2D 和 3D 模型时, 不使用 x 轴的镜像翻转。

本文给出两个结果, 一个结果是训练集上的五轮交叉验证结果, 一个结果是将所有训练数据进行训练后的模型以参赛形式提交给竞赛组织者的结果, 具体见第 4 小节。评价标准采用典型的 DSC 系数及 95% 豪斯道夫 (HD) 距离, 具体定义见官网 <https://structseg2019.grand-challenge.org/Evaluation/>。

4 结 果

表 1 给出了不同模型的分割结果比较, 并与

表 1 不同模型在 22 个头颈器官上的分割 DSC 值对比结果 (加粗为分割较好)

Table 1 Results of different models on 22 HaN organs measured by DSC (the bold fonts indicate best)

标签	器官名称	模型分割结果					
		2D	3D	3D-small	2D-nnU-Net	FocusNet ^[7]	多图谱 ^[19]
1	脑干	0.866	0.873	--	0.854	0.858	0.806
2	左眼	0.886	0.895	--	0.887	0.876	0.787
3	右眼	0.884	0.888	--	0.882	0.912	0.839
4	左晶状体	0.770	0.740	--	0.780	0.808	0.240
5	右晶状体	0.806	0.796	--	0.786	0.790	0.269
6	左视神经	0.708	0.674	--	0.687	0.639	0.477
7	右视神经	0.700	0.714	--	0.683	0.617	0.485
8	视神经交叉	0.497	0.450	0.668	0.492	0.638	0.548
9	左侧颞叶	0.840	0.847	--	0.836	--	--
10	右侧颞叶	0.833	0.838	--	0.837	--	--
11	脑垂体	0.627	0.542	0.776	0.634	0.769	0.446
12	左侧腮腺	0.841	0.861	--	0.829	0.772	0.806
13	右侧腮腺	0.813	0.851	--	0.823	0.800	0.807
14	左侧内耳	0.847	0.855	--	0.845	--	--
15	右侧内耳	0.828	0.815	--	0.834	--	--
16	左侧中耳	0.855	0.841	--	0.862	0.567	0.564
17	右侧中耳	0.834	0.840	--	0.845	0.522	0.562
18	左颞下颌关节	0.741	0.759	--	0.748	0.584	0.469
19	右颞下颌关节	0.782	0.771	--	0.761	0.572	0.503
20	脊椎	0.811	0.832	--	0.813	0.874	0.824
21	左侧下颌骨	0.902	0.903	--	0.896	--	--
22	右侧下颌骨	0.905	0.908	--	0.894	--	--
	平均值	0.799	0.795		0.796	0.725	0.623

注: “--”表示分割结果不存在

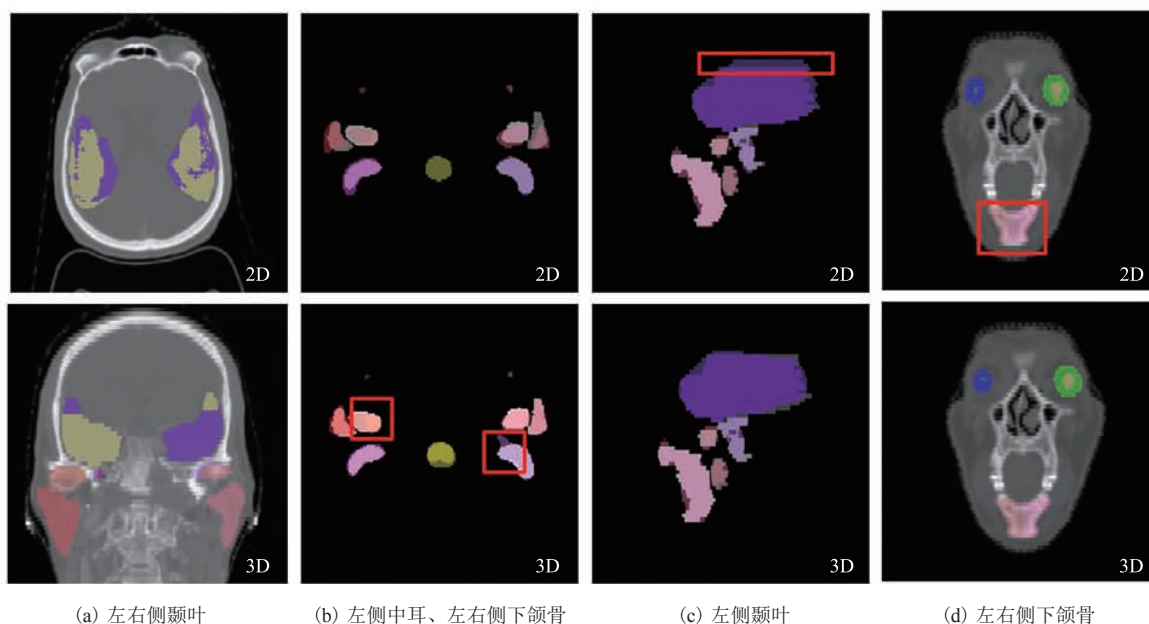
目前常用的多图谱方法及最前沿的 FocusNet 方法进行了比较。从表 1 可以明显看出, 基于深度学习方法在分割精度上优于传统的多图谱方法, 尤其是对如脑垂体(标签 11)、中耳(标签 16 和 17)、晶状体(标签 4 和 5)等微小器官。大多数器官在 3D 模型的表现优于 2D 模型, 少数器官主要依赖横断面切片进行识别分割, 这些器官在 2D 模型中表现更好。同时, 本文针对两个最小器官训练的 3D-small 模型相比于和所有器官同时训练精度提升很多。2D 与原始 2D-nnU-Net 之间的区别是 2D 模型根据器官对称特点在 x 轴不适用镜像翻转数据增强, 而原始的 2D-nnU-Net 模型则使用了镜像翻转。本文也根据器官在不同模型表现的不同, 依据此表制定了最后的模型组合策略, 具体如图 1 所示。这些对比结果可以从图 5 的定性分割具体看出来。图 5(a)中使用了 x 轴上的镜像翻转后, 其中左侧颞叶(标签为 9)和右侧颞叶(标签为 10)容易被混淆。图 5(c)中

第一行红色区域表示左侧颞叶在 2D 模型中产生的过分割。从图 5(d) 2D 模型与 3D 模型对比结果可以看出, 2D 模型分割的左侧下颌骨(标签为 21)和右侧下颌骨(标签为 22)在冠状面上难以分清。图 5(b)则指出了一些器官在 2D 模型中的分割效果优于 3D。

表 2 为不同参赛队伍在看不见的测试集上的结果比较。从表 2 可以看出, 本文方法无论在 DSC 还是在 95% HD 上都取得了第 2 名的成绩, 同时在 DSC 系数上比 nnU-Net 的器官分割平均提高了 0.78%, 表明本文方法的优越性。此外, 竞赛规定每套数据的分割时间不超过 2 min, 本文提出的算法分割一套数据只需要 60 s, 达到竞赛要求。

5 讨论

本文提出一种全自动的头颈 CT 放疗高危及



注: 图中右下角的 2D 和 3D 表示分割结果来自 2D 模型或 3D 模型; (a) 组图为在数据增强中使用 x 轴上的镜像翻转结果; (b)、(c) 和 (d) 组图为 2D 模型和 3D 模型对比结果(红色框表示分割较差的区域)

图 5 定性分割结果比较

Fig. 5 Qualitative results from model

表 2 不同参赛者在 SturctSeg 2019 竞赛测试集结果比较

Table 2 Comparison results from different teams on testing set in StructSeg 2019

参赛队	DSC	DSC 排名	95% HD (mm)	95% HD 排名	最终排名
AIView_sjtu	0.810 9	1	2.804 5	1	1
CAMI_SIAT(ours)	0.806 6	2	2.955 5	2	2
UESTC_501	0.805 2	3	3.043 0	4	3
nnU-Net	0.798 8	5	3.003 4	3	4

器官分割算法, 且获得了较好的分割精度。首先, 与传统的多图谱方法相比, 本文提出的方法精度有了很大的提升, 尤其在微小器官的分割方面。如表 1 所示, 脑垂体(标签 11)、左右中耳(标签 16 和 17)、左右晶状体(标签 4 和 5)微小器官在本文方法上的分割 DSC 值为 0.776、0.855、0.840、0.770、0.806, 远远超过多图谱方法^[19]的 0.446、0.564、0.562、0.240 和 0.269, 和其他基于深度学习的 Focus-net^[7]、nnU-Net^[18]方法的结果相差不大, 这说明了深度学习方法的优越性。其次, 从不同参赛者在测试集上的结果比较来看, 本文方法基于 nnU-Net 方法且使用了与 nnU-Net 方法同样的训练方法, 但结果比原始的 nnU-Net 方法 DSC 值提高了 0.78%(表 2 所示)。这得益于本文方法根据分割图像的特点, 设计出了实用且有效的数据预处理方法及数据增强的方法。通过图 5 的定性对比结果可以看出, 不使用翻转数据增强能够防止左右器官的混淆。因此, 本文重新设计了 3 个 U-net 网络组合型的网络结构。从表 1 中 3 个网络的分割结果可以看出, 不同模型有着各自擅长的分割器官, 如右晶状体和左侧视神经在 2D 模型上的分割 DSC 值分别为 0.806 和 0.708, 而在 3D 模型上的分割结果仅为 0.796 和 0.674; 左眼、左侧腮腺、右侧腮腺、左颞下颌关节、脊椎在 3D 模型上的分割 DSC 值分别为 0.895、0.861、0.851、0.759、0.832, 远远超过了 2D 模型上 0.886、0.841、0.813、0.741、0.811, 验证了本文所提出组合模型方法的实用性, 证明了本文算法思路的可行性。

但本文算法比第 1 名的算法的平均 DSC 值低 0.43%。表 2 中第 1 名算法使用了 2D 和 3D 模型的组合特征融合提取, 而本文算法仅是 2D 与 3D 模型的结果组合, 相比之下第 1 名的方法特征提取更有效。未来本文将在端到端的 2D 与 3D 特征同时融合提取方面进行改进。

6 结 论

本文提出一种全自动头颈风险器官三组合 U-Net 模型分割方法。具体是在 nnU-Net 模型架构基础上, 根据图像特点设计合适的图像预处理、裁剪方法、数据增强方法及网络参数, 最终算法取得了前沿精度, 在 22 个头颈风险器官上分割的平均 DSC 系数为 80.66%, 95% 豪斯道夫 (HD) 距离为 2.96 mm。本文方法结果表明, 得出组合 U-Net 网络就可取得有竞争力的结果, 这将为其他医学图像分割提供一种实用有效的解决方案。

参 考 文 献

- [1] Dawson LA, Sharpe MB. Image-guided radiotherapy: rationale, benefits, and limitations [J]. *The Lancet Oncology*, 2006, 7(10): 848-858.
- [2] Tubiana M. Can we reduce the incidence of second primary malignancies occurring after radiotherapy? A critical review [J]. *Radiotherapy and Oncology*, 2009, 91(1): 4-15.
- [3] Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: a survey [J]. *Medical Image*

- Analysis, 2015, 24(1): 205-219.
- [4] Raudaschl PF, Zaffino P, Sharp GC, et al. Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015 [J]. *Medical Physics*, 2017, 44(5): 2020-2036.
- [5] Ren XH, Xiang L, Nie D, et al. Interleaved 3D CNNs for joint segmentation of small-volume structures in head and neck CT images [J]. *Medical Physics*, 2018, 45(5): 2063-2075.
- [6] Tappeiner E, Pröll S, Hönig M, et al. Multi-organ segmentation of the head and neck area: an efficient hierarchical neural networks approach [J]. *International Journal of Computer Assisted Radiology and Surgery*, 2019, 14(5): 745-754.
- [7] Gao YH, Huang R, Chen M, et al. FocusNet: imbalanced large and small organ segmentation with an end-to-end deep neural network for head and neck CT images [C] // *The Medical Image Computing and Computer Assisted Intervention Society*, 2019: 829-838.
- [8] Tong N, Gou SP, Yang SY, et al. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks [J]. *Medical Physics*, 2018, 45(10): 4558-4567.
- [9] Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks [J]. *Medical Physics*, 2017, 44(2): 547-557.
- [10] Lei Y, Fu YB, Wang TH, et al. Deep learning in multi-organ segmentation [Z/OL]. arXiv Preprint, arXiv: 2001.10619, 2020.
- [11] Zhu WT, Huang YF, Tang H, et al. Anatomynet: deep 3D squeeze-and-excitation U-Nets for fast and fully automated whole-volume anatomical segmentation [Z/OL]. arXiv Preprint, arXiv:1808.05238v2, 2018.
- [12] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C] // *The IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3431-3440.
- [13] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C] // *The Medical Image Computing and Computer Assisted Intervention*, 2015: 234-241.
- [14] Kamnitsas K, Ledig L, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation [J]. *Medical Image Analysis*, 2017, 36: 61-78.
- [15] Zhou YY, Xie LX, Shen W, et al. A fixed-point model for pancreas segmentation in abdominal CT scans [C] // *The Medical Image Computing and Computer Assisted Intervention*, 2017: 693-701.
- [16] Han X. Automatic liver lesion segmentation using a deep convolutional neural network method [Z/OL]. arXiv Preprint, arXiv:1704.07239, 2017.
- [17] Li XM, Chen H, Qi XJ, 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.
- [18] Isensee F, Petersen J, Kohl SAA, et al. nnU-Net: breaking the spell on successful medical image segmentation [Z/OL]. arXiv Preprint, arXiv: 1904.08128, 2019.
- [19] Avants BB, Epstein CL, Grossman M, et al. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain [J]. *Medical Image Analysis*, 2008, 12(1): 26-41.