

## 基于U-net的心脏自动勾画模型的临床应用及改进

常艳奎<sup>1</sup>, 彭昭<sup>1</sup>, 周解平<sup>1,2</sup>, 皮一飞<sup>3</sup>, 吴昊天<sup>4</sup>, 吴爱东<sup>2</sup>, 徐榭<sup>1,4</sup>, 裴曦<sup>1,4</sup>

1. 中国科学技术大学放射医学物理中心, 安徽 合肥 230025; 2. 中国科学技术大学附属第一医院放疗科, 安徽 合肥 230001; 3. 郑州大学第一附属医院放疗科, 河南 郑州 450052; 4. 安徽慧软科技有限公司, 安徽 合肥 230088

**【摘要】目的:**拟分析基于不同医院数据的心脏自动勾画模型在临床应用中的适用性及其改进方法。**方法:**首先,建立基于U-net和Inception模块的心脏自动勾画网络。其次,收集不同治疗中心的患者数据:中国科学技术大学附属第一医院65例(数据1)、MICCAI2019比赛数据50例(数据2)、数据1和2的混合数据(数据3)、郑州大学第一附属医院50例(数据4)和郑州大学第一附属医院100例(数据5),分别训练得到模型1~5。然后,以郑州大学第一附属医院59例患者作为测试集,使用形状相似性系数(DSC)评估该测试集在不同模型上的分割精度,比较模型之间的差别。最后,将模型3作为心脏预训练模型,采用数据5进行模型再训练,分别测试3组实验(20例/次×5次、10例/次×10次、5例/次×20次)对心脏预训练模型的改进情况。**结果:**测试集在模型1~5中的平均DSC为0.926、0.932、0.939、0.941和0.950。在再训练过程中,模型在20例/次×5次的实验中表现更稳定。**结论:**基于不同医院的数据训练模型在心脏自动勾画的临床应用上表现存在差异,使用本地医院数据进行训练的模型预测精度更高。对于非本地数据训练的模型,基于本地数据再训练可以有效提高模型预测的精度,其中以20例/次的再训练方式效果较好。

**【关键词】**U-net; 心脏; 自动勾画; 形状相似性系数

**【中图分类号】**R811.1; R318

**【文献标志码】**A

**【文章编号】**1005-202X(2020)10-1218-06

## Clinical application and improvement of U-net-based model for automatic segmentation of the heart

CHANG Yankui<sup>1</sup>, PENG Zhao<sup>1</sup>, ZHOU Jieping<sup>1,2</sup>, PI Yifei<sup>3</sup>, WU Haotian<sup>4</sup>, WU Aidong<sup>2</sup>, XU Xie<sup>1,4</sup>, PEI Xi<sup>1,4</sup>

1. Center of Radiological Medical Physics, University of Science and Technology of China, Hefei 230025, China; 2. Department of Radiation Oncology, the First Affiliated Hospital of University of Science and Technology of China, Hefei 230001, China; 3. Department of Radiation Oncology, the First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China; 4. Anhui Wisdom Technology Co. Ltd., Hefei 230088, China

**Abstract: Objective** To investigate the clinical applicability of the model established based on the data from different hospitals for the automatic segmentation of the heart and to discuss the methods to improve the model. **Methods** A network based on U-net and Inception module was firstly constructed for the automatic segmentation of the heart, and the clinical data from different hospitals were collected, including 65 cases from the First Affiliated Hospital of University of Science and Technology of China (data 1), 50 cases from MICCAI2019 match data (data 2), the mixed data of data 1 and data 2 (data 3), 50 cases from the First Affiliated Hospital of Zhengzhou University (data 4) and 100 cases from the First Affiliated Hospital of Zhengzhou University (data 5). The collected data were trained for obtaining 5 different models. Then, with the clinical data of another 59 patients from the First Affiliated Hospital of Zhengzhou University as test set, the segmentation accuracies of test set on different models were evaluated using Dice similarity coefficient (DSC), and the differences in segmentation accuracies among different models were also compared. Finally, model 3 was used as a pre-trained model of the heart, and the model was retrained with data 5. Three groups of experiments (20 cases each time × 5 times, 10 cases each time × 10 times, 5 cases each time × 20 times) were carried out to observe the improvement of the pre-trained model. **Results** The average DSC of the test set based on models 1 to 5 was 0.926, 0.932, 0.939, 0.941 and 0.950, respectively. During the retraining of the pre-trained model of the heart, the model was more stable in the experiment of 20 cases each time × 5 times. **Conclusion** The trained model established based on the data from different hospitals has different performances in the automatic segmentation

**【收稿日期】**2020-04-05

**【基金项目】**国家自然科学基金(11575180);安徽省自然科学基金(1908085MA27);安徽省重点研究与开发计划(1804a09020039)

**【作者简介】**常艳奎, 硕士, 主要从事医学图像分割、深度学习等研究, E-mail: cykanhui@mail.ustc.edu.cn

**【通信作者】**裴曦, 博士, 副教授, 主要从事医学物理、人工智能和医学影像等研究, E-mail: xpei@ustc.edu.cn

of the heart, and the model trained with local hospital data has a higher prediction accuracy. For the model based on non-local data training, the retraining with local data can effectively improve the accuracy of model prediction, in which the retraining with 20 cases each time has the optimal performance.

**Keywords:** U-net; heart; automatic segmentation; Dice similarity coefficient

## 前言

随着放疗技术的发展,调强放疗已成为现代放疗的主流技术。为了更好地保护危及器官,降低正常组织并发症的发生,在制定治疗计划过程中需要准确、快速地勾画出危及器官。但是传统定位图像上的手工勾画占用医生大部分时间,且易形成不同勾画者间的差异<sup>[1-3]</sup>,甚至同一名医生在不同时间段勾画的结果也会不同<sup>[4-6]</sup>,这给临床治疗质量带来了很大的负面影响,难以满足未来自适应放疗发展的要求。鉴于以上原因,基于深度学习的自动勾画技术受到了越来越多的关注<sup>[7-11]</sup>,并已经在危及器官的自动勾画方面取得不错的精度<sup>[12-13]</sup>,本课题组使用改进的2D-U-net神经网络结构实现心脏自动分割。在临床实践过程中发现,已经训练好的心脏自动勾画模型直接应用到其他医院难以达到在原先测试集上的精度,也就是说同一个模型在不同医院之间存在差异。基于此,本研究拟探讨基于不同医院数据训练的心脏自动勾画模型的临床应用,比较不同模型的差异,并通过使用医院的本地数据再训练以达到稳定状态,从而实现心脏自动勾画在临床应用上的普适性。

## 1 材料与方法

### 1.1 数据来源

本研究收集来自不同治疗中心的含有心脏勾画信息的DICOM文件,包含患者在放疗前的计算机断层扫描(Computed Tomography, CT)文件和有多年临床经验的医生勾画完成的轮廓(RT Structure, RS)文件,以医生勾画的心脏轮廓作为模型训练的标签和评估自动勾画性能的标准。

具体的数据包括:(1)中国科学技术大学附属第一医院数据65例,分辨率为1.270 mm×1.270 mm,层厚为2.50 mm;(2)MICCAI2019放疗规划自动结构勾画国际挑战赛比赛数据50例,分辨率为1.172 mm×1.172 mm,层厚为5.00 mm;(3)郑州大学第一附属医院数据159例,分辨率为0.977 mm×0.977 mm,层厚为5.00 mm。

### 1.2 自动勾画方法

本研究得到的数据均是以DICOM格式存储的患者数据。首先使用CT分类网络对患者的CT进行分类,然后使用自动勾画网络在包含心脏的CT层上勾画出心脏的轮廓。

**1.2.1 CT分类网络** 使用Python脚本读取患者的DICOM文件可以得到患者的三维数据,其中每张CT的大小为512×512。首先找出每层CT中身体的中心,以此为中心将原先的512×512的图像裁剪成224×224的图像;然后利用卷积神经网络对每张CT进行分类,网络结构如图1所示,网络包括若干卷积层和Inception模块;最后将患者所有的CT分为两类,即含有心脏的CT与不含有心脏的CT。通过CT分类网络筛选出可能包含心脏的CT层可以提高训练效率,而且能够更加准确地完成心脏的自动勾画。

**1.2.2 自动勾画网络** 通过CT分类网络将含有心脏的CT选出,然后找出每层CT中身体的中心,以此为中心将原先的512×512的图像裁剪成256×256的图像,一方面可以提高心脏的勾画效果,另一方面可以有效减少模型参数。

本研究使用的心脏自动勾画网络是基于U-net和Inception模块改进的全卷积神经网络<sup>[14-19]</sup>,网络模型结构如图2所示。将U-net的部分长连接替换为Inception模块,增加网络深度和宽度的同时减少参数,通过多个卷积核提取图像不同尺度的信息,最后进行融合,可以得到图像更好的表征。模型训练过程中,以1-DSC作为loss函数,训练迭代次数设为100, batch的大小设为4,使用Adamax优化方法。

### 1.3 实验方法

根据数据的来源和数量,本研究将数据划分为5组,并使用图2中的网络进行训练得到基于同一个网络不同训练数据的预测模型,具体流程如下:(1)使用中国科学技术大学附属第一医院65例患者数据(数据1)进行训练得到模型1;(2)使用MICCAI 2019放疗规划自动结构勾画国际挑战赛比赛数据50例患者数据(数据2)进行训练得到模型2;(3)使用训练数据1和训练数据2的混合数据(数据3)进行训练得到模型3;(4)使用郑州大学第一附属医院的50例患者数据(数据4)进行训练得到模型4;(5)使用郑州大学第一附属医院的100例患者数据(数据5)进行训练得到模型5。然后将未参与网络训练的59例郑州大学第一附属医院的患者数据作为测试集,分别计算测试集在模型1~5上的形状相似性系数(Dice Similarity Coefficient, DSC),并用配对样本的非参数检验-Wilcoxon符号秩检验对数据进行显著性差异检

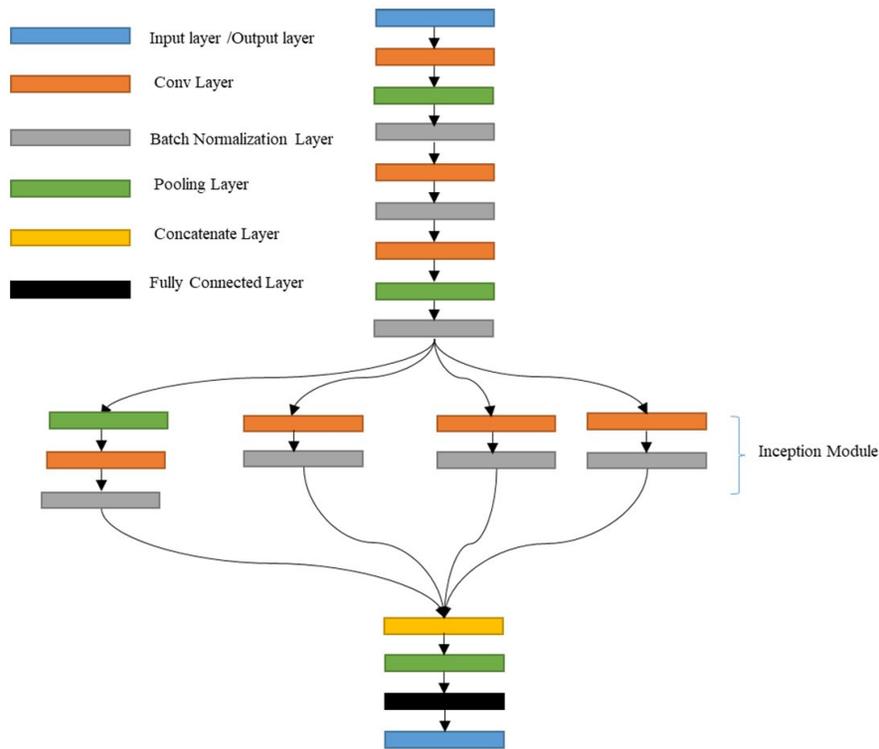


图1 CT分类网络模型结构

Fig.1 CT classification network model structure

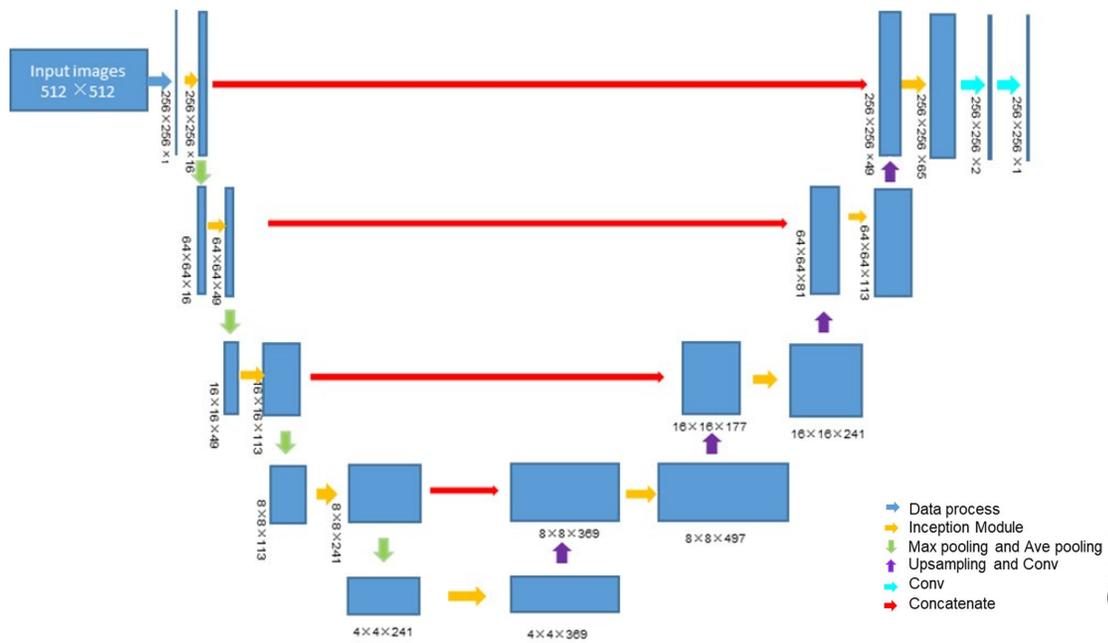


图2 心脏自动勾画网络模型结构

Fig.2 Structure of the network model for the automatic segmentation of the heart

测。在训练数据量接近的前提下,为了表明不同数据中心的差异,将模型1、模型2和模型4预测得到的DSC进行对比分析,并分别计算模型1和模型2得到的DSC与模型4得到的DSC的差异。为了进一步说明模型存在改进的空间,观察并比较测试集在5个不同模型上的差异,将模型1~4得到的DSC与模型5得

到的DSC进行显著性差异分析。

为了改进模型在不同医院临床应用时表现有一定差异的问题,本研究使用一定数量的本地数据对心脏的预训练模型进行再训练,观察预训练模型在临床应用中的改善情况。将模型3作为预训练模型,然后将训练数据5作为本地数据,进行模型的临床应

用改进。在实验过程中,考虑到使用本地数据每次再训练使用的病例数不同,可能会对模型的收敛速度和预测精度造成影响,所以本研究将100例患者的再训练数据分3组:第一组每次再训练采用20例,共进行5次;第二组每次再训练采用10例,共进行10次;第三组每次再训练采用5例,共进行20次。再训练过程中,loss函数、模型训练的迭代次数均与之前保持一致,分别计算每次更新训练模型后在测试集上的DSC。

#### 1.4 评价标准

本研究以资深医生手工勾画的心脏轮廓为“金标准”,使用DSC评价网络自动勾画的心脏轮廓的准确性<sup>[20]</sup>。DSC为0~1,数值越大表示自动勾画与金标准越相似,说明预测精度越高。

$$DSC(A,B) = 2 * \frac{|A \cap B|}{|A| + |B|}, 0 \leq DSC \leq 1 \quad (1)$$

其中, $A$ 为医生手工勾画的心脏区域, $B$ 为自动勾画的心脏区域。

#### 1.5 硬件配置

硬件处理器: Intel(R) Core(TM) i7-8700 CPU@3.70 GHz; 内存(RAM): 16 GB; GPU: NVIDIA GeForce GTX 1060 6 GB; 开发工具: Python Tensorflow 框架。

#### 1.6 统计学分析

本研究对基于不同训练模型预测得到的DSC进行配对样本的非参数检验-Wilcoxon符号秩检验, $P<0.05$ 为差异具有统计学意义。

## 2 结果与讨论

### 2.1 不同训练模型之间的差异

**2.1.1 训练数据数量接近时,不同医院训练模型结果的差别** 模型1、模型2和模型4预测得到的DSC如表1所示,可以看到训练数据量接近的情况下(分别是65例、50例、50例),DSC的比较结果为:模型4>模型2>模型1。

从表1可以看出,基于郑州大学第一附属医院患者数据训练得到的模型在59例郑州大学第一附属医院测试集上表现效果最好,而不使用本地数据进行训练的模型的表现效果劣于本地数据训练的模型。其中,基于中国科学技术大学附属第一医院数据得到的训练模型与基于郑州大学第一附属医院患者数据得到的训练模型在DSC上存在显著性差异,其原因可能是两所医院的CT数据或两所医院的医生勾画习惯存在差异;基于MICCAI 2019放疗规划自动结构勾画国际挑战赛比赛数据的训练模型效果虽然

略低,但是与基于郑州大学第一附属医院患者数据得到的训练模型并没有显著性差别,说明这两个中心数据差异较小或者医生对于心脏的轮廓勾画差别较小,具有相似的心脏勾画规范。

表1 训练模型1、2、4之间的差别

Tab.1 Differences among training models 1, 2 and 4

训练模型	DSC
模型1	0.926±0.030 <sup>a</sup>
模型2	0.932±0.030 <sup>b</sup>
模型4	0.941±0.026

a表示与模型4比较, $P=0.006$ ;b表示与模型4比较, $P=0.070$

**2.1.2 不同医院不同训练数量的模型之间的差别** 从表2和图3可以看到,模型1~5在测试集上的结果存在一定的差异,从DSC的比较结果来看,模型5>模型4>模型3>模型2>模型1,基于郑州大学第一附属医院的100例患者数据训练得到的模型效果最好。

表2 不同训练模型的心脏自动勾画精度

Tab.2 Accuracies of different training models for the automatic segmentation of the heart

训练模型	DSC
模型1	0.926±0.030 <sup>a</sup>
模型2	0.932±0.030 <sup>b</sup>
模型3	0.939±0.023 <sup>c</sup>
模型4	0.941±0.026 <sup>d</sup>
模型5	0.950±0.017

a表示与模型5比较, $P<0.05$ ;b表示与模型5比较, $P<0.05$ ;c表示与模型5比较, $P<0.05$ ;d表示与模型5比较, $P<0.05$

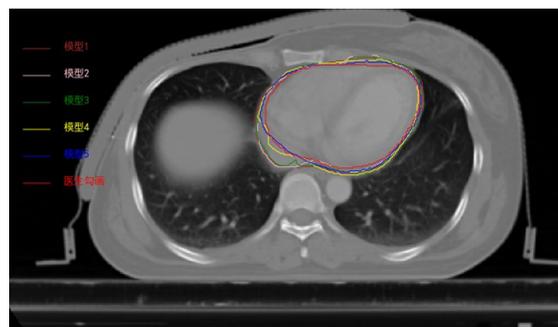


图3 基于不同模型自动勾画与医生手动勾画的心脏轮廓

Fig.3 Heart contours automatically segmented by different models and manually sketched by doctors

模型2得到的DSC低于模型5得到的DSC,并且两者存在显著性差别( $P<0.05$ )。模型3得到的DSC

高于模型1和模型2,模型5得到的DSC高于模型4,说明训练数据量的增多可以提高模型的精度。

### 2.2 模型在临床应用的改进

如表2所示,模型5得到的DSC高于模型4,并且两者存在显著性差别,说明将本模型应用到其他医院的心脏自动勾画时,训练的患者数据量和模型精度有直接联系。但是训练这些数据需要大量的时间,本研究中使用的模型训练耗费的时间如表3所示。本研究提出了模型的再训练改进,使用模型3(115例混合数据)作为心脏的预训练模型,基于本地数据的模型改进结果如图4所示。

表3 不同模型的训练时间  
Tab.3 Training time of different models

训练模型	耗费时间/h
模型1	7.0
模型2	3.5
模型3	10.0
模型4	3.3
模型5	8.5

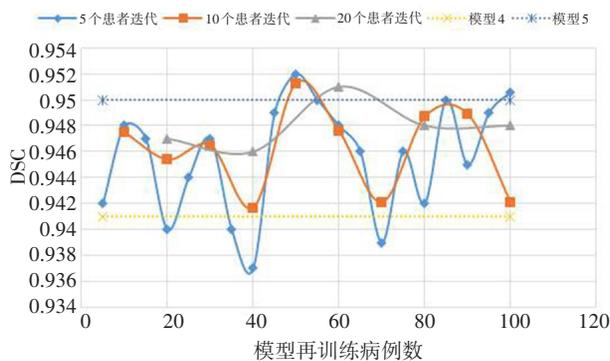


图4 心脏自动勾画模型在郑州大学第一附属医院的自适应改进  
Fig.4 Adaptive improvement of the heart automatic segmentation model in the First Affiliated Hospital of Zhengzhou University

从图4可以看到,在心脏预训练模型的基础上,使用一次本地数据的再训练之后的模型精度就超过了模型4。其中,每次使用5例患者数据再训练会出现较大的波动,甚至出现了低于模型4的效果,平均每次训练时间为0.22 h;每次使用10例患者数据再训练,波动程度减小,效果均优于训练模型4,平均每次训练时间为0.47 h;每次使用20例患者数据再训练,波动程度更小,效果较好,平均每次训练时间为1.08 h。由此可见,模型随着每批次患者数量的增多,波动程度减小,模型趋于稳定提高。

3组实验在再训练的过程中均出现了结果波动

的现象,这可能是由于再训练的患者数据差异性较大,这种差异在每次再训练数据量较小时尤为明显。在已有模型的基础上引入特殊病例会降低模型的精度,但通过提高每次再训练过程中的患者数据可以减小这种特殊病例带来的影响。

在已训练完成的模型应用到其他医院时,模型的再训练改进可以有效提高模型的精度和适用性,在较短的时间内达到较高的精度,在临床上更高效地完成心脏的自动勾画。

### 3 结论

本研究发现将心脏自动勾画模型应用到其他医院时,可以通过在已有的心脏预训练模型的基础上使用本地数据进行模型的再训练方式来提高模型精度与适用性,其中每批次再训练的患者数量越多,得到的模型预测效果越好,但是训练时间也会随之增加。建议在原有的心脏预训练模型的基础上采用每批次20例患者数据进行再训练,可以在较短的时间内适应于新的医院的心脏自动勾画需求。

目前,笔者在已有的勾画模型和数据量的基础上研究了心脏模型的自适应改进过程,从表2可以看到基于不同医院数据训练的模型在测试集上的DSC的最大差别是0.024(模型1:0.926,模型5:0.950),说明对于心脏这种轮廓较为明显的器官,不同医院的勾画存在较小的差异。在后续工作中,笔者将研究不同医院对于靶区的自动勾画是否存在差异,并继续研究通过模型的改进来提高靶区自动勾画精度。

综上所述,本研究证明了不同数据中心的的心脏自动勾画的差异,指出了使用同一模型去预测不同医院患者的心脏轮廓的不足,并就此问题提出了模型的改进方法,从而有效提高模型的临床实用性。

### 【参考文献】

- [1] TEGUH D N, LEVENDAG P C, VOET P W, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck[J]. Int J Radiat Oncol Biol Phys, 2011, 81(4): 950-957.
- [2] HWEE J, LOUIE A V, GAEDE S, et al. Technology assessment of automated atlas based segmentation in prostate bed contouring[J]. Radiat Oncol, 2011, 6(1): 110.
- [3] GAMBACORTA M A, VALENTINI C, DINAPOLI N, et al. Clinical validation of atlas-based auto-segmentation of pelvic volumes and normal tissue in rectal tumors using auto-segmentation computed system[J]. Acta Oncol, 2013, 52(8): 1676-1681.
- [4] PEJAVAR S, YOM S S, HWANG A, et al. Computer assisted, atlas-based segmentation for target volume delineation in whole pelvic IMRT for prostate cancer[J]. Technol Cancer Res Treat, 2013, 12(3):

- 199-206.
- [5] NELMS B E, TOME W A, ROBINSON G. Variation in the contouring of organs at risk: test case from a patient with oropharyngeal cancer [J]. *Int J Radiat Oncol Biol Phys*, 2012, 82(1): 368-378.
- [6] FENG M, DEMIROZ C, KAREN A, et al. Normal tissue anatomy for oropharyngeal cancer contouring variability and its impact on optimization [J]. *Int J Radiat Oncol Biol Phys*, 2012, 84(2): 245-249.
- [7] MEN K, CHEN X Y, ZHANG Y, et al. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning CT images [J]. *Front Oncol*, 2017, 7: 315.
- [8] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. *IEEE Trans Pattern Anal Mach Intell*, 2018, 40(4): 834-848.
- [9] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3431-3440.
- [10] 崔志强, 邱杰, 杨波, 等. 自动勾画技术在前列腺癌自适应放射治疗中的临床应用研究 [J]. *中国医学装备*, 2019, 16(4): 29-33.
- CUI Z Q, QIU J, YANG B, et al. Study on clinical application of automatic delineation technique in adaptive radiotherapy of prostatic cancer [J]. *China Medical Equipment*, 2019, 16(4): 29-33.
- [11] 邓金城, 彭应林, 刘常春, 等. 深度卷积神经网络在放射治疗计划图像分割中的应用 [J]. *中国医学物理学杂志*, 2018, 35(6): 621-627.
- DENG J C, PENG Y L, LIU C C, et al. Application of deep convolution neural network in radiotherapy planning image segmentation [J]. *Chinese Journal of Medical Physics*, 2018, 35(6): 621-627.
- [12] 门阔, 戴建荣. 利用深度反卷积神经网络自动勾画放疗危及器官 [J]. *中国医学物理学杂志*, 2018, 35(3): 256-259.
- MEN K, DAI J R. Automatic segmentation of organs at risk in radiotherapy using deep deconvolutional neural network [J]. *Chinese Journal of Medical Physics*, 2018, 35(3): 256-259.
- [13] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]. *MICCAI*, 2015: 234-241.
- [14] TRAN D, BOURDEV L D, FERGUS R, et al. Deep end2end voxel2voxel prediction [C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015.
- [15] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [J]. *arXiv preprint, arXiv: 1512.00567*, 2015.
- [16] COY H, HSIEH K, WU W, et al. Deep learning and radiomics: the utility of Google TensorFlow™ Inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphase CT [J]. *Abdom Radiol*, 2019. DOI: 10.1007/s00261-019-01929-0.
- [17] CAHALL D E, RASOOL G, BOUAYNAYA N C, et al. Inception modules enhance brain tumor segmentation [J]. *Front Comput Neurosci*, 2019, 13: 44.
- [18] 王巍. 基于Inception模块的医疗图像分类 [D]. 长春: 吉林大学, 2018.
- WANG W. Medical image classification based on inception module [D]. Changchun: Jilin University, 2018.
- [19] 田娟秀, 刘国才, 谷珊珊, 等. 基于3D深度残差全卷积网络的头颈CT放疗危及器官自动勾画 [J]. *中国生物医学工程学报*, 2019, 38(3): 257-265.
- TIAN J X, LIU G C, GU S S, et al. Segmentation of organs at risk on head and neck CT for radiotherapy based on 3D deep residual fully convolutional neural network [J]. *Chinese Journal of Biomedical Engineering*, 2019, 38(3): 257-265.
- [20] CRUM W R, CAMARA O, HILL D L. Generalized overlap measures for evaluation and validation in medical image analysis [J]. *IEEE Trans Med Imaging*, 2006, 25(11): 1451-1461.

(编辑: 谭斯允)