

基于深度学习和图谱库方法自动勾画肿瘤放疗中危及器官的比较

张富利¹, 崔德琪², 王秋生³, 韦凌宇¹, 朱林林², 郁艳军¹, 李海鹏³, 王雅棣¹

1. 解放军总医院第七医学中心, 北京 100700; 2. 北京连心医疗科技有限公司, 北京 100083; 3. 北京航空航天大学自动化科学与电气工程学院, 北京 100083

【摘要】目的:评估比较基于深度学习(DL)和图谱库(Atlas)方法自动勾画不同部位肿瘤放疗中危及器官(OARs)轮廓的几何学精度,为临床应用提供依据。**方法:**选择40例肿瘤患者的CT图像(头颈部、胸部、腹部和盆腔肿瘤患者各10例),由资深放射治疗医师手动勾画OARs,然后再分别使用基于DL和Atlas方法的自动勾画软件勾画OARs。采用形状相似性指数(DC)、Jaccard系数(JC)、Hausdorff距离(HD)、体积差异(VD)等多个指标评价基于DL和Atlas自动勾画与手动勾画OARs的几何学一致性。**结果:**除直肠外,采用DL方法勾画的多数OARs的DC指标高于0.7,优于Atlas方法,差异有统计学意义($P<0.05$)。此外,DL方法的JC值除晶体、直肠、脊髓外也都大于0.7。HD中最大的是脊髓,两种方法均超过20 mm。DL方法中VD绝对值较大的是直肠。**结论:**基于DL方法自动勾画的OARs几何精确度总体上高于Atlas方法。下一步,通过继续增大训练集的数据量可进一步提高基于DL方法模型的鲁棒性,从而更好地辅助放射肿瘤医师,使肿瘤患者获益。

【关键词】深度学习;图谱库;危及器官;自动勾画;肿瘤放射治疗

【中图分类号】R318;R811.1

【文献标志码】A

【文章编号】1005-202X(2019)12-1486-05

Comparative study of deep learning- versus Atlas-based auto-segmentation of organs-at-risk in tumor radiotherapy

ZHANG Fuli¹, CUI Deqi², WANG Qiusheng³, WEI Lingyu¹, ZHU Linlin², YU Yanjun¹, LI Haipeng³, WANG Yadi¹

1. the Seventh Medical Center of Chinese PLA General Hospital, Beijing 100700, China; 2. Beijing Linkingmed Science and Technology Company, Beijing 100083, China; 3. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China

Abstract: Objective To evaluate and compare the geometric accuracy between deep learning (DL)- and Atlas-based auto-segmentation technologies for contouring organs-at-risk (OARs) in radiotherapy for tumors locating in different sites so as to provide a basis for the clinical application. **Methods** The OARs in CT images of 40 patients with tumors in different sites (head and neck, thorax, abdomen, and pelvic cavity) were manually segmented by senior physicians, and then automatically segmented by DL- and Atlas-based auto-segmentation methods. Several evaluation indicators such as Dice coefficient (DC), Jaccard coefficient (JC), Hausdorff distance (HD) and volume difference (VD) were used to evaluate the geometric accuracy between DL- or Atlas-based auto-segmentations and manual segmentation. **Results** The DC values of OARs except for rectum segmented by DL-based method were higher than 0.7, higher than the results obtained by Atlas-based method, with statistical significance ($P<0.05$). In addition, the JC values obtained by DL-based method were also higher than 0.7, except for the JC values of lens, rectum and spinal cord. Spinal cord had the highest HD value, exceeding 20 mm in both methods. The rectum segmented by DL method had relatively high absolute VD. **Conclusion** The geometric accuracy of DL-based auto-segmentation is generally superior to that of Atlas-based auto-segmentation. In the further study, the robustness of DL model will be increased by expanding the training dataset, thereby better assisting radiation oncologists in routine clinical work and bringing benefits to tumor patients.

Keywords: deep learning; Atlas, organs-at-risk; auto-segmentation; tumor radiotherapy

【收稿日期】2019-10-02

【基金项目】首都临床特色应用研究专项课题(Z181100001718011)

【作者简介】张富利,副主任医师/副教授,主要从事多模态影像引导精确放疗、辐射防护与保健物理等临床科研工作, E-mail: radiozfli@163.com

【通信作者】王雅棣,主任医师,主要从事肿瘤精确放疗和综合治疗等临床科研工作, E-mail: wangyadi@hotmail.com

前言

放射治疗是恶性肿瘤的三大治疗手段之一, 超过70%的恶性肿瘤患者需要接受放射治疗^[1]。精确勾画肿瘤靶区和危及器官(OARs)的轮廓是实现精确放射治疗的前提和保障^[2]。通常情况下, 放射肿瘤医师根据患者的定位CT图像手动勾画靶区和OARs, 勾画过程较为耗时费力, 不同医师之间受主观性因素影响导致勾画结果存在较大差异, 而且大量重复性的工作降低了医师对患者进行诊疗的效率^[3-4]。近年来, 自动勾画已成为放疗领域的研究热点之一, 基于图谱库(Atlas)和深度学习(Deep Learning, DL)的自动勾画方法是目前两种主要的自动勾画技术。前者通过构建对比数据模板影像库, 采用刚性和形变配准的方法实现靶区和OARs自动勾画^[5-8]; 后者基于神经网络DL算法, 运用标准数据集对其进行有效训练, 通过算法逻辑实现自动勾画。本研究主要目的是评估基于DL和Atlas的自动勾画方法在不同部位肿瘤放疗中OARs自动勾画的精确度, 为临床应用提供参考依据。

1 资料与方法

1.1 临床资料

随机选取解放军总医院放疗科2016年7月~2019年3月40例肿瘤患者, 其中头颈部、胸部、腹部和盆腔肿瘤患者各10例。定位时均采用仰卧位, 采用CT模拟定位系统(Philips Big bore, USA)进行扫描, 头

颈部肿瘤扫描层厚3 mm, 层间距3 mm, 其余3个部位肿瘤均选择扫描层厚5 mm, 层间距5 mm。

1.2 手动勾画和基于Atlas自动勾画OARs

将扫描获取的CT图像分别传至ADAC Pinnacle³治疗计划系统(Treatment Planning System, TPS, version 9.10, Philips Radiation Oncology systems, Madison, WI)工作站。首先由专业放疗医师参考ICRU 62号报告手动勾画OARs, 包括脑、脑干、眼球、晶体、下颌骨、双肺、心脏、脊髓、双肾、膀胱、直肠、股骨头, 共12类17个OARs。然后使用Pinnacle³ TPS的MBS(Model-based Segmentation)功能模块自动勾画上述病例CT影像中的OARs, 该模块采用基于图谱库的方法实现自动勾画。

1.3 基于U-net的深度卷积神经网络(Convolutional Neural Network, CNN)

U-net网络由Ronneberger等^[9]提出, 对完全卷积神经网络(Fully Convolutional Networks, FCN)进行改进, 不同于FCN的跳层连接方式, U-net使用跨层连接方式, 通过剪裁和复制(crop and copy)方式建立上下文之间的联系, 使得分割结果更为精确、影像细节还原性更好。U-net架构适用于医学影像的语义分割, 一方面医学图像边界模糊、梯度复杂, 需要提取较多的高分辨率信息(浅层特征), 另一方面人体内部结构相对固定, 分割目标在图像中的分布更规律, 语义简单明确, 而低分辨率信息(深层特征)能够提供这一信息, U-net将低分辨率信息和高分辨率信息完美结合在一起, 其架构如图1所示。

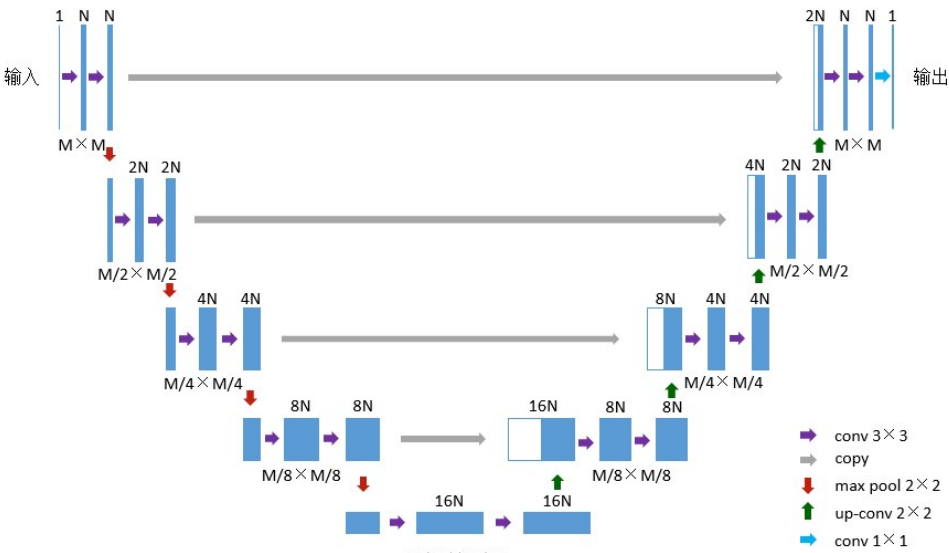


图1 U-net CNN架构
Fig.1 U-net CNN architecture

1.4 基于DL自动勾画OARs

分别选取89例头颈部肿瘤患者数据、596例胸部

肿瘤患者数据、185例腹部肿瘤患者数据和96例盆腔肿瘤患者数据作为训练集, 用以调整CNN的参数, 本

研究选取40例患者数据作为验证集评估模型的性能。所有训练和测试实验均在显存容量11 GB的NVIDIA GeForce GTX 1080 Ti GPU上运行。

1.5 评价指标

以医生手动勾画的OARs轮廓体积作为基准真值,采用形状相似性系数(Dice Coefficient, DC)、Jaccard系数(Jaccard Coefficient, JC)、Hausdorff距离(HD)、体积差异(Volume Difference, VD)指标定量评估两种自动勾画方法获得的OARs轮廓体积的精确性。为了确保手动勾画OARs的精确性,每个病例手动勾画完成后均由副主任医师及以上资格的医师修改、审批。

1.5.1 DC

用于评价 R 与 G 两个集合之间的重合性,其计算公式为:

$$DC(R, G) = \frac{2|R \cap G|}{|R| + |G|} \quad (1)$$

其中, R 为自动勾画结果; G 为手动勾画结果,即真值(ground truth)。其值的范围从0到1,越接近于1,表示两种勾画方式的重合性越好。

1.5.2 JC

用于度量 R 、 G 两个集合之间的相似性,其计算公式为:

$$JC(R, G) = \frac{|R \cap G|}{|R \cup G|} = \frac{|R \cap G|}{|R| + |G| - |R \cap G|} \quad (2)$$

其中, $0 \leq JC(R, G) \leq 1$,JC值越大,样本相似度越高。

1.5.3 HD

用来衡量 R 、 G 在三维空间上的表面距离,其计算公式为:

$$HD(R, G) = \max(h(R, G), h(G, R)) \quad (3)$$

$$h(R, G) = \max_{r \in R, g \in G} (\min \|r - g\|) \quad (4)$$

1.5.4 VD

指分割结果体积与真值(ground truth)体积差异占真值体积的百分比,越接近0越好:

$$VD(R, G) = \frac{|R| - |G|}{|G|} \times 100\% \quad (5)$$

其中, G 为真值, R 是自动分割结果。

1.6 统计学分析

采用SPSS 20.0统计学软件行Wilcoxon符号秩检验,显著性水平 $\alpha=0.05$, $P<0.05$ 表示差异有统计学意义。

2 结果

表1列出了评估基于Atlas和DL方法自动勾画

全脑、脑干、晶体、下颌骨、左右肺、膀胱、直肠等OARs几何精确性的指标。文献[10-11]指出,DC值大于0.7表明两个轮廓重复性尚可,分割效果可以接受。从表1中可以看出,除直肠外,采用DL方法勾画的多数OARs的DC指标高于0.7,优于Atlas方法,差异有统计学意义($P<0.05$)。此外,DL方法的JC、HD、VD普遍优于Atlas方法,其JC值除晶体、直肠、脊髓外也都大于0.7。HD中最大的是脊髓,两种方法均超过20 mm。DL方法中VD绝对值较大的是直肠。

从图2中可以看出,DL方法自动勾画的OARs与手动勾画真值一致性较好,基本上能为临床所接受。

3 讨论

精确的靶区和OARs勾画是放射治疗的重要前提之一。通常情况下,放射肿瘤医师在临床工作中OARs的勾画工作量较大、重复性工作较多,另受放射肿瘤医师主观性以及各自专业水平的限制,难以保证轮廓勾画的一致性。因此,自动勾画已成为目前的研究热点之一,旨在确保勾画精确度的前提下,提高勾画效率、减少不同勾画者间以及勾画者自身的差异。

截至目前,已有多款自动勾画软件应用于放疗临床工作中,这其中主要包括MIM软件(MIM Software Inc., Cleveland, OH)、Eclipse治疗计划系统中的SmartAdapt模块和Velocity模块(Varian Medical Systems, Palo Alto, CA)、ABAS软件(Atlas-based autosegmentation, Elekta Medical System, Stockholm, Sweden)、RayStation治疗计划系统中的RayDeformable模块(RaySearch Laboratories, Stockholm, Sweden)。这几款软件能够执行3DCT、4DCT、MRI或PET/CT等不同模态影像间的形变配准^[12-15]。这些方法的共同点是基于图谱库自动勾画,对待分割的目标影像与包含真值的图谱库影像进行形变配准,将后者中的标记结构自动传播到目标图像上。但是,有研究报道不同配准算法存在较大差异。基于特征的深度学习方法可解决此类问题,通过捕获不同算法间的差异并将其纳入预测模型以解决临床工作中可能遇到的问题,国内外许多学者都对此进行探讨^[16-22]。

本文研究结果显示,基于DL的自动勾画结果总体上优于基于Atlas的自动勾画结果。前者的DC值除直肠外均大于0.7,显示出较高的精确度。JC值除晶体、直肠、脊髓外也都大于0.7,可能原因为晶体、脊髓体积较小,而直肠受个体化差异影响大。HD中,

表 1 基于 Atlas 与 DL 方法自动勾画结果评价指标 ($\bar{x} \pm s$)

Tab.1 Comparison of evaluation parameters between Atlas- and DL-based auto-segmentation results (*Mean±SD*)

危及器官	DC		JC		HD/mm		VD		P 值
	Atlas	DL	Atlas	DL	Atlas	DL	Atlas	DL	
全脑	0.96±0.04	0.98±0.00	0.93±0.07	0.96±0.01	5.69±1.98	4.26±0.74	0.00±0.10	-0.01±0.01	0.172 ^a ,0.173 ^b ,0.022 ^c ,0.074 ^d
脑干	0.67±0.09	0.87±0.03	0.51±0.10	0.76±0.05	7.89±1.05	4.36±1.08	0.27±0.23	-0.07±0.09	0.005 ^a ,0.005 ^b ,0.005 ^c ,0.007 ^d
左眼	0.64±0.12	0.89±0.02	0.48±0.13	0.79±0.03	4.11±1.46	2.70±0.55	-0.54±0.28	-0.01±0.08	0.008 ^a ,0.008 ^b ,0.008 ^c ,0.008 ^d
右眼	0.67±0.11	0.88±0.02	0.52±0.11	0.78±0.03	3.78±1.19	2.81±0.64	-0.59±0.25	-0.03±0.10	0.008 ^a ,0.008 ^b ,0.051 ^c ,0.008 ^d
左晶体	0.43±0.12	0.76±0.06	0.28±0.10	0.62±0.07	2.39±0.46	1.81±0.39	-0.49±0.31	0.02±0.25	0.028 ^a ,0.028 ^b ,0.027 ^c ,0.028 ^d
右晶体	0.46±0.17	0.74±0.09	0.31±0.13	0.59±0.11	2.55±1.21	2.25±0.93	-0.37±0.50	0.12±0.30	0.012 ^a ,0.012 ^b ,0.484 ^c ,0.069 ^d
下颌骨	0.81±0.04	0.88±0.02	0.69±0.06	0.78±0.03	17.23±6.34	5.10±2.59	-0.02±0.15	-0.17±0.06	0.007 ^a ,0.007 ^b ,0.005 ^c ,0.037 ^d
心脏	0.73±0.21	0.90±0.04	0.61±0.22	0.81±0.06	16.03±10.53	10.06±8.62	-0.22±0.31	0.16±0.08	0.001 ^a ,0.001 ^b ,0.030 ^c ,0.001 ^d
左肺	0.95±0.03	0.95±0.02	0.90±0.05	0.91±0.03	18.93±13.90	12.94±14.93	0.03±0.08	-0.05±0.05	0.309 ^a ,0.309 ^b ,0.001 ^c ,0.001 ^d
右肺	0.96±0.02	0.97±0.01	0.91±0.03	0.93±0.02	14.38±6.01	8.64±4.47	0.04±0.05	-0.04±0.03	0.047 ^a ,0.049 ^b ,0.007 ^c ,0.001 ^d
左肾	0.81±0.17	0.91±0.02	0.71±0.20	0.83±0.03	7.42±6.41	3.87±1.17	0.03±0.19	-0.07±0.06	0.015 ^a ,0.015 ^b ,0.003 ^c ,0.049 ^d
右肾	0.82±0.12	0.88±0.04	0.71±0.16	0.79±0.05	6.65±3.37	4.54±1.06	0.15±0.19	-0.08±0.11	0.177 ^a ,0.201 ^b ,0.039 ^c ,0.003 ^d
膀胱	0.84±0.10	0.89±0.07	0.74±0.13	0.81±0.10	11.70±5.46	7.79±5.89	0.12±0.24	0.01±0.12	0.011 ^a ,0.011 ^b ,0.019 ^c ,0.209 ^d
直肠	0.51±0.16	0.65±0.23	0.36±0.15	0.51±0.21	19.80±7.63	14.00±6.87	0.34±0.29	-0.29±0.40	0.034 ^a ,0.012 ^b ,0.041 ^c ,0.005 ^d
左股骨头	0.68±0.20	0.83±0.08	0.55±0.23	0.72±0.11	14.19±8.78	6.93±7.93	0.26±0.24	-0.01±0.20	0.033 ^a ,0.046 ^b ,0.009 ^c ,0.009 ^d
右股骨头	0.61±0.18	0.85±0.07	0.47±0.20	0.75±0.10	14.76±6.74	6.23±7.243	0.31±0.35	-0.05±0.21	0.004 ^a ,0.004 ^b ,0.002 ^c ,0.002 ^d
脊髓	0.75±0.18	0.72±0.17	0.62±0.21	0.59±0.19	24.10±17.41	21.50±22.78	0.27±0.46	0.18±0.48	0.029 ^a ,0.019 ^b ,0.163 ^c ,0.000 ^d

a表示两种自动勾画方法DC值的Wilcoxon 检验P值;b表示两种自动勾画方法JC值的Wilcoxon 检验P值;c表示两种自动勾画方法HD值的Wilcoxon 检验P值;d表示两种自动勾画方法VD值的Wilcoxon 检验P值。

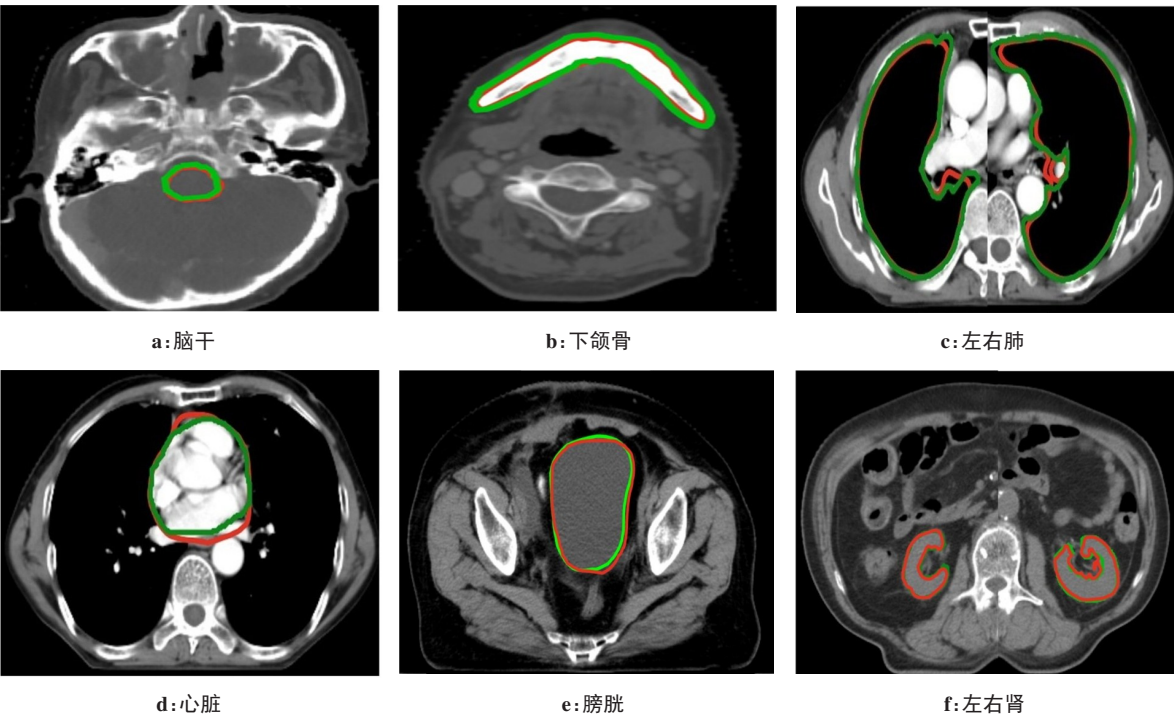


图 2 基于 DL 方法自动勾画 OARs 结果示意图(绿色线条为手动勾画轮廓,红色线条为自动勾画轮廓)

Fig.2 DL-based automatic segmentation results (the green lines were manual segmentations, and the red lines were automatic segmentations)

最大的是脊髓,超过 20 mm,原因可能是脊髓体积较小而层面数又较多所致。VD 值绝对值较大的是晶体、直肠,可能原因基本同上。

尽管在设计治疗计划之前仍需对基于 DL 方法自动勾画的 OARs 进行修改,但已在很大程度上减少了放射肿瘤医师的工作量。下一步,通过继续增大训练集的数据量从而包括更多的肿瘤患者特征,有望进一步提高基于 DL 方法模型的鲁棒性,从而更好地辅助放射肿瘤医师,使肿瘤患者获益。

【参考文献】

- [1] BARTON M B, FROMMER M, SHAFIQ J. Role of radiotherapy in cancer control in low-income and middle-income countries [J]. *Lancet Oncol*, 2006, 7(7): 584-595.
- [2] LEE J G, JUN S, CHO Y W, et al. Deep learning in medical imaging: general overview [J]. *Korean J Radiol*, 2017, 18(4): 570-584.
- [3] 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.
- [4] CHEN A M, CHIN R, BERON P, et al. Inadequate target volume delineation and local-regional recurrence after intensity-modulated radiotherapy for human papillomavirus-positive oropharynx cancer [J]. *Radiother Oncol*, 2017, 23(3): 412-418.
- [5] ELDESOKY A R, YATES E S, NYENG T B, et al. Internal and external validation of an ESTRO delineation guideline-dependent automated segmentation tool for loco-regional radiation therapy of early breast cancer [J]. *Radiother Oncol*, 2016, 121(3): 424-430.
- [6] WANG J, CHEN W, STUDENSKI M, et al. A semi-automated tool for treatment plan-quality evaluation and clinical trial quality assurance [J]. *Phys Med Biol*, 2013, 58(13): N181-N187.
- [7] LA MACCHIA M, FELLIN F, AMICHETTI M, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer [J]. *Radiat Oncol*, 2012, 7(1): 160.
- [8] NIE K, POULIOT J, SMITH E, et al. Performance variations among clinically available deformable image registration tools in adaptive radiotherapy - how should we evaluate and interpret the result? [J]. *J Appl Clin Med Phys*, 2016, 17(2): 328-340.
- [9] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [M]. *Medical Image Computing and Computer-Assisted Intervention*, 2015: 234-241.
- [10] ZIJDENBOS A P, DAWANT B M, MARGOLIN R A, et al. Morphometric analysis of white matter lesions in MR images: method and validation [J]. *IEEE Trans Med Imaging*, 1994, 13(4): 716-724.
- [11] ARTAECHAVARRIA X, MUNOZBARRUTIA A, ORTIZDESO-LORZANO C. Combination strategies in multi-atlas image segmentation: application to brain MR data [J]. *IEEE Trans Med Imaging*, 2009, 28(8): 1266-1277.
- [12] ELDESOKY A R, YATES E S, NYENG T B, et al. Internal and external validation of an ESTRO delineation guideline-dependent automated segmentation tool for loco-regional radiation therapy of early breast cancer [J]. *Radiother Oncol*, 2016, 121(3): 424-430.
- [13] SARUDIS S, KARLSSON A, BIBAC D, et al. Evaluation of deformable image evaluation of deformable image registration accuracy for CT images of the thorax region [J]. *Phys Med*, 2019, 57: 191-199.
- [14] 胡彩容, 林小艺, 阴晓娟, 等. 自动勾画软件在食管癌中的应用 [J]. *中华放射医学与防护杂志*, 2018, 38(9): 684-689.
- [15] HU C R, LIN X Y, YIN X J, et al. Application of autosegmentation software in esophageal cancer [J]. *Chinese Journal of Radiological Medicine and Protection*, 2018, 38(9): 684-689.
- [15] 郑庆增, 王运来, 张建春, 等. 宫颈癌自适应放疗中基于自配准与图谱库的自动勾画比较 [J]. *中华放射肿瘤学杂志*, 2019, 28(4): 292-296.
- [15] ZHENG Q Z, WANG Y L, ZHANG J C, et al. Evaluation of the auto-segmentation based on self-registration and Atlas in adaptive radiotherapy for cervical cancer [J]. *Chinese Journal of Radiation Oncology*, 2019, 28(4): 292-296.
- [16] BAUER S, NOLTE L P, REYES M. Fully automatic segmentation of brain tumor images automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization [J]. *Med Image Comput Assist Interv*, 2011, 14(3): 354-361.
- [17] MEN K, ZHANG T, CHEN X, et al. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning [J]. *Phys Med*, 2018, 50: 13-19.
- [18] MEN K, BOIMEL P, JANOPPAUL-NAYLOR J, et al. Cascaded atrous convolution and spatial pyramid pooling for more accurate tumor target segmentation for rectal cancer radiotherapy [J]. *Phys Med Biol*, 2018, 63(18): 185016.
- [19] LIN L, DOU Q, JIN Y M, et al. Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma [J]. *Radiology*, 2019, 291(3): 677-686.
- [20] GUO Z, GUO N, GONG K, et al. Gross tumor volume segmentation for head and neck cancer radiotherapy using deep dense multi-modality network [J]. *Phys Med Biol*, 2019, 64(20): 205015.
- [21] CHAN J W, KEARNEY V, HAAF S, et al. A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning [J]. *Med Phys*, 2019, 46(5): 2204-2213.

(编辑:陈丽霞)