

基于 U-NET 深度学习算法检测与量化胸部 CT 图像的气胸

王舒皓, 叶丽芳, 杨茗茗, 孟以爽, 李跃华

【摘要】 目的:采用气胸患者的 CT 图像开发了一种使用 U-NET 架构的深度学习分割算法, 并对其性能进行了评估。**方法:**回顾性分析 2018 年—2019 年的急诊胸部 CT 图像, 先由一名放射科医生进行注释, 然后由另一名资深放射科医生修改和审查标注的气胸内容, 作为金标准, 并使用五折交叉验证方法进行深度学习算法的训练和测试。在像素级通过戴斯系数、召回率和符合率来评估分割精度, 并评估了气胸定量的体积误差; 在区域级评估每个患者的气胸区域敏感性和假阳性区域数量。**结果:**共有 200 例气胸患者入组, 平均戴斯系数、召回率和符合率分别为 0.789、0.794 和 0.820。对气胸总量大于 300 mL 的患者, 平均戴斯系数、召回率和符合率都可以达到 0.89 以上。对气胸总量大于 100 mL 的患者, 气胸定量的相对误差小于 10%。对体积大于 30 mL 的气胸区域, 区域敏感性可达 100%。假阳性区域平均体积为 2.3 mL (1.55~3.66 mL), 平均每个病例 2.8 (2.06~4.23) 个假阳性区域。**结论:**U-NET 深度学习分割算法在像素和区域两个层面上都表现出可接受的性能, 这表明在临床实践中可以发挥潜在的辅助作用, 以减轻急诊医务人员的工作量。

【关键词】 体层摄影术, X 线计算机; 人工智能; 深度学习; 气胸

【中图分类号】 R814.42; R-05; R561.4 **【文献标志码】** A **【文章编号】** 1000-0313(2023)03-0290-05

DOI: 10.13609/j.cnki.1000-0313.2023.03.009

开放科学(资源服务)标识码(OSID):



Pneumothorax detection and quantification in chest computed tomography images using a U-NET deep learning segmentation algorithm WANG Shu-hao, YE Li-fang, YANG Ming-ming, et al. Department of Radiology, the Sixth People's Hospital affiliated to Shanghai Jiaotong University School of Medicine, Shanghai 200233, China

【Abstract】 Objective: To construct a deep learning segmentation algorithm using U-NET architecture based on CT images of pneumothorax patients, and evaluate its performance. **Methods:** The emergency chest CT images from 2018—2019 were retrospective collected and annotated by a radiologist, followed by another senior radiologist to revise and review the annotated pneumothorax content as a gold standard. A deep learning algorithm was trained and tested using a five-fold cross-validation method. Segmentation accuracy was evaluated at the pixel level by Dice coefficients, recall and compliance, and volumetric error in pneumothorax quantification was assessed. The sensitivity of pneumothorax regions and the number of false positive regions were evaluated at the region level for each patient. **Results:** A total of 200 pneumothorax patients were enrolled, and the mean Dice coefficient, recall rate and compliance rate were 0.789, 0.794, and 0.820, respectively. For patients with total pneumothorax greater than 300 mL, the average Dice coefficient, recall and compliance rate can reach more than 0.89. For patients with a total pneumothorax volume greater than 100 mL, the relative error of pneumothorax quantification was less than 10%. For the pneumothorax region with volume greater than 30 mL, the regional sensitivity could reach 100%. The average volume of false-positive regions was 2.3 mL (1.55~3.66 mL), and the average number of false-positive regions was 2.8 (2.06~4.23) per case. **Conclusion:** The U-NET deep learning segmentation algorithm showed acceptable performances at both pixel and region levels, which can suggest as a potential tool in assisting emergency medical personnel to reduce the workload.

【Key words】 Tomography, X-ray computed; Artificial intelligence; Deep learning; Pneumothorax

作者单位: 200233 上海, 上海交通大学医学院附属第六人民医院放射科(王舒皓、叶丽芳、杨茗茗、李跃华); 200072 上海, 飞利浦中国投资有限公司(孟以爽)

作者简介: 王舒皓(1989—), 女, 黑龙江绥化县人, 硕士研究生, 住院医师, 主要从事人工智能和心胸影像诊断研究。

通讯作者: 李跃华, E-mail: liyuehua77@sjtu.edu.cn

气胸的定义是胸膜腔内存在异常空气。根据病因分为几种类型,包括原发性和继发性的自发性气胸、医源性及创伤性气胸^[1-2]。任何类型的气胸都可能成为潜在的危及生命的紧急情况,应立即诊断和处理,以减少不良后遗症^[2-3]。由于气胸的临床症状和体征通常没有特异性,32%的ICU患者在首次出现气胸时被误诊^[4],相关死亡率可能高达68%^[5],所以有效和准确地诊断气胸至关重要。

胸部CT是气胸诊断和容量估计的金标准,有些患者可以常规治疗或观察,但有些患者需要立即插入胸管或进行抽吸^[1,6]。由于人力和时间资源的限制,手动注释气胸区域对于放射科医生来说是耗时的,而自动检测系统有可能对医疗提供很大的帮助。

人工智能正越来越多地应用于医学领域的图像分析^[7-8],人工智能通过减少评分者之间的分歧并通过提高基于CT成像的标准化来快速临床诊断^[9-11]。U-NET卷积网络结构因其编码器和解码器排列的形状而得名,此设计用于快速而精确的生物医学图像分割^[12]。编码器提供分割特征,解码器对编码器学习的特征进行分类,同时它以不同的分辨率与编码器连接,并通过下采样来恢复丢失的信息^[12]。以往开发了几种深度学习算法检测胸部X光片上的气胸^[13-16],但使用CT图像开发的算法很少。尽管临床上可以通过多种影像学手段诊断气胸,但胸部CT自动检测系统无疑是必要的。

在本研究中,笔者使用气胸患者的CT图像开发了一种基于U-NET结构的深度学习分割算法,并对其像素和区域级别的性能进行了评估。这项研究的目

的是确定在气胸患者的CT成像解释过程中加入自动检测系统的可能性。

材料与方 法

1.数据收集

回顾性分析2019年—2021年上海第六人民医院急诊科收治的因外伤或胸痛就诊的成人患者(≥ 18 岁)的CT影像资料,并且患者在放射诊断报告中确诊为气胸。建立金标准包括两个阶段:第一阶段,根据诊断报告,由一位有6年工作经验的放射科医生标记病灶的轮廓。在第二阶段,另一位有10年工作经验的资深放射科医生在第一阶段的基础上进行了回顾,并修改了病灶的轮廓位置。由资深放射科医生确认的标签被用作实验中评估结果的金标准。排除图像有成像伪影或运动伪影的患者。

该研究方案由上海市第六人民医院机构评审委员会(IRB)批准,研究是根据赫尔辛基宣言进行的。由于这项研究的回溯性,IRB放弃了患者的知情同意。

2.图像采集及算法建立

图像使用Brilliance 64(Philips Healthcare, Best, 荷兰,层厚5 mm)采集,由两位放射科医师通过ITK-SNAP(www.itksnap.org)进行注释和复查。入组患者根据放射科医生估计的气胸体积被均匀分成五个子数据集,划分方法为先按气胸体积排序,再依次放入这五个子数据集,以确保每个子数据集中的患者气胸体积分布相似。研究使用五折交叉验证方法,即分别以其中四个子数据集的数据为训练集,另一个子数据集为测试集,共进行五次算法的训练和验证来综合评估

训练阶段

测试阶段

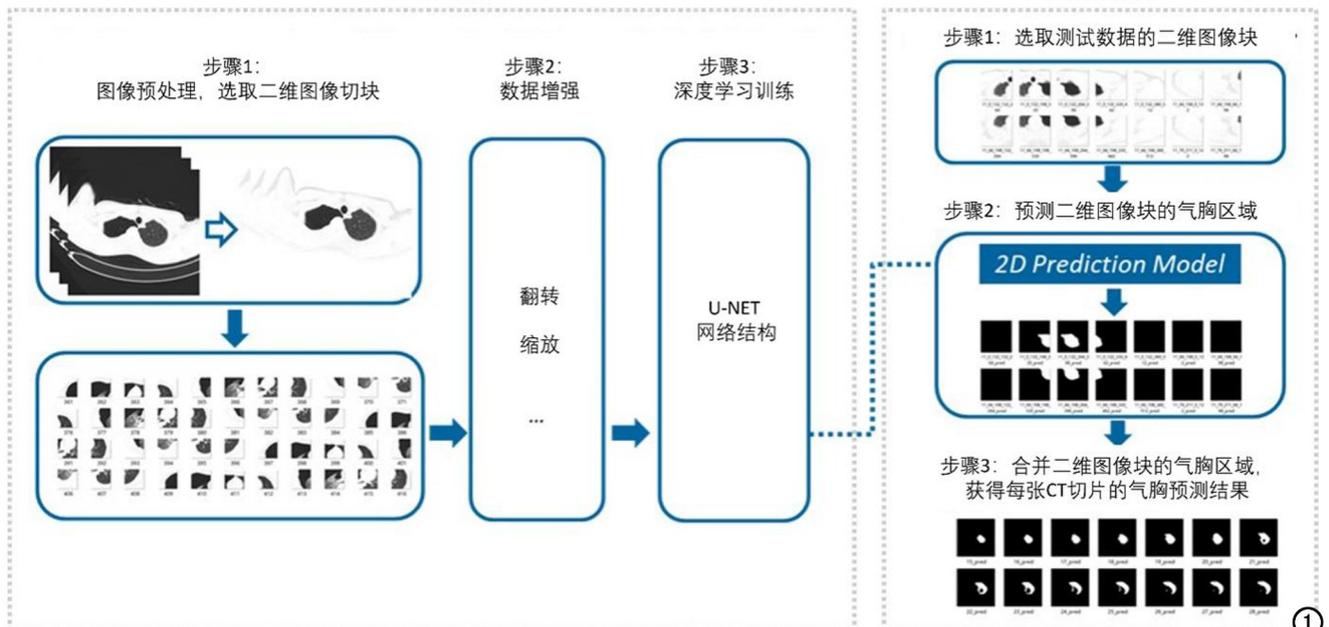


图1 U-NET深度学习算法训练和测试阶段工作流程。

算法性能。

深度学习分割算法基于 U-NET 网络建立,在交叉验证中,每一次的训练和测试步骤如图 1 所示。算法的输入为二维图像块,图像块的选取方法为在每层 CT 二维图像上以 50% 的重叠切块。训练数据通过图像翻转、缩放等方法进行增强,使训练集扩大 5 倍,以避免过拟合。在测试阶段,以相同的方法选取二维图像块,并对患者的所有二维图像块进行气胸预测,再计算每个像素重叠部分的平均概率,将所有预测结果合并得到最终结果。可根据分割结果重建三维图像。

3. 性能评价

研究从像素(体素)和区域两个层面对算法性能进行评估。在像素级,采用戴斯系数、召回率和符合率三种指标评价分割精度。召回率可以反映算法对气胸的检测能力和不漏检率,而符合率则可以反映算法避免误报的能力,公式如下所示。其中,TP 为测试集中的真阳性像素数量(被标记为气胸,并被正确预测为气胸的像素);FP 为假阳性像素数量(被标记为非气胸,但被预测为气胸的像素);FN 为假阴性像素数量(被标记为气胸,但被预测为非气胸像素的像素)。

$$\text{召回率} = \frac{TP}{TP+FN} \quad (1)$$

$$\text{符合率} = \frac{TP}{TP+FP} \quad (2)$$

戴斯系数是衡量召回率和符合率的综合指标,戴斯系数越高,分割模型的整体性能越好,计算公式如下所示,其中 X 是算法的预测结果, Y 是标注的真值。

$$\text{戴斯系数} = \frac{2 \times |X \cap Y|}{|X| + |Y|} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (3)$$

对三维重建,评估算法预测气胸体积的绝对误差和相对误差。体积计算公式为(像素间距与层厚可从 DICOM 数据中获得):

$$\text{体积}(\text{mm}^3) = \sum(\text{相邻切片中的平均气胸面积} \times \text{横向像素间距} \times \text{纵向像素间距} \times \text{层厚}) \quad (4)$$

在区域级,如果算法预测的气胸区域与标注的气胸区域间戴斯系数大于 0.5,则被认为是一个真阳性区域。使用基于真阳性区域和假阴性区域计算得到的区域敏感度和假阳性区域数量来评估算法在区域级的性能。

4. 实施详情

实验用 Pytorch 在 Ubuntu 16.04 环境下进行,图形处理器为 NVIDIA 1080Ti。一个患者的数据处理时间约为 20 秒。

结果

研究共纳入 200 例气胸患者,年龄 18~98 岁,平均(55.1±22.7)岁,男 163 例(81.5%)。五折交叉验证得到的戴斯系数、召回率和符合率平均值分别为 0.789、0.794 和 0.820(表 1)。此外,考虑到算法在像素级的性能可能会受到气胸体积的影响,患者根据气胸总量被分为 7 个亚组进行进一步评估(表 2),不同气胸总量患者百分比分别为小于 10 mL 占 10%,10~50 mL 占 19%,50~100 mL 占 13%,100~300 mL 占 20%,300~500 mL 占 10%,500~1000 mL 占 12%,大于 1000 mL 占 16%。结果表明,在气胸大于 300 mL 的患者中,平均戴斯系数、召回率和符合率均可达到 0.89 以上。气胸少于 10 mL 患者的戴斯系数、召回率和符合率最低,分别为 0.52、0.56 和 0.59。

表 1 不同算法在五折交叉验证中的像素级性能

数据集	戴斯系数	召回率	符合率
数据集 1	0.792	0.788	0.842
数据集 2	0.787	0.765	0.833
数据集 3	0.802	0.805	0.825
数据集 4	0.795	0.811	0.803
数据集 5	0.771	0.803	0.794
所有数据	0.789	0.794	0.820

同样,根据气胸总量将患者分为 7 个亚组进行体积误差评估(表 2)。结果显示气胸总量大于 1000 mL 时绝对误差最大,为 58.2 mL,而该亚组的相对误差仅为 3%。气胸总量少于 10 mL 时相对误差最大,为 57%;当气胸总量大于 100 mL 时,相对误差可小于 10%。通过三维重建图像评估体积的代表性图像见图 2。

由于单个患者可能有多个气胸区域,因此算法也在区域级别上进行性能评估。200 例患者共发现 426 个气胸区域,其中少于 1 mL 占 31.2%,1~10 mL 占 31.9%,10~30 mL 占 7.5%,30~50 mL 占 3.1%,50~100 mL 占 5.2%,100~500 mL 占 10.6%,大于 500 mL 占 10.3%。五折交叉验证中不同体积气胸区域的区域敏感度如表 3 所示。结果表明,当气胸区域超过 30 mL 时,所有子数据集均获得 100% 的区域敏

表 2 不同气胸总量亚组评估结果

参数	气胸总量/mL						
	<10	10~50	50~100	100~300	300~500	500~1000	>1000
戴斯系数	0.52	0.61	0.78	0.86	0.9	0.91	0.95
召回率	0.56	0.66	0.78	0.89	0.89	0.9	0.94
符合率	0.59	0.65	0.82	0.91	0.91	0.93	0.96
绝对误差/mL	3.3	12.3	20.0	18.1	29.5	31.6	58.2
相对误差/%	57	56	22	10	8	5	3

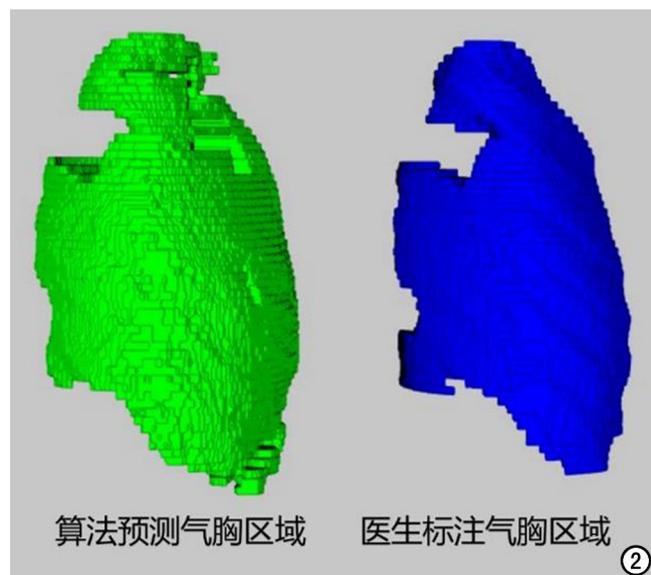


图 2 三维重建的气胸区域。

感度。区域越小,敏感度越低。假阳性区域的平均体积为 2.3 mL(1.55~3.66 mL),平均每病例假阳性区域为 2.8 个(2.06~4.23)。真阳性(TP)、假阳性(FP)和假阴性(FN)区域的典型图像见图 3。

讨论

使用确认的气胸图像进行训练和测试,U-NET 深度学习分割算法在像素级评估中清楚地显示了气胸大小对算法性能的影响。可以预见的是,人工识别和

图像检测算法都难以识别具有异常表现的小病灶区域,但该算法的戴斯系数、召回率和符合率对气胸总量小于 10 mL 和 10~50 mL 的患者仍可分别达到 0.5 和 0.6 左右,与人工识别相比,误差可接受。气胸区域体积分布结果显示,大部分单个气胸区域的体积在 10 mL 以下,说明算法支持对小气胸有一定的识别能力。这是一个具有临床意义的优势,因为气胸有可能迅速扩大,患者的病情可能会迅速恶化,需要紧急干预。

U-NET 架构已被广泛用于处理跨传染病和恶性疾病的 CT 肺部图像^[17-20]。U-NET 网络可以用少量图像进行端到端的训练,并产生更精确的分割。深度学习分割算法的目的是有效、快速地识别气胸及其大小,即作为监测系统的高敏感性气胸自动检测程序。最终,这个系统可以减轻急诊科放射科医生和其他医疗服务提供者的工作量,使他们能够更多地关注高度怀疑的患者,或确认结果,或充当第二判断者。本研究结果表明,当气胸区域大于 30 mL 时,区域敏感度可达 100%。另一项纳入了 703 名钝性创伤患者的研究发现,30 mL 可能是胸导管插入决策中气胸大小的适当临界值^[21],这支持了我们的算法在临床实践中的潜在应用价值。

Röhrich 等^[22]提出了一种基于 U-NET 的深度学习方法对胸部 CT 进行气胸检测和量化,取得了较好

表 3 不同算法在五折交叉验证中的区域级性能

数据集	不同气胸区域的敏感度(%)							平均假阳性区域体积(mL)	每例假阳性区域数量
	0~1mL	1~10mL	10~30mL	30~50mL	50~100mL	100~500mL	>500mL		
数据集 1	10	70	100	100	100	100	100	1.78	2.06
数据集 2	13.6	40	83.3	100	100	100	100	3.24	2.63
数据集 3	13.6	50	100	100	100	100	100	1.55	2.69
数据集 4	15.2	61.8	66.7	100	100	100	100	1.68	4.23
数据集 5	15.4	60	100	100	100	100	100	3.66	2.9
所有数据	13.5	55.9	93.5	100	100	100	100	2.3	2.8

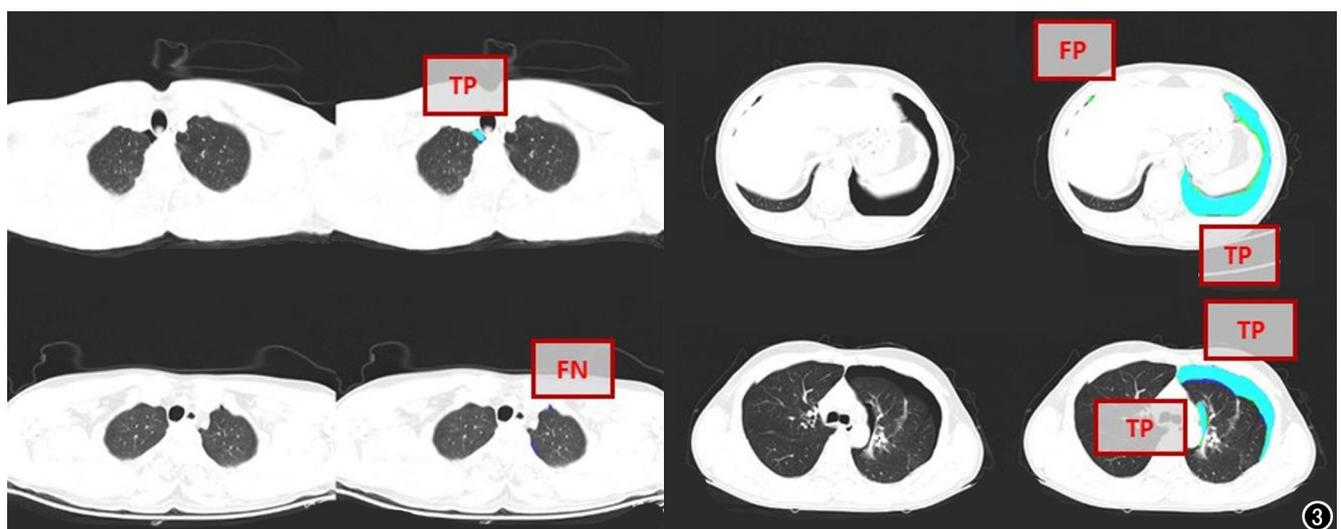


图 3 真阳性(TP)、假阳性(FP)和假阴性(FN)典型图像。

的效果。即使该研究的目的与本研究相似,但本研究设计与该研究的工作有很大的不同。Röhrich 等^[22]的算法主要对患者进行有无气胸的分类,是一项分类研究,而本研究的算法主要是为每个患者找到多个气胸区域。因此,不能对这两项研究进行直接比较。

作为快速监测系统而设计,本研究的算法仅需 20 秒来评估一名患者,这与其他基于 CT 的算法具有可比性,例如 Röhrich 等^[22]提出的算法处理时间为 53 秒, Li 等^[23]提出的算法处理时间为 151 秒, Cai 等^[24-25]提出的算法处理时间约为 3 分钟。需要注意的是,在没有控制变量的情况下直接进行研究之间的比较是不合适的,但在开发人工智能系统之前,应考虑目的和需要解决的临床困难。由于气胸在普通人群中的患病率较低,同时一旦出现就可能危及生命,因此选择工作效率高且敏感度高的监测系统是至关重要的。

应注意本研究的某些局限性,包括样本量较小、回顾性设计及所有患者均从单中心收集。由于所有入组的病例均来自急诊科,对门诊或其他入院的患者使用此算法可能会得到不同的结果。研究的金标准由放射科医生提供,但仍可能发生错误。U-NET 深度学习算法的诊断性能也需要在未来进行进一步的前瞻性评估。

综上所述,本研究提出了一种用于气胸检测和量化的深度学习分割算法,该算法在像素级和区域级都具有良好的性能。本研究结果表明,该算法在临床实践中具有潜在的辅助作用,可以减轻放射科医生和医疗健康工作者的工作量。

参考文献:

[1] Haynes D, Baumann MH. Management of pneumothorax[J]. *Semin Respir Crit Care Med*, 2010, 31(6): 769-780.

[2] Sahn SA, Heffner JE. Spontaneous pneumothorax[J]. *N Engl J Med*, 2000, 342(12): 868-874.

[3] Yarmus L, Feller-Kopman D. Pneumothorax in the critically ill patient[J]. *Chest*, 2012, 141(4): 1098-1105.

[4] Kollef MH. Risk factors for the misdiagnosis of pneumothorax in the intensive care unit[J]. *Crit Care Med*, 1991, 19(7): 906-910.

[5] Chen KY, Jerng JS, Liao WY, et al. Pneumothorax in the ICU: patient outcomes and prognostic factors[J]. *Chest*, 2002, 122(2): 678-683.

[6] Kelly AM, Weldon D, Tsang AY, et al. Comparison between two methods for estimating pneumothorax size from chest X-rays[J]. *Respir Med*, 2006, 100(8): 1356-1359.

[7] Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare[J]. *Nat Med*, 2019, 25(1): 24-29.

[8] 郭丽, 赵凯, 朱逸峰, 等. 使用 U-Net 深度学习网络对腰椎矢状 T₂WI 图像自动分割的可行性研究[J]. *放射学实践*, 2022, 37(2): 229-234.

[9] Bivard A, Churilov L, Parsons M. Artificial intelligence for decision support in acute stroke-current roles and potential[J]. *Nat Rev Neurol*, 2020, 16(10): 575-585.

[10] Kasasbeh AS, Christensen S, Parsons MW, et al. Artificial neural network computer tomography perfusion prediction of ischemic core[J]. *Stroke*, 2019, 50(6): 1578-1581.

[11] Qiu W, Kuang H, Teleg E, et al. Machine learning for detecting early infarction in acute stroke with non-contrast-enhanced CT[J]. *Radiology*, 2020, 294(3): 638-644.

[12] 孙兆男, 刘佳, 崔应谱, 等. 利用深度学习实现 CT 图像上肾脏肿瘤径线自动测量的临床可行性[J]. *放射学实践*, 2022, 37(3): 374-379.

[13] Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study[J]. *PLoS Med*, 2018, 15(11): e1002697.

[14] Wang H, Gu H, Qin P, et al. CheXLocNet: Automatic localization of pneumothorax in chest radiographs using deep convolutional neural networks[J]. *PLoS One*, 2020, 15(11): e0242013.

[15] Kao CY, Lin CY, Chao CC, et al. Automated radiology alert system for pneumothorax detection on chest radiographs improves efficiency and diagnostic performance[J]. *Diagnostics (Basel)*, 2021, 11(7): 1182.

[16] Thian YL, Ng D, Hallinan JTPD, et al. Deep learning systems for pneumothorax detection on chest radiographs: a multicenter external validation study[J]. *Radiol Artif Intell*, 2021, 3(4): e200190.

[17] Hoang-Thi TN, Vakalopoulou M, Christodoulidis S, et al. Deep learning for lung disease segmentation on CT: Which reconstruction kernel should be used? [J]. *Diagn Interv Imaging*, 2021, 102(11): 691-695.

[18] Yu H, Li J, Zhang L, et al. Design of lung nodules segmentation and recognition algorithm based on deep learning[J]. *BMC Bioinformatics*, 2021, 22(Suppl 5): 314.

[19] Saood A, Hatem I. COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet[J]. *BMC Med Imaging*, 2021, 21(1): 19.

[20] Zhang X, Liu X, Zhang B, et al. Accurate segmentation for different types of lung nodules on CT images using improved U-Net convolutional network[J]. *Medicine (Baltimore)*, 2021, 100(40): e27491.

[21] Hefny AF, Kunhivalappil FT, Matev N, et al. Management of computed tomography-detected pneumothorax in patients with blunt trauma: experience from a community-based hospital[J]. *Singapore Med J*, 2018, 59(3): 150-154.

[22] Röhrich S, Schlegl T, Bardach C, et al. Deep learning detection and quantification of pneumothorax in heterogeneous routine chest computed tomography[J]. *Eur Radiol Exp*, 2020, 4(1): 26.

[23] Li X, Thrall JH, Digumarthy SR, et al. Deep learning-enabled system for rapid pneumothorax screening on chest CT[J]. *Eur J Radiol*, 2019, 120: 108692.

[24] Cai W, Lee EY, Vij A, et al. MDCT for computerized volumetry of pneumothoraces in pediatric patients[J]. *Acad Radiol*, 2011, 18(3): 315-323.

[25] Cai W, Tabbara M, Takata N, et al. MDCT for automated detection and measurement of pneumothoraces in trauma patients[J]. *Am J Roentgenol*, 2009, 192(3): 830-836.