

## 基于深度学习单投影的CT断层成像三维重建

张新阳<sup>1,2,3,4</sup>, 贺鹏博<sup>1,2,3,4</sup>, 刘新国<sup>1,2,3,4</sup>, 戴中颖<sup>1,2,3,4</sup>, 马圆圆<sup>1,2,3,4</sup>, 中国盛<sup>1,2,3,4</sup>, 张晖<sup>1,2,3,4</sup>, 陈卫强<sup>1,2,3,4</sup>, 李强<sup>1,2,3,4</sup>  
1. 中国科学院近代物理研究所, 甘肃 兰州 730000; 2. 中国科学院重离子束辐射生物医学重点实验室, 甘肃 兰州 730000; 3. 甘肃省重离子束辐射医学应用基础重点实验室, 甘肃 兰州 730000; 4. 中国科学院大学核科学与技术学院, 北京 100049

**【摘要】目的:**提出一种基于深度学习的计算机断层扫描(CT)单视图断层成像三维(3D)重建方法,在减少数据采集量和降低成像剂量的情况下对不同患者进行CT图像的3D重建。**方法:**对不同患者的CT图像进行数据增强和模拟生成对应的数字重建放射影像(DRR),并进行数据归一化操作。利用预处理后的数据通过卷积神经网络训练出一个普适于不同患者的神经网络模型。将训练好的神经网络模型部署在测试数据集上,使用平均绝对误差(MAE)、均方根误差(RMSE)、结构相似性(SSIM)和峰值信噪比(PSNR)对重建结果进行评估。**结果:**定性和定量分析的结果表明,该方法可以使用不同患者的单张2D图像分别重建出质量较高的3D CT图像,MAE、RMSE、SSIM和PSNR分别为0.006、0.079、0.982、38.424 dB。此外,相比特定于单个患者的情况,该方法可以大幅度提高重建速度并节省70%的模型训练时间。**结论:**构建的神经网络模型可通过不同患者的2D单视图重建出相应患者的3D CT图像。因此,本研究对简化临床成像设备和放射治疗当中的图像引导具有重要作用。

**【关键词】**深度学习;CT单视图断层成像;三维重建;卷积神经网络;稀疏投影成像

**【中图分类号】**R318

**【文献标志码】**A

**【文章编号】**1005-202X(2021)10-1223-06

## Deep learning-based 3D reconstruction of CT tomography from a single projection

ZHANG Xinyang<sup>1,2,3,4</sup>, HE Pengbo<sup>1,2,3,4</sup>, LIU Xinguo<sup>1,2,3,4</sup>, DAI Zhongying<sup>1,2,3,4</sup>, MA Yuanyuan<sup>1,2,3,4</sup>, SHEN Guosheng<sup>1,2,3,4</sup>, ZHANG Hui<sup>1,2,3,4</sup>, CHEN Weiqiang<sup>1,2,3,4</sup>, LI Qiang<sup>1,2,3,4</sup>

1. Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou 730000, China; 2. Key Laboratory of Heavy Ion Radiation Biology and Medicine of Chinese Academy of Sciences, Lanzhou 730000, China; 3. Key Laboratory of Basic Research on Heavy Ion Radiation Application in Medicine, Lanzhou 730000, China; 4. School of Nuclear Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract: Objective** To propose a deep learning-based method for the 3D reconstruction of single-view computed tomography (CT) image, thereby realizing the 3D reconstruction of CT image for different patients under the condition of reducing the amount of data collection and decreasing imaging dose. **Methods** After the CT images of different patients were processed by data enhancement, and the corresponding digitally reconstructed radiograph (DRR) was obtained by simulation, data normalization was carried out. A neural network model universally suitable for different patients was established by training the pre-processed data using deep neural networks, and the trained neural network model was then deployed on the test dataset. Finally, the reconstruction results were evaluated using mean absolute error, root mean square error, structural similarity and peak signal noise ratio. **Results** The qualitative and quantitative analyses showed that the 3D CT images with high quality could be constructed by the proposed method using a single-view 2D image of different patients, with mean absolute error, root mean square error, structural similarity and peak signal noise ratio of 0.006, 0.079, 0.982 and 38.424 dB, respectively. Additionally, compared with the situation specific to a single patient, the proposed method greatly increased the reconstruction speed and save 70% of model training time. **Conclusion** The neural network model established in the study can reconstruct the corresponding 3D CT images using single-view 2D image of different patients. The study may play an important role in simplifying clinical imaging devices and image guidance in radiotherapy.

**【收稿日期】**2021-04-17

**【基金项目】**国家重点研发计划(2017YFC0107500);国家自然科学基金(11875299, 61631001)

**【作者简介】**张新阳,博士研究生,从事重离子束治疗技术基础研究,E-mail: zhangxinyang@impcas.ac.cn

**【通信作者】**李强,研究员,从事重离子治疗相关研究,E-mail: liqiang@impcas.ac.cn

**Keywords:** deep learning; single-view computed tomography; three-dimensional reconstruction; convolutional neural network; sparse projection imaging

## 前言

计算机断层扫描(Computed Tomography, CT)以高空间分辨率对患者或物体进行三维(3D)成像的能力在生命科学领域有着广泛的应用。而传统CT成像需要通过密集扫描获得大量的投影数据,导致重建时间增加,重建速度跟不上扫描速度;另一方面,密集扫描也给患者带来了额外的辐射剂量。随着人们生活水平的提高,辐射对健康的损害逐渐引起人们的注意。此外,在断层成像的许多应用中,成像速度至关重要。于是减少CT扫描次数,降低CT的辐射剂量成为科研人员关注的焦点。近年来,很多研究都在寻找减少采集数据量的同时又不降低图像质量的方法,通过压缩感知<sup>[1-6]</sup>和最大后验<sup>[7]</sup>等技术对稀疏采样的图像重建进行广泛的研究。这些方法在重建反演过程中引入正则化项,以促使在重建图像中出现原本的特征<sup>[8]</sup>。如果不能降低对成像质量的要求,这些方法并不能解决在大幅度减少成像对象所受辐射剂量的同时进行实时成像,而且它们的稀疏性通常是有限的。通过机器学习技术,从大量的数据当中提取出先验知识来提高成像速度和质量是最近的热门研究课题。近年来有很多基于机器学习的图像重建的研究<sup>[9-12]</sup>,但是最具有代表性的是Henzler等<sup>[13]</sup>基于深度学习的卷积神经网络利用2D放射影像重建出了3D图像,以及Shen等<sup>[14]</sup>和Lei等<sup>[15]</sup>利用深度学习和数据驱动的图像重建过程方法,将稀疏采样压缩到了单个投影视图的极限,实现了通过深度学习方法基于特定患者的单视图断层成像。但是在临床上,如果对每一个患者都进行数据增强并重复训练适用于每个患者的模型,时间将会大大增加。

基于以上考虑,本研究利用卷积神经网络首次实现了适用于不同患者的深度学习CT重建方法。这种方法普适性更强,对于不同例数的患者,只需要在治疗前利用他们的CT图像进行数据增强并统一训练,就可以得到适用于每个患者的神经网络模型。相比于重复训练使用该方法使每个患者特定的模型节约70%的时间,在保持重建图像质量的同时大大提高了效率。

## 1 材料与方法

### 1.1 网络模型

深度神经网络因其学习复杂关系的能力以及通

过特征提取和表征学习将现有知识纳入推理模型的能力而备受关注<sup>[16-18]</sup>。这种方法在很多学科都有广泛的应用,如自动驾驶、自然语言处理、计算机视觉和生物医学。而深度残差网络(如ResNet<sup>[19]</sup>)在许多竞赛中表现出良好的性能,它可以加速模型训练,并在很大程度上避免了由于增加网络层数后在反向传播时出现的梯度消失或梯度爆炸现象。目前还没有研究实现适用于不同患者的深度神经网络模型及方法。本研究以Shen等<sup>[14]</sup>研究为基础,对其模型进行改进,在生成网络的部分反卷积残差块加入快捷路径,并尝试适用于不同患者的基于深度学习重建方法。本研究采用的网络模型如图1所示,整个网络架构分为3部分,分别为表征网络、转换模块和生成网络。

表征网络由5个二维卷积残差块组成,以便从2D投影图像中提取层次语义特征。每个2D卷积块由2D卷积层、2D批处理归一化层、ReLU层、2D卷积层、2D批处理归一化层和ReLU层按顺序组成,可以从2D图像中提取出特征信息。第一卷积层使用3×3的卷积核和2×2的滑动步长执行卷积操作将特征图的空间大小下采样一半。为了保持高维特征信息的稀疏性,增加滤波器的数量将特征映射的通道数增加一倍。随后经过批标准化层<sup>[20]</sup>后通过ReLU层将特征信息向后馈送。第二卷积层使用3×3的卷积核和1×1的滑动步长保持特征图的形状大小。另外,在应用第二个ReLU层之前,使用额外的快捷路径将第一个卷积层的输入相加获得最终输出。通过建立身份映射的快捷路径,鼓励第二卷积层学习残差特征表示。

转换模块在表征网络后,通过卷积和反卷积连接表征网络和生成网络,并关联2D和3D特征表示。将表征网络的输出进行整形,跨维度的变换特征表示以便于后续生成3D体积。模型中为了更好地通过该模块进行特征信息的转移,删除了转换模块中的批标准化层。

生成网络由3D反卷积层、3D批归一化层、ReLU层、3D反卷积层、3D批归一化层和ReLU层的3D反卷积块组成。第一个反卷积层使用4×4×4的卷积核和2×2×2的滑动步长将特征图的空间大小上采样2倍。第二个反卷积层使用3×3×3的卷积核和1×1×1的滑动步长保持特征图的形状大小。每个反卷积层之后是3D批归一化层和ReLU层,生成网络主要由4

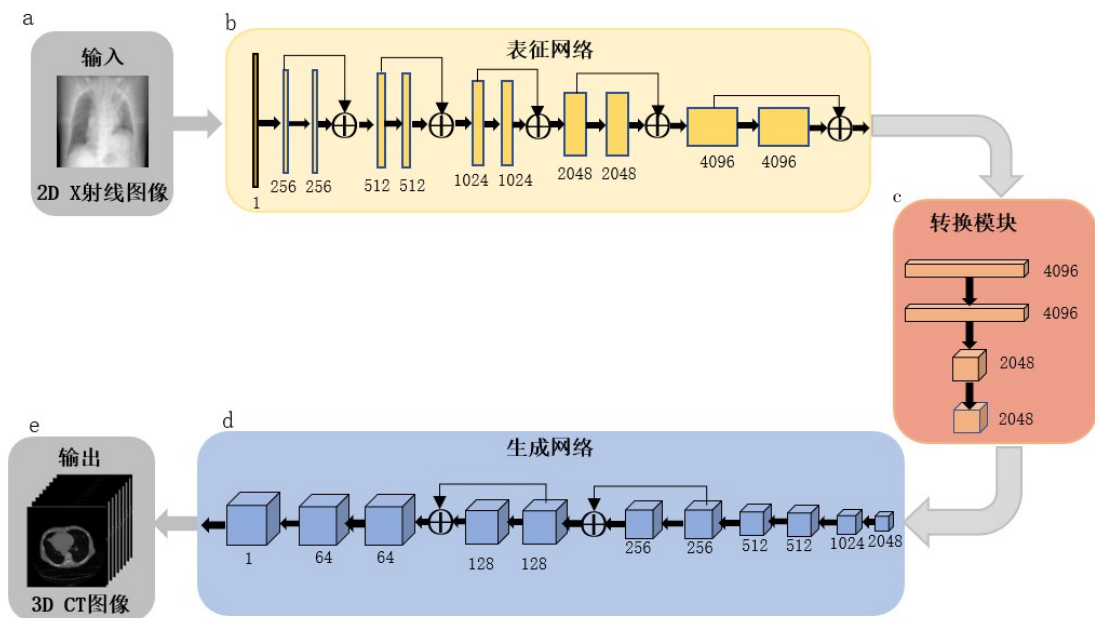


图1 深度学习网络架构

Fig.1 Architecture of deep learning network

a: 模型的输入是单个2D X射线图像;b: 表征网络;c: 转换模块;d: 生成网络;e: 模型的输出是3D CT图像

个反卷积块组成,第二和第三反卷积块使用额外的快捷路径将第一个反卷积层的输入相加获得最终输出。在生成网络的末端,有一个由3D卷积层和2D卷积层(卷积核大小为1)组成的输出变换模块,用于输出3D图像。输出变换模块中的批归一化层被移除,并且最后的卷积层后边去掉了ReLU层。生成网络基于从表征网络学习到的特征信息生成具有细微结构的3D CT图像。

1.2 材料及数据增强方法

本研究采用3例肺癌患者的4D CT图像数据,数据来自上海市质子重离子医院。使用Plastimatch软件将患者的CT图像生成对应的数字重建放射影像(DRR),并使用一个Nvidia Titan RTX图形处理器(Graphics Processing Unit, GPU)训练模型。

由于实际情况限制,我们并不是实际获取了大量的2D X射线图像和对应的相同时刻的CT图像进行监督训练,而是利用患者的CT图像模拟生成的DRR与其对应的CT图像进行训练。在训练深度学习模型时,由于数据有限需要进行数据增强。本研究纯粹从深度学习的数据增强角度出发,对CT数据进行平移、变形以及大角度旋转,最终产生了类似于传统CT重建的多角度投影数据域。在研究中,单例患者时,将其10个时相的4D CT数据按照上述数据增强策略扩充为2 940个CT,并用Plastimatch软件生成对应的DRR,每一个CT及其对应生成的DRR就是一个数据样本,最终共产生2 940个数据样本,将所有数据样本随机打乱后按照6:2:2的比例划分为1 764个训练样本、588个验证样本和

588个测试样本。研究范围是3例患者时,按照上述方法最终共产生8 820个数据样本,其中包括5 292个训练样本、1 764个验证样本以及1 764个测试样本。

1.3 数据预处理

在将数据样本输入网络之前对其进行预处理。与其它利用深度学习方法对图像重建的研究<sup>[11]</sup>类似,由于GPU内存限制以及计算效率的影响,对CT图像进行下采样处理,将CT图像的大小由512×512调整为128×128。所有的数据样本都被调整为相同的大小,将2D DRR图像的大小调整为128×128。单例患者时,将其CT图像仅进行下采样处理,不改变层数,最终得到127×128×128的CT断层图像。研究3例患者时,将3例患者的CT断层图像大小统一调整为123×128×128。具体地说就是将其余两例患者的CT图像从最上边分别减少对应层数与CT图像层数最少的患者相同,这种方法在保证将3例患者的数据顺利输入网络训练的同时还可以确保3例患者的CT图像中的关键信息没有减少。此外,为了使模型在训练时更好的收敛,将2D DRR图像和CT图像的像素强度归一化到[0, 1]。对输入2D DRR图像进行标准差归一化,计算3类样本中所有数据的平均值和标准差,然后用输入2D DRR图像减去对应样本域的平均值,然后除以标准差,使其像素强度的统计分布更接近标准高斯分布,得到最终输入网络的2D图像数据。

1.4 训练策略

输入图像X为2D DRR图像,训练深度网络模型



用来预测CT图像 $Y_{pred}$ ,使预测图像 $Y_{pred}$ 尽可能接近真实图像 $Y_{truth}$ 。将成本函数定义为预测值与真实值之间的均方误差,并通过随机梯度下降迭代优化模型。基于PyTorch库构建网络,并用Adam优化器最小化损失函数,通过反向传播迭代更新网络参数。单例患者时,学习率设置为0.000 03,由于内存限制使用大小为1的小批量进行训练。训练周期为100次,每个训练周期结束后将模型在验证机上进行验证,最后将验证损失最小的模型参数当作最终的模型参数。对3例患者进行研究时,学习率设置为0.000 03,每次向网络输入的批量为1。模型共训练30次,在每个训练周期结束后,将模型在验证集上进行验证,以便监控模型的性能并避免模型在训练集上发生过拟合。如果连续6个训练周期内验证损失没有下降就将学习率自动调整为原来的二分之一,最后将验证集上表现最好也就是验证损失最小的模型参数当作最后的模型参数。在模型训练过程中,网络模型学习从2D DRR图像到CT图像的映射函数。本研究单例患者训练时间大约需要43 h,3例患

者训练时间大约为52 h。在测试过程中,单例患者和3例患者的一个测试样本的3D重建时间约为0.5 s。

1.5 结果评价指标

将最终的模型部署在测试数据集上,使用定性和定量的评估指标分析重建结果并评估模型的性能。使用平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Square Error, RMSE)、结构相似性(Structural Similarity, SSIM)和峰值信噪比(Peak Signal Noise Ratio, PSNR)4种不同的度量标准来衡量预测的CT图像的质量。如表1所示,计算所有测试样本的平均值得到4种度量标准。在机器学习中,通常使用MAE、均方误差(MSE)估计预测图像和真实图像之间的差别,取MSE的平方根得到RMSE。MAE、MSE是 $Y_{pred}$ 和 $Y_{truth}$ 之间的L1范数、L2范数误差。SSIM是通过滑窗的方法计算两幅图像的结构相似性,是衡量两幅图像相似度的常用指标。PSNR是影响图像质量的最大信号功率和噪声功率之间的比率,被广泛用于衡量图像重建的质量。

表1 单例患者和3例患者的重建结果  
Tab.1 Reconstruction results for a patient and 3 patients

研究对象	平均绝对误差	均方根误差	结构相似性	峰值信噪比/dB
1例患者(Shen等 <sup>[14]</sup> 模型)	0.009	0.084	0.971	34.946
1例患者(本研究的模型)	0.008	0.073	0.978	36.153
3例患者(本研究的模型)	0.006	0.079	0.982	38.424

2 结果

图2是单例患者和3例患者训练过程中训练样本

和验证样本的损失曲线,可以看到模型很好地拟合了训练数据,并且在训练集以外的数据中也表现很好。

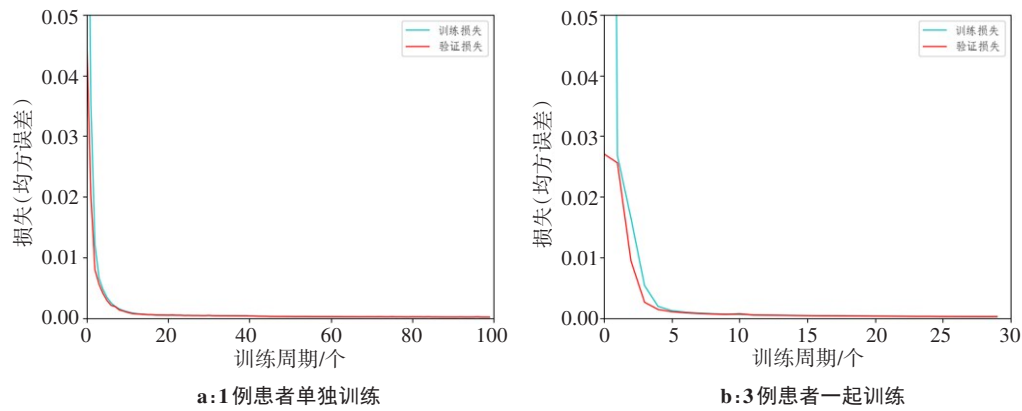


图2 训练损失和验证损失曲线  
Fig.2 Training loss and validation loss curves

表1是各项评估指标在测试样本上的平均值,表中第一行数据是用Shen等<sup>[14]</sup>模型得出的结果,第二

行和第三行是使用本研究改进后的模型以及数据增强方法得出的结果。从表中可以清楚地看出,本研

究所提出的模型和数据增强方法在单独训练单例患者模型中可以得到更好的模型参数。此外,对于不同患者,同样可以得到比较好的模型参数,除了均方根误差略高于单例患者时的结果,其余3个指标都比单例患者时的结果更好。

为了评估该方法的性能,将研究中最终得到的网

络模型部署在测试数据集上。图3展示了单例患者以及适用于不同患者的CT图像重建结果以及真实CT图像与预测CT图像之间的差异图像。结果表明,本研究提出的方法可以对3例不同的患者进行较好的CT图像重建,说明深度学习重建CT图像方法具有很大的潜力,进一步证实了本研究方法的可行性。

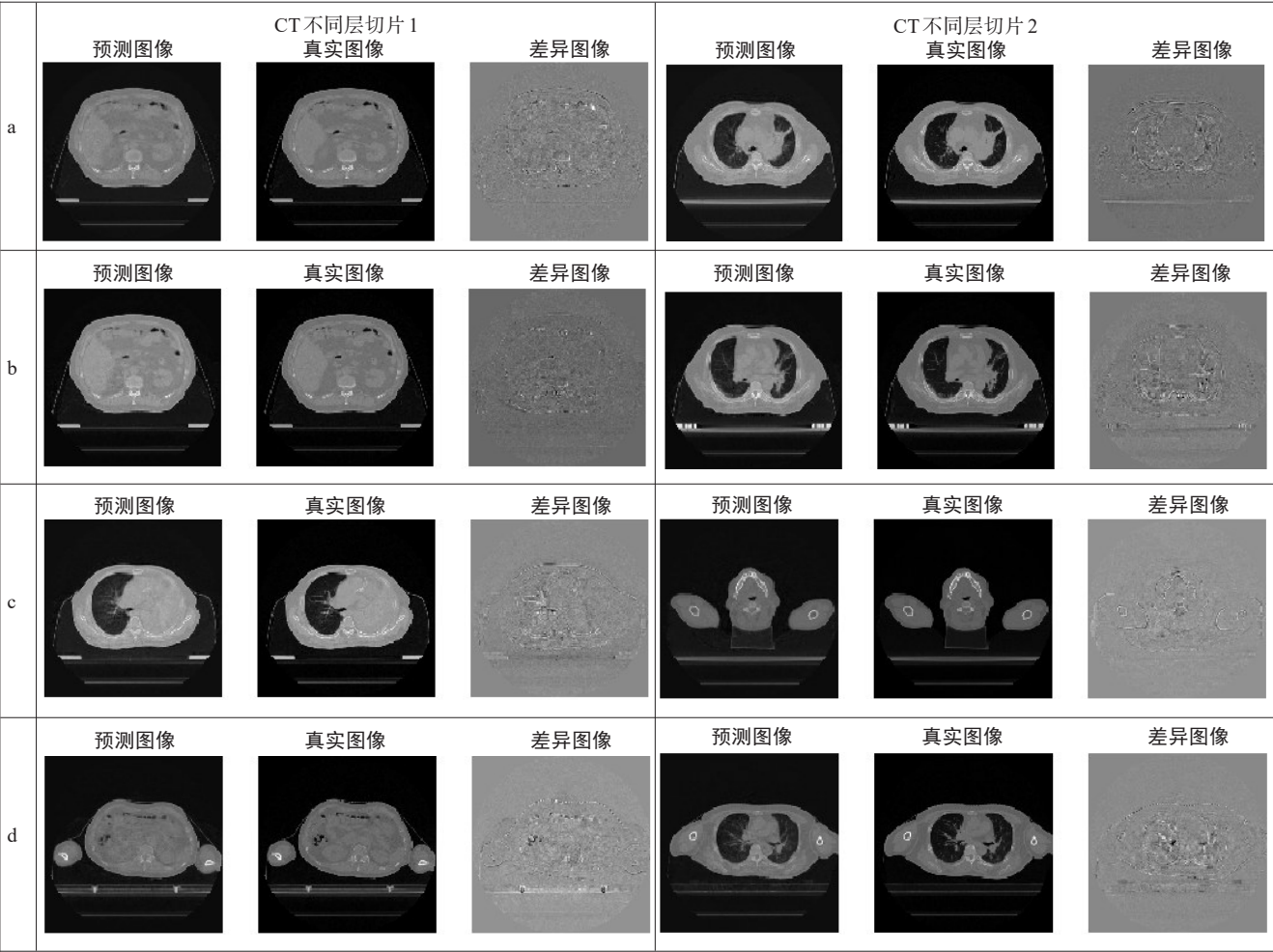


图3 CT重建结果

Fig.3 CT reconstruction results

图a是单例患者的CT重建图像横断面图像;图b~图d分别是3例患者每个患者对应的CT重建图像横断面图像

### 3 讨论

许多不同的3D体积可以产生对应的2D放射影像,然而将这一过程进行反转是非常具有挑战性的。为了减少重建过程中数据采集量和成像剂量并提高重建速度,本研究提出一种基于深度学习的CT单视图断层成像3D重建方法。不同于传统的CT重建方法,本研究利用深度学习,仅使用单个视图就能在短时间内完成CT重建。与现有的深度学习方法相比,本研究将重建对象由1例患者推广为多例患者,实现了针对不同例数患者的CT单视图断层成像3D重

建,而不必再对每例患者单独进行重复的数据增强及模型训练,节省约70%的训练时间。

本研究中的方法证实了通过深度学习构建一个普适于更多患者的网络模型的可行性,有望简化临床上的CT成像设备,并可以为目前放射治疗中图像引导遇到的问题提供潜在的解决方案。未来我们还会对本研究的神经网络模型做进一步的优化,并使用更多例数的小批量患者数据得到普适性进一步提高的模型。

## 【参考文献】

- [1] LUSTIG M, DONOHO D, PAULY J M. Sparse MRI: the application of compressed sensing for rapid MR imaging[J]. Magn Reson Med, 2007, 58(6): 1182-1195.
- [2] SIDKY E Y, PAN X C. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization[J]. Phys Med Biol, 2008, 53(17): 4777-4807.
- [3] CHOI K, WANG J, ZHU L, et al. Compressed sensing based cone-beam computed tomography reconstruction with a first-order method [J]. Med Phys, 2010, 37(9): 5113-5125.
- [4] CANDLES E J, ROMBERG J K, TAO T. Stable signal recovery from incomplete and inaccurate measurements[J]. Commun Pur Appl Math, 2006, 59(8): 1207-1223.
- [5] YU H, WANG G. Compressed sensing based interior tomography[J]. Phys Med Biol, 2009, 54(9): 2791-2805.
- [6] CHEN G H, TANG J, LENG S H. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets[J]. Med Phys, 2008, 35(2): 660-663.
- [7] JI S H, XUE Y, CARIN L. Bayesian compressive sensing[J]. IEEE Trans Signal Proces, 2008, 56(6): 2346-2356.
- [8] STAYMAN J W, FESSLER J A. Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction[J]. IEEE Trans Med Imaging, 2000, 19(6): 601-615.
- [9] ZHU B, LIU J Z, CAULEY S F, et al. Image reconstruction by domain-transform manifold learning[J]. Nature, 2018, 555(7697): 487-492.
- [10] NOMURA Y, XU Q, SHIRATO H, et al. Projection-domain scatter correction for cone beam computed tomography using a residual convolutional neural network[J]. Med Phys, 2019, 46(7): 3142-3155.
- [11] WU Y, MA Y J, CAPALDI D P, et al. Incorporating prior knowledge via volumetric deep residual network to optimize the reconstruction of sparsely sampled MRI[J]. Magn Reson Imaging, 2020, 66: 93-103.
- [12] PREISWERK F, TOEWS M, CHENG C C, et al. Hybrid MRI-ultrasound acquisitions, and scannerless real-time imaging[J]. Magn Reson Med, 2017, 78(3): 897-908.
- [13] HENZLER P, RASCHE V, ROPINSKI T, et al. Single-image tomography: 3D volumes from 2D cranial X-rays[J]. Comput Graph Forum, 2018, 37(2): 377-388.
- [14] SHEN L Y, ZHAO W, XING L. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning[J]. Nat Biomed Eng, 2019, 3(11): 880-888.
- [15] LEI Y, TIAN Z, WANG T, et al. Deep learning-based real-time volumetric imaging for lung stereotactic body radiation therapy: a proof of concept study[J]. Phys Med Biol, 2020, 65(23): 1-14.
- [16] ESLAMI S M A, REZENDE D J, BESSE F, et al. Neural scene representation and rendering[J]. Science, 2018, 360(6394): 1204-1210.
- [17] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [18] SCHMIDHUBER J. Deep learning in neural networks: an overview[J]. Neural Netw, 2015, 61: 85-117.
- [19] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Computer Vision And Pattern Recognition. IEEE, 2016: 770-778.
- [20] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the International conference on machine learning. PMLR, 2015: 448-456.

(编辑:陈丽霞)