A method of rapid quantification of patient-specific organ doses for CT using deep-learning-based multi-organ segmentation and GPU-accelerated Monte Carlo dose computing

Zhao Peng
Department of Engineering and Applied Physics, University of Science and Technology of China, Hefei, Anhui 230026, China

Xi Fang, Pingkun Yan and Hongming Shan
Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Tianyu Liu
Department of Mechanical, Aerospace and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Xi Pei
Department of Engineering and Applied Physics, University of Science and Technology of China, Hefei, Anhui 230026, China
Anhui Wisdom Technology Company Limited, Hefei, Anhui 238000, China

Ge Wang
Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Bob Liu and Mannudeep K. Kalra
Department of Radiology, Massachusetts General Hospital, Boston, MA 02114, USA

X. George Xu[a)]
Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
Department of Mechanical, Aerospace and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

(Received 1 October 2019; revised 6 February 2020; accepted for publication 29 February 2020; published xx xxxx xxxx)

Purpose: One technical barrier to patient-specific computed tomography (CT) dosimetry has been the lack of computational tools for the automatic patient-specific multi-organ segmentation of CT images and rapid organ dose quantification. When previous CT images are available for the same body region of the patient, the ability to obtain patient-specific organ doses for CT — in a similar manner as radiation therapy treatment planning — will open the door to personalized and prospective CT scan protocols. This study aims to demonstrate the feasibility of combining deep-learning algorithms for automatic segmentation of multiple radiosensitive organs from CT images with the GPU-based Monte Carlo rapid organ dose calculation.

Methods: A deep convolutional neural network (CNN) based on the U-Net for organ segmentation is developed and trained to automatically delineate multiple radiosensitive organs from CT images. Two databases are used: The lung CT segmentation challenge 2017 (LCTSC) dataset that contains 60 thoracic CT scan patients, each consisting of five segmented organs, and the Pancreas-CT (PCT) dataset, which contains 43 abdominal CT scan patients each consisting of eight segmented organs. A fivefold cross-validation method is performed on both sets of data. Dice similarity coefficients (DSCs) are used to evaluate the segmentation performance against the ground truth. A GPU-based Monte Carlo dose code, ARCHER, is used to calculate patient-specific CT organ doses. The proposed method is evaluated in terms of relative dose errors (RDEs). To demonstrate the potential improvement of the new method, organ dose results are compared against those obtained for population-average patient phantoms used in an off-line dose reporting software, VirtualDose, at Massachusetts General Hospital.

Results: The median DSCs are found to be 0.97 (right lung), 0.96 (left lung), 0.92 (heart), 0.86 (spinal cord), 0.76 (esophagus) for the LCTSC dataset, along with 0.96 (spleen), 0.96 (liver), 0.95 (left kidney), 0.90 (stomach), 0.87 (gall bladder), 0.80 (pancreas), 0.75 (esophagus), and 0.61 (duodenum) for the PCT dataset. Comparing with organ dose results from population-averaged phantoms, the new patient-specific method achieved smaller absolute RDEs (mean ± standard deviation) for all organs: 1.8% ± 1.4% (vs 16.0% ± 11.8%) for the lung, 0.8% ± 0.7% (vs 34.0% ± 31.1%) for the heart, 1.6% ± 1.7% (vs 45.7% ± 29.3%) for the esophagus, 0.6% ± 1.2% (vs 15.8% ± 12.7%) for the spleen, 1.2% ± 1.0% (vs 18.1% ± 15.7%) for the pancreas, 0.9% ± 0.6% (vs 20.0% ± 15.2%) for the left kidney, 1.7% ± 3.1% (vs 19.1% ± 9.8%) for the gallbladder, 0.3% ± 0.3% (vs 24.2% ± 18.7%) for the liver, and 1.6% ± 1.7% (vs 19.3% ± 13.6%) for the stomach. The trained automatic segmentation tool takes <5 s per patient for all 103 patients in the dataset. The Monte Carlo radiation dose calculations performed in parallel to the segmentation process using the

GPU-accelerated ARCHER code take <4 s per patient to achieve <0.5% statistical uncertainty in all organ doses for all 103 patients in the database.
Conclusion: This work shows the feasibility to perform combined automatic patient-specific multi-organ segmentation of CT images and rapid GPU-based Monte Carlo dose quantification with clinically acceptable accuracy and efficiency. © *2020 American Association of Physicists in Medicine* [https://doi.org/10.1002/mp.14131]

Key words: convolutional neural network, CT organ dose, Monte Carlo, multi-organ segmentation, patient-specific

1. INTRODUCTION

In the United States, the number of computed tomography (CT) examinations increased drastically between the 1980s and 2010s due to rapid improvements in multidetector CT (MDCT) technologies.[1–3] In 2018, about 88.7 million CT examinations were performed in the United States alone, which represented a substantial increase from 21 million exams in 1995.[4] The abdomen and chest regions represent the most frequently scanned body regions, accounting for more than a third of all CT examinations. Given the rising use of CT and concerns over associated radiation risks, the American College of Radiology (ACR) has called for more research and development in patient-specific dose quantification, scanner optimization, and protocol comparison.[1]

Computed tomography dose index volume (CTDIvol) and dose length product (DLP) are technical dose descriptors and do not represent or take into account patient body habitus (size or shape), attenuation, scanned anatomy, age, gender, or actually absorbed radiation doses.[5] Although CTDIvol and DLP provide a good way to compare scanners and scan protocols, they cannot be used to compare, monitor, or assess patient-specific radiation doses from CT. For this reason, size-specific dose estimates (SSDE) have been recommended as an improved approach that take into account patient body habitus.[6] Many methods of generating organ-specific CT dose databases have been reported.[7–15] These methods require Monte Carlo simulations of CT scanner components as well as radiation interactions with whole-body computational phantoms that contain organs/tissues explicitly defined in tiny voxels in accordance with the "Reference-Man" concept — population-averaged anatomical parameters originally defined for radiation protection purposes.[16] However, the process that is required to create such whole-body phantoms is prohibitively complex for routine analysis of patient-specific images. As a result, most clinical end users can only perform CT organ dose assessment using "off-line" software tools, such as VirtualDose,[15] which are based on databases precalculated from a library of population-averaged phantoms. In contrast, in radiation treatment planning, we routinely delineate the target volume and adjacent healthy organs at risk (OARs) using patient-specific images, before performing rapid dose calculation and inverse treatment plan optimization to minimize normal tissue complication probability (NTCP).[17–19] Patient-specific organ dose computing methods already exist. Recently, GPU-based Monte Carlo dose computing codes, including ARCHER,[20,21] for example, have achieved clinically acceptable speeds for both patient CT imaging dose assessment and for treatment planning.

Rising CT utilization has also heightened the concern that patients accrue large cumulative doses from recurrent CT imaging. Sodickson et al. performed a cohort study of 31 462 patients who underwent diagnostic CT in 2007 and had undergone 190 712 CT examinations over the prior 22 yr.[22] The authors discovered that 33% of patients underwent five or more lifetime CT examinations and that 5% of patients underwent between 22 and 132 examinations, leading to the conclusion of the study that, while most patients accrue low radiation-induced cancer risks, a subgroup is potentially at higher risk due to recurrent CT imaging. A recent survey performed in 2019 on 90 146 CT patients at Massachusetts General Hospital found that about 63% of chest scan patients have received at least one previous CT scan between 2014 and 2019.[23] The percentage for recurrent abdomen/pelvis scan patients is 50% and is 40% for head-scan patients. Approximately 50% of patients in the US undergoing CT scan have prior CT images. To take advantage of prior CT scans of the same patient, CT dose optimization will require the current "retrospective dosimetry" paradigm to be replaced by the "prospective dosimetry" paradigm, in which organ dose information is used to guide subsequent CT scans of that patient.

A prospective patient-specific organ dose method will be a game changer in CT dosimetry and can help extend the existing tube current modulation techniques by taking full advantage of organ localization and distribution of organ doses. It can also help imaging physicians make informed and prospective decisions regarding the delivery of doses based on the clinical question, expected disease distribution, and organ dose distribution. Such prospective decisions regarding radiation dose delivery from CT can help usher in personalized scan protocols with truly organ dose-modulated techniques. Among the current technical barriers are the lack of clinically acceptable organ segmentation and rapid organ dose computing tools for CT.

Organ dose quantification for a set of radiosensitive organs in every patient undergoing CT scans is important for radiation protection purposes (instead of cancer treatment planning purposes). Segmentation of radiosensitive organ volumes from CT images has long been a challenging task to the medical physics community.[24] Manual organ segmentation is labor intensive and user dependent, making the

approach impractical for clinical applications involving patient-specific images. Until recently, methods of automatic segmentation of organs relied on low-level image features that require strong prior knowledge about the anatomical structures, both of which are insufficient for clinical use.[25] The advent of deep learning methods involving convolutional neural network (CNN) has brought an unprecedented level of innovation to the field of image segmentation.[26–29] The state-of-the-art models in organ segmentation are variants of encoder–decoder architecture such as the fully convolutional networks (FCNs)[30] and U-Net.[31] However, these models are usually trained for specific organs and cannot be easily extended to multi-organ segmentation needed for CT organ dosimetry. Recently, Trullo et al.[32] used a modified two-dimensional FCN to segment four OARs from CT images and apply conditional random fields to further improve the segmentation performance. Gibson et al.[33] applied a three-dimensional (3D) Dense V-Network to segment eight organs from CT images for navigation purposes in endoscopic pancreatic and biliary procedures. However, these studies did not perform organ dose calculations for patients who receive the CT scans. Recent studies by other groups that did consider CT organs dose evaluations employed traditional organ segmentation algorithm such as feature-based or atlas-based methods.[34,35] Finally, without the necessary accuracy and efficiency, patient-specific dosimetry tools would not become a viable part of the clinical workflow.

This study[36] aims to demonstrate the feasibility of a streamlined fast patient-specific CT organ dose assessment method that performs segmentation of multiple organs from patient-specific CT images using deep CNN algorithms and GPU-accelerated Monte Carlo dose calculations using the ARCHER code in a parallel computational workflow as illustrated in Fig. 1. This is the first study to combine these two tools to achieve the computational accuracy and efficiency required for routine clinical applications. In subsequent sections, we describe steps and methods, summarize results, and discuss limitations before drawing conclusions.

2. MATERIALS AND METHODS

2.A. Organ segmentation

2.A.1. Datasets and image preprocessing

In this study, two publicly available datasets were used: (a) The 2017 lung CT segmentation challenge (LCTSC),[37–39] which contains 60 thoracic CT scan patients with five segmented organs (left lung, right lung, heart, spinal cord, and esophagus), and (b) Pancreas-CT (PCT), which contains 43 abdominal contrast enhanced CT scan patients with eight segmented organs (the spleen, left kidney, gallbladder, esophagus, liver, stomach, pancreas, and duodenum).[28,33,39,40] For each patient in these two datasets, the Hounsfield Unit (HU) values were processed using a minimum threshold of −200 and a maximum threshold of 300 prior to being normalized to yield values between 0 to 1. In order to focus on organs and suppress the background information, we cropped and reserved the regions of interest according to the body contour in the original CT images and used it as training data. Finally, to circumvent the computer memory limitation, data resampling was performed using linear interpolation for CT images and using nearest interpolation for the labels. The interpolation operations are standard routine implementations in common image processing software. In our implementation, we used a python package called scipy.[41] In the linear interpolation, the value of a pixel after resampling is computed as the weighted average of its surrounding pixels, where the weights are calculated based on the distances to the target location. In the nearest-neighbor interpolation, the value of the nearest pixel around the target location is assigned to the target pixel after resampling.[42] For the LCTSC dataset, the original slice resolution is from 1.0 mm × 1.0 mm to 1.4 mm × 1.4 mm and the slice thickness is from 1.0 to 3.0 mm. The resulting resolution after resampling is 2.0 mm × 2.0 mm × 2.5 mm. For the PCT dataset, the original slice resolution is from 0.7 mm × 0.7 mm to 1.0 mm × 1.0 mm, and the original slice thickness is 1.0 mm. Here, we followed the

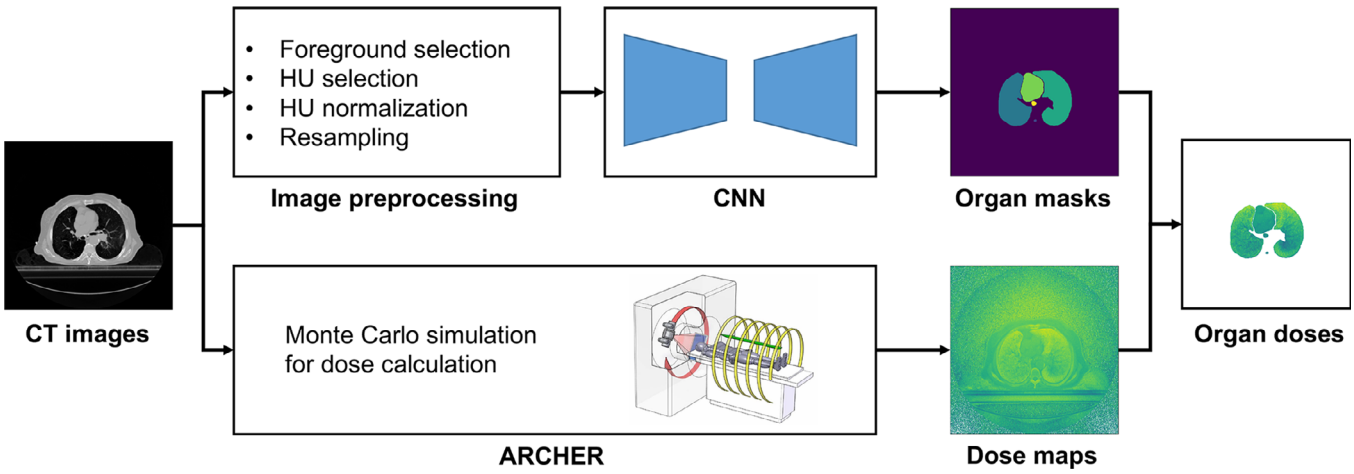

FIG. 1. The overall parallel computational process of the method of patient-specific organ dose assessment for computed tomography combining convolutional neural network-based multi-organ segmentation and a GPU-based Monte Carlo dose engine, ARCHER.

methods described by Gibson et al.,[33] and the size of CT images was resampled to 144 × 144 × 144 pixels. So after resampling, the resolution for each patient is different; specifically, the slice resolution is from 1.8 mm × 1.4 mm to 2.5 mm × 2.1 mm, and the slice thickness is from 1.1 to 1.6 mm.

2.A.2. Network architecture

The proposed network in this study is based on the 3D U-Net.[43] As shown in Fig. 2, the network consists of an encoder and a decoder. The role of the decoder network is to map the low resolution encoder feature maps to full input resolution feature maps for pixel-wise classification.[44] The encoder contains four repeated residual blocks. Each block consists of four convolutional modules and each convolutional module is composed by a convolution layer with the kernel of 3 × 3 × 3, an instance normalization, and a leaky rectified linear unit. For each residual block, the stride of convolution layer in the convolutional modules is 1 × 1 × 1 except for the last convolutional module in which the stride is 2 × 2 × 2 to achieve the purpose of downsampling, and there is a spatial dropout layer between the early two convolutional modules to prevent the network from overfitting. The decoder contains four repeated segmentation blocks. Each block consists of two convolutional modules and one deconvolutional module. Four dashed arrows in the figure indicate four skipping connections that copy and reuse early feature maps as the input to later layers that have the same feature map size by a concatenation operation to preserve high-resolution features. In the final three segmentation blocks, a 1 × 1 × 1 convolution layer is used to map the feature tensor to the probability tensor with the channels of the desired number of classes, n, before all results are merged by the upsampling operation to enhance the precision of segmentation results. Finally, a SoftMax activation is used to output a probability of each class for every voxel.[45]

2.A.3. Training and validation

The fivefold cross-validation method was adopted for this work.[46] The entire dataset is randomly split, using the "random.shuffle()" function in Python, into five non-overlapping subsets for training, validation, and testing in the ratio of 3:1:1 (i.e., three subsets for training, one subset for validation, and one subset for testing). Specifically, for the LCTSC dataset, a total of 60 patients are divided into five subsets (each having 12 patients). For the PCT dataset, a total of 43 patients are divided into five subsets (each having eight or nine patients). The validation process is used to monitor the training process and to prevent overfitting. To reduce potential bias, randomly split five subsets are rotated five times to report the average performance over these five different hold-out testing subsets, as illustrated in Fig. 3. The fivefold cross validation strategy is key to ensuring the independence of the testing data, that is, each sample is used in the testing subsets only once.

At the training stage, patches are first randomly extracted from the resampled CT images to achieve data diversity and to prevent overfitting. The patch size is 96 × 96 × 96 in LCTSC and 128 × 128 × 128 in PCT. Figure 4 shows an example of such patches from LCTSC used in the training in terms of axial, sagittal, and coronal views. The patch-based training method addresses the problem of different sizes of CT images as well as the requirement of data augmentation. An advantage is that it enhances the robustness of the network model. The limitation is that it may negatively impact the predicted performance due to the lack of global information when the patch size is too small. The network can be aware of the Z location of the patches implicitly because the patch image in the different Z location is different, and the patches in any Z location are trained. The orientation of all patches is the same, so the right lung and left lung have different positions, and they have different shapes in the patch. Therefore, the right and left lungs can be differentiated by their position and shape although they have a similar pixel value.

Then, the network is trained by the patch and its corresponding labels. The loss function is defined as the weighted dice similarity coefficient as:

$$\text{Loss} = -\frac{1}{N \times K}\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{2 \times \sum_{v=1}^{V}\left(p_{i,k,v} \times y_{i.k.v}\right) + \varepsilon}{\sum_{v=1}^{V} p_{i,k,v} + \sum_{v=1}^{V} y_{i,k,v} + \varepsilon},$$

where $p_{i,k,v}$ is the predicted probability of the voxel v of the sample i belonging to the class k, $y_{i,k,v}$ is the ground truth label (0 or 1), N is the number of samples, K is the number of classes, V is the number of voxels in one sample, and ε is a smooth factor (set to be 1 in this study). The initial learning rate is 0.0005, and the Adam algorithm[47] is used to update the parameters of the network. The validation loss is calculated for every epoch, and the learning rate is halved when the validation loss no longer decreases after 30 consecutive epochs. To prevent overfitting, the training process is terminated when the validation loss on longer decreases after 50 consecutive epochs.

2.A.4. Testing

In the testing stage, patches are first extracted from each CT images with a moving window with the size of 96 × 96 × 96 in LCTSC and 128 × 128 × 128 in PCT. The stride is 48 in LCTSC and eight in PCT. In other words, multiple patches are extracted from one patient and fed into the network. The output of the network is a probability tensor for each patch. Then, all probability tensors are merged from the same patient with a mean operator in the overlapping area to obtain the final probability tensor. Next, the class of each voxel is determined by the largest probability, which is the preliminary results of organ segmentation, and the value of each voxel is the class number. Last, using the nearest-neighbor interpolation, the preliminary segmentation results are resampled to the size of original CT images to obtain the final organ segmentation results.

All experiments described above were performed on a Linux computer system. Keras with TensorFlow as the

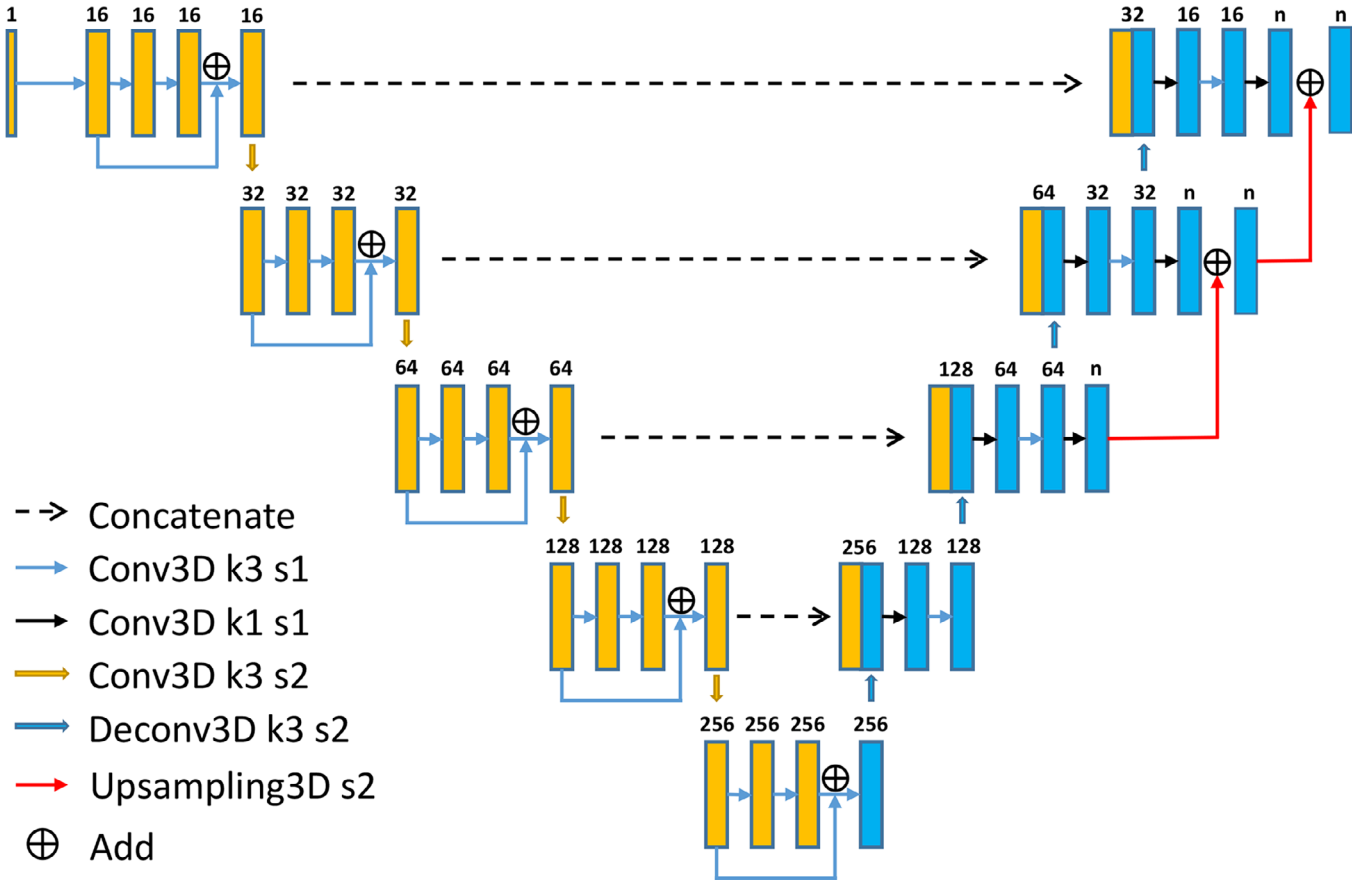

FIG. 2. The network architecture.

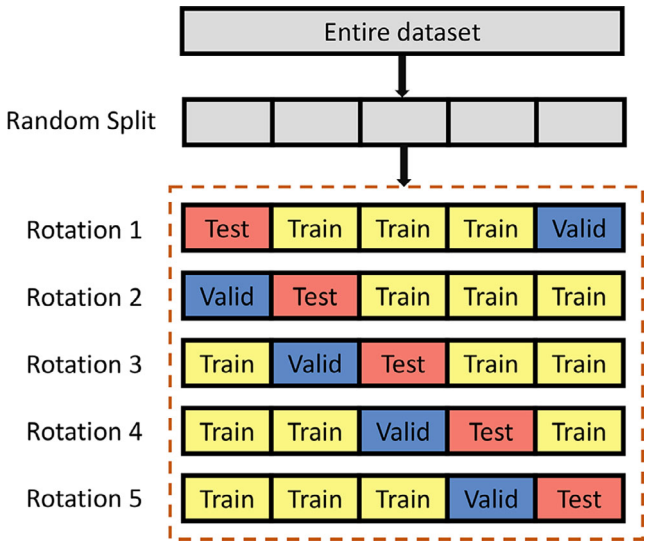

FIG. 3. Example of splitting and rotation using the fivefold cross-validation method for the dataset involving five subsets.

backend was used as the platform for designing and training the neural network.[48] The hardware includes: (a) GPU — Nvidia GeForce Titan X Graphics Card with 12 GB memories, and (b) CPU — Intel Xeon Processor X5650 with 16 GB memories.

2.B. Organ dose calculations

A GPU-accelerated Monte Carlo code, ARCHER, previously developed by members of this group was used in this study to calculate organ doses.[20–21,49] ARCHER simulates the transport of low-energy x-ray photons in heterogeneous media defined by the patient CT images where photoelectric effect, Compton scattering, and Rayleigh scattering can take place. Computed tomography scan protocols are predefined for ARCHER including a combination of scan mode (helical or axial), beam collimation (5, 10, or 20 mm), and kVp (80, 100, 120, or 140). Figure 5 illustrates the simulation model involving a patient and a CT scanner.

This study considers a scanner model representing a GE Lightspeed Pro 16 MDCT that has been validated in our previous studies,[50,51] although a newer scanner model can be similarly created when needed in the future. The scanning protocol includes 120 kVp, 20-mm beam collimation, axial body scan at a constant 100 mAs. A CT scanner's continuous rotational motion is simulated using the step-and-shoot pattern, with each rotation approximated by 16 discrete positions.[50] As shown earlier in Fig. 1, the average absorbed dose for each organ of interest is derived by combining the newly segmented organ masks and voxel-wise dose maps calculated by ARCHER for a specific patient — as is done in radiation treatment planning. The computational speed is evaluated to make sure it is acceptable as part of the clinical workflow.

FIG. 4. An example to illustrate patches from lung computed tomography segmentation challenge the database used in the training in terms of axial, sagittal, and coronal views.

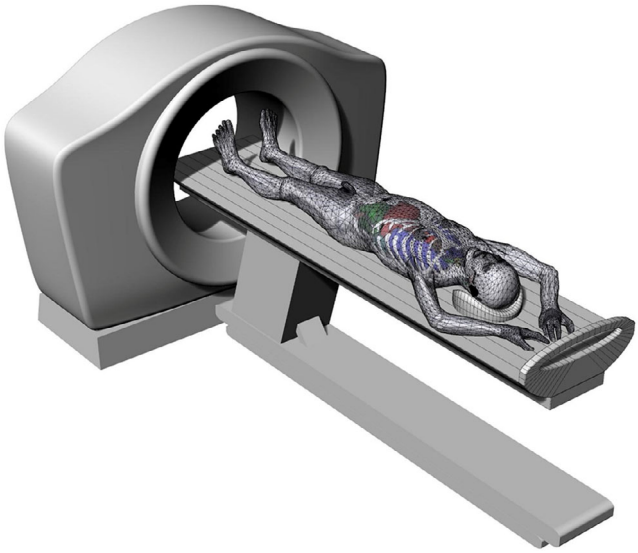

FIG. 5. Computed tomography (CT) dose simulation model of a patient undergoing a CT scan.

To show the potential clinical impact of the new method, patient-specific organ dose results are compared against organ doses derived from population-average phantoms used in the VirtualDose software.[45,52] Figure 6 shows the RPI-Adult Male (73 kg in weight and 176 cm in height) and RPI-Adult Female (60 kg in weight and 163 cm in height) phantoms that were designed in accordance with anatomical parameters for the 50th percentile of the population.[52] When the weight and height of an adult patient are unspecified, the clinical organ dose assessment procedure at Massachusetts General Hospital usually picks these standard adult phantoms from the VirtualDose software to represent that patient.

2.C. Segmentation and organ dose evaluation criteria

The Dice Similarity Coefficient (DSC) is used to evaluate the performance of organ segmentation[53]:

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

where A is the manually segmented organ (i.e., the ground truth) and B is the automatically segmented organ by the network. The DSC ranges from 0 to 1 with the latter indicating a perfect performance. The relative dose error (RDE) was used to evaluate the accuracy of dose calculation for each organ:

$$RDE = \frac{D - D_r}{D_r} \times 100\%$$

where D is the organ dose calculated by ARCHER using either automatically segmented organs in the patient-specific phantom (i.e., our method) or organs in the population-average phantom, and D_r is the reference organ dose calculated by ARCHER using manually segmented organs in the patient-specific phantom.

3. RESULTS

3.A. Organs segmentation

The performance of our network in organ segmentation is evaluated in terms of the DSC. As shown in Fig. 7, the segmentation results of all organs are summarized in these two box plots. For 60 patients from LCTSC, we achieved median DSCs of 0.97 (right lung), 0.96 (left lung), 0.92 (heart), 0.86 (spinal cord), and 0.76 (esophagus), which can be seen in Fig. 7(a). For 43 patients from PCT, we achieved median DSCs of 0.96 (spleen), 0.96 (liver), 0.95 (left kidney), 0.90 (stomach), 0.87 (gall bladder), 0.80 (pancreas), 0.75 (esophagus), and 0.61 (duodenum), which can be seen in Fig. 7(b). Figures 8(a) and 8(b) show visual comparison of manual and automatic multi-organ segmentation results from both LCTSC and PCT, respectively, in axial, sagittal, coronal, and 3D views.

3.B. Organ dose calculations

The accuracy of organ dose calculations is evaluated in terms of RDE for the purposes of CT organ dosimetry, where 10% is generally considered excellent. In the dataset from LCTSC for a total of 60 patients, organs doses are calculated for organs including the lung, heart, and esophagus. The left lung and right lung are treated as one organ, and the RDE of the spinal cord is not considered because it is not segmented

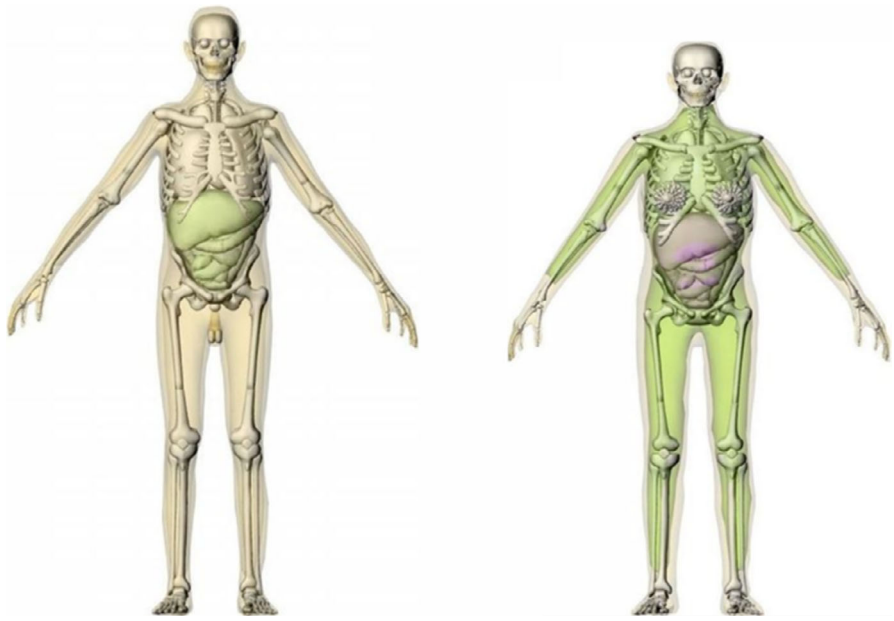

FIG. 6. RPI-Adult Male (left) and RPI-Adult Female (right) phantoms in the VirtualDose software that were designed in accordance with anatomical parameters for the 50th percentile of the population, thus bringing errors when compared with patient-specific organ doses.[52]

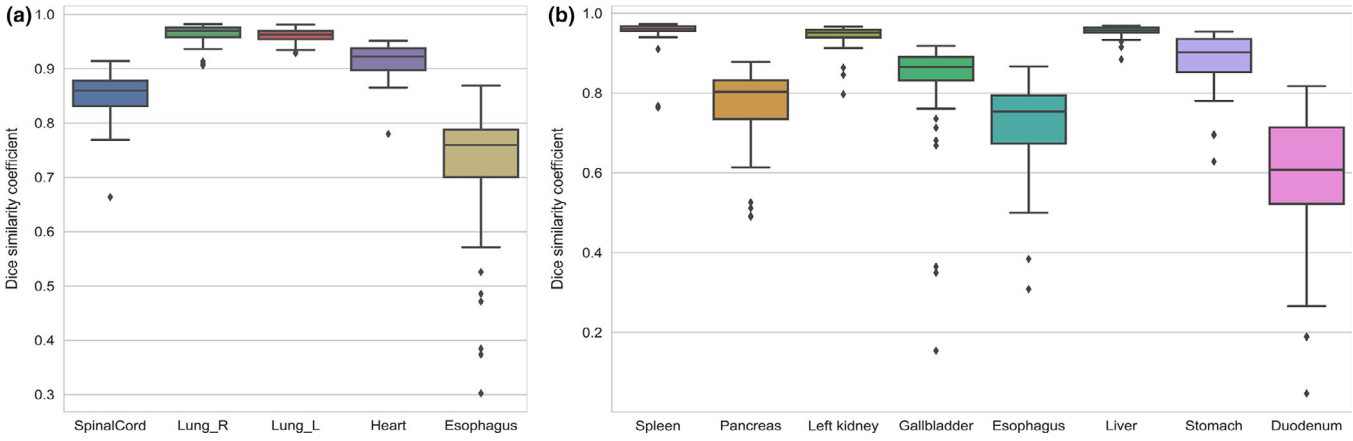

FIG. 7. Evaluation of organ segmentation performance in terms of dice similarity coefficients. (a) Data based on 60 patients from the lung computed tomography segmentation challenge database. (b) Data based on 43 patients from the pancreas-CT database.

in the population-average phantom. In the dataset from PCT for a total of 43 patients, organs doses are calculated for organs including the spleen, left kidney, gallbladder, liver, stomach, and pancreas. The duodenum is not segmented in the population-average phantom, and the esophagus in the specific patient is incomplete in the abdominal CT scanning, so the RDEs are not considered for the duodenum and esophagus. The results are summarized in Table I and further visually compared in Figs. 9(a) and 9(b) using box plots of RDEs for each organ in these two datasets. The ground-truth reference organ doses are calculated for the specific patient using ground-truth segmentation data from the original database. The "proposed method" represents the RDE between organ doses from our automatic segmentation and the reference organ doses, and the "phantom-based method" represents the RDE between organ doses from the population-average phantom and reference organ doses. Comparing with the population-average phantom-based method, our proposed patient-specific method achieved much smaller RDE values. In a CT scan, the height, weight, and organ topology of a patient can influence organ dose values. There is no doubt that it introduces some errors using population-average phantoms to replace a specific patient for organ dose calculation. In the case of dose to the heart, the current method of using population-average phantom in the VirtualDose software is found to have the error range (−15.4%–124.6%) due to the anatomical differences between the phantom and a real patient. The patient-specific method has much smaller errors with the range of −2.9% to 2.6% for the heart due to difference in organ segmentation between the CNN-based method and the

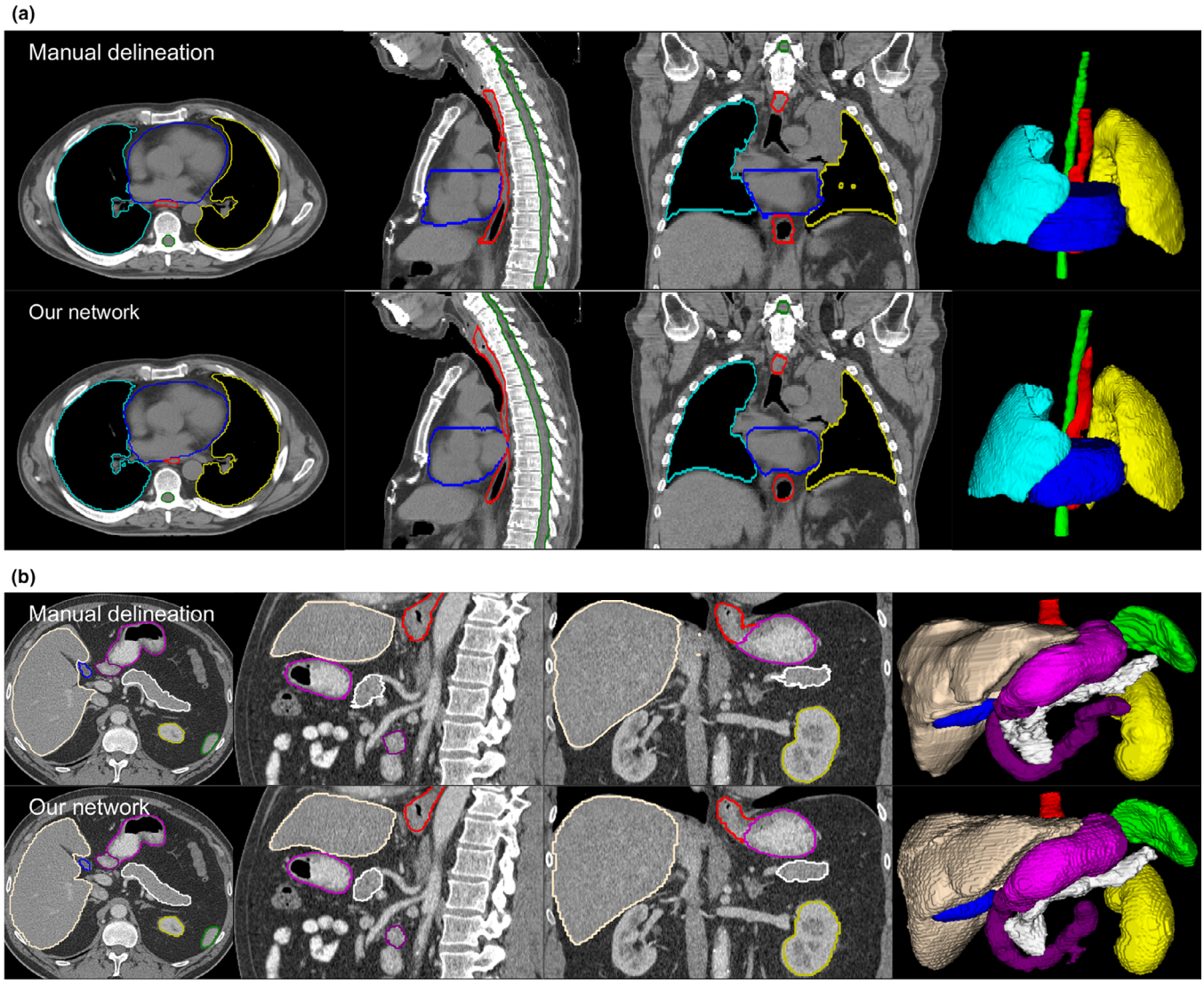

FIG. 8. Examples for visual comparison of organ segmentation between manual methods from lung computed tomography segmentation challenge (LCTSC) or pancreas-CT (PCT) database (showed in the upper row in each panel) and our automatic method (showed in the upper row in each panel), in terms of axial, sagittal, coronal, and 3D views (from left to right). (a) LCTSC database showing left lung (yellow), right lung (cyan), heart (blue), spinal cord (green), and esophagus (red). (b) PCT database showing spleen (green), pancreas (white), left kidney (yellow), gallbladder (blue), esophagus (red), liver (bisque), stomach (magenta), and duodenum (purple).

ground truth. These results suggest that the patient-specific method can bring significant (in the case of dose to the heart, 125/3 times) improvement to the current CT organ dose assessment method that is based on population-average phantoms.

3.C. Computational efficiency

The computing time in our method includes two processes performed in parallel as illustrated previously in Fig. 1. The time for automatic multi-organ segmentation for each patient is <5 s for all 103 patient cases considered in the study. The time to calculate a total of 1×10^8 photons for each patient (for a maximum organ dose statistical uncertainty of 0.5%) using ARCHER code running on an Nvidia Titan RTX GPU card with 24 GB memory is less than 4 s for all 103 patient cases. From our experiences, such computational accuracy and efficiency are expected to be acceptable as part of the routine clinical workflow.

4. DISCUSSION

In this study, we designed a 3D CNN model to automatically segment thoracic and abdominal organs in patient-specific CT images using two publicly available databases. For the duodenum or esophagus, the segmentation performance of our network was found to be relatively poor because the organ and its surrounding tissues have similar pixel values in CT image, making the boundary difficult to detect by the CNN model. Nevertheless, results from this study have clearly demonstrated the accuracy and efficiency of the CNN model in performing the automatic multi-organ segmentation

TABLE I. Comparison of relative dose errors (RDE) of organ doses calculated by the proposed patient-specific method and the population-averaged phantom method.

		Proposed patient-specific method			Population-averaged phantom method		
			Absolute RDE (%)			Absolute RDE (%)	
	Organs	RDE range (%)	Mean	Standard deviation	RDE range (%)	Mean	Standard deviation
Thorax	Lung	−7.5–2.2	1.8	1.4	−21.1–46.4	16.0	11.8
	Heart	−2.9–2.6	0.8	0.7	−15.4–124.6	34.0	31.1
	Esophagus	−9.4–5.0	1.6	1.7	−10.5–125.6	45.7	29.3
Abdomen	Spleen	−7.9–1.0	0.6	1.2	−20.1–57.1	15.8	12.7
	pancreas	−3.4–4.6	1.2	1.0	−20.2–61.1	18.1	15.7
	Left kidney	−2.0–1.9	0.9	0.6	−39.0–70.8	20.0	15.2
	Gallbladder	−15.0–3.9	1.7	3.1	−40.1–14.0	19.1	9.8
	Liver	−0.8–1.3	0.3	0.3	−30.0–72.7	24.2	18.7
	Stomach	−4.6–8.1	1.6	1.7	−47.7–20.8	19.3	13.6

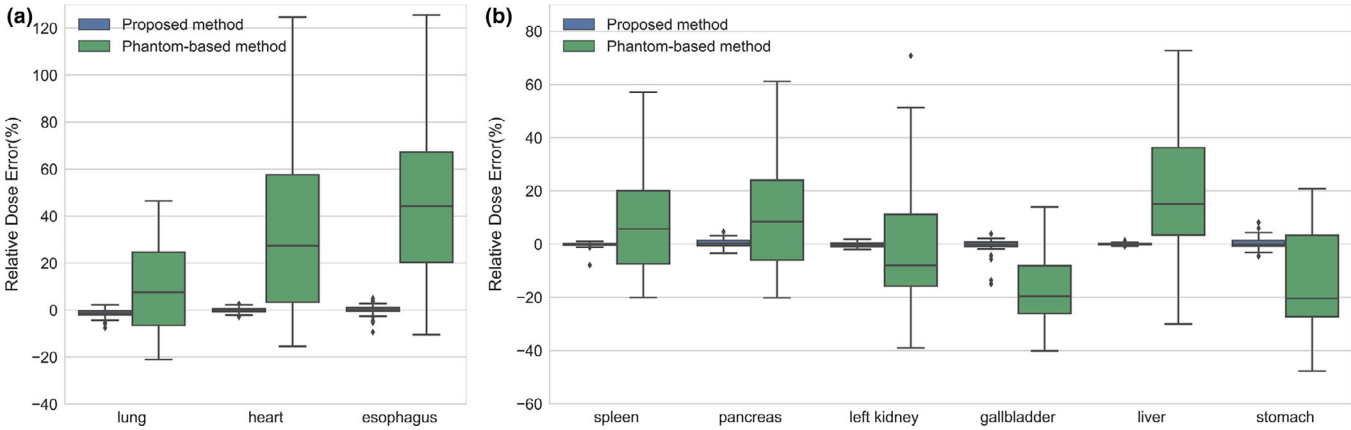

FIG. 9. The box plots of relative dose errors showing that the proposed patient-specific method has much smaller errors than the population-averaged phantom method when evaluated against the ground truth data. (a) For 60 patients from lung computed tomography segmentation challenge and (b) for 43 patients from pancreas-CT.

task for the purposes of assessing patient organ doses. Implementation of the proposed method can lead to significant improvement in the accuracy of organ dose calculation based on the population-average phantoms.

As evidenced in the 2017 AAPM Thoracic Auto-Segmentation Challenge, start-of-the-art automatic segmentation methods, DL-based or atlas-based, can already achieve impressive performances.[38] Therefore, the objective of this study was not to invent a new and better organs segmentation method. Instead, the significance of this study is that, for the first time, we have demonstrated that it is feasible to combine DL-based automatic multi-organ segmentation tool with the GPU-based rapid Monte Carlo dose calculation code in a streamlined process that takes <5 s for each patient. With this newly demonstrated capability of "patient-specific" organ dose assessment, future CT scanners can take advantage of patient- and scan-specific features in a new paradigm — the "prospective" design of tube voltage and current modulation, beam collimation and filtering, and gantry angle — leading to the ultimate goal of achieving low-dose and optimized CT imaging.

There are several limitations in the current study. The variable and somewhat less accurate performance of our approach for segmenting narrow and long structures with poor soft-tissue contrast such as the esophagus and duodenum may be related to the relatively small size of training data in the databases causing irregularities in CT attenuation and position of these structures. Another limitation is that we did not assess the effect of major abnormalities on the organ segmentation — an issue already recognized by organizers of the 2017 lung CT segmentation challenge.[38] Likewise, diffuse abnormalities and paucity of intra-abdominal fat can have a negative effect on the ability of our segmentation algorithm. Further studies should consider larger patient data sizes, covering children and including additional radiosensitive organs in the head and neck regions. One set of unique data already available from MGH is the annotated cadaver CT images that are ideal for testing of DL-based image analysis and dosimetry algorithms.[54–56]

Finally, it is worth noting that, with the patient-specific organ dose information, one can derive the so-called "effective dose" — a quantity that the American Association of

Physicists in Medicine (AAPM) believes to bear significant uncertainty and therefore should be used only for prospective radiologic protection purposes and to help patients understand medical radiation dose in perspective.[57]

5. CONCLUSIONS

In this study, an automatic multi-organ segmentation method has been developed using a CNN model that was trained with two publicly available CT databases involving a total of 103 patients. The method takes <5 s to perform automatic multi-organ segmentation for one patient and, for purposes of CT organ dosimetry, has achieved good segmentation accuracy for the testing cases considered in this study. The organ dose calculation method takes <4 s for a total of 1×10^8 photons using the GPU-based rapid Monte Carlo code, ARCHER, to achieve the organ dose statistical uncertainty of better than 0.5%. These results demonstrate, for the first time, the excellent accuracy and efficiency of a streamlined patient-specific organ dosimetry computational tool. Implementation of such methods as part of the clinical workflow can yield considerable improvement over the current CT organ dose methods that are based on population-average phantoms, thus opening the door to prospective patient-specific optimization features in the future.

ACKNOWLEDGMENTS

Zhao Peng was a visiting Ph.D. student at RPI whose work was supported by a grant from the University of Science and Technology of China and by an International Training Grant from the American Association of Physicists in Medicine (AAPM). Co-authors were supported in part by various grants: NIH/NIBIB (R42EB019265-01A1, U01EB017140, R01EB026646), NIH/NCI (R01CA233888 and R01CA237267), and National Natural Science Foundation of China (11575180).

CONFLICT OF INTEREST

X. George Xu is a co-founder of Virtual Phantoms, Inc (Albany, New York) that commercializes software technologies — VirtualDose™ for medical CT dose reporting and ARCHER™ for real-time Monte Carlo dose computing.

[a]Author to whom correspondence should be addressed. Electronic mail: xug2@rpi.edu; Telephone: 518-276-4014.

REFERENCES

1. Amis ES Jr., Butler PF, Applegate KE, et al. American College of Radiology white paper on radiation dose in medicine. *J Am Coll Radiol*. 2007;4:272–284.
2. de Gonzalez AB, Darby S. Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. *Lancet*. 2004;363:345–351.
3. Brenner DJ, Hall EJ. Computed tomography—an increasing source of radiation exposure. *N Engl J Med*. 2007;357:2277–2284.
4. IMV Medical Information Division, Inc.IMV Benchmark Report for CT; 2018.
5. Shope TB, Gagne RM, Johnson GC. A method for describing the doses delivered by transmission x-ray computed tomography. *Med Phys*. 1981;8:488–495.
6. Newman B, Ganguly A, Kim JE, et al. Comparison of different methods of calculating CT radiation effective dose in children. *Am J Roentgenol*. 2012;199:W232–W239.
7. Deak P, van Straten M, Shrimpton PC, et al. Validation of a Monte Carlo tool for patient-specific dose simulations in multi-slice computed tomography. *Eur Radiol*. 2008;18:759–772.
8. Jarry G, DeMarco JJ, Beifuss U, et al. A Monte Carlo-based method to estimate radiation dose from spiral CT: from phantom testing to patient-specific models. *Phys Med Biol*. 2003;48:2645–2663.
9. Angel E, Yaghmai N, Jude CM, et al. Monte Carlo simulations to assess the effects of tube current modulation on breast dose for multidetector CT. *Phys Med Biol*. 2009;54:497–511.
10. Tian XY, Li X, Segars WP, et al. Pediatric chest and abdominopelvic CT: organ dose estimation based on 42 patient models. *Radiology*. 2014;270:535–547.
11. Lee C, Kim KP, Long DJ, et al. Organ doses for reference pediatric and adolescent patients undergoing computed tomography estimated by Monte Carlo simulation. *Med Phys*. 2012;39:2129–2146.
12. Turner AC, Zhang D, Khatonabadi M, et al. The feasibility of patient size-corrected, scanner-independent organ dose estimates for abdominal CT exams. *Med Phys*. 2011;38:820–829.
13. Li XA, Samei E, Segars WP, et al. Patient-specific radiation dose and cancer risk estimation in CT: part I. Development and validation of a Monte Carlo program. *Med Phys*. 2011;38:397–407.
14. Lee C, Kim KP, Bolch WE, et al. NCICT: a computational solution to estimate organ doses for pediatric and adult patients undergoing CT scans. *J Radiol Protect Off J Soc Radiol Protect*. 2015;35:891–909.
15. Ding A, Gao Y, Liu H, et al. VirtualDose: a software for reporting organ doses from CT for adult and pediatric patients. *Phys Med Biol*. 2015;60:5601.
16. Xu XG. An exponential growth of computational phantom research in radiation protection, imaging, and radiotherapy: a review of the fifty-year history. *Phys Med Biol*. 2014;59:R233.
17. Emami B, Lyman J, Brown A, et al. Tolerance of normal tissue to therapeutic irradiation. *Int J Radiat Oncol*. 1991;21:109–122.
18. Bentzen SM, Constine LS, Deasy JO, et al. Quantitative analyses of normal tissue effects in the clinic (Quantec): an introduction to the scientific issues. *Int J Radiat Oncol*. 2010;76:S3–S9.
19. Emami B. Tolerance of normal tissue to therapeutic radiation. *Rep Radiother Oncol*. 2013;1:123–127.
20. Liu T, Xu XG, Carothers CD. Comparison of two accelerators for Monte Carlo radiation transport calculations, Nvidia Tesla M2090 GPU and Intel Xeon Phi 5110p coprocessor: a case study for X-ray CT imaging dose calculation. *Ann Nucl Energy*. 2015;82:230–239.
21. Su L, Yang Y, Bednarz B, et al. ARCHERRT—a photon-electron coupled Monte Carlo dose computing engine for GPU: software development and application to helical tomotherapy. *Med Phys*. 2014;41:071709.
22. Sodickson A, Baeyens PF, Andriole KP, et al. Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults. *Radiology*. 2009;251:175–184.
23. Xu XG. Personnal Communications. 2019.
24. Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. *Med Phys*. 2014;41:050902.
25. Han M, Ma J, Li Y, et al. Segmentation of organs at risk in CT volumes of head, thorax, abdomen, and pelvis. Medical Imaging 2015: Image Processing: International Society for Optics and Photonics. 2015;94133J.
26. Chen X, Zhang R, Yan P. Feature Fusion Encoder Decoder Network For Automatic Liver Lesion Segmentation. arXiv preprint arXiv:190311834; 2019.
27. Christ PF, Elshaer MEA, Ettlinger F, et al. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin: Springer; 2016:415–423.
28. Roth HR, Lu L, Farag A, et al. Deeporgan: multi-level deep convolutional networks for automated pancreas segmentation. In: *International*

Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer; 2015:556–564.
29. Dou Q, Chen H, Jin Y, et al. 3D deeply supervised network for automatic liver segmentation from CT volumes. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin: Springer; 2016:149–157.
30. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015:3431–3440.
31. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image Computing and Computer-Assisted Intervention*. Berlin: Springer; 2015:234–241.
32. Trullo R, Petitjean C, Ruan S, et al. Segmentation of organs at risk in thoracic CT images using a sharpmask architecture and conditional random fields. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017): IEEE; 2017:1003–1006.
33. Gibson E, Giganti F, Hu Y, et al. Automatic multi-organ segmentation on abdominal CT with dense v-networks. *IEEE Trans Med Imaging*. 2018;37:1822–1834.
34. Schmidt TG, Wang AS, Coradi T, et al. Accuracy of patient-specific organ dose estimates obtained using an automated image segmentation algorithm. *J Med Imaging*. 2016;3:043502.
35. Fu W, Segars W, Abadi E, et al. From patient-informed to patient-specific organ dose estimation in clinical computed tomography. Conference on Medical Imaging - Physics of Medical Imaging. Houston, TX; 2018.
36. Peng Z, Fang X, Yan P, et al. A method of rapid quantification of patient-specific organ dose for CT using coupled deep multi-organ segmentation algorithms and GPU-accelerated Monte Carlo dose computing code. arXiv:190800360v2; 2019.
37. Yang J, Sharp G, Veeraraghavan H, et al. Data from lung CT segmentation challenge. The Cancer Imaging Archive; 2017.
38. Yang J, Veeraraghavan H, Armato SG III, et al. Autosegmentation for thoracic radiation treatment planning: a grand challenge at AAPM 2017. *Med Phys*. 2018;45:4568–4581.
39. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045–1057.
40. Roth HR, Farag A, Turkbey E, et al. Data from Pancreas-CT. The cancer imaging archive;2016.
41. https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.zoom.html
42. Thévenaz P, Blu T, Unser M. 25 - Image interpolation and resampling. In: Bankman IN, ed. *Handbook of Medical Imaging*. San Diego: Academic Press; 2000:393–420.
43. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2016.
44. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal*. 2017;39:2481–2495.
45. Bishop CM. *Pattern Recognition and Machine Learning*. Berlin: Springer; 2006.
46. Arlot S, Celisse A. A survey of cross validation procedures for model selection. *Statistics Surveys*. 2009;4:40–79.
47. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980; 2014.
48. Abadi M, Barham P, Chen J, et al. Tensorflow: a system for large-scale machine learning. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI; 2016;16:265–283.
49. Chen LC, Yang Y, Wang J, et al. Attention to scale: scale-aware semantic image segmentation. *IEEE Conf Comput Vis Pattern Recogn*. 2016;2016:3640–3649.
50. Gu J, Bednarz B, Caracappa P, et al. The development, validation and application of a multi-detector CT (MDCT) scanner model for assessing organ doses to the pregnant patient and the fetus using Monte Carlo simulations. *Phys Med Biol*. 2009;54:2699.
51. Ding A. Development of a radiation dose reporting software for X-ray computed tomography (CT). Rensselaer Polytechnic Institute; 2012.
52. Ding A, Mille MM, Liu T, et al. Extension of RPI-adult male and female computational phantoms to obese patients and a Monte Carlo study of the effect on CT imaging dose. *Phys Med Biol*. 2012;57:2441.
53. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26:297–302.
54. Zhang D, Padole A, Li XH, et al. In vitro dose measurements in a human cadaver with abdomen/pelvis CT scans. *Med Phys*. 2014;41:091911.
55. Li XH, Shi JQ, Zhang D, et al. A new technique to characterize CT scanner bow-tie filter attenuation and applications in human cadaver dosimetry simulations. *Med Phys*. 2015;42:6274–6282.
56. Padole A, Khawaja RDA, Otrakji A, et al. Comparison of measured and estimated CT organ doses for modulated and fixed tube current: a human cadaver study. *Acad Radiol*. 2016;23:634–642.
57. Mccollough C, Cody D, Edyvean S. The measurement, reporting, and management of radiation dose in CT. *Clin Transl Sci*. 2008;8:330–333.