*i*Phantom: A framework for automated creation of individualized computational phantoms and its application to CT organ dosimetry

Wanyi Fu,

Department of Electrical and Computer Engineering, and Carl E. Ravin Advanced Imaging Laboratories, Duke University, Durham, NC 27705 USA

Shobhit Sharma,

Department of Physics and Carl E. Ravin Advanced Imaging Laboratories, Duke University, Durham, NC 27705 USA

Ehsan Abadi,

Department of Electrical and Computer Engineering, and Carl E. Ravin Advanced Imaging Laboratories, Duke University, Durham, NC 27705 USA

Alexandros-Stavros Iliopoulos,

Department of Computer Science, Duke University, Durham, NC 27708, USA

Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Qi Wang,

Department of Radiology, the Fourth Clinical Hospital of Hebei Medical University, Heibei, 050011, China

Joseph Y. Lo,

Departments of Electrical and Computer Engineering, Biomedical Engineering, Medical Physics Graduate Program, and Carl E. Ravin Advanced Imaging Laboratories, Duke University, Durham, NC 27705 USA

Xiaobai Sun,

Department of Computer Science, Duke University, Durham, NC 27708, USA

William P. Segars,

Departments of Biomedical Engineering, Medical Physics Graduate Program and Radiology and the Carl E. Ravin Advanced Imaging Laboratories, Duke University, Durham, NC 27705 USA

Ehsan Samei

Carl E. Ravin Advanced Imaging Laboratories, the Medical Physics Graduate Program, the Departments of Radiology, Electrical and Computer Engineering, Biomedical Engineering, and Physics, Duke University, Durham, NC 27705 USA

Abstract

wanyi.fu@duke.edu .

Objective: This study aims to develop and validate a novel framework, *i*Phantom, for automated creation of patient-specific phantoms or "digital-twins (DT)" using patient medical images. The framework is applied to assess radiation dose to radiosensitive organs in CT imaging of individual patients.

Method: Given a volume of patient CT images, *i*Phantom segments selected anchor organs and structures (e.g., liver, bones, pancreas) using a learning-based model developed for multi-organ CT segmentation. Organs which are challenging to segment (e.g., intestines) are incorporated from a matched phantom template, using a diffeomorphic registration model developed for multi-organ phantom-voxels. The resulting digital-twin phantoms are used to assess organ doses during routine CT exams.

Result: *i*Phantom was validated on both with a set of XCAT digital phantoms (n=50) and an independent clinical dataset (n=10) with similar accuracy. *i*Phantom precisely predicted all organ locations yielding Dice Similarity Coefficients (DSC) 0.6 – 1 for anchor organs and DSC of 0.3–0.9 for all other organs. *i*Phantom showed <10% errors in estimated radiation dose for the majority of organs, which was notably superior to the state-of-the-art baseline method (20–35% dose errors).

Conclusion: *i*Phantom enables automated and accurate creation of patient-specific phantoms and, for the first time, provides sufficient and automated patient-specific dose estimates for CT dosimetry.

Significance: The new framework brings the creation and application of CHPs (computational human phantoms) to the level of individual CHPs through automation, achieving wide and precise organ localization, paving the way for clinical monitoring, personalized optimization, and large-scale research.

Keywords

computational phantoms; organ dose; CT; segmentation; deformable registration

---

I. INTRODUCTION

COMPUTATIONAL human phantoms (CHPs) are mathematical representations of the human anatomy in a digital format. CHPs and their applications have co-evolved in the last six decades [1, 2]. CHP development is driven by important and growing applications, which include retrospective, prospective, or real-time radiation dosimetry, individual cancer risk estimation, diagnostic and interventional radiology studies, monitoring for environmental radiation exposure, assessment of medical imaging protocols, design and verification of shielding protection, and virtual clinical trials for regulatory submissions [1–10]. For these applications, it is essential to have CHPs that realistically reflect individual patients (i.e., patient-specific), as well as the population at large (i.e., population-specific) to echo anatomical variations of real clinical cases and scenarios.

Toward that aim, recent development of CHPs has focused on realistically representing given individuals for patient-specific investigations or assembling many anatomically variable phantoms at large scale for population-based studies. The phantoms are

primarily developed by manually segmenting a limited number of patient cases. However, segmentation is a time-consuming process that can take many months to complete per phantom depending on the level of detail required. Furthermore, the work involves inter- and intra-operator variability that incorporates a certain degree of subjectivity in the process. Even addressing these limitations, the current automatic segmentation techniques based on tissue texture patterns and/or manually-labeled training resources are still limited to only a handful of organs with high-contrast [7, 10–12]. They are generally incapable of differentiating small organs or adjacent organs with similar textures or gray scale (e.g., large intestine). Thus, segmentation alone is inadequate to reflect the extensive range of organs essential for representing the human body. This necessitates a combination of segmentation and deformation to create CHPs.

To facilitate the development of CHPs, deformable techniques have been used to augment existing phantoms to model additional anatomically variable models [1, 2]. Populations of new models can be created by deforming given template phantoms to match variations observed in patient data. However, these methods are still highly constrained by the manual segmentation process. Overall, the current methods cannot be applied to create large sets of CHPs or to create digital representations, so-called digital twins, of clinical cases.

In this study, we develop a framework, *i*Phantom, to address the challenging problem of automatically and accurately creating CHPs in a patient-specific manner toward large scale CHP development and clinical patient-specific implementation. The work draws upon a validated non-automatic pipeline that was used to develop the widely used XCAT phantoms [13–15]. Specifically, the work addresses two major obstacles to generate phantoms directly from patient medical images. The first obstacle is to extract patient specific information. We developed a machine learning model to segment key, or anchor, organs automatically from medical images. Second, a complete CHP requires all major organs, including the low contrast ones. We adapted a deformation morphing technique developed for the multi-organ registration problem in phantom space to embed unsegmented structures from an anatomical template. The template was obtained using a parameterized matching method to draw upon the complete human models from the XCAT phantom library. We systematically validate the framework in terms of phantom geometry and its application for organ radiation dosimetry. This specific application is motivated by the strong need to assess radiation dose and its associated risk to the patient during an exam in a patient-specific manner. Organ dose has been regarded as the most relevant metric to characterize patient risk. However, clinical quantification of this metric has been hindered due to previous limitations in creating accurate CHPs.

In this paper we detail a process to create CHPs from patient images automatically. The preliminary concept was first introduced at the 2018 SPIE Medical Imaging conference [16]. The present manuscript significantly extends that concept into a framework with each individual component specifically developed for the purpose of creating patient-specific CHPs. The framework is further applied to the task of patient-specific organ dose estimation by incorporating a recently released GPU-based Monte Carlo simulation package [17] for CT scans with both fixed and modulated tube current. We demonstrate the clinical utility

of the framework by applying it to unseen clinical patient images to create a new set of patient-specific phantoms and quantitatively validate the results.

The rest of the paper is organized as follows. We introduce the *i*Phantom pipeline and analysis in Section II. Then, we describe its application to radiation dosimetry in CT in Section III. The validation experiments and results are reported in Sections IV and V. Finally, we present conclusions and discussions in Section VI.

II. *i*PHANTOM FRAMEWORK

The proposed framework aims at creating patient-specific phantoms directly from patient-medical images in a fully automated pipeline. The framework first automatically segments a selected set of organs and structures. These organs are rendered in high contrast in CT and can be delineated or segmented with high fidelity. The segmentation of the remaining organs, however, may suffer from a great degree of uncertainty. Thus, the framework fuses the initial segmentation results with an anatomical template that will be used to fill in the missing structures in the newly made phantoms. This addresses the challenge that the body consists of many organs (34 organs were investigated in this study), large and small, with different morphology and textures; and reliably segmenting all of them automatically (even manually) is very challenging. As templates for this work, we used 50 highly detailed adult models in the XCAT phantom library, which was developed in our laboratory and has been widely used for many applications [13, 14]

Figure 1 summarizes the steps in the *i*Phantom framework. First, anchor organs within a given set of patient CT data are automatically segmented, using a learning-based segmentation model developed to segment key organs, to define an initial target (Figure 1. box A.). This person-specific definition of anchor organs is used to guide the fill-in of non-anchor organs. An XCAT phantom template that best matches the partially segmented target is then selected using parameters chosen to reflect anatomical similarities (Figure 1. box B.). Finally, a mapping between the template XCAT model and the patient target was calculated using a registration model adapted for the multi-organ scenario of the phantom space. The mapping is used to transport the non-anchor organs from the template space to define them within the new patient model (Figure 1. box C.). In this study, we developed both linear (affine) and non-linear (diffeomorphic) mapping methods.

A. Enhanced and automated segmentation of anchor organs

Accurate anchor-organ segmentation is critical to the automated creation of individualized phantoms, setting the stage for the subsequent steps. The specific set of anchor organs was determined based on state-of-the-art multi-organ segmentation of CT images of the chest-abdominal-pelvis region [11, 12] and our available labeled training images. Twenty-two organs and structures were selected: thyroid, lungs (L/R), heart, liver, spleen, kidneys (L/R), gallbladder, ribs (L/R), bladder, spine, clavicles, sternum, scapulars, stomach, pancreas, pelvis, femurs, arms, and body. The body represents all organs and tissues not individually segmented but included within the body contour.

3D convolutional neural networks were developed using a U-Net architecture similar to that described by Çiçek et al. [18]. In designing the training objective, or the loss function, we considered multiple classes (organs) as well as the difference in organ volume, i.e., the number of voxels in an organ. For example, the ratio in volume between the lung and thyroid can be up to three orders of magnitude. To overcome the inter-class imbalance problem, we made use of the combined dice loss and cross-entropy loss functions similar to Taghanaki et al. [19]. The dice loss has been used for multi-organ segmentation, even though it may lead to failure to converge for small organs due to the vanishing gradient issue. The cross-entropy loss regularizes the objective function [19]. Specifically, we used the following loss function

$$Loss = w_1\left(1 - \frac{1}{L}\sum_l\left(\frac{2\sum_i p_l^i r_l^i}{\sum_i p_l^i + r_l^i}\right)\right) - w_2\sum_l r_l^i \log\left(p_l^i r_l^i\right), \tag{1}$$

where w_1 and w_2 are weighting coefficients for the dice and cross entropy loss, respectively; p_l^i and r_l^i are the segmentation probability and binary indicator, respectively, for voxel i and class l. In this study, we chose $w_1 = 1$ and $w_2 = \frac{1}{L}$, where L is the total number of classes.

For training the network, we utilized and refined the manually segmented CT data upon which the XCAT library of phantoms were based. In practice, the GPUs we used was a Titan RTX GPU with 24 GB memory, which had limited memory resources. To account for this, we cropped the training images to 128×128×128. Within this input size constraint, to balance sufficient global content for training and resolution for creating phantoms, the images were down-sampled to a longitudinal resolution of 5 mm and an in-plane resolution of 2.5 mm. The inputs were at the size that contained the majority volume of the patient trunk. In the training, CT images were randomly sampled with each structure centered, following Pawlowski et al., [12] to ensure the network was trained on all structures. For inference, whole CT images, rather than cropped segments, were used as inputs to speed up prediction and eliminate prediction window boundary artifacts. Once developed, the segmentation method was validated as described in Section IV.A.

B. Parameterized template matching

After segmentation, a parameterized matching strategy was developed to identify a template phantom that best matches the patient determined by the segmented anchor anatomy. It was assumed that if the anchor layout is similar between a patient and a phantom, the rest of the organs will show a higher likelihood of similarity. This similarity also results in the transformation computation being less expensive. We find phantom i by minimizing the distance to the patient defined by parameters $\Theta_i = \{\theta_{1,i}, \theta_{2,i}, \ldots \theta_{k,i}\}$ derived from the anchor organs as

$$i = arg\min_i \|\Theta_i - \Theta_0\|_2, \tag{2}$$

where Θ_i and Θ_0 are anatomical parameters for phantom i and the target patient, respectively, and G is the set of phantoms satisfying patient-specific constraints (e.g., age range, gender).

In previous work, the trunk height has been shown to be a good indicator of organ distribution [20]. Likewise, the trunk diameter has been shown to be related to the thickness of tissue outside the skeleton [21]. Therefore, in this initial implementation, we defined $\theta_{1,i}$ as the phantom trunk height, and $\theta_{2,i}$ as the phantom trunk effective diameter defined as $\theta_2 = 2\sqrt{\frac{V}{h\pi}}$, where V and h are the segmented trunk contour volume and height, respectively. The matched phantom was then chosen iteratively from G using (2).

C. Registration and embedding of non-anchor organs

With a matching template phantom, the final step was to calculate a mapping from the template space to the target patient space. A mapping may be described as the product composition of an affine mapping for initial global alignment and a non-rigid, non-linear diffeomorphic mapping [22]. The affine mapping has only shifting, scaling, and sheering parameters and can be determined more quickly. The affine-transformed template is then used as input for the diffeomorphic registration to the target. The diffeomorphic transformation enables large deformation calculations while preserving the topology. We developed and evaluated mapping calculations with both (i) affine mapping alone and (ii) combined affine and diffeomorphic mapping (labeled as diffeomorphic mapping).

To compute the diffeomorphic mappings, we adapted the symmetric normalization (SyN) method in the Advanced Normalization Tools (ANTs). ANTs was chosen based on its public availability and its reliable accuracy and wide use in medical research [23, 24]. We modified the transformation calculation parameters from its common utility in brain image registration to calculate the mapping from the template to the target. The mapping was applied to the non-anchor organs located in the template to 'fill-in' these organs in the target.

We choose parameters including similarity metrics, a gradient step, and Gaussian smoothing for velocity and deformation field. The parameters were determined empirically by conducting a large set of experiments over the parameter space to identify those with sufficient registration accuracy. Similarity metrics between the phantoms were calculated in both intensity-based and label-based spaces. Intensity-based metrics are advantageous for regions with more drastic changes (e.g. soft tissue and bone boundaries) while label-based metrics provide additional anatomical similarity knowledge. For intensity-based phantom input, we generated synthesized CT images from the template and target voxel phantoms by assigning organ intensities (HU values) derived from the averaged organ voxels of the XCAT patient CT images. For label-based phantom input, the template and target voxel phantoms were formatted with corresponding unique IDs assigned to the segmented structures. For organs with two separate regions on both left and right sides of the body (e.g., lungs), unique labels were assigned for each side.

For the affine mapping, we used the intensity-based mutual information metric. For the diffeomorphic mapping, we used the combined intensity-based cross-correlation and label-based point-set expectation metrics. The two metrics were weighted (weighting of 0.1 and 0.9 for cross-correlation and point-set expectation, respectively) and summed. For cross-correlation calculation, we used a $3\times3\times3$ voxel neighborhood. The point set expectation was calculated as the expectation of a point x_i on one image being the target point of a set of points $\{y\}$ on the other image, with the expectation calculated by weighted sum [25]. The weighting for point x_i and the points on the other image $y_i \in \{y\}$ was determined by a normal function of the Euclidean distance between x_i and y_i within a neighborhood of 20 voxels.

The runtime for affine transformation was within seconds and diffeomorphic deformation calculation was 1–4 hours on a 2.20 GHz Intel Xeon CPU. Considering segmentation taking only seconds, diffeomorphic deformation is the most time-consuming component of the process.

III. APPLICATION TO PATIENT-SPECIFIC ORGAN DOSIMETRY

The phantoms created using the framework are portable to applications in many domains. In this study, we applied and validated the framework in the classic and actively researched area of *patient-specific* organ dosimetry in computed tomography. Computed tomography has been widely used for diagnosis of major diseases; however, its potentially harmful radiation effect has been a concern for healthcare providers and patients. Patient-specific organ dose has been regarded as the most relevant metric to quantify radiation exposure and the associated risk[26]. However, it has not been widely utilized clinically due to the challenge of automatically creating patient-specific phantoms necessary to estimate reliable organ doses. We attempted to address this limitation.

For this validation, we utilized a CT organ dose estimation module [17] developed in our laboratory. As input, the patient-specific phantoms were converted into voxelized dosimetry phantoms (i.e., reformatted for MC-GPU dose simulation) by assigning materials to each organ and structure. The CT technical parameters are specified in Section IV D, including vendor and CT geometry, bowtie, spectrum, and tube current profile. The absorbed dose was obtained using a validated real-time Monte Carlo (MC) tool, developed using the MC-GPU framework [27], to simulate photon transport across the voxelized dosimetry phantoms.

In general, organ dose estimation requires different levels of accuracy in terms of the phantom geometry for fixed and modulated tube current. It has been demonstrated that the organ dose estimation under fixed tube current is reasonably accurate as long as the patient is matched to a phantom with similar size, as the radiation field under fixed tube current is relatively uniform [21]. Organ dose estimation under the more prevalently used Tube Current Modulation (TCM) requires the organs to be localized more precisely, as the radiation field is more heterogeneous. We evaluated organ dose estimation accuracies under both fixed and modulated tube current.

IV.   VALIDATION

This retrospective study was performed in compliance with the Health Insurance Portability and Accountability Act and was determined to be exempt from Institutional Review Board requirements.

We systematically validated the *i*Phantom framework using the XCAT phantoms and the CT data upon which they were based (XCAT datasets) in a cross-validation approach. We further applied the framework to clinical CT images to evaluate the framework's generalizability in creating new patient-specific phantoms.

For the XCAT datasets, we validated the segmentation stage and data fusion stage (matching and registration) individually. The XCAT datasets were not directly used to validate the full pipeline because the XCAT geometries have been generalized and altered to adapt to many applications (e.g., respiratory and cardiac motion simulations) and thus are no longer fully aligned to their original CT images. We took advantage of the XCAT datasets consisting of an extensive range of organs within a population to validate individual components of the *i*Phantom framework. We used the manually segmented CT data upon which the XCAT library of phantoms were based to validate the anchor organ segmentation accuracy. We further adopted the voxelized XCAT phantoms to assess the registration calculation and non-anchor organ embedding of the framework. A schematic of the validation strategy for the segmentation validation, registration validation, and a baseline dose error estimation is shown in Figure 2.

A.   Segmentation validation

We used fifty chest-abdomen-pelvis CT datasets that were part of the XCAT modeling. We further refined and checked their original manual segmentation under supervision of a radiologist (23 years of experience). The image data show clinical anatomical variations with no abnormalities (20/30 F/M; age range: 18 – 78 y.o.; trunk effective diameter range: 24 – 39 cm). The twenty-two organs and structures listed in Section II.A were segmented in the data.

A five-fold cross-validation was performed to train and validate the segmentation models. Each testing CT dataset was automatically segmented using the trained model of each fold and compared to its corresponding expert manual segmentation in terms of geometrical accuracy and estimated dose (metrics described in Section IV.D.). In each fold, the training, validation, and test set were divided as 30, 10, and 10 of the cases, respectively. The model was implemented by Keras [28] with a Tensorflow backend [29] with Adam optimizer [30]. For training, we used 24,000 iterations with a learning rate of 10e-3 and 10e-4 for the first and second half of iterations, respectively. Training took about 20 hours, and the prediction of one patient took 1–20 seconds using a Titan RTX GPU with 24 GB memory.

B.   Registration validation

As mentioned above, registration was validated independent of segmentation. The registration procedure requires a template and a target image. To generate these images, we voxelized the 50 XCAT phantoms from the existing library. Anchor organs were set to

unique integer IDs while non-anchor organs were set to the background body ID making them indistinguishable. These values were set according to each individual experiment.

To validate the registration accuracy, we performed two experiments: (i) leave-one-phantom-out, and (ii) leave-one-organ-out. The leave-one-phantom-out experiment was used to assess the ability of the framework in predicting the organs for unknown targets. In this experiment, each of 50 XCAT phantoms was used as a target while the remaining 49 were used as the template library. Each target XCAT was matched to a template from the remaining 49 using the methods described in Section II.B. The anchor organ types were those specified in Section II A, while the non-anchor organs included all other radiosensitive structures: thymus, larynx pharynx, trachea bronchi, esophagus, breasts, large intestine, adrenals, small intestine, ovaries, testes, uterus, and vagina.

The leave-one-organ-out experiment was used to evaluate the accuracy of predictions with different anchor organs left out. Within each of the 50 leave-one-phantom-out experiments, we performed multiple leave-one-organ-out experiments. In each leave-one-organ-out experiment, one segmented organ was left out (set to the body ID) and the rest of the organs were used as the anchor organs. The left-out organ was filled in using the *i*Phantom registration module and assessed for accuracy using metrics described in Section IV.D.

Furthermore, to evaluate the isolated effect of diffeomorphic deformation, we assessed the two registration methods, one using only an affine transformation and one with the combined affine and diffeomorphic transformation (labeled as diffeomorphic deformation in the results).

In each experiment, the target phantom was used to provide the target image. As mentioned above, anchor organs were set to unique integer IDs in the voxelized target phantom; non-anchor organ voxels were set to the background body ID. The matched XCAT selected for each target was used to provide the template image. The organs were set to corresponding IDs as the target. Given these images, the framework was used to calculate the transform from the template to the target. The transformation was applied to the template with full anatomy to predict the remaining anatomy of the target. The template phantom with transformation were compared that from the original XCAT target phantom, using the metrics as outlined in Section IV. D. The procedure for each target phantom is detailed in Fig 2b.

In the interest of computational efficiency in terms of the registration as well as the dosimetry calculations, in all registration experiments, the XCAT phantoms were voxelized at an isotropic resolution of 3.45 mm.

C. Application to new CT data

The *i*Phantom framework was tested using new CT datasets, not previously used in the XCAT phantom creation. This HIPPA compliant study included ten randomly selected patients who underwent Chest-abdomen-pelvis CT scans from our institution from January 2017 to May 2017. Ten chest-abdomen-pelvic CT images were included in this experiment (4F/6M; age range: 35–83 y.o.; trunk effective diameter range: 24–44 cm). Two patients

were scanned with the SOMATOM Definition Flash (Siemens Healthcare) and eight patients were scanned with either Revolution CT, Revolution GSI, or Discovery CT750 HD (GE Medical Systems). Nine patients were scanned with iodine contrast and 1 without. The scans used 140 kVp (n=1), 120 kVp (n=8), and 100 kVp (n=1). The exam CTDI$_{vol}$ ranged from 4.67 mGy to 36.91 mGy. For each patient, the anchor organs used in the 1st experiment were manually delineated by a radiologist with 23 years of experience. This annotation was used as the evaluation reference. The *i*Phantom framework was applied to these clinical images to generate patient-specific phantoms and to evaluate the framework accuracy.

Furthermore, using the CT images and the manual annotations as reference, we evaluated the framework accuracy for segmentation and registration components using the metrics described in Section IV. D. In detail, the trained segmentation model by the XCAT dataset (Section IV. A.) was applied to these new CT images to predict organ labels. The predicted and reference segmentation were compared by both geometry and organ dose using the same scheme as in Fig. 2a. The registration model was evaluated using the leave-one-organ-out approach. We used both the manual (reference) and the predicted segmentation as the initial target to evaluate the isolated registration error and the overall segmentation and registration error, respectively. The initial target phantom was then matched to an XCAT and had the left-out organ filled in using the proposed registration module. The template phantom libraries consisted of all the 50 XCAT phantoms with all anchor organs as used in section IV.B. 1). The filled-in organs were compared with its manual delineation using the scheme shown in Fig. 2b. For organ dose estimation, we extracted the actual tube current modulation profile from the CT image DICOM headers to mimic the scans more realistically. For scan simulation, the TCM values were assigned to projection angles according to longitudinal locations with linear interpolation

D. Validation metrics

For each study, comparisons of the predicted anatomy versus the known truth anatomies were made in terms of the geometrical accuracy of the organs and structures as well as estimated CT radiation dose. Geometric accuracy of the predicted anatomy as compared to the known truth was measured by the dice similarity coefficients (DSC).

We measured the dose differences between the phantoms generated from each test set and its reference using the organ dose module. Without loss of generalizability, the organ doses were estimated using a typical clinical protocol. This simulation included a Light Speed VCT scanner (GE Healthcare) with explicitly modeled gantry geometry, bowtie, and spectrum. A chest-abdominal-pelvis protocol was performed with the scan converging 1cm above the lung and 1cm below the pelvis. The CT techniques were 120 kV, pitch of 1.375, and collimation of 40 mm and medium body bowtie filter for simulated scans with both modulated and fixed tube current. The tube current modulation profile was synthesized as a function of patient attenuation at each projection angle, scanner-specific geometry, bowtie type, and kV using the method described by Li et al [31]. In this study, we used the 'strong' TCM configuration (i.e., $a = 1$ in Li et al. [31]) to simulate a scenario with the most heterogeneous radiation field providing an upper bound in organ dose discrepancy.

The organ dose differences between the dose from the ith reference phantom ($D_{0,i}$) and that from the phantom generated from the test set (D_i) were compared by absolute relative error (ARE) as

$$\varepsilon_i = \frac{|D_i - D_{0,i}|}{D_{0,i}} 100\%,$$

(3)

with the mean of ε_i calculated across all the tests and referred to as mean absolute error (MAE).

E. Baseline method

We assessed organ dose errors using an alternative baseline method to provide a dose accuracy reference. In many state-of-the-art approaches in automatically estimating organ doses, the patient anatomy is represented by a matched computational phantom without further post-processing (e.g., registration, rescaling) [20, 32]. For this baseline comparison, we matched each phantom (reference) to another phantom (predicted anatomy) in a leave-one-phantom-out validation approach across the 50 XCAT phantoms. The matched phantom was selected using the proposed parametrized matching method (Section II.B). Instead of using affine or diffeomorphic registration to align the phantoms to the initial target, the matched phantom was aligned to the patient using anatomical landmarks (top of the lung and bottom of pelvis). We reported organ dose differences between each reference phantom and its matched phantom. This method reasonably represented the state-of-the-art in that (i) the XCAT phantom population is relatively large within the literature, and (ii) the matching includes two decisive parameters (i.e., height and width) that influence organ dose.

V. RESULTS

In this section, we present the validation results. We first show geometry and organ dose estimation accuracy of the segmentation and data fusion stages, respectively, using the XCAT dataset. We further provide overall dose error surrogates combining segmentation and registration dose errors and compare them with the baseline dose errors. Finally, we show the clinical data validation results.

A. Geometry validation

Segmentation.—Fig. 3 shows the results of segmentation validation using the XCAT dataset. The results show a sufficient segmentation performance, especially in clear discrimination of organ material types as soft-tissue, bone, and lung, which are essential for accurate dose calculation. Lungs and body show an average DSC greater than 0.98. Large soft tissue organs (liver, spleen, kidneys) show an average DSC greater than 0.9. Soft tissue organs with relatively irregular shape (pancreas, bladder) or smaller size (thyroid, gallbladder) show an average DSC of 0.6–0.8. Bones show an average DSC > 0.85.

Registration leave-one-phantom-out.—Fig. 4 shows the results of registration validation using the leave-one-phantom-out approach. For most anchor organs, affine transformation and diffeomorphic deformation resulted in a DSC of 0.2–0.6 and 0.8–0.9,

respectively. For filled in organs, affine shows a reasonable DSC of 0.2–0.8 and the diffeomorphic deformation improves the results to 0.3–0.9. Indicated by the positive DSC, both affine and diffeomorphic transformations precisely localize organs. We demonstrate this level of accuracy provides sufficient accuracy in organ dose estimation (section V.B.). The improved performance of diffeomorphic deformation is due to the fact that the affine transformation is linear with a limited degree of freedom. The diffeomorphic transformation is non-linear and more flexible, resulting in better anchor organ alignment.

This superior anchor organ framework results in a more accurate prediction of unsegmented organs. This effect can be further observed in that for non-anchor organs bordered by anchor organs in multiple directions, such as the tracheabronchi, thymus, esophagus, and adrenals, the gain from the diffeomorphic deformation is higher. On the contrary, for non-anchor organs with limited constraints, such as the breast and larynx-pharynx, the DSCs were moderately improved from deformation.

Registration leave-one-organ-out.—Fig. 5 shows the results of registration validation using the leave-one-organ-out approach for an example case. Results show a similar trend that both affine (DSC of 0.2–0.8 for both anchor and non-anchor organs) and diffeomorphic (DSC of 0.8–0.9 for anchor organs and DSC of 0.4–0.9 for non-anchor organs) transformation are sufficiently able to fill in organs, with superior performance from the diffeomorphic deformation. The affine transformation results are not substantially affected by whether the organ or its neighboring organs are anchors or left-out. This demonstrates that this non-linear transformation is mainly optimized for whole-body features rather than those from local context. The diffeomorphic deformation shows high transformation accuracy for most anchor organs. This considerable improvement is minimally affected regardless of which organ is left out, demonstrating the flexibility of this nonlinear method.

B. Dosimetry validation

For the segmentation validations, in general, anchor organs with DSC > 0.85 (lung, heart, liver, spleen, kidneys and bones) show a MAE of <1% (those ranges of errors were summarized with precision of 1%) for both fixed and modulated tube current scans. Organs challenging to segment (stomach, pancreas, bladder, gallbladder) show an average error of 1–4% and 2–5% for fixed and modulated scans, respectively. The MAE for thyroid is 12% for both fixed and modulated tube current. The dose errors from segmentation are generally small. The filled-in organs introduce dose errors not only to the filled-in organs, but also to the anchor organs, resulting in anchor organs with a MAE of 1–3% by affine transformation, and <1% by diffeomorphic deformation for both fixed and modulated tube current. Both affine and diffeomorphic transformation show relatively small errors, with errors from diffeomorphic transformation being comparable to those from segmentation.

Fig. 6 shows the mean absolute relative error between reference phantoms and their corresponding predicted anatomy from the *i*Phantom framework averaged across the 50 XCAT patients. The overall errors to anchor organs combining segmentation and registration are 1–6% for affine transformation, and <5% for diffeomorphic deformation for both

fixed and modulated tube current (Fig. 6a). Except for the thyroid, the MAE from the alternative baseline method (Section IV. E) are 7–14% for fixed tube current, and 23–33% for modulated tube current. The results demonstrate the sufficient and superior performance of the proposed approach for anchor organs. Fig. 6b show dose errors for non-anchor organs. For fixed tube current, except for small organs or organs not fully constrained by anchors (breasts, larynx-pharynx, and testes), the affine method shows comparable results to the alternative baseline method (MAE: affine 2–16%, match 3–19%), with the MAE reduced to 2–8% with diffeomorphic deformation. Breasts, larynx-pharynx, and testes show relatively large errors with substantially superior results from diffeomorphic deformation (MAE: match 15–61%, affine 18–69%, deform 14–35%) .For modulated tube current, compared to their corresponding MAE under fixed tube current, phantoms created by transformation show a slightly higher MAE while those by the alternative baseline method showed a substantially larger MAE (match 22–33.5%, affine 2–13%, deform 1–9%, except for the breasts, larynx-pharynx, and testes) for most organs. Results show that, except for limited types of organs, phantoms developed with the proposed transformations offer a high dose accuracy for the most embedded organs, substantially superior compared to those with the match-alone (alternative baseline) method, especially under modulated tube current. This can be explained by that for fixed tube current, by using phantoms with similar size, the organ dose errors are generally guaranteed within 20%, as widely demonstrated by the literature [20, 32]. However, the radiation field under modulated tube current is more heterogeneous, so it requires more precise organ location to obtain accurate dose values. Both affine and diffeomorphic transformation align the matched phantoms to the patients resulting in more accurate anatomical representation.

In the leave-one-organ-out validation, the MAE for filled-in organs, except for the thyroid, are 0–11% for affine, and 0–8% for diffeomorphic transformation for both modulated and fixed tube current (Fig. 7). The MAE is similar between scans with fixed and modulated tube current (<2%). Large filled-in organs and structures (lungs and bones) cause small dose errors on certain neighboring anchor organs (MAE: affine 0.5–4%, deform 0–2%). For example, when the lungs are left out and filled-in by diffeomorphic deformation, the lungs show a MAE of 1.4%, and the anchor organs show a MAE of 1.76% for the heart, 1.02% for the stomach, and 1.33% for the thyroid under tube current modulated scans. Other filled-in organs cause less substantial dose errors to the anchor organs with MAE <0.2% for both affine and deformation, for both fixed and modulated tube current. It can be observed that the thyroid and pancreas show similar DSC; however, they show substantial differences in dose errors. This can be explained by two factors. First, the thyroid is smaller and more sensitive to changes in the radiation field. Second, the thyroid is located in an anatomical region with wide changes in body diameter (neck and shoulder), so the radiation field changes more dramatically compared to the pancreas, which is located in the central area of the abdomen. In general, the filled-in organs show sufficient dose values and a slight effect on the dose to anchor organs.

C. Application to new clinical CT data

Fig. 8 shows renderings of two new example phantoms and phantom anatomy overlaid on the CT images created by applying the proposed framework to the clinical CT datasets. The

plots show that the segmentation and registration perform reasonably to build completed models. The quantitative results show similar trends as those from validation using the XCAT dataset.

Fig. 9 shows the DSCs (Fig. 9a) and MAE (Fig. 9b) between the phantoms predicted from the proposed framework based on CT images and the reference, respectively. Compared to the XCAT dataset results, for organs easy to segment, the DSCs are 0.01 – 0.04 less for lungs, heart, liver, spleen, kidneys, body, and 0.02 – 0.1 less for bones. For organs with larger segmentation uncertainties (stomach, pancreas, bladder, gallbladder, and thyroid), the DSCs are −0.1 – 0.12 different compared to the XCAT validation. These differences are reasonable considering the annotation labels of the XCAT are created from different observers, and that the XCAT phantoms are largely based on normal patients, but the clinical patients were mostly abnormal. The differences in imaging techniques may also result in segmentation discrepancies.

The organ dose errors are comparable to the results from the XCAT dataset with simulations using tube current modulation, with a MEA of 0.5–1.5% for organs easier to segment, and a M EA of 1.5–5.5% for organs challenging to segment.

The DSC between phantoms from the test set and the reference in the leave-one-organ-out validation for the clinical dataset compared to the XCAT is 0 – 0.2 less on the average DSC (except for gallbladder, arms, and scapulars) for both affine and diffeomorphic transformation. The inferior results in the clinical dataset compared to the XCAT dataset may result from some large patients in the clinical dataset (effective diameter range: 24–44 cm for clinical patients versus 24–39 cm for the 50 XCAT patients). For both affine and diffeomorphic transformation, the anchor organ transformation calculation is unaffected by the type of organs left out, similar to the XCAT validation. For both affine and diffeomorphic transformation, the fill-in accuracy is not sensitive to whether the anchor organs are generated from the reference or predicted from the *i*Phantom framework. This demonstrates that the segmentation errors will not generally affect the fill-in accuracy compared to the ground truth.

For dosimetry, the MAE for filled-in organs are similar to those of the XCAT, except for the thyroid, with 0–11% for affine transformation, and 0–8% for diffeomorphic deformation with anchor organs from both the reference and the *i*Phantom prediction. The MAE for anchor organs, when obtained using the *i*Phantom predicted segmentation, is similar to the XCAT anchor organ overall error: 1.5–8% (except for gallbladder) for affine translation, and 1.5–5.0% (except for bladder) for diffeomorphic deformation.

VI. DISCUSSION

In this study, we developed and validated a framework, *i*Phantom, to automatically generate computational human phantoms for individual patients, thus creating a "digital twin" for a patient based on their image data. We demonstrate that the framework provides a good tool for patient-specific organ dosimetry in CT. This specific application targets individualized or population-based image monitoring and protocol optimization. Beyond dosimetry, the

framework also has the potential to efficiently generate large populations of CHPs based on patient data.

The framework generates the phantoms with patient-specific information through organ segmentation. For the organs that are challenging to segment, as the patient-specific knowledge is no longer available, the non-anchor organ anatomies are determined from the template taking advantage of the seminaries in anatomies.

Each module of the *i*Phantom framework can be further optimized or customized. For application to other types of imaging data (e.g., non-contrast CT, MRI, etc.), the segmentation component would just need to be trained on the data. The registration component would operate as is. For the matching component, the parameters could be enhanced with more information from the anchor organ segmentation to be more reflective of the inner organ anatomy [33, 34]. For the registration component, registration algorithms such as the multichannel large deformation diffeomorphic metric mapping (MC-LDDMM) method, used to create the original XCAT phantom library [13–15], or the statistical shape model (SSM) used by Wang et al [35], can also be implemented based on the users' preference. The overall time to generates a phantom takes 1–4 hours dominated by the diffeomorphic calculation. To improve the speed of the framework, machine learning methods can be investigated as alternatives to diffeomorphic registration. Future work will be performed to evaluate different registration methods. Moreover, the proposed framework could be applied to for partial body scans as well. For these scans, the framework defines the organs within the field of view based on template matching and deformation. Structures outside the field of view would be modeled based on template matching. Furthermore, the framework could be potentially extended and validated for creating pediatric and pregnant phantoms. The segmentation component of the framework would need to be trained on the different sets of data. For registration, the XCAT already includes many templates for pediatrics [14]; pregnant patient models are also under development.

For patient-specific organ dosimetry, we combined the framework with a Monte Carlo simulation tool in a fully automated approach. State of the art studies providing similar ranges of organs usually approximate both anatomies (e.g., using a matching based method) and radiation field. One study combined match-based anatomy modeling and a convolution-based radiation field modeling method, and it reported a dose error of about 20–35% [32].

The accuracy of the proposed framework is substantially superior with a MAE less than 10% for most organs even under the more challenging modulated tube current situation. It is generally regarded that organ doses are sufficient with a 10% error limited by the Monte Caro simulation accuracy. Our results further suggest that for dosimetry applications, it is beneficial to apply a tiered approach: When computational resources are sparse, affine transformation offers reasonable accuracy; when computational resources are available, diffeomorphic deformation provides superior results; for the organs outside the field of view, one may use the alternative baseline method.

The geometry validation was performed using DSC. We also evaluated the predicted anatomy and the reference using the mean surface distance and organ volume recovery coefficients with both metrics showing similar results compared to DSC.

This study has some limitations. First, the XCAT dataset is limited in number and the CT images were scanned a decade ago with relatively low image quality. However, this dataset was satisfactory to show the proof of concept of the *i*Phantom framework. In future work, we will incorporate more segmented CT images with current imaging protocols and perform further training to address this limitation. We anticipate the segmentation and organ dose estimation accuracy will be improved. Second, current implementation of diffeomorphic deformation used a global regularization which might limit the accuracy taking into account inter-organ sliding motion; further improvement can be implemented using adaptive regularization [36]. Third, the overall errors of the framework were approximated by combining separated errors from segmentation and deformation. In future work, the framework will be further validated using a leave-multi-organ out strategy and incorporate existing phantoms other than the XCAT with reference manual segmentation of both anchor and non-anchor organs. Fourth, in our previous studies, the MC-GPU code accuracy was reported for only fixed tube current. With the same program used for both reference and predicted phantoms, here we reported more comprehensive dose estimation experiments using both fixed and modulated tube current function. We are also aiming to improve the TCM implementation in our future studies. These limitations are pathways for future improvements of the *i*Phantom methodology with present demonstrated quality and capability.

VII. CONCLUSION

Computational human phantoms (CHPs) are essential for personalized clinical investigations and population-based simulation studies. However, their utility and generalization have been limited by current approaches in creating CHPs using manual segmentation. In this study, we proposed a novel framework, *i*Phantom, for automated and accurate creation of patient-specific CHPs from patient medical images. We showed that the framework precisely localized a wide range of organs, including low contrast organs, in CT images. Specifically, we presented an integrated framework built on fusing patient-specific automated learning-based segmentation with anatomical templates through template matching and diffeomorphic deformation. This framework was applied to patient-specific organ dosimetry, yielding a high accuracy (< 10% organ dose error) across radiosensitive organs. The components of the framework are modular and thus each can be further optimized for customized applications. This methodology may be useful for other applications, for example when dealing with hard-to-segment organs, lack of initial training data, and organ-based image quality evaluation.

Acknowledgments

REFERENCES

[1]. Kainz W. et al. , "Advances in Computational Human Phantoms and Their Applications in Biomedical Engineering—A Topical Review," IEEE transactions on radiation and plasma medical sciences, vol. 3, no. 1, pp. 1–23, 2018.

[2]. Xu XG, "An exponential growth of computational phantom research in radiation protection, imaging, and radiotherapy: a review of the fifty-year history," Physics in Medicine & Biology, vol. 59, no. 18, p. R233, 2014. [PubMed: 25144730]

[3]. Petoussi-Henss N, Zankl M, Fill U, and Regulla D, "The GSF family of voxel phantoms," Physics in Medicine & Biology, vol. 47, no. 1, p. 89, 2001.

[4]. Lee C, Williams JL, Lee C, and Bolch WE, "The UF series of tomographic computational phantoms of pediatric patients," Medical physics, vol. 32, no. 12, pp. 3537–3548, 2005. [PubMed: 16475752]

[5]. Xu XG, Taranenko V, Zhang J, and Shi C, "A boundary-representation method for designing whole-body radiation dosimetry models: pregnant females at the ends of three gestational periods —RPI-P3,-P6 and-P9," Physics in Medicine & Biology, vol. 52, no. 23, p. 7023, 2007. [PubMed: 18029991]

[6]. Badano A. et al., "In silico imaging clinical trials for regulatory evaluation: initial considerations for VICTRE, a demonstration study," in Medical Imaging 2017: Physics of Medical Imaging, 2017, vol. 10132: International Society for Optics and Photonics, p. 1013220.

[7]. Schmidt TG, Wang AS, Coradi T, Haas B, and Star-Lack J, "Accuracy of patient-specific organ dose estimates obtained using an automated image segmentation algorithm," Journal of Medical Imaging, vol. 3, no. 4, p. 043502, 2016. [PubMed: 27921070]

[8]. Segars WP, Tsui BM, Cai J, Yin F-F, Fung GS, and Samei E, "Application of the 4-D XCAT phantoms in biomedical imaging and beyond," IEEE transactions on medical imaging, vol. 37, no. 3, pp. 680–692, 2017. [PubMed: 28809677]

[9]. Cho J. et al. , "Machine learning powered automatic organ classification for patient specific organ dose estimation," in Proceedings of the Society for Imaging Informatics in Medicine Annual Meeting, 2017.

[10]. Peng Z. et al. , "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," Medical Physics, 2020.

[11]. Gibson E. et al. , "Automatic multi-organ segmentation on abdominal CT with dense v-networks," IEEE transactions on medical imaging, vol. 37, no. 8, pp. 1822–1834, 2018. [PubMed: 29994628]

[12]. Pawlowski N. et al. , "Dltk: State of the art reference implementations for deep learning on medical images," arXiv preprint arXiv:1711.06853, 2017.

[13]. Segars W. et al. , "Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization," Medical physics, vol. 40, no. 4, p. 043701, 2013. [PubMed: 23556927]

[14]. Segars W. et al. , "The development of a population of 4D pediatric XCAT phantoms for imaging research and optimization," Medical physics, vol. 42, no. 8, pp. 4719–4726, 2015. [PubMed: 26233199]

[15]. Tward DJ et al. , "Patient specific dosimetry phantoms using multichannel LDDMM of the whole body," International journal of biomedical imaging, vol. 2011, 2011.

[16]. Fu W, Segars WP, Abadi E, Sharma S, Kapadia AJ, and Samei E, "From patient-informed to patient-specific organ dose estimation in clinical computed tomography," in Medical Imaging 2018: Physics of Medical Imaging, 2018, vol. 10573: International Society for Optics and Photonics, p. 1057315.

[17]. Sharma S, Kapadia A, Fu W, Abadi E, Segars WP, and Samei E, "A real-time Monte Carlo tool for individualized dose estimations in clinical CT," Physics in Medicine & Biology, vol. 64, no. 21, p. 215020, 2019. [PubMed: 31539892]

[18]. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, and Ronneberger O, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in International conference on medical image computing and computer-assisted intervention, 2016: Springer, pp. 424–432.

[19]. Taghanaki SA et al. , "Combo loss: handling input and output imbalance in multi-organ segmentation," Computerized Medical Imaging and Graphics, vol. 75, pp. 24–33, 2019. [PubMed: 31129477]

[20]. Tian X, Li X, Segars WP, Frush DP, and Samei E, "Prospective estimation of organ dose in CT under tube current modulation," Medical physics, vol. 42, no. 4, pp. 1575–1585, 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4379759/pdf/MPHYA6-000042-001575_1.pdf. [PubMed: 25832048]

[21]. Li X. et al. , "Effects of protocol and obesity on dose conversion factors in adult body CT," Medical physics, vol. 39, no. 11, pp. 6550–6571, 2012. [PubMed: 23127050]

[22]. Sotiras A, Davatzikos C, and Paragios N, "Deformable medical image registration: A survey," IEEE transactions on medical imaging, vol. 32, no. 7, pp. 1153–1190, 2013. [PubMed: 23739795]

[23]. Avants BB, Epstein CL, Grossman M, and Gee JC, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," Medical image analysis, vol. 12, no. 1, pp. 26–41, 2008. [PubMed: 17659998]

[24]. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, and Gee JC, "A reproducible evaluation of ANTs similarity metric performance in brain image registration," Neuroimage, vol. 54, no. 3, pp. 2033–2044, 2011. [PubMed: 20851191]

[25]. Pluta J, Avants BB, Glynn S, Awate S, Gee JC, and Detre JA, "Appearance and incomplete label matching for diffeomorphic template based hippocampus segmentation," Hippocampus, vol. 19, no. 6, pp. 565–71, 6 2009, doi: 10.1002/hipo.20619. [PubMed: 19437413]

[26]. Samei E, Tian X, and Segars WP, "Determining organ dose: the holy grail," Pediatric radiology, vol. 44, no. 3, pp. 460–467, 2014. [PubMed: 25304705]

[27]. Badal A. and Badano A, "Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit," Medical physics, vol. 36, no. 11, pp. 4878–4880, 2009. [PubMed: 19994495]

[28]. Keras. (2015). [Online]. Available: https://keras.io/

[29]. Abadi M. et al. , "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467, 2016.

[30]. Kingma DP and Ba J, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[31]. Li X, Segars WP, and Samei E, "The impact on CT dose of the variability in tube current modulation technology: a theoretical investigation," Physics in Medicine & Biology, vol. 59, no. 16, p. 4525, 2014. [PubMed: 25069102]

[32]. Tian X, Segars WP, Dixon RL, and Samei E, "Convolution-based estimation of organ dose in tube current modulated CT," Physics in Medicine & Biology, vol. 61, no. 10, p. 3935, 2016. [PubMed: 27119974]

[33]. Akhavanallaf A, Xie T, and Zaidi H, "Development of a library of adult computational phantoms based on anthropometric indexes," IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 3, no. 1, pp. 65–75, 2018.

[34]. Johnson PB, Whalen SR, Wayson M, Juneja B, Lee C, and Bolch WE, "Hybrid patient-dependent phantoms covering statistical distributions of body morphometry in the US adult and pediatric population," Proceedings of the IEEE, vol. 97, no. 12, pp. 2060–2075, 2009.

[35]. Wang H. et al. , "Deformable torso phantoms of Chinese adults for personalized anatomy modelling," Journal of anatomy, vol. 233, no. 1, pp. 121–134, 2018. [PubMed: 29663370]

[36]. Zhou SK, Rueckert D, and Fichtinger G, Handbook of medical image computing and computer assisted intervention. Academic Press, 2019.
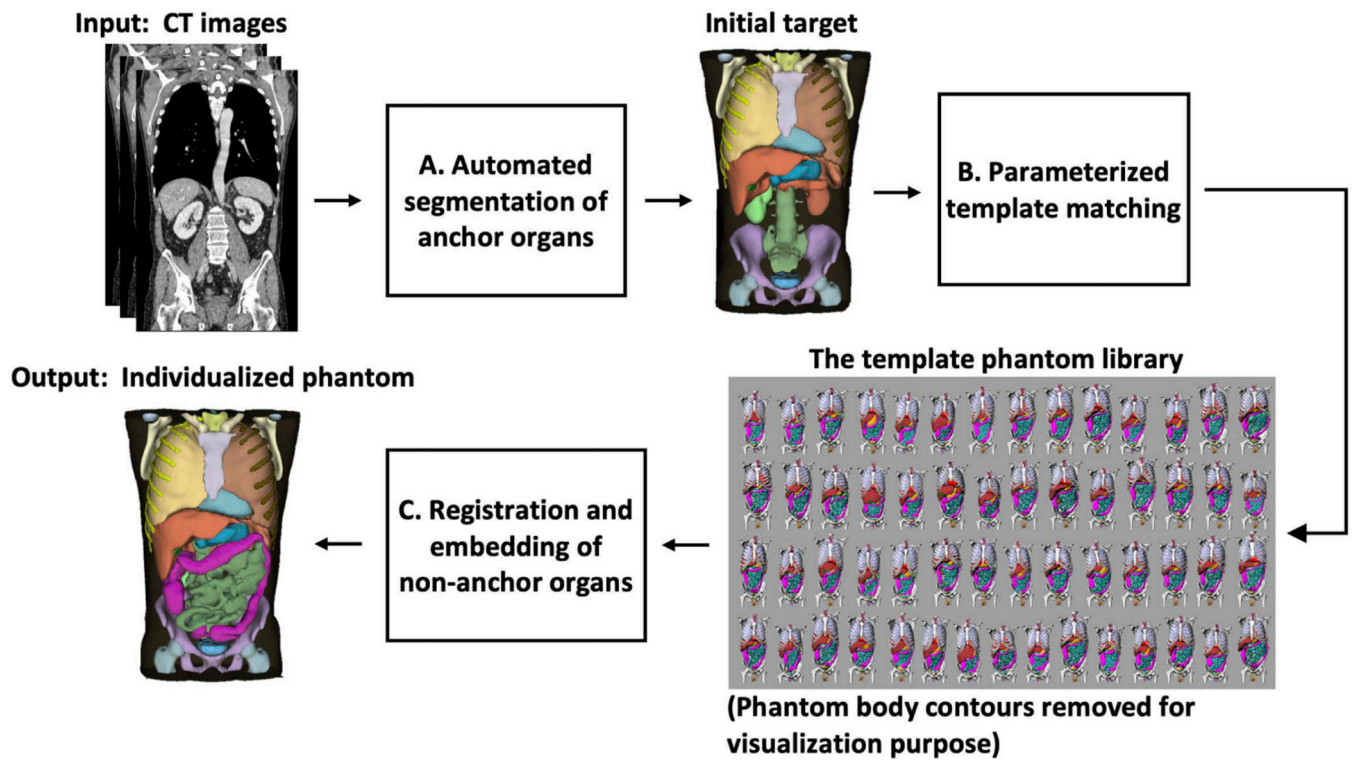
Fig. 1.
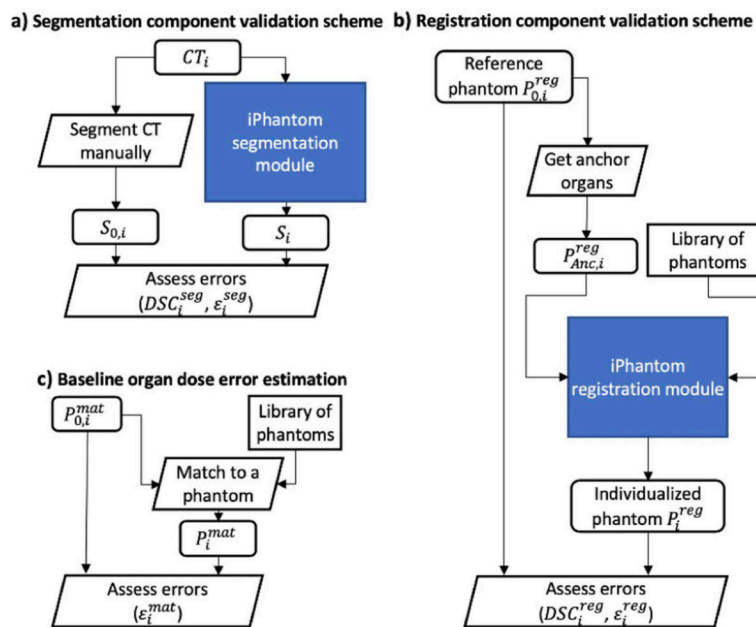Flow-chart of the *i*Phantom framework.

Fig. 2.
Validation scheme for the XCAT experiments with a) the segmentation component, b) registration component, and c) the baseline organ dose error estimation. Oval represents data, parallelogram represents operation, rectangular represents pre-computed functions. Superscripts seg, reg, mat represent studies for segmentation, registration, and matching respectively. Subscripts 0 represents reference, and i represents the ith phantom. The symbols are listed in Table 1.
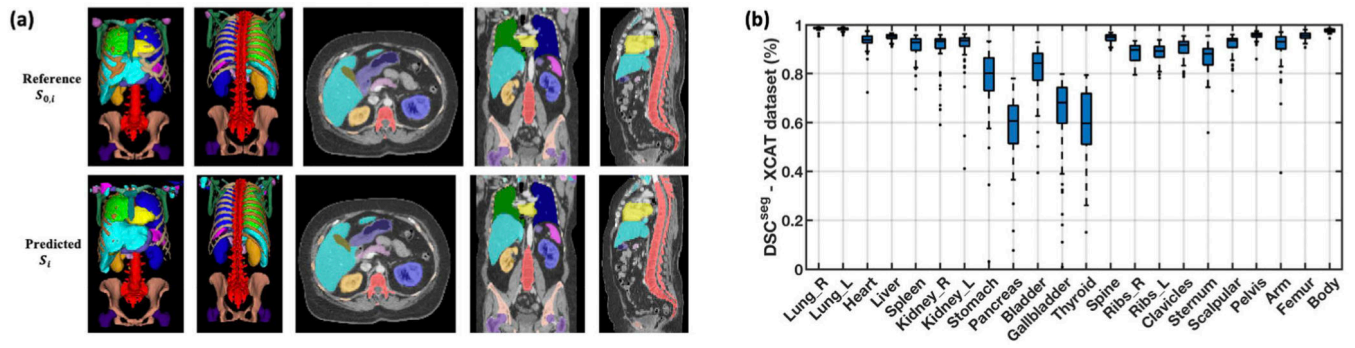
Fig. 3.

Gemometry valiation results of the segmentation component using XCAT datasets in a five-fold cross validation on the test set. a) Example segmentation results overlaid onto the patient CT data for patient with median cross-organ average DSC. b) Box plot summarizing the DSC results calculated from the reference and predicted segmentations from each test case.
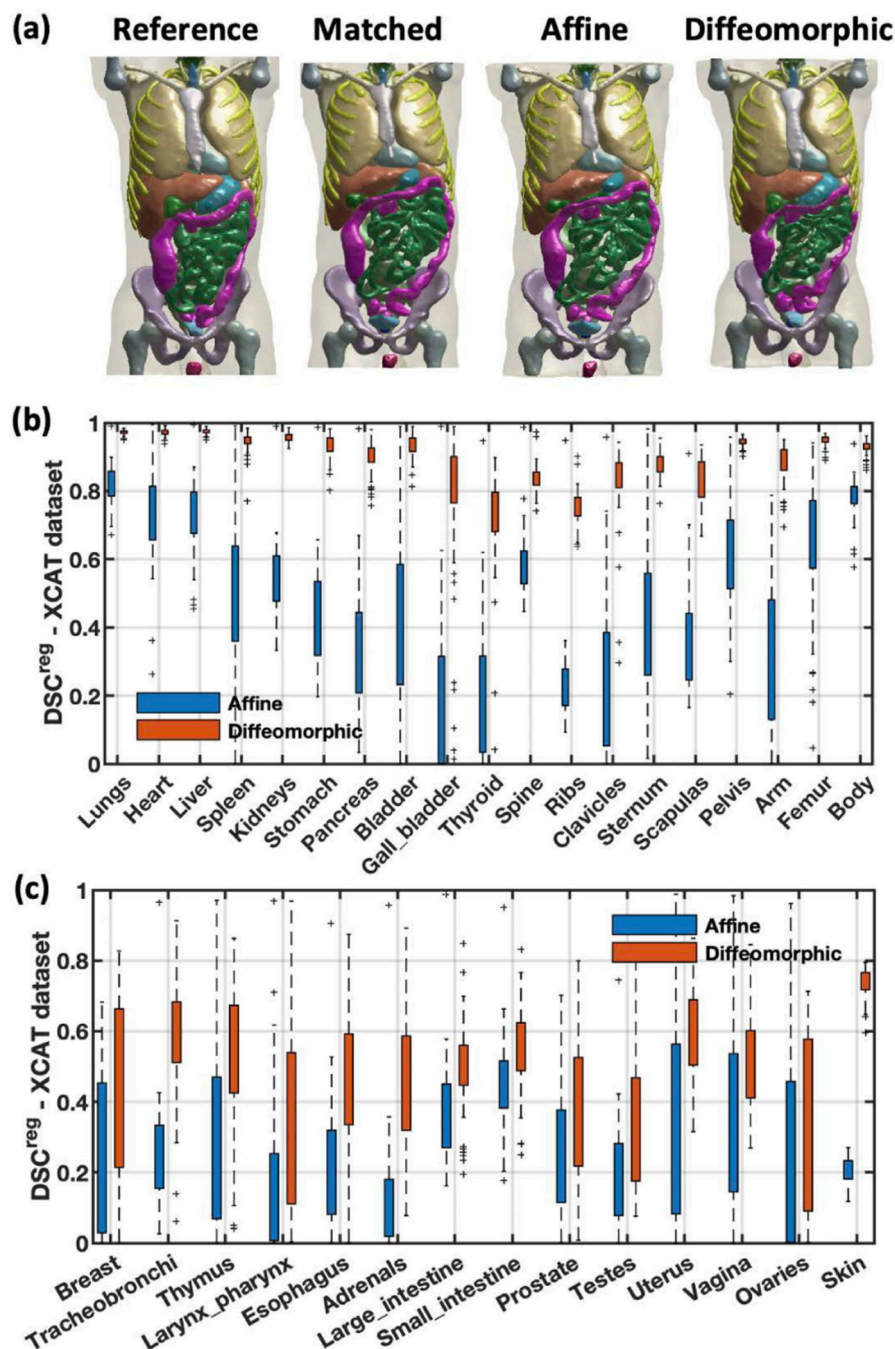
Fig. 4.
Results of the registration component validation using the leave-one-phantom out approach with the XCAT dataset. a) Rendering of a reference, the matched, and the matched model with affine and diffeomorphic transformations. b, c) Box plot of dice similarity coefficients (DSC) between the reference and matched with transformation for b) anchor organs and c) non-anchor organs for all XCAT phantoms. The results show that the diffeomorphic transformation improves the prediction for the filled in organs.
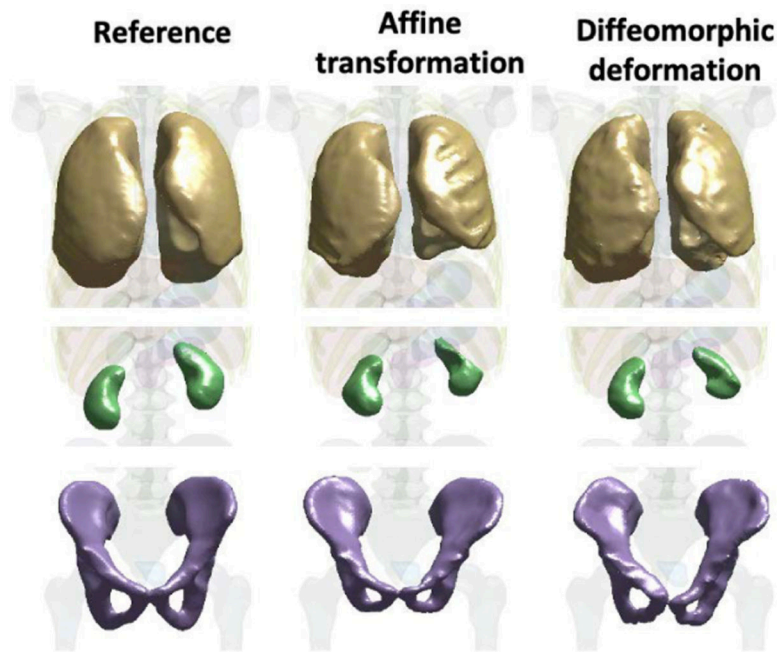
Fig. 5.
Registration geometry-based validation of the XCAT dataset using the leave-one-organ-out approach. Rendering of an example case (same case as in Fig. 4) where the lungs (top), kidneys (middle), and pelvis (bottom) are left out respectively and filled in using affine transformation and diffeomorphic deformation.
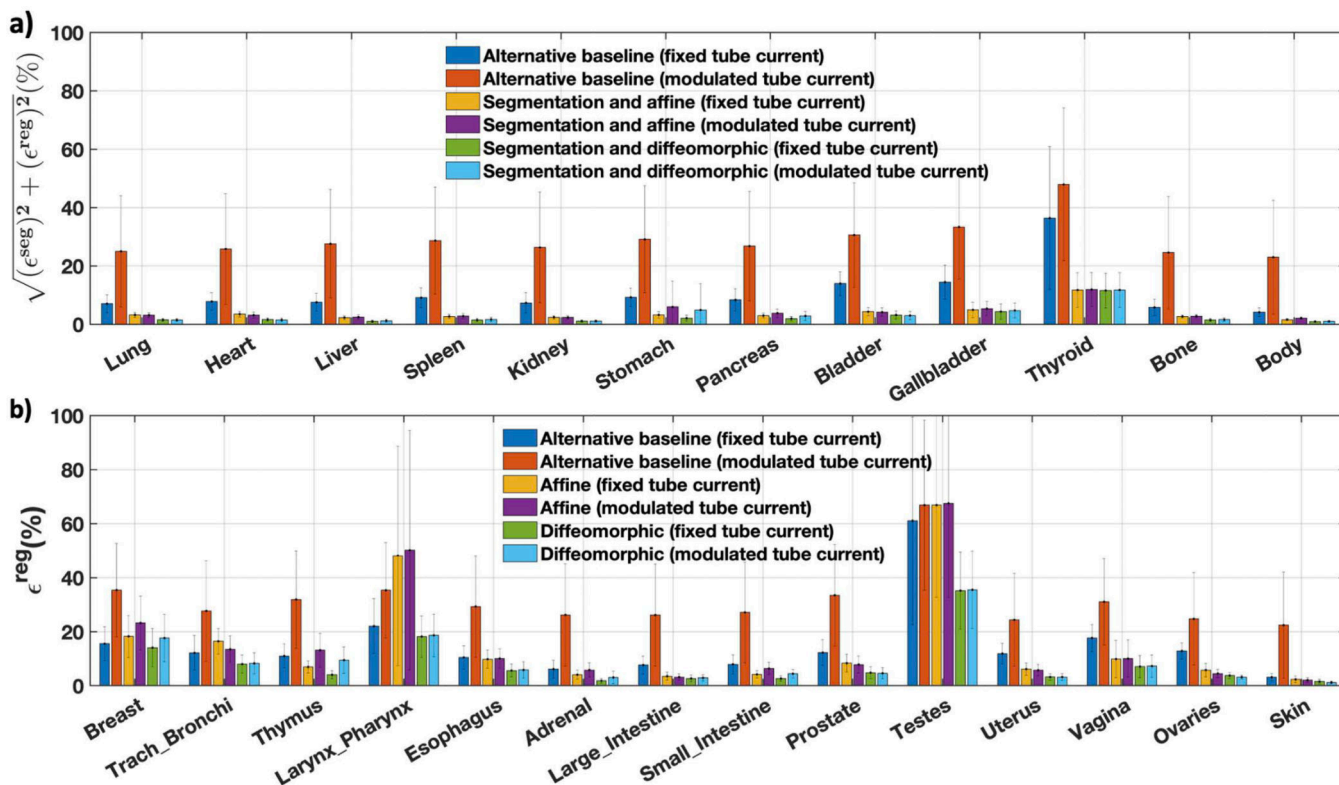
Fig. 6.
Absolute relative error (%) of organ doses from phantoms from the test set and the reference averaged across the 50 XCAT for scans in leave-one-phantom-out validation. a) Organ dose errors (combining errors from segmentation and registration or an alternative baseline method) calculated for anchor organs. b) Organ dose errors calculated for the embedded organs predicted using registration.
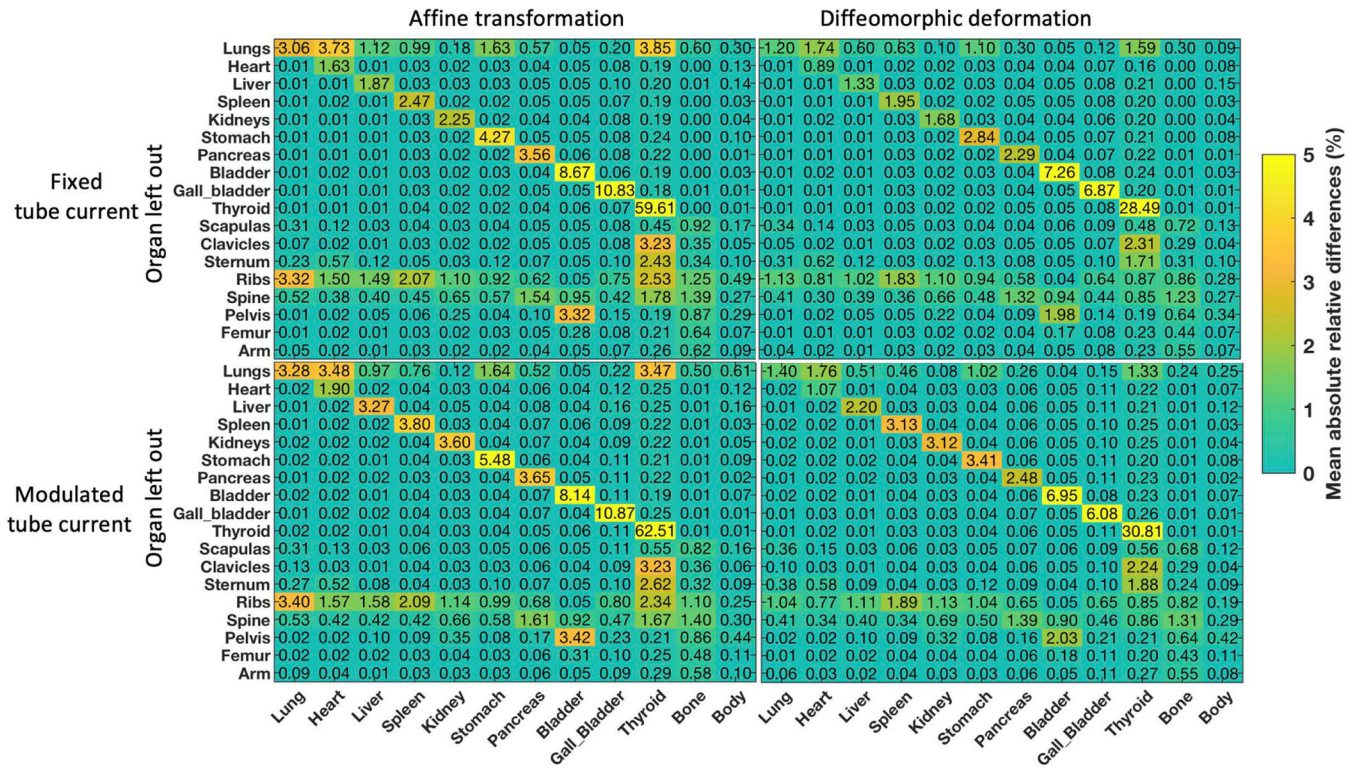
Fig. 7.
Dosimetry validation – XCAT phantom results. Results from the leave-one-organ-out validation where each row represents an experiment with the specified organ out and each column represents the organ dose mean absolute errors (MAE %).
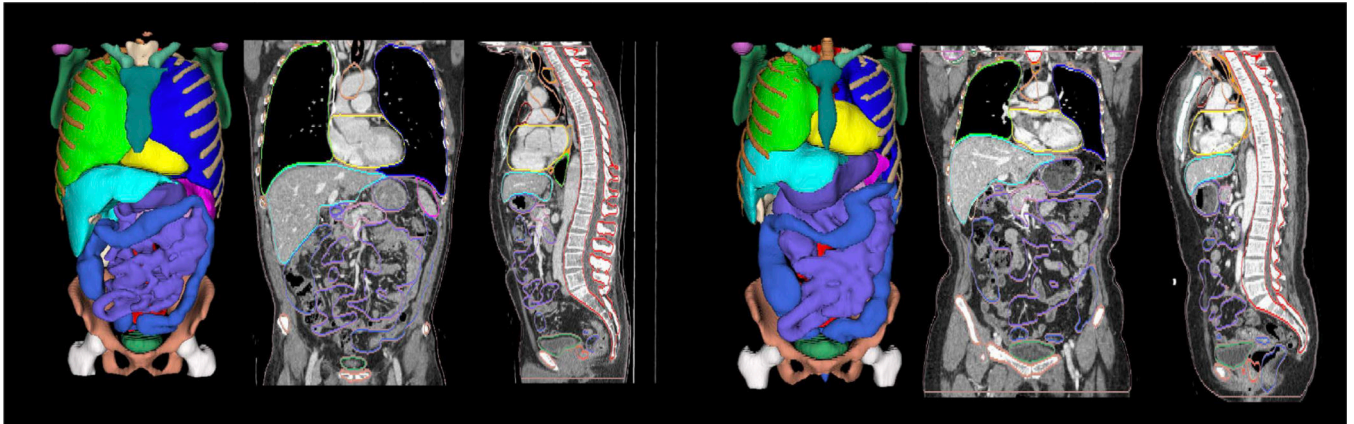
Fig. 8.

Clinical validation. P phantoms generated by automated segmentation and registration from the proposed framework of two example cases. 3D renderings of the phantoms are shown as well as organ outlines from the phantoms overlaid onto coronal and sagittal slices of the patient CT data from which they were generated.
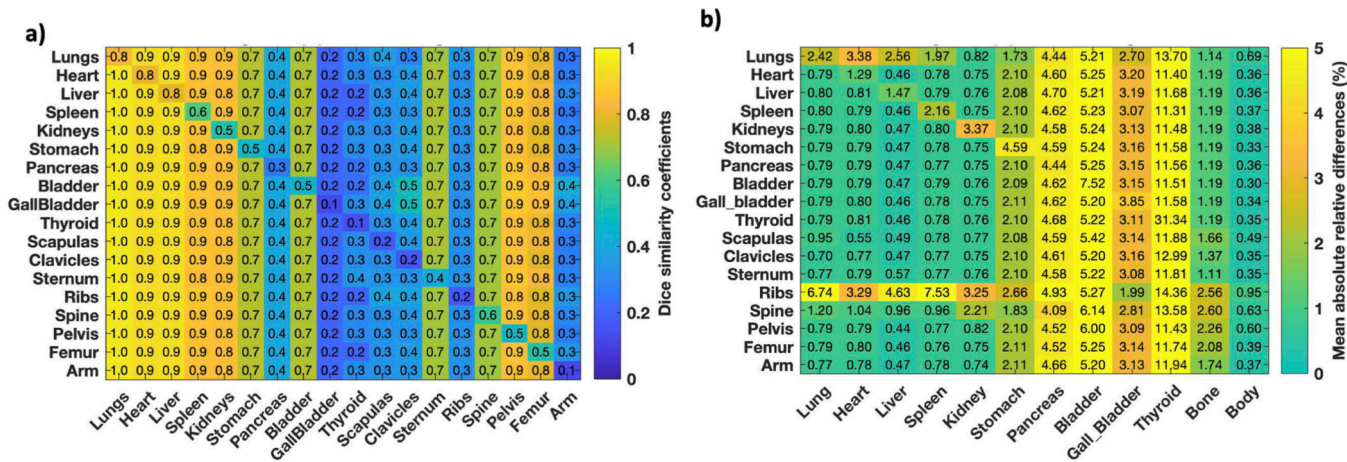
Fig. 9.
Clinical validation: leave-one-organ-out experiment. a) DSC of reference phantoms and the phantoms created using segmentation and diffeomorphic deformation from the 10 clinical datasets. b) Mean absolute dose errors (MAE %) of reference phantoms and the phantoms created using segmentation and diffeomorphic deformation. For both a) and b), each row represents an experiment with the specified organ left out in the segmentations. The organ was subsequently predicted through deformation. Each column shows the resulting values of DSC or MAE to the organs in each experiment.

TABLE 1.

TABLE OF SYMBOLS

i	The ith phantom generated by the test set
$0, i$	The ith reference phantom
anc	Phantom with anchor organs only
non-anc	Phantom with non-anchor organs
MAE	Mean absolute relative error across all test set
DSC	Dice similarity coefficients
DSC_i^{seg}, DSC_i^{reg}	DSC between the ith test set and reference phantom for segmentation and registration validation, respectively.
$\varepsilon_i^{seg}, \varepsilon_i^{reg}, \varepsilon_i^{mat}$	MAE of organ doses between the ith test set and reference phantom for segmentation validation, registration validation, and matching method, respectively