Generating anthropomorphic phantoms using fully unsupervised deformable image registration with convolutional neural networks

Junyu Chen[a)] and Ye Li
Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21287, USA
Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins Medical Institutes, Baltimore, MD 21287, USA

Yong Du
Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins Medical Institutes, Baltimore, MD 21287, USA

Eric C. Frey
Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21287, USA
Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins Medical Institutes, Baltimore, MD 21287, USA

(Received 17 May 2020; revised 28 August 2020; accepted for publication 9 October 2020; published xx xxxx xxxx)

Purpose: Computerized phantoms have been widely used in nuclear medicine imaging for imaging system optimization and validation. Although the existing computerized phantoms can model anatomical variations through organ and phantom scaling, they do not provide a way to fully reproduce the anatomical variations and details seen in humans. In this work, we present a novel registration-based method for creating highly anatomically detailed computerized phantoms. We experimentally show substantially improved image similarity of the generated phantom to a patient image.
Methods: We propose a deep-learning-based unsupervised registration method to generate a highly anatomically detailed computerized phantom by warping an XCAT phantom to a patient computed tomography (CT) scan. We implemented and evaluated the proposed method using the NURBS-based XCAT phantom and a publicly available low-dose CT dataset from TCIA. A rigorous tradeoff analysis between image similarity and deformation regularization was conducted to select the loss function and regularization term for the proposed method. A novel SSIM-based unsupervised objective function was proposed. Finally, ablation studies were conducted to evaluate the performance of the proposed method (using the optimal regularization and loss function) and the current state-of-the-art unsupervised registration methods.
Results: The proposed method outperformed the state-of-the-art registration methods, such as SyN and VoxelMorph, by more than 8%, measured by the SSIM and less than 30%, by the MSE. The phantom generated by the proposed method was highly detailed and was almost identical in appearance to a patient image.
Conclusions: A deep-learning-based unsupervised registration method was developed to create anthropomorphic phantoms with anatomies labels that can be used as the basis for modeling organ properties. Experimental results demonstrate the effectiveness of the proposed method. The resulting anthropomorphic phantom is highly realistic. Combined with realistic simulations of the image formation process, the generated phantoms could serve in many applications of medical imaging research. [https://doi.org/10.1002/mp.14545]

Key words: computerized Phantom, convolutional Neural Networks, deep Neural Networks, image Registration, medical Image Simulation

1. INTRODUCTION

Computerized phantoms for nuclear medicine imaging research have been built based on anatomical and physiological models of human beings. They have played a crucial part in evaluation and optimization of medical image reconstruction, processing, and analysis methods.[1–4] Since the exact structural and physiological properties of the phantom are known, they can serve as a gold standard for the evaluation and optimization process. The four-dimensional extended cardiac-torso (XCAT) phantom[5] was developed based on anatomical images from the Visible Human Project data. This realistic phantom includes parameterized models for anatomy, which allows the generation of a series of phantoms with different anatomical variations. These phantoms have been used in Nuclear Medicine imaging and CT research,[6–12] as well as in various applications of deep learning.[13–15]

In the XCAT phantom, changing the values of parameters that control organ anatomy can be used to vary the volumes and shapes of some tissues. However, the scaling of organs, even when different factors are used in orthogonal directions, does not fully and realistically capture the anatomical

variations of organs within different human bodies. However, for many applications, having a population of phantoms that models the variations in patient anatomy and, in nuclear medicine, uptake realization is essential for comprehensive validation and training of image processing and reconstruction algorithms. To solve this, in Ref. [16]. Segars et al. used a deformable image registration technique to map phantom labels to segmented patient images; the resulting deformation fields were then applied to the phantom, thus creating a population of new XCAT models that capture the anatomical variability among patients. This method relies on the segmentation of patient images, which is tedious and time consuming. In this work, we propose an approach based on Convolutional neural networks (ConvNets) to perform phantom to patient registration without requiring the patient segmentation. The resulting deformation field can then be applied to organ label maps to generate a gold-standard segmentation for the deformed phantom image.

Deformable image registration is a process of transforming two images into a single coordinate system, where one image is often referred to as the moving image, denoted by I_m, and the other is referred to as the fixed image, denoted by I_f. Traditional methods formulate registration as a variational problem for estimating a smooth mapping, ϕ, between the points in one image and those in another. They often tend to iteratively minimize the following energy function [Eq. (1)] on a single image pair[17]:

$$E = E_{sim}(I_m \circ \phi, I_f) + R(\phi), \tag{1}$$

where, E_{sim} measures the level of alignment between the transformed moving image, $I_m \circ \phi$, and the fixed image, I_f. Some common choices for E_{sim} are mean squared error (MSE) or the squared L^2 norm of the difference,[18] sum of squared differences (SSD),[19] cross-correlation (CC),[20] and mutual information (MI).[21] The transformation, ϕ, at every point is defined by an identity transformation with the displacement field $\mathbf{u}$, or $\phi = Id + \mathbf{u}$, where Id represents the identity transform.[22] The second term, $R(\phi)$, is referred to as the regularization of the deformation, ϕ, which enforces spatial smoothness. Many regularization designs have been proposed previously based on different applications and prior knowledge about the deformation field. However, they are usually characterized by the gradients of $\mathbf{u}$. In some applications, regularizers were designed to take sliding organs into account,[23–15] where instead of enforcing global smoothness, they preserve motion discontinuities allowing multiple organs to move independently. In most applications, a common assumption is that similar structures are present in both moving and fixed images. Hence, a continuous and invertible deformation field (a diffeomorphism) is desired, and the regularization term, $R(\phi)$, is designed to enforce or encourage this. Diffeomorphisms can be essential in some studies, for which the registration field is analyzed further. However, in the application of registration-based segmentation, the quality of the segmentation propagation is more critical than the diffeomorphic property of the underlying deformation fields.[26] In this study, due to the large interior and exterior shape variations between digital phantoms and patients, diffeomorphism is less important. However, we show that by introducing various regularizers to the proposed model, the number of noninvertible voxel transformations in the resulting deformation field can be substantially reduced.

Recently, many deep learning-based methods have been proposed to perform registration tasks (summarized in Table I). Some of the methods rely on ground truth deformation fields,[27–29] which are often obtained by simulated deformations or applying classical registration algorithms on a pair of images. However, ground truth registration fields are time-consuming to obtain and can limit the types of deformations that are learned. While other methods, such as,[22, 30–34], were introduced as unsupervised (or more precisely, self-supervised) techniques, but they still require a prior training stage with a large amount of training data. These methods assume that neural networks can provide a universal and generalized model for image registration by minimizing the registration energy function over a dataset of images. This is a common assumption with deep-learning-based approaches. Yet, such an assumption could be unreliable according to a recent study from Zhang et al.,[35] where they showed that a well-generalized CNN classifier trained by a large dataset can still easily overfit a random labeling of the training data. Other studies on fooling deep neural networks (DNNs) with adversarial images also suggest that the well-trained networks can be unstable to small or even tiny perturbations of the data.[36–40] On the other hand, the proposed registration method is fully unsupervised, meaning that ***no previous training is required***. Instead of following the conventional pattern of training a network on a large dataset of training images, we show that a CNN can estimate an optimal deformation field for a single image pair by minimizing the energy function described in Eq. (1) iteratively. This idea was inspired by Lempitsky et al.'s work on the Deep Image Prior[41] (DIP), where they showed that learning from a large amount of data is not necessary for building useful image priors, but the structure of a ConvNet itself is sufficient to capture image statistics. They treated the training of ConvNets with random initialization as a regularization prior, and in order to achieve good solutions in their application of image

TABLE I. Conceptual comparisons among various registration methods and the proposed method (UnsupConvNet)

Registration methods	DNN-based	Supervision	Operating mode
VoxelMorph[22,30,33]	Yes	Self-supervised	Whole image
DLIR[32]	Yes	Self-supervised	Patch-based
DIRNet[31]	Yes	Self-supervised	Patch-based
FAIM[62]	Yes	Self-supervised	Whole image
RegNet[29]	Yes	Supervised	Patch-based
RobustRegNet[28]	Yes	Supervised	ROI-based
SyN[20]	No	Unsupervised	Whole image
LDDMM[18]	No	Unsupervised	Whole image
UnsupConvNet	Yes	Unsupervised	Whole image

denoising, determining early stopping points was often required. Whereas in image registration, instead of starting from a random initialization (i.e., random noise images), it makes logical sense to initialize the ConvNet with a moving image. Since one would like to transform the moving image so that it is similar to a target image as possible, early stopping is not desired. In this work, we treat ConvNet as an optimization tool, where it generates a deformation field that minimizes the difference between deformed moving and fixed images by updating its parameter values in each iteration. The deformation is realized with a spatial transformer constructed based on the spatial transformer networks[42] and VoxelMorph.[22] It differs from the B-Spline grids used in,[29,31,32] which are only demonstrated on sub-regions (patches) of images, support only small transformations, and impose implicit regularization defined by interpolation methods.[22,23] On the contrary, the control points of the spatial transformer used in this work were applied to all the pixel locations. This, thus, enables large deformations and allows for external regularization.

2. MATERIALS AND METHODS

2.A. Computerized phantom generation

The phantom used in this study was created from the three-dimensional (3D) attenuation distribution of the realistic NURBS-based XCAT phantom.[43] Attenuation values were computed based on the material compositions of the materials and the attenuation coefficients of the constituents at 140 keV, the photon energy of Tc-99m. This single 3D phantom image was deformed to multiple patient CT images. The simulated attenuation map image can be treated as the template image, and phantom label map can then be thought of as the atlas in the traditional paradigm of medical image registration. The aim is to first register the phantom attenuation map image to patient CT images. Next, the registration parameters would be applied to the XCAT phantom label map (used to define organs and thus the activity distribution) to create new anthropomorphic phantoms. For the nuclear medicine imaging application, new images would be generated from the resulting phantoms using conventional physics-based simulation codes.[44–49]

2.B. Image registration with ConvNet

Let the moving image be I_m, and the fixed image be I_f; we assume that they are grayscale images defined over a n-dimensional spatial domain $\Omega \subset \mathcal{R}^n$ and affinely aligned. This paper primarily focuses on the two-dimensional (2D) case (i.e., $n = 2$), but the implementation is dimension independent (Notice that for $n > 2$ cases, the required GPU memory will be significant increased). We model the computation of the displacement field, ϕ, given the image pair, I_m and I_f, using a deep ConvNet with parameters θ, that is, $f_\theta(I_m, I_f) = \phi$. Figure 1 describes the architecture of the proposed method; it consists of a ConvNet that outputs a registration field, and a B-spline spatial transformer. First, the ConvNet generates the ϕ for the given image pair, I_m and I_f. Second, the deformed moving image is obtained by applying a B-spline spatial transformer that warps I_m with ϕ (i.e., $I_m \circ \phi$). Finally, we backpropagate the loss computed from the similarity measure between $I_m \circ \phi$ and I_f to update θ in the ConvNet. The steps are repeated iteratively until the loss converges; the resulting ϕ then represents the optimal registration field for the given image pair. The loss function ($\mathcal{L}$) of this problem can be formulated mathematically as:

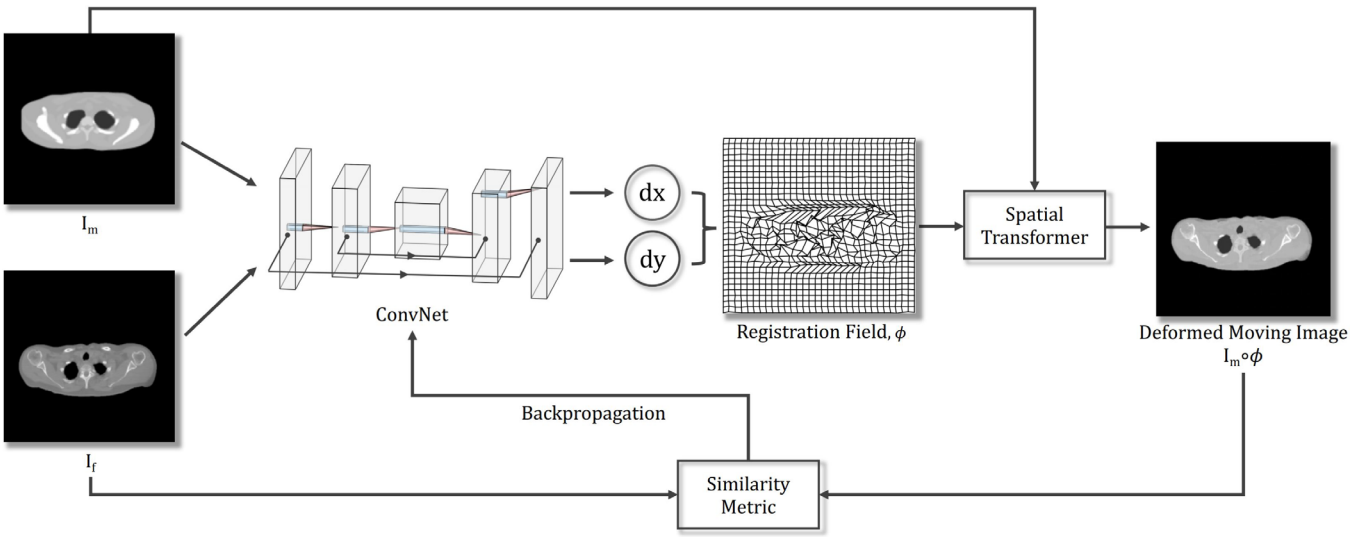

FIG 1. Schematic of the proposed method. The network takes a pair comprised of one moving and one fixed image as its inputs. The ConvNet learns from a single image pair and generates a deformation field, ϕ. We then warp the moving image I_m with ϕ using a B-spline spatial transformer. The loss determined by the image similarity measure between $I_m \circ \phi$ and I_f is then backpropagated to update the parameters in the ConvNet. Since no aspect of the ConvNet is learned from a prior training stage, the method follows a fully unsupervised paradigm.

$$\begin{aligned}\mathcal{L}(I_m, I_f, \phi; \theta) &= \mathcal{L}_{sim}(I_m \circ \phi, I_f; \theta) + \lambda R(\phi; \theta) \\ &= \mathcal{L}_{sim}(I_m \circ f_\theta(I_m, I_f), I_f; \theta) \\ &+ \lambda \mathcal{R}(f_\theta(I_m, I_f), \theta).\end{aligned} \tag{2}$$

where $\mathcal{L}_{sim}$ is the image similarity measure and $\mathcal{R}$ represents the regularization of ϕ Then, the parameters θ that generate the optimal registration field can be estimated by the minimizer:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}(I_m, I_f, \phi; \theta), \tag{3}$$

and the optimal ϕ is then given by:

$$\phi^* = f_{\theta^*}(I_m, I_f). \tag{4}$$

Different choices of image similarity metrics and registration field regularizers ($R(\phi)$) were also studied in this work, and they are described in detail in a later section. The next subsection describes the design of ConvNet architecture.

2.B.1. ConvNet architecture

The ConvNet had a U-Net-like "hourglass" architecture.[50] The network consisted of one encoding path, which takes a single input formed by concatenating the moving and fixed images into a $2 \times M \times M$ volume, where $M \times M$ represents the shape of one image. Each convolutional layer had a 3×3 filter followed by a rectified linear unit (ReLU), and the downsampling was performed by 2×2 max pooling operations. In the decoding stage, the upsampling was done by "up-convolution."[50] Each of the upsampled feature maps in the decoding stage was concatenated with the corresponding feature map from the encoding path. The output registration field, ϕ, was generated by the application of sixteen 3×3 convolutions followed by two 1×1 convolutions to the 16 feature maps. This is a relatively small network with 98 794 trainable parameters in total. The network architecture is shown schematically in Fig. 2.

2.B.2. Spatial transformer

The spatial transformer was implemented on the basis of the spatial transformer networks,[42] which applies a nonlinear warp to the moving image, where the warp is determined by a flow field of displacement vectors ($\mathbf{u}$) that define the correspondences of pixel intensities in the output image to the pixel locations in the moving image. The number of control points in the flow field, ($\mathbf{u}$), is equal to the image's size; thus, the spacing of the control points grid is 1. The intensity at each pixel location, $\mathbf{p}$, in the output image, $I_m \circ \phi(\mathbf{p})$, is defined by:

$$I_m \circ \phi(\mathbf{p}) = I_m(\mathbf{p} - \mathbf{u}(\mathbf{p})). \tag{5}$$

Notice that $\mathbf{p}-\mathbf{u}(\mathbf{p})$ is not necessarily an integer, and pixel intensities are only defined at integer locations in the image. Therefore, the value of $I_m \circ \phi(\mathbf{p})$ was obtained by applying interpolation methods to the nearest pixels around $\mathbf{p}-\mathbf{u}(\mathbf{p})$:

$$I_m \circ \phi(\mathbf{p}) = \sum_{\mathbf{q} \in \mathcal{Z}(\mathbf{p}')} I_m(\mathbf{q}) \prod_{d \in \{x,y\}} k(\mathbf{p}'_d - \mathbf{q}_d), \tag{6}$$

where $\mathbf{p}' = \mathbf{p} - \mathbf{u}(\mathbf{p})$, $\mathcal{Z}(\mathbf{p}')$ represents the neighboring pixels of $\mathbf{p}'$, d iterates over dimensions of the spatial domain Ω, and $k()$ is a generic sampling kernel, which defines the image interpolation. We used, respectively, bi-linear and nearest-neighbor interpolation to obtain pixel values in the deformed XCAT attenuation map and the deformed organ labels. The nearest-neighbor interpolation reduces Eq. (6) to:

$$I_m \circ \phi(\mathbf{p}) = \sum_{\mathbf{q} \in \mathcal{Z}(\mathbf{p}')} I_m(\mathbf{q}) \prod_{d \in \{x,y\}} \delta(\lfloor \mathbf{p}'_d + 0.5 \rfloor - \mathbf{q}_d), \tag{7}$$

where $\delta()$ represents the Kronecker delta function, and $\lfloor \mathbf{p}'_d + 0.5 \rfloor$ rounds $\mathbf{p}'_d$ to nearest integer value. If $k()$ is a bi-linear sampling kernel, Eq. (6) is reduced to:

$$I_m \circ \phi(\mathbf{p}) = \sum_{\mathbf{q} \in \mathcal{Z}(\mathbf{p}')} I_m(\mathbf{q}) \prod_{d \in \{x,y\}} (1 - |\mathbf{p}'_d - \mathbf{q}_d|). \tag{8}$$

2.B.3. Image similarity metrics

Over the years, considerable effort has been expended designing image similarity metrics. We mentioned some of the metrics that have been widely adopted in image registration in the previous section. In this work, we studied the effectiveness of five different loss functions, we also propose a new $\mathcal{L}_{sim}$ that combines the advantages of both Pearson's

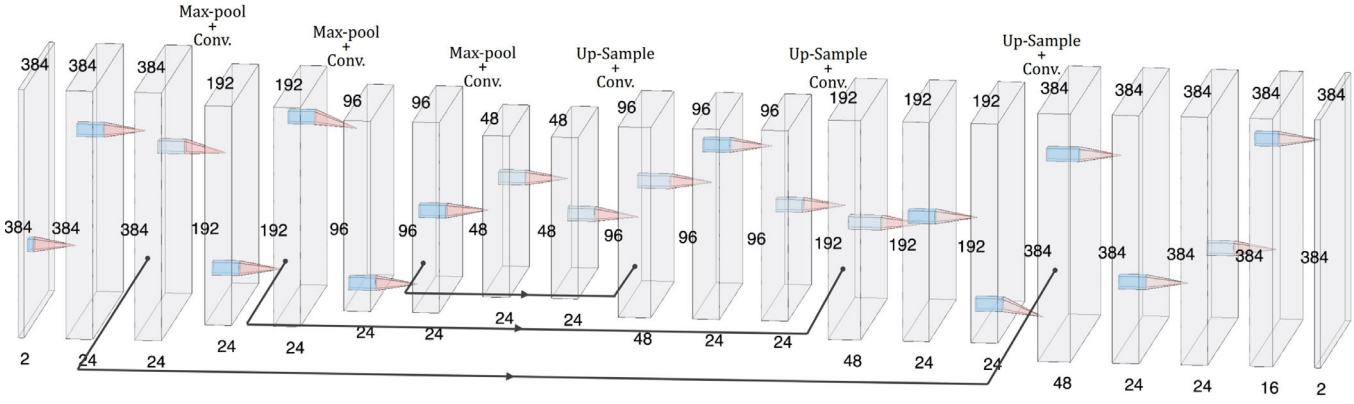

FIG 2. The ConvNet has a U-Net-like architecture.

Correlation Coefficient (PCC) and the Structural Similarity Index (SSIM). In the following subsections, we denote the deformed moving image as I_d (i.e., $I_d = I_m \circ \phi$) for simplicity.

Mean squared error (MSE): MSE is a measurement of fidelity, and indicates the degree of agreement of intensity values in images; it is applicable when I_f and I_m have similar contrast and intensity distributions. The loss function is defined as $\mathcal{L}_{sim}(I_m, I_f, \phi; \theta) = \text{MSE}(I_d, I_f)$.

Local cross correlation (CC): Another popular image similarity metric is CC, due to its robustness to intensity variations between images.[20,35,51] Since $\text{CC} \geq 0$, we minimize the negative CC, the loss function is $\mathcal{L}_{sim}(I_m, I_f, \phi; \theta) = -\text{CC}(I_d, I_f)$.

Mutual information (MI): MI was first applied to image registration in.[52] It measures the statistical dependence between the intensities of corresponding pixels in both moving and fixed images. Let $p_{I_f}(a)$ and $p_{I_d}(b)$ be the marginal probability distributions of the fixed and deformed moving images. MI is a measure of the Kullback-Leibler Divergence[53] between the joint distribution $p_{I_f I_d}(a,b)$ and the distribution associated with the case of complete independence $p_{I_f}(a) \cdot p_{I_d}(b)$[52]:

$$\text{MI}(I_d, I_f) = \sum_{a,b} p_{I_f I_d}(a,b) \log \frac{p_{I_f I_d}(a,b)}{p_{I_f}(a) \cdot p_{I_d}(b)}. \tag{9}$$

The joint distribution, $p_{I_f I_d}(a,b)$, can be computed as:

$$p_{I_f I_d}(a,b) = \frac{1}{\Omega} \sum_{i \in \Omega} \delta(a - I_f(i)) \delta(b - I_d(i)), \tag{10}$$

Notice that the Kronecker delta function, $\delta()$, is not differentiable. Therefore, the resulting loss cannot be back-propagated in the network. To solve this issue, we approximate $p_{I_f I_d}(a,b)$ with the Parzen windowing functions as described by Mattes et al. in [54]:

$$p_{I_f I_d}(a,b) = \alpha \sum_{i \in \Omega} \psi(a - I_f(i)) \chi(b - I_d(i)), \tag{11}$$

where α is a normalization factor that ensures $\sum p_{I_f I_d}(a,b) = 1$, and $\psi()$ and $\chi()$ are kernel functions of Parzen window. There is a broad choice for kernel functions, such as the first order or the third order B-Spline kernel,[54] and Gaussian kernel.[21] Since Cubic B-Spline has a close relationship with Gaussian functions,[55,56] we chose $\psi()$ and $\chi()$ to be Gaussian kernels in this work. The joint distribution, Eq. (11), can be rewritten as:

$$p_{I_f I_d}(a,b) = \alpha \sum_{i \in \Omega} \frac{1}{\Delta w_d \Delta w_m \pi} e^{\frac{(a - I_f(i))^2}{2\Delta w_f^2}} e^{\frac{(b - I_d(i))^2}{2\Delta w_d^2}}, \tag{12}$$

where Δw_f and Δw_d are the widths of each intensity bin in image I_f and I_d. A larger bin width potentially leads to an improvement in computational efficiency, because the number of intensity bins used in the estimation of the marginal probability distributions could be reduced. Finally, the two marginal probability distributions can be derived using $p_{I_f I_d}(a,b)$. Maximizing MI is equivalent to minimizing the negative of the MI. Thus, the loss function is formulated as $\mathcal{L}_{sim}(I_m, I_f, \phi; \theta) = -\text{MI}(I_d, I_f)$.

Pearson's correlation coefficient (PCC): PCC measures the linear correlation between two images. Unlike MSE, PCC is less sensitive to linear transformations of intensity values from one image to another. Its application to medical image registration is described in Ref. [57]. PCC is defined as the covariance between images divided by the product of their standard deviations:

$$\text{PCC}(I_d, I_f) = \frac{\sum_{i \in \Omega} (I_f(i) - \bar{I}_f)(I_d(i) - \bar{I}_d)}{\sqrt{\sum_{i \in \Omega} (I_f(i) - \bar{I}_f)} \sqrt{\sum_{i \in \Omega} (I_d(i) - \bar{I}_d)}} \tag{13}$$

where $\bar{I}_f$ and $\bar{I}_d$ represent the mean intensities. PCC has a range from -1 to 1, where 0 implies that there is no linear correlation, and -1 and 1 correspond, respectively, to the maximum negative and positive correlations between two images. Since a positive correlation is desired, we can define the loss function to be: $\mathcal{L}_{sim}(I_m, I_f, \phi; \theta) = 1 - \text{PCC}(I_d, I_f)$.

Structural similarity index (SSIM): SSIM was proposed in Ref. [58] for robust image quality assessments based on the degradation of structural information. Within a given image window, SSIM is defined by:

$$\text{SSIM}(I_d, I_f) = \frac{(2\mu_{I_d}\mu_{I_f} + C_1)(2\sigma_{I_f I_d} + C_2)}{(\mu_{I_f}^2 + \mu_{I_d}^2 + C_1)(\sigma_{I_f}^2 + \sigma_{I_d}^2 + C_2)}, \tag{14}$$

where C_1 and C_2 are small constants needed to avoid instability, μ_{I_f} and μ_{I_d}, and σ_{I_f} and σ_{I_d} are local means and standard deviations of the images I_f and I_d, respectively. The SSIM has a range from -1 to 1, where 1 indicates a perfect structural similarity. Thus, $\mathcal{L}_{sim}(I_m, I_f, \phi; \theta) = 1 - \text{SSIM}(I_d, I_f)$.

PCC + SSIM: While PCC is robust to noises, it was also found to be less sensitive to blurring. A motivating example is shown in Fig. 3, where in (b), the "Shepp-Logan" phantom image[50] was corrupted with Gaussian noise, and in (c), the image was blurred by a Gaussian filter. Both (b) and (c) yield a lower SSIM and a higher PCC. If we think of (a) as a moving image, and (b) and (c) as fixed images, SSIM would impose the ConvNets to model the details, including noises and artifacts. Whereas, using PCC alone as the loss function might converge to a less accurate result. Hence, there is a need to balance the two similarity measures. Both PCC and SSIM are bounded with a range from -1 to 1, where 1 indicates the most similar. Thus, we propose to combine SSIM and PCC with an equal weight:

$$\mathcal{L}_{sim}(I_m, I_f, \phi; \theta) = 0.5 * (1 - \text{SSIM}(I_d, I_f)) + 0.5 * (1 - \text{PCC}(I_d, I_f)) \tag{15}$$

2.C. Deformation regularization

The spacing of the control point grid is 1 pixel, thus the spatial transformer does not implicitly enforce any regularization (i.e., smoothness of the deformation field). Because each pixel can move freely, optimizing the image similarity metrics solely would encourage the deformed moving image, I_f, to be as close as possible to the fixed image, I_m. However, the resulting deformation field might not be smooth or realistic. To impose smoothness and weakly enforce diffeomorphism in the deformation field, we tested several different regularizers.

2.C.1. Diffusion regularizer

Balakrishnan, et al. used a diffusion regularizer in a ConvNet-based image registration model, VoxelMorph.[22] In this method, the regularization is applied on the spatial gradients of the displacement field $\mathbf{u}$:

$$\mathcal{R}_{Diffusion}(\phi; \theta) = \sum_{i \in \Omega} \|\nabla \mathbf{u}(i)\|^2, \tag{16}$$

where the spatial gradients are approximated by the forward difference, that is $\nabla \mathbf{u}(i) \simeq \mathbf{u}(i+1) - \mathbf{u}(i)$. Minimizing this the value of this regularizer leads to smaller spatial variations in the displacements, resulting in a smooth deformation field.

2.C.2. Total variation regularizer

Instead of using the squared L^2 norm as the diffusion regularizer, the total variation norm regularizes the L^1 norm on the spatial gradients of $\mathbf{u}$[60]:

$$\mathcal{R}_{TV}(\phi; \theta) = \sum_{i \in \Omega} \|\nabla \mathbf{u}(i)\|_1 \tag{17}$$

Penalizing the TV of the displacement field constrains its spatial incoherence without forcing it to be smooth. Detailed properties of TV regularization of displacements were studied by Vishnevskiy et al.[61]

2.C.3. Non-negative Jacobian

The determinants of the Jacobian represent the amount of transformation under a certain deformation. In Ref. [62], Kuang et al. proposed a regularizer that specifically penalizes "folding" or noninvertable deformations, that is, the spatial locations where the Jacobian determinants are less than 0. This regularizer is formulated as:

$$\mathcal{R}_{Jacobian}(\phi; \theta) = \sum_{i \in \Omega} (|det(\mathbf{J}_\phi(i))| - det(\mathbf{J}_\phi(i))). \tag{18}$$

Combined this with diffusion regularization to constrain the overall smoothness results in a regularizer that produces deformations with fewer folded pixels:

$$\mathcal{R}_{reg}(\phi; \theta) = \mathcal{R}_{Diffusion}(\phi; \theta) + \alpha \mathcal{R}_{Jacobian}(\phi; \theta), \tag{19}$$

where α is a weighting parameter.

2.C.4. Gaussian smoothing

A direct way to constrain a deformation field to be smooth is to convolve the displacements with a Gaussian smoothing filter parameterized by its standard deviation, σ [34]:

$$\hat{\mathbf{u}} = G_\sigma * \mathbf{u}, \tag{20}$$

where a larger σ gives a smoother deformation, and vice versa.

2.D. Registration procedure

The overall algorithm for the proposed method is shown in Algorithm 1. In the beginning, we initialized an untrained ConvNet (f_θ) for a given pair of moving and fixed images, I_m and I_f. First, the untrained f_θ produces an initial deformation field, ϕ. Second, we deform the moving image with ϕ (i.e., $I_m \circ \phi$). Then, the registration loss is computed as:

$$\ell = \mathcal{L}_{sim}(I_d, I_f; \theta) + \lambda \mathcal{R}(\phi; \theta), \tag{21}$$

where $\mathcal{L}_{sim}$ represents the similarity measure between I_d and I_f, $\mathcal{R}$ represents the value of the regularizer applied to the deformation field, and λ is a user-defined weighting parameter to control the effectiveness of $\mathcal{R}$. The loss is back-propagated to update the parameters in f_θ. The above procedure is repeated for a pre-specified number of iterations.

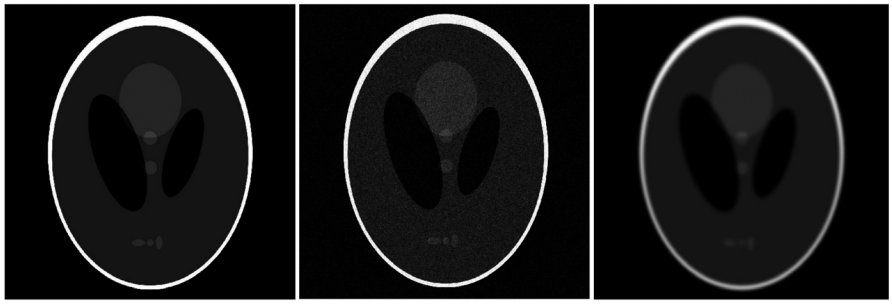

FIG 3. Comparison of "Shepp-Logan" phantom images[59] with different types of distortions. (a) Original Image. (b) Image corrupted by Gaussian noise. SSIM: 0.14, PCC: 0.96. (c) Gaussian blurred image. SSIM: 0.9, PCC: 0.94.

Algorithm 1 ConvNet Registration

```
procedure CNNREG(I_m, I_f)                          ▷ Input I_m and I_f
    f_θ = Initialize(ConvNet)
    while i < iter do                               ▷ For iter number of iterations
        φ = f_θ([I_m, I_f])                         ▷ Predict deformation, φ
        I_d = I_m ∘ φ                               ▷ Deform moving image, I_m
        ℓ = L(I_d, I_f; θ) + λR(φ; θ)               ▷ Compute loss
        f_θ = BackPropagate(f_θ, ℓ)                 ▷ Update ConvNet
        i = i + 1
    return I_d, φ
```

Since no information other than the given image pair is needed, the proposed method requires no prior training and is thus fully and truly unsupervised. The ConvNet is capable of learning an "optimal" deformation from a single pair of images. In the next section, we discuss a series of experiments that were performed to study the effectiveness of the proposed method.

2.E. Experiments

The goal of this work was to create anthropomorphic phantoms by registering the XCAT phantom attenuation map with patient CT images, and then using the deformed of XCAT phantom label map. Nine clinical low-dose whole-body CT patient scans were used in this study; for those, only the torso part of the scans was extracted, resulting in 1153 2D-transaxial slices in total. We resampled the 2D slices into the size of 384 × 384 using bi-linear interpolation to match the size of an XCAT slice. The proposed method was implemented using Keras[63] with a Tensorflow[64] backend on an NVIDIA Quadro P5000 GPU (with 16 GB memory). We applied the proposed method to register a pair of 2D XCAT and CT images slice by slice. The required GPU memory to register a pair of 2D slices was 2693 MB. The patient CT data were obtained from a publicly available dataset (NaF Prostate,[65]) from The Cancer Imaging Archive (TCIA,[66]). The dataset contains 44 baseline and follow-up studies of nine patients, where we randomly extracted one PET/CT scan from the studies of each nine patients to form the dataset used in this work. We first compared the performance produced by the ConvNets with different image similarity metrics. Then, we compared the proposed method to state-of-the-art registration algorithms: the symmetric image normalization method (SyN) [20] from the ANTs package,[67] and a learning-based self-supervised method, VoxelMorph.[22,23]

2.F. Evaluation metrics

We used the following metrics to evaluate the quality of the registration:

1. **Mean Squared Error**: MSE is measured as the mean squared difference between every pixel in I_d and the corresponding pixel in I_f.
2. **Structural Similarity Index**: SSIM indicates the average of perceived change in structural information between the I_d and I_f.
3. **Number of nonpositive Jacobian Determinants**: $\mathbf{J}_\phi$ gives the portion of the registration transformation that resulting from the deformation, ϕ. The quantity measurement of the nonpositive $|\mathbf{J}_\phi|$ indicates the number of pixel transformations that are not inveratible. A small number of nonpositive $|\mathbf{J}_\phi|$ means a smoother deformation, and vice versa.

3. RESULTS

We first compared the effectiveness of different loss functions (without any regularization) in Section 3.A. Then, we showed that some qualitative results generated by different regularizers in Section 3.B. Finally, we comprehensively studied the proposed method and compared it to several state-of-the-art registration methods. The corresponding empirical results are shown in Section 3.C.

3.A. Loss function comparisons

Some examples of the registered XCAT phantom attenuation map images resulting from the six loss functions are shown in Fig. 4. Images (a) and (h) represent the same moving image, and (b) and (i) are the target images from the same CT slice, where the later was blurred by a low-pass Gaussian filter to reduce the effects streaking artifacts. The images in both (c)–(h) and (k)–(p) show results from using, respectively, PCC, SSIM, PCC+SSIM, MSE, CC, and MI. As highlighted in the boxes, MSE, CC, and MI all failed in leading to acceptable results. Whereas, the results produced by SSIM+PCC exhibits fewer image artifacts and provided the best structural match to the target image [as shown in (m)]. Combined with the Gaussian pre-filtering to suppress streaking artifacts in the target image, SSIM+PCC generated the best qualitative results among the loss functions evaluated.

3.B. Regularization comparisons

Figure 5 shows results generated using different regularizers. The three images in the first column are a slice of the XCAT attenuation map (moving image), a slice of patient CT image (fixed image), and a slice of the XCAT label map. The second through last columns show the deformed moving image (first row), transformed labels (second row), and deformed grids (last row) by using no regularization, diffusion regularization, TV regularization, diffusion with non-negative Jacobian regularization, and Gaussian smoothing, respectively. The deformed moving image using no regularization had a virtually identical appearance compared to the fixed image. However, the deformed label maps were unrealistic: in the regions highlighted in rectangles, the bone marrow appeared outside of the cortical bone. Applying regularization to the deformation field helped with this issue, but there was a clear trade-off between the similarity to the fixed image and the smoothness of the deformation field. This trade-off was quantitatively studied, and the results are

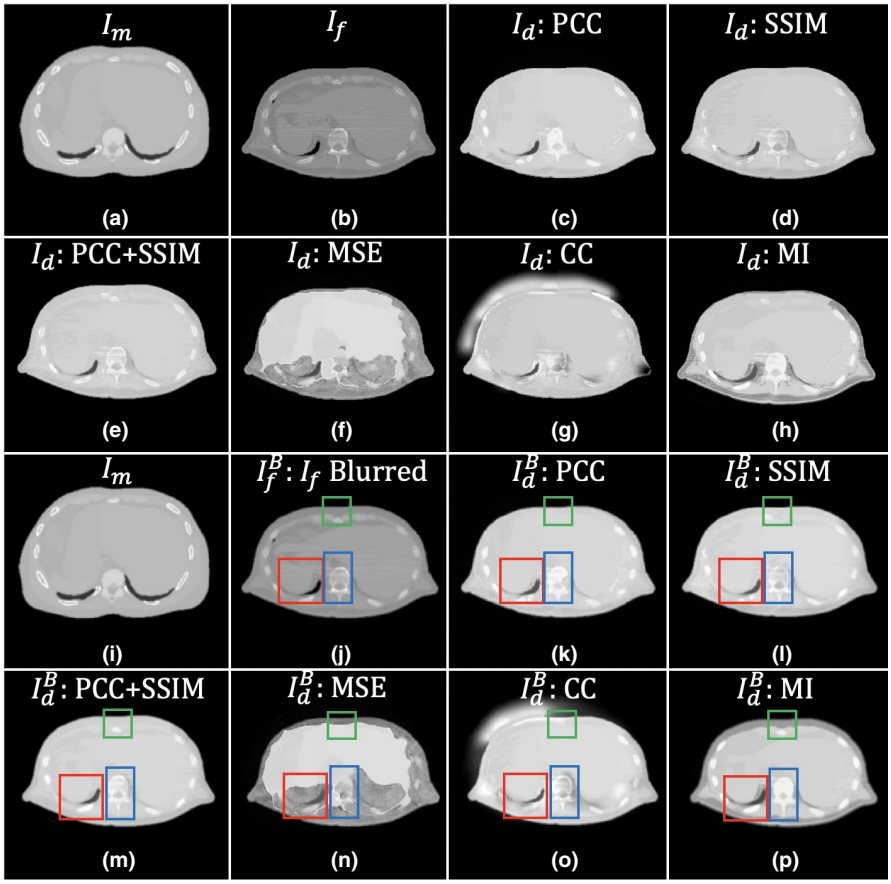

FIG 4. In comparing registered XCAT attenuation map images generated using different loss functions (without any regularization), some differences are highlighted by colored rectangles. The top two rows show the results generated without pre-filtering the fixed image; the bottom two rows show the results generated using the pre-filtered fixed image. The images in (a) and (i) exhibit the same slice of the attenuation map generated from the XCAT phantom, which served as the moving image, I_m; (b) and (j) are the same patient CT images, but prior to use in the registration. Image (j) is (b) blurred with Gaussian filter ($\sigma = 0.8$) to reduce noise and artifacts. The images in (b) and (j) were used as the fixed image, I_f. Images shown in (c)–(h) and (k)–(p) resulted from applying the ConvNet using six different loss functions: (c) and (k) PCC; (d) and (l) SSIM; (e) and (m) PCC + SSIM; (f) and (n) MSE; (g) and (o) CC; and (h) and (p) MI.

discussed in the next section. We specifically quantified the regularity of the field by counting all the pixel deformations for which the transformation was not diffeomorphic (i.e., folding or $|\mathbf{J}_\phi| \leq 0$).[22,68]

3.C. Registration performance comparisons

In this subsection, we compared the proposed method with the SyN[20] and VoxelMorph[22,23] algorithms. Since VoxelMorph requires prior training, it was evaluated using a leave-one-out method: images from eight patients were used for training ($\sim$1024 2D slices), and images from one patient were treated as the test set ($\sim$128 2D slices). Then, we altered patients whose image was used for training and testing, rendering nine possible combinations of the patient images. Figures 6 and 7 show comparisons of the proposed method with different regularizations, the SyN, and the VoxelMorph methods, respectively. The first column shows the moving and fixed images. The second to the last column shows the deformed XCAT images (upper row) and deformed labels (lower row), respectively. Based on these qualitative results, the proposed method provides a more detailed deformation than other methods, especially in terms of the anatomy of the bone structures and soft tissues. Since a gold-standard bone segmentation was not available for the NaF Prostate dataset,[65] the registration performance was evaluated quantitatively based on MSE and SSIM between $I_m \circ \phi$ and I_f. The results are shown in Table II. Without any regularization of the deformation field, the proposed method gave a mean SSIM of 0.955 and a mean MSE of 37.340, which outperformed the SyN and VoxelMorph by a significant margin (with $P < 0.0001$ from the paired t-test).

The plots in Fig. 8 exhibit the impact of different regularization parameters on the SSIM, MSE, and the number of folded pixels. A decreasing trend in registration accuracy and the number of folded pixels was generally seen with increasing weighting parameter values (λ and σ) for regularizers. Among the different regularization methods, the diffusion regularizer (column 1 in Fig. 8) with $\lambda = 1$ yielded the best balance between registration accuracy and the number of folded pixels. Overall, the method achieved comparable performances to VoxelMorph in terms of deformation regularity

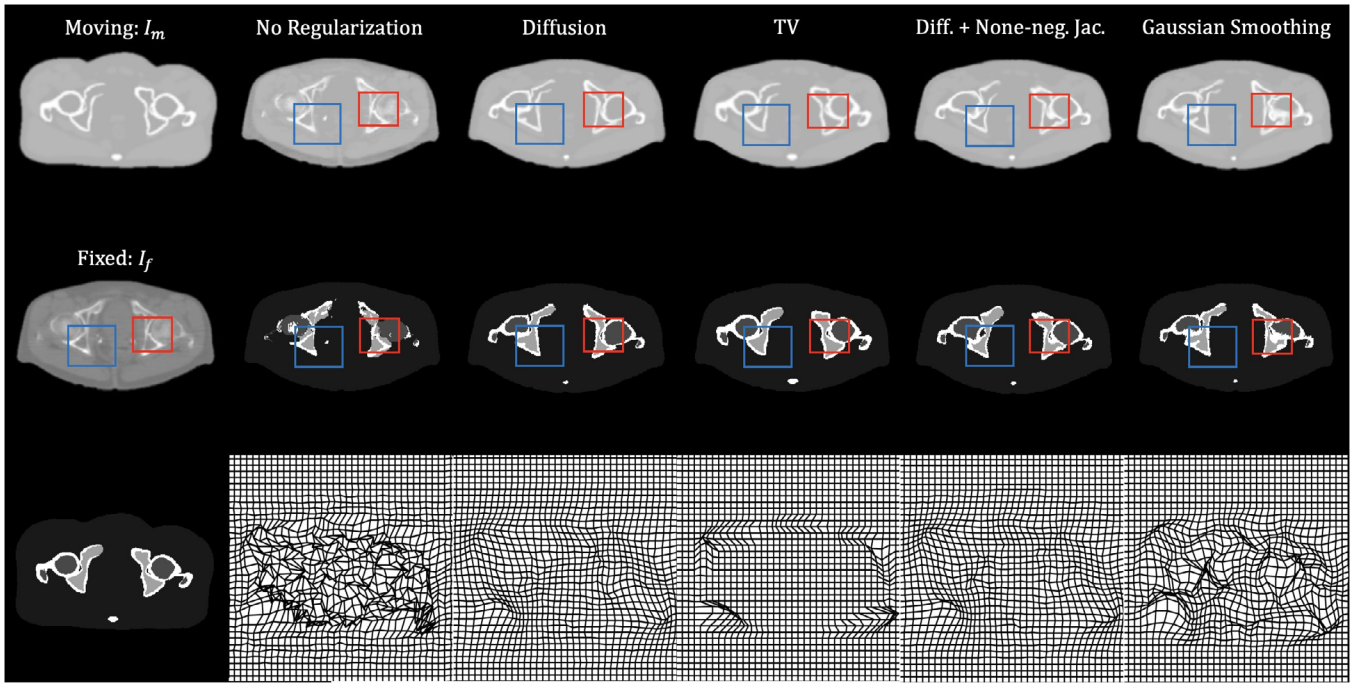

FIG 5. Example results from different regularization techniques. The three figures in the first column represent moving image, fixed image, and the corresponding label map, respectively. The second to last column shows deformed images: the first row shows the deformed XCAT phantom, the second row shows the deformed label map, and the last row shows the deformed grid.

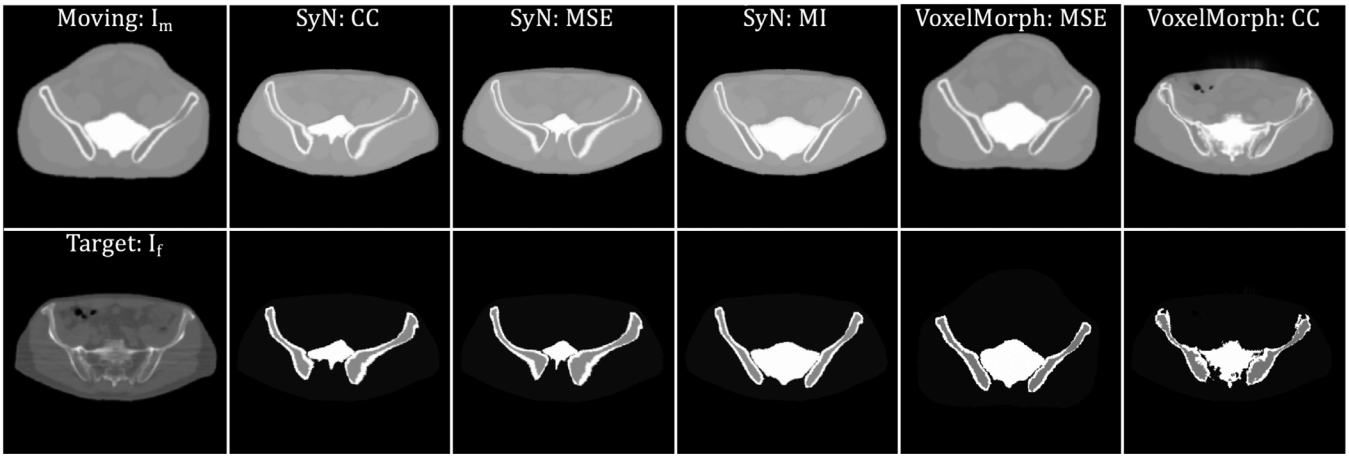

FIG 6. Example results generated by two baseline methods, SyN and VoxelMorph. The 1st column: moving (the XCAT attenuation map) and target image (patient CT). For the second to the last column, the first row corresponds to the deformed moving images, and the second row shows the deformed label map. The 2nd through 4th columns show results from SyN using CC, MSE, and MI. The 5th column through the last column shows results from VoxelMorph using MSE and CC.

(as measured by the number of pixels where there was folding) while providing better registration accuracy.

4. DISCUSSION

4.A. Fully unsupervised ConvNet

We have developed a fully unsupervised U-Net-based registration framework for generating highly detailed, anatomically realistic phantoms. The proposed method works on an image pair consisting of the XCAT attenuation map and patient CT. Thus, it does not require prior training using a large dataset. This makes the proposed method different from the previously proposed self-supervised ConvNet-based registration algorithms, such as VoxelMorph,[22,30,33] DIRNet,[32] or DLIR,[31] which require a training stage. However, while the proposed method does not demand training data, there is a trade-off: like traditional techniques,[18,20] the proposed method minimizes a loss function iteratively for the given inputs, which leads to the increased computational time and complexity. The runtime for performing registration on a pair of 3D volumes (with size 192 × 192 × 128) is roughly an

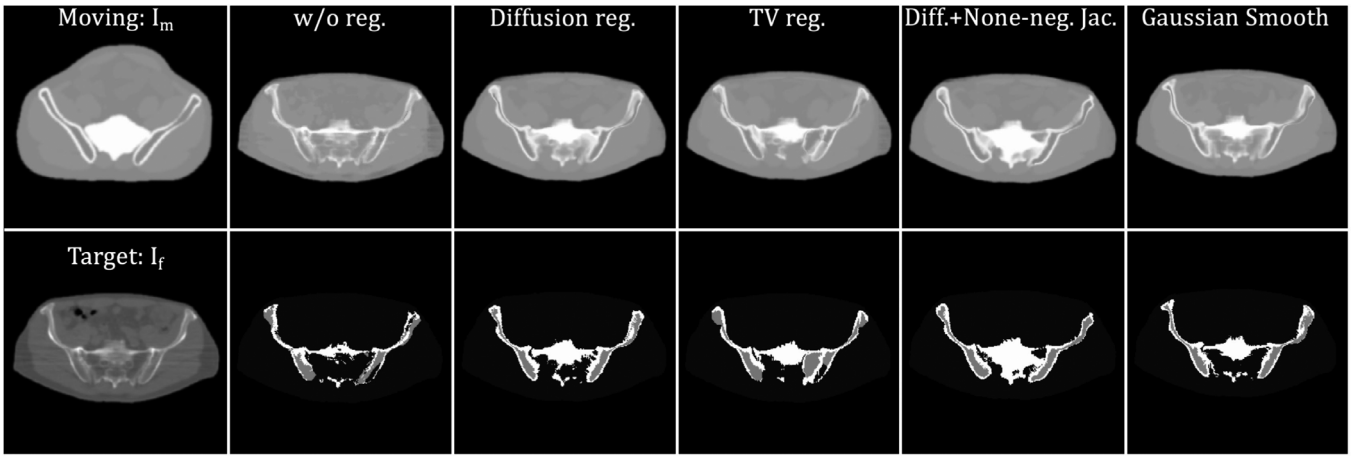

FIG 7. Example results generated by the proposed method with different regularization techniques. The 1st column exhibits moving (XCAT attenuation map) and target image (patient CT). The 2nd column to the last column display the deformed results. These correspond, respectively, to no regularization, diffusion regularization, TV norm, diffusion with none-negative Jacobian regularization, and Gaussian smoothing.

TABLE II. Comparison of SSIM, MSE, and the number and percentage of pixel locations with nonpositive Jacobian determinant among the proposed method (UnsupConvNet), SyN, and VoxelMorph

Method	SSIM	MSE	$\lvert\mathbf{J}_\phi\rvert \leq 0$ (counts)	% of $\lvert\mathbf{J}_\phi\rvert \leq 0$ (%)
Affine only	0.83 ± 0.008	69.2 ± 2.7	–	–
VoxelMorph (MSE)[22]	0.88 ± 0.003	47.0 ± 2.4	685 ± 185	0.5 ± 0.1
VoxelMorph (CC)[22]	0.92 ± 0.006	43.5 ± 4.8	2754 ± 370	1.9 ± 0.3
SyN (MSE)[20]	0.88 ± 0.011	52.0 ± 4.1	–	–
SyN (MI)[20]	0.88 ± 0.011	55.1 ± 4.0	–	–
SyN (CC)[20]	0.89 ± 0.011	52.8 ± 4.1	–	–
UnsupConvNet (w/o regularization)	**0.96** ± 0.007	**37.3** ± 5.1	21082 ± 3938	14.3 ± 2.7
UnsupConvNet (w/ diffusion regularization)	0.93 ± 0.008	42.6 ± 5.4	1202 ± 225	0.8 ± 0.1
UnsupConvNet (w/ diff. + None-neg. Jac. reg.)	0.92 ± 0.009	44.8 ± 4.9	518 ± 74	0.4 ± 0.1
UnsupConvNet (w/ TV regularization)	0.87 ± 0.030	54.7 ± 9.3	659 ± 459	0.4 ± 0.3
UnsupConvNet (w/ Gaussian filtering)	<u>0.94</u> ± 0.008	<u>41.5</u> ± 5.4	8500 ± 1829	5.7 ± 1.3

The top three results in SSIM and MSE are shown in **bold**, <u>*underline*</u>, and *italics*, respectively. Evaluations were done on 2D images with size 384 × 384.

hour on an NVIDIA Quadro P5000 GPU. The computational time may be reduced either by using a smaller ConvNet (i.e., few trainable parameters) or using a faster GPU.

4.B. Loss function choice

We demonstrated in the previous sections that the proposed SSIM-based loss function yielded the best qualitative performance and thus was more suitable for XCAT-to-CT registration. While other loss functions, MSE, CC, or MI, are commonly used in other image registration tasks, they all performed poorly in this task. Specifically, the failure of MSE is likely due to the lack of spatial information, and it is sensitive to linear transformations of the mean intensity values (as shown in the third to last column in Fig. 4). Another commonly used metric, CC, also produced sub-optimal results that exhibited image artifacts (as shown in the seventh column in Fig. 4). We also showed that using PCC or SSIM alone did not produce good results: while PCC loss was robust to image artifacts, it produced "discretized" results around the spine (see the regions highlighted in rectangles in (c) and (j) in Fig. 4). On the other hand, the SSIM loss function produced an image that reproduced even the noise and artifacts in the target image [as shown in (d) in Fig. 4].

4.C. Regularization choice

Despite the fact that the proposed method without any regularization generated a deformed image that was almost identical in appearance to the fixed image, the warped label maps were not realistic (see the second column in Fig. 7). This was mainly caused by the nonsmooth and noninvertible deformation field that produced a large number of folded pixels. Adding a regularizer to the loss function enforced the smoothness in the deformation field, and thus produced more realistic warped label maps. However, as shown in Figs. 8 and 9, a decreasing trend in the similarity measure to the fixed image was observed with improved deformation regularity (i.e., a decrease in the number of nonpositive Jacobian determinants). Images in (b), (f), and (j) in the right panel of Fig. 9

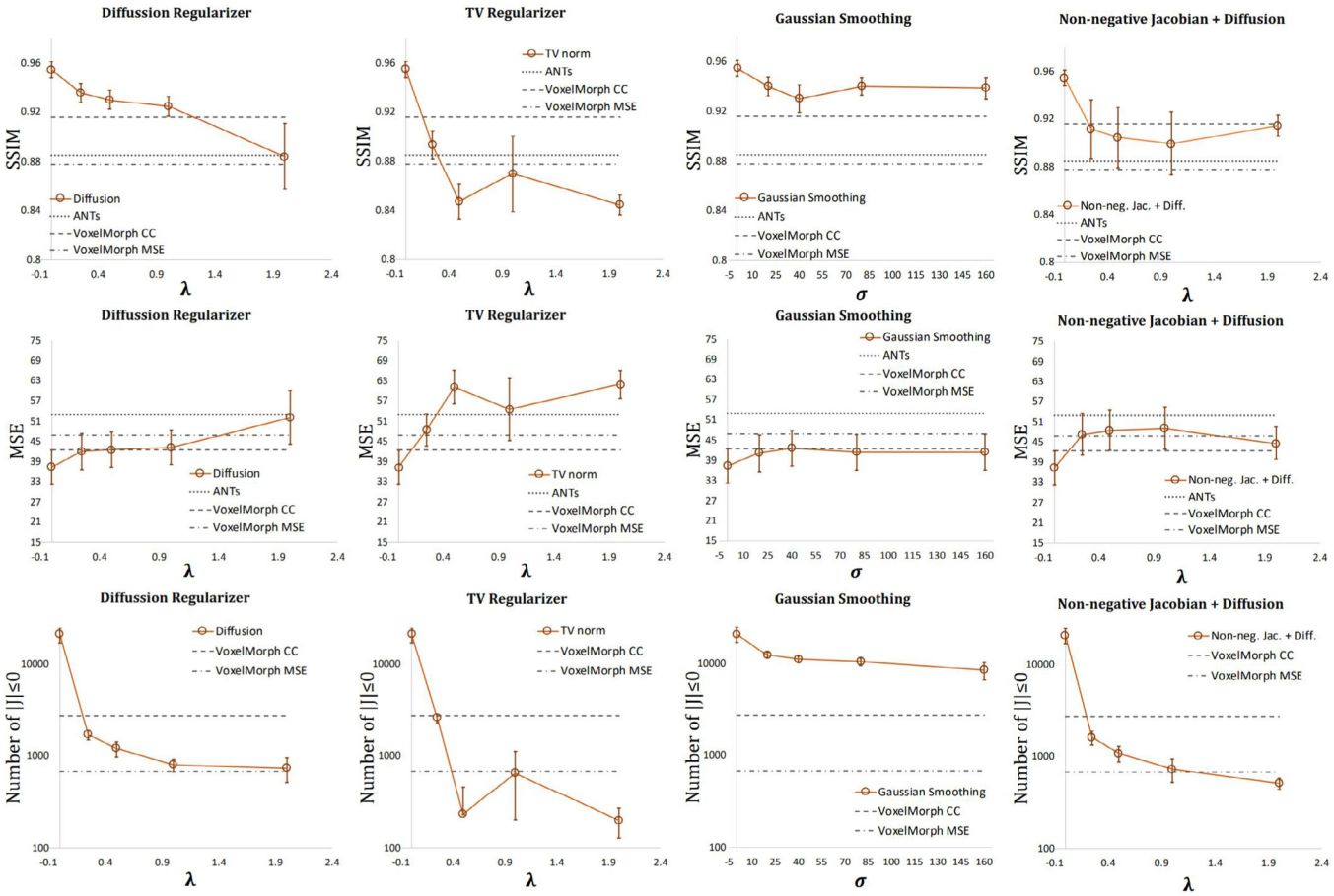

FIG 8. The effect of different regularization parameters on the registration accuracy (SSIM and MSE), and deformation regularity (number of folded pixels, i.e., where the determinant of Jacobian was ≤0). First to last rows indicate the performances in SSIM, MSE, and number of folded pixels, respectively. The columns, from left to right, are the results generated using the diffusion regularizer, TV regularizer, Gaussian smoothing, and non-negative Jacobian + diffusion regularizer, respectively.

show the deformed XCAT attenuation map using different weights, λ, of the diffusion regularization. For $\lambda = 0$ (no regularization), the soft-tissue components (e.g., liver) in (b) showed good alignment with those in the target image ((c) in the left panel of Fig. 9). When λ was increased (as shown in the second and the last row in the right panel of Fig. 9), the smoothness of the resulting deformation fields gradually improved (as shown in (d), (h), and (I)). However, the mismatch between the soft-tissue components in the deformed and the fixed image began to appear, while regions with higher contrast (e.g., bone and outer surface) retained the good alignment. The experimental results in Fig. 8 show that the diffusion regularization worked best among the regularizers investigated. It sacrificed a modest amount of image similarity (0.93 in mean SSIM and 42.6 in mean MSE) in exchange for smoother deformation fields with a smaller number of nonpositive Jacobian determinants (< 0.8% of pixel transformations were noninvertible). It should be noted that the regularizers used in this work only consider global smoothing, thus we did not allow motion discontinuities between organs. Regularization methods that preserve sliding-motion[23–25] could be included in the future to address the problem of sliding organs.

4.D. Example application: SPECT simulations

In this section, we demonstrate a 3D application of the proposed method to generate realistic medical image simulations. We employed the proposed registration method to map the 3D volume of the XCAT attenuation map to a 3D patient CT scan acquired from a clinical bone SPECT/CT acquisition. We then generated a realistic simulated SPECT image on the basis of the resulting deformed XCAT attenuation map as described below. The patient scan was acquired using a clinical whole-body SPECT/CT scan protocol; the CT scan was a low-dose one designed to provide an attenuation map. Both SPECT and CT images were reconstructed using scanner software. Two sample coronal slices of the patient scans are shown in the second and the third columns of the left panel in Fig. 10; the second and third columns show the CT and SPECT images, respectively. We used the proposed method with a diffusion regularizer ($\lambda = 1$) to perform the 3D registration. We resampled the image volumes to a size of $192 \times 192 \times 128$ in order to fit into the GPU memory of 16 GB. The memory required to perform this 3D registration was 15 GB. The resulting deformed XCAT attenuation map and the corresponding SPECT simulation are shown in the

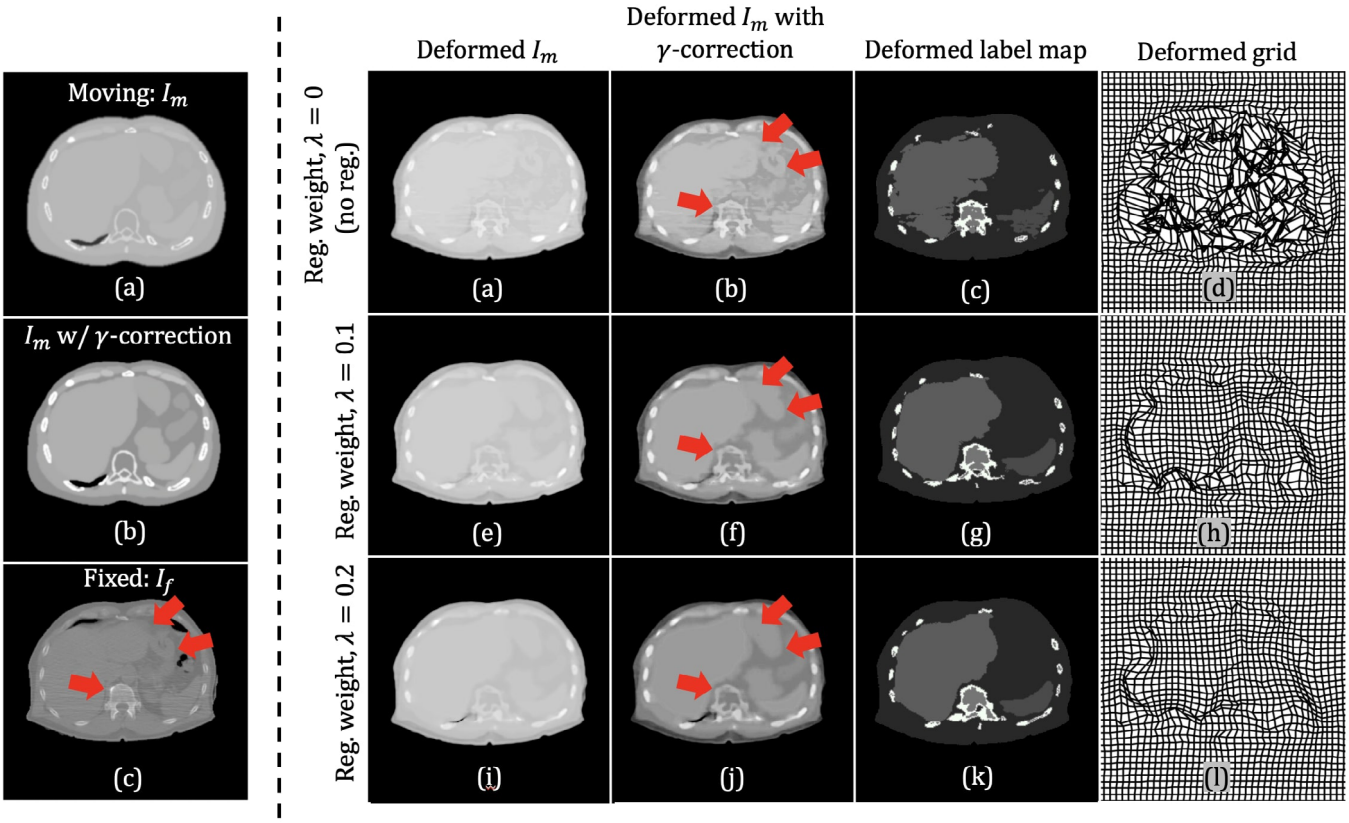

FIG 9. Visual effects of the diffusion regularization on the deformed image. Left panel: Images in (a), (b), and (c) corresponds to, respectively, moving image, moving image with Gamma-correction (γ = 1.8), and fixed image. Right panel: Each row represents the results obtained using different regularization weights. The first through last columns, from left to right, represent deformed moving image, deformed moving image with Gamma-correction, deformed label map, and the corresponding deformation field.

fourth and the fifth columns, respectively. We then added several artificial lesions to the phantoms, and two example slices of the resulting SPECT simulation are shown in the last column. SPECT projections were simulated by an analytic projection algorithm that realistically models attenuation, scatter, and the spatially varying collimator-detector response.[47,48] We computed attenuation values on the basis of the material compositions and the attenuation coefficients of the constituents at 140 keV, the photon energy of Technetium-99m. We inserted several artificial sclerotic bone lesions to random bone regions with increased attenuation coefficient and radio-pharmaceutical uptake. The cortical and trabecular bones had an uptake of, respectively, 12.6 and 23.2 times that of the soft-tissue background, and bone lesions had an uptake of 3.5–4.5 times that of normal bone. These scale factors were computed based on the patient SPECT scan. SPECT simulations were reconstructed using a the ordered subsets-expectation maximization algorithm (OS-EM)[69,70] using 5

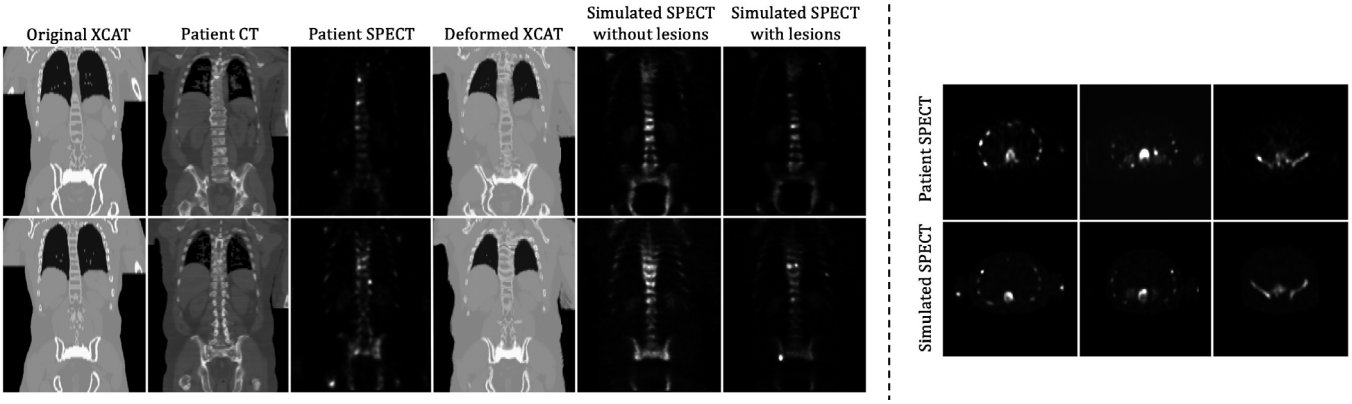

FIG 10. Visualization of the deformed XCAT attenuation map and SPECT simulations. Left: Two coronal slices of the original XCAT, patient CT, and SPECT scans, the registered XCAT, the SPECT simulation, and the simulated SPECT with lesion added. Right: Comparison of transverse slices between patient SPECT scan and SPECT simulation.

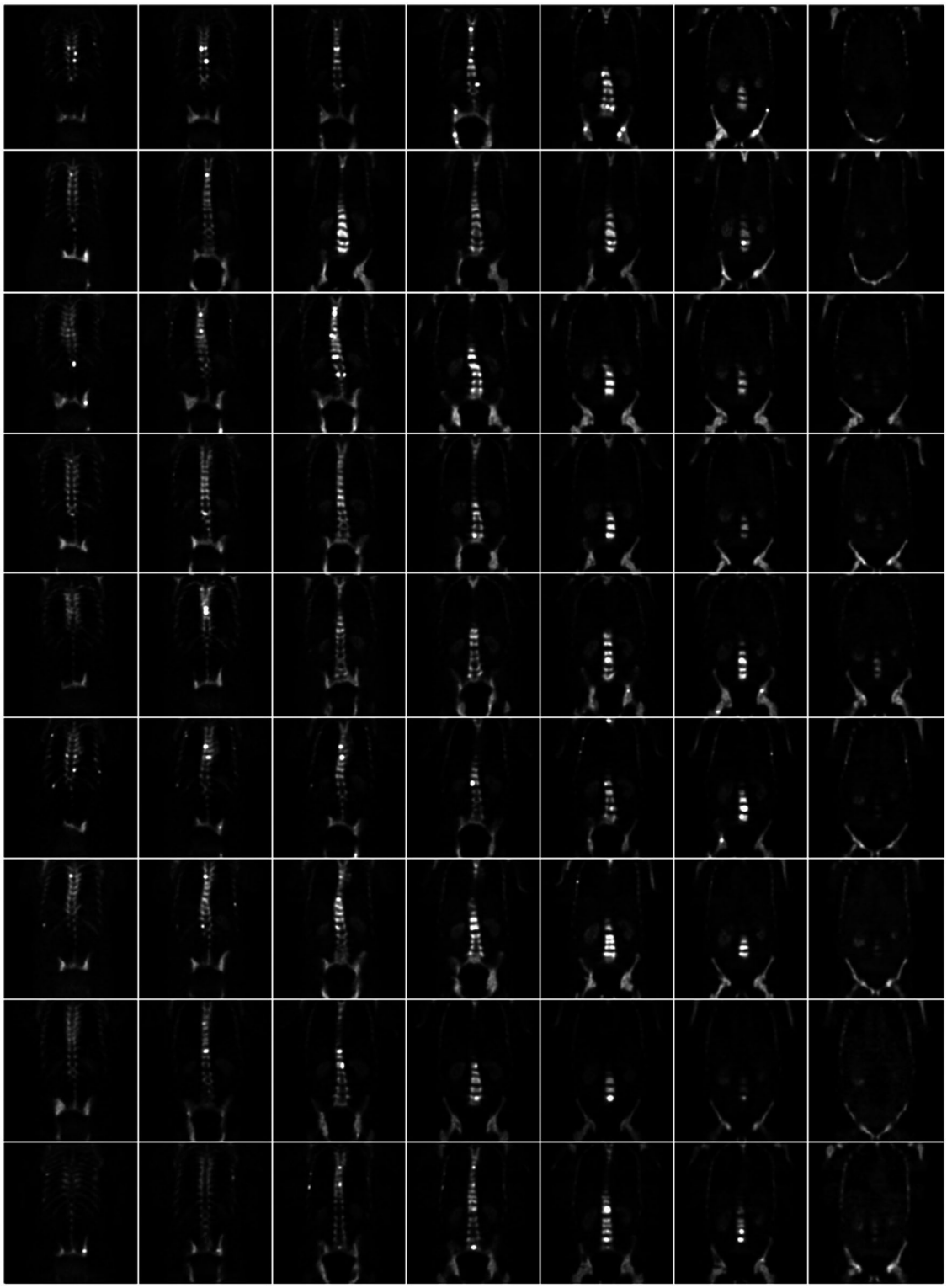

FIG 11. Example coronal slices of the SPECT simulations generated by the proposed method, where each row represents a different patient.

iterations and 10 subsets. The figure on the right panel in Fig. 10 shows three transverse slices of the patient SPECT (top row) and the simulated SPECT (bottom row). Because the deformed XCAT phantom was able to successfully capture the anatomical structures in the patient scan, when combined with realistic physics models of image formation processing, the resulting SPECT simulation appears quite realistic compared to the patient SPECT scan. In addition, the relationship between the generated phantom activity distribution and the projection data are quantitatively realistic

because of the method used to generate the projections. Figure 11 shows some additional SPECT simulations (with lesions added to various bone locations) generated using the proposed method with the nine clinical CT scans from the TCIA dataset. In that figure, each row represents a different patient.

5. CONCLUSION

We have developed a method to create anthropomorphic phantoms using an unsupervised, ConvNet-based, end-to-end registration technique. Unlike existing deep-learning-based registration methods, the proposed method requires no prior training. While classical registration methods also do not require training data, they work in a lower-dimensional parameter space; the proposed approach operates directly in the high-dimensional parameter space common to deep-learning-based methods but without any prior training. Compared to the commonly used loss functions in ConvNet-based registration, we demonstrated that the registration performance can be improved by use of the combination of SSIM and PCC as a loss function for updating the parameters in the ConvNet. The proposed method was evaluated for the application of registering the XCAT attenuation map with real patient CT scans as part of a process to simulate realistic nuclear medicine images. We compared the registration performance of the proposed technique in terms of SSIM and MSE to conventional state-of-the-art image registration methods. Both quantitative and qualitative analyses indicated that the proposed method provided the best registration results. We also demonstrated that the proposed method, combined with accurate simulation tools, provided a highly realistic anthropomorphic medical image with known truth that faithfully represents the image formation process and qualitatively matches the appearance of a real patient image.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Cancer Institute, U01-CA140204. The views expressed in written materials or publications and by speakers and moderators do not necessarily reflect the official policies of the NIH; nor does mention by trade names, commercial practices, or organizations imply endorsement by the U.S. Government. We would like to show our gratitude to Dr. Daniel Tward and Shuwen Wei for sharing pearls of wisdom with us during the course of this research.

[a)][a]Author to whom correspondence should be addressed. Electronic mail: jchen245@jhmi.edu.

REFERENCES

1. Christoffersen CP, Hansen D, Poulsen P, Sorensen TS. Registration-based reconstruction of four-dimensional cone beam computed tomography. *IEEE Trans Med Imaging*. 2013;32:2064–2077.
2. Zhang Y, Ma J, Iyengar P, Zhong Y, Wang J. A new CT reconstruction technique using adaptive deformation recovery and intensity correction (ADRIC). *Med Phys*. 2017;44:2223–2241.
3. Chen J, Jha AK, Frey EC, Incorporating CT prior information in the robust fuzzy C-means algorithm for QSPECT image segmentation. In: Medical Imaging 2019: Image Processing, edited by E. D. Angelini and B. A. Landman, page 66, SPIE; 2019.
4. Abdoli M, Dierckx RAJO, Zaidi H. Contourlet-based active contour model for PET image segmentation. *Med Phys*. 2013;40:082507.
5. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT phantom for multimodality imaging research. *Med Phys*. 2010;37:4902–4915.
6. He B, Wahl RL, Du Y, et al. Comparison of residence time estimation methods for radioimmunotherapy dosimetry and treatment planning-Monte Carlo simulation studies. *IEEE Trans Med Imaging*. 2008;27:521–530.
7. Ghaly M, Du Y, Links JM, et al. Collimator optimization in myocardial perfusion SPECT using the ideal observer and realistic background variability for lesiondetection and joint detection and localization tasks. *Phys Med Biol*. 2016;61:2048–2066.
8. Li Y, O'Reilly S, Plyku D, et al. A projection image database to investigate factors affecting image quality in weight-based dosing: Application to pediatric renal SPECT. *Phys Med Biol*. 2018;63:145004.
9. Nakada K, Taguchi K, Fung GSK, Amaya K. Joint estimation of tissue types and linear attenuation coefficients for photon counting CT. *Med Phys*. 2015;42:5329–5341.
10. Lee O, Kappler S, Polster C, Taguchi K. Estimation of basis line-integrals in a spectral distortion-modeled photon counting detector using low-order polynomial approximation of x-ray transmittance. *IEEE Trans Med Imaging*. 2017;36:560–573.
11. Lee O, Kappler S, Polster C, Taguchi K. Estimation of basis line integrals in a spectral distortion-modeled photon counting detector using low-rank approximation-based x-ray transmittance modeling: k-edge imaging application. *IEEE Trans Med Imaging*. 2017;36:2389–2403.
12. Kidoh M, Shen Z, Suzuki Y, et al. False dyssynchrony: problem with image-based cardiac functional analysis using x-ray computed tomography. In: Medical Imaging 2017: Physics of Medical Imaging, edited by T. G. Flohr, J. Y. Lo, and T. G. Schmidt, volume 10132, International Society for Optics and Photonics, SPIE; 2017:449–455.
13. Gong K, Guan J, Liu C-C, Qi J. PET image denoising using a deep neural network through fine tuning. *IEEE Trans Radiat Plasma Med Sci*. 2018;3:153–161.
14. Gong K, Guan J, Kim K, et al. Iterative PET image reconstruction using convolutional neural network representation. *IEEE Trans Med Imaging*. 2019;38:675–685.
15. Lee H, Lee J, Cho S. View-interpolation of sparsely sampled sinogram using convolutional neural network. In: Medical Imaging 2017: Image Processing, volume 10133, SPIE; 2017:1013328.
16. Segars WP, Bond J, Frush J, et al. Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization. *Med Phys*. 2013;40:043701.
17. Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: A survey. *IEEE Trans Med Imaging*. 2013;32:1153–1190.
18. Beg MF, Miller MI, Trouvé A, Younes L. Computing large deformation metric mappings via geodesic ows of diffeomorphisms. *Int J Comput Vision*. 2005;61:139–157.
19. Wolberg G, Zokai S. Robust image registration using log-polar transform. In: Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101), volume 1; 2000:493–496.
20. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008;12:26–41.
21. Viola P, Wells WM. Alignment by maximization of mutual information. *Int J Comput Vision*. 1997;24:137–154.
22. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans Med Imaging*. 2019;38:1788–1800.
23. Pace DF, Enquobahrie A, Yang H, Aylward SR, Niethammer M. Deformable image registration of sliding organs using anisotropic diffusive regularization. in 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE; 2011:407–413.

24. Pace DF, Aylward SR, Niethammer M. A locally adaptive regularization 620621 based on anisotropic diffusion for deformable image registration of sliding organs. *IEEE Trans Med Imaging*. 2013;32:2114–2126.
25. Papiez BW, Heinrich MP, Fehrenbach J, Risser L, Schnabel JA. An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. *Med Image Anal*. 2014;18:1299–1311.
26. Rueckert D, Aljabar P, Heckemann RA, Hajnal JV, Hammers A. Diffeomorphic registration using B-splines, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 4191 LNCS, Springer Verlag; 2006:702–709.
27. Cao X, Yang J, Zhang J, et al. Deformable image registration based on similarity-steered CNN regression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer; 2017:300–308.
28. Krebs J, Mansi T, Delingette H, et al. Robust non-rigid registration through agent-based action learning. In: International Conference on Medical Image Computing and Computer Assisted Intervention, Springer; 2017:344–352.
29. Sokooti H, De Vos B, Berendsen F, Lelieveldt BP, I. Išgum, and M. Staring. Nonrigid image registration using multi-scale 3D convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2017:232–239.
30. Dalca AV, Balakrishnan G, Guttag J, Sabuncu MR. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Med Image Anal*. 2019;57:226–236.
31. de Vos BD, Berendsen FF, Viergever MA, Staring M, Išsgum I. End-to-end unsupervised deformable image registration with a convolutional neural network. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10553 LNCS; 2017:204–212.
32. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Ivsgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal*. 2019;52:128–143.
33. Balakrishnan G, Zhao A, Sabuncu MR, Dalca AV, Guttag J. An Unsupervised Learning Model for Deformable Medical Image Registration. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2018:9252–9260.
34. Krebs J, Delingette H, Mailhe B, Ayache N, Mansi T. Learning a probabilistic model for diffeomorphic registration. *IEEE Trans Med Imaging*. 2019;38:2165–2176.
35. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning requires rethinking generalization, arXiv; 2016.
36. Su J, Vargas DV, Sakurai K. One pixel attack for fooling deep neural networks. *IEEE Trans Evol Comput*. 2019;23:828–841.
37. Moosavi-Dezfooli SM, Fawzi A, Frossard P. DeepFool: A Simple and Accurate Method to Fool Deep Neural Networks. IEEE Computer Society, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 2016-December; 2016:2574-2582–2582.
38. Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples, arXiv; 2014.
39. Papernot N, McDaniel P, Jha S, et al. The limitations of deep learning in adversarial settings. In: Proceedings - 2016 IEEE European Symposium on Security and Privacy, EURO S and P 2016, Institute of Electrical and Electronics Engineers Inc.; 2016:372–387.
40. Szegedy C. Intriguing properties of neural networks, arXiv; 2013.
41. Lempitsky V, Vedaldi A, Ulyanov D. Deep Image Prior. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society; 2018:9446–9454.
42. Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks. In: Advances in neural information processing systems; 2015:2017–2025.
43. Segars WP, Mahesh M, Beck TJ, Frey EC, Tsui BM. Realistic CT simulation using the 4D XCAT phantom. *Med Phys*. 2008;35:3800–3808.
44. Ljungberg M, Strand S, King M. The SIMIND Monte Carlo program, Monte Carlo calculation in nuclear medicine: Applications in diagnostic imaging; 2012:145–163.
45. Jan S, Benoit D, Becheva E, et al. GATE V6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy. *Phys Med Biol*. 2011;56:881.
46. Reilhac A, Lartizien C, Costes N, et al. PET-SORTEO: a Monte Carlo-based simulator with high count rate capabilities. *IEEE Trans Nucl Sci*. 2004;51:46–52.
47. Frey EC, Tsui BMW. A practical method for incorporating scatter in a projector-backprojector for accurate scatter compensation in SPECT. *IEEE Trans Nucl Sci*. 1993;40:1107–1116.
48. Kadrmas DJ, Frey EC, Tsui BMW. An SVD investigation of modeling scatter in multiple energy windows for improved SPECT images. *IEEE Trans Nucl Sci*. 1996;43:2275–2284.
49. Du Y, Frey EC, Wang WT, et al. Combination of MCNP and SimSET for Monte Carlo simulation of SPECT with medium- and high-energy photons. *IEEE Trans Nucl Sci*. 2002;49:668–674.
50. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv; 2015.
51. Zhu W, Myronenko A,Xu Z, et al. NeurReg: Neural Registration and Its Application to Image Segmentation, arXiv; 2019.
52. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging*. 1997;16:187–198.
53. Vajda I. *Theory of Statistical Inference, in Theory of Statistical Inference*. Dordrecht: Kluwer Academic Publishers; 1989.
54. Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W, PET-CT image registration in the chest using free-form deformations. IEEE transactions on medical imaging 22; 2003:120–128.
55. Wang Y-P, Lee SL. Scale-space derived from B-splines. *IEEE Trans Pattern Anal Mach Intell*. 1998;20:1040–1055.
56. Unser M, Aldroubi A, Eden M. On the asymptotic convergence of B-spline wavelets to Gabor functions. *IEEE Trans Inf Theory*. 1992;38:864–872.
57. Saad ZS, Glen DR, Chen G, et al. A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *NeuroImage*. 2009;44:839–848.
58. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600–612.
59. Shepp LA, Logan BF. Fourier reconstruction of a head section. *IEEE Trans Nucl Sci*. 1974;NS-21:21–43.
60. Li H, Fan Y. Non-rigid image registration using self-supervised fully convolutional networks without training data. Proceedings - International Symposium on Biomedical Imaging 2018-April, 2018:1075–1078.
61. Vishnevskiy V, Gass T, Szekely G, Tanner C, Goksel O. Isotropic total variation regularization of displacements in parametric image registration. *IEEE Trans Med Imaging*. 2017;36:385–395.
62. Kuang D, Schmah T. FAIM – A ConvNet Method for Unsupervised 3D Medical Image Registration, arXiv; 2019:646–654.
63. Chollet F. Keras. https://keras:io; 2015.
64. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, Software available from tensor ow.org; 2015.
65. Kurdziel KA, The kinetics and reproducibility of 18F-sodiumuoride for oncology using current PET camera technology. *J Nucl Med*. 2012;53:1175–1184.
66. 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.
67. Avants BB, Tustison NJ, Song G, Gee JC, ANTS: open-source tools for normalization and neuroanatomy. *IEEE Trans Biomed Eng*. 2009;10:1–11.
68. Ashburner J. A fast diffeomorphic image registration algorithm. *NeuroImage*. 2007;38:113.
69. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med Imaging*. 1994;13:601–609.
70. He B, Du Y, Song X, Segars WP, Frey EC. A Monte Carlo and physical phantom evaluation of quantitative In-111 SPECT. *Phys Med Biol*. 2005;50:4169–4185.