A new deep convolutional neural network design with efficient learning capability: Application to CT image synthesis from MRI

Abass Bahrami
Faculty of Physics, University of Isfahan, Isfahan, Iran

Alireza Karimian[a)]
Department of Biomedical Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran

Emad Fatemizadeh
School of Electrical Engineering, Sharif University of Technology, Tehran, Iran

Hossein Arabi
Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva CH-1211, Switzerland

Habib Zaidi
Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva CH-1211, Switzerland
Geneva University Neurocenter, Geneva University, Geneva 1205, Switzerland
Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands
Department of Nuclear Medicine, University of Southern Denmark, Odense DK-500, Denmark

(Received 31 March 2020; revised 3 July 2020; accepted for publication 17 July 2020; published xx xxxx xxxx)

Purpose: Despite the proven utility of multiparametric magnetic resonance imaging (MRI) in radiation therapy, MRI-guided radiation treatment planning is limited by the fact that MRI does not directly provide the electron density map required for absorbed dose calculation. In this work, a new deep convolutional neural network model with efficient learning capability, suitable for applications where the number of training subjects is limited, is proposed to generate accurate synthetic computed tomography (sCT) images from MRI.

Methods: This efficient convolutional neural network (eCNN) is built upon a combination of the SegNet architecture (a 13-layer encoder-decoder structure similar to the U-Net network) without softmax layers and the residual network. Moreover, maxpooling indices and high resolution features from the encoding network were incorporated into the corresponding decoding layers. A dataset containing 15 co-registered MRI-CT pairs of male pelvis (1861 two-dimensional images) were used for training and evaluation of MRI to CT synthesis process using a fivefold cross-validation scheme. The performance of the eCNN model was compared to an atlas-based sCT generation technique as well as the original U-Net model considering CT images as reference. The mean error (ME), mean absolute error (MAE), Pearson correlation coefficient (PCC), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) metrics were calculated between sCT and ground truth CT images.

Results: The eCNN model exhibited effective learning capability using only 12 training subjects. The model achieved a ME and MAE of 2.8 ± 10.3 and 30.0 ± 10.4 HU, respectively, which is substantially lower than values achieved by the atlas-based (-0.8 ± 35.4 and 64.6 ± 21.2) and U-Net (7.4 ± 11.9 and 44.0 ± 8.8) methods, respectively.

Conclusion: The proposed eCNN model exhibited efficient convergence rate with a low number of training subjects, while providing accurate synthetic CT images. The eCNN model outperformed the original U-Net model and showed superior performance to the atlas-based technique. © *2020 American Association of Physicists in Medicine* [https://doi.org/10.1002/mp.14418]

Key words: ATLAS, deep learning, machine learning, MRI, pseudo-CT generation

1. INTRODUCTION

Computed tomography (CT) imaging is commonly employed in external radiation therapy for delineation of treatment volumes and dose calculation taking advantage of the direct availability of electron density map. Moreover, CT adequately depicts bony structures, most often used for patient positioning and definition of anatomical landmarks. Nevertheless, CT images suffer from poor soft-tissue contrast, hampering accurate delineation of structures and tissue/organ discrimination. Conversely, magnetic resonance imaging (MRI) provides high soft-tissue contrast, thus allowing excellent tissue discrimination and is a multiparametric imaging modality by nature. In addition to superior soft-tissue visualization, MRI does not use ionizing radiation as opposed to CT, thus making online radiation planning adjustment and tumor monitoring possible with no extra exposure. These features of MRI are so promising that radiation treatment planning is being

revisited to be based solely on MRI.[1,2] Besides, the combination of MRI with other modalities such as positron emission tomography (PET) (PET/MRI) is gaining momentum owing to above-mentioned benefits of MRI.[3,4]

However, eliminating CT from radiation treatment planning or replacing PET/CT with PET/MRI is not trivial and could be challenging since electron density maps are not readily provided by MRI. To address this issue, various strategies were proposed the literature to derive electron density maps from MRI rely on three generic approaches.[5–7] Tissue segmentation-based techniques employ image segmentation algorithms to delineate a number of tissue classes from MRI. This is followed by assignment of a single predefined density value to each tissue class. Organ/tissue segmentation is commonly performed to identify soft-tissue, fat, air, lung, and in some cases bones from MRI.[8–10] Delineation of bony structures is the major challenge of this approach since conventional MR sequences are not capable of discriminating bone from air. To this end, specialized MR sequences, including ultra-short echo time (UTE) and zero-echo-time were devised to pinpoint bone signals. However, these approaches suffer from long acquisition times, low signal-to-noise ratio and the fact that bulk segmentation of tissues, does not take into account the natural heterogeneity of bony structures, namely cortical and spongy bones.[11–14]

Template-based methods rely on aligned CT/MR image pairs covering a reasonable range of anatomical variability, commonly performed using a combination of rigid and non-rigid image registration.[15] Subsequently, MR atlas images are registered pairwise to the target MR image followed by mapping the corresponding CT images to the target MR image using the obtained transformation maps. The generation of synthetic CT images from the transformed atlas CT is commonly performed using image fusion techniques (voxel-wise weighting or averaging).[16,17] The performance of atlas-based methods for cases with abnormal anatomies is restricted. Machine learning techniques cover a wide range of algorithms that attempt to establish a nonlinear relationship between MRI intensities and electron density maps. Among these approaches, convolutional neural networks (CNNs) exhibited great potential to accurately estimate electron density maps or achieve automated MR image segmentation. This approach has witnessed great success and tremendous growth in the image analysis framework over the years.[18] Nevertheless, much effort has been made to improve the performance and robustness of this approach in the framework of CT image synthesis from MRI owing to its dependence of the characteristics of the training datasets, such as noise and intensity variation, which would lead to gross errors.[19]

Nie et al.[20] used a generative adversarial network to train a fully three-dimensional (3D) convolutional neural network with the aim to produce a more realistic target for synthetic CT images. Their pelvic dataset consisted of 22 subjects, each with MR and CT images. They reported a mean absolute error (MAE) and peak signal-to-noise ratio (PSNR) of 39.0 ± 4.6 HU and 34.1 ± 1.0, respectively. Xiang et al.[21] proposed a very deep network architecture for synthesizing CT images from T1-weighted MR images. Their model had a transform and reconstruction steps featured by an intermediate block which embeds the tentative synthesis of CT images into feature maps. They trained their model using a prostate dataset consisting of 22 subjects, achieving a MAE and PSNR of 42.5 ± 3.1 HU and 33.5 ± 0.8, respectively. It is worth emphasizing that a higher PSNR does not necessarily imply perceptually better results.[22]

U-Net, SegNet, and Visual Geometry Group 16 (VGG16) models are among the most efficient convolutional neural network architectures.[23–25] The original VGG16 architecture benefits from 13 convolutional layers using small kernels of 3×3 at each layer connected to three fully connected layers. This model has in total 138 million trainable parameters and has been incorporated in many state-of-the-art deep convolutional neural network designs owing to its promising performance and robustness.[26,27] Similar to the VGG16 model, the U-Net architecture proposed by Ronneberger et al.[23] for biomedical image segmentation has exhibited high performance for a wide range of applications. Moreover, the SegNet model benefits from a deep convolutional encoder-decoder architecture and has shown promising performance in the context of image segmentation.[24]

These three models were frequently exploited for different applications owing to their efficient convergence even when using a small of number training datasets. The U-Net architecture had a contracting path to capture the context of the input shape and a symmetric expanding path for the reconstruction of segments in biomedical imaging. For precise localization, the high-resolution features from the contracting path were combined with the upsampled output in the expanding path. Inspired by the U-net architecture, Badrinarayanan et al.[24] proposed SegNet, a deep convolutional neural network for image segmentation using an encoder-decoder framework with pooling indices shortcut between them. Pooling indices indicate the locations where the feature maps in the encoder show high values and make major contribution to better reconstruct the output shape. The fundamental structures of the encoder and decoder blocks in this model were constructed based on the original U-Net network. These state-of-the-art architectures of the convolutional neural network aim at increasing the accuracy of the outcomes while avoiding dramatic increase in the complexity of the algorithm and training parameters.

Han employed a similar encoder-decoder structure and applied maxpooling indices shortcut between them which enabled an end-to-end CT image synthesis from MRI in the brain region.[28] This method was further evaluated in the pelvis region against state-of-the-art atlas-based methods, demonstrating comparative performance for synthetic CT estimation[29,30]. In addition to the above mentioned methods, generative adversarial networks (GAN) have shown great potential in a broad range of applications, including image reconstruction,[31,32] partial volume correction,[33] and super-resolution imaging.[34] In this regard, the adversarial semantic structure deep learning proposed in Ref. [35] resulted in reliable synthetic CT generation and clinically tolerable PET

quantification bias. Despite the promising performance of the above-mentioned approaches, a relatively large number of trainable parameters requires large training dataset to ensure efficient training while avoiding underfitting/overfitting. This also adds to the complexity of the optimization process to escape from the local minima and slow down the convergence rate of the training process.

Building on our previous work,[36] a new convolutional neuralnetwork architecture was proposed to achieve accurate and robust CT synthesis through efficient training using a small number of training subjects. This architecture is inspired from the U-net and SegNet structures with encoder-decoder compartments. The encoder-decoder compartments were structured based on the U-Net architecture modified by residual networks, deconvolutional layers and scaled exponential linear unit (SeLU) to achieve effective training and to minimize the hazard of overfitting. The proposed algorithm was evaluated in the context of CT image synthesis from pelvis MRI. Among the advantages of the technique is a robust network for efficient training using a small number of training subjects for applications where generating a large training dataset is challenging.

2. MATERIALS AND METHODS

2.A. Image acquisition and preprocessing

The dataset used in this study consists of 15 co-registered MRI-CT pairs of male pelvis scans (1861 two-dimensional images). The cohort included patients aged between 56 and 76 yr (68 $\pm$ 3) with body mass indices ranging from 18.9 to 34.8 kg/m^2 (25 $\pm$ 2.5). The CT scans were acquired on a GE LightSpeed RT (Milwaukee, USA) with a voxel size of 1.5625 $\times$ 1.5625 $\times$ 2 mm^3 and stored in a matrix of 256 $\times$ 256 $\times$ 128. The CT scans were performed with empty rectum and full bladder. The MRI scans were acquired on a Siemens Skyra 3T scanner (Erlangen, Germany) using a 3D T2-weighted 1.4 mm isotropic sampling perfection with application optimized contrast covering the whole pelvis area. The MRI voxel size was originally 1.4 $\times$ 1.4 $\times$ 2 mm^3 that was converted after coregistration to the corresponding CT image resolution. The patients were referred to the department of radiation therapy for the treatment of prostate cancer. The CT and MR images were acquired in the same day or with one day difference maximum. MRI to CT image registration was performed using a combination of rigid and non-rigid transformations and the normalized mutual information criterion. MR images were aligned to CT images using B-spline transform functions implemented within the Elastix* package. Prior to image registration, MR images underwent intensity nonuniformity (intra-patient) correction using N4 ITK software followed by image denoising using a bilateral edge preserving filter. Inter-patient MRI intensity variation was addressed by histogram matching to a common histogram template. Two-dimensional (2D) slices of MR and CT images were stacked in two separate tensors with dimensions of (1861, 256, 256, 1). The normalization of each tensor was performed using the following formula:

$$\boldsymbol{y}(\boldsymbol{Normalized}) = (\boldsymbol{x} - \boldsymbol{x}_{\boldsymbol{min}})/(\boldsymbol{x}_{\boldsymbol{max}} - \boldsymbol{x}_{\boldsymbol{min}}) \quad (1)$$

where x_{max} and x_{min} denote maximum and minimum values of image pixels in the tensor, respectively. Hence, the range of image intensities would be within the range [0–1] for both MR and CT images. In the next step, we divided each tensor into training tensor (1550, 256, 256, 1) and validation tensor (311, 256, 256, 1).

The evaluation of the proposed method includes four additional patients (in addition to 15 patients used for training and evaluation).These four patients were scanned using the same acquisition parameters and were solely included in the evaluation process.

2.B. Network architecture

The overall architecture of our model is inspired from the works of Ronneberger et al.[23] and Badrinarayanan[24] through the combination of U-net and encoder-decoder structures. The efficient CNN (eCNN) model was built based on the encoder-decoder networks in the U-Net model where the convolutional layers were replaced with the building structures (aiming at extracting image features from the input MRI[25]) as illustrated in Fig. 1 and Fig. S1. The number of filters in each of the building structures was set the same as those of the corresponding convolutional layers in the U-Net model.

The building structure has two 3 $\times$ 3 convolutional layers, wherein each layer is followed by batch normalization and SeLU activation layers to avoid dying rectified linear unit (ReLU) effects. ReLU was initially proposed to cope with the challenge of vanishing gradients, a difficulty faced by the neural networks which utilize gradient-based learning approaches (e.g., back propagation). This issue renders the parameters tuning in the earlier layers of the architecture complicated and becomes worse as the number of layers increase. ReLUs effectively tackled the vanishing gradient issue through converting the negative values to zero. In fact, ReLU acts as an identity map for positive inputs while negative inputs are mapped to zero. Dying ReLU effect or dead state occurs when this function gets stuck in the negative side. Since the slope of the negative side in ReLU is zero, once a neuron falls in this region, it is very unlikely to escape/recover from dead state. As such, these neurons do not play any role in the learning process. Although ReLU helps the deep neural network to handle the vanishing gradient issue, it inherently bears the risk of falling or getting stuck in a dead state. This could take place when changes in weights cause very small changes in the output of the next iteration in the sense that ReLU barely operates in the linear part (identical mapping or positive side). As such, the related cells are not able to contribute effectively to the learning of the network and their gradients remain almost equal to zero. If this phenomenon occurs in a considerable number of cells, the network could fail from operating properly. SeLU was introduced[37,38] by Klambauer et al.[39] to address this issue.

*http://elastix.isi.uu.nl/.

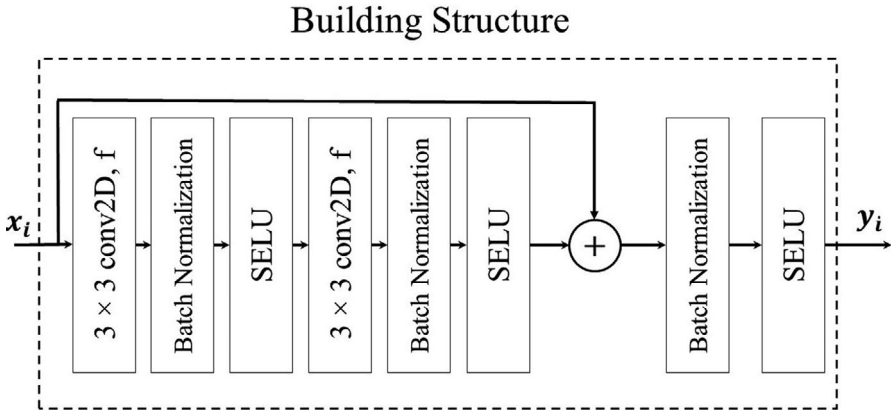

FIG. 1. The building structure in the proposed model. *f* denotes the number of filters in each convolutional layer of this structure.

$$SeLU(x) = \lambda \begin{cases} \alpha(e^x - 1) & x \leq 0 \\ x & x > 0 \end{cases} \quad (2)$$

where α is a constant value (equal to 1.6733) and x denotes the input. For input values greater than zero, SeLU operates like ReLU but multiplied by a factor λ (a positive number ~ 1.05). For a negative input, the output is different from zero and follows an exponential curve. This characteristic of SeLU promotes a self-normalizing property during the learning process and weight updates which helps to circumvent the dead state.

In this model, wherever a connection is established between two convolutional layers with different dimensions (or number of filters), a matching layer is inserted to adapt the dimensions. The depth of the network is critical for proper features extraction as deeper networks lead to higher order feature extraction.[37,38] Deep networks inherently integrate low/mid/high features into the end-to-end learning process. It should be noted that increasing the depth of the network by simply inserting a series of plain convolutional, batch normalization, namely SeLU activation and Maxpooling layers, is not enough for image segmentation or image classification tasks.[40] In fact, there are two major issues, which can potentially impair the performance of the network: overfitting and gradient vanishing/exploding. These issues prevent the deep networks from efficient convergence and affect the accuracy of the outcome. In convolutional neural networks, the number of learning parameters increases exponentially with the depth of the network and as such, apart from the computational cost, the training of the network would become more challenging. As discussed earlier, to overcome these issues, we replaced each convolutional layer in plain U-Net architecture with a building block architecture,[38] referred to as building structure in this work (Fig. 1).

This reformulation of layers caused more straightforward optimization and efficient convergence using a small dataset. Equation (3) formulates the core of the proposed model.

$$y_i = SeLU(SeLU(w_{i2}.SeLU(w_{i1}x_i + b_{i1}) + b_{i2}) + x_i) \quad (3)$$

x_i and y_i denote the input and output vectors of layer (i), respectively. In Fig. 1, f indicates the number of filters in each layer, w_{i1} and w_{i2} are the learning weights and b_{i1} and b_{i2} indicate the biases inside each building structure. Figure 2 summarizes the overall structure of our proposed deep CNN model. As explained earlier, the number of filters in our model is similar to the original U-Net model.

In the encoding part, whenever the number of filters is doubled, a maxpooling layer with 2 × 2 window and stride 2 (nonoverlapping window) was used in the next layer to reduce the size of the feature by half to avoid unnecessary computational cost. At the maxpooling layers, the pooling mask indices were saved for use at the corresponding decoding network as a shortcut connection (Fig. 2). The decoding layers were modified according to the corresponding encoding structures where the maxpooling layers were replaced with deconvolutional layers. This architecture contained in total 52 (26 encoding and 26 decoding) 3 × 3 convolutional layers enabling efficient feature extraction from input MRI and CT synthesis. During the learning procedure, the deep encoder network learns to extract a hierarchy of complex features from the input MRI. As shown in Fig. 1, before each SeLU layer, a batch normalization layer is set to reduce the internal covariate shifts and improving the training of the eCNN model. The existence of batch normalization layer enabled the use of higher learning rates and caused less sensitivity to the initialization of the training parameters.[41] The decoding network is a mirroring of the encoding network except that instead of downsampling by a maxpooling layer, a 2D convolutional transpose (deconvolution) layer with 2 × 2 window and stride 2 was used for upsampling. This allowed efficient update of the parameters of this layer during training. At the end of the decoding network, a 1 × 1 convolutional layer reconstructs the sCT image with the same resolution as the input MRI.

2.C. Model implementation

The proposed deep convolutional neural network model is implemented using the open source Keras TensorFlow

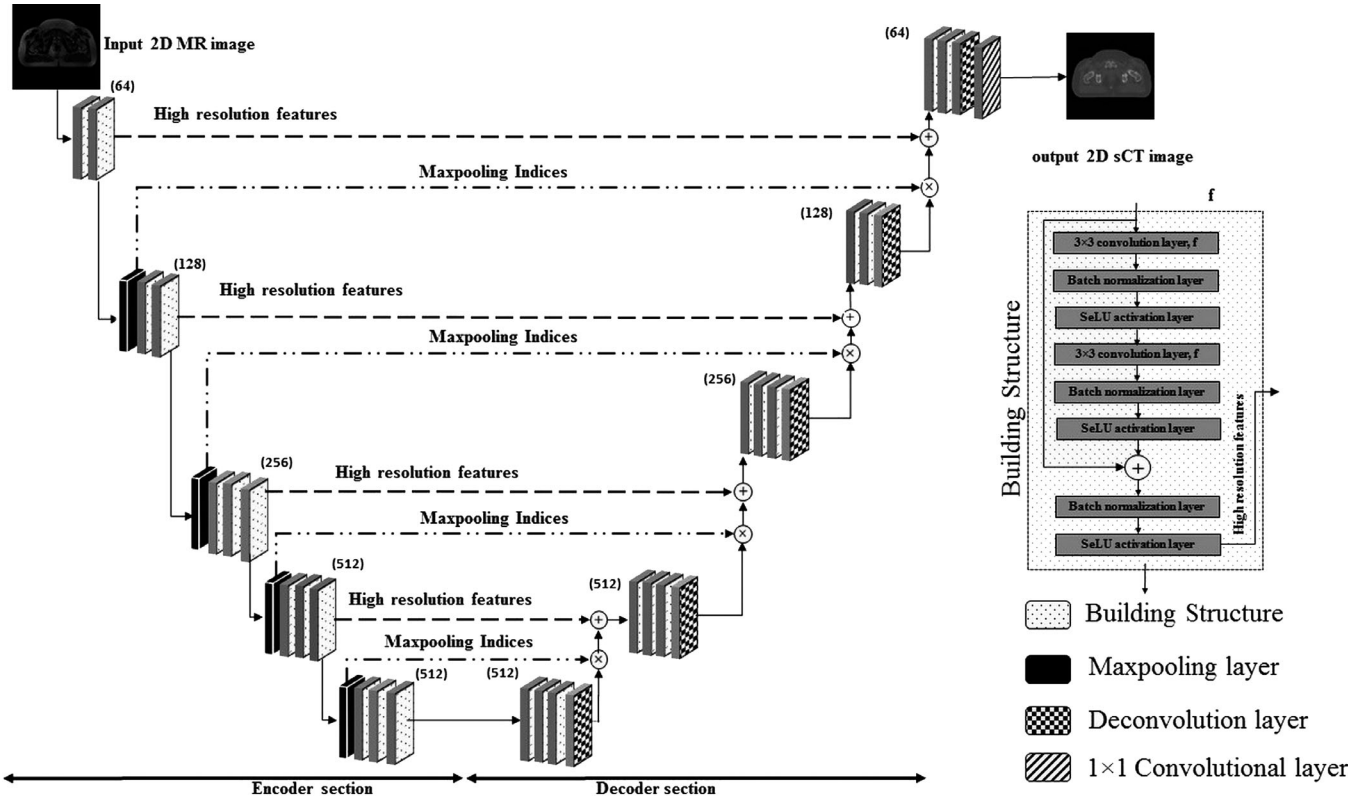

FIG. 2. Architecture of the proposed model. The digits shown next to each building structure denote the number of filters used in the convolutional layers. The dashed arrows labeled with high resolution features indicate the connections between encoding and decoding blocks to transfer high-resolution features. The dashed arrows labeled with maxpooling indices depict the connections between the maxpooling layers and decoding networks to transfer the corresponding indices. The architecture of inside the building structures are indicated by the dashed rectangles.

backend package.[42] The algorithm was run on an NVIDIA GTX GEFORCE 1080 Ti with 11 GB graphics memory. The training was performed using mean absolute error as loss function and back-propagation algorithm with Adam stochastic optimization method. A batch size of 13 was used for the training of the model. A higher batch size was not possible owing to limitations in the graphic memory. The bias and kernel initializer were set at "zeros" and "he- normal", respectively, for a better convergence rate in the eCNN model. Using batch normalization layers reduced the internal covariate shift which is the change in the input distribution of each layer. The input to each layer might be affected by certain parameters which could lead to fluctuation of the input to the next layer. By using batch normalization layers, the internal covariate shift was reduced trough minimizing the changes in the input distribution of each layer and fluctuation of the input to the next layer. For this network, the learning rate was set to 0.01 and momentum to 0.9 for proper training. In total, the eCNN model has 52 two-dimensional 3×3 convolutional layers and 66996609 trainable parameters. Without using any pre-trained model for encoding and decoding parts or any data augmentation, the eCNN model learned to efficiently generate sCT images from MRI. The 3D images of the 15 patients were converted to 1861 two-dimensional 256×256 slices among which 1550 were used for training and the rest for evaluation using a fivefold cross-validation scheme. The eCNN model is able to converge without any significant overfitting after less than 200 epochs. For an eloquent comparison, the basic encoder-decoder model based on the U-Net architecture (Fig. S2) was also evaluated in this work to provide a bottom line for performance assessment of the proposed eCNN model. To this end, the U-Net model was trained using the same dataset and a sufficient number of iterations (200 epochs) to ensure proper convergence. The hyperparameters were separately fine-tuned for a fair comparison. Figure 3 shows the training and validation loss of the proposed and the U-Net models for 200 training epochs.

2.D. Atlas-based method

The proposed deep learning-based technique was compared to an atlas-based method to provide insight to the level of accuracy achieved using the eCNN model. A representative atlas-based approach was implemented in this work, which involved pairwise registration of the MR atlas images to the target subject in a leave-one-out cross-validation scheme.[43] To this end, MR images of the 14 patients were deformed to match the MR image of the target patient using a combination of rigid and nonrigid registrations. Image registration was performed using the B-spline transform function and a normalized mutual information criterion as loss function implemented within the Elastix package.[44] Thereafter,

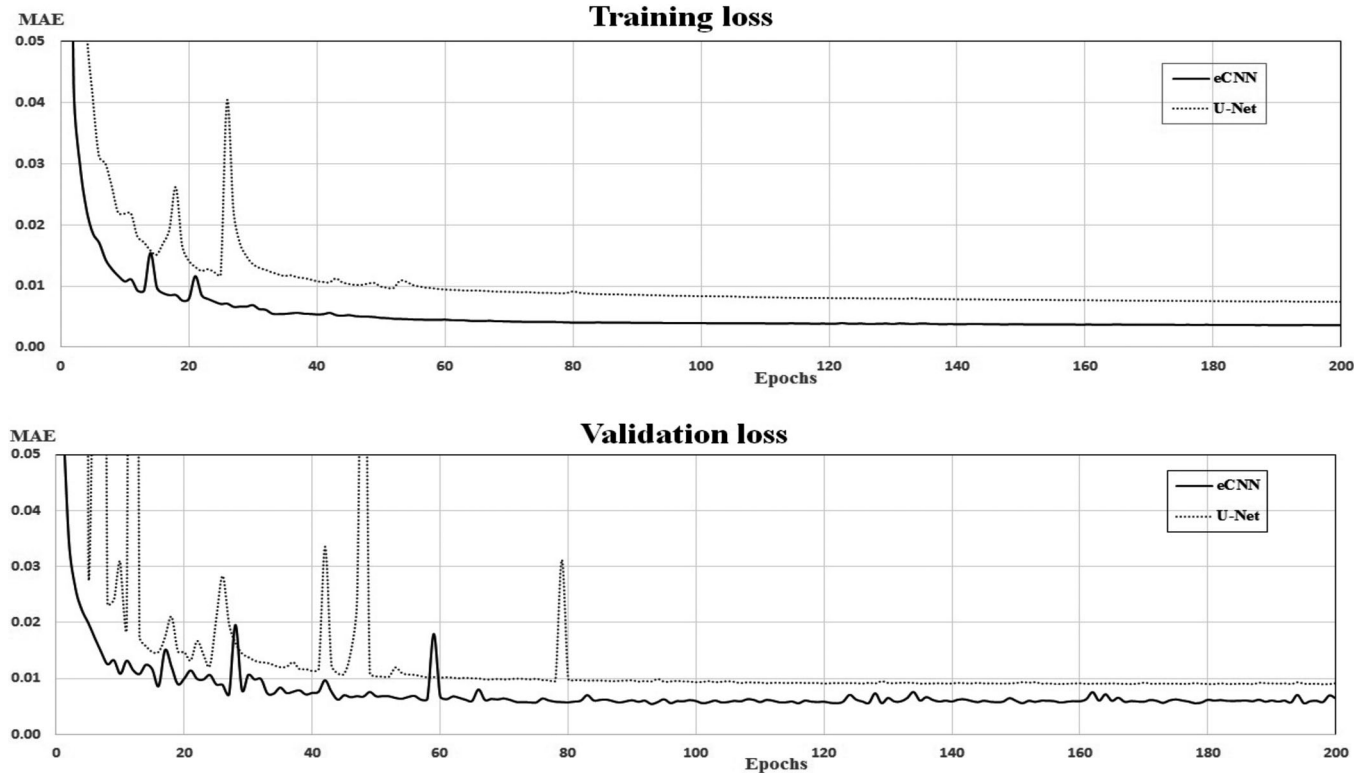

FIG. 3. Training and validation losses of efficient convolutional neural network and U-net architectures within 200 epochs.

using the obtained transformation maps, the corresponding CT atlas images were mapped to the target subject (for each target patient, 14 atlas CT images were transformed to a common coordinate of the target subject). The final atlas-based synthetic CT images were generated by taking the average of the all transformed CT images in a voxel-wise manner.[45]

2.E. Evaluation strategy

The accuracy of our eCNN model was evaluated by comparing the generated sCT images to the ground truth CT images using the MAE and mean error (ME) metrics. Furthermore, the Pearson correlation coefficient (PCC) and structural similarity index (SSIM) were also computed between the ground truth CT and sCT. PCC is a measure of the linear correlation between two samples whereas the SSIM is a measure for predicting the perceived quality of digital images and videos. The calculation of the above-mentioned metrics was carried out only for voxels within the body contour using the following equations:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|CT(i) - sCT(i)| \tag{4}$$

$$ME = \frac{1}{N}\sum_{i=1}^{N}(CT(i) - sCT(i)) \tag{5}$$

$$PCC(CT, sCT) = \frac{\sum_{i=1}^{N}\left(CT(i) - \overline{CT}\right)\left(sCT(i) - \overline{sCT}\right)}{\sqrt{\sum_{i=1}^{N}\left(CT(i) - \overline{CT}\right)^2}\sqrt{\sum_{i=1}^{N}\left(sCT(i) - \overline{sCT}\right)^2}} \tag{6}$$

$$SSIM(CT, sCT) = \frac{\left(2\overline{CT}\cdot\overline{sCT} + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\overline{CT}^2 + \overline{sCT}^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{7}$$

$$PSNR(dB) = 20.\log_{10}\left(\frac{1}{\sqrt{MSE}}\right) \tag{8}$$

where $\overline{CT}$ and $\overline{sCT}$ are means of reference CT and synthetic CT images, respectively. c_1 and c_2 are constants and σ_{xy}, σ_x and σ_y denote the covariance of $\overline{CT}$ and $\overline{sCT}$, variance of $\overline{CT}$ and variance of $\overline{sCT}$ samples, respectively. MSE indicates pixelwise mean squared error between synthetic CT and reference CT images.

The above-mentioned metrics were also calculated separately for air, soft-tissue, and air cavities. These tissues were segmented from the reference CT and sCT images by applying the following intensity thresholds: bone >160 HU, air cavity <−400 HU inside the body contour, soft-tissue between −400 and 160 HU.[35] Moreover, given the segmented bone, air cavities and soft-tissue from the synthetic and reference CT images, the dice similarity coefficient[46] was calculated to evaluate the tissue identification accuracy

using the atlas method, the U-Net architecture and the eCNN model.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{9}$$

In Eq. (9), X and Y denote the binary mask of tissues segmented from the synthetic CT and reference CT images, respectively.

All quantitative metrics were calculated in 2D on each slice (rather than on the whole 3D volume). Hence, the mean and standard deviation reflect the performance of the different approaches on a 2D slice basis. Tissue segmentation was performed as a post-processing procedure on synthetic CT images generated by eCNN, the U-Net architecture and the atlas-based methods. The segmentation threshold levels utilized in Ref. [35] were adopted for the delineation of contours.

To investigate the impact of data augmentation on the performance of the eCNN and U-Net models, affine transformations using the following sub-transforms were implemented: ±5° rotation, ±5% translations, ±5° shearing and 5% zooming. The quantitative results before and after data augmentation were compared for both models.

To evaluate the models using the four unseen external subjects (512 2D images), the training was carried out using the 15 subjects of the training dataset (1550 2D images). The results of the external dataset are reported separately.

3. RESULTS

The training of the eCNN and original U-Net models was performed using a fivefold cross validation scheme where 1550 two-dimensional images were used for training and 311 slices for evaluation within 200 training epochs. Figure 4 illustrates representative views of the generated synthetic CT images along with the target MRI and the ground truth CT images. The visual inspection revealed the superior quality of the synthetic CT generated by the eCNN model compared to the atlas-based and original U-Net methods in terms of anatomical details (in particular air pockets) and bone delineation. Figure S3 also depicts synthetic CT images of another subject along with the target MRI and references CT images, wherein the bladder and rectum are visible and highlighted.

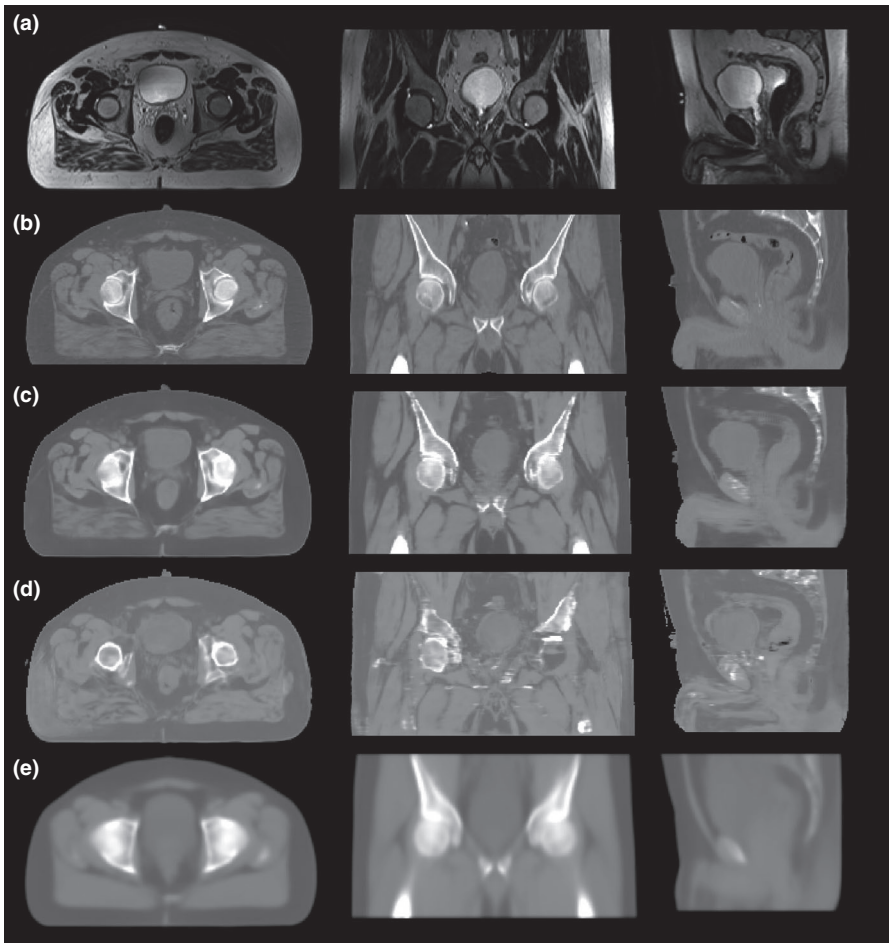

FIG. 4. Qualitative comparison of synthetic computed tomography (sCT) images generated using the efficient convolutional neural network, U-Net and atlas-based methods against ground truth CT together with the original input MRI shown in axial, coronal and sagittal planes from left to right, respectively. (a) Input MRI, (b) ground truth CT, (c) sCT generated using the eCNN method, (d) sCT generated using the U-Net model, and (e) atlas-based synthetic CT.

The results of the quantitative analysis are summarized in Tables I and II which report the MAE, ME, PCC, SSIM, and PSNR of the CT synthesis results for 15 training/validation patients and four additional patients, respectively, over the whole pelvis region. Tables III and IV present the results of the same evaluation performed in bone, air cavities and soft-tissue regions for 15 training/validation and 4 additional patients, separately. Dice similarity coefficients are reported in Table V for the segmented air cavities, bone, and soft-tissue from synthetic CT images. Representative slices of segmented air cavities, bone and soft-tissue from and sCT and ground truth CT images are shown in Fig. 5.

The eCNN and U-Net models were revaluated with and without data augmentation where the results for the four (unseen) external patients are presented in Table VI. In contrast to the original U-Net model, which performed much better after data augmentation, the eCNN model exhibited no improvement.

To verify the effectiveness of the major components added to the model, notably SeLU and building structure, the eCNN model was re-implemented several times to put into perspective the contribution of each of these components. SeLU was replaced with ReLU and building structure with a plain 3×3 convolutional layer (plain structure). Figures S4 and S5 illustrate the training and validation losses for the eCNN model with ReLU activation function and plain structure, respectively. The superior performance of the SeLU activation function is clearly visible in Fig. S4, wherein remarkably less fluctuations are observed in the outcome of the eCNN model with the SeLU activation layer. The building structure component led to improved convergence of the model and significantly higher prediction accuracy (lower loss values in both training and validation dataset). To sum up, Table S2 summarizes the quantitative metrics calculated in the whole pelvis, air, bone, and soft-tissue content regions, respectively, for the four external patients (512 2D images) using the eCNN model with ReLU activation function and without building

TABLE I. Comparison of different quantitative metrics across the entire pelvic region between efficient convolutional neural network (eCNN), U-Net, and atlas-based methods for the 15 patients in a fivefold cross validation scheme.

	MAE (HU) (Std. Dev.)			ME (HU) (Std. Dev.)			PCC (Std. Dev.)			SSIM (Std. Dev.)			PSNR (Std. Dev.)		
Patient	eCNN	Atlas	U-Net	eCNN	Atlas	U-Net	eCNN	Atlas	U-Net	eCNN	Atlas	U-Net	eCNN	Atlas	U-Net
1	21.8 (4.2)	55.5 (22.0)	36.8 (13.2)	1.5 (1.8)	11 (29.3)	1.9 (4.6)	0.95 (0.03)	0.80 (0.09)	0.82 (0.11)	0.98 (0.01)	0.95 (0.03)	0.96 (0.02)	35.6 (1.8)	23.1 (1.1)	31.2 (2.7)
2	27.6 (6.4)	45.9 (5.8)	45.7 (12.4)	4.1 (4.2)	13.0 (7.4)	5.6 (8.5)	0.90 (0.05)	0.85 (0.05)	0.71 (0.12)	0.97 (0.01)	0.96 (0.01)	0.96 (0.02)	31.0 (1.8)	22.1 (0.5)	27.7 (1.7)
3	24.0 (3.0)	46.7 (3.7)	38.2 (6.2)	2.1 (2.6)	9.5 (7.4)	11.0 (9.9)	0.94 (0.02)	0.87 (0.03)	0.84 (0.06)	0.98 (0.00)	0.96 (0.01)	0.96 (0.01)	33.8 (1.8)	24.3 (0.2)	30.0 (1.2)
4	39.1 (5.2)	52.9 (11.2)	54.0 (8.9)	16.1 (8.0)	−9.4 (14.1)	28.2 (8.0)	0.88 (0.04)	0.83 (0.08)	0.76 (0.06)	0.96 (0.01)	0.95 (0.01)	0.93 (0.02)	29.7 (1.3)	24.9 (1.0)	27.0 (1.1)
5	23.1 (2.5)	55.0 (22.9)	35.9 (4.2)	−1.6 (2.6)	5.4 (29.7)	9.33 (5.39)	0.94 (0.02)	0.82 (0.10)	0.84 (0.03)	0.98 (0.01)	0.93 (0.03)	0.95 (0.01)	33.1 (1.5)	20.8 (0.8)	29.2 (1.1)
6	23.5 (6.4)	74.2 (60.0)	35.1 (8.0)	0.9 (3.0)	4.6 (72.3)	8.7 (9.6)	0.95 (0.02)	0.81 (0.15)	0.85 (0.05)	0.98 (0.01)	0.95 (0.03)	0.97 (0.01)	34.7 (2.0)	22.3 (2.4)	30.8 (1.6)
7	24.2 (5.8)	55.0 (15.5)	46.5 (7.7)	1.5 (5.3)	6.7 (21.8)	19.2 (5.5)	0.95 (0.03)	0.86 (0.02)	0.82 (0.07)	0.99 (0.00)	0.96 (0.01)	0.96 (0.01)	35.1 (2.1)	24.4 (0.5)	29.3 (1.7)
8	22.0 (2.2)	54.7 (3.7)	41.1 (5.8)	−0.1 (2.3)	22.2 (5.3)	−5.4 (7.1)	0.94 (0.02)	0.88 (0.01)	0.81 (0.05)	0.98 (0.01)	0.96 (0.01)	0.95 (0.02)	33.1 (1.2)	23.3 (0.2)	28.3 (1.2)
9	17.5 (1.3)	56.9 (2.9)	26.3 (2.4)	0.9 (1.6)	34.4 (5.1)	5.8 (2.9)	0.96 (0.01)	0.86 (0.03)	0.91 (0.03)	0.99 (0.00)	0.92 (0.01)	0.97 (0.01)	36.0 (1.2)	16.0 (0.2)	32.5 (1.1)
10	23.4 (5.0)	43.3 (5.0)	38.9 (7.1)	1.5 (3.6)	4.0 (8.8)	9.0 (7.0)	0.94 (0.03)	0.85 (0.03)	0.81 (0.07)	0.98 (0.01)	0.95 (0.01)	0.95 (0.02)	33.3 (2.0)	25.1 (0.4)	29.2 (1.3)
11	25.9 (5.0)	55.9 (52.0)	38.5 (7.7)	5.9 (5.3)	−4.1 (53.0)	13.3 (4.5)	0.91 (0.03)	0.82 (0.16)	0.8 (0.05)	0.98 (0.00)	0.95 (0.04)	0.96 (0.01)	32.4 (2.0)	25.9 (2.7)	29.2 (1.5)
12	37.5 (5.3)	66.1 (2.8)	56.7 (13.0)	−7.0 (4.4)	27.2 (8.9)	−8.3 (11.8)	0.86 (0.05)	0.81 (0.05)	0.71 (0.08)	0.97 (0.01)	0.95 (0.01)	0.94 (0.02)	30.9 (1.3)	22.2 (0.3)	28.3 (1.1)
13	42.3 (8.9)	106.5 (25.6)	53.6 (13.1)	13.9 (6.0)	−91.2 (27.3)	12.0 (9.8)	0.80 (0.07)	0.83 (0.09)	0.71 (0.08)	0.95 (0.02)	0.95 (0.02)	0.93 (0.02)	27.6 (1.6)	28.3 (2.7)	26.2 (1.2)
14	55.5 (6.7)	82.7 (43.6)	57.4 (9.8)	−22.7 (6.4)	34.2 (61.0)	−21.0 (8.2)	0.80 (0.06)	0.82 (0.11)	0.76 (0.06)	0.96 (0.02)	0.95 (0.03)	0.95 (0.01)	29.0 (1.2)	22.4 (1.5)	28.4 (1.4)
15	43.2 (6.4)	117.4 (194.1)	50.8 (9.2)	24.8 (5.6)	−79.4 (203.3)	22.4 (8.6)	0.82 (0.06)	0.67 (0.35)	0.73 (0.11)	0.95 (0.02)	0.90 (0.12)	0.94 (0.02)	28.0 (1.5)	22.7 (4.8)	27.0 (1.3)
Average	30.0 (10.4)	64.6 (21.2)	44.0 (8.8)	2.8 (10.28)	−0.8 (35.4)	7.4 (11.9)	0.90 (0.06)	0.83 (0.05)	0.79 (0.06)	0.97 (0.01)	0.95 (0.02)	0.95 (0.01)	32.2 (2.7)	23.2 (2.7)	28.9 (1.7)

TABLE II. Comparison of different quantitative metrics across the entire pelvic region between efficient convolutional neural network (eCNN), U-Net, and atlas-based methods for the four additional patients.

	MAE(HU) (Std. Dev.)			ME(HU) (Std. Dev.)			PCC (Std. Dev.)			SSIM (Std. Dev.)			PSNR (Std. Dev.)		
Patient	eCNN	Atlas	U-Net	eCNN	Atlas	U-Net	eCNN	Atlas	U-Net	eCNN	Atlas	U-Net	eCNN	Atlas	U-Net
1	46.2 (6.5)	128.1 (213.3)	44.5 (7.8)	20.31 (8.1)	−108.1 (217.8)	19.7 (8.3)	0.83 (0.06)	0.66 (0.36)	0.79 (0.08)	0.95 (0.01)	0.90 (0.13)	0.95 (0.02)	28.8 (1.6)	23.4 (5.4)	26.7 (1.3)
2	40.0 (12.2)	65.2 (56.0)	55.6 (16.1)	−15.4 (10.0)	−1.8 (64.7)	−24.5 (10.5)	0.80 (0.06)	0.77 (0.21)	0.77 (0.10)	0.95 (0.02)	0.93 (0.04)	0.95 (0.02)	27.8 (2.2)	23.1 (2.3)	27.2 (2.3)
3	33.6 (7.7)	86.1 (2.6)	41.7 (6.0)	6.0 (6.6)	67.0 (7.6)	17.1 (9.1)	0.84 (0.07)	0.90 (0.04)	0.74 (0.07)	0.97 (0.01)	0.97 (0.01)	0.96 (0.01)	30.7 (1.8)	27.1 (0.4)	28.2 (0.9)
4	32.2 (6.6)	72.6 (4.2)	42.1 (6.5)	13.0 (4.2)	51.7 (8.2)	12.3 (6.8)	0.83 (0.09)	0.89 (0.05)	0.7 (0.08)	0.97 (0.01)	0.97 (0.01)	0.95 (0.01)	30.7 (1.8)	27.9 (1.2)	27.5 (1.3)
Average	38.0 (5.6)	88.0 (24.3)	46.0 (5.7)	6.0 (13.4)	2.2 (68.6)	6.2 (17.9)	0.83 (0.02)	0.81 (0.10)	0.75 (0.03)	0.96 (0.01)	0.94 (0.03)	0.95 (0.00)	29.5 (1.3)	25.4 (2.1)	27.4 (0.6)

TABLE III. Summary of quantitative metrics including mean absolute error (MAE), mean error (ME), Pearson correlation coefficient (PCC), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) in air, bone and soft tissue regions for the efficient convolutional neural network (eCNN), U-Net, and atlas-based techniques over the 15 patients.

Region	Method	MAE(HU) (Std. Dev.)	ME(HU) (Std. Dev.)	PCC (Std. Dev.)	SSIM (Std. Dev.)	PSNR (Std. Dev.)
Air	eCNN	548.1 (115.1)	−495.6 (143.2)	0.17 (0.12)	0.97 (0.01)	12.9 (2.1)
	Atlas	592.8 (91.0)	−320.93 (116.6)	0.37 (0.23)	0.94 (0.02)	11.6 (2.2)
	U-Net	576.4 (113.8)	−620.0 (147.5)	0.12 (0.11)	0.97 (0.01)	11.9 (1.8)
Bone	eCNN	144.51 (54.02)	85.0 (55.7)	0.73 (0.12)	0.99 (0.00)	23.1 (3.1)
	Atlas	236.2 (85.6)	−101.1 (136.3)	0.62 (0.17)	0.95 (0.02)	20.3 (1.8)
	U-Net	218.5 (76.8)	161.4 (83.1)	0.64 (0.11)	0.98 (0.01)	20.6 (2.8)
Soft tissue	eCNN	21.8 (6.2)	−4.4 (9.4)	0.84 (0.05)	0.98 (0.00)	36.6 (1.4)
	Atlas	66.6 (21.2)	−53.7 (30.6)	0.72 (0.06)	0.96 (0.02)	23.3 (3.5)
	U-Net	23.1 (6.6)	−7.0 (10.3)	0.82 (0.05)	0.98 (0.00)	36.1 (1.5)

TABLE IV. Summary of quantitative metrics including mean absolute error (MAE), mean error (ME), Pearson correlation coefficient (PCC), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) in air, bone, and soft tissue regions for the four external patients.

Region	Method	MAE(HU) (Std. Dev.)	ME(HU) (Std. Dev.)	PCC (Std. Dev.)	SSIM (Std. Dev.)	PSNR (Std. Dev.)
Air	eCNN	699.7 (64.9)	−692.7 (70.1)	0.04 (0.03)	0.97 (0.01)	10.6 (0.6)
	Atlas	731.3 (13.5)	−305.04 (121.0)	0.33 (0.19)	0.96 (0.01)	9.9 (1.0)
	U-Net	713.0 (48.8)	−781.8 (39.7)	0.03 (0.04)	0.97 (0.01)	9.98 (0.6)
Bone	eCNN	176.9 (20.9)	124.4 (25.9)	0.60 (0.05)	0.99 (0.00)	20.2 (1.0)
	Atlas	254.8 (121.3)	50.77 (132.8)	0.51 (0.29)	0.97 (0.01)	19.5 (1.2)
	U-Net	202.4 (36.9)	219.5 (57.3)	0.53 (0.06)	0.99 (0.00)	19.7 (1.0)
Soft tissue	eCNN	26.3 (4.6)	2.7 (10.7)	0.88 (0.04)	0.98 (0.00)	34.8 (1.0)
	Atlas	70.5 (16.2)	−61.3 (38.9)	0.69 (0.08)	0.95 (0.02)	26.0 (2.2)
	U-Net	28.4 (2.4)	0.18 (13.0)	0.77 (0.02)	0.98 (0.00)	34.7 (0.7)

structure. Combination of the SeLU activation function and the building structure in the eCNN model enhanced remarkably the performance of this model.

4. DISCUSSION

The use of deep learning techniques for CT synthesis from MRI sequences has witnessed rapid growth over the years owing to their promising performance compared to state-of-the-art methods.[19,47] The primary aim of this study was to introduce a robust deep convolutional neural network presenting with efficient convergence in the training phase without compromising the CT synthesis accuracy. The eCNN framework described in this work incorporates the encoder-decoder architecture into the U-Net model. For the sake of effective training, the encoder-decoder architecture was modified by the residual networks through establishing extra connection between the encoder and decoder compartments. Using SeLU as activation layer and establishing the parameter free identity shortcut connections in each building structure enabled avoiding overfitting and the gradient vanishing/exploding phenomena while achieving efficient training. Moreover, multiplication of the maxpooling indices extracted from the encoding compartment to the upsampling layers in

TABLE V. Comparison of average Dice coefficient indices over the 15 patients and four additional patients for air, bone and soft-tissue regions using efficient convolutional neural network (eCNN), atlas-based and U-Net methods with respect to the ground truth computed tomography (CT).

	eCNN			Atlas			U-Net		
	Air	Bone	Soft tissue	Air	Bone	Soft tissue	Air	Bone	Soft tissue
DSC for 15 patients	0.77 (0.09)	0.84 (0.07)	0.98 (0.01)	0.51 (0.22)	0.75 (0.06)	0.97 (0.01)	0.50 (0.24)	0.71 (0.06)	0.90 (0.03)
DSC for 4 Ext. patients	0.16 (0.11)	0.77 (0.03)	0.98 (0.00)	0.59 (0.26)	0.75 (0.03)	0.95 (0.02)	0.13 (0.12)	0.70 (0.09)	0.98 (0.00)

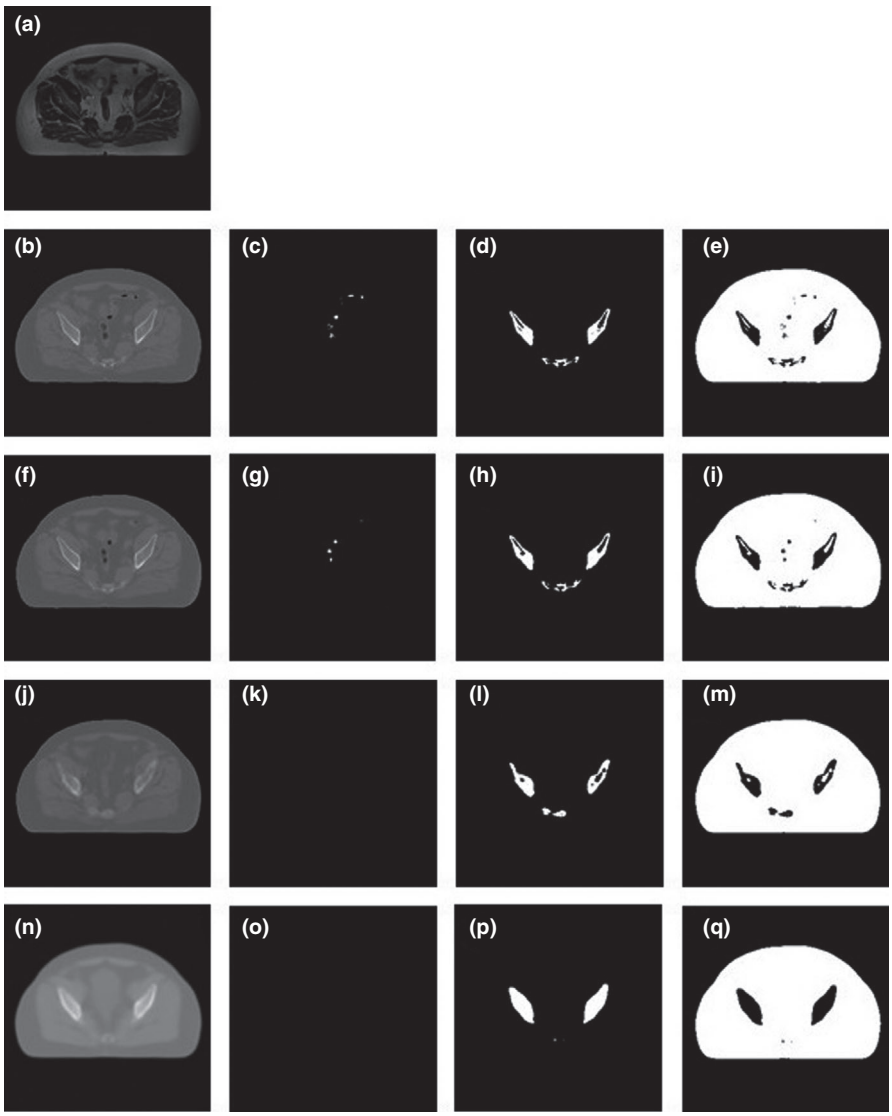

FIG. 5. Representative slices of ground truth computed tomography (CT), magnetic resonance imaging (MRI)-based synthetic CT (sCT) image in axial plane generated using the efficient convolutional neural network (eCNN) and U-Net models as well as atlas-based method together with binary masks of air, bone and soft tissue. (a) Input MRI, (b) Ground truth CT, (c) ground truth air mask, (d) ground truth bone mask, (e) ground truth soft-tissue mask, (f) eCNN sCT, (g) eCNN air mask, (h) eCNN bone mask, (i) eCNN soft-tissue mask, (j) U-Net sCT, (k) U-Net air mask, (l) U-Net bone mask, (m) U-Net soft-tissue mask, (n) atlas-based sCT, (o) atlas-based air mask, (p) atlas-based bone mask, and (q) atlas-based soft-tissue mask.

the decoder compartment created a sparse/over-complete representation.

The over-complete data representation facilitates the process of solution finding whereas the sparse representation enables the model to converge to a unique and accurate solution. The extra connections established between the encoder and decoder compartments allowed the model to exchange the high resolution features and created a robust CT synthesis network with lower trainable parameters and complexity.

TABLE VI. Summary of the quantitative metrics, including mean absolute error (MAE), mean error (ME), Pearson correlation coefficient (PCC), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) in whole pelvis region with and without data augmentation for the four external patients (512 2D slices) when using the efficient convolutional neural network (eCNN) and the U-Net architectures based on 13 initial layers of the U-NET model, including shortcut connection between encoder and decoder parts, ReLU activation layers, upsampling layers, and using maxpooling indices.

	MAE(HU) (Std. Dev.)	ME(HU) (Std. Dev.)	PCC (Std. Dev.)	SSIM (Std. Dev.)	PSNR (Std. Dev.)
eCNN					
Without data augmentation	38.0 (5.6)	6.0 (13.4)	0.83 (0.02)	0.96 (0.01)	29.5 (1.2)
With data augmentation	41.1 (7.0)	8.7 (13.7)	0.79 (0.03)	0.96 (0.01)	28.8 (1.5)
U-net					
Without data augmentation	46.0 (5.7)	6.2 (17.9)	0.75 (0.03)	0.95 (0.00)	27.7 (0.4)
With data augmentation	42.7 (9.3)	0.7 (19.9)	0.77 (0.04)	0.95 (0.01)	28.2 (1.6)

Our experimental results showed that using SeLU as activation function resulted in a more efficient learning behavior (lower training and evaluation loss) within less than 200 epochs of training (Fig. 3). The network is capable of reaching the plateau without any significant overfitting while the models proposed by Han,[28] Emami et al.[48] and Fu et al.[49] required 600, 300, and 200 epochs, respectively, to achieve proper training (minimizing the loss function).

In addition, the simultaneous use of maxpooling indices and U-Net shortcut connections between encoding and decoding networks together with replacing conventional plain connection with residual network in the eCNN model resulted in robust CT synthesis using a limited number of training datasets. Comparing the results of this study to the work of Arabi et al.[19] revealed improved statistical metrics measured in the entire pelvis region. The proposed eCNN exhibited superior performance to the atlas-based method achieving a MAE of 30.0 ± 10.4 HU and ME of 2.8 ± 10.3 HU for the entire pelvis region while the atlas-based method resulted in a MAE and ME of 64.6 ± 21.2 HU and -0.8 ± 35.4 HU, respectively. Fu et al.[49] proposed a similar model to the work of Han[28] where the batch normalization and upsampling layers were replaced with the instance norm and deconvolutional layers. The modified model resulted in MAEs of 40.5 ± 5.4 HU, 28.9 ± 4.7 HU and 159.7 ± 22.5 HU for the whole pelvis, soft-tissue, and bone, respectively. Conversely, the eCNN model proposed in this work exhibited MAEs of 30.0 ± 10.4 HU, 21.8 ± 6.2 HU, and 144.5 ± 24.0 HU for the same regions, respectively, thus demonstrating better performance than the model proposed by Fu et al. as both models were trained on 2D images. Considering bone extraction accuracy, Fu et al. reported a DSC of 0.81 ± 0.04 for bone segmented using an intensity threshold of 150 HU.[49] To facilitate the comparison, the evaluation of bone extraction was repeated using the same intensity threshold where the eCNN model resulted in a DSC of 0.84 ± 0.07, while the original U-Net model led to DSC of 0.71 ± 0.06. Overall, the eCNN method outperformed the atlas-based and original U-Net methods in terms of CT value estimation and tissue delineation.

A limited number of CT synthesis studies were conducted in the pelvis region. Hence, the performance of the proposed approach was compared to previous works in the brain region. Compared to a MAE and ME of 30.0 ± 10.4 HU and 2.8 ± 10.3 HU achieved by the eCNN, Han[28] reported values of 84.8 ± 17.3 HU and -3.1 ± 21.6 HU for the same metrics, respectively, using a typical U-Net network architecture. Emami et al.[48] reported a MAE of 89.30 ± 10.25 HU, SSIM of 0.83 ± 0.03 and PSNR of 26.64 ± 1.17, respectively, using a generative adversarial network. In this regard, the eCNN model exhibited better performance leading to SSIM and PSNR of 0.97 ± 0.01 and 32.20 ± 2.65, respectively.

The comparison of the results obtained in this work with other articles might not be fair since they are not evaluated on the same datasets with the same pre-processing steps. Hence, the models proposed by Han[28] and Fu et al.[49] were implemented in this work to conduct a fair comparative assessment. Figure S6 depicts the training and validation losses of these models in comparison with eCCN model. The eCNN model exhibited faster convergence with noticeably less fluctuation in the validation loss. Moreover, Table S1 summarizes the quantitative metrics, including MAE, ME, PCC, SSIM, and PSNR in the whole pelvis region, air, bone, and soft-tissue obtained from the different eCNNs as well as Han[28] and Fu et al.[49] models for the four extra patients. eCNN resulted in a MAE of 38.0 ± 5.6 (HU) for the whole pelvis, thus outperforming Han's and Fu's 2D models which achieved a MAE of 144.8 ± 27.7 (HU) and 197.0 ± 43.5 (HU), respectively.

To investigate the performance of the eCCN and U-net models with a smaller training dataset, the training of the network was repeated using only 900 training samples (selected randomly) from the training dataset and the models were evaluated on the same test dataset. The entire dataset contained 1861 co-registered MR and CT image pairs. The original eCNN model was trained using 1550 and tested on 311 samples, respectively. Table VII compares the results of the eCNN and U-models before and after reducing the size of the training dataset. Despite reducing the training dataset by almost half, the accuracy of the eCNN model did not change dramatically. Evidently, the U-net model cannot tolerate a reduction in the training dataset and the results were significantly degraded, particularly in bone and soft-tissue regions. Moreover, Supplemental Figure S7 compares the visual quality of the generated synthetic CT images before and after reducing the size of the training dataset for eCNN and U-Net models.

A possible extension of this work could be to employ state-of-the-art architectures of the VGG19 network and evaluate its performance in different body regions, particularly the lung region, which is challenging for accurate CT synthesis.

TABLE VII. Summary of quantitative metrics, including mean absolute error (MAE), mean error (ME), Pearson correlation coefficient (PCC), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) calculated in the pelvis, air, bone and soft-tissue regions for the four external patients when using the efficient convolutional neural network (eCNN) and U-models before and after reducing the size of the training dataset

		MAE(HU)		ME(HU)		PCC		SSIM		PSNR	
Model	Region	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
eCNN (trained with 1550 samples)	Whole pelvis	38	5.59	6	13.35	0.83	0.02	0.96	0.01	29.48	1.26
eCNN (trained with 900 samples)		40	4.03	−14	4.5	0.82	0.01	0.96	0.01	24.97	0.42
U-Net (trained with 1550 samples)		45	7.67	9	10.75	0.8	0.03	0.97	0	27.87	0.78
U-Net (trained with 900 samples)		123	10.32	107	9.62	0.75	0.01	0.96	0	24.1	0.34
eCNN (trained with 1550 samples)	Air	700	64.86	−693	70.07	0.04	0.03	0.97	0.01	10.57	0.64
eCNN (trained with 900 samples)		711	76.73	−730	94.54	0.03	0.02	0.95	0.02	9.99	0.64
U-Net (trained with 1550 samples)		356	81.04	−376	112.85	0.04	0.11	0.97	0	13.39	1.4
U-Net (trained with 900 samples)		309	64.5	−232	85.98	0.2	0.11	0.98	0	16.54	1.75
eCNN (trained with 1550 samples)	Bone	177	20.92	124	25.92	0.6	0.05	0.99	0	20.24	1.02
eCNN (trained with 900 samples)		201	32.87	132	29.23	0.59	0.02	0.98	0	17.36	0.65
U-Net (trained with 1550 samples)		220	24.14	207	58.57	0.49	0.07	0.99	0	18.1	1.44
U-Net (trained with 900 samples)		384	13.14	380	13.07	0.39	0.06	0.99	0	14.77	0.21
eCNN (trained with 1550 samples)	Soft tissue	26	4.59	3	10.71	0.88	0.04	0.98	0	34.79	0.96
eCNN (trained with 900 samples)		28	5.87	−6	14.01	0.87	0.06	0.98	0.01	26.88	0.79
U-Net (trained with 1550 samples)		26	4.9	7	10.55	0.79	0.02	0.98	0	34.11	0.68
U-Net (trained with 900 samples)		102	9.5	98	10.26	0.74	0.04	0.98	0	26.99	0.72

The main objective of this work was to propose a deep learning-based approach featuring an efficient training model using a limited number of training datasets. The aim was to design a deep convolutional neural network enabling robust and effective extraction of key features, enhancing the accuracy of the prediction. This is especially important in applications where the number of training dataset is limited.

The inclusion of data augmentation within the training of the original U-Net and eCNN models led to opposite outcome. Data augmentation enhanced the learning performance of the original U-Net network since the features extracted by the U-Net after data augmentation were invariant to the affine transformations. Conversely, no improvement (if not worse performance) was observed when using data augmentation for the eCNN model. The latter was able to effectively extract distinctive features even prior to the application of data augmentation. Hence, data augmentation did not help the eCCN or added to the complexity of the training, to reach a better solution for the CT synthesis problem. Even after application of data augmentation, the U-net model was outperformed by the eCNN model. However, eCNN exhibited sub-optimal performance after data augmentation. There is no convincing/conclusive explanation for this observation owing to the black-box nature of deep learning approaches. A plausible justification is that eCNN reached its optimal performance before applying data augmentation. Data augmentation not only did not help the network to converge to a more optimal solution, but also disturbed the relatively optimal solution achieved without using data augmentation.

The performance of the eCNN model was compared to an atlas-based approach as well as the U-Net model. The motivation behind this comparison was that previous comparative studies demonstrated the dependable performance of atlas-based methods in the context of CT synthesis for the purpose of MRI-guided treatment planning[19,50] and PET attenuation correction.[35] Hence, the atlas-based technique could serve as baseline for comparison to provide an insight into the overall performance of the proposed model.

5. CONCLUSIONS

A novel eCNN model with efficient learning performance was proposed for the generation of synthetic CT images from MRI. This model relies on a combination of U-net, SegNet, residual connection, and SeLU as activation layer for efficient synthetic CT generation from MR images. The quantitative evaluation revealed promising performance of the proposed method compared to atlas-based techniques. This model exhibited efficient learning capability using only a small number of training dataset, outperforming both the atlas-based method and U-Net model.

ACKNOWLEDGMENTS

This work was supported by the University of Isfahan, the Swiss National Science Foundation under grant SNFN 320030-176052, the Swiss Cancer Research Foundation under Grant KFS-3855-02-2016 and by the Eurostars programme of the European Commission under grant E! 12326 ILLUMINUS.

[a)]Author to whom correspondence should be addressed. Electronic mail: karimian@eng.ui.ac.ir; Telephone: +983137934059.

REFERENCES

1. Das IJ, McGee KP, Tyagi N, Wang H. Role and future of MRI in radiation oncology. *Br J Radiol*. 2019;92:20180505.
2. Mittauer K, Paliwal B, Hill P, et al. A new era of image guidance with magnetic resonance-guided radiation therapy for abdominal and thoracic malignancies. *Cureus*. 2018;10:e2422.
3. Boss A, Bisdas S, Kolb A, et al. Hybrid PET/MRI of intracranial masses: Initial experiences and comparison to PET/CT. *J Nucl Med*. 2010;51:1198–1205.
4. Zaidi H, Becker M. The promise of hybrid PET/MRI: technical advances and clinical applications. *IEEE Sign Process Mag*. 2016;33:67–85.
5. Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol*. 2017;12:28.
6. Mehranian A, Arabi H, Zaidi H. Vision 20/20: magnetic resonance imaging-guided attenuation correction in PET/MRI: challenges, solutions, and opportunities. *Med Phys*. 2016;43:1130–1155.
7. Johnstone E, Wyatt JJ, Henry AM, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only radiation therapy. *Int J Radiat Oncol Biol Phys*. 2018;100:199–217.
8. Chin AL, Lin A, Anamalayil S, Teo BK. Feasibility and limitations of bulk density assignment in MRI for head and neck IMRT treatment planning. *J Appl Clin Med Phys*. 2014;15:100–111.
9. Arabi H, Rager O, Alem A, Varoquaux A, Becker M, Zaidi H. Clinical assessment of MR-guided 3-class and 4-class attenuation correction in PET/MR. *Mol Imaging Biol*. 2015;17:264–276.
10. Paulus DH, Quick HH, Geppert C, et al. Whole-body PET/MR imaging: quantitative evaluation of a novel model-based MR attenuation correction method including bone. *J Nucl Med*. 2015;57:1061–1066.
11. Berker Y, Franke J, Salomon A, et al. MRI-based attenuation correction for hybrid PET/MRI systems: a 4-class tissue segmentation technique using a combined Ultrashort-Echo-Time/Dixon MRI sequence. *J Nucl Med*. 2012;53:796–804.
12. Johansson A, Karlsson M, Nyholm T. CT substitute derived from MRI sequences with ultrashort echo time. *Med Phys*. 2011;38:2708–2714.
13. Keereman V, Fierens Y, Broux T, De Deene Y, Lonneux M, Vandenberghe S. MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences. *J Nucl Med*. 2010;51:812–818.
14. Sekine T, Ter Voert EE, Warnock G, et al. Clinical evaluation of ZTE attenuation correction for brain FDG-PET/MR imaging-comparison with atlas attenuation correction. *J Nucl Med*. 2016;57:1927–1932.
15. Hofmann M, Bezrukov I, Mantlik F, et al. MRI-based attenuation correction for whole-body PET/MRI: quantitative evaluation of segmentation- and Atlas-based methods. *J Nucl Med*. 2011;52:1392–1399.
16. Arabi H, Zaidi H. One registration multi-atlas-based pseudo-CT generation for attenuation correction in PET/MRI. *Eur J Nucl Med Mol Imaging*. 2016;43:2021–2035.
17. Arabi H, Zaidi H. Magnetic resonance imaging-guided attenuation correction in whole-body PET/MRI using a sorted atlas approach. *Med Image Anal*. 2016;31:1–15.
18. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
19. Arabi H, Dowling JA, Burgos N, et al. Comparative study of algorithms for synthetic CT generation from MRI: consequences for MRI-guided radiation planning in the pelvic region. *Med Phys*. 2018;45:5218–5233.
20. Nie D, Trullo R, Lian J, et al. Medical image synthesis with deep convolutional adversarial networks. *IEEE Trans BioMed Eng*. 2018;65:2720–2730.
21. Xiang L, Wang Q, Nie D, et al. Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image. *Med Image Anal*. 2018;47:31–44.
22. Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR); 2017.
23. Ronneberger O, Fischer P, U-net BT. Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical image computing and computer-assisted intervention; 2015, pp. 234–241.
24. Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39:2481–2495.
25. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556; 2014.
26. Alippi C, Disabato S, Roveri M. Moving convolutional neural networks to embedded systems: the alexnet and VGG-16 case. In: Proceedings of the 17th ACM/IEEE International Conference on Information Processing in Sensor Networks; 2018; Porto, Portugal, pp. 212–223.
27. Xie X, Han X, Liao Q, Shi G. Visualization and Pruning of SSD with the base network VGG16. Proceedings of the 2017 International Conference on Deep Learning Technologies; Chengdu, China, June 2017 Pages 90–94.
28. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44:1408–1419.
29. Maspero M, Savenije MH, Dinkla AM, et al. Fast synthetic CT generation with deep learning for general pelvis MR-only Radiotherapy. arXiv preprint arXiv:1802.06468v2; 2018.
30. Leynes AP, Yang J, Wiesinger F, et al. Zero-echo-time and Dixon deep pseudo-CT (ZeDD CT): direct generation of pseudo-CT images for pelvic PET/MRI attenuation correction using deep convolutional neural networks with multiparametric MRI. *J Nucl Med*. 2018;59:852–858.
31. Yang G, Yu S, Dong H, et al. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging*. 2017;37:1310–1321.
32. Seitzer M, Yang G, Schlemper J, et al. Adversarial and perceptual refinement for compressed sensing MRI reconstruction. International conference on medical image computing and computer-assisted intervention; 2018:232–240.
33. Zhu J, Yang G, Lio P. Lesion focused super-resolution. Medical Imaging: Image Processing; 2019 vol. 10949.
34. Zhu J, Yang G, Lio P. How Can We Make GAN Perform Better in Single Medical Image Super-Resolution? A Lesion Focused Multi-Scale Approach. arXiv preprint. *arXiv:1901.03419*; 2019.
35. Arabi H, Zeng G, Zheng G, Zaidi H. Novel adversarial semantic structure deep learning for MRI-guided attenuation correction in brain PET/MRI. *Eur J Nucl Med Mol Imaging*. 2019;46:2746–2759.
36. Bahrami A, Karimian A, Fatemizadeh E, Arabi H, Zaidi H. A novel convolutional neural network with high convergence rate: Application to CT synthesis from MR images. Paper presented at: IEEE Nuclear Science Symposium and Medical Imaging Conference. (NSS, MIC); 26 October 2, . 2019. UK: Manchester; November 2019.
37. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. Paper presented at: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 7–12 June 2015. pp.1-9.
38. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27–30 June 2016, pp 770–778.
39. Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-normalizing neural networks. arXiv. *preprint arXiv:* 1706.02515v5*;* 2017.
40. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; 2010; Proceedings of Machine Learning Research: 249–256.
41. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Paper presented at: Proceedings of the 32nd International Conference on International Conference on Machine Learning2015; Lille, France arXiv preprint arXiv 1502.03167; 2015.
42. Keras CF. The Python Deep Learning library; 2015. GitHub.
43. Malone IB, Ansorge RE, Williams GB, Nestor PJ, Carpenter TA, Fryer TD. Attenuation correction methods suitable for brain imaging with a PET/MRI scanner: a comparison of tissue atlas and template attenuation map approaches. *J Nucl Med*. 2011;52:1142–1149.
44. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: A toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29:196–205.
45. Arabi H, Zaidi H. Comparison of atlas-based techniques for whole-body bone segmentation. *Med Image Anal*. 2017;36:98–112.

46. Martinez-Möller A, Souvatzoglou M, Delso G, et al. Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data. *J Nucl Med*. 2009;50:520–526.
47. Largent A, Barateau A, Nunes J-C, et al. Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning. *Int J Radiat Oncol Biol Phys*. 2019;105:1137–1150.
48. Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. *Med Phys*. 2018;45:3627–3636.
49. Fu J, Yang Y, Singhrao K, et al. Deep learning approaches using 2D and 3D convolutional neural networks for generating male pelvic synthetic computed tomography from magnetic resonance imaging. *Med Phys*. 2019;46(9):3788–3798.
50. Arabi H, Koutsouvelis N, Rouzaud M, Miralbell R, Zaidi H. Atlas-guided generation of pseudo-CT images for MRI-only and hybrid PET–-MRI-guided radiotherapy treatment planning. *Phys Med Biol*. 2016;61:6531–6552.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Full architecture of the eCNN model. The numbers in maxpooling and deconvolution boxes denote spatial resolution reduction and increasing ratio, respectively. The digits shown in each building structure represents the number of filters in each convolutional layer in this structure.
Fig. S2. Architecture of the original U-net model. The digits shown next to each building structure denote the number of filters used in the convolutional layers.
Fig. S3. Qualitative comparison of sCT images generated using the eCNN, U-net and atlas-based methods against standard of reference CT images along with the original input MRI shown in axial, coronal and sagittal planes. (A) Input MRI, (B) standard of reference CT, (C) sCT generated using eCNN, (D) sCT generated using the U-net model, (E) atlas-based synthetic CT, and (F) manual delineation of bladder (white) and rectum (red) obtained from the reference CT image.
Fig. S4. Training and validation losses for the eCNN model with SeLU (red) and ReLU (green) activation functions.
Fig. S5. Training and validation losses for the eCNN model with (red) and without (green) building structure (Residual block).The building structure is replaced with plain 3×3 convolutional layer.
Fig. S6. Training and validation losses for eCNN, Fu and Han models within 200 epochs.
Fig. S7. (a) Original MRI, (b) reference CT, (c) synthetic CT generated using eCNN model trained with 1550 samples (full training dataset), (d) synthetic CT generated using eCNN model trained with 900 samples (reduced training dataset), (e) synthetic CT generated using U-net model trained with 1550 samples and (f) synthetic CT generated using the U-net model trained with 900 samples.
Table S1. Summary of the quantitative metrics, including MAE, ME, PCC, SSIM, and PSNR measured in the whole pelvis region, air, bone, and soft-tissue obtained from the different variations of the eCNN model for the four external patients (512 2D images).
Table S2. Summary of the quantitative metrics, including MAE, ME, PCC, SSIM, and PSNR in the whole pelvis region, air, bone, and soft-tissue obtained from the different models for the four extra patients (512 2D images).