A novel convolutional neural network with high convergence rate: Application to CT synthesis from MR images

Abbas Bahrami, Alireza Karimian, Emad Fatemizadeh, Hossein Arabi and Habib Zaidi *IEEE Fellow*

Abstract– **Synthetic CT (sCT) generation from MR images is yet one of the major challenges in the context of MR-guided radiation planning as well as quantitative PET/MR imaging. Deep convolutional neural networks have recently gained special interest in large range of medical imaging applications including segmentation and image synthesis. In this study, a novel deep convolutional neural network (DCNN) model is presented for synthetic CT generation from single T1-weighted MR image. The proposed method has the merit of highly accelerated convergence rate suitable for applications where the number of training dataset is limited while highly robust model is required. This algorithm exploits a Visual Geometry Group (VGG16) model without fully connected layer coupled to a residual network in the form of encoder-decoder structure. The training of the proposed algorithm was performed using pelvis image of only 15 patients in a five-fold cross-validation scheme. No network pre-training and data augmentation was used. The outcome of the proposed algorithm was compared to an atlas-based approach in terms of accuracy of CT intensity estimation within different body tissues. After only 100 epochs, the proposed algorithm resulted in mean absolute error (MAE) and mean error (ME) of 40.64 ± 12.66 and -2.80 ± 10.98 (HU) for the entire pelvis region, respectively. While atlas-based method led to MAE of 82.06 ± 52.59 and ME of -13.00 ± 60.19 (HU). Within the soft-tissue the atlas-based method and the proposed algorithm achieved MAEs of 51.01±14.3 and 25.51±7.72 (HU), respectively. Likewise in bony tissue, MAEs of 212.65±78.45 and 221.99±76.28 (HU) were obtained when using DCNN and atlas-based methods, respectively. The proposed algorithm showed superior performance to the atlas-based method with only relying on a limited number of training subjects. The proposed algorithm is suitable for the clinical applications where accurate models are required while accessing a large number of training cases is limited.**

I. INTRODUCTION

MRI-guided attenuation correction on PET/MRI systems and radiation planning require reliable algorithms for synthetic CT (sCT) generation from MRI [1, 2]. The primary aim of this work is to develop a deep convolutional neural network with high convergence rate capability to enable robust training on a limited number of subjects. Such a model would be of interest in applications where collecting a large number of training dataset is not feasible. We evaluate the

proposed algorithm in the context of CT synthesis from MRI. The proposed method differs from previous techniques in the sense that it is able to generate comparable sCT images from a single MRI sequence using only few training subjects (exp. 12 subjects) [3]. The proposed DCNN algorithm is based on a Visual Geometry Group (VGG16) model without fully connected layers coupled to a residual network. This model was implemented in an encoder-decoder structure where some high resolution features from the encoding network are linked to the decoding counterpart. Fifteen aligned CT/MRI image pairs of male subjects in the pelvic region were used for the training and evaluation using a five-fold cross-validation scheme. The proposed algorithm was compared to an atlas-based synthetic CT generation method involving pairwise registration of the atlas dataset to the target MRI in a one-leave-out cross-validation scheme [4, 5]. The accuracy of CT value estimation was evaluated for different tissue types in the pelvic region. Comparison to a previously developed atlas-based method provides valuable insight into the accuracy of the proposed model for synthetic CT generation.

II. MATERIAL AND METHODS

A. Clinical CT and MRI studies

CT and MRI pelvic scans of 15 patients were retrospectively employed for the quantitative evaluation of the DCNN sCT generation technique. The patients underwent a 3D T1-weighted magnetization prepared rapid gradient-echo. In a separate session, the patients underwent a CT scan (120 kVp, 400 mA) on a Siemens 64-slice Sensation scanner. The CT images were acquired at 1.5 mm slice thickness and in plane resolution of 0.85×0.85 mm.

B. Proposed deep learning-based approach

Overall, the architecture of the proposed model is inspired from the U-net model [6] and the encoder-decoder network [7]. For the encoder network, convolutional layers similar to VGG16 model were employed. However, due to the complexity of anatomical structures in MR images, the convolutional layers in the VGG16 model were extended to achieve deeper network. Practically, increasing the depth of the network by simply adding series of convolutional layers coupled to ReLu, Maxpooling and batch normalization is not efficient enough. Such a network would suffer from overfitting and gradient vanishing or exploding issues. To overcome these issues, the convolutional layers in the VGG16 architecture were replaced with a bottleneck structure (referred to as Building Structure) as described in [8]. The modified VGG architecture features accurate and fast convergence rate on a small

dataset. Equation 1 shows the formulation of a modified layer where X_i is the input vector of the layer number i, Y_i is the output vector of the layer and ReLu is the rectified linear unit.

$$Y_i = ReLu\left(ReLu\left(wi2\left(ReLu(wi1 * X_i + bi1) + bi2\right)\right) + X_i\right) \quad (1)$$

w_{i1} and w_{i2} denote the learning weights and b_{i1} and b_{i2} are the bias in each building structure. Fig. 1A illustrates the building structure, where f indicates the number of the filter, which is similar to that of VGG16. At the upper layers (high resolution), the pooling mask indices were incorporated at the corresponding decoding network as a shortcut connection.
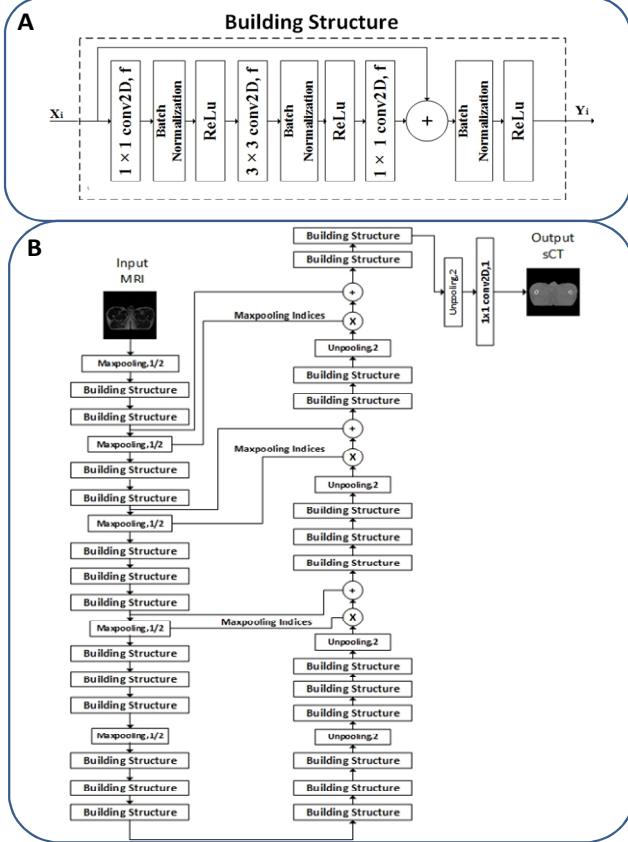

Fig. 1. A) The design of building structure. B) The acrchitecture of the proposed model (DCNN).

In total, the encoding network contains 92 2D convolutional layers. As shown in Fig.1.B, before each ReLu layer a batch normalization layer is inserted to reduce the internal covariate shifts and improving the convergence rate of the model. The decoding network is the mirror of the encoding network except that downsampling in the encoding network is replaced with a 2D deconvolutional layer with 2×2 window and stride size of 2 for upsampling. There are two major modifications in this model. First, the insertion of a maxpooling layer (2×2 window and stride 2 for downsampling of the input image) before the first convolutional layer and a deconvolution operator for upsampling before the last 1×1 convolutional layer. Second, the intermediate maxpooling inputs in the encoding network are connected to the input of corresponding layers in the decoding network to enhance the accuracy of CT synthesis. The mean absolute error was used as the loss function using the Adam stochastic optimization method. This model has 41,707,713 trainable parameters in total. No data augmentation and pre-training was used in this work. The training was

performed at maximum in 100 epochs. The 3D images of the 15 patients were converted to 1800 2D 256×256 slices, where 1500 slices were used as training dataset in a five-fold cross-validation scheme.

C. Performance evaluation

The performance of the proposed algorithm was compared with a conventional atlas based method using a leave-one-out cross-validation scheme [9]. To this end, the MR atlas images were pairwise registered to the target MR image. The obtained transformation maps were used to map the corresponding CT atlas images. The final atlas-based sCT was generated by taking the average of all transformed CT atlas images. The CT value estimation was evaluated through comparison between voxel intensity of ground truth CT and the sCT images. The bone and soft-tissue were segmented from synthetic CTs and ground CT images using intensity thresholding. Then, the mean error (ME) and mean absolute error (MAE) were calculated between the ground truth CT (R_{CT}) and sCT (A_{CT}) for whole pelvic region as well as for bone and soft-tissue.

$$ME_{CT} = \frac{1}{P}\sum_{i=1}^{P}(A_{CT}(i) - R_{CT}(i)) \quad (2)$$

$$MAE_{CT} = \frac{1}{P}\sum_{i=1}^{P}|A_{CT}(i) - R_{CT}(i)| \quad (3)$$

The structural similarity index (SSIM) was calculated between the ground truth and synthetic CTs using Eq. 4, where M_R and M_A denote the mean value of the images R_{CT} and A_{CT}, respectively. $\delta_{R,A}$ indicates the covariance and δ_R and δ_A are the variance of the M_R and M_A images, respectively. K_1=0.01 and K_2=0.02 are the constant parameters.

$$SSIM = \frac{(2M_R M_A + K_1)(2\delta_{R,A} + K_2)}{(M_R^2 + M_A^2 + K_1)(\delta_R^2 + \delta_A^2 + K_2)} \quad (4)$$
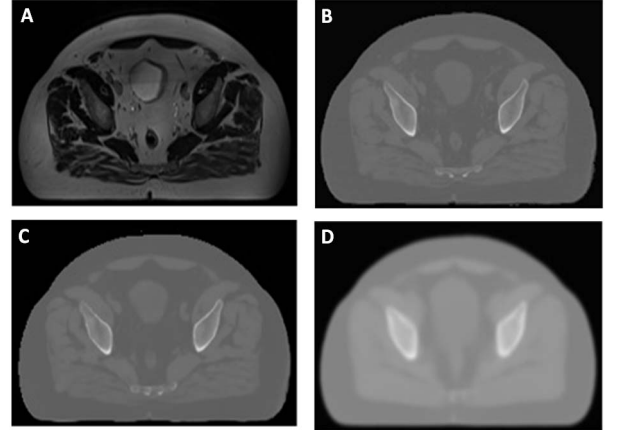
III. RESULTS AND DISCUSSION

Fig. 2. A) Target MR. B) Reference CT. C) Deep learning-based sCT. D) Atlas-based sCT.

Representative slices of the synthetic CT images are shown in Fig. 2 together with ground truth CT and MR images. For the whole pelvic region, the DCNN method resulted in a ME±SD (MAE±SD) of -2.80±10.98 (40.64±12.66) HU while the Atlas method led to -13.00±60.19 (82.06±52.59). The accuracy of CT value estimation for the whole pelvic region is presented in Table 1. Likewise, Table 2 provides the accuracy of CT value estimation within bone and soft-tissue.

TABLE 1. ACCURACY OF CT VALUE ESTIMATION USING THE
ATLAS AND DCNN METHODS.

Whole pelvis	MAE(HU)	ME(HU)	SSIM
DCNN	40.64±12.66	-2.80±10.98	0.95±0.02
Atlas	82.06±52.59	-13.00±60.19	0.91±0.05

TABLE 2. ACCURACY OF CT VALUE ESTIMATION WITHIN BONE
AND SOFT-TISSUE.

Bone	MAE(HU)	ME(HU)
DCNN	212.65±78.45	144.59±83.5
Atlas	221.99±76.28	168.41±83.66
Soft-tissue	MAE(HU)	ME(HU)
DCNN	25.51±7.72	-5.07±10.3
Atlas	51.01±14.3	-12.45±22.72

The quantitative evaluation of the synthetic CT images demonstrated the superior performance of the DCNN algorithm with respect to the atlas method. The visual inspection of the generated synthetic CT (Fig. 2) also showed the ability of the proposed method to extract fine details from the MR image.

IV. Conclusions

The proposed deep convolutional neural network-based DCNN algorithm is capable of generating superior synthetic CTs images compared to the atlas-based method while providing high convergence rate, thereby enabling efficient training on small datasets.

References

[1] A. Mehranian, H. Arabi, and H. Zaidi, "Vision 20/20: Magnetic resonance imaging-guided attenuation correction in PET/MRI: Challenges, solutions, and opportunities," *Med Phys,* vol. 43, pp. 1130-55, 2016.

[2] H. Arabi, J. A. Dowling, N. Burgos, X. Han, P. B. Greer, N. Koutsouvelis*, et al.*, "Comparative study of algorithms for synthetic CT generation from MRI: Consequences for MRI-guided radiation planning in the pelvic region," *Med Phys,* vol. 45, pp. 5218-5233, 2018.

[3] H. Arabi, G. Zeng, G. Zheng, and H. Zaidi, "Novel adversarial semantic structure deep learning for MRI-guided attenuation correction in brain PET/MRI," *Eur J Nucl Med Mol Imaging,* vol. 46, pp. 2746-2759, 2019

[4] H. Arabi, N. Koutsouvelis, M. Rouzaud, R. Miralbell, and H. Zaidi, "Atlas-guided generation of pseudo-CT images for MRI-only and hybrid PET–MRI-guided radiotherapy treatment planning," *Phys Med Biol,* vol. 61, p. 6531-6552, 2016.

[5] H. Arabi and H. Zaidi, "Comparison of atlas-based techniques for whole-body bone segmentation," *Med Image Anal,* vol. 36, pp. 98-112, 2017.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234-241.

[7] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans Pattern Anal Mach Intell* vol. 39, pp. 2481-2495, 2017.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[9] H. Arabi and H. Zaidi, "One registration multi-atlas-based pseudo-CT generation for attenuation correction in PET/MRI," *Eur J Nucl Med Mol Imaging,* vol. 43, pp. 2021-35, 2016.