Improved Wasserstein GAN based Methods for pseudo-CT Synthesis from MR DIXON Images: Application to MR only Radiotherapy

P. Ceccaldi[1], S. Biswas[2], H. Chandarana[3,4], H. Wang[5], I. Das[5], M. Fenchel[6], H. Bhat[6], B. Odry[6], M.S. Nadar[6]

[1]Siemens Healthcare, LTD, [2]University of Iowa, [3]Center for Advanced Imaging Innovation and Research (CAI2R), Department of Radiology, New York University School of Medicine, New York, NY, USA, [4]Bernard and Irene Schwartz Center for Biomedical Imaging, Department of Radiology, New York University School of Medicine, New York, NY, USA, [5]Department of Radiation Oncology, New York University School of Medicine, & Laura and Isaac Perlmutter Cancer Center, New York, NY, USA, [6]Siemens Medical Solutions, Inc.

Introduction: There is interest in using deep learning methods for developing CT like images (pseudo-CT) directly from MRI for radiation dose planning. This will enable in MR-only treatment planning for radiation therapy. Recent work has explored deep learning generative networks to synthesize pseudo-CT images from MR Dixon [1,2] images of brain and pelvis. We propose to develop and evaluate three Deep Learning methods based on Generative Adversarial Networks (GAN) for imaging of the abdomen and assess the feasibility of using pseudo-CT images for MR-based radiation treatment planning for abdominal malignancy.

Methods: We present three 2D Image to Image GAN methods that are composed of an Encoder-Decoder network G that generates pseudo-CTs, and a Discriminator Network D (DenseNet) to differentiate original from pseudo-CTs: 1) a modified Pix2Pix, based on the Pix2Pix[3] architecture, which consists of a conditional GAN used for image translation, thereby providing MR Dixon images to the discriminator D, in addition to the pseudo or real CT. We brought some improvements to the original architecture by tweaking the network parameters, but mainly by using a Wasserstein GAN [4] loss to improve the stability of the training while gaining details in the generation. 2) Our second approach (FCD) still is GAN based using a Fully Convolutional Dense network [5] (FCD) as our generator. The FCD is a combination of the U-Net [4] architecture with the DenseNet [5] technique. 3) A Context based approach that is an extension of the FCD. Before the generator G, we insert an initial FCD to produce a preliminary segmentation of the bones from the MR Dixon images. This segmentation serves as contextual information, to facilitate the synthesis of the bones in the produced pseudo-CT. Ground truth for bones segmentation is obtained by thresholding the original CT. We trained all our models using two MR Dixon images, In-phase and Out-of-phase (excluding water and fat images), as a two-channel input and regressed to the corresponding full ranged CT. The adversarial loss is defined as: $\max_f E_{\{x\sim realCT\}}[f(x)] - E_{\{x\sim synCT\}}[f(x)] + \lambda_{\{L_1\}}\, E_{\{x,y\}}||y - G(x)||_1 + \lambda_{\{gp\}} E_{\{x\sim\hat{x}\}}\left(\nabla_{\{\hat{x}\}} f(\hat{x}) - 1\right)^2$, with the first term expressing the Wasserstein loss; f is a discriminator function that satisfies the Lipchitz constraint $||f||_L \leq 1$, the second term is a L_1 penalty applied on G [4ref], and the third term being the gradient penalty that enforces the Lipchitz constraint for f [6]. The loss used for the contextual FCD producing the bone segmentation is the categorical cross entropy applied after a softmax output. IRB was obtained for twenty-two patients randomly selected who underwent a research whole-body PET/MR in addition to a clinical PET/CT exam. MR Dixon images were based on a 3D spoiled steady state sequence (VIBE) acquired at 3T with a dual in-and opposed-phase echo times (1.23 ms and 2.46 ms). Further acquisition parameters: TR 2.6ms, FoV 500x328mm, matrix 192x126, BW 960 Hz/px, flip angle 10 deg). MR images were acquired on a Siemens Biograph mMR system. CT was acquired according to a standard clinical protocol on a PET/CT scanner (Biograph mCT, Siemens AG Healthcare). The CT acquisition used a setting of 100–140 kVp and had a pixel size of 1.52 × 1.52 mm2 or 1.37 × 1.37 mm2 with a slice thickness of 5.0 mm. MR Dixon In-phase and Out-of-phase, and CT images were co-registered [7] (and resampled to 2.5mm x 2.5mm x 3.2mm. Occasional residual alignment errors still could be observed in some areas in cases of larger differences in respiratory motion or patient positioning than the registration was able to compensate, e.g. arms up in the CT and arms down in the MR. Fat and Water Dixon maps were not used because of remnant fat/water tissue swaps in the Dixon images. 14 patients were used for training (9000 slices after data augmentation), and 2 patients for testing. SSIM and MSE were utilized as metrics for evaluation of quality of pseudoCT for each model.

Results: Computed SSIM (MSE) were 95.40 +/- 05.02 (52.92 +/- 25.05) for our modified Pix2Pix model, 95.59 +/-02.3 (60.10 +/-18.56) for our FCD model and 96.71 +/- 1.89 (45.19 +/- 15.47) for our Context based approach. Figure 1 shows that bone synthesis appears sharper using the Context based approach while tissues are better captured with the FCD. This indicates an expected bias towards the bone with the context network. To address radiation therapy planning, we generated Digitally Reconstructed Radiographs (DRR) from both the synthetic and original CTs and evaluated the simulated dose for each of the generated pseudoCTs at several regions of the body. Figure 2 shows examples of DRRs from each method and the original CT. Mean Dose for PTV and OAR was 105.2%, 15.8%, 2.8% for original CT based plan, 105%, 15.5%, 2.9% for Context-based pseudo-CT plan, 105.3%, 15.8%, 2.9% for FCD based pseudo-CT plan and 105.4%, 15.8%, 2.8% for modified Pix2Pix-based pseudo-CT plan.

Conclusions: We evaluated three GAN based methods for image translation and applied it to pseudo-CT synthesis from MR Dixon images. With In-phase and Out-of-phase Dixon images, we are able to show very good translation results based on SSIM and MSE metrics, with the Context based FCD network getting an average error of 45.19 HU, by adding focus on the bones. Ultimately, radiation therapy planning based on synthetic pseudo-CT shows very little difference between the simulated radiation doses, which is a promising result towards adoption of MR based dosimetry for radiation therapy planning. Future work with training on all Dixon contrasts, more cases and improved registrations should help to outperform these results. This feature is based on research, and is not commercially available. Due to regulatory reasons, its future availability cannot be guaranteed.

References: [1] Maspero et al "Fast synthetic CT generation with deep learning for general pelvis MR-only radiotherapy" arXiv1802.06468. [2] Wolterink et al. "Deep MR to CT Synthesis using Unpaired Data" arXiv:1708.01155 [3] Isola et al "Image-to-Image Translation with Conditional Adversarial Networks" arXiv:1611.07004. [4] Arjovsky et al "Wasserstein GAN" arXiv:1701.07875. [5] Jegou et al "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation" arXiv:1611.09326. [6]Gulrajani et al. "Improved Training of Wasserstein GANs" arXiv:1704.00028. [7] Hermosillo et al "Variational methods for multimodal image matching" Int J Comput Vis 2002; 50:329-343

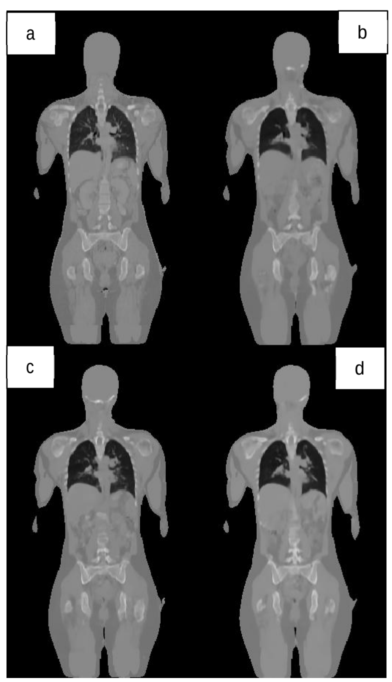

Figure 1: Comparison of original CT (a), modified pix2pix (b), FCD (c), Context based (d). Bones are better defined in the spine and shoulders for Context.

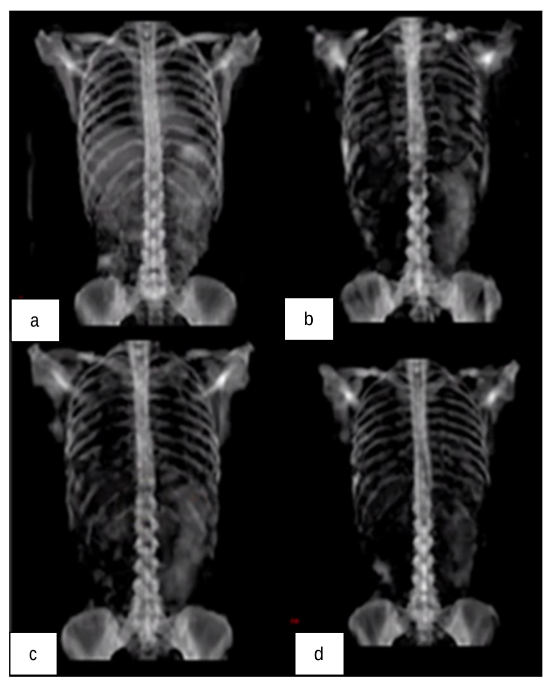

Figure 2: Comparison of DDR from original CT (a), modified Pix2Pix (b), FCD (c) and Context (d)