2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)
April 3-7, 2020, Iowa City, Iowa, USA

MRI TO CT SYNTHESIS OF THE LUMBAR SPINE FROM A PSEUDO-3D CYCLE GAN

Reda Oulbacha[⋆] *and Samuel Kadoury*[⋆†]

[⋆]MedICAL Laboratory, Polytechnique Montreal, Canada
[†] CHU Sainte-Justine Research Center, Montreal, Canada

ABSTRACT

In this paper, we introduce a fully unsupervised approach for the synthesis of CT images of the lumbar spine, used for image-guided surgical procedures, from a T2-weighted MRI acquired for diagnostic purposes. Our approach makes use of a trainable pre-processing pipeline using a low-capacity fully convolutional network, to normalize the input MRI data, in cascade with FC-ResNets, to segment the vertebral bodies and pedicles. A pseudo-3D Cycle GAN architecture is proposed to include neighboring slices in the synthesis process, along with a cyclic loss function ensuring consistency between MRI and CT synthesis. Clinical experiments were performed on the SpineWeb dataset, totalling 18 patients with both MRI and CT. Quantitative comparison to expert CT segmentations yields an average Dice score of 83 ± 1.6 on synthetic CTs, while a comparison to CT annotations yielded a landmark localization error of 2.2 ± 1.4mm. Intensity distributions and mean absolute errors in Hounsfield units also show promising results, illustrating the strong potential and versatility of the pipeline by achieving clinically viable CT scans which can be used for surgical guidance.

Index Terms— CT-MRI synthesis, Cycle GAN, Lumbar spine, Pseudo-3D, Image-guided spine surgery

1. INTRODUCTION

Accurate planning and delineation of vertebrae from diagnostic imaging is a crucial preliminary task in image-guided spine surgery. Describing in detail the morphology of the vertebral anatomy during the planning stages can be of significant value to plan pedicle screw trajectories by locating critical nerves to avoid, and identify potential complications for inserting pedicle screws in the cortical bone.

In the past decade, magnetic resonance imaging (MRI) has become the standard for pre-operative imaging which is used not only for diagnosis but also for planning and assessment purposes, as it is typically acquired as routine during patient evaluation. Computed tomography (CT) is used only for specific surgical indications when image-guided surgery is performed. However, it involves additional radiation exposure and is avoided as much as possible.

On the other hand, integrating MRI for surgical applications, particularly those guided with fluoroscopy, presents several challenges as it exhibits higher variability in spatial resolution with different intensity distribution from different manufacturers and clinical sites. Over the recent years, deep learning approaches, and particularly Generative Adversarial Networks (GANs), have become the tool of predilection for tackling challenging image synthesis and domain (multimodal) adaptation problems. For the specific problem of MRI-based pseudo-CT generation, studies using classical machine learning algorithms such as the K-Nearest Neighbours [1] and the random forest regression [2] were proposed. More studies leveraging deep learning [3] were also proposed. However, image synthesis methods were primarily developed for the generation of medical images with normal anatomy [4], and provided single-slice prediction without incorporating the context from the neighbouring three dimensional anatomical context.

In this work, we propose an automated MRI to CT synthesis method of the lumbar spine from a pseudo-3D Cycle GAN architecture, to train a model that learns the structure of vertebral bodies and pedicles from a training set of MR images and corresponding CT images, which uses contextual information with a pseudo-3D inference from neighboring slices.

2. MATERIALS AND METHODS

2.1. Data

The dataset consists of MRI and CT images of the lumbar spine from 18 different patients, obtained from the public SpineWeb library (http://spineweb.digitalimaginggroup.ca). The vertebral bodies and pedicles were automatically segmented with a trainable pre-processing pipeline using FC-ResNets followed by a low-capacity fully convolutional network (FCN) which acts as a normalizing pre-processor of the MR data. The normalized data is then iteratively refined by an FC-ResNet to segment the vertebral bodies and pedicles [5]. For training purposes, volumes were then cropped around the spine area, rigidly registered using normalized mutual information and resampled in the same physical space

Funding provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Authorized licensed use limited to: Makerere University Library. Downloaded on July 29,2023 at 06:42:15 UTC from IEEE Xplore. Restrictions apply.

with a sagittal slice thickness of 1.5mm. Finally, MRIs were standardized to have zero mean and unit variance and we then downscaled the MRIs and CTs to the open range $(-1, 1)$ using the arbitrary ranges $[-10, 10]$ and $[-1000, 4000]$ respectively, as in [3].

2.2. MR-CT Synthesis Network Architecture

We used a Cycle GAN architecture inspired from the original design [6], using a 9-Residual Block ResNet Generator (Table 1) and a Fully Convolutional 64x64 Patch discriminator (Table 2). Unlike the original network which was designed for individual 2D images, we propose to train on a 3D volume consisting of 4 neighbouring 2D slices stacked together along the channel dimensions, forming a thickened slice. One benefit of such an approach is to capture nearby 3D information without resorting to the memory-heavy 3D convolutions. A volume of N singleton slices equivalates to $N - 3$ thickened slices. To recombine the thickened slices into a volume of N singleton slices, we take the first three slices of the thick slice. For each of following thickened slices, we chose to keep only the 3rd slice. Finally, the last thickened slice is created from remaining slices to complete a total of N slices. This recombination strategy ensures that all the slices that are not in the extremities come from thick slices with 3 overlapping slices, allowing a continuity in the shared 3D information.

Fig. 1: Overview of the pseudo-3D Cycle GAN with thickened slices. G_{CT} and G_{MRI} are the MRI $\rightarrow$ CT and CT $\rightarrow$ MRI generative networks respectively, while D_{CT} and D_{MRI} are the CT and MRI discriminative networks. Left to right path is the *Real MRI* $\rightarrow$ *Synthetic CT* $\rightarrow$ *Reconstructed MRI* cycle, while the right to left path is the *Real CT* $\rightarrow$ *Synthetic MRI* $\rightarrow$ *Reconstructed CT* cycle.

For the proposed network, we used Instance Normalization without learnable affine parameters, using Dropout with an activation probability of 0.5. The residual blocks are the same as those from [6]. All LeakyReLUs in the discriminator had a slope of 0.2. Due to the strides and the fully convolutional nature of the network, it performs 64x64 sliding window discrimination using a stride of 16.

Table 1: 9-Block Resnet Architecture Generator

Block	Description
2D Reflection Padding	Padding size 3
Conv - InstanceNorm - ReLU	64 7x7 kernels, stride 1, pad 0
Conv - InstanceNorm - ReLU	128 3x3 kernels, stride 2, pad 1
Conv - InstanceNorm - ReLU	256 3x3 kernels, stride 2, pad 1
9 x Residual block	-
ConvTranspose - InstanceNorm - ReLU	128 3x3 kernels, stride 2, pad 1
ConvTranspose - InstanceNorm - ReLU	64 3x3 kernels, stride 2, pad 1
2D Reflection Padding	Padding size 3
Conv - Tanh	7x7 kernels, stride 1, pad 0

Table 2: 64x64 Fully Convolutional Discriminator

Block	Description
Conv - Dropout - LeakyReLU	64 4x4 kernels, stride 2, pad 1
Conv - InstanceNorm - Dropout - LeakyReLU	128 4x4 kernels, stride 2, pad 1
Conv - InstanceNorm - Dropout - LeakyReLU	256 4x4 kernels, stride 2, pad 1
Conv - InstanceNorm - Dropout - LeakyReLU	512 4x4 kernels, stride 2, pad 1
Conv - Sigmoid	1 4x4 kernel, stride 1, pad 0

2.3. Training Procedure

We separate the loss between the generators and discriminator such that:

$$\begin{aligned}\mathcal{L}(G_{CT}, G_{MRI}) = &\|D_{CT}\left(G_{CT}(MRI)\right) - 1\|_2^2 + \\ &\|D_{MRI}\left(G_{MRI}(CT)\right) - 1\|_2^2 + \\ &|G_{CT}\left(G_{MRI}(CT)\right) - CT|_1 + \\ &|G_{MRI}\left(G_{CT}(MRI)\right) - MRI|_1\end{aligned} \tag{1}$$

$$\begin{aligned}\mathcal{L}(D_{CT}, D_{MRI}) = &\|D_{CT}\left(CT\right) - 1\|_2^2 + \\ &\|D_{MRI}\left(MRI\right)) - 1\|_2^2 + \\ &\|D_{CT}\left(G_{CT}(MRI)\right)\|_2^2 + \\ &\|D_{MRI}\left(G_{MRI}(CT)\right)\|_2^2\end{aligned} \tag{2}$$

From Eqs.(1) and (2) above, D_{CT} and D_{MRI} designate the CT and MRI discriminators respectively, while G_{CT} and G_{MRI} designate the MRI to CT and CT to MRI generators, respectively. At each iteration, we minimize Eq.(1) with respect to the generators, and then minimize Eq.(2) with respect to the discriminators. We used 600 epochs for training with a batch size of 1, a learning rate of 0.0002 linearly decaying to 0 starting as of epoch 200. We use the Adam optimizer for all networks with β_1 of 0.2. For data augmentation, we perform random flipping and cropping with sizes from 25% to 100% of the original sagittal images. All sagittal images were resized to [256x256] before being fed to the networks. The discriminator and generator were trained separately at each iteration, and the discriminator's loss was scaled by a factor of 0.5. We also trained the discriminator with images randomly sampled from a history of the latest 50 images produced by the generator for regularization, as in [7]. At inference time, we set Dropout to evaluation mode as opposed to the original network, as keeping it introduces noise in the axial and coronal planes when grouping the slices together to form a volume.

Authorized licensed use limited to: Makerere University Library. Downloaded on July 29,2023 at 06:42:15 UTC from IEEE Xplore. Restrictions apply.

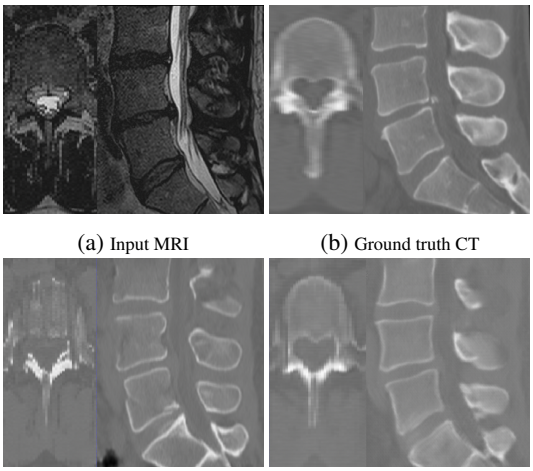
2.4. Validation Procedure

During training, generated samples were monitored and qualitative validation was performed on held-out samples for model hyper-parameter selection. Although the MRIs and CTs were registered, perfect correspondence is difficult to achieve due to spine movement from MRI to CT, which hinders the voxelwise validation metric. The same applies for the GAN loss function, as it is a well established fact that the adversarial loss does not perfectly correlate with the visual quality of the produced images. After training, we performed a cross-validation split into 15-3 for training and testing, giving 6 different splits for which we trained 6 different models and aggregate their output on the test set, yielding a total of 18 volumes. All visible vertebrae were annotated with 6 landmarks each (pedicles and vertibral body), and registered the synthetic CT to the real CT with for each vertebra. As the real CTs had segmentations of the vertebral bodies and pedicles, we trained a U-Net on that data and produced predictions on the synthetic CT dataset to compare with real CT segmentations. To ensure that the ground truth segmentations did not overfit the training data, we trained the U-Net on patches rather than on entire slices. A final visual inspection and manual fine tuning of each vertebral segmentation was performed to ensure the segmentations corresponded to the visible structure on the synthetic images. We used the segmentations to compute the Dice score, and validated the technique with landmark-based localization errors. Finally, we show histogram distribution and voxel intensity comparison metrics.

3. RESULTS AND DISCUSSION

3.1. Pseudo-3D and Single-slice Comparison

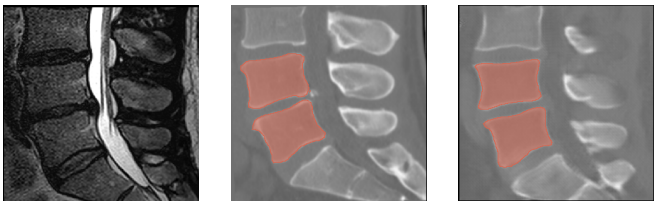
A five-fold cross-validation procedure was performed, where at each fold the pseudo-3D GAN model was re-trained to fine-tune the hyper-parameters, producing predictions for all 18 test MR volumes. Results were compared to the model in the study proposed by [3], which is a 2D single-slice Cycle GAN model, to assess the effects of a pseudo-3D approach for image synthesis as opposed to a more classical 2D approach. For all patients, the pseudo-3D model produced volumes which were visually coherent, both in 2D and 3D, while the single slice model volumes were unusable in 6 out of the 18 patients. For the single-slice model, the remaining 12 volumes were still noisy in the axial and coronal planes, due to the assumption that the sagittal slices are all independent. Figure 2 shows results where we clearly notice an improvement in the axial direction for the pseudo-3D model compared to the single-slice Cycle GAN model. The sagittal slices are also smoothed with continuous vertebrae contours. Integrating a pseudo-3D approach in fact augments the training data that is fed to the network, as each slice in the thickened volume can have several positions within the quadruplet. The model can thus see different combinations and permutations.



(a) Input MRI (b) Ground truth CT

(c) Single-slice model prediction (Wolterink et al., 2017) (d) Pseudo-3D model prediction (Ours)

Fig. 2: Pseudo-3D Cycle GAN and single-slice Cycle GAN model output comparison along both sagittal and axial planes.



(a) Input MRI (b) Ground truth CT (c) Synthetic CT

Fig. 3: Left to right: input MRI, manually segmented ground truth CT and automatically segmented pseudo-3D synthetic CT.

3.2. Quantitative Validation

Having established that the single-slice Cycle GAN model does not produce usable 3D volumes, we chose to move on with the pseudo-3D approach for a quantitative evaluation through landmark-based registration followed up by an evaluation of segmentation accuracy of vertebrae on CT based on Dice scores to assess the predicted vertebral morphology. Despite slight inaccuracies in voxelwise correspondence with the ground truth due to pose change from MRI to CT, we computed some comparison metrics for both voxels and histograms. From Figure 3, we can observe that the FC-ResNet was able to reproduce the segmentation trained on the synthetic data. As demonstrated with the segmentations which seem visually accurate, we obtained an average dice 0.83 ± 0.16 as presented in Figure 4, which is in the expected range given the intrinsic differences of volume for the vertebrae between MRI and CT. The landmark error between real and

Authorized licensed use limited to: Makerere University Library. Downloaded on July 29,2023 at 06:42:15 UTC from IEEE Xplore. Restrictions apply.

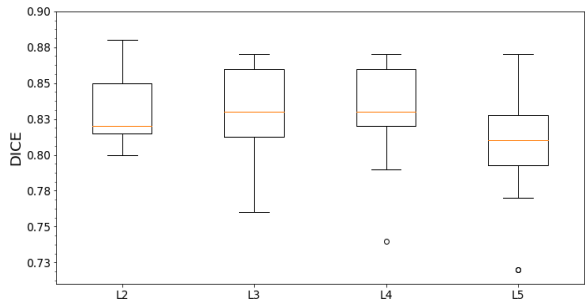


Fig. 4: Dice score coefficients for each vertebra level.

Table 3: MAE (in HU) and NHI ratio computed between the real and synthetic CT volumes, per patient.

Volume	MAE	NHI	Volume	MAE	NHI
1	114.3	0.90	11	141.1	0.77
2	131.3	0.92	12	129.4	0.80
3	120.3	0.83	13	134.9	0.73
4	130.3	0.80	14	138.1	0.78
5	111.3	0.78	15	111.3	0.81
6	105.9	0.83	16	119.0	0.86
7	181.0	0.83	17	93.0	0.87
8	153.9	0.73	18	148.4	0.74
9	126.3	0.83	Average MAE	125.65 ± 10.07	
10	139.0	0.85	Average NHI	0.82 ± 0.03	

synthetic CT was 2.2 ± 1.4mm. For a comparison in intensity distributions, we use the mean absolute error (MAE) as our voxel metric, as well the normalized histogram intersection (NHI). Table 3 presents the average values. Although the NHI seems satisfying, and Figure 5 shows a good correspondence in histogram shape, Figure 6 illustrates promising difference maps in Hounsfield units between real and synthesized CT images. When compared to previous work from [3], one can observe that that the HU errors in their study stem primarily from regions of bone structure where the signal varies abruptly, which is what we observe in figure 6 too. In fact, given that our data is focused on the lumbar spine, with larger bone anatomy, the higher MAE is expected.

4. CONCLUSION

We presented a fully unsupervised approach for MRI-based synthesis of CT images of the lumbar spine from T2-weighted MR images. We first showed the significant differences in image quality between a classical Cycle GAN approach, where slices are assumed to be independent, and a pseudo-3D approach, which takes into account each slice's neighbouring spatial information. From a quantitative analysis of our approach, using both segmentations and manual landmark annotations, the proposed method demonstrates a good potential for surgical guidance workflows.

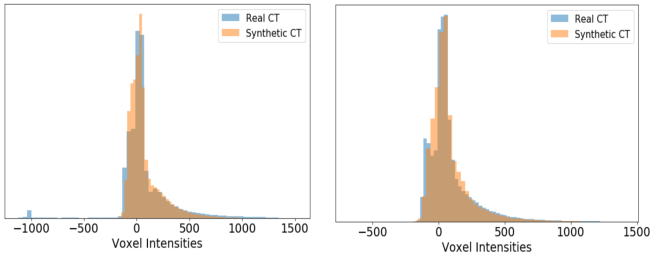


Fig. 5: Sample histogram distributions for two patient images.

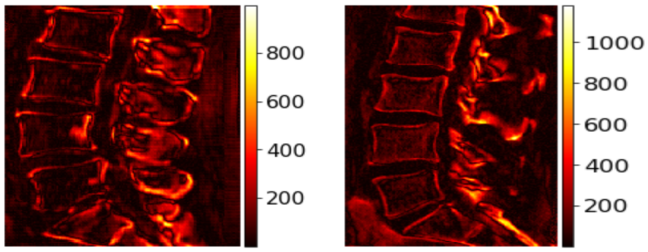


Fig. 6: Absolute error maps (in HU) from two different synthesized patient images.

5. REFERENCES

[1] M.J Van der Bom *et al.*, "Registration of 2d x-ray images to 3d mri by generating pseudo-ct data," *Physics in Medicine & Biology*, vol. 56, no. 4, pp. 1031, 2011.

[2] T. Huynh *et al.*, "Estimating ct image from mri data using structured random forest and auto-context model," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 174–183, 2015.

[3] J.M Wolterink *et al.*, "Deep mr to ct synthesis using unpaired data," in *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2017, pp. 14–23.

[4] AF Frangi, SA Tsaftaris, and JL Prince, "Simulation and synthesis in medical imaging," *IEEE Trans. on Med. Imag.*, vol. 37, no. 3, pp. 673, 2018.

[5] M Shakeri, I Nahle, E Finley, and S Kadoury, "Intervertebral disk modelling from pairs of segmented vertebral models using trainable pre-processing networks," in *Proc. IEEE ISBI*. IEEE, 2018, pp. 1122–1125.

[6] J-Y Zhu, T Park, P Isola, and A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. of the IEEE ICCV*, 2017, pp. 2223–32.

[7] A. Shrivastava *et al.*, "Learning from simulated and unsupervised images through adversarial training," in *Proc. of IEEE Conf. CVPR*, 2017, pp. 2107–2116.

Authorized licensed use limited to: Makerere University Library. Downloaded on July 29,2023 at 06:42:15 UTC from IEEE Xplore. Restrictions apply.