Multi-Planar Convolutional Neural Networks for MRI and CBCT to CT Translation

Gustav Müller-Franzes*[1(✉)][0000−0002−7413−2570], Firas Khader*[1][0000−0001−5089−3589], and Daniel Truhn[1][0000−0002−9605−0728]

University Hospital RWTH Aachen, Aachen, Germany
{gumueller,fkhader,dtruhn}@ukaachen.de
https://www.ukaachen.de/

Abstract. Computed tomography (CT) is the primary imaging modality for dose estimation in radiotherapy. However, their frequent use can result in increased radiation exposure to patients, causing potential radiation-induced complications. In contrast, while magnetic resonance imaging (MRI) and cone-beam computed tomography (CBCT) provide additional information crucial for radiotherapy treatment planning, they fail to provide reliable measures for precise dose calculations. As part of the SynthRAD 2023 challenge, the goal of this study is to develop and evaluate a method for translating MR and CBCT images to conventional CT images, allowing for accurate dose calculations without the need for additional CT acquisitions. Within the context of this study, we propose a multi-planar convolutional neural network, which projects the 3D volumes into 2D slices across the axial, sagittal, and coronal views, and subsequently predicts separate CT images for each view that are fused in a later stage. Preliminary results show that multi-planar fusion enhances CT image quality considerably as compared to single-plane predictions.

Keywords: Image-to-Image Translation · CT · Multi-planar CNN

1 Introduction

Medical imaging has witnessed remarkable growth in recent years, especially in its capacity to diagnose and treat cancer patients. Computed tomography (CT) stands out as a foundational tool in radiotherapy, offering unparalleled insights into patient anatomy and facilitating accurate dose calculations via electron density conversion. Cone Beam CT (CBCT), a more recent variation of traditional CT imaging, has also gained traction in the radiotherapy realm, becoming a go-to choice for image-guided adaptive radiation therapy (IGART). However, CBCT is susceptible to severe artifacts and distortions which can compromise image quality, making it less reliable for precise dose calculations, a crucial aspect of radiotherapy. Moreover, both CT and CBCT imaging expose patients to high-energy radiation that can lead to unintended tissue damage. Especially

* equal contribution

in radiotherapy, where patients may undergo such procedures on multiple occasions, cumulative radiation exposure can become significant.
Magnetic resonance imaging (MRI) holds the promise to provide physicians with more accurate tumor delineation due to its increased soft-tissue contrast, without compromising patients' health due to increased radiation exposure.

Although MRI and CBCT imaging each offer unique insights, essential for effective radiotherapy, their limitations in delivering crucial information on radiation doses constrain their application. This has prompted increased research into deep learning techniques, as potential solutions to bridge this gap. Notably, generative deep learning models are being harnessed to produce synthetic CT images by leveraging existing MRI or CBCT imaging data. These synthetic CT images aim to approximate the radiation dosimetry properties of conventional CT scans, offering the potential to improve radiotherapy treatment planning without the need for additional CT imaging.

To further promote research in this direction, the SynthRAD 2023 challenge [1] was initiated. By providing researchers with datasets containing paired MRI-CT and CBCT-CT images, the objective is to develop machine learning models that allow 1) the conversion of MR to CT images, facilitating MRI-only radiation therapy, and 2) the conversion of CBCT to CT images, enabling CBCT-only IGART. In this study, we present our approach, designed for performing the aforementioned image-to-image conversions, that was developed in the context of the SynthRAD 2023 challenge.

2 Materials and Methods

2.1 Dataset

We trained and evaluated our models using the publicly available data from the SynthRAD 2023 challenge [1]. This data encompasses two distinct tasks: Task 1 provides 180 brain and 180 pelvis MRT-CT paired images, while task 2 provides 180 brain and 180 pelvis CBCT-CT paired images. In addition, each MR-CT and CBCT-CT pair is supplemented with an additional binary mask, roughly outlining the location of the brain and pelvis, respectively.

2.2 Preprocessing

Prior to inputting the images into the neural network, several preprocessing steps were performed. For MR images, we first clipped the signal intensity to the 99th percentile of values within the area delineated by the binary mask. Subsequently, each MR image was individually normalized to a range between -1 and 1. In the case of CBCT images, the signal intensity was adjusted by shifting all values by the minimal value present in the image, clipping its range between 0 and 3000, and then scaling each image from -1 to 1. For CT images, the signal intensity

was clipped between -1024 and 3000. Afterwards, based on the clipped range (-1024 to 3000), the images were rescaled to fall between -1 and 1. Lastly, all 3D images were divided into slices across the axial, coronal, and sagittal planes and stored as 2D images.

2.3 Architecture

The multi-planar approach consisted of three identical fully convolutional neural networks (see Figure 1), employing a symmetric encoder-decoder design. Each network featured an input layer with 64 channels, followed by three down-sampling layers having 128, 256, and 512 channels respectively. The bottleneck layer, i.e., the point of highest compression, comprised 9 convolutional layers, each with 512 channels. Subsequent to the bottleneck layer were three up-sampling layers with transposed convolutions, followed by an output layer.

2.4 Training

The dataset of each task and anatomical region (n=180 images) was split into a training (n=162) and validation set (n=18) with a 9-to-1 ratio. Images were padded to a size divisible by 8 before feeding them into the neural network. The network's predictions were then cropped back to match the original input size. The training process was conducted separately for both tasks 1 and 2, as well as for predicting the brain and pelvis regions.

To supervise the training, the sum over the resulting Mean Absolute Error (MAE), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) between the synthetic CT and ground truth CT images were used as the loss function. This loss function was computed and averaged at both the full and half output resolutions. Importantly, the loss computation was confined to the masked region. To further refine the images according to the challenge metrics (i.e., MAE, SSIM, and PSNR), we decided to conduct additional epochs of fine-tuning on the neural network, omitting the LPIPS loss component.

To update the weights and minimize the loss, the AdamW optimizer with a learning rate of $\alpha = 10^{-4}$ was employed. Moreover, an early stopping criterion was employed, which terminated the training process once the average over the SSIM and MAE losses did not decrease over five consecutive epochs.

2.5 Inference

For inference, the 3D input images were divided into individual slices along the axial, coronal, and sagittal planes. These slices were then sequentially input into the model and later concatenated along their respective planes to form a 3D volume. As outlined in section 3.2, test-time augmentation was applied to each

of the images prior to inputting them into the model, and the resulting predictions were averaged together. In the final step, the 3D volumes corresponding to the axial, sagittal, and coronal predictions were averaged to arrive at the final prediction for the synthetic CT.

3 Results & Discussion

3.1 2D Axial, Coronal, and Sagittal Predictions

On our internal validation set, we observed comparable MAE and SSIM outcomes among the predictions for the brain region in task 1 across the axial, coronal, and sagittal planes (Table 1). This is in contrast to the plane predictions for the pelvis region in task 1, where the axial plane displayed superior MAE and SSIM scores compared to the coronal and sagittal planes. The disparity was likely attributed to the anisotropic voxel spacing ($1 \times 1 \times 2.5$mm) that we encountered when observing the images of the pelvic region. Due to the impending deadline of the SynthRAD 2023 challenge on October 22, 2023, further assessments in the coronal and sagittal planes were not pursued.

Table 1. Results on the validation set. NA = Not available.

	Task 1				Task 2			
	Brain		Pelvis		Brain		Pelvis	
	MAE	SSIM	MAE	SSIM	MAE	SSIM	MAE	SSIM
2D Axial	73.9	0.92	54.1	0.87	55.9	0.94	59.9	0.85
2D Coronal	75.4	0.92	62.3	0.84	58.5	0.93	NA	NA
2D Sagittal	72.1	0.92	62.9	0.84	59.2	0.93	NA	NA

3.2 Test-Time Augmentation

We applied test-time augmentation to our models to increase the predictive performance. This procedure involved horizontal and vertical flipping of the input images before feeding them through the model and subsequently applying the inverse transformation on the output. The resulting images were then aggregated by averaging the voxel values over all images. When test-time augmentation was implemented, we observed an improvement in performance across all examined planes, tasks, and regions, reflected by a reduction in the MAE by approximately -2 (see Table 2).

3.3 Checkpoint Ensemble

During the training process, the model weights corresponding to the best three checkpoints as measured by the lowest validation loss were retained. These checkpoints were leveraged in an ensemble, in which the predictions of the models were averaged. This led to a reduction in the MAE score by about 0.5 (see Table 3).

Table 2. Results on the validation set with test-time augmentation. NA = not available.

	Task 1				Task 2			
	Brain		Pelvis		Brain		Pelvis	
	MAE	SSIM	MAE	SSIM	MAE	SSIM	MAE	SSIM
2D Axial	70.7	0.924	51.9	0.866	53.6	0.940	58.8	0.855
2D Coronal	70.3	0.925	NA	NA	55.7	0.937	NA	NA
2D Sagittal	68.8	0.926	NA	NA	60.2	0.922	NA	NA

Table 3. Results on the validation set after ensembling. NA = not available.

	Task 1				Task 2			
	Brain		Pelvis		Brain		Pelvis	
	MAE	SSIM	MAE	SSIM	MAE	SSIM	MAE	SSIM
2D Axial	69.5	0.926	51.4	0.867	52.6	0.941	57.4	0.860
2D Coronal	68.2	0.928	NA	NA	55.3	0.938	NA	NA
2D Sagittal	67.4	0.929	NA	NA	56.0	0.937	NA	NA

3.4 Multi-Planar Ensemble

To quantify the performance gain when employing the multi-planar approach, the mean signal intensities of the 3D CT predictions from the axial, sagittal, and coronal plane predictions were computed. We observed that the quality of the brain region improved in both MAE and SSIM, while the quality of the pelvis region diminished (see Table 4).

Table 4. Results on the validation set after fusing the axial, sagittal, and coronal plane predictions. NA = not available; (*) Without test-time augmentation.

	Task 1				Task 2			
	Brain		Pelvis		Brain		Pelvis	
	MAE	SSIM	MAE	SSIM	MAE	SSIM	MAE	SSIM
Multi-planar	64.4	0.933	53.6(*)	0.862(*)	51.4	0.944	NA	NA

References

1. Thummerer, A., Van Der Bijl, E., Galapon, A., Verhoeff, J.J.C., Langendijk, J.A., Both, S., Van Den Berg, C.N.A.T., Maspero, M.: SynthRAD2023 Grand Challenge dataset: Generating synthetic CT for radiotherapy. Medical Physics **50**(7), 4664–4674 (Jul 2023). https://doi.org/10.1002/mp.16529

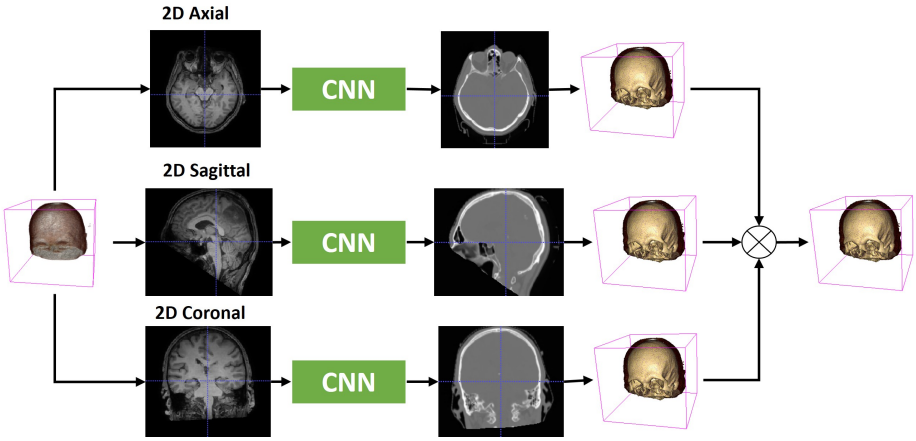

Fig. 1. Network Architecture