



<http://www.diva-portal.org>

This is the published version of a paper published in *Zeitschrift für Medizinische Physik*.

Citation for the original published paper (version of record):

Simkó, A., Bylund, M., Jönsson, G., Löfstedt, T., Garpebring, A. et al. (2023)

Towards MR contrast independent synthetic CT generation

Zeitschrift für Medizinische Physik

<https://doi.org/10.1016/j.zemedi.2023.07.001>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Published under a CC BY-NC-ND 4.0 licence.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-214270>

Towards MR contrast independent synthetic CT generation

Attila Simkó^{a,*}, Mikael Bylund^{a,*}, Gustav Jönsson^a, Tommy Löfstedt^b, Anders Garpebring^a, Tufve Nyholm^a, Joakim Jonsson^a

^a Department of Radiation Sciences, Umeå University, Umeå, Sweden

^b Department of Computing Science, Umeå University, Umeå, Sweden

Received 1 December 2022; accepted 4 July 2023

Abstract

The use of synthetic CT (sCT) in the radiotherapy workflow would reduce costs and scan time while removing the uncertainties around working with both MR and CT modalities. The performance of deep learning (DL) solutions for sCT generation is steadily increasing, however most proposed methods were trained and validated on private datasets of a single contrast from a single scanner. Such solutions might not perform equally well on other datasets, limiting their general usability and therefore value. Additionally, functional evaluations of sCTs such as dosimetric comparisons with CT-based dose calculations better show the impact of the methods, but the evaluations are more labor intensive than pixel-wise metrics.

To improve the generalization of an sCT model, we propose to incorporate a pre-trained DL model to pre-process the input MR images by generating artificial proton density, T1 and T2 maps (i.e. contrast-independent quantitative maps), which are then used for sCT generation. Using a dataset of only T2w MR images, the robustness towards input MR contrasts of this approach is compared to a model that was trained using the MR images directly. We evaluate the generated sCTs using pixel-wise metrics and calculating mean radiological depths, as an approximation of the mean delivered dose. On T2w images acquired with the same settings as the training dataset, there was no significant difference between the performance of the models. However, when evaluated on T1w images, and a wide range of other contrasts and scanners from both public and private datasets, our approach outperforms the baseline model.

Using a dataset of T2w MR images, our proposed model implements synthetic quantitative maps to generate sCT images, improving the generalization towards other contrasts. Our code and trained models are publicly available.

Keywords: Synthetic CT generation; MRI contrast; Robust machine learning

1 Introduction

Radiotherapy treatment planning is commonly based on a combination of MRI and CT data. The high soft tissue contrast of MRI makes it ideal for treatment volume delineation while the electron density information from CT is used for dose calculations for the volume. This pipeline requires

two scans, with the two modalities registered before dose planning to compensate for movement between the scans.

Machine learning has a great potential in medical imaging, especially in radiotherapy [1], where the amount of available data and need for tedious manual work is large. Machine learning solutions for generating synthetic CT (sCT) images from MRI have proven extremely successful,

* Corresponding authors: Attila Simkó, Mikael Bylund, Department of Radiation Sciences, Umeå University, Umeå, Sweden.

E-mail addresses: attila.simko@umu.se (A. Simkó), mikael.bylund@umu.se (M. Bylund).

¹ These authors contributed equally.

with several models being commercially and clinically available [2]. The available machine learning models usually focus on a specific anatomy and MR images with a specific contrast [3]. The contrast settings of an MRI scan defines the significance of the underlying tissues in the acquired signal, therefore the image characteristics can vary widely between settings. The bias of the models towards a specific MR contrast is generally not listed as a limitation, as this implication is considered obvious.

For a spin-echo MRI sequence, a contrast is defined by two settings, the echo (T_E) and repetition times (T_R). The signal of a spin-echo sequence is based on three underlying quantitative maps: proton density (PD), the T1- (T_1) and T2-relaxation times (T_2) according to the signal equation,

$$s = PD \cdot \left(1 - e^{-\frac{T_R}{T_1}}\right) \cdot e^{-\frac{T_E}{T_2}}. \quad (1)$$

Hence the MR signal is defined using the underlying quantitative maps, which are by definition independent of the imaging contrast.

Reported sCT generation models are most often evaluated using pixel-wise metrics [3]. Although differences in dose calculations would provide a more useful, practical performance metric [4], they are performed less often as they require dose calculations, tedious manual work, and dose planning software.

In our presented work we investigated how augmented information such as synthetic PD , T_1 and T_2 maps can help sCT generation, and how it affects generalization towards other contrasts and scanners. We also build on previous work by [5] and use radiological depths to evaluate the dose calculation accuracy, by adding correction factors to the radiological depth which allows for dose estimation. The trained models and source code are made publicly available².

2 Materials and methods

We collected pelvic MRI scans from 375 patients with a 3T Signa PET/MR scanner (GE Healthcare, Chicago, Illinois, United States) at the University Hospital of Umeå, Sweden (ethical approval Dnr: 2019–02666) and corresponding CT scans with a Philips Brilliance Big Bore (Philips Medical Systems, Cleveland, OH, USA). The CT images were registered to the MRs using non-rigid registrations to account for patient movement between the scans. Registrations were performed in Hero Imaging³, which uses the Elastix [6] software package for image registration, based on the Insight Toolkit [7] code. The registrations were performed with the default non-rigid registration

preset. For each scan, 131 axial slices of size 512×512 were used. The T_2w images in the dataset use echo times of approximately 90 ms and repetition times of around 14,000 ms. A mask was created for each patient using the voxels from the CT image above the value of $-200HU$, and the largest connected component was extracted. This ensures that only the patient was included in the mask. This mask was then applied to both the CT and MR images, and saved for later use during the evaluations. The MR images were bias field corrected in Hero Imaging, using N4ITK with a shrink factor of 4, and with a number of control points that produced visually pleasing results. Afterwards, the MR images were Z-normalized and the CT images were trimmed similar to [4], between the values -1000 and $1000 HU$, and then scaled between -1 and 1 . The patients were split randomly between training (80%) and validation (10%), while the remaining 10% were used for testing. For evaluations we also used the publicly available Gold Atlas dataset [8] covering 3 sites (3 different scanners) and two different image contrasts (T_1w and T_2w).

Using the method proposed in [9] we deconstructed the signal from each slice into synthetic PD , T_1 , and T_2 maps. These synthetic quantitative maps ($sQMs$) were used as a contrast-independent representation of the scanned anatomy.

For all experiments, the difference between the evaluated methods was tested for significance using a Friedman test of equivalence followed by a Nemenyi post hoc test.

2.1 Model architectures

Our model architecture was a SRResNet [10], 12 blocks deep with 37.7 million trainable parameters. The architecture included Dropout layers with a rate of 0.2, which can be used to capture the uncertainty in the model, known as Monte-Carlo Dropout [11,12]. The SRResNet was selected due to its extensive use and established good performance in medical imaging applications [13–16]. Two models were trained for sCT generation; the sCT images generated by our baseline model trained on MR images directly are denoted sCT_{MR} while the sCT images generated by the model trained on the decomposed signal—on the $sQMs$ —are denoted sCT_{sQM} .

The sCT_{MR} model has a single channel input, the MR images, seen as “Baseline” in Fig. 1. Whereas for the sCT_{sQM} model, the MR slices were downsampled using Lanczos interpolation to 256×256 and then scaled between 0 and 1 as per the pre-processing requirements of the signal decomposition model. The outputs of this model—the $sQMs$ —were then upsampled using the same interpolation to 512×512 , where the previously saved patient masks were applied on each map to set values outside of

² <https://github.com/attilasimko/contrast-sct>.

³ www.heroimaging.com.

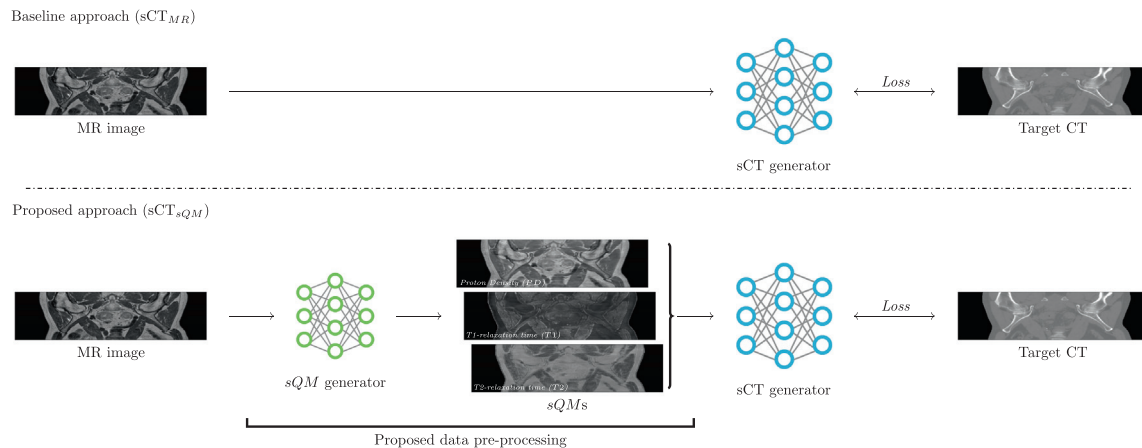


Figure 1. A summary of the two approaches for training an sCT generator model (in blue). The baseline approach (top) uses the input MR to train the sCT model, whereas our proposed approach (bottom) uses a deep learning model proposed in [9] to first decompose the MR image into PD , $T1$ and $T2$ maps, which are then used as input. For both approaches, the CT images are used as target data during training.

the anatomy to zero, and then all three maps were individually Z-normalized. These three maps were then used as a 3 channel image input for the sCT_{sQM} model for sCT generation, seen as our “Proposed approach” in Fig. 1.

The hyperparameters regarding model training were tuned for both models using a range of values—the loss function (including mean squared error, mean absolute error and mean absolute percentage error), optimizer (including Adam, Nadam and SGD) and the learning rate (including 0.05, 0.02, 0.01, 0.005, ..., 0.00001)—to achieve the best validation performance. A grid search approach was used to evaluate all hyperparameter combinations. In both cases, the best results were achieved using mean squared error, and the Adam optimizer with a learning rate of 0.0001.

2.2 Evaluating MAE

The models were evaluated on the testing dataset by comparing the generated sCTs to their corresponding original CTs using MAE. The errors were categorized using the HU values of the original CT into errors coming from reconstructing air (below -100 HU), soft tissue (between -100 HU and 100 HU) and bone (above 100 HU).

Afterwards, the models were evaluated on the publicly available Gold Atlas dataset. As the training dataset is bias field corrected, while the Gold Atlas dataset is not, we perform the evaluations on the original data, and also a bias field corrected version, using N4ITK with 6 control points, and a shrink factor of 2. The sCT generation model in [17] was also evaluated on a subset of this dataset, without using bias field correction, therefore their reported MAE results can be compared to our methods.

We acquired scans using nine other contrasts, to further evaluate the robustness towards the input MR contrast. This in-house pelvic dataset used scans from 6 patients captured

with a 3T Signa PET/MR scanner (GE Healthcare, Chicago, Illinois, United States) at the University Hospital of Umeå, Sweden (ethical approval nr. 2019-02666) and corresponding CT scans with a Philips Brilliance Big Bore (Philips Medical Systems, Cleveland, OH, USA). The dataset contains registered MR images from nine contrasts using a combination of T_E values from [8, 75, 120] ms, and T_R values from [400, 750, 4500] ms and a corresponding CT image. The MAE of the sCTs generated from the different contrasts were plotted against T_E and T_R to visualize the robustness of the models against the settings.

Apart from the nine evaluation points using the acquired dataset, we created synthetic contrasts of further T_E and T_R combinations, by the contrast transfer method proposed in [9]. This yields an artificial, but more varied set of MR images to evaluate sCT generation on. In this case, when evaluating sCT_{sQM} , the contrast transfer model is applied twice, once when transferring a signal of a specific contrast, and then for decomposing the synthetic contrast to use as an input for sCT generation. The evaluation points of the original contrasts are visualized with white squares in Fig. 2, surrounded by the evaluation points of the synthetic contrasts.

2.3 Approximation of dose calculation accuracy

In the pelvic area where there are few large areas with markedly different tissue densities, the deposited dose to a point p within the patient can be reasonably accurately modeled using first order corrections, i.e. correcting for the attenuation of the primary beam. This is usually done by calculating the effective pathlength (radiological depth) to the point p through the patient anatomy and using that depth to calculate the attenuation, used for similar evaluations in [5]. The radiological depth z' is calculated as function of the geometrical depth z as

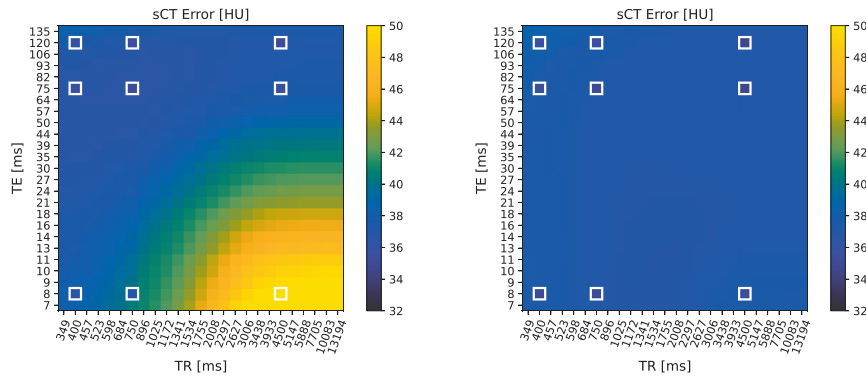


Figure 2. Evaluating the performance of sCT generation for a wide range of MR contrasts using real (in white squares) and synthetic contrasts. The plot shows the MAE results for sCT_{MR} (left) and sCT_{sQM} (right) for the individual contrasts. The values are plotted against T_E and T_R to show how the contrast settings influence the performance of the two models. A logarithmic axes was used for both T_E and T_R following their effect on the MR signal as seen in Eq. 1.

$$z' = \frac{1}{\rho_w} \int_0^z \rho(z'') dz'' \quad (2)$$

where ρ_w is the electron density of water and $\rho(z'')$ is the local electron density estimated from the CT image via a look up table [18].

For small differences in effective pathlength, i.e. below 10 mm, the attenuation scales approximately linearly to the difference in radiological depth. However, the total dose difference at the point p will also depend on the amount of radiation that is delivered in each ray line. To take different beam weights into account, we extracted the cumulative beam weights and control point beam angles from 200 prostate VMAT arcs from the clinical quality archive of our radiotherapy department, and created a standard prostate VMAT arc. Other factors that influence the dose error is the field size and energy. To account for these, we calculated the dose deviations for differences in radiological depths for several field sizes and energies in a water phantom using the dosimetric QA software EqualDose [19], that were used as scaling factors for the dose differences.

The radiological depths were calculated from the original CTs of the testing dataset and the generated sCTs for the center of each slice for every fifth slice for the angles of a 4-field conventional method, and for 36 gantry angles of a full VMAT arc. The mean radiological depths were also calculated for all Gold Atlas patients for the center of mass of each slice for every fifth slice for the VMAT arcs, and translated to an approximate mean difference in dose.

3 Results and discussion

The results for evaluating the MAE on the testing dataset are collected in Table 1, while evaluations on the Gold Atlas dataset are reported in Table 2.

The results for evaluating on a wide range of contrasts are visualized in Fig. 2. Comparison using the $T2w$ images from the testing dataset in Table 1 shows that while the sCT_{MR} achieves a lower MAE for soft tissue, the performance of the two models are not significantly different according to the Nemenyi post hoc test. For an in-depth analysis of the robustness, on all three sites of the Gold Atlas dataset, the difference in performance between using $T2w$ and $T1w$ images is larger for sCT_{MR} than for sCT_{sQM} . Introducing N4ITK bias field correction on the MR images improved their perceptual quality, however this did not translate to significantly better sCT results. As bias field correction introduced no significant improvement to our approach, all later experiments were done on the original Gold Atlas scans. As shown in Table 2, the robust model performs better on $T2w$ images as well on images from all sites, possibly because of the difference in the T_E and T_R values used, compared to the training dataset, despite all three being $T2w$. The uncertainty of sCT_{MR} on the $T1w$ images is not only signaled by the higher MAE but also by the higher standard deviations. Both models significantly outperform the model proposed in [17] as seen in Table 2, which was evaluated on the $T2w$ images from Site 3 of the Gold Atlas dataset, and resulted in a MAE of 40.4 ± 4.7 . Their model was trained on a much smaller dataset, and also resampled the images to 256×256 , which both affect the MAE results, making their model less relevant for further comparison. Examples of generated images for both approaches are visualized in Fig. 4.

The robustness of sCT_{sQM} is further underlined by the evaluations on the wide range of interpolated synthetic contrasts. The training dataset contains images of $T_E = 90$ ms and $T_R = 14,000$ ms and both models perform similarly in that range of echo and repetition times. The visualization of the results in Fig. 2 shows that the performance of the baseline model decreases when either T_E or T_R is decreased

Table 1

Results of MAE and their standard errors for evaluating the sCTs generated by both models on the testing dataset. The results in bold indicate the best performance across rows without significant differences between them following a Nemenyi post hoc test.

	Air MAE [HU]	Soft Tissue MAE [HU]	Bone MAE [HU]
sCT _{MR}	1.357 ± 0.048	20.093 ± 0.182	102.789 ± 1.262
sCT _{sQM}	1.312 ± 0.049	21.342 ± 0.189	103.392 ± 1.082

Table 2

Results of MAE and their standard errors for evaluating the sCTs generated by both models on all the Gold Atlas patients from all three sites and both contrasts. sCTs were generated from both the original (top two rows), and the bias field corrected images (bottom two rows, denoted "N4ITK"). The results in bold indicate the best performance across rows without significant differences between them following a Nemenyi post hoc test.

	Site 1		Site 2		Site 3	
	T2w	T1w	T2w	T1w	T2w	T1w
sCT _{MR}	32.01 ± 0.27	33.81 ± 0.39	39.82 ± 0.24	48.80 ± 0.53	36.75 ± 0.42	38.47 ± 0.42
sCT _{sQM}	31.00 ± 0.23	31.27 ± 0.24	33.89 ± 0.25	39.14 ± 0.22	34.41 ± 0.34	35.27 ± 0.35
	T2w + N4ITK	T1w + N4ITK	T2w + N4ITK	T1w + N4ITK	T2w + N4ITK	T1w + N4ITK
sCT _{MR}	31.87 ± 0.27	33.83 ± 0.38	39.46 ± 0.25	51.55 ± 0.53	36.74 ± 0.42	36.13 ± 0.42
sCT _{sQM}	30.91 ± 0.23	31.04 ± 0.24	33.81 ± 0.24	39.05 ± 0.21	34.40 ± 0.34	34.99 ± 0.34

compared to the *T2w* settings of the training data (for both real and synthetic contrasts), while the robust model shows no change in performance for the explored range of contrast settings.

The calculated radiological depths from the testing dataset are collected in Table 3.

The mean beam weights from the 200 clinical beams and the radiological depth results for the Gold Atlas dataset are visualized in Fig. 3 and their averages are collected in Table 4.

Evaluations on the testing *T2w* dataset collected in Table 3 show that both sCT models are worse for reconstructing the radiological depths from angles 90° and 270° which correspond to the sides of the patients, where the depths are expected to be the largest. However, the other angles and the average results for both the 4-field conventional calculations and VMAT show no statistical differ-

ences between using the original CTs or either generated sCTs.

The radiological depth evaluations on the Gold Atlas from Table 4 show that while sCT_{MR} performs worse for both contrasts than sCT_{sQM}, the average absolute δ values are always below 1 mm, even when evaluating on *T1w* images. The same holds for the weighted δ s as well. Evaluating for the individual angles on Fig. 3 shows that the weighted δ values are always below 1 mm for sCT_{sQM}. For sCT_{MR} the weighted δ values are above 1 mm for several angles between 90° and 270° for *T2w* and around 90° and 270° for *T1w*. Approximating the dose differences from radiological depth values is described in Appendix A. Our results align well with other evaluated models collected in [20] showing that most models trained on prostate datasets achieve a dose difference < 1%. Similarly for MAE, the standard deviation of the results increase when evaluated

Table 3

Radiological depths and their standard errors calculated on the testing dataset from different angles according to 4-field conventional and VMAT methods. The results highlighted in bold were not deemed significantly different from the depths calculated from the original CTs, following a Nemenyi post hoc test.

	4-field conventional				VMAT	
	0°	90° [mm]	180° [mm]	270° [mm]	Mean [mm]	Mean [mm]
CT	96.8 ± 0.67	179.85 ± 0.64	122.15 ± 0.58	182.06 ± 0.64	145.23 ± 1.80	144.03 ± 1.66
sCT _{MR}	96.09 ± 0.65	178.57 ± 0.61	121.80 ± 0.52	180.68 ± 0.64	144.29 ± 1.77	143.04 ± 1.63
sCT _{sQM}	95.86 ± 0.64	178.02 ± 0.61	121.49 ± 0.50	180.10 ± 0.64	143.87 ± 1.77	142.56 ± 1.63

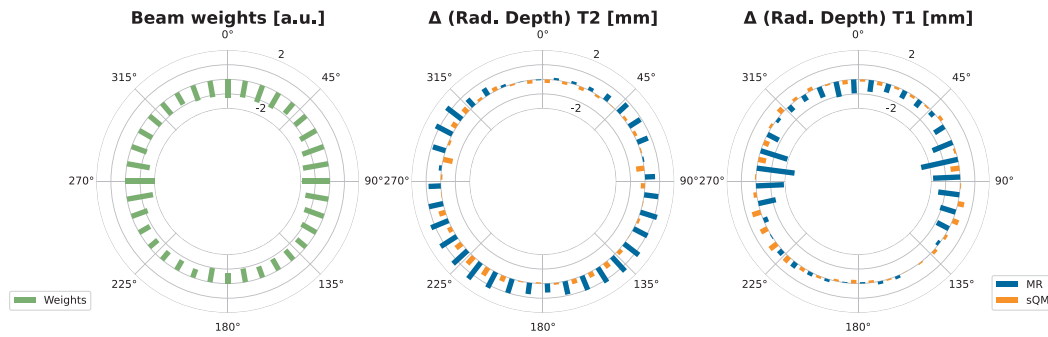


Figure 3. Results of the radiological depth calculations on Gold Atlas dataset for each VMAT arc: On the left the average beam weights are plotted, the plot in the middle shows the difference between the radiological depths for both models generated from $T2w$ images, while on the right, the same results are plotted generated from corresponding $T1w$ MR images.

Table 4

Results of the radiological depth δ calculations and their standard errors, the weighted δ calculations, and the patient-wise mean δ in dose evaluated on the Gold Atlas dataset. The results highlighted in bold were not deemed significantly different from the depths calculated from the original CTs, following a Nemenyi post hoc test.

	Mean δ [mm]		Mean weighted δ [mm]		Mean dose δ [%]	
	T2w	T1w	T2w	T1w	T2w	T1w
sCT_{MR}	0.804 ± 0.078	-0.679 ± 0.127	0.759 ± 0.079	-0.888 ± 0.138	0.234 ± 0.050	-0.567 ± 0.349
sCT_{sQM}	0.019 ± 0.076	-0.059 ± 0.099	0.035 ± 0.075	-0.035 ± 0.103	0.020 ± 0.041	0.015 ± 0.058

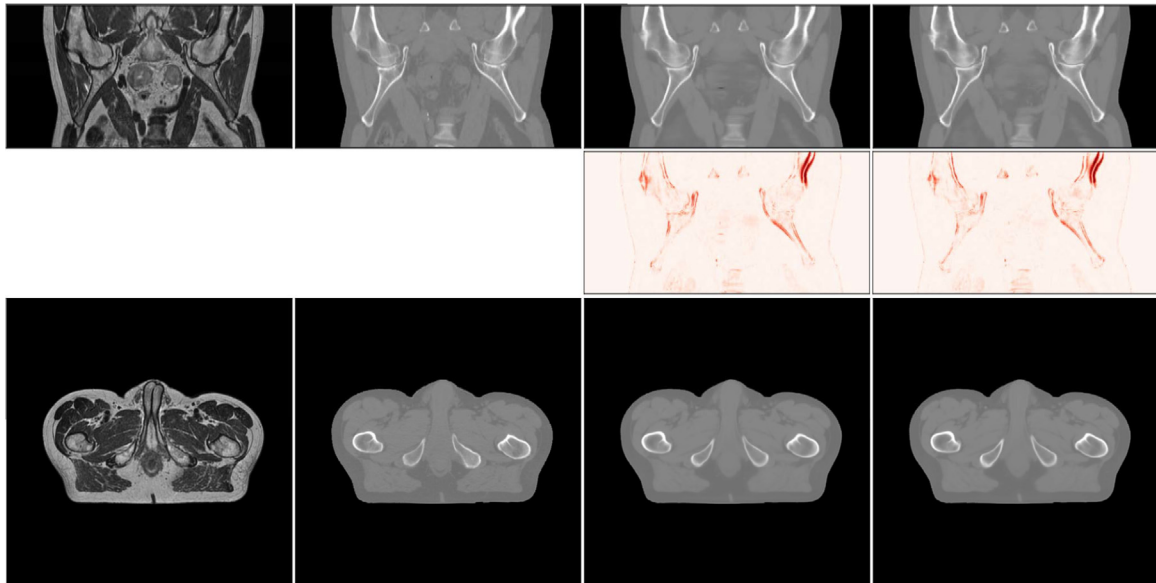


Figure 4. Results for an example patient. From left to right: MR image, the registered CT image, sCT_{MR} , sCT_{sQM} . The top row shows all slices for a patient from the coronal plane, the middle row shows the MAE between the real and synthetic CTs, and the bottom row shows the axial slices that the models operate on.

on $T1w$ for sCT_{MR} , while for sCT_{sQM} the increase is less prominent.

4 Conclusions

We improve the robustness of deep learning-based sCT generation with regards to the input MR image contrast by introducing a pre-processing step of synthetically deconstructing the MR signal into proton density, T1, and T2 maps—contrast-independent quantitative maps. The performance of such a model remains robust towards a wide range of evaluated contrasts, even across different scanners, when compared to a model trained on the same $T2w$ MR images directly. Although all training data was bias field corrected using N4ITK, both approaches also performed well on biased data. Hence, the evaluations show that bias field correction is not a required pre-processing step of the proposed methods. We further show that the differences in radiological depths and approximate dose calculation accuracy are minimal, providing an easy to automate alternative for dosimetric evaluations.

An sCT generation method that works well across different scanners and scanner settings holds great potential for widespread application, removing the need for site-specific model retraining. To ensure that the model works well in new scenarios, the radiological depth calculations provide an easy to automate validation metric, closely connected to practical dosimetric results. Additionally, the available source code and model weights ensure reproducibility and the easy application of our method.

A possible continuation of the project could be to evaluate the method for other anatomies, to investigate if similar robustness could be achieved there. Additionally, an uncertainty metric could be added to provide an estimated reliability of the voxel value assignments. This could help highlight parts of the image that are hard for the sCT generation to handle, such as inputs that for some reason are dissimilar from regularly encountered images. Such a metric could be part of an automated QA process to help identify potential cases where the quality of the sCT generation is uncertain, and additional investigation might be necessary before using the images in the treatment of patients.

Data Availability Statement

The code used to extract the data is distributed by the authors as open-source. The patient data cannot be made available on request due to privacy/ethical restrictions.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [All model evaluations were performed using Hero Imaging, NONPI Medical AB, where Attila Simkó is an employee, Tommy Löfstedt, Anders Garpebring, Tufve Nyholm and Joakim Jonsson are co-founders.]

Acknowledgements

All model evaluations were performed using Hero Imaging, NONPI Medical AB, where Attila Simkó is an employee and Tommy Löfstedt, Anders Garpebring, Tufve Nyholm, and Joakim Jonsson are co-owners.

We are grateful for the financial support obtained from the Cancer Research Foundation in Northern Sweden (LP 18-2182, AMP 18-912, AMP 20-1014, and LP 22-2319), the Västerbotten regional county, and from Karin and Krister Olsson. Computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at the High Performance Computing Center North (HPC2N) in Umeå, Sweden, partially funded by the Swedish Research Council through grant agreement No. 201–05973.

Appendix A Connecting the radiological depths to dose calculations

The calculated scaling factors are shown in Fig. A.5. For the 200 beams that the weights were extracted from, 198 beams were 10 MV. The mean field size was 74×76 mm. By interpolating, the relative dose δ found in Table 4 could be calculated.

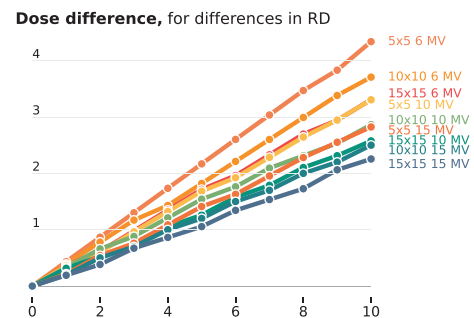


Figure A.5. The calculated scaling factors for the different field sizes and energies to translate the differences in calculated dose.

The scaling factors offer a direct way to connect the differences in radiological depths to dose differences. They can also be used to translate the difference in radiological depths to other field sizes and energies.

References

- [1] Maier A, Syben C, Lasser T, Riess C. A gentle introduction to deep learning in medical image processing. *Z Med Phys* 2019;29(2):86–101.
- [2] Lerner M, Medin J, Gustafsson CJ, Alkner S, Siverson C, Olsson LE. Clinical validation of a commercially available deep learning software for synthetic CT generation for brain. *Radiat Oncol* 2021;16(1):1–11.
- [3] Boulanger M, Nunes JC, Chourak H, Largent A, Tahri S, Acosta O, De Crevoisier R, Lafond C, Barateau A. Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review. *Phys. Medica* 2021;89(July):265–281.
- [4] Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jurgensliemk-Schulz IM, Kerkmeijer LGW, Cornelis AT, Van Den Berg. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol* 2018;63(18):0–11.
- [5] Andreasen D, Van Leemput K, Hansen RH, Andersen JAL, Edmund JM. Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. *Med Phys* 2015;42(4):1596–1605.
- [6] Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: A toolbox for intensity-based medical image registration. *IEEE Trans Medical Imag* 2010;29(1):196–205.
- [7] McCormick M, Liu X, Jomier J, Marion C, Ibanez L. Itk: Enabling reproducible research and open science. *Front Neuroinformat* 2014;8(FEB):1–11.
- [8] Nyholm T, Svensson S, Andersson S, Jonsson J, Sohlén M, Gustafsson C, Kjellén E, Söderström K, Albertsson P, Blomqvist L, Zackrisson B, Olsson LE, Gunnlaugsson A. MR and CT data with multiobserver delineations of organs in the pelvic area-Part of the Gold Atlas project. *Med Phys* 2018;45(3):1295–1300.
- [9] Simko AT, Löfstedt T, Garpebring A, Bylund M, Nyholm T, Jonsson J. Changing the contrast of magnetic resonance imaging signals using deep learning. *Proceedings of the Fourth Conference on Medical Imaging with Deep Learning*, vol. 143. p. 713–727.
- [10] Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, Aitken AP, Tejani A, Totz J, Wang Z, et al. Photo-realistic single image super-resolution using a generative adversarial network. *Cvpr* 2017;2(3):4.
- [11] Gal Y. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. *ICML'16: Proceedings of the 33rd International Conference on International Conference on Machine Learning*. 2016; 48: 1050–1059.
- [12] Melucci M. Relevance feedback algorithms inspired by quantum detection. *IEEE Trans Knowl Data Eng* 2016;28(4):1022–1034.
- [13] Prabhat KC, Zeng R, Mehdi FM, Myers Kyle J. Deep neural networks-based denoising models for CT imaging and their efficacy. *Medical Imaging 2021: Physics of Medical Imaging*. 2021; February.
- [14] Dwarikanath M, Bozorgtabar B, Hewavitharanage S, Garnavi R. Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis. In: *Medical Image Computing and Computer Assisted Intervention, (Proceedings, Part III)*. p. 382–391.
- [15] Jönsson, G. "Impact of MR training data on the quality of synthetic CT generation." 2022.
- [16] Zhang H, Shinomiya Y, Yoshida S. 3D Mri reconstruction based on 2D generative adversarial network super-resolution. *Sensors* 2021;21(9):1–20.
- [17] Vajpayee R, Agrawal V, Krishnamurthi G. Structurally-constrained optical-flow-guided adversarial generation of synthetic CT for MR-only radiotherapy treatment planning. *Sci Rep* 2022;12(1):1–10.
- [18] Ahnesjö A, Mania Aspradakis M. Dose calculations for external photon beams in radiotherapy. *Phys Med Biol* 1999;44(11).
- [19] Nyholm T, Olofsson J, Ahnesjö A, Karlsson M. Photon pencil kernel parameterisation based on beam quality index. *Radiat Oncol* 2006;78(3):347–351.
- [20] Johnstone E, Wyatt JJ, Henry AM, Short SC, Sebag-Montefiore D, Murray L, Kelly CG, McCallum HM, Speight R. 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(1):199–217.

Available online at: www.sciencedirect.com

ScienceDirect