Utrecht University

Feasibility of Cone Beam Computed Tomography with Invertible Recurrent Inference Machines

Minor Research Project
MSc Medical Imaging

Konstantinos Drymas Vrakidis

Examination Committee:

Dr. Matteo Maspero
Assistant Professor, UMC Utrecht

Dr. Alessandro Sbrizzi
Associate Professor, UMC Utrecht

Utrecht, 21st October 2022

Feasibility of cone beam computed tomography with invertible recurrent inference machines

Konstantinos Drymas Vrakidis

MSc Student, Medical Imaging, Utrecht University
E-mail: k.d.vrakidis@students.uu.nl, Student no.: 3187909

Supervisor: Matteo Maspero, Assistant Professor, UMC Utrecht

Abstract

Introduction: Online adaptive radiation therapy (ART) relies on high-quality in-room volumetric imaging. Conventional linear accelerators (linac) are generally equipped with onboard orthovoltage cone beam computed tomography (CBCT) systems. Due to their suboptimal image quality, however, their use is limited to patient positioning tasks. Deep learning-based methods can potentially generate synthetic-CTs (sCT) directly from CBCT projection data, and enable CBCT-based online ART to be brought to conventional treatment delivery systems. Designed for solving inverse problems, invertible recurrent inference machines (iRIM) can be considered a suitable candidate for this task. To this end, the feasibility of using iRIM frameworks for CBCT image reconstruction is examined in this work.

Materials & Methods: 2D and 3D iRIM models were implemented and trained on datasets composed of CBCTs from sixty-two head and neck patients who underwent head and neck image-guided radiation therapy (IGRT). In the 2D models, the CBCT inverse problem was approximated as a parallel-beam CT, while in their 3D counterparts, a CBCT geometry was emulated. In parallel, trainings using sparse CBCT data were performed on both 2D and 3D models by applying two-fold and four-fold reductions in acquisition angles. Utilizing the structural similarity index measure (SSIM) as a metric, the performances of the trained models were evaluated and subsequently compared using the paired Mann-Whitney U test.

Results: The 2D and 3D iRIMs respectively achieved SSIMs (mean ± std) of (0.96 ± 0.02) and (0.94 ± 0.04) in the case of the complete CBCT acquisitions. The two-fold and four-fold reduction in acquisition angles yielded (0.94 ± 0.04) and (0.93 ± 0.05) for the 2D iRIM, and (0.93 ± 0.05) and (0.92 ± 0.05) for the 3D iRIM, respectively. In all cases, the performances of the 2D iRIMs were superior to their corresponding 3D ones, with the differences not being found significant.

Conclusion: iRIMs can be used for CBCT reconstruction in 2D and 3D cases, even for undersampled acquisitions. This makes iRIM an excellent candidate to obtain high-quality CT-grade reconstructions from CBCT data, and potentially bring online ART to conventional clinical linacs.

Keywords: adaptive radiotherapy, cone beam computed tomography, image reconstruction, deep learning

Introduction

Since the inception of the therapeutic potential of ionizing radiation in 1896 [1], [2], radiation therapy (RT) has progressively become one of the most important treatment strategies for oncological disorders. Globally, 50% of cancer patients receive RT as part of their treatment [3]. The most common RT delivery system is the electron linear accelerator (linac), with over 10,000 currently in use worldwide [4].

Adaptive RT (ART) [5] is a process in which irradiation treatment plans are adapted to anatomical or functional variations based on regular imaging [6]. Patient-specific variations can be of physiologic origins, such as organ deformation, filling change, respiratory and peristaltic motion [7], or they can be radiation-induced, such as tumor shrinkage, weight loss, and morphological changes in organs [8]. Adapting the original treatment plan to address these changes maximizes the dose delivered to the target and minimizes the treatment-related toxicity while sparing the organs at risk [7].These characteristics have been linked to a lower risk of late complications and the potential to improve disease-free survival [8].

Treatment replanning, dose calculations, and patient-specific quality assurance performed in ART require high-quality volumetric imaging, which can take place at multiple time points throughout a treatment that might take weeks. In offline ART, treatment replanning takes place between treatment fractions using computed tomography (CT) or magnetic resonance imaging (MRI) scans. This method mainly addresses systematic and progressive variations due to the elapsed time between imaging and treatment delivery [7]. Adaptations in online ART take place immediately before the treatment fraction using in-room imaging, allowing the incorporation of temporal and stochastic anatomy variations [7]. It requires the use of in-room imaging which most commonly involves either a dedicated on-rails CT system or makes use of the imaging modality integrated into specialized high-end linacs [9].

Simultaneously, the majority of the conventional linacs have integrated orthovoltage (kV) cone beam CT (CBCT) systems. kV-CBCT is a volumetric imaging technique that relies on a pair of X-ray tubes, producing a two-dimensional collimated cone-shaped beam and a flat panel detector. Mounted on robotic arms perpendicularly to the radiation beam, the transaxial rotation of the pair acquires projections from multiple directions. The reconstructed images provide adequate spatial resolution and allow for precise target localization [10] and are therefore used for patient position verification and minimization of setup errors [10].

Despite CBCT providing volumetric imaging, it is primarily used in the context of image-guided RT (IGRT). CBCT is generally considered not suitable for online ART, due to the inferior image quality it provides [7], [9]. Comparisons with CT found CBCT to be less capable of producing noise-free and anatomically correct images and of providing inferior soft tissue contrast [11]. Multiple image artifacts also plague CBCT, with the most notable ones being caused by scatter, beam hardening, and metallic implants [12].

Different possible directions can be followed to bring CBCT to a CT-grade quantitative accuracy required for online ART. Improving the image quality of CBCT is one of them. Several methods have been proposed, among which notable are the statistical iterative reconstruction methods [13], the use of total variation penalty terms for edge preservation and noise suppression [14], and Monte Carlo -based scatter correction methods [15]. Post-reconstruction mapping to CT using deformable image registrations or Hounsfield Unit (HU) quantification using look-up tables can be used [16].

Deep learning-based CT synthesis is considered an alternative direction, in which synthetic-CTs (sCT) are generated based on CBCTs. Multiple neural network (NN) architectures have been applied for this task. Cycle generative adversarial networks (CyleGAN) [17] have been used for unpaired unsupervised learning in multiple studies [16], [18], from which the dose accuracies based on the sCTs were found comparable to the ones based on CT ones. Two studies [19], [20] that compared U-Net [21], CycleGAN and pix2pix [22] models for CT-synthesis for breast cancer RT and nasopharyngeal carcinoma RT, found the U-Net and the CycleGAN as the best-performing models, respectively.

Novel data-driven methods have recently achieved CT synthesis directly based on CBCT projection data. A significant such method is the iterative CBCT (iCBCT) model [23]. iCBCT combines a statistical iterative CBCT algorithm with a trainable total variation penalty for noise suppression. Based on a variation of iCBCT, a commercial CBCT-linac system was recently introduced capable of CBCT-based online ART [24].

Invertible recurrent inference machines (iRIM) [25], [26] were recently introduced as frameworks for solving inverse problems. iRIMs combine characteristics of recurrent neural networks (RNNs), such as multiple steps and hidden memory states, with physics-informed forward-model operators to iteratively learn to reconstruct from the training data. Notable performances have been demonstrated with iRIMs in accelerated MRI reconstructions [26]–[29], in reconstructions of astronomical observations [30], and in parallel-beam CT reconstructions using simulated low-dose data [31].

Inspired by the recent clinical adoption of iCBCT, and the potential of iRIM frameworks, we wish to investigate the use of iRIM-based frameworks for an end-to-end CT synthesis approach using CBCT projection data. Provided that sufficient HU accuracy can be achieved, this approach could enable the use of onboard kV CBCT integrated, rendering the dependencies to CT replanning.

As an intermediate step towards that objective, the feasibility of using iRIM frameworks for CBCT reconstructions is investigated in this work. iRIM models with 2D and 3D architectures were trained to generate CBCT reconstructions using synthetic CBCT projection data. Additional investigations by training with sparse projection data were used to assess the learning capacity of the iRIM models.

Background

The CBCT inverse problem

Image reconstruction belongs to the class of inverse problems, in which the acquired information is used to recover the spatial distribution of the imaged object. In essence, the CBCT inverse problem can be formulated using a linear system of equations in which projections, y, are generated from X-rays passing through an object, x. The forward model is then defined as:

$$\mathrm{y} = \mathcal{F}\mathrm{x} + \mathrm{n}$$

Where n is the additive noise inserted into the acquisitions, and $\mathcal{F}$ the forward projection operator that models the acquisition. Image reconstruction aims to find a back projection operator $\mathcal{B}$ that by acting on the projections yields the spatial distribution of the object as:

$$\mathrm{x} = \mathcal{B}\mathrm{y}$$

Tasks as such are not trivial. Accurate modeling of the stochastic nature of the acquisition and finding suitable forward and backward operators is the subject of a broad field in mathematics and physics. The computational sequences that approximate the backward operator $\mathcal{B}$ are called image reconstruction algorithms, for which more details for the case of CBCT can be found in [32], [33].

Conventional CBCT reconstruction

In CT, analytical reconstruction algorithms use the Radon transform $\mathcal{R}$ as a forward operator of the acquisition, which estimates the line integrals through the imaged object. The filtered backprojection algorithm (FBP) is the simplest form of reconstruction and backward operator, which involves the adjoint Radon transform $\mathcal{R}^{-1}$ (backprojection) along with a filtering process. FBP is regarded as a highly efficient and reliable algorithm suitable for many imaging applications [12]. However, it is also prone to generate noisy reconstructions with image artifacts. The Feldkamp-Davis-Kress (FDK) [34] algorithm, which is an extension of FBP for circular CBCT geometry, is considered the most important algorithm for CBCT image reconstruction [12]. Despite the inheritance of many FBP issues, FDK has a highly parallelized structure and exhibits enough robustness to render it the most used algorithm in clinical applications [12].

iRIM

Invertible recurrent inference machines (iRIM) [25], [26] were introduced for the fastMRI challenge [35], and are RNN-based frameworks for solving inverse problems. Imitating iterative reconstruction methods, they unroll the inference procedure into multiple RNN steps. Each step consists of a group of convolutional layers that process an input to produce an output and an update of a hidden memory state [30]. The mismatch between the output of each RRN step and the ground truth contributes to the final loss function. iRIMs are designed to optimize towards a maximum a posteriori (MAP) solution without the explicit use of a likelihood function or a prior within each time step. Instead, these are considered implicit learnable parameters of the model. Their gradients over time steps, however, are stored and used during model training in backpropagation. The estimation of the likelihood gradients is a key component of iRIMs and requires the direct comparison of acquired data and image estimates of each RRN step. This is enabled by the integration of a pair of forward and backward operators into the model architecture. The additional feature of invertibility allows for memory saving and it removes training instabilities when using large training datasets [26]. For a detailed description of the technical aspects of iRIM frameworks, the reader is referred to [25], [26], [31].

Materials & Methods

Data

In this study, we collected kV-CBCT from 62 patients who underwent head and neck IGRT, and divided them into groups of 44 training, 5 validation, and 13 test patients. CBCTs were acquired using the Elekta XVI system v5.0.2.b72 (Elekta, AB, Sweden) equipped with a Si-flat panel detector with an active imaging area of 409.6 x 409.6 mm positioned at the distances of 536 mm from the axial isocenter, and 1536 mm from the kV X-Ray source [36]. CBCTs were acquired using the S20 collimator, a field-of-view (FOV) of 276.7 x 276.7 x 276.7 mm at the isocenter, and the center of the kV detector aligned to the center axis of the kV source.

The raw data from each scan consisted of a set of 2D projections acquired in gantry angles ranging from 190 to 210 projections, with the size of each projection being 504 x 504 pixels. Sinograms are constructed by concatenating the projection slices for all angles along a third dimension and re-slicing the resulting 3D array along a different direction. As a result, each scan obtained a set of 504 sinograms of sizes 504 x (190-210) pixels. Apart from the inconsistency in angular dimensions of the sinograms, additional inconsistencies were present, which are showed in Appendix I. Different acquisition arcs with total lengths in the range of 190° to 220°, clockwise and counterclockwise rotation directions and inhomogeneous angular sampling were observed across different scans.

Data preprocessing

Two variations of datasets were created based on the available data: the 2D and the 3D ones. With the exemption of an additional interpolation step at the end of the preprocessing of the 2D dataset, the same preprocessing steps were followed. The inconsistencies in size and contextual information contained in raw sinograms generated highly heterogeneous training data. To simplify iRIM trainings, we standardized the input by simulating acquisitions with a fixed number of projections and acquisition arcs across different scans according to the pre-processing steps found in Figure 1.

Starting from the raw sinogram data and their corresponding angular sampling information, we generated CBCTs by employing the FDK analytical reconstruction algorithm. The algorithm implementation was based on the Reconstruction Toolkit (RTK) v2.3 Python package [37]. Parker-weighting [38] was applied to compensate for the partial arc of acquisition and truncation correction to suppress artifacts outside the FoV of interest. To be in line with clinical reconstructions, the chosen size of final image volumes was 270x264x270 of isotropic voxels of sides 1 mm, and the smaller dimension corresponds to the axis of the XVI gantry rotation.

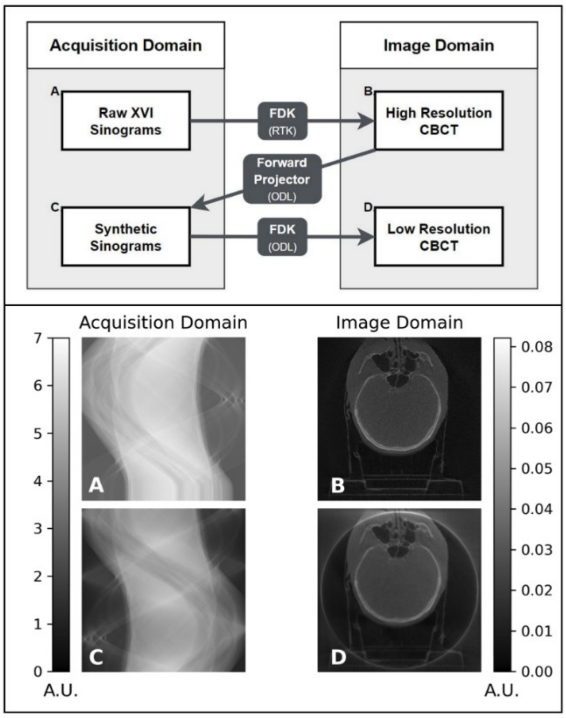


Figure 1. (top) Schematic flow diagram of the preprocessing pipeline. (bottom) An example of a visualization of the results of each of the preprocessing steps. Sinograms (A), and (C) respectively correspond to the raw ones and the synthetic that were used as inputs in iRIM trainings. CBCTs (B) and (D) correspond to the high- and low-quality images used for ground truth and priors in trainings, respectively.

Synthetic sinograms are generated by imitating a simplified acquisition process using a Forward Projection operator on the reconstructed image volume. For this purpose, the Radon transformation as implemented in the Operator Discretization Library (ODL) v1.0 [39] was used, which also required the construction of a geometric model of the XVI system. For the acquisition process, a fixed set of 200 projection angles equally spaced at an angle interval of 0° to 234° was selected across all scans. Ultimately, projection data consisted of a set of 504 sinograms with sizes of 200x504 were obtained.

Since iRIMs optimize towards a MAP solution, the use of an initial estimate of the image to be used as a prior is recommended, to reduce the solution space [25]. As such, low-quality CBCTs were provided as additional inputs for the training and the inferences of the iRIM models. Henceforth, these CBCT images are referred to as priors. These were reconstructed using an ODL implementation of the Parker-weighted FDK algorithm on the synthetic sinograms with a Hann filter with a relative cut-off frequency of 0.6. Their grid and voxel sizes were identical to that of the target images. As can be seen in Figure 1.D, image artifacts [12] were present due to the lack of truncation correction.

For the training of the 3D iRIM, the complete 3D sinogram of size 200x504x504 was used along with an image estimate and ground truth of size 270x264x270 each, to be obtained from each original scan and, subsequently, to be used for the 3D training. The 2D case requires further preprocessing of the sinogram and image data. From both types of data, 2D slices along the axis of the gantry rotation were generated. Since images have lower resolution than their corresponding sinograms along this dimension, nearest-neighbor interpolation was used to match their dimensions. As a result, 2D data comprised of 504 sinograms of sizes 200x504 were obtained for each original scan, along with 504 initial image estimates and ground truths of sizes 270x270.

For the 2D datasets, the split of scans into training, validation, and test datasets led to iRIM trainings over 22176 sets of sinogram, prior and ground truth image slices, and parameter tuning on 2520 slices. For the 3D dataset, the number of patients in each dataset also represented the number of data used.

Network Architecture

The core architecture for the 2D and 3D networks used was based on the iRIM model implemented for a submission [40] to the fastMRI challenge [35], which dealt with undersampled MRI data for accelerated MRI reconstruction. Further modifications to the original model were made separately for the 2D and 3D iRIMs. Developed with different architectural choices and operators, these networks approach the CBCT inverse problem with distinct strategies and thus were used independently.

The 2D iRIM model approximated the task as a simplified 2D parallel-beam CT problem. This was inspired by previous work on modifying the original iRIM model for image reconstruction of low-dose 2D CT data [31]. That change in imaging modality, and the dependence of the likelihood gradient estimations on modality-specific forward and backward projection operators, required the replacement of the Fourier-based operators necessary for the MRI inverse problem with Radon-based ones that are utilized in transmission tomography. For the implementation of the 2D Radon transformations and model of the parallel-beam CT geometry, the ODL library was chosen. The latter is due to ODL's gradient-preserving capabilities, which render it suitable for back-propagation operations during neural network training.

The 2D iRIM model was constructed with 4 RNN steps, with each step composed of 12 convolutional layers. The number of channels in these layers was kept constant at 64, while the number of channels of the hidden states varied, as noted in Table 1. Multiplicity, a factor determining the number of iterations over the likelihood gradient at every time step, was set to 4. Weight sharing across different iRIM steps that would limit the depth of the network and, by extension, the number of trainable parameters was not used. The resulting 2D iRIM consisted of 138 million parameters in total. With float32 precision, the VRAM requirements were 8.1 GiB.

For the 3D iRIM, complete 3D CBCT acquisitions were regarded for the inverse problem by implementing the 3D CBCT operators for the likelihood gradient in ODL. To maintain training stability, the outputs of the forward and backward operators were always divided by 100. For practical limitations, the extension of the model to the third dimension requires different architectural hyperparameters. While the number of RNN steps was kept the same as the 2D iRIM, the channels of both the convolutional layers and the hidden states were halved, and multiplicity was set to 1. Despite these changes, the resulting 3D iRIM was larger than the 2D version, with 167 million parameters in total. The 3D model occupied 31.5 GiB of VRAM with float32 precision.

Training

The models were developed and trained using PyTorch v1.8.1, enabled with CUDA v11.1, on an NVIDIA Tesla V100 Tensor Core GPU equipped with 32 GB of VRAM. The training process was performed on the training datasets for 80 epochs using the Adam optimizer [41]. The mean squared error (MSE) was chosen as the objective function, which for images with n number of voxels it is given by:

$$MSE_{(x,\hat{x})} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$$

Where x_i and $\hat{x}_i$ the ground truth and the estimated images, respectively. The learning rate was initially set at 10^{-4} and was scheduled to decay by a factor of 10 every 30 epochs. Due to memory limitations, the batch size was set to 8 for the 2D iRIMs and 1 for the 3D iRIMs. In each epoch during training, the MSE loss was additionally estimated on the validation dataset. The best-performing model was selected based on the epoch with the lowest validation loss to prevent overfitting. Early stopping was used in cases of slow convergence with a trigger of a smaller than 1% change in the validation loss for three consecutive epochs.

	Architectural Hyperparameters							Training Hyperparameters			
Architecture	Conv. Layer Channels	Hidden state channels	RNN steps	Shared weights	Multiplicity	Number of Parameters	VRAM (float32)	Batch size	Loss Function	Optimizer	Initial Learning Rate
2D	[64, 64, 64, 64, 64, 64, 64, 64, 64, 64, 64, 64]	[64, 64, 128, 128, 256, 1024, 1024, 256, 128, 128, 64, 64]	4	False	4	138 mil.	8.1 GB	8	MSE	Adam	10^{-4}
3D	[32, 32, 32, 32, 32, 32, 32, 32, 32, 32, 32, 32]	[32, 32, 64, 64, 128, 512, 512, 128, 64, 64, 32, 32]			1	167 mil.	31.5 GB	1			

Table 1 The chosen options of hyperparameters of the 2D and 3D iRIM models used.

Evaluation

The performance of the networks was assessed based on the 13 scans of the reserved test dataset. For the case of the 2D iRIM model, the generated output and ground truth images were additionally concatenated into image volumes of size 270x504x270 voxels, which were subsequently resized to 270x264x270 using bilinear interpolation. The generated image volumes were quantitatively compared with their corresponding ground truth using three metrics.

Mean absolute error (MAE) is a metric used for regression tasks and quantifies the dissimilarity of the estimated image $\hat{x}$ to the ground truth x based on the mean absolute difference between their corresponding voxels. For images of n voxels, it is defined as:

$$MAE_{(x,\hat{x})} = \frac{1}{n}\sum_{i}^{n}|x_i - \hat{x_i}|$$

Peak signal-to-noise ratio (PSNR) is a computer vision metric used to measure reconstruction loss. It is defined as the ratio between the maximum possible value of a signal, and the amount of noise present in an image, as described by the MSE. PSNR is expressed in terms of a logarithmic decibel scale, with higher values representing a better match between images, and it is estimated by:

$$PSNR_{(x,\hat{x})} = 10 \cdot \log_{10}\left(\frac{\max(x)^2}{MSE_{(x,\hat{x})}}\right)$$

The structural similarity index measure (SSIM) [42] is a metric based on the similarity of two images as perceived by humans. The perceptual quality is modeled using the means $\mu, \hat{\mu}$, the variances $\sigma, \hat{\sigma}$, and the covariance $\sigma_{x\hat{x}}$ of the two images $x, \hat{x}$ as follows:

$$SSIM_{(x,\hat{x})} = \frac{(2\mu\hat{\mu} + c_1)(2\sigma_{\mu\hat{\mu}} + c_2)}{(\mu^2 + \hat{\mu}^2 + c_1)(\sigma^2 + \hat{\sigma}^2 + c_2)}$$

The two variables are given as $c_1 = (0.01 \cdot L)^2$ and $c_2 = (0.03 \cdot L)^2$, with L the dynamic ranges of the voxel values. The single-scale SSIM was used, as implemented in the Scikit-Image v0.19.2 [43] Python package. The estimation was applied regionally within a window of 7x7x7 voxels moving pixel-by-pixel over the entire image volumes, with the mean SSIM value being reported.

All metrics used for the quantitative comparison were estimated within a cylindrical volume with a height of 230 mm and diameter of 256 mm, aligned with the rotation axis. This mask is in line with the clinical reconstructions, limiting the FoV and eliminating reconstruction artifacts [44] from the evaluation. In addition, the metrics were also used to compare the low-resolution CBCT before the ground truth, which will act as the baseline comparison. Whisker plots based on these metrics were created, with the median, the 1st (Q1), and 3rd (Q3) quartile ranges of the distributions being annotated.

From the resulting mean and standard deviations of all metrics, SSIM was chosen as the representative metric to be reported. Any discrepancies with the rest of the metrics will be explicitly noted. The performances of the trained models on the test data were compared by using the paired Mann-Whitney U test on their calculated metrics, for which the SciPy v1.7.3 library [45] in Python was used. A difference in performance is characterized as significant for p-values smaller than 0.05.

Experiments

In parallel to the training and the assessment of the previously described 2D and 3D iRIMs, several experiments were performed to find the optimal training parameters and to assess the learning capabilities of models with iRIM architectures. Several additional iRIM models were trained and evaluated using the 2D and 3D datasets previously described. The objectives and the evaluation procedures followed in these investigations are described in the following sections.

Loss Function

A separate 2D iRIM training took place using a SSIM-based loss function instead of the MSE loss. Contrary to the SSIM metric used for evaluations, a sliding window was not used for the estimation of the SSIM loss. The performance of the SSIM-trained 2D iRIM was compared against the original MSE-trained model.

iRIM depth

To assess the impact of the relatively increased depth of the 2D architecture when compared to the 3D one, a 2D iRIM model was trained using the architectural hyperparameters used for the 3D iRIM. The resulting 2D network had 34.4 million parameters and was trained with a batch size of 8. It required 4.3 GB of VRAM with float32 precision. The significance of the decreased network depth was evaluated based on the performances of the trained model against the original 2D iRIM.

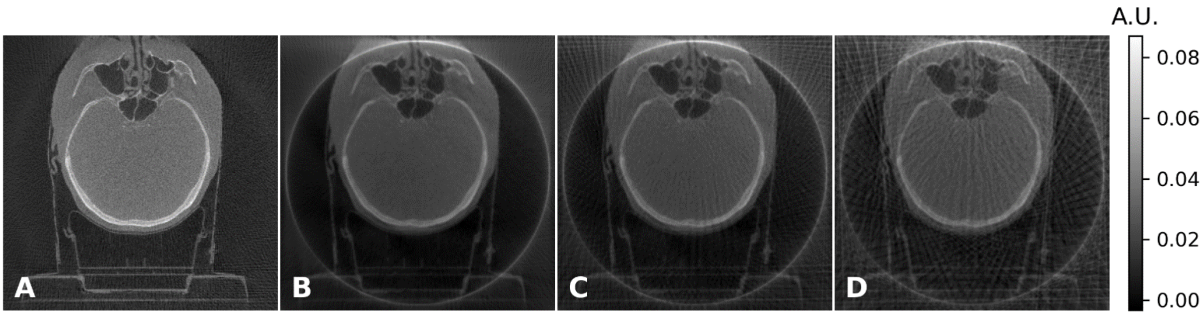


Figure 2. A showcase example of the priors of the undersampled datasets. In (A), the ground truth of an axial slice from the training dataset is shown, while in (B), (C), and (D), the training priors are shown as generated with 200, 100, and 50 projection angles, respectively.

Sparse CBCT acquisitions

Investigations on the impact of the number of projection angles chosen for the generation of synthetic sinograms were performed. While 200 projections are typical for clinical CBCT acquisition, the learning capability of the iRIM architecture would be further tested in cases of heavily undersampled acquisitions. The preprocessing steps described in the data pre-processing section were repeated for 50 and 100 projection angles, creating four additional datasets of both 2D and 3D types. After adapting their operators to the new projection angles, a corresponding number of IRIM models were trained with these datasets. The outputs of the trained networks and the priors used for each training were compared against the related ground truth, and their differences were statistically analyzed. A plethora of combinations of inputs and outputs were statistically assessed to evaluate the effects of undersampling in image reconstruction using iRIMs.

iRIM: 2D vs 3D

Since different approaches were followed for the 2D and 3D iRIMs using distinct forward and backward operators, finding the best-performing approach was of special interest. Therefore, the similarities of the 2D and 3D outputs to the ground truth were statistically evaluated. To gather further evidence of any performance difference, this assessment was also extended to the undersampled cases of 100 and 50 projection angles.

Feasibility

In this work, the 2D and 3D iRIM models were implemented to investigate the feasibility of using iRIM frameworks for CBCT reconstruction. This feasibility was determined by gathering evidence of significant improvements in the iRIM-generated reconstructions, over the low-quality CBCTs that were used as priors in the training process. To that end, parallel comparisons of priors and iRIM outputs against the ground truth were performed using the evaluation metrics previously described. Subsequently, their differences were statistically compared to measure their significance. This process was performed on all cases of 2D and 3D iRIMs trained with the datasets of 200, 100 and 50 projection angles.

The time required for the generation of a full 3D reconstruction was also assessed. The average inference speed was measured over inferences of 1000 slices for the 2D iRIMs and of 13 full scans for the 3D iRIM. Lastly, a qualitative visual assessment of the full inferred iRIM outputs was performed on the second-best, median, and the second worst performing patient scans, as measured by SSIM. From these, random axial slices and mid-sagittal slices were chosen to showcase the characteristics of the generated reconstructions.

Results

Due to the early stopping, the number of iterations over the training data varied within models. Similarly, the mean inference speed also depended on the architecture of the trained models. These numbers are reported in Table 2. MAE, PSNR, and SSIM exhibited very similar behavior over all the experiments, which validates the choice of reporting mainly on SSIM. The complete set of results in tabular forms can be found in Appendices II and III.

iRIM Model	Trained Epochs	Mean Inference Times	
		Per slice [ms]	Per volume (264 slices) [s]
2D SSIM	11	152.5	40.3
2D Reduced	7	54.5	14.4
2D 200	16	154.2	40.7
2D 100	16	150.7	39.8
2D 50	18	150.9	39.8
3D 200	34	-	12.1
3D 100	33	-	10.2
3D 50	30	-	9.5

Note: The accuracy of inference speed measurements between the 2D and the 3D models might be influenced by external factors, such as CPU and GPU processes running in the background.

Table 2. Trained epochs and mean inference speeds of the trained iRIM models.

2D iRIM loss function and depth

The use of the SSIM loss function for the training of the 2D iRIM accelerated the convergence of the learning when compared to the MSE-trained one, with early stopping being triggered at the 10th epoch down from the 15th one. As can be seen in Figure 3, with the SSIM-trained model we obtained a mean (±std) SSIM of 0.96 (±0.22). No significance (p-value = 0.88) was found between them and the original MSE-trained model, which also obtained 0.96 (±0.22).

As seen in Figure 3, inferences with the reduced iRIM achieved the mean SSIM value of 0.96 (±0.23). Comparison with the corresponding value of the 2D iRIM was found to be statistically insignificant (p-value = 0.76). A notable change in the speed of inferences was observed between the reduced and the original models, with 14.4s and 40.7s mean time for inference of 264 slices, respectively.

Undersampled CBCT

The complete set of the results of the undersampled experiments can be found in Appendices II-IV. A statistical comparison over the priors used as inputs showed that only the 50 projections priors had a significant (p-value = 0.036) degradation over the 200 projections prior, with mean SSIMs of 0.8643 (±0.0583) and (0.9041+/-0.0406), respectively. The effect of reducing the 200 projection angles by half an insignificant (p-value=0.38) effect, which yielded the SSIM of 0.8943 (±0.0460). MAE was in line with these findings, while PSNR showed no significant impact on any undersampled prior.

Significance tests to assess the 2D iRIM outputs of the undersampled training against the complete acquisition prior measured p-values of 0.006 and 0.003 for the SSIMs, respectively for the 100 and 50 projection angles datasets. Therefore, the outputs of the 2D iRIM trained with 100 and 50 projection angles were significantly superior to the prior generated with the full 200 projection angles. Similar comparisons for the 3D iRIM measured p-values of 0.028 and 0.24 for the 100 and 50 projection angles datasets, respectively. In the 3D case, only the iRIM trained with 100 projection angles was found significantly superior to the full acquisition prior.

Lastly, a cross-comparison was performed between the outputs of the 200, 100, and 50 projections cases. For the 2D iRIM model, only the 50 projection outputs were found significantly (p-value) inferior to their 200 projection

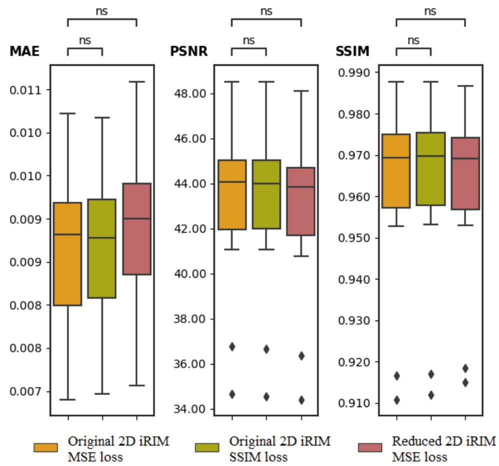


Figure 3. Whisker plots of the results of the 2D iRIM Experiments. Statistical comparisons using the paired Mann-Whitney U test are annotated with brackets on the top of the plots. With ns, p-values greater than 0.05 are signified.

counterparts using SSIM. With MAE, the 100 projections case was also found significantly (p-value) inferior to the complete acquisition, and with PSNR, no significance was found in any of the patients. In the 3D iRIM model outputs, all differences were insignificant across all metrics.

iRIM: 2D vs 3D

For the full acquisition, the 2D iRIM achieved an average SSIM of 0.96 (±0.02), while the 3D iRIM a 0.94 (±0.04). In the 100 projections case, the 2D network achieved the mean values of SSIM 0.94 (±0.04). Similarly, for the heaviest undersampling case of 50 projections, a mean SSIM of 0.93 (±0.05) was obtained with the 2D iRIM, and 0.92 (±0.05) with the 3D iRIM. None of these 2D-3D differences per sample case were found significant using SSIM, PSNR and, with one exemption, MAE. In the case of the 200 projection angles, using MAE the 2D iRIM was found significantly (p-value=0.005) superior to the 3D iRIM, with 0.009 (±0.001) and 0.0102 (±0.012), respectively.

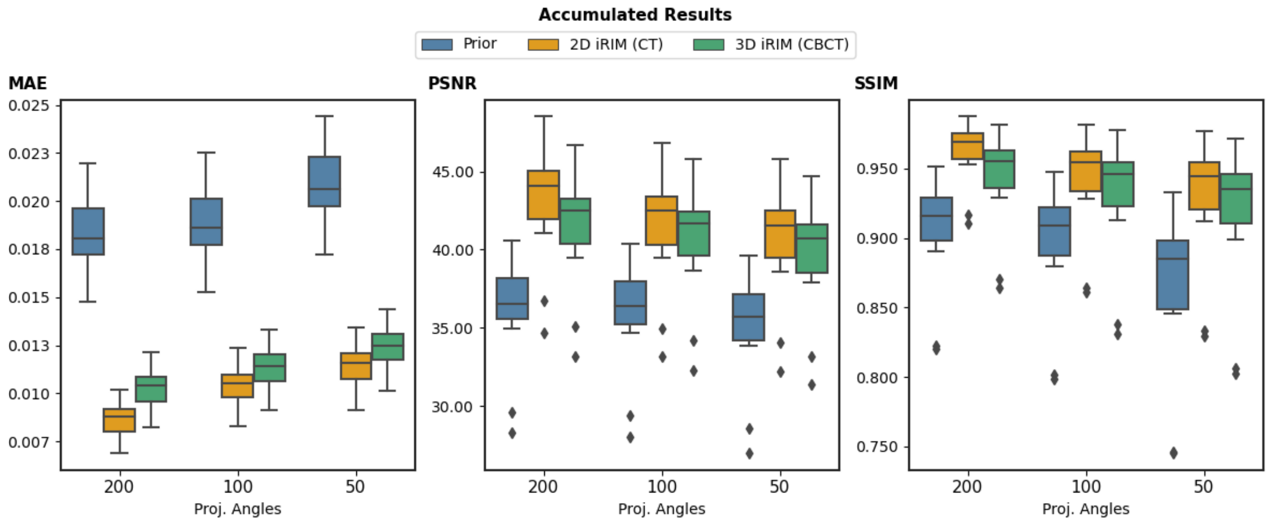


Figure 4. Whisker plots of the quantitative performance evaluation of the 2D and 3D iRIM models trained with datasets of 200, 100, and 50 projections against the ground truth.

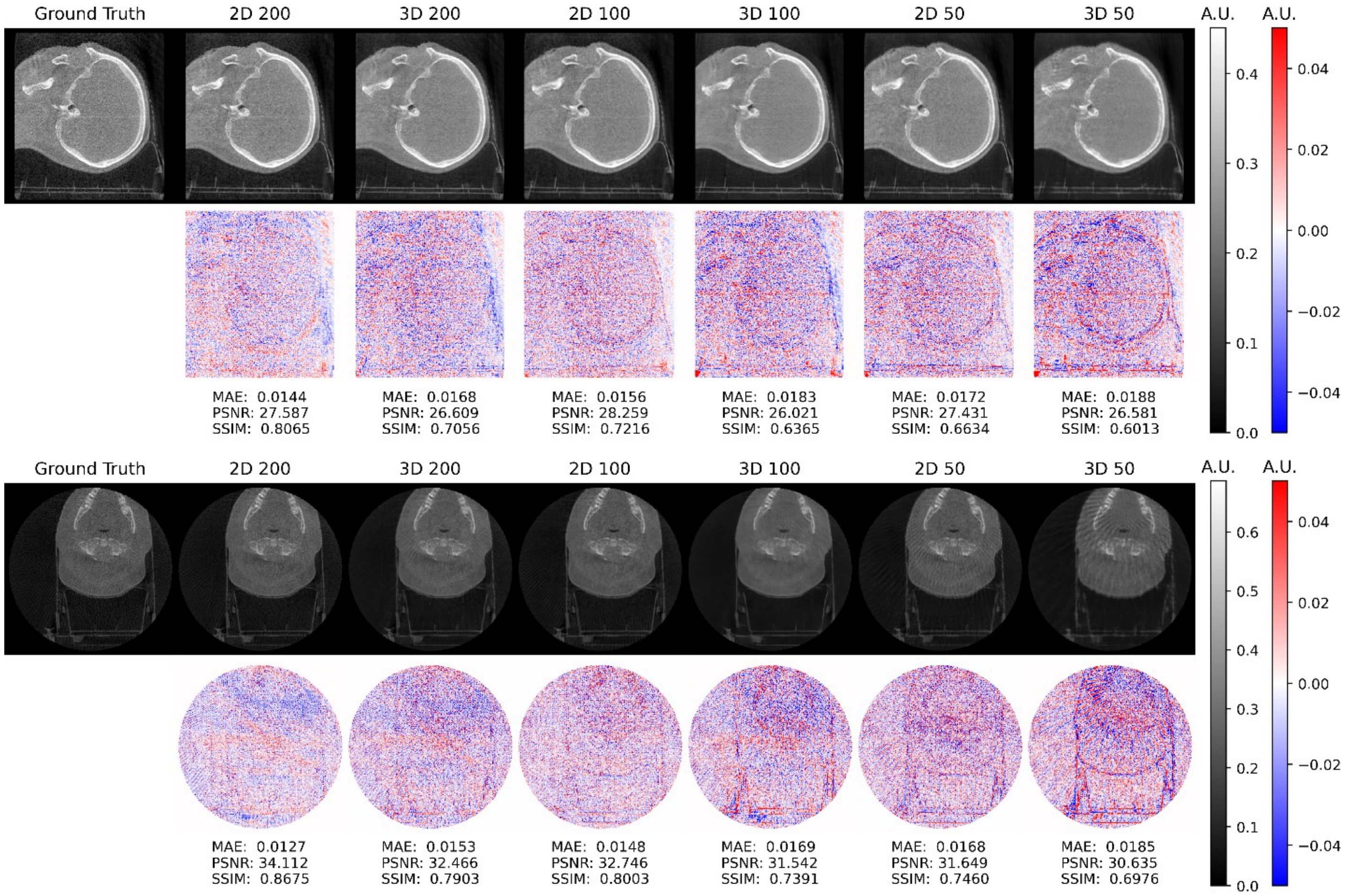


Figure 5. (Top) Sagittal slices from the 2nd best performing sample from the test dataset. (Bottom) Axial slices from the median performing sample from the test dataset. In the grayscale figures, the ground truth and the outputs of 2D and 3D iRIM models are sequentially depicted for all 200, 100 and 50 projection angle datasets. In the color-scaled images, their corresponding differences are depicted by subtracting them from the ground truth.

Feasibility

As found in Appendix II, the outputs of the 2D models yielded mean SSIMs of 0.96 (±0.02), 0.94 (±0.04), and 0.93 (±0.05 were obtained for the 200, 100, and 50 projection angle datasets, respectively. Similarly, the respective SSIM values for the 3D iRIM outputs were 0.94 (±0.04), 0.93 (±0.05) and 0.92 (±0.05). Comparisons of these outputs with their corresponding priors used as inputs consistently found their superior quantitative resemblance to the ground truth significant.

Qualitative analysis

Several observations could be noted from the qualitative assessment performed, the results of which can be found in Appendix V. As can be seen in the examples of Figure 5, the ground truth images contained a considerable amount of noise. Noise with similar characteristics could be observed in the 2D iRIM reconstructions. On the opposite side, the 3D iRIM reconstructions appeared to have smoother distributions with less noise.

A large number of the patients inspected had metallic implants in their oral cavities. These were found to introduce image artifacts such as streaking ones or the loss of contrast in the vicinity of the implants. The iRIM-based reconstructions qualitatively appeared to inherit these artifacts to an extent. In the qualitative comparison of their differences with their respective ground truths, however, the oral cavities with metal implant accumulated larger degrees of errors. These errors were observed to be propagated in the transaxial slices containing the implants.

Focusing on the effects of undersampling, image quality degradation could be observed across all cases. Increasing the degree of undersampling, increasingly made the noise and the streaking artifacts more prominent. The perceived image resolution was observed to be degrading with undersampling, with the effect being more visible in the axial slices and less visible in sagittal ones. Accurate edge preservation appeared to increasingly pose a challenge with undersampling. This was particularly prominent in high-contrast interfaces, such as between the soft tissue and air or bone.

Generally, a particularly clear visual difference between the second-best and median-performing patient scans could not be observed. In the sagittal slice of the second-worst-performing case, band patterns were observed in the extreme ends of the axial direction.

Discussion

CBCT reconstructions to the ones used as priors in model trainings. Similar was the case when two-fold and four-fold CBCT data undersamplings were used. Visual investigations verified the high level of agreement of the iRIM-based reconstructions to the FDK-generated ground truth, in the case of the full data. Full CBCT inferences were obtained within 12 s and can potentially be further reduced with hyperparameter tuning, as shown in this work, or pruning techniques. This is faster than typical FDK-based reconstruction which requires approximately 20 s for similar grid and voxel sizes [46]. It is therefore evident that an iRIM-based CBCT reconstruction method could be possible in the future.

In this work, the 3D CBCT inverse problem was approached in two different ways by using 3D and 2D iRIMs. In the former case, an attempt to accurately model the acquisition was made using 3D Radon-based forward and backward operators, and modeling the system geometry. In the latter case, the inverse problem was grossly approximated as a 2D parallel-beam CT inverse problem, using 2D Radon-based transforms as operators. Quantitatively, both approaches yielded comparable results. An insignificant but repeating pattern of the 2D iRIM models outperforming their 3D counterparts was observed. Considering the simplifications made in the 2D approach, this was unexpected. Visual inspections revealed that the 2D model generated reconstructions with similar noise characteristics as the provided ground truth, while the 3D ones generated smoother distributions. Therefore, it remains unclear as to which approach was superior, and training using noise-free CBCT images as ground truth is recommended. Noise correlation metrics such as the noise power spectrum [12], are used generally used for CT-based image reconstructions and their use should be considered in the future.

Regardless of the optimal approach, the overall performance of the 2D iRIM models is noteworthy. It might indicate that approximate modeling of the operators describing the acquisition can suffice and can be compensated for in the iRIM framework. Along this line of reasoning, we trained the 2D iRIM model with a four-fold parameter reduction. The resulting negligible impact on the performance of the model strongly indicates that the learning capacity of the original 2D iRIM by far exceeds the complexity of this task. To further investigate this hypothesis, trainings with further reduced depth are suggested. In the opposite direction, an increase in the depth of the 3D iRIM model followed by improved performance could also validate this hypothesis. Unfortunately, the size of the 3D iRIM can not be expanded due to current memory size limitations. In any case, the consistent performance after the four-fold parameter reduction of the 2D iRIM showcases indicates the existence of a large margin for architectural hyperparameter tuning.

To further investigate the robustness of the iRIM frameworks, trainings using sparse CBCT data were performed with both 2D and 3D models. By observing the increasing prevalence of image artifacts and noise in the priors due to the undersampling of the CBCT acquisitions, an impact on the performance was anticipated. Indeed, the performance of the iRIMs became progressively worse with an increasing amount of undersampling. Despite that, the inferences of iRIMs remained superior to their corresponding priors. Quantitatively, even the inferences based on the heaviest undersampled datasets were found superior to the full acquisition priors. These findings were regarded as evidence of the robustness of iRIMs.

Considerable attention would be interesting to be given to heavily undersampled acquisitions with fewer projection data than the number of unknowns. In that case, the linear system of the inverse problem becomes undetermined and this threshold was estimated at 76 projection angles using our current setup. Despite the case of 50 projection angles being included in our experiments, several more undersampled cases are required for such analysis. In this direction, an iRIM-based approach could potentially be applied for sparse-view CBCT reconstructions that aim to lower the radiation dose delivered to the patients by acquiring a limited number of projection angles. Due to its clinical relevance in diagnostic and surgical imaging applications, it is an actively researched topic with several conventional [47] and DL-based approached proposed [48], and even sparse-view 4D CBCT reconstruction challenge recently taking place [49].

With the primary focus of this work being the implementation and testing of iRIM models, simplifications were made that deviated from a realistic clinical application. Most notable is the decision to use synthetic sinograms as inputs that were generated based on the high-quality CBCT image. This was motivated by the practical limitations of directly training on raw sinograms, such as different sinogram sizes and acquisition arcs, which ought to be resolved in future studies.

Another concern stems from the absence of a clinic-grade high-quality image reconstruction method to base our feasibility criteria upon. Clinically implemented reconstruction algorithms have embedded, among others, scatter correction [15] and noise suppression algorithms [50]. Such additional post-reconstruction steps were not included in the generation of ground truth CBCT used in this study. Therefore, steps towards a clinically realistic scenario would include the use of raw sinograms as inputs and clinic-grade CBCTs as ground truth.

Specifically for ART purposes, accurate dose calculations are based on HU [7]. Therefore, a natural extension of the current work would be the use of high-quality CTs as ground truths. Due to different acquisition times, FoVs, and patient localization, a mismatch of the information present in CTs and CBCT projection data is expected. Since overcoming this mismatch is the motivation behind the use of unpaired unsupervised methods for CT-synthesis, such as GANs [23], interesting would be the impact on supervised methods such as iRIMs. This added complexity could be proven a valid test case for 2D-3D comparisons of iRIM architectures to be based on. Eventually, the HU accuracies of the synthetic CTs and their impact on dose estimations will determine the viability of any iRIM architecture to be used as an end-to-end method.

Conclusions

In this work, the feasibility of using iRIM frameworks for CBCT reconstruction has been showcased in multiple investigations. iRIM models with 2D and 3D architectures were able to provide reconstructions with high fidelity to the provided ground truth images based on simulated CBCT data. The hypothesized potential of iRIMs for an end-to-end approach was verified. Further investigations to integrate iRIMs into current clinical workflows for further assessment should be performed in prospective studies.

References

[1] G. M. (George M. MacKee, *X-rays and radium in the treatment of diseases of the skin*. Philadelphia, Lea & Febiger, 1921. Accessed: Oct. 18, 2022. [Online]. Available: http://archive.org/details/xraysandradiumi00mackgoog

[2] "News of Science," *Science*, vol. 125, no. 3236, pp. 18–22, Jan. 1957, doi: 10.1126/science.125.3236.18.

[3] M. K. Thompson *et al.*, "Practice-changing radiation therapy trials for the treatment of cancer: where are we 150 years after the birth of Marie Curie?," *Br J Cancer*, vol. 119, no. 4, Art. no. 4, Aug. 2018, doi: 10.1038/s41416-018-0201-z.

[4] "Linacs to narrow radiotherapy gap," *CERN Courier*, Dec. 21, 2021. https://cerncourier.com/a/linacs-to-narrow-radiotherapy-gap/ (accessed Oct. 18, 2022).

[5] D. Yan, F. Vicini, J. Wong, and A. Martinez, "Adaptive radiation therapy," *Phys. Med. Biol.*, vol. 42, no. 1, pp. 123–132, Jan. 1997, doi: 10.1088/0031-9155/42/1/008.

[6] O. L. Green, L. E. Henke, and G. D. Hugo, "Practical Clinical Workflows for Online and Offline Adaptive Radiation Therapy," *Semin Radiat Oncol*, vol. 29, no. 3, pp. 219–227, Jul. 2019, doi: 10.1016/j.semradonc.2019.02.004.

[7] C. K. Glide-Hurst *et al.*, "Adaptive Radiation Therapy (ART) Strategies and Technical Considerations: A State of the ART Review From NRG Oncology," *International Journal of Radiation Oncology, Biology, Physics*, vol. 109, no. 4, pp. 1054–1075, Mar. 2021, doi: 10.1016/j.ijrobp.2020.10.021.

[8] B. Bak *et al.*, "Criteria for Verification and Replanning Based on the Adaptive Radiotherapy Protocol 'Best for Adaptive Radiotherapy' in Head and Neck Cancer," *Life (Basel)*, vol. 12, no. 5, p. 722, Dec. 2022, doi: 10.3390/life12050722.

[9] S. Lim-Reinders, B. M. Keller, S. Al-Ward, A. Sahgal, and A. Kim, "Online Adaptive Radiation Therapy," *International Journal of Radiation Oncology, Biology, Physics*, vol. 99, no. 4, pp. 994–1003, Nov. 2017, doi: 10.1016/j.ijrobp.2017.04.023.

[10] D. Zheng, "The Use of Cone Beam Computed Tomography in Image-Guided Radiotherapy," *OMICS J Radiology*, vol. 01, no. 01, 2012, doi: 10.4172/2167-7964.1000e104.

[11] L. Lechuga and G. A. Weidlich, "Cone Beam CT vs. Fan Beam CT: A Comparison of Image Quality and Dose Delivered Between Two Differing CT Imaging Modalities," *Cureus*, vol. 8, no. 9, Sep. 2016, doi: 10.7759/cureus.778.

[12] C. C. Shaw, "3 Reconstruction algorithms," in *Cone Beam Computed Tomography*,

[13] J. Wang *et al.*, "An experimental study on the noise properties of x-ray CT sinogram data in Radon space," *Phys. Med. Biol.*, vol. 53, no. 12, pp. 3327–3341, Mar. 2008, doi: 10.1088/0031-9155/53/12/018.

[14] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, Dec. 2008, doi: 10.1088/0031-9155/53/17/021.

[15] Y. Xu *et al.*, "A practical cone-beam CT scatter correction method with optimized Monte Carlo simulations for image-guided radiation therapy," *Phys. Med. Biol.*, vol. 60, no. 9, pp. 3567–3587, Dec. 2015, doi: 10.1088/0031-9155/60/9/3567.

[16] M. Maspero *et al.*, "CBCT-to-CT synthesis with a single neural network for head-and-neck, lung and breast cancer adaptive radiotherapy," *Physics and Imaging in Radiation Oncology*, vol. 14, pp. 24–31, Apr. 2020, doi: 10.1016/j.phro.2020.04.002.

[17] I. J. Goodfellow *et al.*, "Generative Adversarial Networks." arXiv, Jun. 10, 2014. doi: 10.48550/arXiv.1406.2661.

[18] X. Liang *et al.*, "Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy," *Phys. Med. Biol.*, vol. 64, no. 12, p. 125002, Mar. 2019, doi: 10.1088/1361-6560/ab22f9.

[19] X. Wang *et al.*, "Synthetic CT generation from cone-beam CT using deep-learning for breast adaptive radiotherapy," *Journal of Radiation Research and Applied Sciences*, vol. 15, no. 1, pp. 275–282, Mar. 2022, doi: 10.1016/j.jrras.2022.03.009.

[20] X. Xue *et al.*, "Cone Beam CT (CBCT) Based Synthetic CT Generation Using Deep Learning Methods for Dose Calculation of Nasopharyngeal Carcinoma Radiotherapy," *Technol Cancer Res Treat*, vol. 20, p. 15330338211062416, Jan. 2021, doi: 10.1177/15330338211062415.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015, doi: 10.48550/arXiv.1505.04597.

[22] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," Nov. 2016, doi: 10.48550/arXiv.1611.07004.

[23] B. Chen, K. Xiang, Z. Gong, J. Wang, and S. Tan, "Statistical Iterative CBCT Reconstruction Based on Neural Network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1511–1521, Jun. 2018, doi: 10.1109/TMI.2018.2829896.

[24] Y. Archambault, "Making on-line adaptive radiotherapy possivle using artificial intelligenece andmachine learning for efficient daily dose re-planning.," p. 10, 2020, [Online]. Available: http://mpijournal.org/pdf/2020-02/MPI-2020-02-p077.pdf

[25] P. Putzky and M. Welling, "Recurrent Inference Machines for Solving Inverse Problems." arXiv, Jun. 13, 2017. doi: 10.48550/arXiv.1706.04008.

[26] P. Putzky and M. Welling, "Invert to Learn to Invert." arXiv, Nov. 25, 2019. doi: 10.48550/arXiv.1911.10914.

[27] D. Karkalousos, S. Noteboom, H. E. Hulst, F. M. Vos, and M. W. A. Caan, "Assessment of data consistency through cascades of independently recurrent inference machines for fast and robust accelerated MRI reconstruction," *Phys. Med. Biol.*, vol. 67, no. 12, p. 124001, Mar. 2022, doi: 10.1088/1361-6560/ac6cc2.

[28] K. Lønning, P. Putzky, J.-J. Sonke, L. Reneman, M. W. A. Caan, and M. Welling, "Recurrent inference machines for reconstructing heterogeneous MRI data," *Medical Image Analysis*, vol. 53, pp. 64–78, 2019, doi: 10.1016/j.media.2019.01.005.

[29] E. R. Sabidussi *et al.*, "Recurrent inference machines as inverse problem solvers for MR relaxometry," *Medical Image Analysis*, vol. 74, p. 102220, Dec. 2021, doi: 10.1016/j.media.2021.102220.

[30] W. R. Morningstar *et al.*, "Data-Driven Reconstruction of Gravitationally Lensed Galaxies using Recurrent Inference Machines," *ApJ*, vol. 883, no. 1, p. 14, Sep. 2019, doi: 10.3847/1538-4357/ab35d7.

[31] A. M. Gurusamy Muthuvelrabindran, "Invertible recurrent inference machines for low-dose computed tomography," Jul.

2021. https://essay.utwente.nl/86820/ (accessed Oct. 18, 2022).

[32] J. A. Fessler, "Statistical Image Reconstruction Methods for Transmission Tomography," vol. PM80, pp. 1–71, Jun. 2000, doi: 10.1117/3.831079.ch1.

[33] "Principles of Computerized Tomographic Imaging." https://www.slaney.org/pct/pct-toc.html (accessed Oct. 18, 2022).

[34] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A, JOSAA*, vol. 1, no. 6, pp. 612–619, Jun. 1984, doi: 10.1364/JOSAA.1.000612.

[35] J. Zbontar *et al.*, "fastMRI: An Open Dataset and Benchmarks for Accelerated MRI." arXiv, Dec. 11, 2019. doi: 10.48550/arXiv.1811.08839.

[36] Elekta Ltd., *XVI R5.0 Instructions for Use p.64*.

[37] S. Rit, M. Vila Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, and G. C. Sharp, "The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK)," in *International Conference on the Use of Computers in Radiation Therapy (ICCR) 2013*, Melbourne, Australia, Feb. 2013, vol. 489, no. 1, p. 012079. doi: 10.1088/1742-6596/489/1/012079.

[38] D. L. Parker, "Optimal short scan convolution reconstruction for fan beam CT," *Medical Physics*, vol. 9, no. 2, pp. 254–257, 1982, doi: 10.1118/1.595078.

[39] J. Adler, H. Kohr, and O. Öktem, "Operator Discretization Library (ODL)." Zenodo, Jan. 17, 2017. doi: 10.5281/zenodo.249479.

[40] P. Putzky *et al.*, "i-RIM applied to the fastMRI challenge." arXiv, Oct. 20, 2019. doi: 10.48550/arXiv.1910.08952.

[41] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization." arXiv, Jan. 29, 2017. doi: 10.48550/arXiv.1412.6980.

[42] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[43] S. van der Walt *et al.*, "scikit-image: image processing in Python," *PeerJ*, vol. 2, p. e453, Jun. 2014, doi: 10.7717/peerj.453.

[44] A. K. Nagarajappa, N. Dwivedi, and R. Tiwari, "Artifacts: The downturn of CBCT image," *J Int Soc Prev Community Dent*, vol. 5, no. 6, pp. 440–445, 2015, doi: 10.4103/2231-0762.170523.

[45] "SciPy 1.0: fundamental algorithms for scientific computing in Python | Nature Methods." https://www.nature.com/articles/s41592-019-0686-2 (accessed Oct. 18, 2022).

[46] M. van Herk *et al.*, "First clinical experience with cone-beam CT guided radiation therapy; evaluation of dose and geometric accuracy," *International Journal of Radiation Oncology, Biology, Physics*, vol. 60, no. 1, p. S196, Sep. 2004, doi: 10.1016/j.ijrobp.2004.06.137.

[47] C. Xu, B. Yang, F. Guo, W. Zheng, and P. Poignet, "Sparse-view CBCT reconstruction via weighted Schatten p-norm minimization," *Opt. Express, OE*, vol. 28, no. 24, pp. 35469–35482, Nov. 2020, doi: 10.1364/OE.404471.

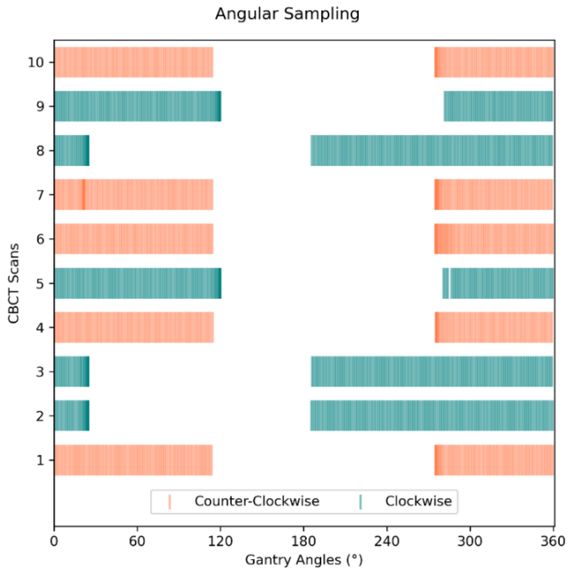
[48] L. Chao, Z. Wang, H. Zhang, W. Xu, P. Zhang, and Q. Li, "Sparse-view cone beam CT reconstruction using dual CNNs in projection domain and image domain," *Neurocomputing*, vol. 493, pp. 536–547, Jul. 2022, doi: 10.1016/j.neucom.2021.12.096.

[49] C.-C. Shieh *et al.*, "SPARE: Sparse-view reconstruction challenge for 4D cone-beam CT from a 1-min scan," *Medical Physics*, vol. 46, no. 9, pp. 3799–3811, 2019, doi: 10.1002/mp.13687.

[50] X. Jia, Y. Lou, R. Li, W. Y. Song, and S. B. Jiang, "GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation," *Medical Physics*, vol. 37, no. 4, pp. 1757–1760, 2010, doi: 10.1118/1.3371691.

Appendix I – Angular sampling inhomogeneity

For the figure below, 10 CBCT scans were selected randomly out of our datasets. For each scan, the angles of the projections acquired are indicated with a single line, color-coded for the orientation of the gantry rotation. We can observe that despite the use of a standardized preset, each acquisition was characterized by a unique set of projection angles, ultimately leading to inconsistent geometric information contained in sinograms across different scans. As can be observed in figure below, multiple inconsistencies can be noted. Different directions of the gantry rotation across scans, as well as differences in the starting and ending angles of the acquisition arc, induce a large variability in sinograms that correspond to virtual rotations of the imaged geometry along the axis of the system. In addition, inhomogeneous angular sampling during the gantry rotation is observed. This is caused by missing samples or decelerations of the rotation, which are especially pronounced towards the last samples of the acquisition. As a result, sinograms have different dimensions across different scans, with numbers of angular samples within the range of 190 and 210, and different acquisition arcs in the range of 190-220 degrees.



Appendix II – iRIM evaluation results

Results of performance evaluation, as defined by the output - ground truth difference, are listed in the following table for all iRIM cases used in this study.

Samples: 13			**MAE**		**PSNR**		**SSIM**	
Projections Angles	**Type of Data compared to GT**	**Training**	Mean	Std	Mean	Std	Mean	Std
200	**Prior**	-	0.0184	0.0022	35.99	3.52	0.9041	0.0406
	iRIM Outputs	2D	0.0085	0.0010	42.95	3.81	0.9613	0.0231
		2D, SSIM	0.0085	0.0010	42.92	3.84	0.9617	0.0229
		2D, Reduced	0.0088	0.0010	42.67	3.81	0.9614	0.0219
		3D	0.0102	0.0012	41.34	3.77	0.9423	0.0362
100	**Prior**	-	0.0189	0.0023	35.76	3.53	0.8943	0.0460
	iRIM Outputs	2D	0.0103	0.0012	41.34	3.81	0.9408	0.0377
		3D	0.0112	0.0013	40.50	3.77	0.9289	0.0455
50	**Prior**	-	0.0209	0.0023	34.94	3.58	0.8643	0.0583
	iRIM Outputs	2D	0.0113	0.0013	40.43	3.79	0.9276	0.0463
		3D	0.0123	0.0013	39.59	3.77	0.9156	0.0532

Appendix III – P-tables

The results (p-values) of all comparisons using paired Mann-Whitney U tests are listed in the following tables.

200	Prior									
	2D	1.7e-5								
	3D	1.4e-5	3.5e-3							
100	Prior	4.4e-1								
	2D	1.7 e-5	2.5e-3		1.7e-5					
	3D	1.7 e-5		5.1e-2	7.3e-2	7.3e-2				
50	Prior	1.3e-1			8.1e-1					
	2D	9.7e-5	9.7e-5					1.7e-5		
	3D	1.7 e-5		1.5e-3			5.8e-2	1.7e-5	8.1e-2	
		Prior	2D	3D	Prior	2D	3D	Prior	2D	3D
MAE		200			100			50		

200	Prior									
	2D	5.9e-4								
	3D	2.1e-3	2.0e-1							
100	Prior	7.2e-1								
	2D	2.5e-3	2.0e-1		1.8e-3					
	3D	3.5e-3		3.8e-1	2.9e-3	3.8e-1				
50	Prior	2.8e-2			4.1e-1					
	2D	3.5e-3	6.4e-2			4.4e-1		2.1e-3		
	3D	7.7e-3		1.8e-1			3.7e-1	2.9e-3	4.1e-1	
		Prior	2D	3D	Prior	2D	3D	Prior	2D	3D
PSNR		200			100			50		

200	Prior									
	2D	3.3e-4								
	3D	4.8e-3	6.5e-2							
100	Prior	3.8e-1								
	2D	5.6e-3	5.8e-2		3.5e-3					
	3D	2.7e-2		2.8e-1	1.0e-2	3.6e-1				
50	Prior	3.6e-2			5.8e-2					
	2D	3.1e-2	1.4e-2			3.3e-1		3.5e-3		
	3D	2.4e-1		8.1e-2			3.6e-1	5.6-3	3.8-1	
		Prior	2D	3D	Prior	2D	3D	Prior	2D	3D
SSIM		200			100			50		

Original MSE	1.7e-5	
Original SSIM	1.7e-5	4.7e-5
Reduced MSE	1.7e-5	9.6e-1
MAE	Prior 200	Original MSE

Original MSE	5.9e-4	
Original SSIM	7.1e-4	5.7e-1
Reduced MSE	5.9e-4	9.6e-1
PSNR	Prior 200	Original MSE

Original MSE	3.1e-4	
Original SSIM	2.7e-4	8.8e-1
Reduced MSE	3.3e-4	7.6e-1
SSIM	Prior 200	Original MSE

Appendix IV – Undersampled investigations

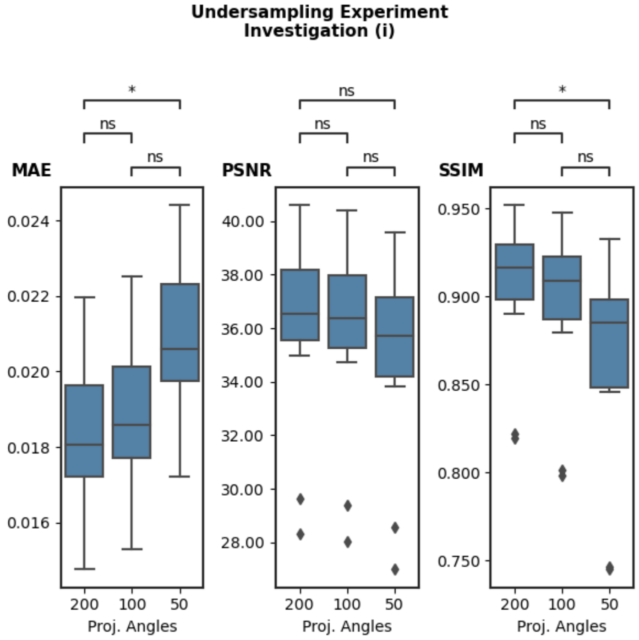
To assist the analysis of the undersampled investigation on the overall results of Figure 4, Appendix II and III, a series of whisker plots were made, with annotated statistical significance as estimated on the paired Mann-Whitney U test. These were grouped as:

i) Comparison of the Priors(angles), for angles in 200, 100, 50.

ii) Prior(angles)-Output(angles), for angles in 200, 100, 50, and both 2D and 3D.

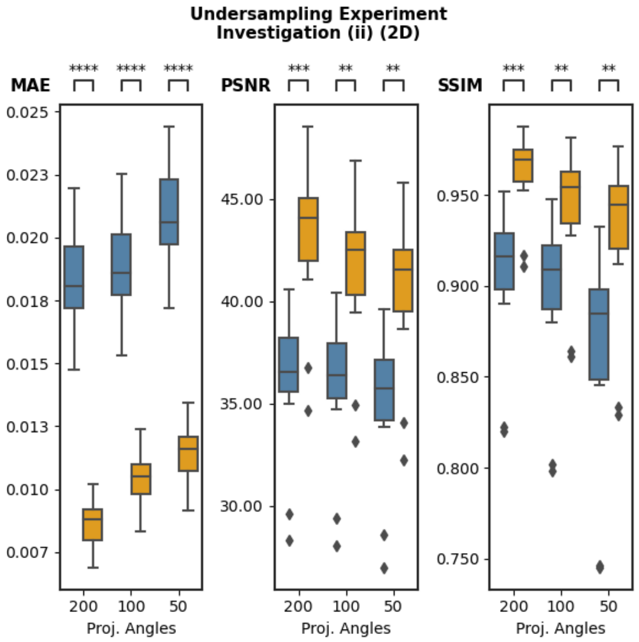
iii) Prior(200)-Output(angles), for angles in 100, 50, and both 2D and 3D.

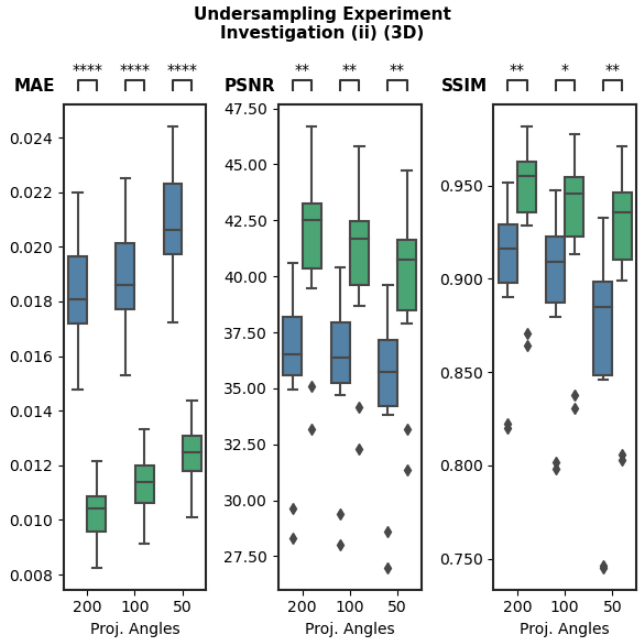
iv) Comparison of the Outputs(angles), for angles in 200, 100, 50, and both 2D and 3D.

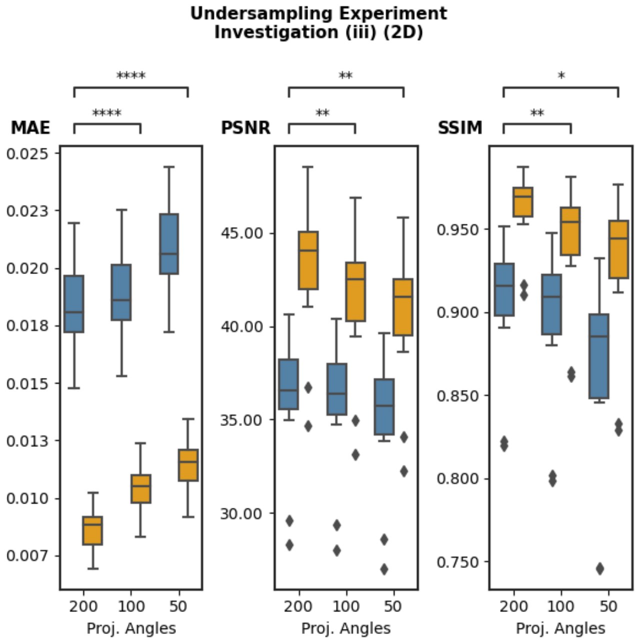


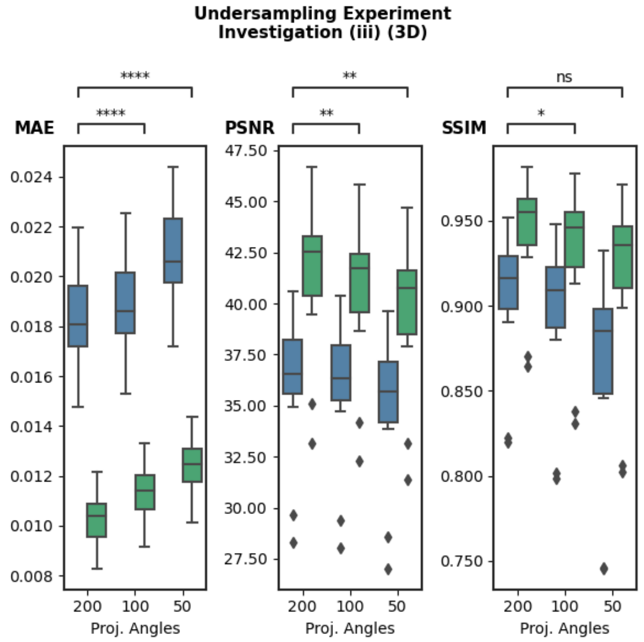
P-Values Legend

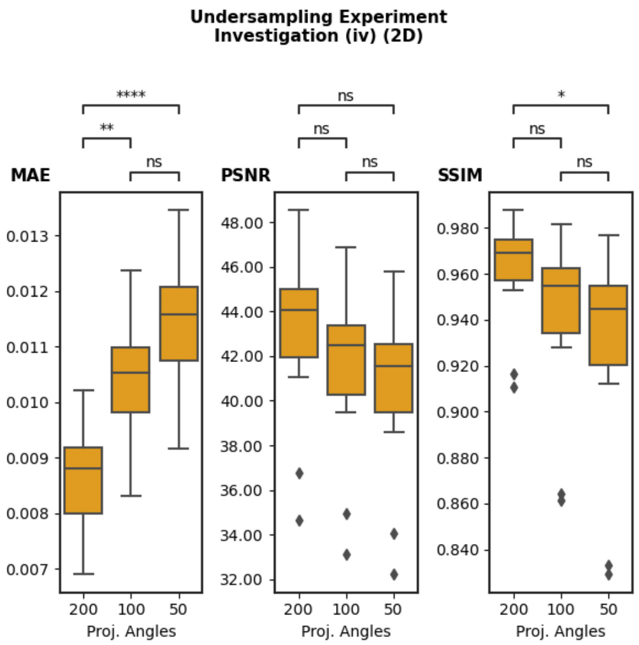
	From	To
ns	0.05	1
*	0.01	0.05
**	0.001	0.01
***	0.0001	0.001
****	0	0.0001

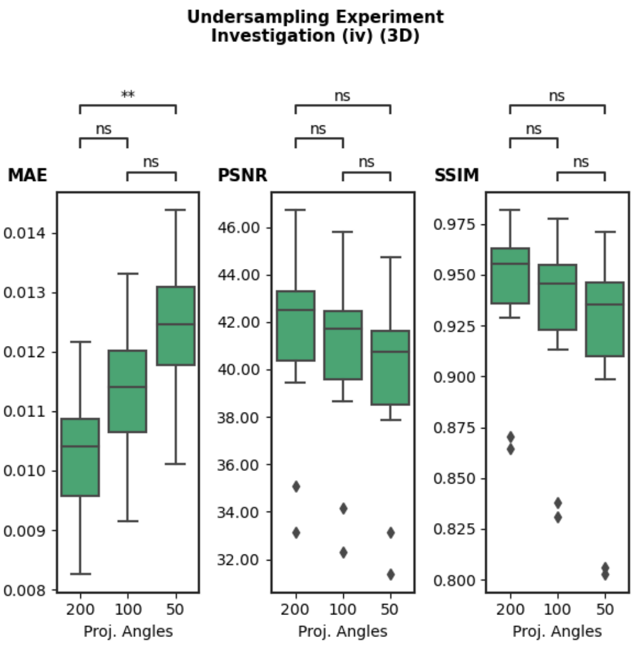





Undersampling Experiment
Investigation (iii) (2D)
MAE
PSNR
SSIM
Proj. Angles


Undersampling Experiment
Investigation (iii) (3D)
MAE
PSNR
SSIM
Proj. Angles


Undersampling Experiment
Investigation (iv) (2D)
MAE
PSNR
SSIM
Proj. Angles

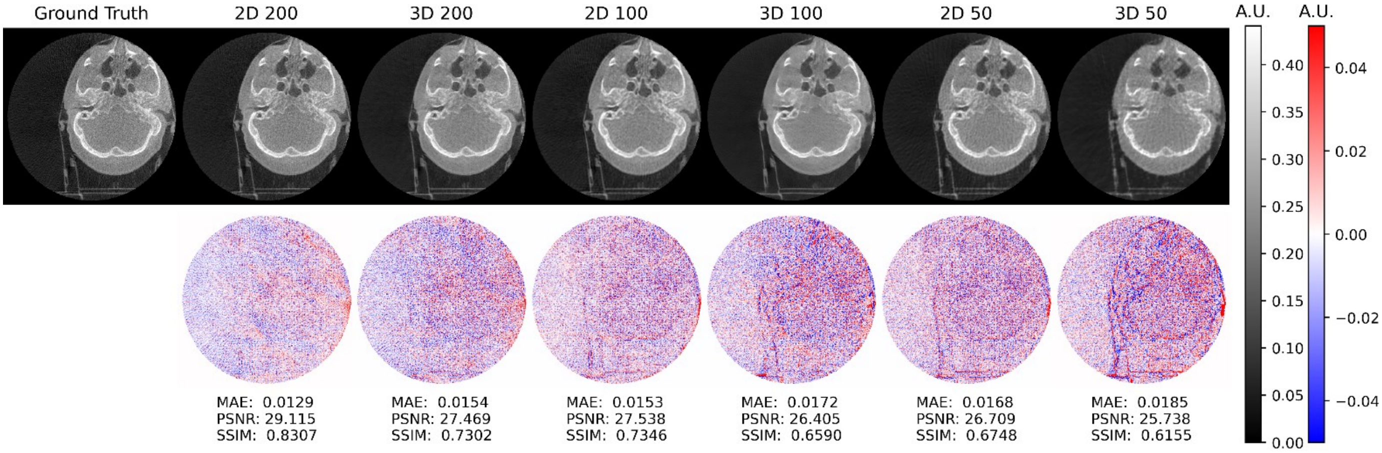

Undersampling Experiment
Investigation (iv) (3D)
MAE
PSNR
SSIM
Proj. Angles

Appendix V – Visualizations

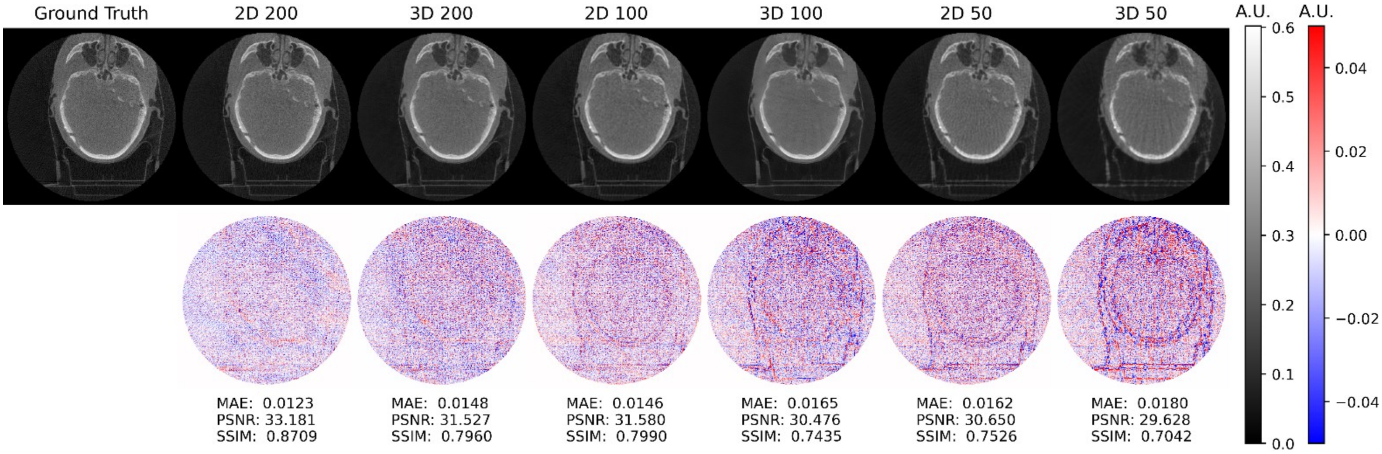
Axial and sagittal slices from the second-best, median, and second-worst rated patient scan out of the test data of 13 patient scans are visualized, as rated by SSIM. The axial slices were selected arbitrarily, while the mid-slice was selected for the sagittal ones. For each example, the ground truth and the outputs of the 2D, 3D iRIM models trained with the 200, 100, and 50 projection angles datasets are visualized, along with their differences from the ground truth (ground truth-output).

Axial Slices

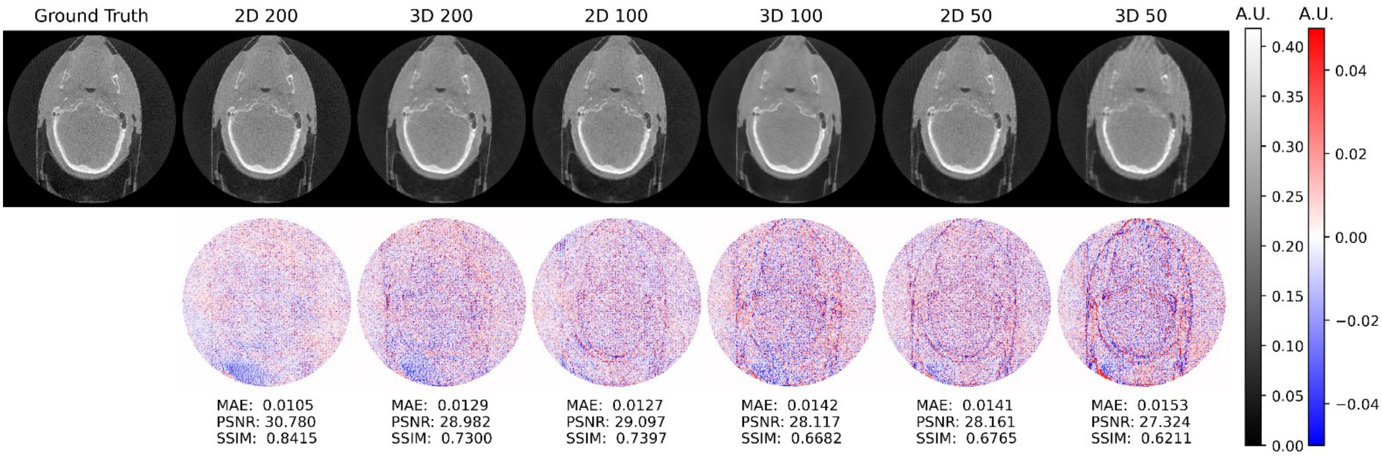
2nd Best Case



Median Case

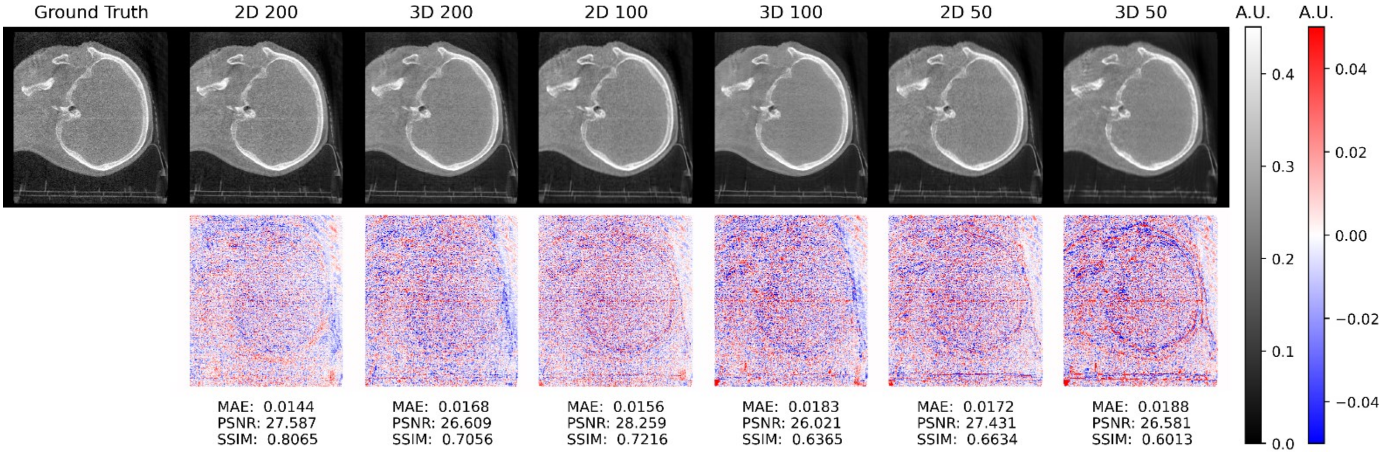


2nd Worst Case

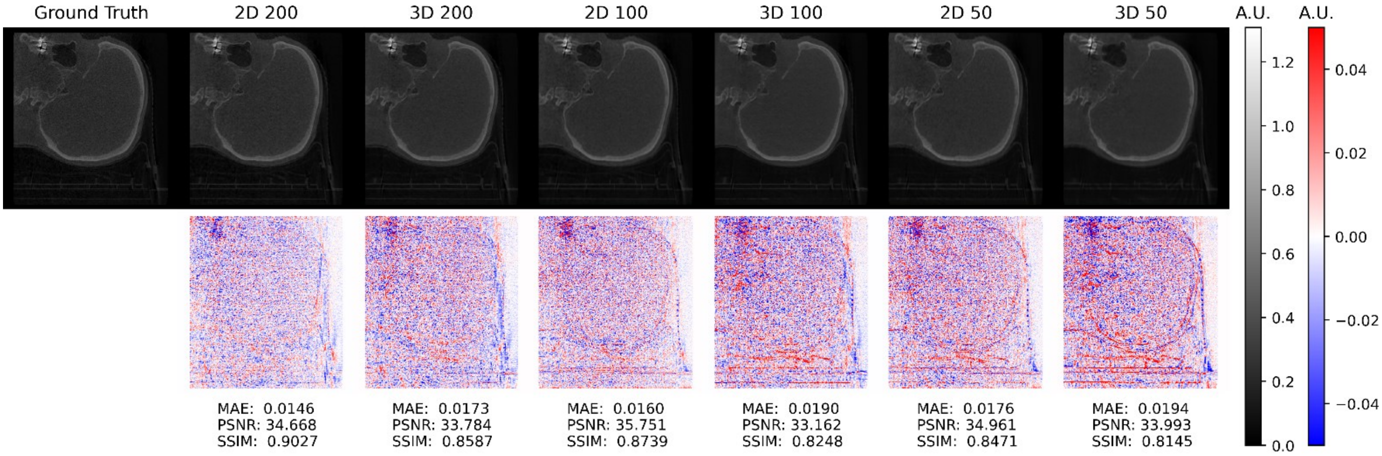


Sagittal Slices

2nd Best Case



Median Case



2nd Worst Case