Title

On-board Image Driven View Synthesis and Volume Reconstruction by Integrating Neural Representation with Iterative Reconstruction

Author

Tsai, Yun-Han

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

On-board Image Driven View Synthesis and Volume Reconstruction by Integrating Neural Representation with Iterative Reconstruction

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Bioengineering

by

Yun-Han Tsai

2024

ABSTRACT OF THE THESIS

On-board Image Driven View Synthesis and Volume Reconstruction

by Integrating Neural Representation with Iterative Reconstruction

by

Yun-Han Tsai

Master of Science in Bioengineering

University of California, Los Angeles, 2024

Professor Dan Ruan, Chair

Onboard imaging (OBI) is widely used in radiotherapy for alignment, delivery verification, and adaption. However, its acquisition is subject to configuration constraints and pulsing control, compromising view optimality and decision quality. This study combines the complementary advantages of inference power from Generative Radiance Fields (GRAF) representation and robustness from iterative reconstruction targeting instance; and alleviates their respective limitations in volume reconstruction and efficiency. Initialize with a pre-trained GRAF representation, an online instance learning module refines the GRAF using the instantaneous projection from an arbitrary angle acquired during treatment. A view from an alternative angle is generated to provide clinically preferable triggers for real-time setup review or beam adaptation decisions. To support offline delivery verification, dense synthesized projections are fed into an

iterative reconstruction module to yield the full volumetric rendering. Qualitative and quantitative evaluations were performed for 2D projection query-generation and 3D volume reconstruction. Vessel bifurcation landmarks were used as surrogates for tumor targets, and their localization accuracy was analyzed to reflect motion-tracking efficacy. Detailed characterization of dependency on query angle, number of inputs, and operational conditions was reported. Our assessment indicates that paired query with instance learning enhances model's adaptation to target subjects in reducing nMSE by 17% for synthesized projections and 33% for volumetric rendering, and improving localization accuracy by 47% in bifurcation landmark analysis. This study shows the promise of using single or pair real-time projection for view toggling and volume tracking, demonstrating its utility for onboard monitoring, adaptation, and retrospective reconstruction in radiotherapy. We are working on further improving absolute accuracy and reconstruction resolution.

The thesis of Yun-Han Tsai is approved.

Daniel A. Low

Holden H. Wu

Dan Ruan, Committee Chair

University of California, Los Angeles

2024

To my parents Jack and Cellina,

to my sister Winnie,

and to my beloved Cassidy,

for their love, understanding, support, sacrifices, and companionship.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

ACKNOWLEDGEMENTS

This thesis would not have been possible without the countless people I have had the privilege of calling my peers, friends, and family. First and foremost, I would like to thank my advisor, Prof. Dan Ruan, for her endless guidance and support throughout my time at her lab. She has given me countless opportunities to prove myself while always pushing me toward better. Her passion for academia inspired me to enjoy learning, and her profound knowledge guided me to become a more professional researcher. She is a respected scholar, a caring teacher, and a great role model in life. I am very happy and grateful to have Prof. Dan Ruan as my advisor.

I would also like to thank my committee members, Prof. Daniel A. Low and Prof. Holden H. Wu, for their valuable feedback and supervision on both research and thesis. Their insights and assistance have also been a source of great support for me. My sincere appreciation also goes to all the incredible members at the Radiation Oncology Laboratory. I am fortunate to have worked with such a talented and warm team. I am sincerely grateful for the moments of mutual support and the memories filled with laughter.

Last but more importantly, I would like to extend my gratitude to my loving family and girlfriend, for their unwavering, unconditional support and encouragement in everything. They cheer my small and big moments and made a lot of sacrifices to let me concentrate on my academic pursuits. Distance has never been a barrier to feeling your love. Thank you all for being there through the ups and downs of my journey.

Chapters 2 to 4 and part of Chapter 1 are versions of a manuscript submitted for publication to Medical Physics, co-authored by myself and my advisor, Prof. Dan Ruan. The manuscript is currently under review.

Chapter 1: Introduction

1.1 Background

Advancements in radiation therapy have enabled highly targeted and controlled energy delivery to eliminate tumor cells while sparing surrounding healthy tissue. However, the complexity of physiological movements makes it challenging for clinicians to guarantee accuracy in treatment. Contemporary onboard imaging (OBI) systems and therapeutic apparatus are frequently limited by their fixed configuration to the gantry and the risk of collisions, restricting their capability to provide the most clinically advantageous viewpoints[1], which are essential for radiologists to adjust the radiation beam as tumors and organs move during treatment. Additionally, volume tracking and motion management remain a critical challenge. Hence, we aim to address two high-demand tasks: (1) to virtually toggle viewing angles to align with the treatment beam or other specific clinical requirements, and (2) to achieve spatiotemporal full-volume reconstruction.

1.2 Neural Radiance Fields (NeRF) in Radiation Therapy

Recent innovations in deep learning have demonstrated promising potential in radiotherapy applications. Neural Radiance Fields (NeRF) have been instrumental in estimating complete three-dimensional volumetric representation with encoded radiance field and density of given scenes[2]. It represents scenes as a continuous function and is optimized to reconstruct images from multiple viewpoints.

NeRF operates by sampling points along the rays passing through the pixels in input images. These spatial coordinates and viewing directions query a neural network that outputs the color and density for each point, then are differentially accumulated along each ray to produce

the final pixel values in the synthesized image. This process allows NeRF to render novel views that were unseen in the training data.

NeRF offers promising potential for improving visualization in the context of radiation therapy. It can be used to synthesize alternative views from arbitrary angles, aiding in real-time setup review and beam adaptation decisions. Additionally, its ability to query detailed volumetric representations can achieve motion management and dose accumulation tracking. These capabilities are crucial for addressing the challenges posed by physiological movements during treatment.

Subsequently, Generative Radiance Fields (GRAF) integrated these characteristics within the Generative Adversarial Network (GAN) model to accomplish scene reconstructions with self-supervision[3-5]. Nevertheless, GRAF-based models are observed to be insensitive to details, and insufficient to capture complex anatomy in medical imaging. The values and limitations of these advanced models must be appreciated with well-defined clinical goals.

1.3 Alternative Solutions and Proposed Method

There are other possible methods that are commonly used in this field, including statistical iterative reconstruction, compressive sensing, and direct deep learning reconstruction[6-12]. Iterative algorithms are known for their robustness and ability to produce high-quality images, especially in scenarios where data is sparse or noisy; Compressive sensing leverages the sparsity of medical images in various transform domains such as wavelet and Fourier to efficiently recover high-quality reconstructions from a small set of measurements; On the other hand, direct deep learning reconstruction approaches offer potential improvements in speed and accuracy. These techniques are beneficial in applications where high-quality reconstructions are essential for accurate diagnosis and treatment planning. Despite their promise, these methods also come

with their limitations, such as demanding computational complexity and availability in real-time scenarios.

This study positions the proposed method as a complementary approach to these alternative solutions. To address the challenges mentioned above within a GRAF-based deep network, we explored the practical implications of employing instance learning with a quantitative-focused objective. Iterative statistical reconstruction was further implemented to obtain a robust volumetric rendering and mitigate the loss of image quality due to downsampling in both angle and detector dimensions. This pipeline allows the system to dynamically adapt to patient-specific changes and render projections from various angles shortly; and further provide a comprehensive solution for onboard monitoring and retrospective reconstruction in radiation therapy. Even though it is hard to solve reconstruction problems based on a single or a pair of projections with iterative algorithms directly, they offer an effective framework for integrating volumetric consistency and prior knowledge. We extensively investigated various aspects of synthesizing diverse projections with different query projection numbers and angles to comprehend their clinical applicability.

To assess the contribution of instance learning in different clinical scenarios, we conducted thorough quantitative and qualitative assessments of 2D digitally reconstructed radiograph (DRR) renderings and 3D computed tomography (CT) reconstruction. In addition, vessel bifurcation landmark analysis was used to reflect tumor localization accuracy for treatment management. In summary, this study aims to provide a more accurate and dynamic visualization to monitor and adapt to target movement during therapeutic procedures.

Chapter 2: Method and Materials

We have developed a complete pipeline to translate information from a single-view X-ray projection into a prescribed view angle or the underlying volumetric CT.

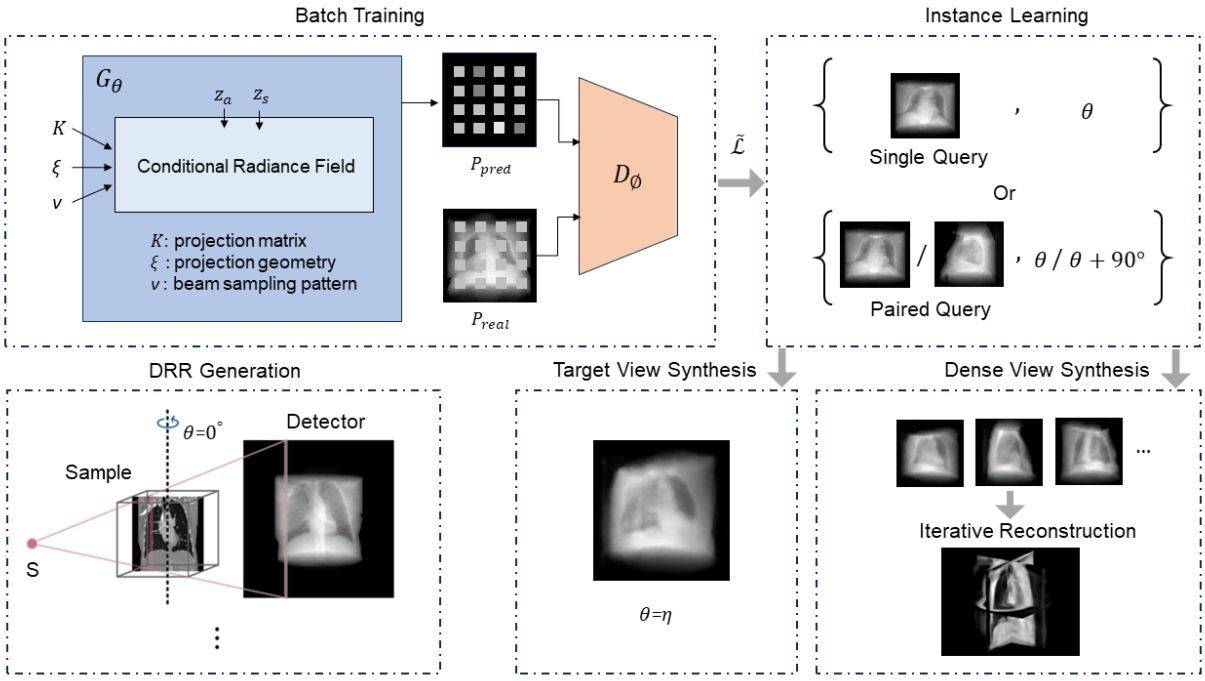


Figure 2-1: Overall schematic. The DRR Generation module creates ground-truth projections from the reference 3D CT for GRAF batch training. In the instance learning stage, two setups (1) a single coronal view, and (2) a combination of coronal and sagittal views are introduced to refine the generator with objective $\tilde{\mathcal{L}}$. Projections are synthesized at both toggle-target angle η for online decision-making, or at dense angles for retrospective spatiotemporal reconstruction.

Figure 2-1 illustrates the overall schematic of our process. To prepare data for model training and systematic evaluation, we forward project the reference CT data to generate DRRs. These projections provide benchmarks for quality during the view-toggle inference stage or the three-dimensional volume rendering. Structural identity and the implicit neural attenuation field representation are learned with a baseline GRAF network. Finally, instance learning is applied to adapt the underlying field and to generate projections at alternative angles with single/paired

query input; or dense projections that are then used with iterative reconstruction to yield an instantaneous estimate of the CT volume.

2.1 Generative Radiance Fields (GRAF) Overview

Generative Radiance Fields (GRAF) represents an advancement by integrating the strengths of Generative Adversarial Networks (GAN) with Neural Radiance Fields (NeRF), promising 3D-aware image synthesis with precise control over shape and appearance.

In the adapted GRAF framework in our study, the generator G_θ is tasked with predicting image patches using inputs such as the projection matrix K, projection geometry ξ, a beam sampling pattern ν, and latent codes for shape z_s and appearance z_a. These inputs are encoded by a multi-layer perceptron (MLP) to create a conditional radiance field. It maps 3D locations and viewing directions, conditionally influenced by the latent codes, to radiation attenuation response c and volume density σ of the sampled points. The discriminator D_ϕ, serving as the adversarial counterpart to the generator, compares the synthesized image P_{pred} patch against a real patch P_{real} extracted from the training dataset. It is implemented as a convolutional neural network that classifies patches as real or synthetic, to provide gradients for updating the generator towards more realistic image synthesis.

2.2 On-query Instance Learning for Quantitative Alignment

It is important to update the NeRF representation with instantaneous information to reflect the inter- or intra-subject changes and dynamics from the static reference representation. We perform instance learning of single query to fine-tune part of the generator parameters together with the latent vectors for shape z_s and appearance z_a[13-14]. This process is guided by a weighted objective of perceptual loss with respect to feature encoding from VGG 16 and mean squared error (MSE) from the acquired OBI image[15]. This form is typically used to improve the trade-

off between perceptual quality and distortion in GAN models[16]. The instance learning generator objective is illustrated in **Equation 2-1**.

Equation 2-1

$$\mathcal{L}_{\text{gen}} = \lambda_1 \mathcal{L}_{\text{perception}}(\text{G};\ \text{VGG16}) + \lambda_2 \mathcal{L}_{\text{MSE}}(\text{G})\ ,$$

Where weights λ_1 and λ_2 determine the relative importance of perception feasibility and goodness of fit.

We investigate the impact of instance learning on the quality of the rendered projection images and the resultant volume reconstruction from two scenarios: (1) a single coronal view, and (2) paired coronal and sagittal views. Instance learning of various query view angles is also systemically evaluated. These assessments are meant to provide insights into the adaptability and precision of the model from limited and varied viewpoints across different imaging configurations.

Our application context of both view angle toggling and CT reconstruction strongly focuses on quantitative accuracy. Since the instance learning is performed based on one or two query view images, it is reasonable to expect the compliance to human visual perception and consistency across different views to be inherited from the NeRF training, and won't be compromised much by the proposed refinement.

2.3 Experiments and Implementation Details

We utilized the available public TCIA lung CT datasets[17-19]. Among the comprehensive repositories that include a variety of scan types, we selected non-contrast standard CT studies to ensure consistency and homogeneity of appearance. We used CT scans of 6 training subjects and 5 testing subjects.

Forward projection was performed according to cone-beam geometry, employing a 50 keV mono-energetic X-ray source (based on NIST standards), with Source-to-Axis Distance (SAD) and Source-to-Detector Distance (SDD) as specified in the patient DICOM files, to generate 72 digitally reconstructed radiographs (DRRs) for each subject, at intervals of 5 degrees for complete circumferential coverage throughout a full rotation of 360 degrees. Projections were recorded with a size of 128×128 pixels. The coordinate system was established using the right-hand rule, with the coronal view set as the baseline at zero degrees and the Z-axis oriented in the superior direction.

For the main GRAF model, the training was conducted using 432 instances with a batch size of 8 over 150,000 iterations. It took approximately 50 hours to train on a single Quadro RTX 8000 GPU. The module configuration and hyperparameters were closely aligned with the original specifications suggested by Schwarz et al[5,20].

For online instance learning, the relative weight of perception and MSE loss λ_1 and λ_2 was explored over various combinations. To assess how a single query from different angles affects the synthesized output, we systematically adjusted the query angle in increments of 30 degrees, reaching a maximum of 150 degrees. Furthermore, the paired query from two orthogonal angle projections (coronal and sagittal views) was introduced to analyze the potential advantages of incorporating information from an additional query angle. The instance learning process proceeded with 1,500 iterations for a single query and 3,000 iterations for the paired query.

For volume rendering, a dense set of synthesized projections at 72 distinct angles were synthesized and the Order Subsets Adaptive Steepest Descent Projection Onto Convex Set (OS-ASD-POCS) algorithm[21] was used for reconstruction. The configuration was optimized, with the

subset size, the number of total variation (TV) steps within each iteration, and the TV weight were set to values of 10, 10, and 0.001, respectively.

We quantitatively and qualitatively evaluated the model's performance on the 2D projection rendering and 3D volumetric reconstruction. For two-dimensional assessment, the synthesized outputs were compared against ground-truth DRRs. The normalized mean squared error (nMSE) and normalized mean absolute error (nMAE) were calculated to measure the quantitative difference, and the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were computed to analyze visual similarity and conspicuity. The MSE, nMSE, nMAE, PSNR, and SSIM calculations were given in **Equation 2-2**.

Equation 2-2

$$\mathrm{MSE} = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(I_{\mathrm{gt},i} - I_{\mathrm{recon},i}\right)^2,$$

$$\mathrm{nMSE} = \frac{\sum_{i=1}^{N}\left(I_{\mathrm{gt},i} - I_{\mathrm{recon},i}\right)^2}{\sum_{i=1}^{N} {I_{\mathrm{gt},i}}^2},$$

$$\mathrm{nMAE} = \frac{\sum_{i=1}^{N}\left|I_{\mathrm{gt},i} - I_{\mathrm{recon},i}\right|}{\sum_{i}\left|I_{\mathrm{gt},i}\right|},$$

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\max {I_{gt}}^2}{\mathrm{MSE}}\right),$$

$$\mathrm{SSIM} = \frac{(2\mu_{\mathrm{gt}}\mu_{\mathrm{recon}} + c_1)(\sigma_{\mathrm{gt}}\sigma_{\mathrm{recon}} + c_2)(\sigma_{\mathrm{gt,recon}} + c_3)}{({\mu_{\mathrm{gt}}}^2 + {\mu_{\mathrm{recon}}}^2 + c_1)({\sigma_{\mathrm{gt}}}^2 + {\sigma_{\mathrm{recon}}}^2 + c_2)(\sigma_{\mathrm{gt}}\sigma_{\mathrm{recon}} + c_3)},$$

Where $I_{\mathrm{gt},i}$ represents the intensity value of the ground-truth image at pixel i, and $I_{\mathrm{recon},i}$ denotes the pixel values of the reconstructed image; μ_{gt}, μ_{recon}, σ_{gt}, σ_{recon} are the mean and standard deviations of the ground-truth and the reconstructed images, respectively; $\sigma_{gt,recon}$ represents the covariance of the ground-truth and reconstructed images. The luminance, contrast, and structure are equally weighted to unity. All configuration constants are set according to convention: L being

the dynamic range of the pixel-values; $\kappa_1 = 0.01$ and $\kappa_2 = 0.03$, $c_1 = (\kappa_1 L)^2$, $c_2 = (\kappa_2 L)^2$, and $c_3 = c_2/2$.

For volumetric accuracy, nMSE, nMAE, PSNR, and SSIM were compared between the query-driven CT reconstruction and its corresponding benchmark. Moreover, we conducted vessel bifurcation landmark identification and compared the displacement accuracy.

Chapter 3: Results and Discussion

3.1 Evaluation of 2D View Toggle Performance

To assess the 2D synthesis performance of our model, we investigated the weighting of perception (λ_1) vs MSE (λ_2) components in the objective for instance learning, with extreme cases where instance learning was guided solely by one of them. **Table 3-1** shows that a composite with $\lambda_1 = 3, \lambda_2 = 1$ was most favorable, and we adopt this setting for subsequent analyses.

$\boldsymbol{\lambda_1}$	$\boldsymbol{\lambda_2}$	**MSE ($\mu \pm \sigma$)**	**MAE ($\mu \pm \sigma$)**	**PSNR ($\mu \pm \sigma$)**	**SSIM ($\mu \pm \sigma$)**
0	1	0.977 ± 0.004	1.511 ± 0.054	32.06 ± 0.23	0.74 ± 0.01
1	0	0.969 ± 0.066	1.454 ± 0.387	32.15 ± 0.08	0.76 ± 0.04
1	3	0.943 ± 0.083	1.294 ± 0.243	32.62 ± 0.44	**0.81 ± 0.02**
3	1	**0.938 ± 0.055**	**1.275 ± 0.336**	**32.71 ± 0.52**	0.81 ± 0.04

Table 3-1: 2D renderings evaluation for the optimal balance of perception and MSE within instance learning.

Figure 3-1 provides a visual comparison of 2D renderings for view toggling tasks without and with instance learning. Although the original GRAF model represents a reasonable anatomical appearance without instance learning, it lacks the capability to adjust to the new instantaneous anatomical change. On the other hand, instance learning with a single query improves faithfulness to the specific patient across the query angle and the synthesized angles. It demonstrates a discernible improvement in accuracy and image quality while preserving 3D consistency.

The potential value of query input is reported in **Figure 3-2** where single query at 0° or paired query inputs at 0° or 90° are used to generate projections at a rendering angle of $\theta = 45^\circ$. It shows that on top of the benefit from instance learning with a single query, additional projection (orthogonal to the original in our experiment) further enhances the rendering performance. **Table 3-2** reports the various metrics employed to quantitatively compare the renderings from different scenarios against the ground-truth projections. Instance learning driven by a single query

significantly improves across all the metrics. Additional query input further enhances the overall results, leading to a 17% reduction in nMSE.

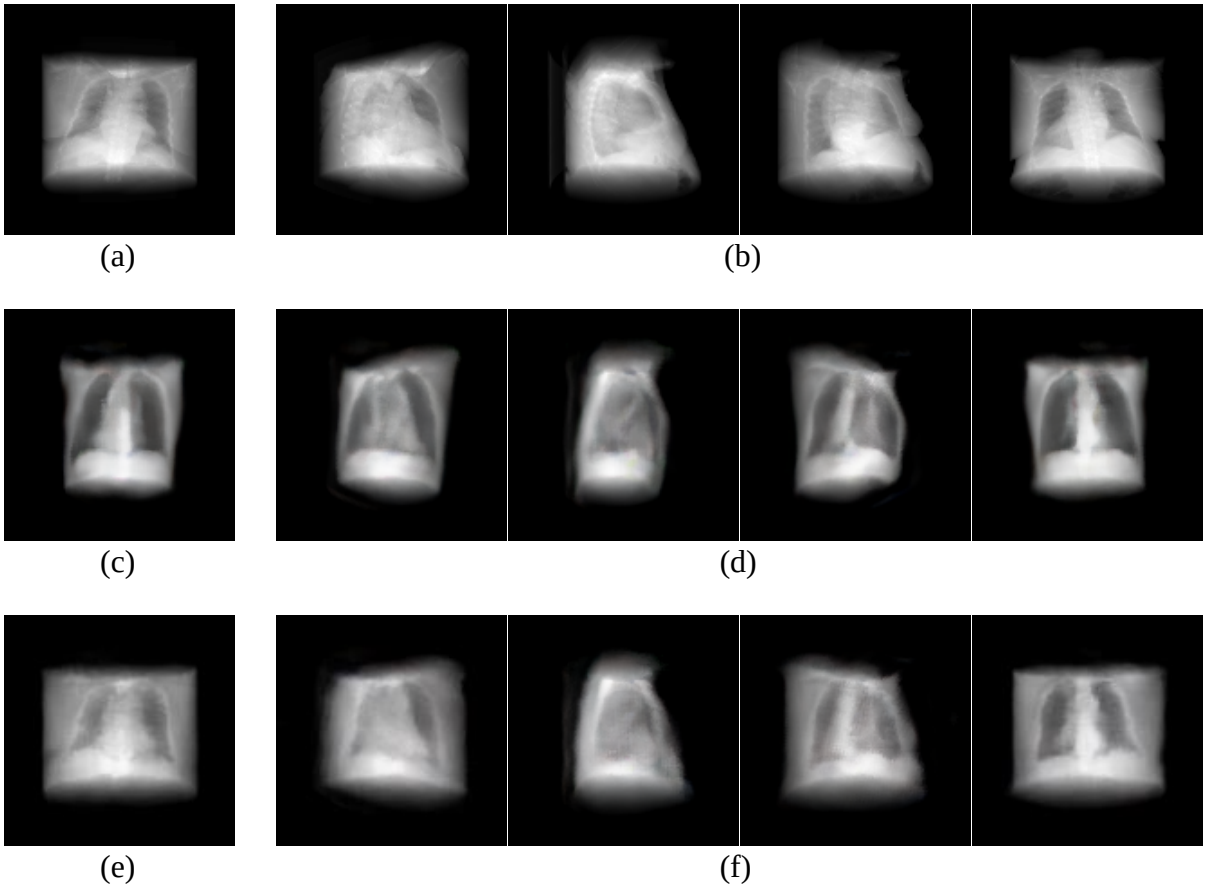
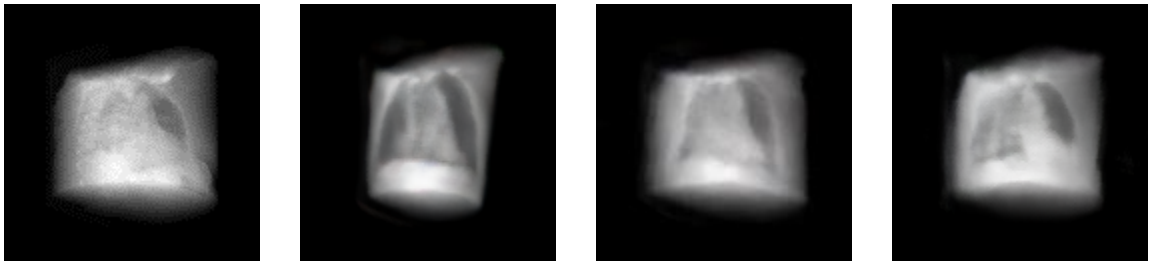


Figure 3-1: Comparison of view toggling performance w/o and w/ instance learning. Upper row displays ground-truth DRR images: (a) input query image at 0°, followed by (b) DRR images at 45°, 90°, 135°, and 180°. (c)(d) show the results of direct GRAF queries, while (e)(f) present the outcomes of GRAF queries employing instance learning based on (a).

To illustrate the information content and adaptation contribution depending on query angle, **Table 3-3** compares the quantitative results of single query instance learning from various query angles. **Figure 3-3** illustrates nMSE and SSIM as a function of input query angle. The coronal view is found to yield the optimal overall performance. Furthermore, it is observed that the quality of the generated images decreases as the query angle approaches the sagittal view, and the model exhibits similar performance with modules π due to symmetry as expected. This observation

aligns with expectations, given that the sagittal views are subject to more pronounced obstructions from the rig-cage and longer path length, thereby diminishing the intrinsic information content in driving evolution in instance learning. Consequently, the transition from the sagittal to the coronal view establishes a 13% decrease in nMNE for these synthesized projections.

Figure 3-4 shows the changes in the synthesized projections as the optimization progresses, which offers a reference for expected outcomes from instance learning under various conditions.



(a) (b) (c) (d)

Figure 3-2: Query output results at 45° for various conditions, where (a) ground truth (b) single query w/o instance learning (c) single query w/ instance learning (d) paired query w/ instance learning.

Method	MSE ($\mu \pm \sigma$)	MAE ($\mu \pm \sigma$)	PSNR ($\mu \pm \sigma$)	SSIM ($\mu \pm \sigma$)
single query	1.071 ± 0.055	1.385 ± 0.405	32.12 ± 0.68	0.73 ± 0.06
single query + IL	0.938 ± 0.055	1.275 ± 0.336	32.71 ± 0.52	0.81 ± 0.04
paired query + IL	**0.893 ± 0.043***	**1.205 ± 0.168***	**32.91 ± 0.56***	**0.82 ± 0.03***

Table 3-2: Comparative evaluation of 2D renderings applying instance learning (IL). (*: $p < 0.0001$, one-tailed paired t-test, compared to single query w/o instance learning)

Query angle	MSE ($\mu \pm \sigma$)	MAE ($\mu \pm \sigma$)	PSNR ($\mu \pm \sigma$)	SSIM ($\mu \pm \sigma$)
0°	0.938 ± 0.055 ↓	1.275 ± 0.336 ↓	32.71 ± 0.52 ↑	0.81 ± 0.04 ↑
30°	0.942 ± 0.050	1.334 ± 0.256	32.69 ± 0.61	0.79 ± 0.03
60°	0.983 ± 0.051	1.471 ± 0.098	32.50 ± 0.48	0.78 ± 0.03
90°	1.077 ± 0.109 ↑	1.495 ± 0.330 ↑	32.12 ± 0.27 ↓	0.74 ± 0.02 ↓
120°	0.974 ± 0.049	1.379 ± 0.231	32.54 ± 0.84	0.78 ± 0.05
150°	0.945 ± 0.069	1.291 ± 0.056	32.67 ± 0.39	0.79 ± 0.02

Table 3-3: Comparative evaluation of 2D renderings from various query angles.

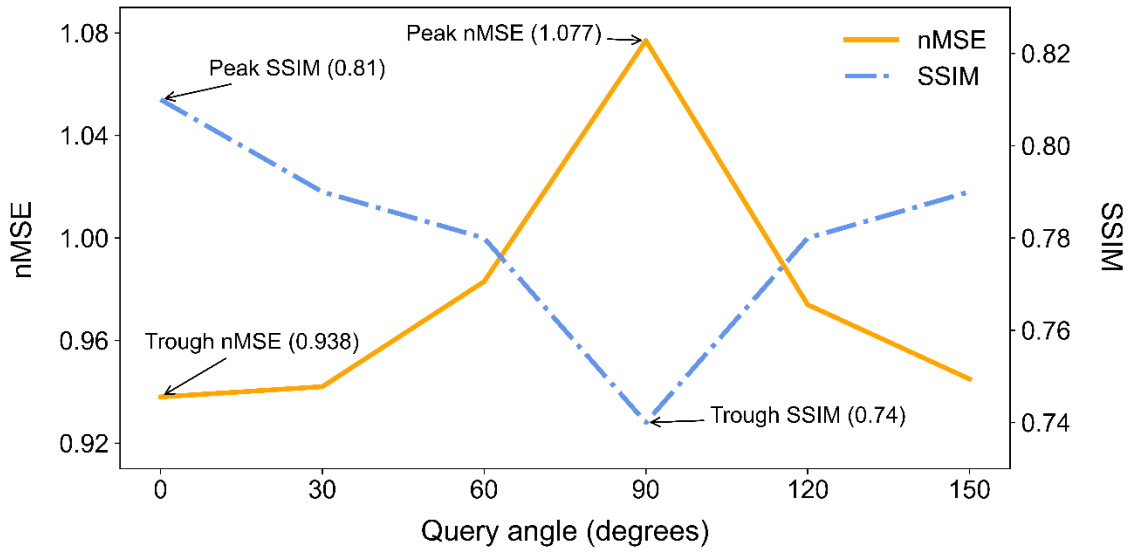
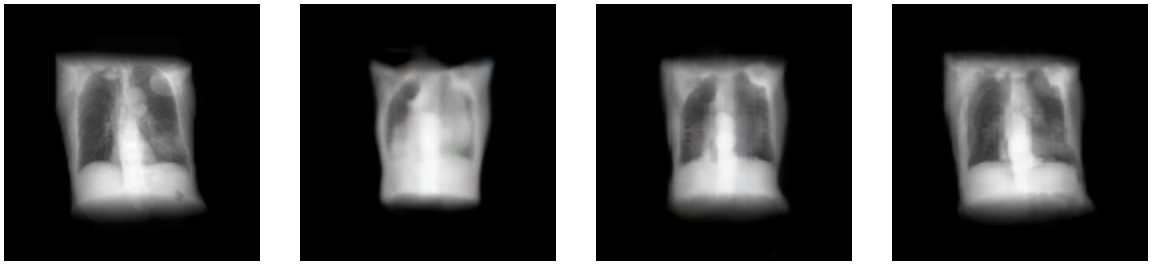


Figure 3-3: Variation of nMSE and SSIM in 2D renderings with respect to input query angles. Optimal quality is achieved at 0° (coronal) with the minimal nMSE and maximal SSIM; while the worst quality is observed at 90° (sagittal) with the maximal nMSE and minimal SSIM.



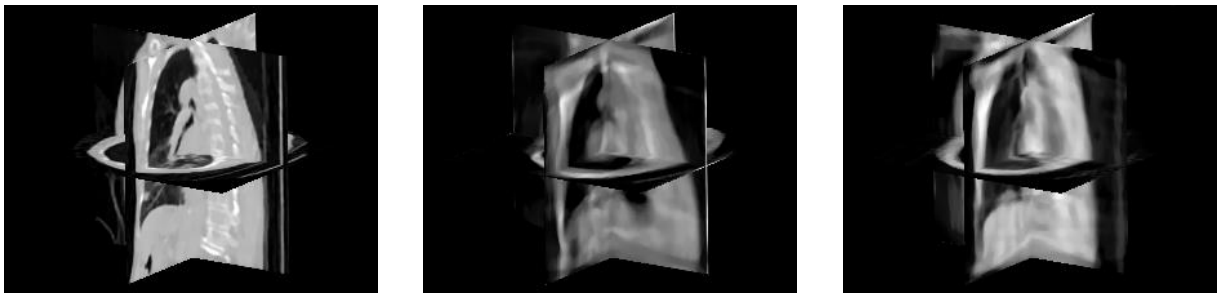
(a) (b) (c) (d)

Figure 3-4: Evolution of synthetic projections as the number of instance learning iterations increases. (a) DRR ground truth. (b) w/o instance learning (c) 750 iterations (d) 1,500 iterations.

3.2 Evaluation of 3D Reconstruction Accuracy

Figure 3-5 displays the reconstructed CT scans both without and with instance learning, compared against the benchmark reconstruction. Instance learning effectively incorporates the instantaneous information of specific targets compared to the initial GRAF model, enabling the model to align more closely with the actual 3D anatomical features of the target.

The empirical results of applying single and paired instance learning to the reconstructed volumes are displayed in **Tables 3-4**, **3-5**, and **Figure 3-6**. The outcomes are consistent with observations for 2D rendering. Instance learning with paired query images lead to an average 33% reduction in nMSE. In addition, instance learning employing the coronal view, as opposed to the sagittal view, results in reconstruction with a 12% decrease in nMSE.



(a) (b) (c)

Figure 3-5: 3D ortho-slice view of the volumetric reconstruction. (a) benchmark reconstructed from 72 ground-truth projections. CT volume reconstruction from a single query (b) without and (c) with instance learning.

Method	**MSE ($\mu \pm \sigma$)**	**MAE ($\mu \pm \sigma$)**	**PSNR ($\mu \pm \sigma$)**	**SSIM ($\mu \pm \sigma$)**
single query	0.719 ± 0.118	0.848 ± 0.064	10.86 ± 0.96	0.38 ± 0.02
single query + IL	0.622 ± 0.049	0.785 ± 0.029	11.45 ± 0.65	0.46 ± 0.01
paired query + IL	**0.482 ± 0.029***	**0.697 ± 0.017***	**12.55 ± 0.49***	**0.49 ± 0.01***

Table 3-4 Comparative evaluation of 3D CT reconstruction applying instance learning. (*: $p < 0.01$, one-tailed paired t-test, compared to the method w/o instance learning)

Query angle	**MSE ($\mu \pm \sigma$)**	**MAE ($\mu \pm \sigma$)**	**PSNR ($\mu \pm \sigma$)**	**SSIM ($\mu \pm \sigma$)**
0°	0.622 ± 0.049 ↓	0.785 ± 0.029 ↓	11.45 ± 0.65 ↑	0.46 ± 0.01 ↑
30°	0.644 ± 0.159	0.794 ± 0.102	11.32 ± 1.71	0.45 ± 0.02
60°	0.687 ± 0.164	0.806 ± 0.104	11.23 ± 1.47	0.45 ± 0.03
90°	0.707 ± 0.135 ↑	0.829 ± 0.100 ↑	11.08 ± 0.77 ↓	0.42 ± 0.03 ↓
120°	0.705 ± 0.091	0.794 ± 0.060	11.34 ± 0.43	0.43 ± 0.01
150°	0.655 ± 0.067	0.792 ± 0.042	11.38 ± 0.66	0.44 ± 0.01

Table 3-5: Comparative evaluation of 3D CT reconstructions from various query angles.

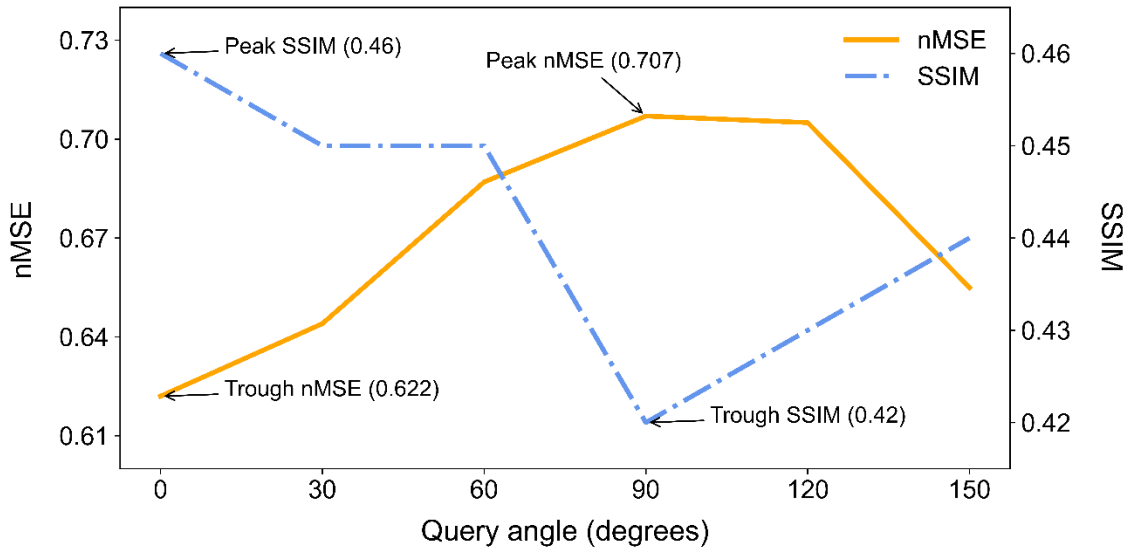
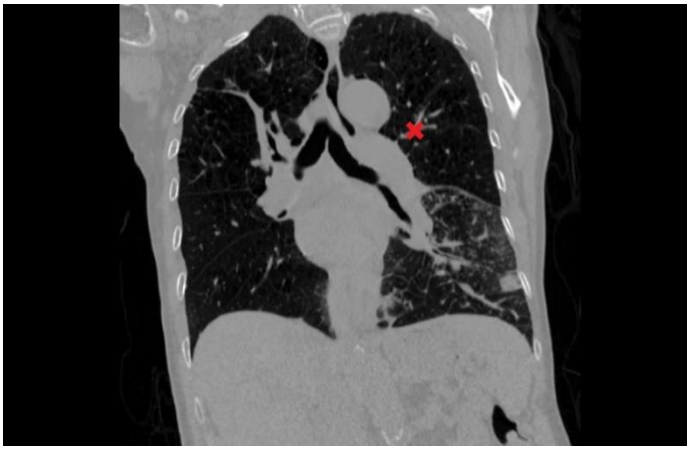


Figure 3-6: Variation of nMSE and SSIM in 3D CT reconstruction with respect to input query angles.



Method	**Displacement, mm ($\mu \pm \sigma$)**
single query	26.50 ± 9.12
single query + IL	19.42 ± 4.74
paired query + IL	**14.00 ± 8.18**

Table 3-6: Bifurcation landmark evaluation of reconstructed volumetric w/o and w/ instance learning

Table 3-6 illustrates and reports the localization on the reference and displacement corresponding to ground truth under various conditions. Paired query instance learning improves the localization accuracy by approximately 47%. However, observed errors due to blurriness remain above clinical requirements, which is possibly inherited from GRAF.

Table 3-7 demonstrates an enhancement in 3D volumetric outcomes when applying OS-ASD-POCS algorithm in conjunction with synthesized projections. Compared to the sampled volume rendering from GRAF and its subsequent improvement with instance learning, we observe a reduction in nMSE by 14% without instance learning and 7% with instance learning, respectively. These metrics are calculated by comparing the volume against the corresponding ground-truth CT data, indicating a more accurate and robust estimation of an iterative algorithm. At the same time, it is crucial to note that this approach generally incurs a higher time cost relative to direct GRAF volume rendering. Illustrative visualizations for each method can be found in **Figure 3-7**.

Method	**MSE($\mu \pm \sigma$)**	**MAE($\mu \pm \sigma$)**	**PSNR($\mu \pm \sigma$)**	**SSIM($\mu \pm \sigma$)**
direct GRAF	0.845±0.133	1.046±0.164	10.83±0.88	0.25±0.04
GRAF+OS-ASD-POCS	0.721±0.117	0.853±0.061	10.93±1.04	0.35±0.02
direct GRAF+IL	0.671±0.002	0.827±0.103	11.05±0.71	0.36±0.04
GRAF+IL+OS-ASD-POCS	**0.624±0.049***	**0.788±0.026***	**11.52±0.77**	**0.42±0.01***

Table 3-7: Comparison of 3D volumetric reconstruction or outcome: iterative reconstruction using synthetic projections versus GRAF direct volume rendering. (*: $p < 0.05$, one-tailed paired t-test, compared to the direct GRAF+IL method)

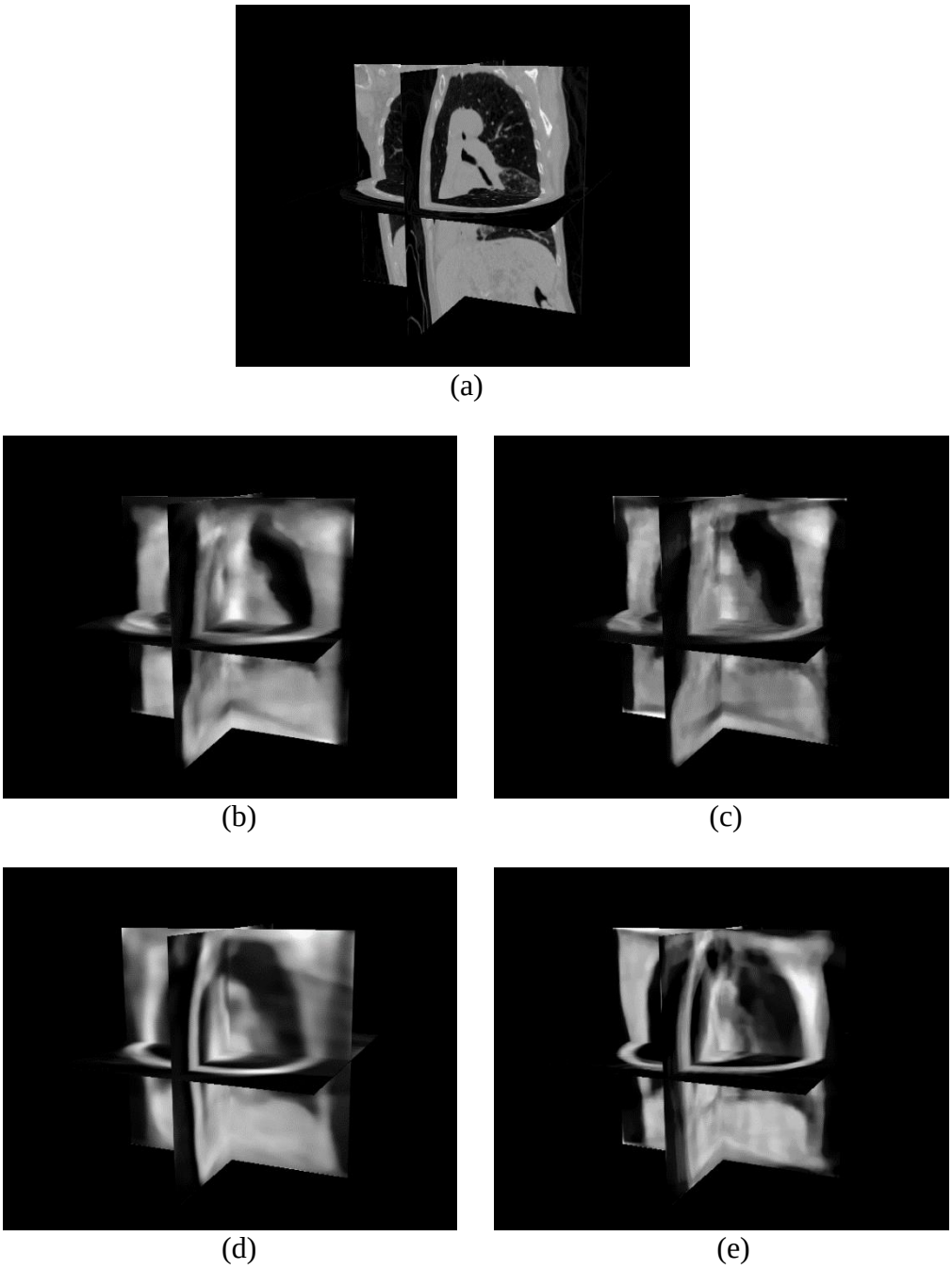


Figure 3-7: Comparative visualization of direct GRAF volume rendering and CT reconstruction. (a) Ground-truth CT ortho-slice. (b) direct GRAF volume rendering without instance learning. (c) OS-ASD-POCS without instance learning. (d) direct GRAF volume rendering with instance learning (e) OS-ASD-POCS with instance learning.

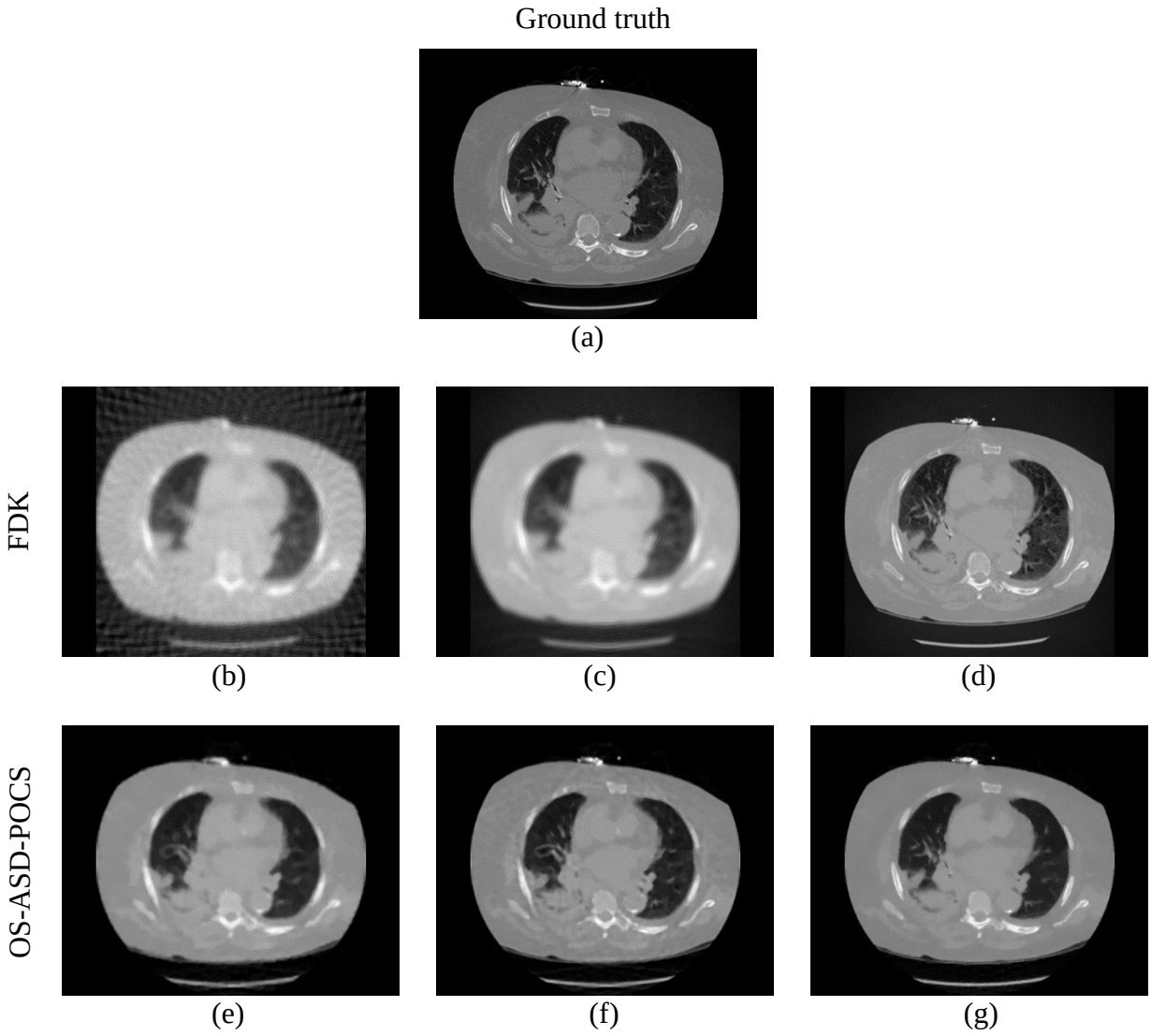


Figure 3-8. Comparative visualization of CT. (a) ground-truth reference data. (b) FDK from benchmark configuration of 72 angular projections with 128×128 resolution. (c) FDK with increased angular sampling density of 630 angular projections with 128×128 resolution. (d) FDK with 72 angular projections at 1000×1000 resolution. (e)(f)(g) showcase corresponding enhancements using OS-ASD-POCS for the configurations described in (b)(c)(d), respectively.

Chapter 4: Concluding Remarks and Summary

To demonstrate the core functionality of view toggling and reconstruction, we have used DRRs to provide training samples and benchmark synthesis performance. While DRRs are known to differ from realistic radiographs in scatter and specific patterns, they are similar in characteristics to X-ray images enough for clinical alignment and structural depiction, and specifically for motion management where the precise integrated attenuation is incidental compared to the localization of the tumor target in the context of surrounding tissue.

It is known that insufficient angular and detector resolution can induce Gibbs, streaking, and Moire artifacts, among others (shown in **Figure 3-8**). While GRAF based model allows us to synthesize at dense angular angles, the intrinsic information is limited. OS-ASD-POCS improves robustness and achieves decent performance under basic benchmark configurations, allowing us to strike a balance between image quality and time cost trade-offs.

Given that our process integrates sequential DRR generation and reconstruction effectively, we possess the freedom and flexibility to explore alternative reconstruction approaches, allowing us to adjust the framework according to different needs and time requirements. For example, we could consider utilizing deep network reconstruction[22-23], which provides both the benefit of enhanced reconstruction quality and time effectiveness, to offer clinically acceptable dose calculation accuracy. One needs to take spatial caution against deep-learning related hallucination for structure tracking, a direction we are actively pursuing.

In summary, This research develops an end-to-end pipeline for projection synthesis across arbitrary viewing angles and renderings of volumetric CT images. Based on our results, instance learning facilitates the adaptation to capture inter- and intra-subject dynamic change. Instance learning with paired query images demonstrates improvement in 2D rendering and notable

enhancement in localization accuracy during bifurcation landmark analysis. The overall performance of the model has room for further improvement to meet the demands of clinical therapy applications. Our current investigations are dedicated to refining the resolution and accuracy of the volumetric reconstructions, to optimize the structure's conspicuity for better identification.

References

[1] Coffey, M., & Vaandering, A. (2010). Patient setup for PET/CT acquisition in radiotherapy planning. Radiotherapy and Oncology, 96(3), 298-301.

[2] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), 99-106.

[3] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

[4] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

[5] Schwarz, K., Liao, Y., Niemeyer, M., & Geiger, A. (2020). Graf: Generative radiance fields for 3d-aware image synthesis. Advances in Neural Information Processing Systems, 33, 20154-20166.

[6] Jia, X., Dong, B., Lou, Y., & Jiang, S. B. (2011). GPU-based iterative cone-beam CT reconstruction using tight frame regularization. Physics in Medicine & Biology, 56(13), 3787.

[7] Bian, J., Wang, J., Han, X., Sidky, E. Y., Shao, L., & Pan, X. (2012). Optimization-based image reconstruction from sparse-view data in offset-detector CBCT. Physics in Medicine & Biology, 58(2), 205.

[8] Ye, J. C. (2019). Compressed sensing MRI: a review from signal processing perspective. BMC Biomedical Engineering, 1(1), 8.

[9] Liu, L. (2014). Model-based iterative reconstruction: a promising algorithm for today's computed tomography imaging. Journal of Medical imaging and Radiation sciences, 45(2), 131-136.

[10] Chen, G., Hong, X., Ding, Q., Zhang, Y., Chen, H., Fu, S., ... & Gao, H. (2020). AirNet: fused analytical and iterative reconstruction with deep neural network regularization for sparse-data CT. Medical physics, 47(7), 2916-2930.

[11] Monga, V., Li, Y., & Eldar, Y. C. (2021). Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. IEEE Signal Processing Magazine, 38(2), 18-44.

[12] Koetzier, L. R., Mastrodicasa, D., Szczykutowicz, T. P., van der Werf, N. R., Wang, A. S., Sandfort, V., ... & Willemink, M. J. (2023). Deep learning image reconstruction for CT: technical principles and clinical prospects. Radiology, 306(3), e221257.

[13] Pan, X., Zhan, X., Dai, B., Lin, D., Loy, C. C., & Luo, P. (2021). Exploiting deep generative prior for versatile image restoration and manipulation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), 7474-7489.

[14] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 586-595).

[15] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[16] Blau, Y., & Michaeli, T. (2018). The perception-distortion tradeoff. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6228-6237).

[17] Hugo, G. D., Weiss, E., Sleeman, W. C., Balik, S., Keall, P. J., Lu, J., & Williamson, J. F. (2016). Data from 4D lung imaging of NSCLC patients. The Cancer Imaging Archive, 10, K9.

[18] Tsai, E. B., Simpson, S., Lungren, M. P., Hershman, M., Roshkovan, L., Colak, E., ... & Wu, C. C. (2021). The RSNA international COVID-19 open radiology database (RICORD). Radiology, 299(1), E204-E213.

[19] Tsai, E. B., Simpson, S., Lungren, M. P., Hershman, M., Roshkovan, L., Colak, E., Erickson, B. J., Shih, G., Stein, A., Kalpathy-Cramer, J., Shen, J., Hafez, M. A. F., John, S., Rajiah, P., Pogatchnik, B. P., Mongan, J. T., Altinmakas, E., Ranschaert, E., Kitamura, F. C.,Topff, L., Moy, L., Kanne, J.P., Wu, C. (2021). Medical Imaging Data Resource Center (MIDRC) – RSNA International COVID-19 Open Radiology Database (RICORD) Release 1b – Chest CT Covid- (MIDRC-RICORD-1B) [Data set]. The Cancer Imaging Archive.

[20] Corona-Figueroa, A., Frawley, J., Bond-Taylor, S., Bethapudi, S., Shum, H. P., & Willcocks, C. G. (2022, July). Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 3843-3848). IEEE.

[21] Sidky, E. Y., & Pan, X. (2008). Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Physics in Medicine & Biology, 53(17), 4777.

[22] Chen, G., Hong, X., Ding, Q., Zhang, Y., Chen, H., Fu, S., ... & Gao, H. (2020). AirNet: fused analytical and iterative reconstruction with deep neural network regularization for sparse-data CT. Medical physics, 47(7), 2916-2930.

[23] Cheung, J. P., Shugard, E., Mistry, N., Pouliot, J., & Chen, J. (2019). Evaluating the impact of extended field-of-view CT reconstructions on CT values and dosimetric accuracy for radiation therapy. Medical physics, 46(2), 892-901.