

Sparse View CT Reconstruction Based on a Dual-domain Convolutional Neural Network

Xin Tie (xtie@wisc.edu, 9081372113)

Minyi Dai (mdai26@wisc.edu, 9079876687)

Hao Zhang (h Zhang978@wisc.edu, 9084624668)

Introduction

In medical imaging, reconstruction from acquired data typically presumes that the image object is stationary during the acquisition process. However, in dynamic imaging scenarios, this assumption may not hold true. Failing to account for the object motion leads to the emergence of motion artifacts, which can significantly affect radiologists' diagnosis. To deal with the dynamic acquisitions, two potential solutions have been proposed: upgrading the hardware system to shorten the scan time and developing image reconstruction algorithms to generate artifact-free images from an undersampled data set [1].

Computed tomography (CT) is a non-invasive diagnostic imaging tools, known for fast acquisitions, high spatial resolution, and broad availability. Sparse-view CT reconstruction in the dynamic acquisitions is an intriguing yet challenging problem, which involves acquiring only a portion of the full projection data. This low-cost and efficient technique has the benefits, including reducing radiation dose, reducing scan time and improving time-resolving capability in the Cardiac CT [2]. The main challenge in sparse-view reconstruction is angular undersampling that violates Nyquist's criterion, causing aliasing artifacts [3]. These artifacts manifest as streaks in the CT images and can obscure low-contrast objects of interest, such as lesions. As the number of acquired view projection data decreases, the resulting artifacts become more pronounced and severe. Figure 1 depicts the effect of reducing the sampled data to 1/4, 1/8 and 1/16 of the full data set on the image quality. Such artifacts greatly impact the clinical assessment. Therefore, sophisticated reconstruction algorithms are required to restore the anatomical details and maintain the diagnostic

performance. Recent advances in deep learning have shown remarkable progress in various medical imaging tasks, including disease classification, lesion detection, image denoising and artifact removal, etc. The powerful regression capability of the deep neural network enables us to tackle extremely challenging problems. In this study, we introduce a dual-domain convolutional neural network (CNN) to solve the sparse-view CT reconstruction problem.

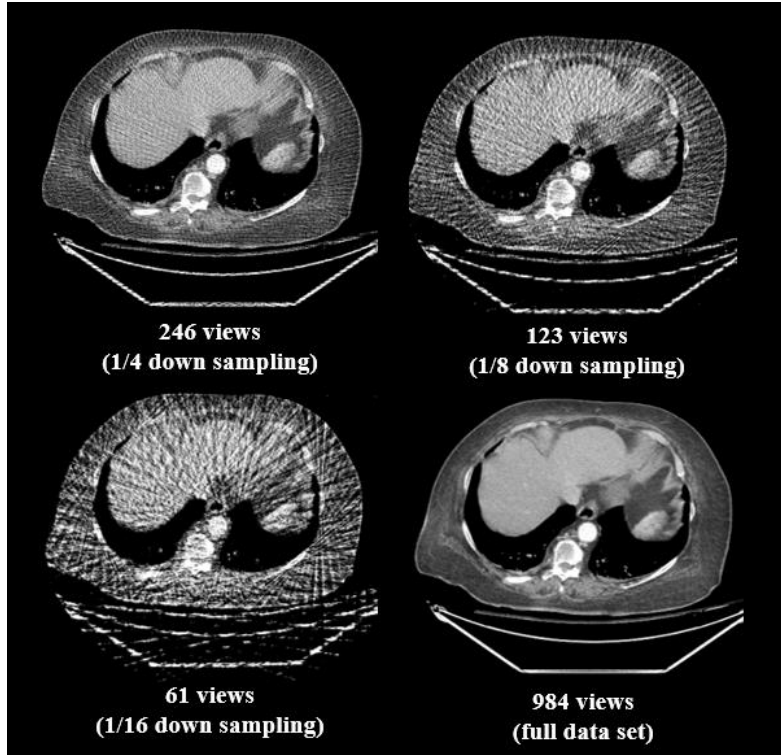


Figure 1: Impact of undersampling on CT image quality. Left: sparse-view CT with 123 views; Right: dense-view CT (or fully sampled CT) with 984 views.

Related works

Sparse-view CT reconstruction is an ill-posed inverse problem and there does not exist an analytical solution. In the past, two paradigms have been proposed to recover high-quality images from severely undersampled data. The first paradigm, compressed sensing (CS) [3-5], reformulates an image reconstruction problem to a convex optimization problem with two terms: the data fidelity term, which enforces consistency with the acquired data, and the regularization term, which transforms the image to a new space and promotes the sparsity in that space [3]. This optimization problem can be solved by a range of gradient-descent based methods, such as conjugate gradient and ADMM. In the end, the reconstructed image balances data fidelity and artifact removal [1]. The second paradigm, deep-learning-based methods [1-2,6-10], enables fast and higher-quality image reconstruction compared to CS-based methods. This class of methods can be categorized into three groups: (1) converting artifact-contaminated images to artifact-free images (2) inpainting undersampled projection data to generate a full data set, then applying filter backprojection (FBP), and (3) directly reconstructing images from undersampled projection data without explicit use of the classic reconstruction algorithms (e.g., FBP). In most cases, deep neural networks trained purely on the image domain performed reasonably well, like RED-CNN [11]. However, they did not fully leverage the information from the projection domain. Lack of consistency check with the acquired data may cause false negative lesions and false positive lesion-like structures [1] in the reconstructed images. Similarly, only using the projection data to train the model loses information of spatial correlations in the image domain.

In this work, we propose a dual-domain sparse view CT reconstruction pipeline that leverages both sinograms and images. Specifically, we first use a deep neural network to convert the artifact-contaminated CT image to an artifact-reduced image, followed by a Radon transform to convert the image to the sinogram. In the sinogram domain, we develop another model to correct the projection data to obtain a more accurate estimate of the full data set. Finally, we reconstruct the image using FBP. Our contributions are summarized as follows:

1. Introduction of a dual-domain CNN to estimate the full projection data.
2. Comparison with a purely image-domain model to highlight the importance of incorporating sinogram-domain information.
3. Model evaluation on acquisition scenarios different from the training conditions to test for generalizability.

Dataset

This study employs a retrospective dual-energy dataset [12], comprising 44 clinical pulmonary CT angiography (CTA) exams. The dataset encompasses both sinograms and images obtained at 140 kVp and 80 kVp. The former was used to train the model while the latter was used to test the generalizability in terms of the acquisition condition. It's worth mentioning that despite being acquired simultaneously with minimal registration errors, scans at two kV levels have significant differences in CT numbers, image contrast and beam hardening artifacts, etc. All data were acquired with a 64-slice MDCT scanner under the Gemstone Spectral Imaging (GSI) dual-energy CT (DECT) mode. The scans were performed in helical mode with the pitch of 0.984. All images were reconstructed by first applying the helical rebining and then performing 2D fan-beam FBP. The in-plane pixel size is 0.98mm×0.98mm and the slice thickness is 2.5mm. The standard ramp kernel was used to filter the sinogram before backprojection. To simulate the sparse-view CT data, one-eighth of the total view angles (984 views) were sampled, resulting in a total of 123 views. The examples of sparse-view CT images with standard FBP reconstruction are shown in

Figure 1. Among 44 human subjects, 36 (4531 slices) were randomly selected for model development and the remaining 8 cases (938 slices) were used for testing.

Method

The workflow of the dual-domain CNN is presented in Figure 2. The image-domain network, denoted as N1, accepts the artifact contaminated images reconstructed from undersampled view angles as inputs. The purpose of this network is to mitigate the streak artifacts in the sparse-view CT images. In this work, we used the Attention-UNet [13] which applied the attention gates to the higher semantic level representations and reweighted the features maps. Models with attention gates have been shown to suppress irrelevant regions and focus more on salient features, resulting in more efficient training and improved performance [13-14]. The output of N1 is fed into a forward projection (FP) module that maps the artifact-reduced images to the sinogram \tilde{Y}_{N1} . Note that this FP module is differentiable, and its gradient is the direct backprojection (BP). The implementation of FP and BP was based on the torch-radon. In the sparse-view acquisition, 123 view angles were obtained and should be utilized as real data for full projection data estimation. This is accomplished by masking the estimated sinogram at the acquired view angles and replacing them with the measured data:

$$\tilde{Y} = \tilde{Y}_{N1} \odot (1 - M) + Y_{sparse} \odot M$$

Where \tilde{Y}_{N1} denotes the estimated sinogram from N1, Y_{sparse} denotes the undersampled sinogram with zeros at the unacquired view angles, M is the binary mask with ones at the acquired view angles and zeros at the unacquired angles. \tilde{Y} is then fed to a separate Attention-UNet (N2) to estimate the full projection data. At both domains, a combination of least squared loss and gradient loss was used to backpropagate the errors with respect to the ground true images and sinograms. As the scales of these two losses were found to be similar, we assigned equal weights to each of them. The final loss function is written as

$$L_{tot} = L_2(X_{GT}, X_{Est}) + L_{grad}(X_{GT}, X_{Est}) + \lambda [L_2(Y_{GT}, Y_{Est}) + L_{grad}(Y_{GT}, Y_{Est})]$$

Where L_2 denotes the least squared loss, L_{grad} denotes the gradient loss, X_{GT} denotes the reference CT images reconstructed from complete projection data, X_{Est} denotes the output images of N1, Y_{GT} denotes the fully sampled sinogram and Y_{Est} denotes the estimated sinogram from N2. λ is a hyperparameter that controls the contributions of the image-domain loss and the sinogram-domain loss. In this work, λ was set to 1e-5 to balance the differences in magnitude between images and sinograms. The proposed dual-domain approach allows for joint optimization of both image- and sinogram-domain networks, leading to better information integration and utilization.

During the training, the mini-batch size was set as 8, and 200 epochs were used. The ADAM optimizer with an initial learning rate of 1e-4 was empirically selected. If there was no decrease in the validation loss for 10 consecutive epochs, the learning rate was reduced by a factor of 0.2. This process was repeated until the learning rate reached 1e-6. Training data augmentation was performed by applying the horizontal/vertical flip and rotation operations to the paired images.

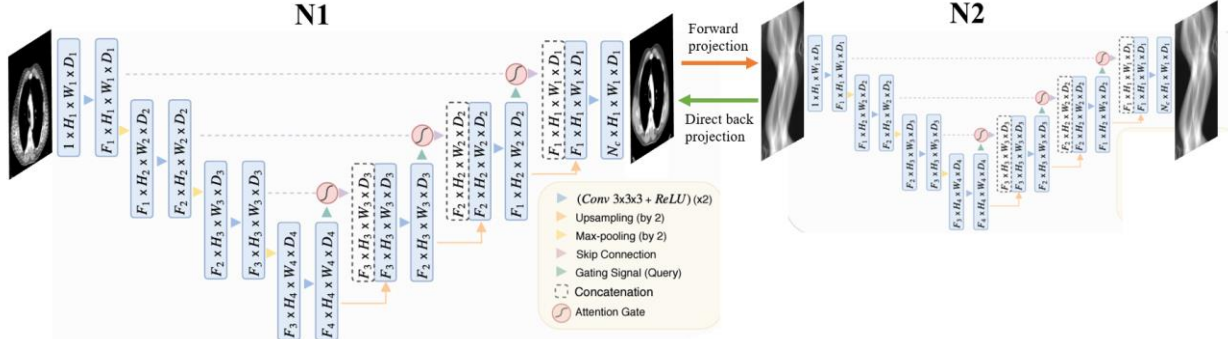


Figure 2: The workflow of the dual-domain CNN

Evaluation

Three evaluation metrics were used to quantify the performance of our models. Two of them reflect the accuracy of CT numbers compared to fully sampled images: relative root mean squared error (rRMSE) and mean absolute error (MAE). The third metric measures the similarity between the images in terms of perceived quality: structure similarity index metric (SSIM). The SSIM metric takes into account the structural information present in the images, such as pixel correlations, brightness, and image contrast. To highlight the importance of integrating information from sinogram data, we compared the results achieved by the dual-domain CNN with those from the purely image-domain model.

Generalizability testing

Our current model was trained on the 140 kV dataset without exposure to any CT data acquired at other tube potentials. It is crucial to test the model generalizability since the conventional FBP and compressed sensing approaches could achieve consistent results across various acquisition scenarios. In addition to the tube potential, we also tested the model capability in handling even fewer view angles (61 views). Since the model was not trained on a 61-view dataset and fewer view angles made the sparse-view problem even more challenging, it is reasonable to expect that our model may not attain satisfactory performance in this scenario.

Current Progress

Figure 3 shows the reconstruction results of 4 cases obtained from the dual-domain model (2nd column) and the image-domain model (3rd column). The resulting images demonstrate that the streak artifacts present in the original sparse-view images have been effectively eliminated, and the anatomical details are well-preserved. However, the image-domain model tends to overly smooth the images while the results of the dual-domain model closely resemble the fully-sampled FBP images. The first 2 cases in Figure 3 demonstrate the superior preservation of tiny and high-contrast objects by the dual-domain model. And the last case shows that the dual-domain model eliminates the spurious structure that appears as a dark cluster in the image generated by the image-domain model. Table 1 summarizes the quantitative results obtained by each model on 8 testing cases. The dual-domain model achieved consistently better results across all evaluation metrics.

| Testing data | Dual-domain model | | | Image-domain model | | |
|--------------|----------------------|----------------------|-------------------------|----------------------|----------------------|-------------------------|
| | MAE (HU) | rRMSE (%) | SSIM | MAE (HU) | rRMSE (%) | SSIM |
| Case 1 | 21.8 [18.9, 25.1] | 1.69 [1.45, 1.87] | 0.890 [0.867, 0.909] | 26.0 [23.8, 27.3] | 1.86 [1.76, 2.04] | 0.860 [0.842, 0.869] |
| Case 2 | 22.6 [20.1, 26.9] | 1.83 [1.65, 2.02] | 0.874 [0.854, 0.895] | 28.8 [24.6, 30.8] | 2.06 [1.94, 2.31] | 0.824 [0.811, 0.841] |
| Case 3 | 24.9 [22.4, 28.9] | 1.74 [1.56, 1.97] | 0.857 [0.826, 0.885] | 29.2 [27.1, 32.9] | 2.01 [1.85, 2.18] | 0.812 [0.799, 0.829] |
| Case 4 | 18.0 [16.1, 20.3] | 1.44 [1.25, 1.84] | 0.907 [0.894, 0.926] | 21.1 [19.7, 22.0] | 1.47 [1.42, 2.16] | 0.881 [0.874, 0.886] |
| Case 5 | 21.4 [18.8, 23.7] | 1.59 [1.40, 1.77] | 0.901 [0.882, 0.920] | 27.0 [25.5, 28.2] | 1.86 [1.78, 1.91] | 0.843 [0.835, 0.850] |
| Case 6 | 16.9 [14.9, 18.3] | 1.34 [1.16, 1.62] | 0.928 [0.910, 0.941] | 20.7 [18.4, 21.3] | 1.66 [1.39, 1.93] | 0.898 [0.893, 0.901] |
| Case 7 | 20.5 [18.1, 24.0] | 1.65 [1.49, 1.86] | 0.883 [0.863, 0.905] | 25.2 [22.7, 28.1] | 1.95 [1.84, 2.16] | 0.824 [0.817, 0.835] |
| Case 8 | 17.3 [15.9, 19.0] | 1.45 [1.26, 1.74] | 0.916 [0.902, 0.929] | 20.9 [20.3, 21.5] | 1.62 [1.43, 2.01] | 0.888 [0.885, 0.892] |
| Total | 20.4 [17.5, 23.8] | 1.62 [1.39, 1.86] | 0.897 [0.869, 0.917] | 24.5 [21.0, 28.0] | 1.89 [1.69, 2.06] | 0.848 [0.824, 0.883] |

Table 1: the quantitative results achieved by the dual-domain model and the purely image-domain model. Results are presented in the form of median [25th quantile, 75th quantile].

Revisions and Timeline

Initially, we planned to implement the diffusion model for sinogram inpainting. However, several challenges held us back: (1) The training and inference were slow. (2) The generated images looked like CT, but the anatomical structures got distorted and the image quality was not as good. Incorporating the idea of diffusion model to sparse view CT reconstruction is fascinating but we felt that we might be too ambitious. Therefore, instead of implementing the diffusion model, we decided to use the convolutional neural networks to reduce the streak artifacts and noise caused by the sparse-view acquisitions. And in the sinogram domain, we used a similar but shallower network architecture to incorporate the information from acquired view angles. Additionally, we conducted the generalizability testing for the data acquired under different scenarios, including lower tube potential and fewer view angles.

| | |
|-------|--|
| 03/15 | Finish the training of the image-domain CNN model |
| 03/31 | Finish the training of the dual-domain CNN model |
| 04/02 | Perform preliminary evaluation of the trained models |
| 04/04 | Submit the midterm report |
| 04/15 | Finish the model evaluation and comparison |
| 04/20 | Finish the slides for project presentation |
| 05/05 | Finish the project website |

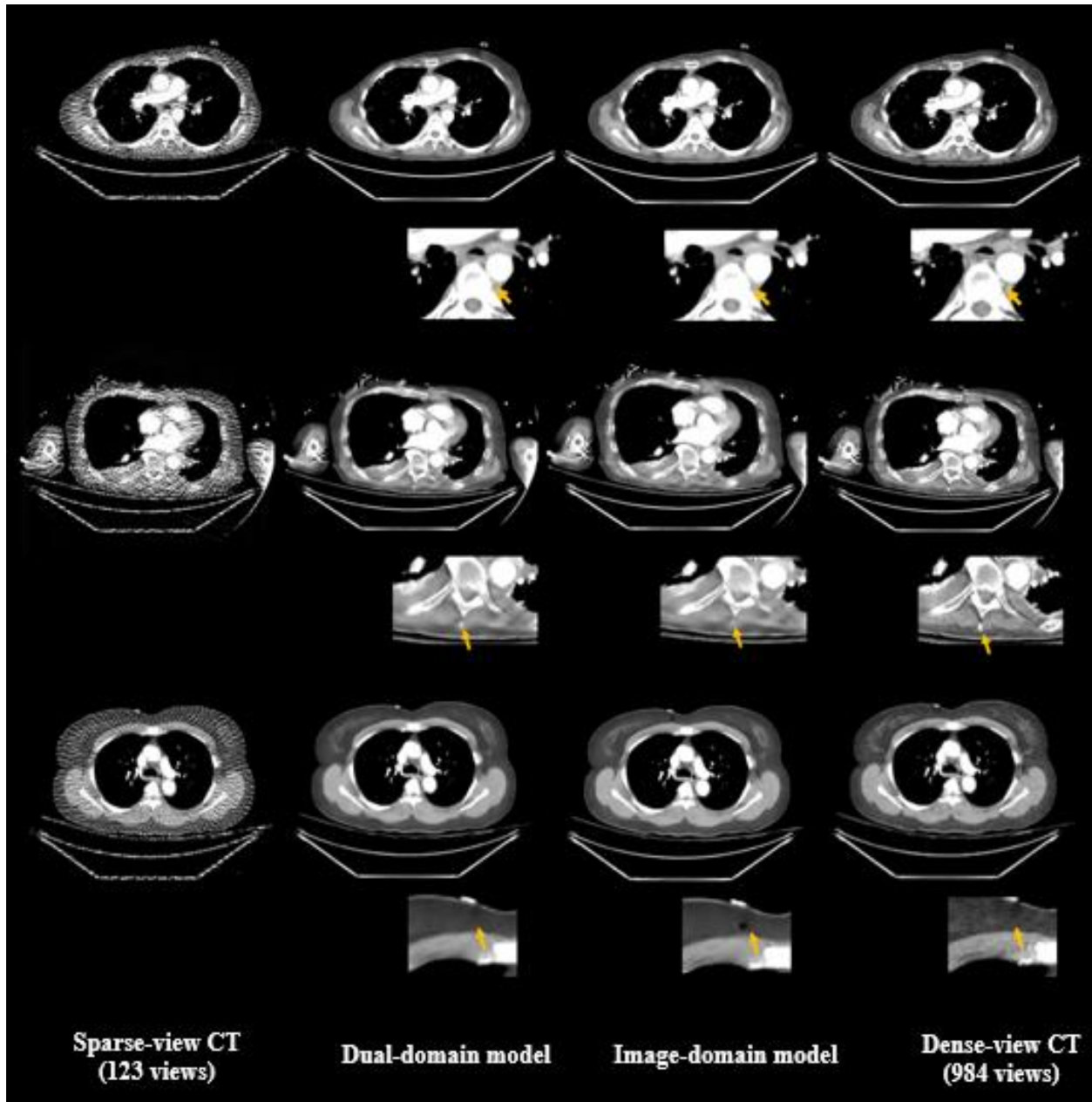


Figure 3: Visual assessment of results obtained by the dual-domain model and the image-domain model.

Difficulties

Implementing the diffusion model took us a while but the progress was not made until we switched to the convolutional neural network. CNN has been employed extensively in the medical imaging reconstruction. Implementing a purely image-domain model is straightforward but integrating information from both images and sinograms is more challenging because it requires a careful design of the domain transformation module otherwise the errors cannot be backpropagated to update the model parameters.

References:

- [1] C. Zhang, Y. Li, and G. Chen, “Accurate and robust sparse-view angle CT image reconstruction using deep learning and prior image constrained compressed sensing (DL-PICCS),” *Medical Physics*, vol. 48, no. 10, pp. 5765–5781, Oct. 2021, doi: [10.1002/mp.15183](https://doi.org/10.1002/mp.15183).
- [2] H. Kudo, T. Suzuki, and E. A. Rashed, “Image reconstruction for sparse-view CT and interior CT— introduction to compressed sensing and differentiated backprojection,” *Quantitative Imaging in Medicine and Surgery*, vol. 3, no. 3, 2013.
- [3] W. Xia, W. Cong, and G. Wang, “Patch-Based Denoising Diffusion Probabilistic Model for Sparse-View CT Reconstruction.” arXiv, Nov. 18, 2022. Accessed: Feb. 17, 2023. [Online]. Available: <http://arxiv.org/abs/2211.10388>
- [4] J. Bai, Y. Liu, and H. Yang, “Sparse-View CT Reconstruction Based on a Hybrid Domain Model with Multi-Level Wavelet Transform,” *Sensors*, vol. 22, no. 9, p. 3228, Apr. 2022, doi: [10.3390/s22093228](https://doi.org/10.3390/s22093228).
- [5] Z. Qu, X. Yan, J. Pan, and P. Chen, “Sparse View CT Image Reconstruction Based on Total Variation and Wavelet Frame Regularization,” *IEEE Access*, vol. 8, pp. 57400–57413, 2020, doi: [10.1109/ACCESS.2020.2982229](https://doi.org/10.1109/ACCESS.2020.2982229).
- [6] Z. Fu, H. W. Tseng, S. Vedantham, A. Karellas, and A. Bilgin, “A residual dense network assisted sparse view reconstruction for breast computed tomography,” *Sci Rep*, vol. 10, no. 1, p. 21111, Dec. 2020, doi: [10.1038/s41598-020-77923-0](https://doi.org/10.1038/s41598-020-77923-0).
- [7] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, “A Sparse-View CT Reconstruction Method Based on Combination of DenseNet and Deconvolution,” *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1407–1417, Jun. 2018, doi: [10.1109/TMI.2018.2823338](https://doi.org/10.1109/TMI.2018.2823338).
- [8] W. Xia, Z. Yang, Q. Zhou, Z. Lu, Z. Wang, and Y. Zhang, “A Transformer-Based Iterative Reconstruction Model for Sparse-View CT Reconstruction,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13436, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 790–800. doi: [10.1007/978-3-031-16446-0_75](https://doi.org/10.1007/978-3-031-16446-0_75).
- [9] B. Zhou, X. Chen, S. K. Zhou, J. S. Duncan, and C. Liu, “DuDoDR-Net: Dual-domain data consistent recurrent network for simultaneous sparse view and metal artifact reduction in computed tomography,” *Medical Image Analysis*, vol. 75, p. 102289, Jan. 2022, doi: [10.1016/j.media.2021.102289](https://doi.org/10.1016/j.media.2021.102289).
- [10] Y. Han and J. C. Ye, “Framing U-Net via Deep Convolutional Framelets: Application to Sparse-view CT.” arXiv, Mar. 28, 2018. Accessed: Feb. 17, 2023. [Online]. Available: <http://arxiv.org/abs/1708.08333>
- [11] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, and P. Liao, “Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network (RED-CNN)”.
- [12] Li, Y, Tie, X, Li, K, et al. A quality-checked and physics-constrained deep learning method to estimate material basis images from single-kV contrast-enhanced chest CT scans. *Med Phys*. 2023; 1- 21. <https://doi.org/10.1002/mp.16352>
- [13] Schlemper J, Oktay O, Schaap M, et al.. Attention gated networks: learning to leverage salient regions in medical images. ArXiv1808.08114 [preprint] <https://arxiv.org/abs/1808.08114>. Posted August 22, 2018. Accessed February 2019.
- [14] Kearney V, Ziemer BP, Perry A, Wang T, Chan JW, Ma L, Morin O, Yom SS, Solberg TD. Attention-Aware Discrimination for MR-to-CT Image Translation Using Cycle-Consistent Generative Adversarial Networks. *Radiol Artif Intell*. 2020 Mar 25;2(2):e190027. doi: 10.1148/ryai.2020190027. PMID: 33937817; PMCID: PMC8017410.