

# Proj2Proj: self-supervised low-dose CT reconstruction

Mehmet Ozan Unal<sup>1</sup>, Metin Ertas<sup>2</sup> and Isa Yildirim<sup>1</sup>

<sup>1</sup> Department of Electronics and Communication Engineering, Istanbul Technical University, Istanbul, Turkey

<sup>2</sup> Department of Electrical and Electronics Engineering, Istanbul University-Cerrahpasa, Istanbul, Turkey

## ABSTRACT

In Computed Tomography (CT) imaging, one of the most serious concerns has always been ionizing radiation. Several approaches have been proposed to reduce the dose level without compromising the image quality. With the emergence of deep learning, thanks to the increasing availability of computational power and huge datasets, data-driven methods have recently received a lot of attention. Deep learning based methods have also been applied in various ways to address the low-dose CT reconstruction problem. However, the success of these methods largely depends on the availability of labeled data. On the other hand, recent studies showed that training can be done successfully without the need for labeled datasets. In this study, a training scheme was defined to use low-dose projections as their own training targets. The self-supervision principle was applied in the projection domain. The parameters of a denoiser neural network were optimized through self-supervised training. It was shown that our method outperformed both traditional and compressed sensing-based iterative methods, and deep learning based unsupervised methods, in the reconstruction of analytic CT phantoms and human CT images in low-dose CT imaging. Our method's reconstruction quality is also comparable to a well-known supervised method.

**Subjects** Bioinformatics, Artificial Intelligence, Computer Vision, Neural Networks

**Keywords** Deep learning, Low-dose CT, Image reconstruction, Self-supervised learning

Submitted 13 July 2023

Accepted 10 January 2024

Published 29 February 2024

Corresponding author

Mehmet Ozan Unal,  
unalmehmet@itu.edu.tr

Academic editor

Jyotismita Chaki

Additional Information and  
Declarations can be found on  
page 13

DOI 10.7717/peerj-cs.1849

© Copyright  
2024 Unal et al.

Distributed under  
Creative Commons CC-BY 4.0

**OPEN ACCESS**

## INTRODUCTION

Computed Tomography (CT) imaging is one of the most common tools used in the diagnosis of diseases. The inside of the human body can be monitored and problematic tissues, deformities, and lesions can be detected *via* CT. Despite these benefits, CT imaging has one crucial downside: ionizing radiation. The radiation dose reduction in CT imaging is possible in several ways such as: *i*) decreasing the number of projections, *ii*) reducing of X-ray tube current, *iii*) narrowing the viewing angle.

During CT imaging, the source emits X-ray, and the detector data at different angles are collected. These gathered data are known as CT projections, and the image created from them is known as a CT image. The reconstruction in CT is the creation of images from these noisy indirect projections. Filtered back projection (FBP), the most traditional method used in CT imaging, works by projecting CT projections back to the image domain. With the help of a filter, since low frequencies are sampled more than high frequencies due to the system geometry, it reduces the effect of low frequencies and

increases the weight of high frequencies. However, this operation increases the sensitivity of the FBP method to noise. Therefore, when the projections are noisy or incomplete, FBP method does not produce satisfactory results.

To overcome this limitation of the FBP method, iterative methods were developed. Simultaneous algebraic reconstruction technique (SART), one of the most popular of these methods, simultaneously back-projects the error for all projections ([Andersen & Kak, 1984](#)). Recently compressed sensing based iterative reconstruction methods have been proposed. Compressed sensing states that: if the information is sparse on a known basis, this information can most likely be recovered by incomplete measurements ([Donoho, 2006](#); [Candès, Romberg & Tao, 2006](#)). Given that the information to be recovered is not sparse in its original space, sparsifying transforms can be used. Among them, total variation (TV) has been extensively used. TV minimization method aims to minimize the gradient magnitude of the images, which constrains the solution set to piece-wise smooth images ([Rudin, Osher & Fatemi, 1992](#)). Iterative and compressed-sensing-based methods were merged and proposed for low-dose CT problem ([Yu & Wang, 2009](#); [Sidky & Pan, 2008](#)). Image domain denoisers were also used to address this problem. Non-Local means ([Buades, Coll & Morel, 2005](#)) and Block Matching 3D (BM3D) ([Dabov et al., 2007](#)) exploit the non-local similarities of the images and use these similarities for denoising. These methods are quite hyperparameter dependent and use simplistic hand-crafted priors.

Nowadays, deep learning based studies have become attractive with the increase in computing capacity and the availability of large datasets. Numerous deep learning based methods have been proposed for image domain inverse problems such as denoising, deconvolution, and inpainting. Deep learning based methods have also been applied to the low-dose CT problem. Convolutional neural networks (CNN) based methods were used as denoisers in the image domain ([Jin et al., 2017](#); [Chen et al., 2017](#); [Buzug, 2008](#); [Yang et al., 2018](#); [Liu et al., 2020](#)). Deep learning methods were also extended by applying iteratively with classical reconstruction techniques ([Adler & Öktem, 2018](#); [Wu et al., 2017](#); [He et al., 2019](#)).

Advancements in the deep learning field made it possible to define sophisticated learning methods and learned constraints to further optimize low-dose CT reconstructions. [Wu et al. \(2021\)](#) applied a residual-based network method by optimizing the network using both measurement consistency and image quality awareness. Spectral2Spectral ([Guo et al., 2023](#)) study used the similarity prior within the image-spectral domain as a regularization term to constrain the network training. Deep learning methods were also used in metal artifact reduction by applying the loss in the dual domain ([Zhou et al., 2022](#)). Another method to improve the image quality of low-dose CT images is to denoise the projections. [Yang et al. \(2023\)](#) applied this method by using transformer architecture with self-attention. [Wu et al. \(2023\)](#) proposed a wavelet-improved denoising for low-dose CT problems that exploit the score-based generative model approach.

For most of the deep learning based methods given, one of the most important factors of the success is the availability of large datasets. Another field of focus aims to tackle this problem by enabling denoising without the need for noisy-denoised pairs datasets.

In order to address this issue, deep image prior (DIP) study suggested to use the architecture of CNNs as a regularizer. [Baguer, Leuschner & Schmidt \(2020\)](#) combined this approach with TV regularizer and applied this unsupervised deep CNN based method to low-dose CT problem. Noise2Noise proposes a method for the denoising problem that does not require noise-free target images ([Lehtinen et al., 2018](#)). However, it still requires two independent noisy measurements of the same information. Noise2Self ([Batson & Royer, 2019](#)) and Noise2Void ([Krull, Buchholz & Jug, 2019](#)) studies, on the other hand, suggested that denoising can be performed using only the noisy measurement itself. In other words, the image itself is used as the target image during the model training. One of the problems that may arise at this point is that the model can converge to an identity function. Noise2Self study suggested a kind of perturbation mechanism called  $J_{th}$  invariant principle. Noise2Void study approached the denoising problem on a pixel scale and defined a receptive field so that the pixel was not used during the estimation of that pixel. Subsequently, studies ([Quan et al., 2020](#); [Xie, Wang & Ji, 2020](#)) were also proposed to prevent this convergence in different ways.

The opportunity of training on noisy target images could be quite valuable for low-dose CT reconstruction. Since it is not always easy to obtain low-dose/normal-dose pairs to create big datasets, various learning methods without noise-free targets have been proposed. Noise2Inverse ([Hendriksen, Pelt & Batenburg, 2020](#)) study proposed a learning method that exploits Noise2Noise ([Lehtinen et al., 2018](#)) principle by grouping projections and using them as targets against each other. To learn a reconstruction method from a single image, Noise2Filter ([Lagerwerf et al., 2020](#)) study combined Noise2Inverse and NN-FBP ([Pelt & Batenburg, 2013](#)) methods.

In our study, we propose a method so-called Proj2Proj which customizes Noise2Self method for the low-dose CT reconstruction problem by applying the self-supervision principle in the projection domain. We used the self-supervision to train a neural network that denoises the reconstruction. We were able to train this neural network with only low-dose projections without the need for low-dose/normal-dose pairs whose availability might be a big concern in this field. Our method outperformed traditional and compressed sensing-based iterative methods and an unsupervised deep CNN based method in the reconstruction of analytic phantoms and human CT images both qualitatively and quantitatively. Furthermore, it produced such results that were comparable with a well-known supervised approach, FBP+U-Net ([Jin et al., 2017](#)). In addition to peak signal-to-noise (PSNR) and Structural Similarity (SSIM), the Learned Perceptual Image Patch Similarity (LPIPS) ([Zhang et al., 2018](#)) metric was also used for the qualitative comparisons, which focuses on the perceptual similarity of the results. The preliminary results of this study were published in a leading conference in biomedical imaging. The present article extends ([Unal, Ertas & Yildirim, 2021](#)) by detailing the mathematical definition of the method and adding theoretical justification of the method, by training the network with two different datasets, by extending the experiments with self-supervised and supervised deep learning based methods and new metrics.

The article is organized as follows. The rationale and implementation of the self-supervised training method are described in “Method”. The experiment datasets

and settings, as well as the findings, are presented in “Experiments”. “Discussion” discusses further details of the method. Finally, the last section brings the article to a conclusion.

## METHOD

Normal-dose CT image acquisition process can be defined as follows:

$$y_N = A_N x^* + \eta, \quad x_N = \text{FBP}(y_N) \quad (1)$$

where  $x^*$  is the ground-truth X-ray attenuation of the body,  $A_N$  is the normal-dose data acquisition matrix,  $\eta$  is the data acquisition noise,  $y_N$  is the normal-dose CT projections. Conventionally, the normal dose CT image is reconstructed with the FBP method.

The low-dose reconstruction process can be represented as:

$$y_L = A_L x^* + \eta, \quad x_L = \text{FBP}(y_L) \quad (2)$$

where  $x^*$  is the ground-truth image,  $A_L$  is the low-dose data acquisition matrix,  $y_L$  is the low-dose CT projections. Similarly, the low-dose CT image can be reconstructed with FBP method and represented here as  $x_L$ . However,  $x_L$  does not usually have sufficient image quality when it is compared with the ground truth image  $x^*$ .

In the presence of normal-dose CT projections, a supervised reconstruction model can be shown with the following equation:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_x [ \|f_{\theta}(x_L) - x_N\|_2^2 ] \quad (3)$$

where  $f_{\theta}$  is the deep neural network (DNN) which is parameterized with  $\theta$ ,  $\mathbb{E}_x$  is the expectation over different  $x$  images,  $x_L$  is the reconstructed low-dose image defined in (2), and  $x_N$  is the reconstructed normal-dose image defined in (1). The supervised model given in (3) can be used in projection domain as follows:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_y [ \|A_N f_{\theta}(x_L) - y_N\|_2^2 ] \quad (4)$$

where  $\mathbb{E}_y$  is the expectation over different  $y$  projections. For the supervised solution to the low-dose CT problem, both low-dose and normal-dose pairs of the same information are required. DNN which is defined with  $f_{\theta}$  learns how to map a low-dose image to a normal-dose one. However, it is not easy to collect such data due to the sensitivity of human tissues to ionizing radiation. The proposed method requires only the low-dose projections  $y_L$  without the need of  $y_N$  or  $x_N$ . This idea was formulated by defining a self-supervised loss that uses only low-dose projections and images,  $y_L$  and  $x_L$ .

$$\mathbb{E}_{x,y} [ \|A_L f_{\theta}(x_L) - y_L\|_2^2 ] \quad (5)$$

where  $\mathbb{E}_{x,y}$  is an expectation over  $x$  and  $y$  random variables and the rest of the variables are defined in (1) and (2). Our proposed self-supervised loss can be proven to converge the supervised solution. Let's replace  $y_L$  in (5) with  $A_L x^* + \eta$  using (2):

$$\mathbb{E}_{x,n} [ \|A_L f_{\theta}(x_L) - A_L x^* - \eta\|_2^2 ]. \quad (6)$$

If (6) is rewritten by expanding the  $l_2$ -norm square:

$$\mathbb{E}_x[\|(A_L f_\theta(x_L) - A_L x^*)\|_2^2] + \mathbb{E}_\eta[\|\eta\|_2^2] - \mathbb{E}_{x,\eta}[2\langle A_L f_\theta(x_L) - A_L x^*, \eta \rangle] \quad (7)$$

The second term is the expectation over  $\eta$  of  $\|\eta\|_2^2$  which is the variance of the noise and not taken into account during the minimization since its gradient equals to zero. The third term of the equation is expectation over  $x, \eta$  of  $-\mathbb{E}_{x,\eta}[2\langle A_L(f_\theta(x_L) - x^*), \eta \rangle]$  whose weight becomes minimal in (7) since  $f_\theta(x_L) - x^*$  ideally converges to zero. As a result, the first term,  $\mathbb{E}_x[\|(A_L f_\theta(x_L) - A_L x^*)\|_2^2]$ , becomes the most effective term in this optimization problem which is actually correlated with the supervised loss defined in (3) and uses  $x^*$  random variable as the target.

In the previous paragraph, it is shown why self-supervised loss can be an effective candidate to learn a mapping from low-dose to normal-dose. However,  $f_\theta$  might converge to an identity function in (5). To overcome this issue,  $J_{th}$  invariant principle was applied (Batson & Royer, 2019) which proposes that  $N$  subset of the pixels are selected from noisy target images and the selected pixels are perturbed using the neighbor pixels, and the loss is defined based on only these pixels as in (8). In this way, the convergence of  $f_\theta$  to an identity function is prevented. To apply this principle to our Proj2Proj loss function, (5) is rewritten as follows:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_y[\|J(A_L f_\theta(\text{FBP}(y_{Lc})) - y_L)\|_2^2] \quad (8)$$

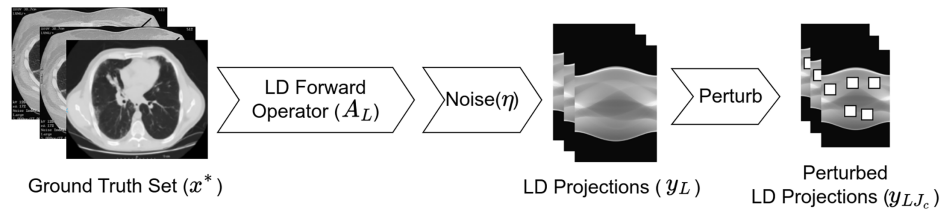
where  $J$  is the mask which filters only the  $J_{th}$  subset of the pixels,  $y_{Lc}$  is the low-dose projections whose  $J_{th}$  subset of pixels are perturbed. Finally, the proposed loss function can learn from only low-dose projections without the need of normal-dose images or projections.

The complete working schema of the proposed Proj2Proj reconstruction method is also given in Fig. 1. The process can be examined in three parts as preprocessing, training, and reconstruction and evaluation. In preprocessing, low-dose projections are created from the CT images *via* low-dose CT forward operator. The obtained projection dataset is split into training, validation, and test sets. For the training set projections, the  $J_{th}$  subset of the pixels is perturbed and given as the input to FBP reconstruction algorithm. The perturbations are done *via* the following steps:

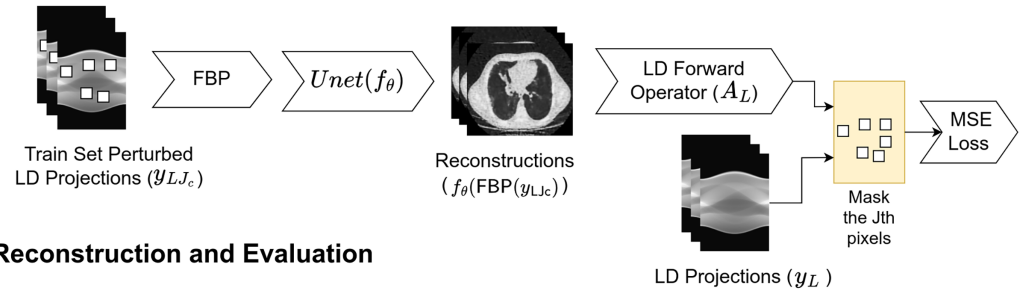
- The input image is divided into square pieces with the help of a grid. In our setup, it is selected as  $4 \times 4$ .
- The pixel ( $i_{th}$ ) to be perturbed is selected by taking the modulo of the number of iterations over the number of pixels in the grid (in our case  $4 \times 4 = 16$ ). For example, if the iteration number is 17,  $17 \bmod (16)$  equals to 1.
- $i_{th}$  pixel of all the grids are perturbed with the average of all four neighbor pixels excluding the  $i_{th}$  pixel itself.

The output of FBP is denoised with the neural network  $f_\theta$  and forward projected to projection space *via* low-dose CT forward transform. The error is calculated between the

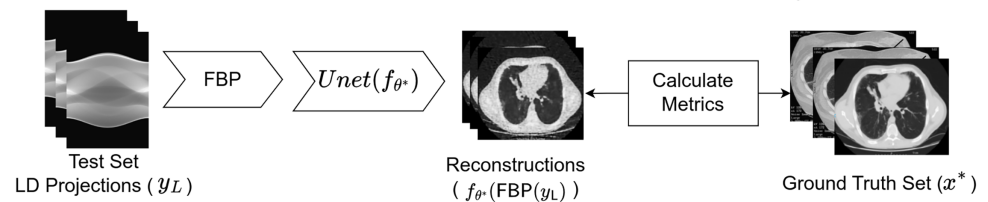
### Preprocessing



### Training



### Reconstruction and Evaluation



**Figure 1** Proposed working schema for self-supervised low-dose CT reconstruction. Image source credit: Yang et al. (2018). Full-size DOI: 10.7717/peerj-cs.1849/fig-1

output and the perturbed projections after the same  $J_{th}$  subset mask is applied to the error. The neural network is optimized by minimizing the loss function in (8) with the Adam (Kingma & Ba, 2015) optimizer. After the network weights ( $\theta$ ) are optimized by the training, they are used for the reconstruction. In the reconstruction phase, the input projections are given to the FBP operator in raw form without perturbation. The output of the FBP operator is denoised with the learned neural network  $f_{\theta^*}$  whose output is the reconstructed image by the proposed approach.

## EXPERIMENTS

The details of experiments and code are given at the code repository (<https://github.com/mozanunal/SparseCT>).

### Experimental setup

Deep lesion dataset (Yan et al., 2017) as human CT data and Shepp-Logan phantom and ellipses dataset as synthetic data were used in the experiments. The image resolution was selected as  $512 \times 512$ . The ellipses dataset consists of 36,400 artificially generated images and it was split into training (32,000 slices), validation (3,200 slices), and test (3,200 slices) datasets. DeepLesion dataset consists of 32,120 CT slices from 10,594 studies of 4,427 unique patients. It was split into three datasets as training (60%), validation (20%), and test (20%). For simulated low-dose CT setup, 64-view parallel beam projections were generated

using Radon transform, and detector resolution was selected as  $512 \times 1$  pixels. The projections were contaminated with the noise level between 30–40 dB.

### Comparison methods

FBP, SART (*Andersen & Kak, 1984*), SART+TV (*Yu & Wang, 2009*), SART+BM3D (*Dabov et al., 2007*), DIP+TV (*Baguer, Leuschner & Schmidt, 2020*) and FBP+U-Net (*Jin et al., 2017*) methods were compared with our self-supervised reconstruction method. Inverse Radon transform was used with a ramp filter for FBP. Iteration number = 40 and relaxation parameter = 0.15 were chosen for SART hyperparameters. The TV weight was set to 0.9 and sigma parameter of BM3D was chosen as 0.35. All these hyperparameters were chosen to generate the highest PSNR values in the validation sets. In addition to these methods, another BM3D (sigma = 0.20) reconstructor was used in the comparisons which gives the highest SSIM (*Wang, Simoncelli & Bovik, 2003*) for the validation set. For the implementation of the FBP+U-Net, the same network architecture of the proposed method was used and the number of parameters and iteration number were selected as the same for a fair comparison. For the realization of DIP+TV method, the reference study was used (*Baguer, Leuschner & Schmidt, 2020*) and Skipnet (*Wang et al., 2018*) was used as the network architecture. The network overfitted to a single image for 4,000 iterations with a learning rate of 0.02. The hyperparameters were obtained from the validation dataset.

### Network architecture and training

U-Net, a commonly used architecture in denoising applications, was selected for the denoiser neural network. Since U-Net is an auto-encoder, it is successful at extracting features of the images and then decoding them from these latent representations (*Chen et al., 2017*). Besides, thanks to the skip connections between the scales of the encoder and the decoder, the gradient flow is facilitated and the higher-order structure can be transferred to the output more easily. In our model, a five-scale structure was preferred with the convolutional filter numbers being 64, 128, 256, 512, 1,024 from the first layer to the last layer.

The number of parameters and training iterations were chosen as 2,160,000 and 200,000, respectively. In total, the self-supervised method was trained for 28 hours on a computer with an RTX2080 TI graphics card. The learning rate was chosen as 0.0001 and ADAM (*Kingma & Ba, 2015*) optimizer was used.

The proposed self-supervised method was trained with two different schemes: *i*) Proj2Proj method trained on ellipses dataset, *ii*) Proj2Proj method trained on human CT dataset. Each was used in the comparisons of the corresponding dataset.

## RESULTS

In order to validate the proposed method, the experiments were done on Shepp-Logan, ellipses dataset and human CT dataset with different noise levels. PSNR, SSIM, and LPIPS metrics were used for quantitative assessment. In addition to the numerical analysis, 1-D profile analysis and visual examination for qualitative assessment were used.

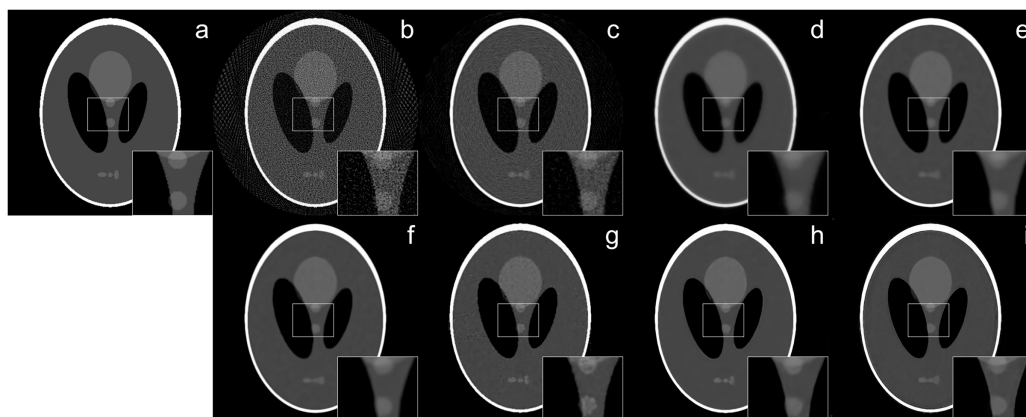
**Table 1** The ellipses dataset results for 64 projections with the noise level 30 to 40 dB SNR.

	30 dB SNR			33 dB SNR			37 dB SNR			40 dB SNR		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
FBP	12.14 ± 1.66	0.25 ± 0.01	0.23 ± 0.03	14.97 ± 1.62	0.29 ± 0.02	0.28 ± 0.02	18.52 ± 1.54	0.38 ± 0.03	0.37 ± 0.01	20.84 ± 1.55	0.46 ± 0.04	0.39 ± 0.02
SART	19.38 ± 1.62	0.41 ± 0.04	0.50 ± 0.05	21.76 ± 1.56	0.50 ± 0.05	0.51 ± 0.04	24.36 ± 1.61	0.63 ± 0.05	0.55 ± 0.06	25.77 ± 1.69	0.71 ± 0.05	0.61 ± 0.06
SART+TV	27.59 ± 1.93	0.85 ± 0.03	0.60 ± 0.02	27.84 ± 1.95	0.87 ± 0.03	0.68 ± 0.07	27.99 ± 1.97	0.88 ± 0.02	0.73 ± 0.05	28.04 ± 1.97	0.88 ± 0.02	0.81 ± 0.02
BM3D 0.35	27.95 ± 2.08	0.90 ± 0.03	0.60 ± 0.09	28.45 ± 2.01	0.91 ± 0.02	0.68 ± 0.08	28.70 ± 1.94	0.92 ± 0.02	0.79 ± 0.04	28.76 ± 1.94	0.92 ± 0.02	0.84 ± 0.04
BM3D 0.20	26.89 ± 2.26	0.85 ± 0.07	0.62 ± 0.07	28.27 ± 2.13	0.90 ± 0.03	0.71 ± 0.07	28.76 ± 2.03	0.92 ± 0.02	0.77 ± 0.05	28.87 ± 2.06	0.92 ± 0.02	0.86 ± 0.03
DIP+TV	25.84 ± 2.21	0.79 ± 0.08	0.50 ± 0.12	27.36 ± 1.83	0.85 ± 0.05	0.65 ± 0.07	28.55 ± 1.78	0.90 ± 0.03	0.79 ± 0.06	28.82 ± 1.76	0.91 ± 0.02	0.89 ± 0.02
FBP+U-Net	30.48 ± 1.80	0.93 ± 0.02	0.81 ± 0.05	31.27 ± 1.79	0.94 ± 0.01	0.83 ± 0.05	31.90 ± 1.85	0.95 ± 0.01	0.86 ± 0.03	32.21 ± 1.89	0.95 ± 0.01	0.90 ± 0.01
Proj2Proj	28.12 ± 1.70	0.91 ± 0.02	0.81 ± 0.04	28.36 ± 1.76	0.92 ± 0.01	0.85 ± 0.02	28.54 ± 1.85	0.92 ± 0.01	0.87 ± 0.02	28.63 ± 1.85	0.93 ± 0.01	0.92 ± 0.01

Table 1 summarizes the quantitative performance of reconstruction methods with 64-view CT projections from the ellipses dataset at different noise levels. FBP produced the poorest results based on these metrics for all noise levels. The performance of SART was significantly improved when its output image was denoised *via* TV or BM3D with different  $\sigma$  values. As an unsupervised deep CNN based method, DIP+TV, showed slightly worse results than iterative and regularized reconstruction methods in general. However, the performance of DIP+TV increases with decreasing noise level. Considering the unsupervised architecture, Proj2Proj outperforms DIP+TV in all metrics, especially in higher noise levels which is the main challenge in low-dose CT reconstruction. However, the supervised method generated significantly better results than Proj2Proj in terms of PSNR and SSIM values. LPIPS aims to mimic human perception in terms of measuring image similarity. Therefore, it is a good candidate for measuring the quality of CT images in terms of clinical usability. For the LPIPS metric, the proposed Proj2Proj method produced significantly higher results than iterative and regularized methods and a competitor unsupervised method, DIP+TV. Moreover, it produced slightly higher values than the supervised equivalent of our method, FBP+U-Net and this is also supported by the visual assessments.

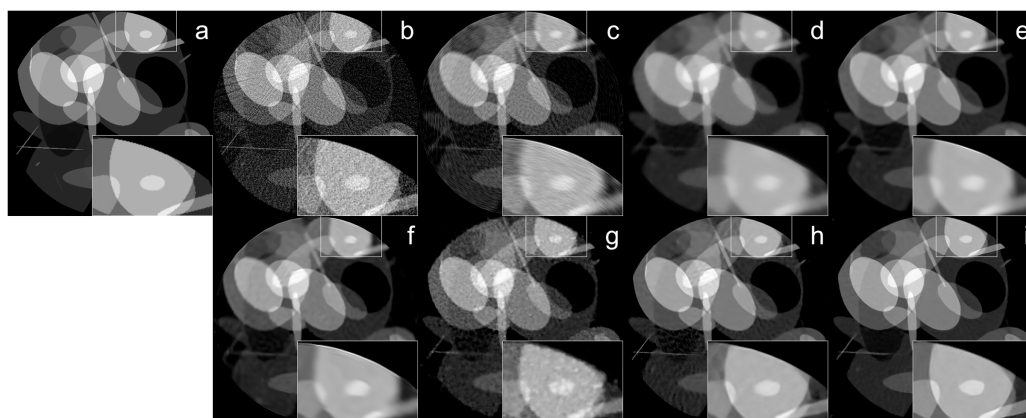
The reconstruction of Shepp–Logan phantom (*Shepp & Logan, 1974*) results are given in Fig. 2. FBP reconstruction suffered from severe artifacts. SART generated better results than FBP however it still contains significant background noise. Although TV and BM3D suppress this background noise successfully to some extent, the detectability of fine details was adversely affected due to their over-smoothing effect. Compared to these methods, Proj2Proj method performed the best in both suppressing the background noise and recovering the fine details. Considering deep CNN based reconstructions, DIP+TV suffers from deformities in details though it provides a smoother background than SART. However, Proj2Proj and supervised FBP+U-net produced better results than all other methods with a smoother background in Proj2Proj. For a better qualitative comparison, Fig. 2 is zoomed which favored our method in recovering sharper features. In Fig. 3, the reconstruction results of an ellipses image with a noise level of 33 dB are given. When the performance of all methods are compared, there is a strong analogy with those obtained for





**Figure 2** Shepp-Logan phantom reconstruction results from 64-view projections with 37 dB noise level: (A) ground truth, (B) FBP, (C) SART, (D) SART+TV, (E) SART+BM3D ( $\sigma = 0.35$ ), (F) SART+BM3D ( $\sigma = 0.20$ ), (G) DIP+TV, (H) FBP+U-Net, (I) Proj2Proj trained on ellipses dataset.

Full-size DOI: 10.7717/peerj-cs.1849/fig-2



**Figure 3** Ellipses image reconstruction results from 64-view projections with 33 dB noise level: (A) ground truth, (B) FBP, (C) SART, (D) SART+TV, (E) SART+BM3D ( $\sigma = 0.35$ ), (F) SART+BM3D ( $\sigma = 0.20$ ), (G) DIP+TV, (H) FBP+U-Net, (I) Proj2Proj trained on ellipses dataset. Image source credit: Ellipses dataset, <https://github.com/jleuschn/dival/tree/master/dival/datasets>.

Full-size DOI: 10.7717/peerj-cs.1849/fig-3

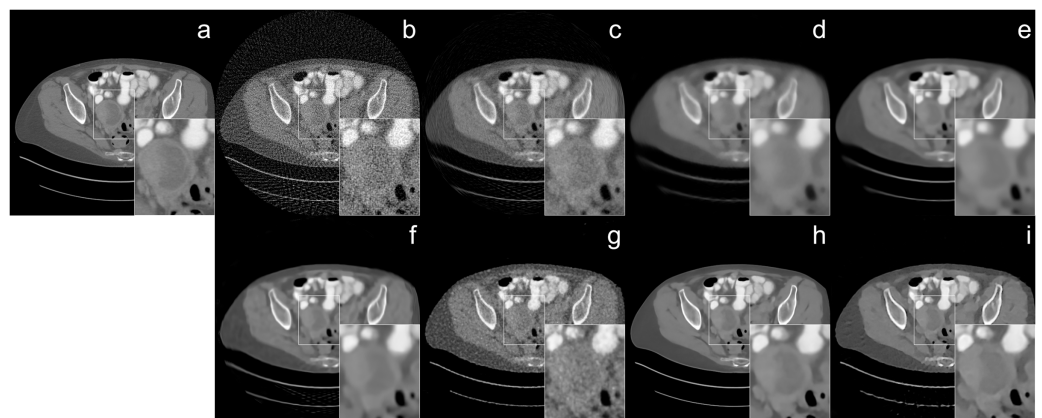
Shepp-Logan phantom given in Fig. 2. Our proposed Proj2Proj method gave superior results in both suppressing background noise and recovering the morphological structures.

Table 2 summarizes the quantitative performance of reconstruction methods with 64-view CT projections from human CT dataset at different noise levels. Proj2Proj method gave better performance than the state-of-the-art regularized iterative methods and DIP+TV. Since FBP+U-Net was trained in a supervised setting on the human CT data, it produced higher metrics in terms of the PSNR and SSIM. It should be emphasized that these metrics totally rely on the perfectness of ground truth images. However, in terms of LPIPS values, Proj2Proj produced similar results with FBP+U-Net.

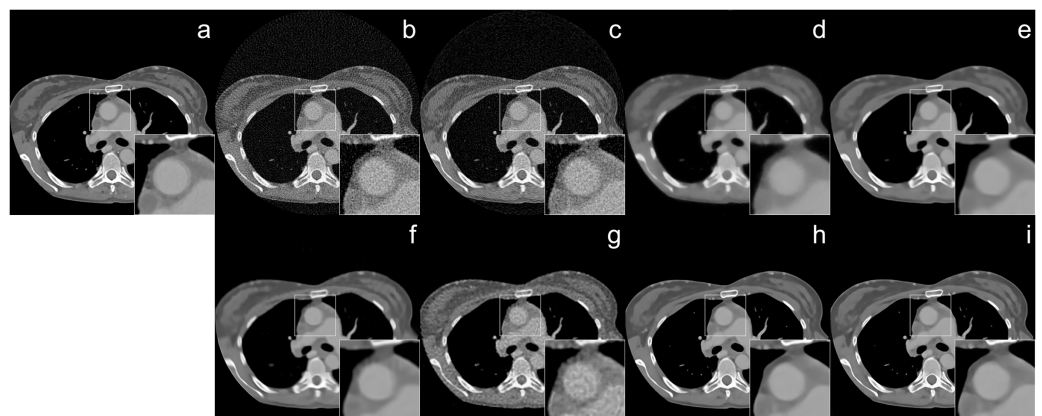
The reconstruction of human CT image results were given in Figs. 4 and 5 with different noise levels. When the reconstruction quality of all methods was compared, a strong

**Table 2** The human CT dataset results for 64 projections with the noise level 30 to 40 dB SNR.

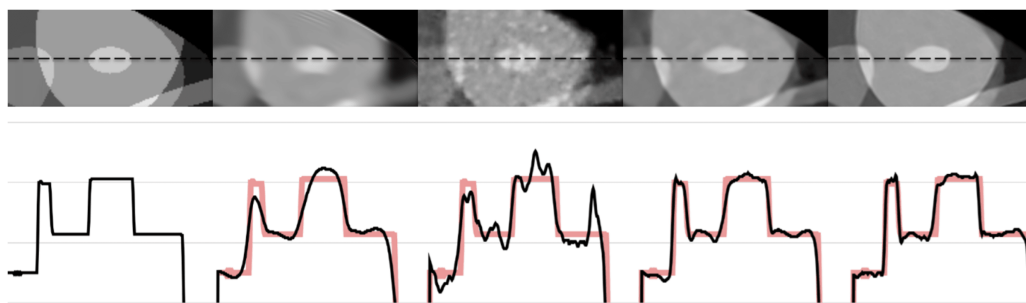
	30 dB SNR			33 dB SNR			37 dB SNR			40 dB SNR		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
FBP	15.03 ± 1.16	0.30 ± 0.02	0.27 ± 0.02	17.51 ± 0.95	0.36 ± 0.02	0.28 ± 0.02	20.28 ± 0.63	0.44 ± 0.02	0.31 ± 0.01	21.84 ± 0.57	0.50 ± 0.02	0.34 ± 0.01
SART	22.20 ± 1.03	0.51 ± 0.04	0.67 ± 0.09	24.29 ± 0.79	0.62 ± 0.04	0.68 ± 0.08	26.41 ± 0.69	0.73 ± 0.03	0.69 ± 0.04	27.44 ± 0.78	0.78 ± 0.03	0.71 ± 0.05
SART+TV	25.95 ± 0.80	0.84 ± 0.01	0.66 ± 0.07	26.02 ± 0.81	0.85 ± 0.01	0.72 ± 0.05	26.04 ± 0.82	0.86 ± 0.01	0.79 ± 0.03	26.05 ± 0.82	0.87 ± 0.01	0.84 ± 0.02
BM3D 0.35	28.29 ± 0.93	0.92 ± 0.01	0.75 ± 0.06	28.38 ± 0.96	0.92 ± 0.01	0.79 ± 0.05	28.44 ± 0.98	0.93 ± 0.01	0.82 ± 0.04	28.45 ± 0.99	0.93 ± 0.01	0.86 ± 0.02
BM3D 0.20	28.80 ± 0.96	0.91 ± 0.01	0.75 ± 0.06	29.13 ± 0.98	0.93 ± 0.01	0.79 ± 0.05	29.25 ± 1.02	0.93 ± 0.01	0.82 ± 0.04	29.27 ± 1.03	0.93 ± 0.01	0.86 ± 0.02
DIP+TV	25.49 ± 0.69	0.86 ± 0.03	0.79 ± 0.05	26.14 ± 0.84	0.89 ± 0.02	0.81 ± 0.04	26.40 ± 0.89	0.91 ± 0.01	0.84 ± 0.04	26.45 ± 0.91	0.91 ± 0.01	0.85 ± 0.03
FBP+U-Net	34.07 ± 1.01	0.97 ± 0.01	0.87 ± 0.02	35.09 ± 1.07	0.97 ± 0.01	0.89 ± 0.03	36.14 ± 1.13	0.98 ± 0.01	0.93 ± 0.01	36.65 ± 1.20	0.98 ± 0.01	0.95 ± 0.01
Proj2Proj	26.00 ± 0.77	0.90 ± 0.01	0.84 ± 0.03	26.10 ± 0.81	0.90 ± 0.01	0.88 ± 0.02	26.16 ± 0.83	0.91 ± 0.01	0.91 ± 0.01	26.21 ± 0.85	0.91 ± 0.01	0.94 ± 0.02



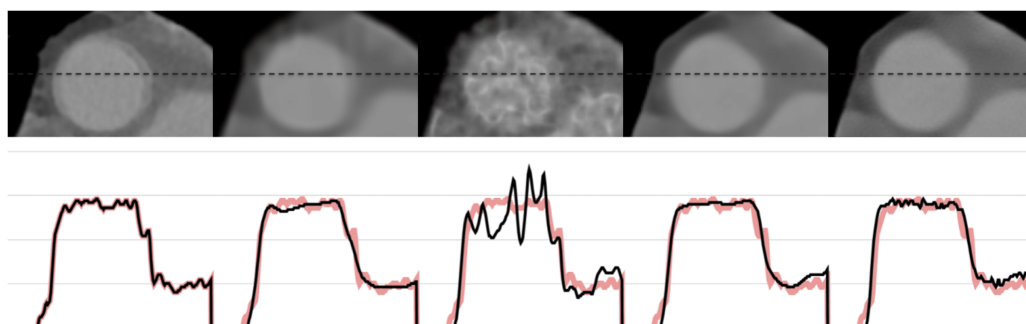
**Figure 4** Human CT image reconstruction results from 64-view with 33 dB SNR noise level: (A) ground truth, (B) FBP, (C) SART, (D) SART+TV, (E) SART+BM3D ( $\sigma = 0.35$ ), (F) SART+BM3D ( $\sigma = 0.20$ ), (G) DIP+TV, (H) FBP+U-Net, (I) Proj2Proj trained on human CT dataset. Image source credit: *Yang et al. (2018)*. [Full-size !\[\]\(7c47b229ca7bdb95c18f544ee7ceb332\_img.jpg\) DOI: 10.7717/peerj-cs.1849/fig-4](https://doi.org/10.7717/peerj-cs.1849/fig-4)



**Figure 5** Human CT image reconstruction results from 64-view 37 dB SNR noise level: (A) ground truth, (B) FBP, (C) SART, (D) SART+TV, (E) SART+BM3D ( $\sigma = 0.35$ ), (F) SART+BM3D ( $\sigma = 0.20$ ), (G) DIP+TV, (H) FBP+U-Net, (I) Proj2Proj trained on human CT dataset. Image source credit: *Yang et al. (2018)*. [Full-size !\[\]\(0f04eac0f44b45c27c855fb13a3dc0b4\_img.jpg\) DOI: 10.7717/peerj-cs.1849/fig-5](https://doi.org/10.7717/peerj-cs.1849/fig-5)



**Figure 6** The 1-D profiles of the reconstructions from left to right: ground truth, SART+BM3D ( $\sigma = 0.20$ ), DIP+TV, FBP+U-Net, proposed Proj2Proj method. Image source credit: *Yang et al. (2018)*. Full-size DOI: 10.7717/peerj-cs.1849/fig-6



**Figure 7** The 1-D profiles of the reconstructions from left to right: ground truth, SART+BM3D ( $\sigma = 0.20$ ), DIP+TV, FBP+U-Net, proposed Proj2Proj method. Image source credit: *Yang et al. (2018)*. Full-size DOI: 10.7717/peerj-cs.1849/fig-7

analogy was observed with those obtained from Shepp-Logan and ellipses dataset. In Fig. 4, streak artifact is observed in the image (i). Since FBP method is used as the initial reconstruction for our method, the structured artifact that arises from the initial reconstructor (FBP) is prolonged by the proposed network. The noise level also plays an important role in the formation of structural artifacts. The noise level in Fig. 4 is higher than the one in Fig. 5 which adversely affects the reconstruction quality of FBP the most. As a result, the image (i) reconstructed by Proj2Proj in Fig. 5 does not have similar artifacts.

Figures 6 and 7 show the 1-D intensity profiles passing through the dashed line from the reconstructed images which enables us to see how smoothly the intensities change. As the numerical and visual evaluation suffers from poor results, FBP, iterative, and regularized reconstruction methods are not given in the comparisons except for SART+BM3D with  $\sigma = 0.20$ . The black line represents the 1-D intensity of ground truth and other methods while the red line is a copy of ground truth to show how well the reconstruction methods fit over the ground truth. In Fig. 6, SART+BM3D shows an over-smoothing through the line while DIP+TV shows undesired spikes along the same line. Proj2Proj and FBP+U-Net have similar line intensity profiles but Proj2Proj has slightly sharper edges than FBP+U-Net. In Fig. 7, a similar analogy is observed with SART+BM3D and DIP+TV methods. The ground truth image in Fig. 7 is the one reconstructed with normal-dose taken from the

**Table 3** The execution times of the reconstructions in seconds.

FBP	SART	SART+TV	SART+BM3D	DIP+TV	FBP+U-Net	Proj2Proj
0.38	34.16	36.91	41.95	320.20	1.21	1.27

dataset and thus it consists of small minor spikes throughout the line which can be considered as the noise by its nature. FBP+U-Net here smooths out these spikes and creates a smoother image however in our proposed Proj2Proj method, it is evident that the small spikes are also preserved.

Another parameter that should be considered is the reconstruction time given in Table 3. Although the training times are quite long for the self-supervised method, the reconstruction times on the trained neural networks are very short thanks to modern computer technology.

## DISCUSSION

Widely used image quality measurement metrics, SSIM and PSNR, have some limitations such as smoothness-biased and pixel-to-pixel difference measurement rather than focusing on perceptual similarity. Besides, in CT imaging, the reference images obtained from normal-dose CT scans contain noise and artifacts by its nature. The supervised method learns a mapping from low-dose to normal-dose images including all these imperfections as well. This might lead the supervised method to generate very high PSNR and SSIM scores even though it may not give the best reconstruction quality. Recently a new approach was proposed (Zhang et al., 2018) to measure the similarity between two images which outperformed traditional metrics when we compare them with human preference. For CT reconstruction problem, the perceptual difference is more important than pixel-to-pixel differences. Therefore LPIPS metrics were also added to the quantitative results.

One of the challenges of our proposed training method is the convergence of the network weights to an identity function. To overcome this problem, our method uses the  $J_{th}$  invariance method proposed in the Noise2Self (Batson & Royer, 2019) study, which performs a pixel-based denoising assuming the independence of inter-pixel noise. In a low-dose CT image, noise cannot be modeled independently on a pixel-by-pixel basis due to the back projection operator. Therefore, the  $J_{th}$  invariance method was applied in the projection domain. When the self-supervision principle was also tested in the image domain, the network converged to an identity function, which led the network to produce a result almost identical to the output of the FBP operator.

One of the questions that may arise is why noise reduction is done in the image domain. Although there are studies that eliminate noise in the sinogram domain (Lee et al., 2019; Anirudh et al., 2018), CNN neural networks are generally designed for natural images, and sinogram images have different characteristics from natural images. Besides, DIP study (Ulyanov, Vedaldi & Lempitsky, 2018) proposes an unsupervised method for image domain inverse problems by exploiting the structure of Deep CNNs.

Another discussion regarding the implementation of the proposed method might be choosing FBP as the initial reconstruction method. FBP, the most conventional reconstruction method, is used in FBP+U-Net (Jin *et al.*, 2017), and a very similar architecture with FBP+U-Net was chosen for a fair comparison. On the other hand, iterative or regularized iterative methods for the initial reconstruction might help to obtain improved results. Among various supervised low-dose CT reconstruction methods, FBP+U-Net was found one of the best and fairest candidates to be used in the comparisons considering its network architecture and reconstruction approach.

Clinical applicability, which is one of the biggest factors in measuring the success of a reconstruction method, increases when all details can be reconstructed and structural details can be easily examined by radiologists. In our study, the deep lesion (Yan *et al.*, 2017) dataset was selected as medical CT data and we aimed to predict the lesion detection performance of reconstruction methods. In addition to the superiority of the Proj2Proj method in quantitative analysis, the fact that it is clearly more successful than the other methods in the recovery of fine details, which shows its clinical applicability potential.

## CONCLUSION

In this study, it was shown that low-dose CT reconstruction problem can be tackled by defining a training scheme to use low-dose projections as their own training targets even if the low-dose/normal-dose pairs of large datasets are not available. Even though traditional metrics might favor other methods depending on test setups and noise levels, considering detailed visual assessments including 1-D profiles and a novel deep CNN based quality assessment metric which evaluates the clinical usability of the results the best compared to the other metrics, Proj2Proj method is the most favorable one among all others including a well-known supervised method, FBP+U-net. It has great potential in other medical imaging problems where the same limitations exist and the assumptions used in this study are valid. As an advanced stage of this study, the method can be adapted to other imaging modalities.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The authors received no funding for this work.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Mehmet Ozan Unal conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Metin Ertas analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

- Isa Yildirim conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

### Data Availability

The following information was supplied regarding data availability:

The DeepLesion dataset is available at NIH Box (Ronald Summers): <https://nihcc.app.box.com/v/DeepLesion>.

The simulated data, Ellipses dataset, is available at GitHub (dival/dival/datasets/ellipses\_dataset.py): <https://github.com/jleuschn/dival/tree/master/dival/datasets>.

## REFERENCES

- Adler J, Öktem O. 2018.** Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging* 37(6):1322–1332 DOI 10.1109/TMI.2018.2799231.
- Andersen A, Kak A. 1984.** Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm. *Ultrasonic Imaging* 6(1):81–94 DOI 10.1177/016173468400600107.
- Anirudh R, Kim H, Thiagarajan JJ, Mohan KA, Champley K, Bremer T. 2018.** Lose the views: limited angle ct reconstruction via implicit sinogram completion. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 6343–6352.
- Baguer DO, Leuschner J, Schmidt M. 2020.** Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Problems* 36(9):094004 DOI 10.1088/1361-6420/aba415.
- Batson J, Royer L. 2019.** Noise2Self: blind denoising by self-supervision. In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9–15 June 2019*. Vol. 97: Long Beach, California, USA: Proceedings of Machine Learning Research, 524–533.
- Buades A, Coll B, Morel J-M. 2005.** A non-local algorithm for image denoising. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 2, Piscataway: IEEE, 60–65.
- Buzug T. 2008.** *Computed tomography: from photon statistics to modern cone-beam CT*. Berlin: Springer Science & Business Media.
- Candès E, Romberg J, Tao T. 2006.** Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics* 59(8):1207–1223 DOI 10.1002/cpa.20124.
- Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, Zhou J, Wang G. 2017.** Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging* 36(12):2524–2535 DOI 10.1109/TMI.2017.2715284.
- Dabov K, Foi A, Katkovnik V, Egiazarian K. 2007.** Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing* 16(8):2080–2095 DOI 10.1109/TIP.2007.901238.
- Donoho DL. 2006.** Compressed sensing. *IEEE Transactions on Information Theory* 52(4):1289–1306 DOI 10.1109/TIT.2006.871582.
- Guo X, Li L, Chang D, He P, Feng P, Yu H, Wu W. 2023.** Spectral2Spectral: image-spectral similarity assisted spectral CT deep reconstruction without reference. ArXiv preprint DOI 10.48550/arXiv.2210.01125.

- He J, Yang Y, Wang Y, Zeng D, Bian Z, Zhang H, Sun J, Xu Z, Ma J. 2019.** Optimizing a parameterized plug-and-play admm for iterative low-dose ct reconstruction. *IEEE Transactions on Medical Imaging* **38(2)**:371–382 DOI [10.1109/TMI.2018.2865202](https://doi.org/10.1109/TMI.2018.2865202).
- Hendriksen AA, Pelt DM, Batenburg KJ. 2020.** Noise2Inverse: self-supervised deep convolutional denoising for tomography. *IEEE Transactions on Computational Imaging* **6**:1320–1335 DOI [10.1109/TCI.2020.3019647](https://doi.org/10.1109/TCI.2020.3019647).
- Jin KH, McCann MT, Froustey E, Unser M. 2017.** Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing* **26(9)**:4509–4522 DOI [10.1109/TIP.2017.2713099](https://doi.org/10.1109/TIP.2017.2713099).
- Kingma DP, Ba J. 2015.** Adam: a method for stochastic optimization. In: *3rd International Conference on Learning Representations, ICLR 2015, May 7–9, 2015, Conference Track Proceedings*. San Diego, CA, USA.
- Krull A, Buchholz T, Jug F. 2019.** Noise2void-learning denoising from single noisy images. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, June 16–20, 2019*. Piscataway: IEEE, 2129–2137.
- Lagerwerf MJ, Hendriksen AA, Buurlage J-W, Batenburg KJ. 2020.** Noise2Filter: fast, self-supervised learning and real-time reconstruction for 3d computed tomography. ArXiv preprint DOI [10.48550/arXiv.2007.01636](https://doi.org/10.48550/arXiv.2007.01636).
- Lee H, Lee J, Kim H, Cho B, Cho S. 2019.** Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences* **3(2)**:109–119 DOI [10.1109/TRPMS.2018.2867611](https://doi.org/10.1109/TRPMS.2018.2867611).
- Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M, Aila T. 2018.** Noise2Noise: learning image restoration without clean data. In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, July 10–15, 2018, Proceedings of Machine Learning Research*. Vol. 80: Stockholmsmässan, Stockholm, Sweden: PMLR, 2971–2980.
- Liu Z, Bicer T, Kettimuthu R, Gursoy D, Carlo FD, Foster I. 2020.** Tomogan: low-dose synchrotron x-ray tomography with generative adversarial networks: discussion. *Journal of the Optical Society of America A* **37(3)**:422–434 DOI [10.1364/JOSAA.375595](https://doi.org/10.1364/JOSAA.375595).
- Pelt DM, Batenburg KJ. 2013.** Fast tomographic reconstruction from limited data using artificial neural networks. *IEEE Transactions on Image Processing* **22(12)**:5238–5251 DOI [10.1109/TIP.2013.2283142](https://doi.org/10.1109/TIP.2013.2283142).
- Quan Y, Chen M, Pang T, Ji H. 2020.** Self2self with dropout: learning self-supervised denoising from single image. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 1887–1895.
- Rudin LI, Osher S, Fatemi E. 1992.** Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* **60(1)**:259–268 DOI [10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F).
- Shepp LA, Logan BF. 1974.** The fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science* **21(3)**:21–43 DOI [10.1109/TNS.1974.6499235](https://doi.org/10.1109/TNS.1974.6499235).
- Sidky EY, Pan X. 2008.** Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine and Biology* **53(17)**:4777–4807 DOI [10.1088/0031-9155/53/17/021](https://doi.org/10.1088/0031-9155/53/17/021).
- Ulyanov D, Vedaldi A, Lempitsky V. 2018.** Deep image prior. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE.
- Unal MO, Ertas M, Yildirim I. 2021.** Self-supervised training for low-dose ct reconstruction. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. Piscataway: IEEE, 69–72.

- Wang Z, Simoncelli EP, Bovik AC. 2003.** Multiscale structural similarity for image quality assessment. In: *The Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003*. Vol. 21398–1402.
- Wang X, Yu F, Dou Z, Darrell T, Gonzalez JE. 2018.** SkipNet: learning dynamic routing in convolutional networks. In: *Computer Vision-ECCV 2018-15th European Conference, September 8–14, 2018, Proceedings, Part XIII*. Vol. 11217: Munich, Germany: Springer, 420–436.
- Wu W, Hu D, Niu C, Yu H, Vardhanabhuti V, Wang G. 2021.** Drone: dual-domain residual-based optimization network for sparse-view CT reconstruction. *IEEE Transactions on Medical Imaging* **40(11)**:3002–3014 DOI [10.1109/TMI.2021.3078067](https://doi.org/10.1109/TMI.2021.3078067).
- Wu D, Kim K, El Fakhri G, Li Q. 2017.** Iterative low-dose ct reconstruction with priors trained by artificial neural network. *IEEE Transactions on Medical Imaging* **36(12)**:2479–2486 DOI [10.1109/TMI.2017.2753138](https://doi.org/10.1109/TMI.2017.2753138).
- Wu W, Wang Y, Liu Q, Wang G, Zhang J. 2023.** Wavelet-improved score-based generative model for medical imaging. *IEEE Transactions on Medical Imaging* **1** DOI [10.1109/TMI.2023.3325824](https://doi.org/10.1109/TMI.2023.3325824).
- Xie Y, Wang Z, Ji S. 2020.** Noise2same: Optimizing A self-supervised bound for image denoising. In: Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H, eds. *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, December 6–12, 2020*. Virtual: NeurIPS.
- Yan K, Wang X, Lu L, Summers RM. 2017.** Deeplesion: automated deep mining, categorization and detection of significant radiology image findings using large-scale clinical lesion annotations. ArXiv preprint DOI [10.48550/arXiv.1710.01766](https://doi.org/10.48550/arXiv.1710.01766).
- Yang L, Li Z, Ge R, Zhao J, Si H, Zhang D. 2023.** Low-dose CT denoising via sinogram inner-structure transformer. *IEEE Transactions on Medical Imaging* **42(4)**:910–921 DOI [10.1109/TMI.2022.3219856](https://doi.org/10.1109/TMI.2022.3219856).
- Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra MK, Zhang Y, Sun L, Wang G. 2018.** Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging* **37(6)**:1348–1357 DOI [10.1109/TMI.2018.2827462](https://doi.org/10.1109/TMI.2018.2827462).
- Yu H, Wang G. 2009.** Compressed sensing based interior tomography. *Physics in Medicine & Biology* **54(9)**:2791–2805 DOI [10.1088/0031-9155/54/9/014](https://doi.org/10.1088/0031-9155/54/9/014).
- Zhang R, Isola P, Efros AA, Shechtman E, Wang O. 2018.** The unreasonable effectiveness of deep features as a perceptual metric. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Zhou B, Chen X, Xie H, Zhou SK, Duncan JS, Liu C. 2022.** DudoufNet: dual-domain under-to-fully-complete progressive restoration network for simultaneous metal artifact reduction and low-dose CT reconstruction. *IEEE Transactions on Medical Imaging* **41(12)**:3587–3599 DOI [10.1109/TMI.2022.3189759](https://doi.org/10.1109/TMI.2022.3189759).