Computer Methods and Programs in Biomedicine 226 (2022) 107167

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Sparse-view and limited-angle CT reconstruction with untrained networks and deep image prior☆

Ziyu Shu*, Alireza Entezari

CISE Department, University of Florida, Gainesville, FL 32611-6120, USA

ARTICLE INFO

Article history:
Received 18 January 2022
Revised 27 September 2022
Accepted 29 September 2022

Keywords:
X-ray tomography
Sparse-view
Limited-angle
Deep image prior
Iterative reconstruction
Neural network

ABSTRACT

Background and objective: Neural network based image reconstruction methods are becoming increasingly popular. However, limited training data and the lack of theoretical guarantees for generalizability raised concerns, especially in biomedical imaging applications. These challenges are known to lead to an unstable reconstruction process that poses significant problems in biomedical image reconstruction. In this paper, we present a new framework that uses untrained generator networks to tackle this challenge, leveraging the structure of deep networks for regularizing solutions based on a technique known as Deep Image Prior (DIP).

Methods: To achieve a high reconstruction accuracy, we propose a framework optimizing both the latent vector and the weights of a generator network during the reconstruction process. We also propose the corresponding reconstruction strategies to improve the stability and convergent performance of the proposed framework. Furthermore, instead of calculating forward projection in each iteration, we propose implementing its normal operator as a convolutional kernel under parallel beam geometry, thus greatly accelerating the calculation.

Results: Our experiments show that the proposed framework has significant improvements over other state-of-the-art conventional, pre-trained, and untrained methods under sparse-view, limited-angle, and low-dose conditions.

Conclusions: Applying to parallel beam X-ray imaging, our framework shows advantages in speed, accuracy, and stability of the reconstruction process. We also show that the proposed framework is compatible with all differentiable regularizations that are commonly used in biomedical image reconstruction literature. Our framework can also be used as a post-processing technique to further improve the reconstruction generated by any other reconstruction methods. Furthermore, the proposed framework requires no training data and can be adjusted on-demand to adapt to different conditions (e.g. noise level, geometry, and imaged object).

1. Introduction

Neural networks have achieved unprecedented success in a wide range of applications. They have also emerged as a new tool in CT reconstruction with the potential to change the field. Reconstruction methods in this area generally use neural networks to find a mapping from raw inputs to specific outputs. One such example is the mapping from sinogram data to reconstructed images. Neural networks can not only build end-to-end image reconstruction algorithms [1,2], but also enhance the performance of any procedures in conventional reconstruction methods like plugins [3–5]. Furthermore, neural networks allow for implementing complicated prior [6] and have the potential to enable high-quality low-dose, or sparse-measurement (e.g. sparse-view and limited-angle) CT reconstruction [3,4].

Still, neural network related methods face their own challenges. First, as data-driven methods, neural networks require plenty of training data. However, getting enough training data is challenging in specific biomedical imaging applications. In practice, high-dose reconstructions obtained with classical methods are considered ground truth, which implies that patients have to be exposed to high doses of X-ray radiation. Second, neural network related methods lack classical guarantees. These methods are built on the assumption that the distribution of test input should be the same as that of the training input. Although the loss functions for the

☆ This work was supported in part by the NSF [Grant number CCF-2210866].
* Corresponding author.
E-mail address: ziyushu@ufl.edu (Z. Shu).

https://doi.org/10.1016/j.cmpb.2022.107167
0169-2607/

training data is minimized during the training process, there is no guarantee that such loss functions can also be minimized for each inference input. Third, networks have to be retrained for each specific setting (e.g. reconstruction resolution, sinogram domain sampling ratio, noise level, imaging objects, etc.). Hence, in practice, hundreds of different networks are required to facilitate that many different settings. As a result, implementing neural networks in medical image reconstruction imposes the risk of missing patient-specific features. Small abnormal changes, which would be considered symbols of illness by radiologists, may be ignored by neural network related algorithms and cause severe consequences [7].

In this paper, we propose a new framework for CT reconstruction. It uses an untrained generator network as a prior, and both the weights and the latent vector of the generator are optimized iteratively during the reconstruction process to match the observed measurements. While reference images (optional) can be used in the regularizers to guide and accelerate the reconstruction, the proposed framework has no training process and does not require a training dataset. Thus, the problems caused by the training dataset and process can be eliminated. The extraordinary ability of neural networks together with a strategy to increase the stability of the reconstruction process minimizes the difference between the measurement and reconstruction result. Also, the DIP helps generate a more natural result without a training process. The experiments show that the proposed framework provides significant improvements over other state-of-the-art conventional methods, pre-trained and untrained models, especially under sparse-view, limited-angle, and low-dose conditions.

The remainder of this paper is organized as follows: the details of the CT reconstruction problem, related works, and proposed framework are introduced in Section 2; the proposed framework is compared with conventional methods, pre-trained models, and an untrained method on phantoms and real CT images under different conditions in Section 3. We also show that the proposed framework can improve the results generated by other algorithms. The corresponding discussion is in Section 4; the conclusion of this study is presented in Section 5.

2. Materials and methods

2.1. Computed tomography and MBIR

Computed tomography (CT) is an essential technology with a wide range of applications in biomedical imaging. Since its introduction in the 1970s, multiple methods have been proposed to improve its speed and accuracy. However, conventional methods cannot provide quality images under low-dose or sparse-measurement (sparse-view and limited-angle) conditions, which is necessary to reduce the potentially harmful radiation. In our work, we compare our proposed method with conventional reconstruction methods, pre-trained and untrained neural network related methods on these non-ideal scenarios for a basic 2D parallel beam geometry. In that case, the forward operator is given by the Radon transform [8], which can be expressed as:

$$\boldsymbol{g}(y) = \mathcal{P}_{\boldsymbol{\theta}}\{f\}(y) = \int_{\mathbb{R}} f(t\boldsymbol{\theta} + \boldsymbol{P}_{\boldsymbol{\theta}^{\perp}}^{\mathrm{T}}\boldsymbol{y})\mathrm{d}t,$$

where $\boldsymbol{P}_{\boldsymbol{\theta}^{\perp}}$ is a 1×2 transformation matrix that geometrically projects the 2D x-coordinate system onto the 1D y-coordinate system perpendicular to $\boldsymbol{\theta}$. The attenuation map of an imaged object $f(\boldsymbol{x})$ can be represented by a discretization kernel ψ as:

$$f_{\psi}(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^2} c_{\boldsymbol{k}}\psi(\boldsymbol{x}-\boldsymbol{k}),$$

where $c_{\boldsymbol{k}}$ is a set of coefficients of total number n^2. Thus, the forward model can be represented as a matrix:

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{c} = \begin{bmatrix} \mathcal{P}_{\boldsymbol{\theta}_1}\{\psi\}(y_1 - \boldsymbol{P}_{\boldsymbol{\theta}_1^{\perp}}\boldsymbol{k}_1) & \dots & \mathcal{P}_{\boldsymbol{\theta}_1}\{\psi\}(y_1 - \boldsymbol{P}_{\boldsymbol{\theta}_1^{\perp}}\boldsymbol{k}_{n^2}) \\ \vdots & & \vdots \\ \mathcal{P}_{\boldsymbol{\theta}_1}\{\psi\}(y_l - \boldsymbol{P}_{\boldsymbol{\theta}_1^{\perp}}\boldsymbol{k}_1) & \dots & \mathcal{P}_{\boldsymbol{\theta}_1}\{\psi\}(y_l - \boldsymbol{P}_{\boldsymbol{\theta}_1^{\perp}}\boldsymbol{k}_{n^2}) \\ \vdots & & \vdots \\ \mathcal{P}_{\boldsymbol{\theta}_m}\{\psi\}(y_1 - \boldsymbol{P}_{\boldsymbol{\theta}_m^{\perp}}\boldsymbol{k}_1) & \dots & \mathcal{P}_{\boldsymbol{\theta}_m}\{\psi\}(y_1 - \boldsymbol{P}_{\boldsymbol{\theta}_m^{\perp}}\boldsymbol{k}_{n^2}) \\ \vdots & & \vdots \\ \mathcal{P}_{\boldsymbol{\theta}_m}\{\psi\}(y_l - \boldsymbol{P}_{\boldsymbol{\theta}_m^{\perp}}\boldsymbol{k}_1) & \dots & \mathcal{P}_{\boldsymbol{\theta}_m}\{\psi\}(y_l - \boldsymbol{P}_{\boldsymbol{\theta}_m^{\perp}}\boldsymbol{k}_{n^2}) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_{n^2} \end{bmatrix},$$

where y_1 to y_l indicate l sampling points in the sinogram domain, $\boldsymbol{\theta}_1$ to $\boldsymbol{\theta}_m$ indicate m projection angles, and $\boldsymbol{k}_1$ to $\boldsymbol{k}_{n^2}$ indicate the coordinates of n^2 pixels. The CT reconstruction problem aims to get $\boldsymbol{c}$ from the measurement $\boldsymbol{g}$, which requires inverting the matrix $\boldsymbol{A}$. However, under the sparse-measurement condition, the inverse problem is ill-posed.

Conventional model based iterative reconstruction (MBIR) related methods calculate the residual between the reconstruction and the measurement ($\boldsymbol{g}-\boldsymbol{A}\boldsymbol{c}$), then back project it to the image domain ($\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{g}-\boldsymbol{A}\boldsymbol{c}) = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{g} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{c}$) in each iteration to obtain reconstruction results. Furthermore, the sparsity of the original images can be utilized to improve the reconstruction quality. As a result, multiple regularizers such as total variation (TV) [9], anisotropic total variation (ATV) [10], reweighted anisotropic total variation (RwATV) [11], and anisotropic relative total variation (ARTV) [12] are proposed to improve the reconstruction quality under sparse-measurement or low-dose condition. Other researchers proposed using Markov random field (MRF) to pursue similar results, such as Gaussian MRF [13], Gaussian mixture MRF (GM-MRF) [14], and q-generalized Gaussian MRF (qGGMRF) [15]. In that case, the MBIR methods involve solving an unconstrained optimization problem:

$$\arg\min_{\boldsymbol{c}} |\boldsymbol{g}-\boldsymbol{A}\boldsymbol{c}|^2 + \lambda R(\boldsymbol{c}), \tag{1}$$

where R indicates regularizers. However, these methods are far from ideal, especially when the number of views or the angular range of the projection angles is too small. According to the central slice theorem (Fig. 1(a)), the 1D Fourier transform of a projection is equal to a slice in the frequency domain of the original image. Similarly, the back projection operation is equivalent to updating the corresponding frequency slice in the frequency domain. Thus, a missing projection is equal to a missing slice in the frequency domain (Fig. 1(b)), and the sparse-measurement CT reconstruction is equal to recovering the whole frequency image from the few slices indicated by the blue area (Fig. 1(c)) or lines (Fig. 1(d)). However, MBIR methods with regularizations such as total variation cannot solve the problem very well. This is because the forward and back projection scenario itself can only update the frequency pixels in the measured area, and the unknown area cannot be accurately inpainted with only the regularizations such as TV. To tackle this challenge, multiple neural network related methods are proposed.

2.2. Neural network related methods

Neural network approaches can be used for CT image reconstruction both directly and indirectly. Zhu et al. [1] proposed a unified, end-to-end reconstruction framework called AUTOMAP. This framework uses two convolutional layers and three large fully connected layers to learn the mapping between the measured sinogram and the reconstructed image. The author claimed that AUTOMAP outperforms conventional reconstruction methods on noise

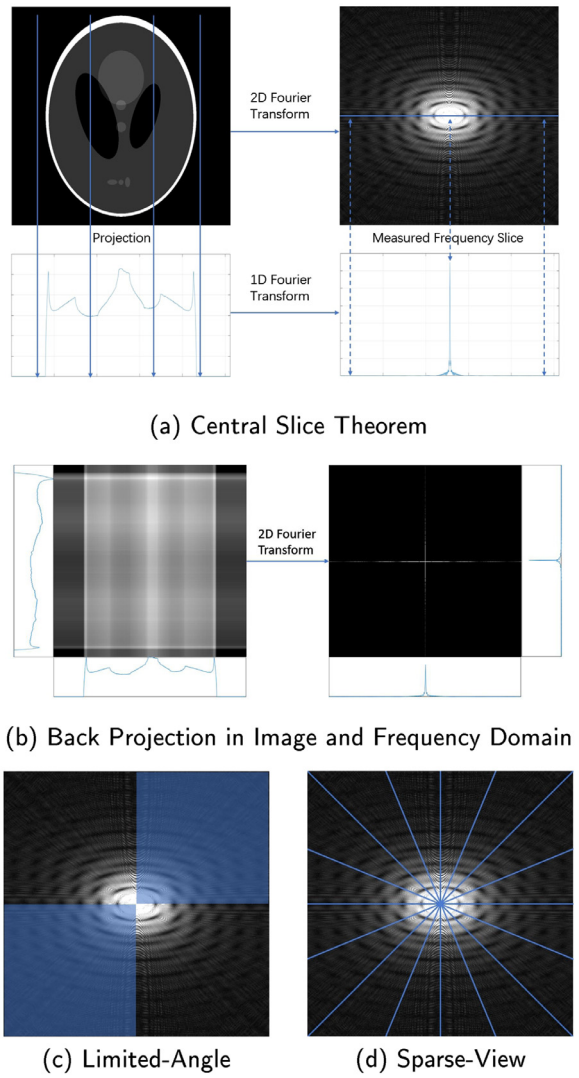

Fig. 1. (a) The Fourier transform of the projection of an image is equal to a slice of the Fourier transform of that image through the origin in the Fourier space perpendicular to the projection angle. (b) Back projection (2 views) in the image (left) and frequency (right) domains. (c) Limited-angle ($0^\circ - 90^\circ$ in this case) CT reconstruction is equal to recovering the whole frequency image from the frequency slices in the blue area. (d) Sparse-view CT reconstruction is equal to recovering the whole frequency image from the few slices indicated by the blue lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and artifact reduction tasks and can be used in multiple image reconstruction areas such as CT, PET, and MRI. The main shortcoming of AUTOMAP is that its main building blocks consist of large fully connected layers. Depending on the problem, the number of parameters can grow quickly with the data dimension, which makes efficient training, storage, and inference of the model infeasible. In the past several years, more research was done to improve its performance and reduce its size. Ell50 and MED50 [16] are networks for CT reconstruction or any Radon transform-based inverse problems. They use multiple convolutional layers to reduce the number of parameters, and U-net [17] structure to extract features from multiple resolutions. Furthermore, measured data is first processed by filtered back projection (FBP), since the authors believe that the FBP encapsulates information about the physics of the inverse problem, provides a warm start to the CNN, and thus simplifies the learning procedure. As a result, the sizes of the networks are much smaller than that of AUTOMAP while achieving better results. IRadonMAP [18] proposed designing specific layers to imitate the procedures of the FBP algorithm and also achieved impressive results.

Neural networks can also work as plugins for conventional reconstruction methods to improve their performances. For example, a set of all images that satisfies regularizations (priors) such as non-negativity and data-mismatch is defined as a feasible set. Projections onto convex sets algorithms (POCS) such as ASD-POCS [19] require such a set. However, the set is hard to obtain under conditions such as sparse-view, limited-angle, and low-dose. Furthermore, conventional convex regularizations may be unable to generate an optimal feasible set. By using neural networks, some more complicated or non-convex priors can also be used to better define the feasible set. The "projection onto feasible set" operation can also be easily implemented. Algorithm 1 explains the idea in more detail, where line (2) is used to solve the conventional convex regularizations such as non-negativity, data-mismatch, and TV. The algorithm C can be a simple gradient descent rule or any iterative reconstruction (IR) algorithm. Line (3) is the realization of projection onto the feasible set, where the neural network G can be pre-trained to model complicated non-convex regularizations. During the reconstruction, G is fixed and the latent vector $\boldsymbol{z}$ is trainable so that the "projection onto feasible set" operation can be achieved. The loop can be repeated multiple times [20,21] so that G can steer the solution in each iteration. It can also be repeated just once so that G can be regarded as a post-processing method, which will be discussed in the next paragraph.

Algorithm 1 Neural network in MBIR.

Input: measurement matrix $\boldsymbol{A}$, measurement $\boldsymbol{g}$, initial guess $\boldsymbol{c}$, algorithm C solving the convex regularizations, and a pre-trained neural network G.

1: **Repeat:**
2: $\quad \boldsymbol{w} \leftarrow C(\boldsymbol{g}, \boldsymbol{Ac})$
3: $\quad \boldsymbol{c} \leftarrow G(\arg\min_z ||\boldsymbol{w} - G(\boldsymbol{z})||)$
4: **Until:** convergence, or a fixed number of iterations is reached.

Neural networks can act as pre or post-processing methods. Lee et al. [22] used a fully convolutional U-net to complete the incomplete sinogram from sparse-view measurement, improving the accuracy of sparse-view CT reconstruction. Anirudh et al. [23,24] achieved similar results using GAN. Chen et al. [5] removed noise from low-dose CT images with a three-layer convolutional neural network (CNN). A similar three-layer CNN can also be used for artifact reduction in limited-angle CT reconstruction [4]. Xie et al. [3] proposed using GAN to remove the artifacts for limited-angle CT reconstruction.

Recently, researchers proposed using multiple neural networks at different parts of the reconstruction process to obtain better results. Yin et al. [25] proposed using two neural network models to denoise in the sinogram domain and image domain respectively. Hu et al. [26,27] shared similar ideas with [25], the difference being the introduction of a discriminator to further guide the model. All these newly proposed methods have complex structures, loss functions, and multiple sub-networks. For example, the method proposed in Zhang et al. [27] uses two ResUNets for image and sinogram domain denoising, one discriminator for image domain discrimination, and five losses:

- The loss between denoised sinogram signal and ground truth sinogram signal.
- The loss between the filtered back projection (FBP) of the denoised sinogram signal and ground truth.
- The loss between the denoised FBP of the denoised sinogram signal and ground truth.

- The loss between the sinogram of the denoised FBP of the denoised sinogram signal and ground truth sinogram signal.
- The loss for the discriminator.

As mentioned in Section 1, these methods are problematic for large-scale implementation. The fundamental weakness is the requirement for a well-trained model. It is difficult, if not impossible, to properly train these models in the medical imaging area. Popular image processing benchmark such as CIFAR-10 contains 60,000 small images of size 32×32, while CT reconstruction benchmark such as the LIDC-IDRI [28] (Lung Image Database Consortium Image Collection) dataset only contains 1018 patients, which is insufficient to properly train a deep neural network for patients of all demographics. The fact that the training dataset may be imbalanced further complicates this issue. On top of that, network related methods require the inference input to share the same distribution with the training data, which may lead to severe consequences. For example, pre-trained models tend to ignore details like abnormal structures or tiny perturbations, and replace them with the ubiquitous textures learned from the training datasets. However, such details may correspond to the symptoms of an illness with an extremely low incident rate, making their replacements unacceptable [7]. The cascade of multiple sub-networks makes the problem even worse, as the perturbations may be amplified level by level. Furthermore, the complex structures of these models impede efficient training. As a result, all these models lack versatility and have to be trained for each specific problem at a very high cost.

In order to overcome these challenges, some researchers proposed using inverse GAN [29] related methods, where the latent vector $\boldsymbol{z}$ of a pre-trained generator $G(\boldsymbol{z}; \boldsymbol{w})$ is optimized by solving the problem:

$$\boldsymbol{z}^* = \arg\min_{\boldsymbol{z}} ||\boldsymbol{g} - \boldsymbol{A}G(\boldsymbol{z}; \boldsymbol{w})||_2^2, \ \hat{\boldsymbol{c}} = G(\boldsymbol{z}^*; \boldsymbol{w}). \tag{2}$$

This guarantees that at least a local minimum of the objective function can be found in the space spanned by the generator G. However, to obtain a high-quality result, an appropriate pre-trained model is still necessary.

2.3. Deep image prior

Methods requiring less or even no training data are also proposed to tackle the problem. Ulyanov et al. [30] pointed out that the structure of a convolutional network itself is sufficient to capture plenty of low-level image statistical priors. Thus, high-quality images can be generated in standard inverse problems such as denoising, inpainting, and super-resolution with no training process. Researchers also claimed that convolutional image generators fit natural images faster than noise and learn to construct them from low to high frequencies [30–32]. Bojanowski et al. [33] proposed assigning latent vectors to each training image and training a generator or decoder solely by these image-vector pairs. Impressive results were obtained in the absence of the corresponding discriminator or encoder (which is necessary for the general autoencoder and GAN framework). Thus, the author claimed that the proposed method shares many desirable properties with autoencoder and GAN, such as interpolating meaningfully between samples and performing linear arithmetic with noise vectors. These researches imply that one can generate images with relatively simple network structures and training processes. A preliminary implementation called the CS-DIP algorithm is available in Veen et al. [34]. Instead of using a pre-trained generator and optimizing the latent z-space, the author proposed using an untrained generator and optimizing the generators weights while keeping the latent z-space fixed, which can be expressed as:

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} ||\boldsymbol{g} - \boldsymbol{A}G(\boldsymbol{z}; \boldsymbol{w})||_2^2, \ \hat{\boldsymbol{c}} = G(\boldsymbol{z}^*; \boldsymbol{w}). \tag{3}$$

The proposed algorithm was tested by using Gaussian measurement and Fourier measurement (common in MRI applications). Although it requires a sufficient number of measurements and is not stable enough (the author proposed to run the algorithm multiple times and choose the best result), impressive results were produced. Another attempt with U-net is available in Baguer et al. [35].

2.4. Proposed method

Let $\boldsymbol{c}^* \in \mathbb{R}^{n^2}$ be the coefficients that we are trying to get, $\boldsymbol{A} \in \mathbb{R}^{m \times n^2}$ be the measurement matrix. Given $\boldsymbol{A}$ and the corresponding observations $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{c}^*$, we want to get a $\hat{\boldsymbol{c}}$ which is close to $\boldsymbol{c}^*$. A generator/decoder is a convolutional neural network which can be represented as $G(\boldsymbol{z}; \boldsymbol{w}) : \mathbb{R}^k \to \mathbb{R}^{n^2}$. It takes a latent vector $\boldsymbol{z} \in \mathbb{R}^k$ as the input and is parameterized by the weights $\boldsymbol{w}$. These models have shown an impressive ability to generate not only natural images but also CT images. In this paper, an untrained convolutional neural network will be used to reconstruct CT images of size 256×256.

2.4.1. Proposed framework

The basic idea of the proposed framework is to find a latent vector and its pairing weights for a randomly initialized generator, so that the generated image and the imaged object are consistent under the same measurements.

As described in Eq. (3), the CS-DIP algorithm will randomly initialize and then freeze the latent vector $\boldsymbol{z}$. Only the weights of the generator will be optimized during the reconstruction process. In our framework, both the latent vector $\boldsymbol{z}$ and the weights of the generator $\boldsymbol{w}$ will be optimized. We believe that by making both $\boldsymbol{z}$ and $\boldsymbol{w}$ trainable, we can enhance the models' ability to generate arbitrary images. Thus, the corresponding optimization problem can be written as:

$$(\boldsymbol{z}^*, \boldsymbol{w}^*) = \arg\min_{\boldsymbol{z},\boldsymbol{w}} ||\boldsymbol{g} - \boldsymbol{A}G(\boldsymbol{z}, \boldsymbol{w})||_2^2, \ \hat{\boldsymbol{c}} = G(\boldsymbol{z}^*, \boldsymbol{w}^*). \tag{4}$$

To solve Eq. (4) and other similar problems, one of the most important challenges is that the forward projection needs to be computed in each iteration, corresponding to the calculation of $\boldsymbol{A}G(\boldsymbol{z}, \boldsymbol{w})$. However, the size of matrix $\boldsymbol{A}$ in the X-ray CT reconstruction problem is too large to be calculated efficiently: its number of rows equals the number of measurements; its number of columns equals the number of pixels of the reconstructed image. We propose using the normal operator of matrix $\boldsymbol{A}$ to accelerate the calculation. Thus, Eq. (4) can be written as:

$$(\boldsymbol{z}^*, \boldsymbol{w}^*) = \arg\min_{\boldsymbol{z},\boldsymbol{w}} ||\boldsymbol{A}^{\mathrm{T}}\boldsymbol{g} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}G(\boldsymbol{z}, \boldsymbol{w})||_2^2, \ \hat{\boldsymbol{c}} = G(\boldsymbol{z}^*, \boldsymbol{w}^*), \tag{5}$$

where $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{g}$ is the back projection of sinogram signal $\boldsymbol{g}$, and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ indicates the combination of forward and back-projection. The advantage of Eq. (5) over Eq. (4) is that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ can be calculated efficiently. The calculation of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}G(\boldsymbol{z}, \boldsymbol{w})$ can be implemented with a freezing convolution kernel of size $(2n-1)^2$ in a neural network [36,37].

Although Eq. (5) is a non-convex problem, high-quality results are still obtainable using gradient-based optimizers. Ulyanov et al. [30] pointed out that generators/decoders such as DCGAN and autoencoder tend to produce smooth, natural images because of their convolutional structures. Veen et al. [34] further claimed that this property is also applicable in the general linear measurement process. As a result, a high-quality reconstructed image $\hat{\boldsymbol{c}} = G(\boldsymbol{z}^*, \boldsymbol{w}^*)$ can be obtained with a small number of measurements without pre-training.

It is worth mentioning that the proposed framework is more like a conventional IR method instead of a neural network related method, as it requires no training process and the result is updated

iteratively. For IR methods, linear optimization algorithms are used as solvers and priors: The algorithms guarantee that the objective function can be minimized. Also, the result generated by the solver will be selected as the best result when the system is under-determined. For the proposed framework, gradient-based optimizers and the structures of CNNs are used as solvers and priors. The ability of the neural network guarantees the minimization of the objective function, and the DIP helps generate the best result when the system is under-determined.

2.4.2. Regularizations and references

In the discussion above, only one data-mismatch term is included in the objective function. In practice, we would like to account for more factors. Thus, Eq. (5) can be rewritten as:

$$(\boldsymbol{z}^*, \boldsymbol{w}^*) = \arg\min_{\boldsymbol{z},\boldsymbol{w}} ||\boldsymbol{A}^{\mathrm{T}}\boldsymbol{g} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}G(\boldsymbol{z},\boldsymbol{w})||_2^2 + \lambda R(G(\boldsymbol{z},\boldsymbol{w})),$$
$$\hat{\boldsymbol{c}} = G(\boldsymbol{z}^*, \boldsymbol{w}^*), \tag{6}$$

where $R(\cdot)$ is a penalty term with a weighted parameter λ. Compared with the conventional optimization framework, neural network based framework is similar to the Superiorization Method (SM) [38], where a proximity function is explicitly designed and minimized in each iteration to steer the algorithm to a solution that not only minimizes the data-mismatch term but also fits the regularization. In our proposed framework, the penalty term does not need to be explicitly designed as all the differentiable functions are usable. Such a property greatly helps the reconstruction process, especially when the measurements are of low quality. Although many other complex functions or even other neural networks [3,29] can be used as regularizations, in this paper, we shall focus on the following three regularizations:

Total variation (TV) It is one of the most popular regularizations and has been proven useful in recovering piece-wise smooth images and denoising. The TV of an image is defined as the sum of image gradients [9]:

$$\mathrm{TV}(f) = ||\nabla f||_1 = \sum_{i,j} \sqrt{(\partial_x f_{i,j})^2 + (\partial_y f_{i,j})^2}, \tag{7}$$

where $||\cdot||_1$ indicates the l_1 norm. Thus, TV counts the summation of image gradient magnitude. In practice, $\partial_x f_{i,j}$ and $\partial_y f_{i,j}$ are approximated by difference operators, for example, $\partial_x f_{i,j} \approx f_{i,j} - f_{i-1,j}$ and $\partial_y f_{i,j} \approx f_{i,j} - f_{i,j-1}$. TV regularization and its variants such as anisotropic total variation (ATV) [10], adaptive-weighted total variation (AwTV) [39] and anisotropic relative total variation (ARTV) [11] are widely used in low-dose, sparse-view and limited-angle CT reconstructions. In this paper, for simplicity, we shall only consider the TV regularization, but its variants can also be implemented in our framework.

Indication mask In most cases, the imaged object is located at the center of the imaging area, and the non-zero area is less than 50% of the total area. It is evident that the reconstruction performance can be further improved if such property can be utilized as a prior. We propose using a binary mask to roughly indicate the non-zero area of the ground truth. It can be used to reduce the number of unknown pixels and eliminate the artifact caused by sparse-measurement. Also, such a mask is easy to obtain.

Reference images Reference images such as templates, images from adjacent slices and other similar objects can also contribute to the reconstruction. They can help the untrained network to find an intermediate result that is close to the ground truth. If multiple reference images are available, the soft-max function can be used to normalize the losses, so that the reconstructed image can be guided to the most similar reference image. It is worth mentioning that reference images can also be the reconstruction results of FBP, conventional IR algorithms, or even other pre-trained models. In that case, our proposed framework can be regarded as a post-processing method.

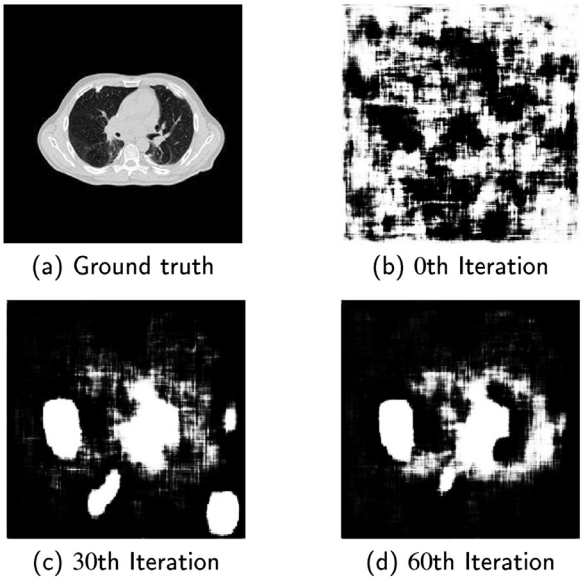

Fig. 2. The reconstructed images at the early stages of the reconstruction process. (a) Ground truth image; (b), (c) and (d) reconstructed images after 0, 30, and 60 iterations.

2.4.3. Reconstruction strategies

One of the biggest challenges faced by the current untrained neural network based reconstruction methods is achieving stability. The CS-DIP algorithm proposed by Veen et al. [34] has a high probability to generate extremely abnormal or even entirely black output, especially when the number of views is small. We propose using the following approaches to tackle this problem.

The first approach deals with the dying ReLU [40] problem. In the DIP related methods, all the models' weights are randomly initialized and then updated to minimize an objective function (loss function). This is equivalent to training a neural network with only one input-output pair. As a result, ReLU neurons of the neural network may become inactive at the early stage of the reconstruction process and cannot be reactivated. We propose using leaky ReLU [41] to solve this problem. It has a small slope for negative values, so that the dying ReLU problem can be solved without introducing any extra trainable parameters.

The second approach is about regularization. A unique image that minimizes the data mismatch term can be found when the number of measurements is sufficient. Adding the regularization sacrifices the data fidelity for the image regularity. However, for incomplete data where multiple images have equivalent data fidelity, regularization takes on the additional role of selecting the most probable result under the same data fidelity. For an untrained neural network whose parameters are initialized randomly, regularization may steer the reconstruction process into a local minimum. In fact, one of the reasons that CS-DIP generates output containing entirely black or white blocks is the improper using of TV. An example is shown in Fig. 2, where the CS-DIP algorithm is used to reconstruct an image (Fig. 2(a)). At the start of the reconstruction process, the output quickly converges to a shape where several non-zero blocks cluster at the center of the image (Fig. 2(c) and (d)). At this stage, minimizing the TV loss can easily generate some totally white or black areas. This will not increase the loss of data mismatch since the loss is so unoptimized that it can be reduced even if the weights are updated in a direction away from the global minimum. Such a problem may cause the algorithm to get stuck in a local minimum. Furthermore, zero pixel intensity always implies that a ReLU neuron is deactivated and cannot be reactivated. As a result, high-quality reconstruction results cannot be

obtained. To solve this problem, we propose to dynamically set the weight of the regularization term λ.

At the start of the training process, the weights corresponding to regularizations such as TV are set to 0 to avoid the local minimum and dying ReLU problems; the weights corresponding to reference images can be set to 1 to guide the model closer to the global minimum. Then, during the reconstruction process, the weights of regularizations increase while the weights of reference images decrease. In the final stage, the weights of regularizations are set to a proper value to achieve a good trade-off between data mismatch and image regularity. Artifacts caused by sparse-view or limited-angle projection can also be minimized in this stage by utilizing different kinds of regularization. The weights of reference images can be set to zero to avoid potential interference.

Another potential method is to use reference images to pre-train the model. Note that the goal of the training here is to better initialize the model so that the initial output is closer to the ground truth. Thus, no more than several images are needed. Also, since the model is still being optimized during the reconstruction process, incorrect information learned from reference images can be removed.

3. Results

In this section, we will compare our method with the state-of-the-art untrained model, CS-DIP [34], the well-known pre-trained models ELL50 and MED50 [16], conventional reconstruction methods such as Lasso in DCT basis [42] and Daubechies wavelet basis [43], as well as TVAL3 [44,45]. Shepp–Logan phantom [46], LIDC-IDRI [28] (the lung image database consortium image collection) dataset, and random ellipses dataset [16] are used in our experiments. All computations were done on one PC with an i7-8700K CPU, 32 GB of RAM, and an NVIDIA GeForce RTX 2080 GPU using Python.

3.1. Comparing with CS-DIP and conventional methods

3.1.1. Reconstruction performance

We first test the reconstruction performance of the proposed framework. To make a fair comparison, we use a generator network whose structure is the same as the CS-DIP algorithm (Fig. 3). The differences are the objective functions (Eqs. (5) and (3)), reconstruction strategies (described in Section 2.4.3), and the use of leaky ReLU. RMSProp with 0.9 momentum and 0 weight decay is used as the optimizer. The learning rate is 10^{-2} for cold start reconstruction and 10^{-3} for warm start reconstruction; both learning rates decrease by a factor of 0.8 per 500 iterations. Two of the most important properties are measured, and the results are as follows:

Stability To test the stability of reconstruction, we rerun the reconstruction process 1000 times and count the number of abnormal outputs (e.g. entirely black). It turns out that the average error rate of the CS-DIP with the number of views between 5 and 50 is 21%, while ours is 3.3%. Furthermore, the proposed framework can achieve a 0% error rate if a reference image is used. This result shows that the proposed framework is more stable. It is worth mentioning that abnormal outputs will be removed manually and won't be taken into account in the following experiments.

Convergence The cold start convergent performance is shown in Fig. 4(a), where both networks are untrained; the warm start convergent performance is shown in Fig. 4(b), where an adjacent slice of the reconstructed slice is used as a reference image to pre-train the networks. Our method shows a consistent improvement in both cases. This indicates that the performance of the untrained network can be improved by making the latent vector trainable.

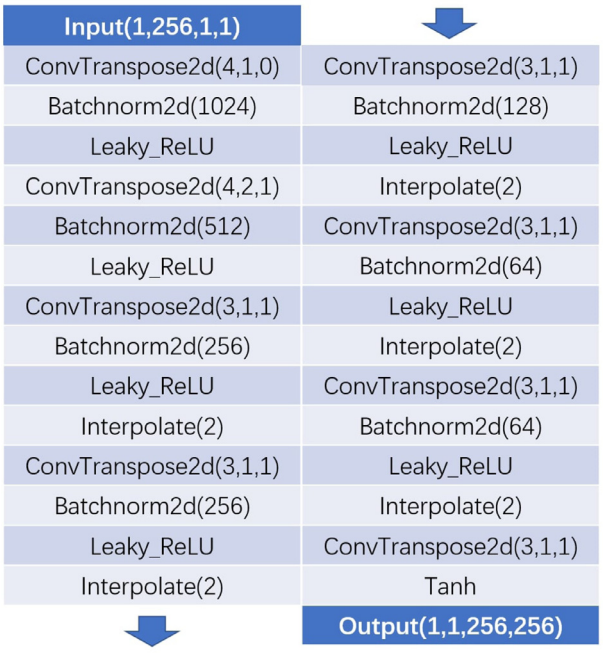

Fig. 3. The network structure.

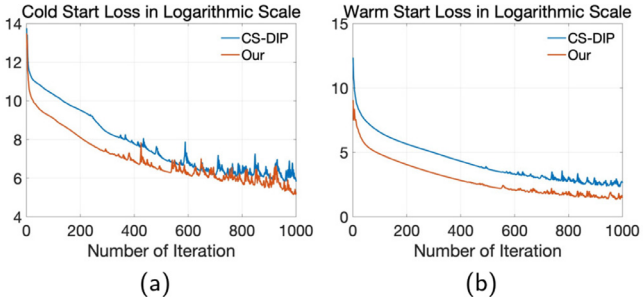

Fig. 4. The reconstruction loss in logarithmic scale. (a) Cold start, the generator network is completely untrained; (b) warm start, the generator network is pre-trained by one reference image.

3.1.2. Sparse-view reconstruction

Now the reconstruction accuracy of the proposed framework is tested. Shepp–Logan phantom and images from LIDC-IDRI are used as the ground truth, and all the images are of size 256×256. In all the following experiments, the sinogram sampling step is equal to the pixel size, and each projection has 513 sampling points so that projections with different angles can be sampled completely. The weight (λ) of the TV regularization is set to 0 at the beginning and linearly increases to 10^{-2} at the final stage. The weight of reference images, if applicable, is set to $\frac{1}{1+e^{(\frac{n}{n_s}-n_c)}}$, where n indicates the current number of iterations. It is a sigmoid function centered at $n_c = 5$ and stretched by the factor $n_s = 1000$. As a result, the weight is close to 1 at the early stage of the reconstruction process ($\frac{n}{n_s} \ll n_c$), and will decrease to 0 at the end ($\frac{n}{n_s} \gg n_c$) to avoid interference. It is worth mentioning that all these hyperparameters can be adjusted on-demand for each specific reconstruction problem since the proposed framework acts as a conventional IR method and requires no training process. However, for simplicity, the hyperparameters are fixed in our experiments, and the implementation detail of the proposed algorithm can be described by Algorithm 2.

For sparse-view CT reconstruction, the projection angles uniformly distribute from 0° to 180°, and the number of projections

Algorithm 2 Proposed algorithm using an Untrained NN.

Input:measurement matrix $\boldsymbol{A}$, measurement $\boldsymbol{g}$, randomly initialized neural network $G(\boldsymbol{z}, \boldsymbol{w})$ (shown in Fig. 3), hyperparameters λ for penalty term.

1: **Repeat:**
2: $\boldsymbol{c} \leftarrow G(\boldsymbol{z}, \boldsymbol{w})$
3: $\boldsymbol{z}, \boldsymbol{w} \leftarrow$ update $\boldsymbol{z}$ and $\boldsymbol{w}$ by the loss function $||\boldsymbol{A}^{\mathrm{T}}\boldsymbol{g} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}G(\boldsymbol{z}, \boldsymbol{w})||_2^2 + \lambda R(G(\boldsymbol{z}, \boldsymbol{w}))$ using backpropagation algorithm.
4: update λ
5: **Until:** convergence, or a fixed number of iterations is reached.

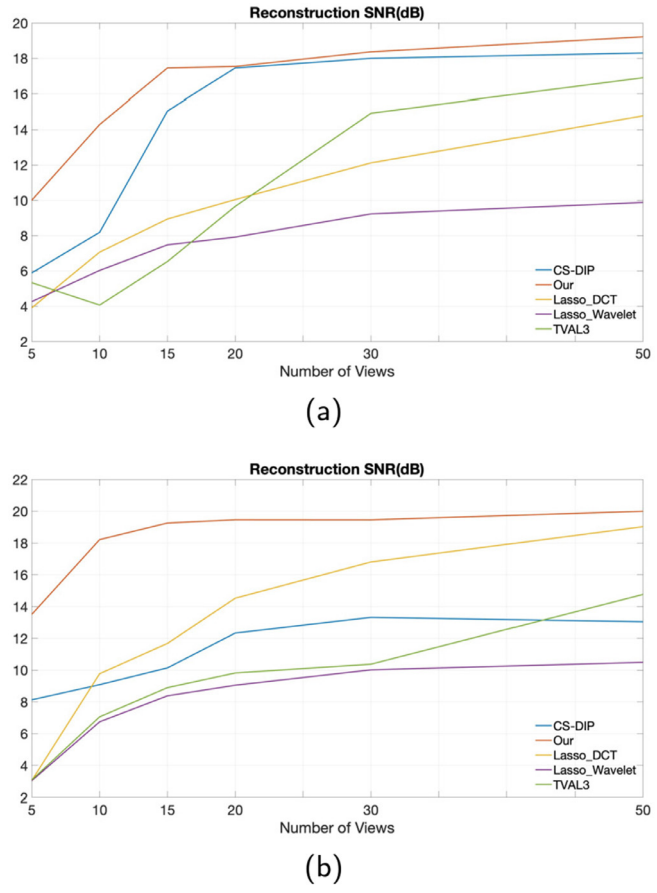

Fig. 5. SNR of sparse view reconstruction result. (a) Shepp–Logan phantom; (b) Real images from LIDC-IDRI.

goes from 5 to 50. The performance of the reconstruction is shown in Fig. 5, and some of the reconstruction results are shown in Figs. 6 and 7.

3.1.3. Limited-angle reconstruction

To test the proposed framework's performance on limited-angle reconstruction, we redo the experiment in Section 3.1.2 with the projections uniformly distributing from 0° to 90°. In this experiment, only the LIDC-IDRI dataset will be used, since the Shepp–Logan phantom has no reference image.

Fig. 8 shows the ground truth (Fig. 8(a) and (b)), the reference image (Fig. 8(c), and the indication mask (Fig. 8(d)). It is worth mentioning that the reference image is closer to Fig. 8(b) than Fig. 8(a), so a complete comparison can be shown. Four methods are compared in our experiment. The first two methods are the CS-DIP method and the proposed framework with only TV regularization. The third and fourth methods are the proposed frameworks with extra customized regularizations: the third calculating the l_2 distance between the reconstructed image and the reference image, and the fourth using an indication mask. The results are shown in Fig. 9. Previous conventional methods with TV regularization and its variants are not included since they can only generate acceptable reconstruction results when the angular range and the number of projections are large (e.g. 140 projections uniformly distributed from 15° to 155° in Chen et al. [47]).

3.2. Comparing with ELL50 and MED50

In this section, we compare our framework with the well-known pre-trained models ELL50 and MED50 [16], which correspond to the same network trained by two different datasets (the random ellipses dataset and real CT images). ELL50 and MED50 take FBP images as input and generate reconstructed images directly. To make a fair comparison, our proposed framework takes FBP images as references. We also use the reconstructed images produced by ELL50 and MED50 as references to see if the proposed framework can be used as a post-processing method to further improve the reconstruction quality. It is worth mentioning that the system is determined when the number of views is close to 100 and highly under-determined when the number of views is smaller than 50. The performances on the random ellipses dataset and LIDC-IDRI dataset are shown in Figs. 10 and 11 respectively. Some of the reconstruction results of the LIDC-IDRI dataset are shown in Fig. 12.

3.3. Noise in the sinogram

In this section, we repeat the experiments in Section 3.2 by using the sinogram data polluted with Poisson distributed noise to test the proposed framework's performance under low-dose conditions. The average number of X-ray photons received by the ith detector can be expressed as:

$$E_i = I_0 e^{[\mathcal{P}f]_i}, \tag{8}$$

where $I_0 > 0$ is the blank measurement ($[\mathcal{P}f]_i = 0$). It is worth mentioning that the sinogram data in our experiment is simulated by the Radon transform instead of obtained from a real instrument. Thus, I_0 here is a parameter for relative measurement.

The experiment results on real images and phantoms are shown in Fig. 13, where the results of FBP are used as a baseline to help understand the effect of the noise.

4. Discussion

4.1. Convergence and the selection of hyperparameters

Without considering extra regularizations, the global minimum can be achieved by conventional linear optimization algorithms. However, such convergence doesn't reach optimal reconstruction due to the under-determined system and noise. Thus, regularizations are introduced to the loss function despite making convergence more difficult and adding extra hyperparameters (λ).

According to the universal approximation theorem [48], deep neural networks can be used to approximate any functions. Thus, both pre-trained models and the proposed framework use neural networks to guarantee convergence. However, pre-trained models are not optimized for inference images, since inference image is excluded from the training dataset; neither are they optimized for every single training image, since the model is optimized for all training images on average. On the other hand, the proposed method's neural network is optimized for a single inference image, so the aforementioned problems can be avoided. As a result, the convergence performance of the proposed method is at least not

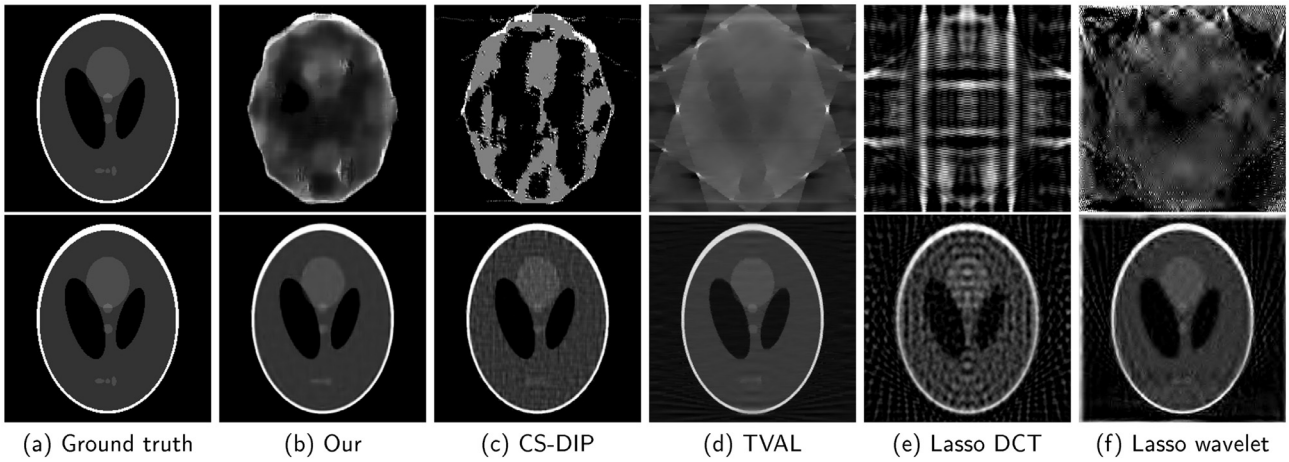

Fig. 6. The reconstructed results of Shepp–Logan phantom for different methods under sparse-view conditions. The first and second rows show the results generated from 5 and 30 projections respectively. These projections distribute uniformly from 0° to 180°.

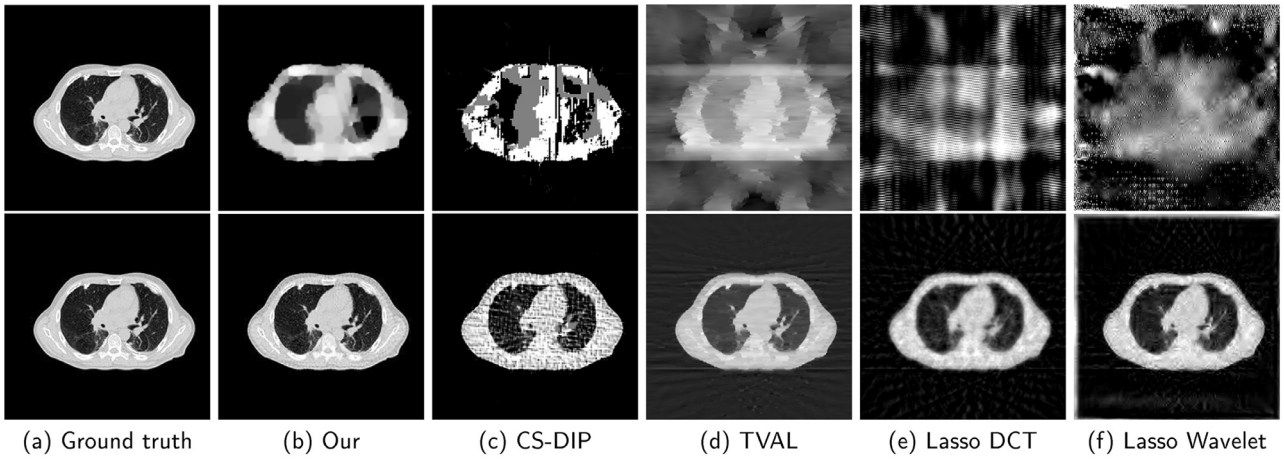

Fig. 7. The reconstructed results of an image from LIDC-IDRI dataset for different methods under sparse-view conditions. The first and second rows show the results generated from 5 and 30 projections respectively. These projections distribute uniformly from 0° to 180°.

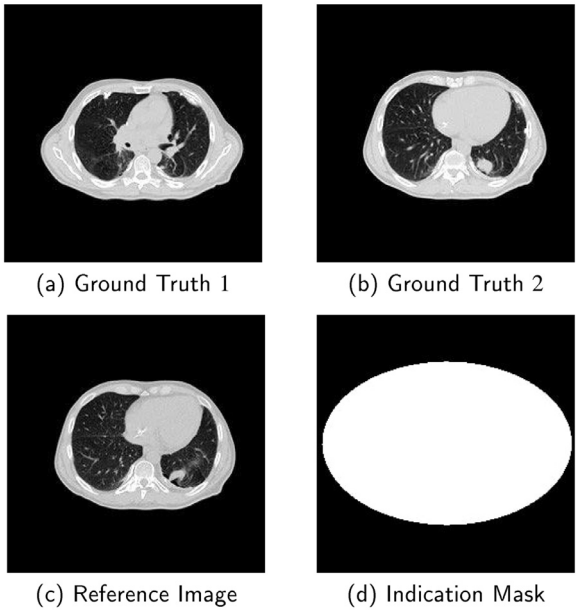

Fig. 8. Ground truth images, reference image, and indication mask.

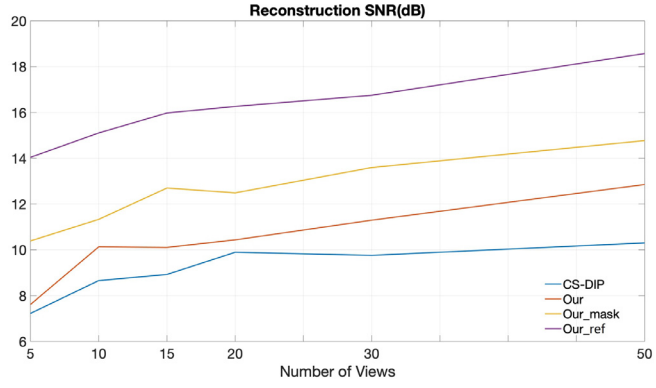

Fig. 9. SNR of limited-angle reconstruction. CS-DIP: the method proposed by Veen et al. [34]; Our: The proposed framework; Our_mask: The proposed framework with regularization using indication mask Fig. 8(d); Our_ref: The proposed framework with regularization using reference image Fig. 8(c).

worse than that of the current pre-trained neural network related methods.

The selection of the hyperparameters (λ) is another problem for all the reconstruction algorithms using extra regularizations. Ideally, we can adjust such hyperparameters and their corresponding

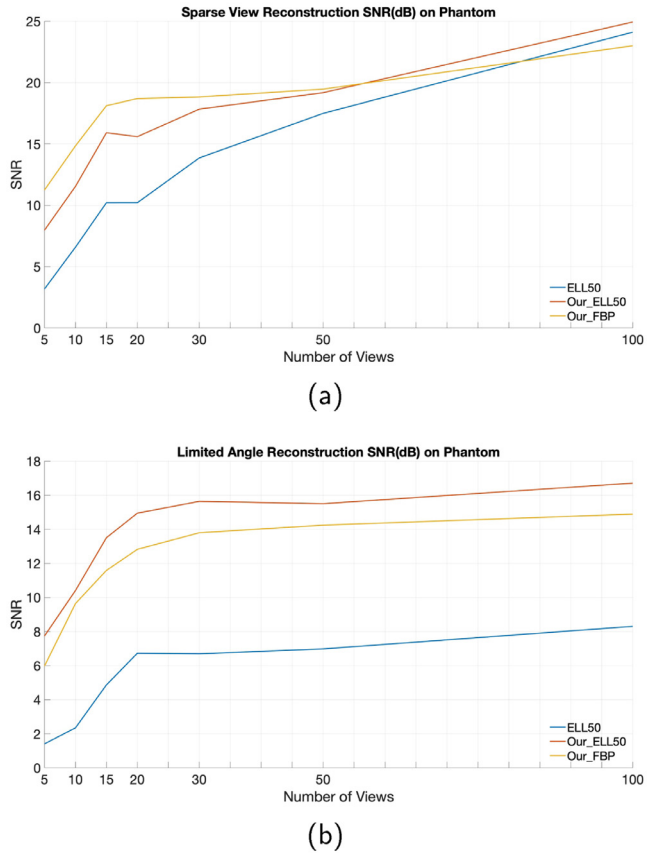

Fig. 10. SNR of reconstruction result under sparse-view and limited-angle ($0 \sim \frac{\pi}{2}$) conditions for phantom. Our_FBP indicates the result of using filtered back projection images as references in our proposed framework; Our_ELL50 indicates the result of using the images reconstructed by ELL50 as references. All the algorithms are tested on the random ellipses dataset.

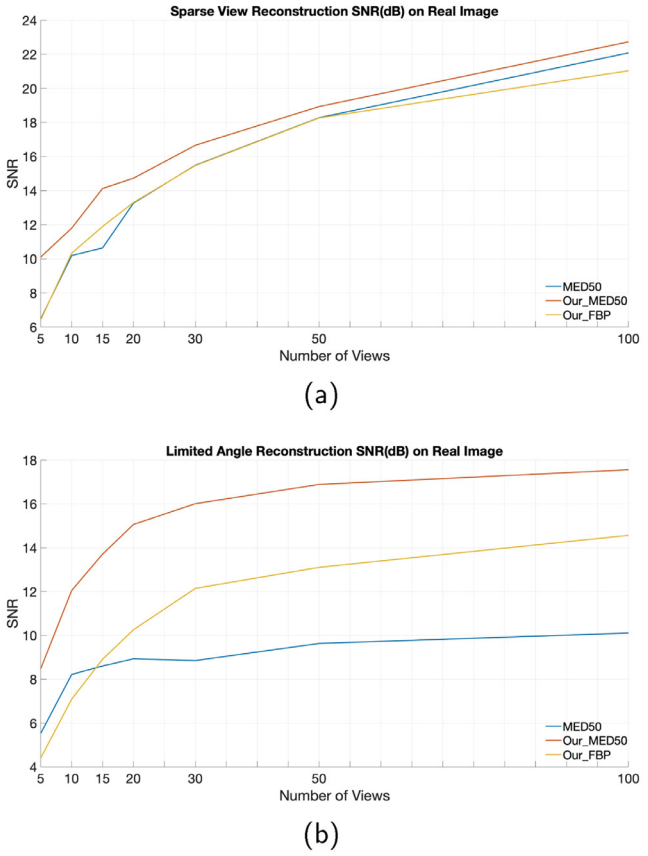

Fig. 11. SNR of reconstruction results under sparse-view and limited-angle ($0 \sim \frac{\pi}{2}$) conditions for real CT images. Our_FBP indicates the result of using filtered back projection images as references in our framework; Our_MED50 indicates the result of using the images reconstructed by MED50 as references. All the algorithms are tested on the LIDC-IDRI dataset.

regularizations to adapt to different conditions (e.g. different noise levels, number of views, angular ranges, imaged objects). Unlike other pre-trained models, whose hyperparameters and corresponding regularizations cannot be changed after the training process, ours can be adjusted on-demand as the proposed framework requires no training process.

4.2. Comparing with CS-DIP and conventional methods

As discussed at the end of Section 2.4.1, the proposed framework is in fact an IR method. In that case, a fair comparison should be among the proposed framework and other conventional MBIR methods. As shown in Fig. 5, the proposed framework shows a consistent improvement over all the other methods, especially in reconstructing real images. Furthermore, our method requires fewer views than others. Figs. 6 and 7 show the reconstruction results of the Shepp–Logan phantom and an image from LIDC-IDRI dataset under sparse-view conditions. From the first row of these two figures (reconstruction results from 5 projections), it is evident that both the proposed framework and CS-DIP method generate more reasonable results than the other methods. This clearly indicates the effectiveness of DIP. Also, it is obvious that the improper use of ReLU and TV regularization makes the reconstruction results of CS-DIP too piece-wise constant (the TV of the reconstruction results are too small) and thus downgrades the performance. The same problem can also be found in the second row of these two figures (reconstruction results from 30 projections), where both the proposed framework and CS-DIP method correctly capture the shape of the reconstructed image, but the CS-DIP algo-

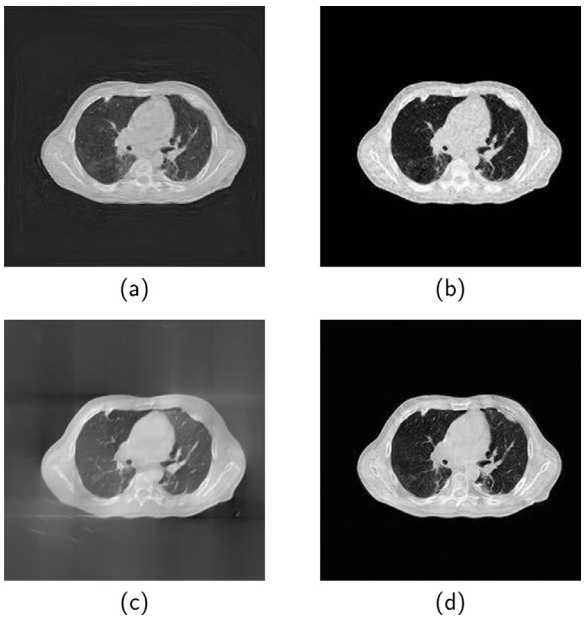

Fig. 12. The sparse-view (first row, 30 projections distributed uniformly from 0° to 180°) and limited-angle (second row, 90 projections distributed uniformly from 0° to 90°) CT reconstruction results of MED50 (first column) and our proposed framework using FBP as the reference (second column). The ground truth is shown in Fig. 8(a).

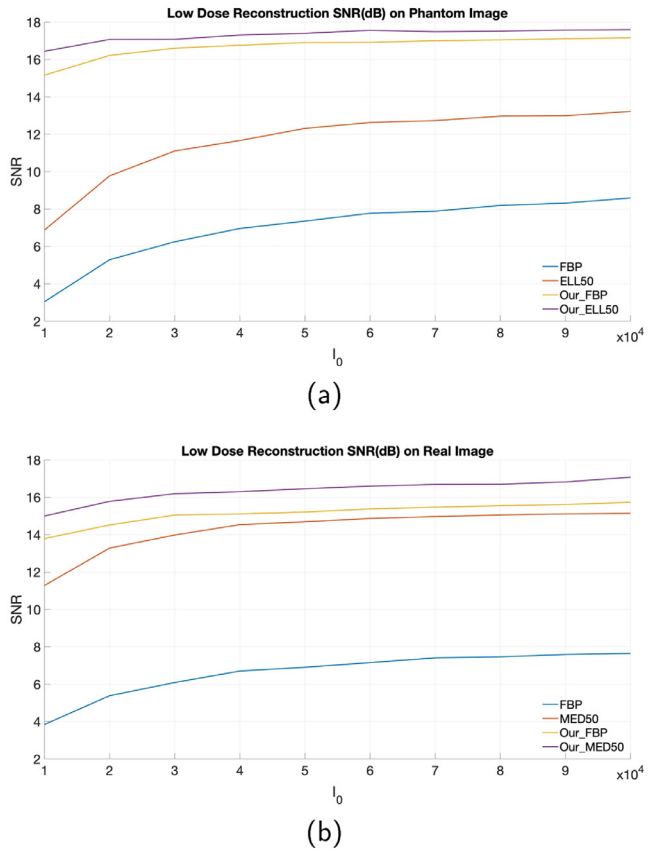

Fig. 13. Effect of quantum noise modeled by Poisson distribution in the sinogram on reconstruction SNR. The first plot shows results on random ellipses phantom and the second plot shows results on images from LIDC-IDRI dataset.

rithm gets stuck in a local minimum and cannot generate a high-quality result.

From Fig. 9, it is evident that our proposed framework shows consistent improvement over the CS-DIP method under the limited-angle condition. Customized regularizations do help the reconstruction process a lot, even an indication mask can provide an improvement of 5 dB. Furthermore, if the similarity between the reference image and the imaged object is relatively high, accurate reconstruction can be obtained with extremely sparse measurements.

It is worth mentioning that the size of the measurement matrix $\boldsymbol{A}$ is 513 m $\times$ 256^2, which is too large to be stored. In that case, the current untrained reconstruction algorithm based on Eq. (3) has to use complicated algorithms such as De Man and Basu [49], Long et al. [50], Ha and Mueller [51] to calculate $\boldsymbol{A}G(\boldsymbol{w};\boldsymbol{z})$. However, our algorithm based on Eq. (5) only needs a fixed convolution kernel of size 511 $\times$ 511 to calculate $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}G(\boldsymbol{z},\boldsymbol{w})$, which can be handled efficiently with a GPU.

4.3. Comparing with ELL50 and MED50

Although the proposed framework is in fact an IR algorithm, comparisons with pre-trained models MED50 and ELL50 are made for a complete analysis. There is no doubt that other pre-trained models with more complicated network structures and loss functions have better performance than that of MED50 and ELL50, but our goal is not to outperform all the pre-trained models. A fair comparison should focus on the structural properties as well as the difference caused by the training process, instead of the complexity of the networks. Therefore, MED50 and ELL50 are selected since they have a similar network complexity to the proposed network.

The reason for choosing MED50 and ELL50 is that the proposed framework has a similar network complexity, and

Fig. 10 (a) shows that our framework using FBP images as reference (Our_FBP, yellow line) improves the reconstruction results significantly under the sparse-view condition (number of views $\leq$ 50). It is worth mentioning that the proposed framework using the reconstruction result of ELL50 as reference (Our_ELL50, red line) outperforms the original ELL50 (blue line). This indicates that when being used as a post-processing method, our untrained network has the capability to utilize only the correct information from the reference images to further improve the reconstruction quality.

In the limited-angle reconstruction problem (Fig. 10(b)), where the system is always under-determined, the proposed framework using FBP images as reference (Our_FBP, yellow line) outperforms the pre-trained ELL50 model (blue line) by about 8 dB. This indicates that high-quality results can be achieved by our untrained network directly. Furthermore, our framework can achieve higher reconstruction accuracy by using the reconstructed image as a reference (Our_ELL50, red line), where the improvement is about 10 dB. The comparison with MED50 on real CT images (Fig. 11) shows the same trend.

It is worth mentioning that the random ellipses dataset is relatively easier for the proposed untrained network than the pre-trained network. However, it is the opposite for the LIDC-IDRI dataset. The reasons are:

1. Compared with the real CT images, ellipses are relatively easier for the untrained model to generate.
2. In the LIDC-IDRI dataset, one of the most obvious features is that the imaged objects are always at the center of the images, which can be utilized by a pre-trained model to improve its reconstruction accuracy easily.

Those may be the reasons for the slight differences between Figs. 10 and 11.

Fig. 12 shows some of the reconstruction results of MED50 and the proposed framework under sparse-view and limited-angle conditions. The system is highly under-determined (30 projections for sparse-view, and 90° angular range for limited-angle), so an exact reconstruction may be unobtainable. However, it is evident that our proposed method still generates high quality results (Fig. 12(b) and (d)). Comparing to others, the results generated by the proposed method have a totally black background and much fewer artifacts. There are two main reasons for such a huge improvement:

1. The deep image prior is much more powerful than a pre-trained neural network with a similar network structure under sparse-measurement conditions.
2. Although more than 2000 images from the LIDC-IDRI dataset are used for training the pre-trained model, the majority of the images are different from the inference images and may even interfere with the training process since they may correspond to different cross-sections. This also implies that instead of learning the correct reconstruction method, pre-trained models actually generate results from similar training images directly. As a result, a much larger training dataset is necessary for a well-trained model, which is impractical in the field of medical imaging.

4.4. Noise in the sinogram

Fig. 13 shows that the noise in the sinogram has little effect on the proposed framework, which indicates that our framework has better noise resistance performance than others when doing low-dose CT reconstructions.

4.5. Versatility

In this paper, for simplicity and fairness, only the 2D CT reconstruction in parallel-beam geometry of the size 256×256 is discussed. The reason to look into parallel-beam geometry is that the forward projection $\boldsymbol{A}$ and its normal operator $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ can be calculated exactly and efficiently under the parallel-beam geometry [37,52]. We use the size 256×256 because the forward projection $\boldsymbol{A}$ in the compared MBIR algorithms and CS-DIP algorithm will be too large to compute when the reconstruction resolution increases to 512×512. It is worth mentioning that the proposed framework can be used for multiple scenarios, since the key point of our proposed framework is to use an untrained model to do reconstruction directly, and the forward projections in different geometries are also well analyzed. The proposed framework is compatible with all regularizations used in both IR and neural network related methods. Furthermore, unlike other pre-trained models, these regularizations can be modified in the proposed framework on-demand (e.g. increasing the weight of total variation regularization when the noise level is high).

5. Conclusion

In this paper, we introduce a new neural network related framework for X-ray CT reconstruction. We show that better reconstruction results can be obtained without a training process by making all the neural network parameters trainable and using a new reconstruction strategy. We significantly reduce the computational cost in parallel-ray X-ray CT reconstruction by using the normal operator of the forward model. We also show that the proposed framework is compatible with multiple regularizations, and these regularizations can be adjusted on-demand for different scenarios. Furthermore, such a framework can also be applied to any other neural network based image reconstruction methods.

Most of our effort is focused on sparse-view and limited-angle CT reconstruction. We discover that the results can be improved significantly by using customized regularization, including but not limited to total variation and l_2 distance to reference images. It is worth mentioning that the incorrect information from reference images can be removed since the proposed framework guarantees the minimization of objective functions during the reconstruction process.

In our experiments, the proposed framework outperforms the conventional methods, the CS-DIP algorithm, and pre-trained models with similar network complexity. This improvement will be more evident under sparse-measurement conditions with a real object. The proposed framework can also act as a post-processing method to further improve the reconstruction results generated by these algorithms. Furthermore, our framework shows impressive noise resistance performance when solving the low-dose CT reconstruction problem.

With these results, we conclude that under sparse-view, limited-angle, and low-dose conditions, the proposed framework is better than all the methods discussed above, especially when there is insufficient training data to obtain a well-trained model.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Zhu, J.Z. Liu, S.F. Cauley, B.R. Rosen, M.S. Rosen, Image reconstruction by domain-transform manifold learning, Nature 555 (7697) (2018) 487–492.

[2] G. Shen, K. Dwivedi, K. Majima, T. Horikawa, Y. Kamitani, End-to-end deep image reconstruction from human brain activity, Front. Comput. Neurosci. 13 (2019) 21.

[3] S. Xie, H. Xu, H. Li, Artifact removal using GAN network for limited-angle CT reconstruction, in: 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA), IEEE, 2019, pp. 1–4.

[4] H. Zhang, L. Li, K. Qiao, L. Wang, B. Yan, L. Li, G. Hu, Image prediction for limited-angle tomography via deep learning with convolutional neural network, 2016, arXiv:1607.08707.

[5] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, G. Wang, Low-dose CT denoising with convolutional neural network, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), IEEE, 2017, pp. 143–146.

[6] G. Ma, C. Shen, X. Jia, Low dose CT reconstruction assisted by an image manifold prior, arXiv preprint arXiv:1810.12255 (2018).

[7] V. Antun, F. Renna, C. Poon, B. Adcock, A.C. Hansen, On instabilities of deep learning in image reconstruction and the potential costs of AI, Proc. Natl. Acad. Sci. 117 (48) (2020) 30088–30095.

[8] J. Radon, On the determination of functions from their integral values along certain manifolds, IEEE Trans. Med. Imaging 5 (4) (1986) 170–176.

[9] E.Y. Sidky, C.-M. Kao, X. Pan, Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT, J. X-ray Sci. Technol. 14 (2) (2006) 119–139.

[10] X. Jin, L. Li, Z. Chen, L. Zhang, Y. Xing, Anisotropic total variation for limited-angle CTreconstruction, in: IEEE Nuclear Science Symposuim & Medical Imaging Conference, IEEE, 2010, pp. 2232–2238.

[11] T. Wang, K. Nakamoto, H. Zhang, H. Liu, Reweighted anisotropic total variation minimization for limited-angle CT reconstruction, IEEE Trans. Nucl. Sci. 64 (10) (2017) 2742–2760.

[12] C. Gong, L. Zeng, Self-guided limited-angle computed tomography reconstruction based on anisotropic relative total variation, IEEE Access 8 (2020) 70465–70476.

[13] C. Bouman, K. Sauer, A generalized gaussian image model for edge-preserving map estimation, IEEE Trans. Image Process. 2 (3) (1993) 296–310.

[14] R. Zhang, C.A. Bouman, J.-B. Thibault, K.D. Sauer, Gaussian mixture Markov random field for image denoising and reconstruction, in: 2013 IEEE Global Conference on Signal and Information Processing, IEEE, 2013, pp. 1089–1092.

[15] S.J. Kisner, E. Haneda, C.A. Bouman, S. Skatter, M. Kourinny, S. Bedford, Model-based CT reconstruction from sparse views, in: Second International Conference on Image Formation in X-ray Computed Tomography, 2012, pp. 444–447.

[16] K.H. Jin, M.T. McCann, E. Froustey, M. Unser, Deep convolutional neural network for inverse problems in imaging, IEEE Trans. Image Process. 26 (9) (2017) 4509–4522.

[17] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-assisted intzervention, Springer, 2015, pp. 234–241.

[18] J. He, Y. Wang, J. Ma, Radon inversion via deep learning, IEEE Trans. Med. Imaging 39 (6) (2020) 2076–2087.

[19] E.Y. Sidky, X. Pan, Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization, Phys. Med. Biol. 53 (17) (2008) 4777.

[20] C. Shen, G. Ma, X. Jia, Low-dose CT reconstruction assisted by a global CTimage manifold prior, in: 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, vol. 11072, International Society for Optics and Photonics, 2019, p. 1107205.

[21] V. Shah, C. Hegde, Solving linear inverse problems using GAN priors: an algorithm with provable guarantees, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 4609–4613.

[22] H. Lee, J. Lee, H. Kim, B. Cho, S. Cho, Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction, IEEE Trans. Radiat. Plasma Med. Sci. 3 (2) (2018) 109–119.

[23] R. Anirudh, H. Kim, J.J. Thiagarajan, K.A. Mohan, K. Champley, T. Bremer, Lose the views: Limited angle CT reconstruction via implicit sinogram completion, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6343–6352.

[24] Z. Zhao, Y. Sun, P. Cong, Sparse-view CT reconstruction via generative adversarial networks, in: 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), IEEE, 2018, pp. 1–5.

[25] X. Yin, Q. Zhao, J. Liu, W. Yang, J. Yang, G. Quan, Y. Chen, H. Shu, L. Luo, J.-L. Coatrieux, Domain progressive 3D residual convolution network to improve low-dose CT imaging, IEEE Trans. Med. Imaging 38 (12) (2019) 2903–2913.

[26] D. Hu, Y. Zhang, J. Liu, C. Du, J. Zhang, S. Luo, G. Quan, Q. Liu, Y. Chen, L. Luo, Special: single-shot projection error correction integrated adversarial learning for limited-angle CT, IEEE Trans. Comput. Imaging 7 (2021) 734–746.

[27] Y. Zhang, D. Hu, Q. Zhao, G. Quan, J. Liu, Q. Liu, Y. Zhang, G. Coatrieux, Y. Chen, H. Yu, Clear: comprehensive learning enabled adversarial reconstruction for subtle structure enhanced low-dose CT imaging, IEEE Trans. Med. Imaging 40 (11) (2021) 3089–3101.

[28] S.G. Armato III, G. McLennan, L. Bidaut, M.F. McNitt-Gray, C.R. Meyer, A.P. Reeves, B. Zhao, D.R. Aberle, C.I. Henschke, E.A. Hoffman, et al., The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans, Med. Phys. 38 (2) (2011) 915–931.

[29] R. Anirudh, H. Kim, J.J. Thiagarajan, K.A. Mohan, K.M. Champley, Improving limited angle CT reconstruction with a robust GAN prior, arXiv preprint arXiv:1910.01634 (2019).

[30] D. Ulyanov, A. Vedaldi, V. Lempitsky, Deep image prior, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9446–9454.
[31] P. Chakrabarty, S. Maji, The spectral bias of the deep image prior, arXiv preprint arXiv:1912.08905 (2019).
[32] R. Heckel, M. Soltanolkotabi, Denoising and regularization via exploiting the structural bias of convolutional generators, arXiv preprint arXiv:1910.14634 (2019).
[33] P. Bojanowski, A. Joulin, D. Lopez-Paz, A. Szlam, Optimizing the latent space of generative networks, arXiv preprint arXiv:1707.05776 (2017).
[34] D.V. Veen, A. Jalal, M. Soltanolkotabi, E. Price, S. Vishwanath, A.G. Dimakis, Compressed sensing with deep image prior and learned regularization, 2018. arXiv:1806.06438.
[35] D.O. Baguer, J. Leuschner, M. Schmidt, Computed tomography reconstruction using deep image prior and learned reconstruction methods, Inverse Probl. 36 (9) (2020) 094004.
[36] M.T. McCann, M. Unser, High-quality parallel-ray X-ray CT back projection using optimized interpolation, IEEE Trans. Image Process. 26 (10) (2017) 4639–4647.
[37] Z. Shu, A. Entezari, Gram filtering and sinogram interpolation for pixel-basis in parallel-beam X-ray CT reconstruction, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 624–628.
[38] Y. Censor, R. Davidi, G.T. Herman, R.W. Schulte, L. Tetruashvili, Projected subgradient minimization versus superiorization, J. Optim. Theory Appl. 160 (3) (2014) 730–747, doi:10.1007/s10957-013-0408-3.
[39] Y. Liu, J. Ma, Y. Fan, Z. Liang, Adaptive-weighted total variation minimization for sparse data toward low-dose X-ray computed tomography image reconstruction, Phys. Med. Biol. 57 (23) (2012) 7923.
[40] L. Lu, Y. Shin, Y. Su, G.E. Karniadakis, Dying ReLU and initialization: theory and numerical examples, arXiv preprint arXiv:1903.06733 (2019).
[41] B. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectified activations in convolutional network, arXiv preprint arXiv:1505.00853 (2015).
[42] N. Ahmed, T. Natarajan, K.R. Rao, Discrete cosine transform, IEEE Trans. Comput. 100 (1) (1974) 90–93.
[43] I. Daubechles, Orthonormal bases of compactly supported wavelets, in: Fundamental Papers in Wavelet Theory, Princeton University Press, 2009, pp. 564–652.
[44] C. Li, W. Yin, Y. Zhang, Users guide for TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms, CAAM Rep. 20 (46–47) (2009) 4.
[45] J. Zhang, S. Liu, R. Xiong, S. Ma, D. Zhao, Improved total variation based image compressive sensing recovery by nonlocal regularization, in: 2013 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 2013, pp. 2836–2839.
[46] L.A. Shepp, B.F. Logan, The fourier reconstruction of a head section, IEEE Trans. Nucl. Sci. 21 (3) (1974) 21–43.
[47] Z. Chen, X. Jin, L. Li, G. Wang, A limited-angle CT reconstruction method based on anisotropic TV minimization, Phys. Med. Biol. 58 (7) (2013) 2119.
[48] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Netw. 2 (5) (1989) 359–366.
[49] B. De Man, S. Basu, Distance-driven projection and backprojection in three dimensions, Phys. Med. Biol. 49 (11) (2004) 2463.
[50] Y. Long, J.A. Fessler, J.M. Balter, 3D forward and back-projection for X-ray CT using separable footprints, IEEE Trans. Med. Imaging 29 (11) (2010) 1839–1850.
[51] S. Ha, K. Mueller, A look-up table-based ray integration framework for 2-D/3-D forward and back projection in X-ray CT, IEEE Trans. Med. Imaging 37 (2) (2017) 361–371.
[52] Z. Shu, A. Entezari, Exact gram filtering and efficient backprojection for iterative CT reconstruction, Med. Phys. 49 (5) (2022) 3080–3092.