A Dataset-free Deep Learning Method for Low-Dose CT Image Reconstruction

Qiaoqiao Ding[1], Hui Ji[2], Yuhui Quan[3], and Xiaoqun Zhang[1,4]
[1] Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China
[2] Department of Mathematics, National University of Singapore, 119076, Singapore
[3] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
[4] School of Mathematical Sciences and MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, China
{dingqiaoqiao,xqzhang}@sjtu.edu.cn
matjh@nus.edu.sg
csyhquan@scut.edu.cn

August 29, 2022

Abstract

Low-dose CT (LDCT) imaging attracted a considerable interest due to its reduction of exposure to X-ray radiation. In recent years, supervised deep learning has been extensively studied for LDCT image reconstruction, which trains a neural network (DNN) over a dataset containing many pairs of normal-dose and low-dose images. However, the challenge on collecting many such pairs in the clinical setup limits the application of supervised-learning-based methods for LDCT image reconstruction in practice. Aiming at addressing the challenges raised by the collection of a training dataset, this paper proposed an unsupervised deep learning (DL) method for LDCT image reconstruction, which does not require any external training data. The proposed method is built on a re-parametrization technique for Bayesian inference via a DNN with random weights, combined with total variational (TV) regularization. The experiments show that the proposed method noticeably outperforms existing dataset-free image reconstruction methods.

keywords: X-ray CT, Low-Dose CT, Deep Neural Networks

1 Introduction

X-ray Computed Tomography (CT) has been widely applied in clinical imag-

ing, for its ability of providing high-resolution images of internal anatomical structures. High-quality CT images are very useful for prevention, diagnosis and treatments of human diseases. However, many studies indicated that excessive exposure to radiation from a X-ray CT scanner may be responsible for the increasing risk of getting cancer, thus there is the demand for reducing radiation dose when run CT scanning. There are two main techniques for radiation dose reduction: decreasing the radiation exposure time, *i.e.* decreasing the number of projection views [33], and lowering the X-ray tube current [38], *i.e.* LDCT. In comparison to normal dose CT (NDCT), the signal-to-noise ratio (SNR) of measurements in LDCT is much lower. As a result, the quality of images reconstructed using conventional methods is not satisfactory for LDCT. There are often noticeable streaky artifacts and random patterns appearing in reconstructed images.

The image reconstruction problem for LDCT can be formulated as solving a linear inverse problem

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{A}$ denotes the projection matrix of CT imaging, $\boldsymbol{y}$ denotes the available measurements, $\boldsymbol{x}$ denotes the image to be reconstructed, and $\boldsymbol{n}$ denotes the measurement noise which is often modeled by i.i.d. random variables. The inverse problem (1) from LDCT imaging is ill-posed. Certain regularization on $\boldsymbol{x}$ needs to be introduced to address solution ambiguity and to suppress noise magnification when solving (1).

In recent years, deep learning has been a very promising tool for developing effective image reconstruction methods for CT, including LDCT. The majority of existing deep learning-based methods are built upon supervised learning; see e.g. [1, 6, 7, 10, 11, 13, 17, 19, 25, 28, 42]. Supervised learning requires a dataset with many training samples, i.e., the pairs of a low-dose image (used as input) and a normal dose image (assumed to be ground-truth). The construction of such image pairs requires two-times scan as well as registration for alignment of every image/projection pair, which is costly and troublesome in practice. In addition, the number of real-world images for LDCT is also very limited. As a result, there is an increasing interest on the development of powerful unsupervised deep learning methods for LDCT imaging which work well in data-limited environments.

1.1 Related works

In existing literature, many methods, e.g., analytical filtering methods [14,22,26] have been proposed to improve image quality of LDCT imaging. Due to low SNR of measurement data, these methods, including the ones equipped with adaptive filtering [3] and bilateral filtering [30], are not capable of producing high-quality CT images. Over the past decades, iterative reconstruction algorithm is a popular approach adopted in LDCT imaging, which is derived by minimizing a cost function. The cost function usually is composed of a fidelity term determined by statistical characteristics of noise and a regularization term

induced by some pre-defined prior on the image. In the past, many regularizations have been proposed for LDCT image reconstruction, *e.g.*, total variation (TV) [34, 47], wavelet tight frame-based sparsity prior [23], nonlocal total variation [24], and low-rank based patch prior [5].

In past few years, deep learning has emerged as a prominent tool for developing powerful image reconstruction methods for LDCT imaging. Earlier work on DL-based LDCT image reconstruction used DL as a post-processing tool [6, 7, 25, 28], which trains a deep neural network (DNN) to denoise images reconstructed from some existing works. The denoising network is trained by using many pairs of images reconstructed from LDCT and the corresponding NDCT. Different network architectures have been exploited in these works, *e.g.*, convolution neural network (CNN) [7], encoder-decoder CNN [6], residual network [28], and U-Net [25]. As the artifacts in a reconstructed image often cannot be modeled as independent random noise, the performance gain brought by such a post-process is limited. A more effective approach is the so-called optimization unrolling scheme [1, 10, 13, 17] and plug-and-play [19, 41]. Such a scheme follows some iterative image reconstruction scheme derived from some regularization methods, and replaces the related regularization steps by a or multiple learnable/pre-trained denoising DNN. The main difference among these methods lies in which iterative scheme is used for unrolling and how to train the denoising network embedded inside the iterations.

Recently, the development of dataset-free DL methods has drawn a lot of attention for LDCT imaging. By using a generative adversarial network (GAN), the works [39, 40] conducted training on a dataset containing both low-dose and normal-dose images, but which are not paired. Inspired by recent works on unsupervised learning for generic image denoising, there are two approaches to extending these unsupervised learning approaches to inverse problems such as low-dose CT image reconstruction.

One approach is treating CT image reconstruction as a denoising process which post-processes the reconstructed images. In [18, 45], the Noise2Noise (N2N) [27], a denoising network trained over the image pairs with independent noise, is introduced to remove artifacts of the reconstructed LDCT images. In [21], the Noise2Self (N2S) [4], an unsupervised denoising network trained over noisy images, is extended to the tomography problem by modeling image artifacts as independent random noise. Such an approach suffers from the same performance issue as its supervised counterparts that rely on a denoising network for removing artifacts of the reconstructed images. The reason is that the artifacts of a reconstructed image are indeed highly correlated to the entries of the image, which cannot be well modeled by simple random variables, such as i.i.d. noise, assumed by the unsupervised denoising networks. As a result, the performance of these unsupervised methods, derived from those unsupervised denoising networks above, is not very competitive to the state-of-the-art supervised methods for LDCT imaging.

Another unsupervised approach is built on the so-called deep image prior (DIP) [35]. DIP is originally proposed for image denoising. It is empirically observed that when training a CNN to fit a noisy image, regular image struc-

tures appear before random patterns. Thus, one can train a denoising network on a noisy image by early stopping. In other words, early stopping can be an effective technique for regularizing a denoising network. DIP has been exploited in various medical imaging tasks, such as PET reconstruction [16,43], MRI [44], diffraction tomography [48], and compressed sensing [36]. In [2], DIP is combined with TV-based regularization for CT image reconstruction. While DIP is simple and effective for image denoising, there are issues regarding DIP-derived methods for CT imaging. The artifacts in a reconstructed images are not random noise. They are also regular patterns correlated with the image structures. As a result, the early stopping adopted in DIP cannot prevent the appearance of artifacts in the images reconstructed by DIP.

1.2 Our idea

In this paper, we present an unsupervised deep learning method for LDCT image reconstruction, without requiring any external training samples with ground truths. Such an unsupervised method certainly can see its great value in practice. The proposed method is built on the Bayesian inference where the prior distribution of an image is re-parametrized by a DNN with random weights.

Recall that in Bayesian inference, we have two representative Bayesian estimators. One is the maximum a posterior (MAP) estimator:

$$\boldsymbol{x}_{\text{MAP}} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y}), \tag{2}$$

and the other is the minimum mean squared error estimator (MMSE) estimator, or equivalently the conditional mean estimator:

$$\boldsymbol{x}_{\text{CM}} = \mathbb{E}_{(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x}|\boldsymbol{y}) = \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x}, \tag{3}$$

where $p(\boldsymbol{x}|\boldsymbol{y})$ denotes the posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{y}$. The key to both estimators is about deriving the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$ which models the data well. A common practice in Bayesian inference is to re-express $p(\boldsymbol{x}|\boldsymbol{y})$ by Bayesian rule:

$$p(\boldsymbol{x}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})/p(\boldsymbol{y}),$$

where the likelihood term $p(\boldsymbol{y}|\boldsymbol{x})$ can be expressed as

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2,$$

in the presence of i.i.d Gaussian white noise $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$. Then, the study of the estimators turns to defining a prior distribution $p(\boldsymbol{x})$ that accurately models statistical characteristics of images for reconstruction.

In traditional regularization methods, to be computationally tractable, the prior distribution $p(\boldsymbol{x})$ usually is modeled by mean-field approximation which

assumes the independence of all image pixels. For instance, the well-known TV-based regularization assumes

$$p(\boldsymbol{x}) = \prod_i q((\nabla x)_i),$$

where q is the density function of a Laplacian distribution: $q(z) \sim e^{-\frac{|z|}{\lambda}}$. There are two concerns in the assumptions of the prior distribution used in TV-based regularization. One is the mean-field assumption and the other is the Laplacian assumption.

The over-simplified mean-field model for the prior distribution $p(\boldsymbol{x})$ motivates us to study a different approach to modeling the prior distribution $p(\boldsymbol{x})$ or the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$. Inspired by the advance of optimization techniques for solving the optimization problems of DNN training, we propose to adopt a re-parametrization technique for Bayesian inference, which re-expresses the variable $\boldsymbol{x}$ by a DNN with random weights

$$\boldsymbol{x} = f(\boldsymbol{x}_0; \boldsymbol{\theta}),$$

where $\boldsymbol{x}_0$ is some initial seed and $\boldsymbol{\theta}$ are random variables. It can be seen that after re-parametrization, the prior distribution of $p(\boldsymbol{x})$ can be very complicated, even though the variable $\boldsymbol{\theta}$ is modeled by mean-field approximation.

After re-parametrization, the variables for inference now are random network weights $\boldsymbol{\theta}$. Again, the key for a Bayesian inference now is to define an appropriate posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ for $\boldsymbol{\theta}$. As in general, it is not computationally tractable in high dimension, we adopt variational approximation to approximate $p(\boldsymbol{\theta}|\boldsymbol{y})$ by a set of approximation distributions $q(\boldsymbol{\theta}|\boldsymbol{\mu})$ parametrized by $\boldsymbol{\mu}$. The optimal approximation with distribution parameters $\boldsymbol{\mu}^*$ is then estimated by minimizing the Kullback–Leibler (KL) divergence between two distributions. Once the approximation to posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ is obtained, we can utilize the Bayesian inference to estimate the image.

This paper is organized as follows. Section 2 describes the proposed method and algorithm, with the NN architecture and implementation details given in Section 2.4. Section 3 is devoted to the experimental evaluation and comparison to other methods. Section 4 concludes the paper.

2 Method

In this section, we give a detailed discussion on the proposed self-supervised method for LDCT reconstruction from noisy measurements, which is built on the DNN-based re-parametrization for Bayesian inference. Recall that CT reconstruction problems can be formulated as the following inverse problem: Given an observed image $\boldsymbol{y} \in \mathbb{R}^m$ corrupted according to forward model and noise, $\boldsymbol{n}$, find the unknown image $\boldsymbol{x} \in \mathbb{R}^n$ which satisfies the observation

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}. \tag{4}$$

Considering a DNN with random weights for the re-parametrization:

$$\boldsymbol{x} = f(\boldsymbol{x}_0; \boldsymbol{\theta}).$$

Then, the inference of $\boldsymbol{x}$ from noisy measurement $\boldsymbol{y}$ is now about inferring the network weights $\boldsymbol{\theta}$ from $\boldsymbol{y}$. In order to perform Bayesian inference for $\boldsymbol{\theta}$, the key is to derive the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$. As $p(\boldsymbol{\theta}|\boldsymbol{y})$ is in general computationally intractable, we propose to approximate it by the following set of distributions $q(\boldsymbol{\theta}|\boldsymbol{\mu})$ defined by

$$\boldsymbol{\theta} = \boldsymbol{\mu} \odot \boldsymbol{b}: \quad \theta_i = \mu_i \cdot b_i, 1 \leq i \leq N, \tag{5}$$

where μ_i denotes the distribution parameter of θ_i and $b_i \sim \mathbf{B}(p_i)$ follows a Bernoulli distribution with probability p_i. In other words, the probability density function of b_i is defined as

$$p(b_i) = p_i^{b_i}(1-p_i)^{1-b_i} \qquad b_i = \{0, 1\}. \tag{6}$$

In other words, the DNN with random weights used in this paper is the widely used the network with dropout. It is noted that the idea of using the network with dropout also has been exploited in S2S [8, 29, 32] for self-supervised image denoising and deconvolution.

In the next, we gives a detailed discussion on how to train the network by minimizing the KL divergence between $q(\boldsymbol{\theta}|\boldsymbol{\mu})$, and how to use the trained model for testing by using Monte-Carlo sampling.

2.1 Training

As we use $q(\boldsymbol{\theta}|\boldsymbol{\mu})$ to approximate $p(\boldsymbol{\theta}|\boldsymbol{y})$, the optimal approximation is estimated by minimizing the KL-divergence between $q(\boldsymbol{\theta}|\boldsymbol{\mu})$ and $p(\boldsymbol{\theta}|\boldsymbol{y})$:

$$\begin{aligned}
&\min_{\boldsymbol{\mu}} \mathrm{KL}(q(\boldsymbol{\theta}|\boldsymbol{\mu})||p(\boldsymbol{\theta}|\boldsymbol{y})) \\
=&\min_{\boldsymbol{u}} \mathbb{E}_{\boldsymbol{\theta}\sim q(\boldsymbol{\theta}|\boldsymbol{\mu})}[\log q(\boldsymbol{\theta}|\boldsymbol{\mu}) - \log p(\boldsymbol{\theta}|\boldsymbol{y})] \\
\propto&\min_{\boldsymbol{\mu}} \mathbb{E}_{\boldsymbol{\theta}\sim q(\boldsymbol{\theta}|\boldsymbol{\mu})}[\log q(\boldsymbol{\theta}|\boldsymbol{\mu}) - (\log p(\boldsymbol{y}|\boldsymbol{\theta}) + \log p(\boldsymbol{\theta}))] \\
=&\min_{\boldsymbol{\mu}} \mathrm{KL}(q(\boldsymbol{\theta}|\boldsymbol{\mu})||p(\boldsymbol{\theta})) - \mathbb{E}_{\boldsymbol{\theta}\sim q(\boldsymbol{\theta}|\boldsymbol{\mu})} \log p(\boldsymbol{y}|\boldsymbol{\theta}).
\end{aligned} \tag{7}$$

For the first term, suppose that $p(\boldsymbol{\theta})$ is a uniform distribution in a sufficient larger region Ω. Here, we abuse the notion $\frac{0}{0} = 1$. We have $q(\theta_i|\mu_i) = p_i^{\frac{\theta_i}{\mu_i}}(1-p_i)^{1-\frac{\theta_i}{\mu_i}}$, $\theta_i = \{0, \mu_i\}$ and $p(\theta_i) = 1/s_i$, where s_i is the length of the domain of

definition about θ_i. Then,

$$\begin{aligned} D_{KL}(q(\boldsymbol{\theta}|\boldsymbol{\mu})||p(\boldsymbol{\theta})) &= \sum_i D_{KL}(q(\theta_i|\mu_i)||p(\theta_i)), \\ &= \sum_i q(\theta_i|\mu_i) \log \frac{q(\theta_i|\mu_i)}{p(\theta_i)}, \\ &= \sum_i (1-p_i)\log(1-p_i) + p_i \log p_i + \log s_i. \end{aligned}$$

Finally, we obtain

$$D_{KL}(q(\boldsymbol{\theta}|\boldsymbol{\mu})||p(\boldsymbol{\theta})) = c_0, \quad \boldsymbol{\theta} \in \Omega, \tag{8}$$

where c_0 is a constant.

In the second term, suppose that the measurement noise $\boldsymbol{n}$ is Gaussian white noise such that $p(\boldsymbol{n}) \propto \prod_i \exp(\frac{-n_i^2}{2\tilde{\sigma}^2})$, we have

$$\log(p(\boldsymbol{y}|\boldsymbol{\theta})) \propto -\frac{1}{2\tilde{\sigma}^2}\|\boldsymbol{A}f(\boldsymbol{x}_0,\boldsymbol{\theta}) - \boldsymbol{y}\|_2^2.$$

Then, we have

$$\min_{\boldsymbol{\mu}} D_{KL}(q(\boldsymbol{\theta}|\boldsymbol{\mu})||p(\boldsymbol{\theta}|\boldsymbol{y})) \propto \min_{\boldsymbol{\mu}} \mathbb{E}_{\boldsymbol{\theta}\sim q(\boldsymbol{\theta}|\boldsymbol{\mu})}\|\boldsymbol{A}f(\boldsymbol{x}_0,\boldsymbol{\theta}) - \boldsymbol{y}\|_2^2. \tag{9}$$

It can be seen from (9) that the KL divergence only constrains the estimation in the range space of the projection matrix $\boldsymbol{A}$. To avoid possible overfitting, we introduce an additional regularization on the estimation, and we adopt the widely-used TV regularization to the loss function. Recall (5) and consider the definition of $q(\boldsymbol{\theta}|\boldsymbol{\mu})$ and $\boldsymbol{B}(\boldsymbol{p})$, we deduce that

$$\begin{aligned} &\min_{\boldsymbol{\mu}} \mathbb{E}_{\boldsymbol{\theta}\sim q(\boldsymbol{\theta}|\boldsymbol{\mu})}\|\boldsymbol{A}f(\boldsymbol{x}_0,\boldsymbol{\theta}) - \boldsymbol{y}\|_2^2 \\ &\quad = \min_{\boldsymbol{\mu}} \int \|\boldsymbol{A}f(\boldsymbol{x}_0,\boldsymbol{\theta}) - \boldsymbol{y}\|_2^2 \; q(\boldsymbol{\theta}|\boldsymbol{\mu}) d\boldsymbol{\theta} \\ &\quad \overset{d\boldsymbol{\theta}=\boldsymbol{\mu}\odot d\boldsymbol{b}}{=\!=\!=} \min_{\boldsymbol{\mu}} \int \|\boldsymbol{A}f(\boldsymbol{x}_0,\boldsymbol{\mu}\odot\boldsymbol{b}) - \boldsymbol{y}\|_2^2 \; \boldsymbol{B}(\boldsymbol{p}) d\boldsymbol{b} \\ &\quad = \min_{\boldsymbol{\mu}} \mathbb{E}_{\boldsymbol{b}\sim\boldsymbol{B}(\boldsymbol{p})}\|\boldsymbol{A}f(\boldsymbol{x}_0,\boldsymbol{\mu}\odot\boldsymbol{b}) - \boldsymbol{y}\|_2^2. \end{aligned}$$

The final loss function for training the network now is

$$\min_{\boldsymbol{\mu}} \mathbb{E}_{\boldsymbol{b}\sim\boldsymbol{B}(\boldsymbol{p})}\|\boldsymbol{A}f(\boldsymbol{x}_0,\boldsymbol{\mu}\odot\boldsymbol{b}) - \boldsymbol{y}\|_2^2 + \alpha\|\nabla f(\boldsymbol{x}_0,\boldsymbol{\mu}\odot\boldsymbol{b})\|_1, \tag{10}$$

where α is a pre-defined hyper-parameter. When training the network, the loss function is minimized by using MC dropout [15], *i.e.*, randomly dropping out nodes during the training with dropout rate $\boldsymbol{p}$.

2.2 Testing

Once the NN is trained via minimizing the loss function given in (10), we have an approximation to the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$, denoted by $q(\boldsymbol{\theta}|\boldsymbol{\mu}^*)$. In our approach, we estimate the image $\boldsymbol{x}$ using conditional mean estimator. Recall that given the measurement $\boldsymbol{y}$, its conditional mean estimator for $\boldsymbol{x}$ reads

$$\boldsymbol{x}_{\text{CM}} = \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x}.$$

By re-parametrization: $\boldsymbol{x} = f(\boldsymbol{x}_0; \boldsymbol{\theta})$, we have

$$\boldsymbol{x}_{\text{CM}} = \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x} = \int f(\boldsymbol{x}_0; \boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}.$$

By approximating $p(\boldsymbol{y}|\boldsymbol{\theta})$ using $q(\boldsymbol{\theta}|\boldsymbol{\mu}^*)$, we have an approximate conditional mean estimator of $\boldsymbol{x}$ given by

$$\boldsymbol{x}^*_{\text{CM}} = \int f(\boldsymbol{x}_0; \boldsymbol{\theta}|\boldsymbol{\mu}^*) q(\boldsymbol{\theta}|\boldsymbol{\mu}^*) d\boldsymbol{\theta}. \tag{11}$$

The integration above is calculated by using Monte Carlo (MC) integration in practice. That is, after the network is trained, we take K random samples of the networks with dropout:

$$f(\boldsymbol{x}_0; \boldsymbol{\theta}_k) = f(\boldsymbol{x}_0; \boldsymbol{\mu}^* \odot \boldsymbol{b}_k), \quad \boldsymbol{b}_k \sim \boldsymbol{B}(\boldsymbol{p}).$$

Then, the estimate is defined by taking the the average of these K samples:

$$\boldsymbol{x}^* = \frac{1}{K} \sum_{k=1}^{K} f(\boldsymbol{x}_0; \boldsymbol{\theta}_k) = \frac{1}{K} \sum_{k=1}^{K} f(\boldsymbol{x}_0; \boldsymbol{\mu}^* \odot \boldsymbol{b}_k). \tag{12}$$

2.3 Discussion

In Section 2.1 and Section 2.2, we present a DNN-based re-parametrization $\boldsymbol{x} = f(\boldsymbol{x}_0; \boldsymbol{\theta})$ for facilitating the Bayesian inference of LDCT image reconstruction. In the proposed approach, the corresponding posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ is approximated by a network with dropout $q(\boldsymbol{\theta}|\boldsymbol{\mu})$ via minimizing their KL divergence. After the network is trained with dropout. The network is sampled with dropout to have a MC-based approximation to the conditional mean estimator of $\boldsymbol{x}$.

In addition, as only the noisy measurement $\boldsymbol{y}$ is available which only measures the image $\boldsymbol{x}$ in the range space of $\boldsymbol{A}$, a TV-regularization is introduced in the loss function for regularizing the network to avoid possible overfitting. As a result, the loss function (10) is closely connected to the non-learning TV regularization method for solving inverse problems. Indeed, based on the loss function (10), the proposed method can be viewed as learning multiple solvers to the TV-regularization model, and each solver differentiate itself from others

by using different network architectures (by random dropout). From the perspective of ensemble learning, the proposed method can be also interpreted as an ensemble learning method that is built on TV-related regularization. It is likely that the artifacts from each instance of the solver to TV regularization have certain degree of independence. Then, the average of the results from these solvers will benefit such artifact independence to have an estimate with less artifacts. In short, the proposed method provides an efficient ensemble learning method for LDCT image via dropout.

2.4 Implementation details

2.4.1 NN architecture

To evaluate the effectiveness of the proposed method, we test it using an encoder-decoder with skip-connection as the backbone network, whose diagram is illustrated in Fig. 1 (a). In the diagram of the network, the notation D_i, U_i and S_i represent the downsampling, upsampling and skip-connection blocks in the NNs. In the decoder-encoder architecture, $c_u[i]$, $c_d[i]$, $c_s[i]$ correspond to the number of filters at depth i for the upsampling, downsampling, skip-connections respectively. The values $k_u[i]$, $k_d[i]$, $k_s[i]$ correspond to the respective kernel sizes. The values $p_u[i]$, $p_d[i]$, $p_s[i]$ are the drop probability of dropout for the upsampling, downsampling, skip-connections respectively. Note that there is no upsampling layer in U_1 and the NN structure is similar to a U-Net.

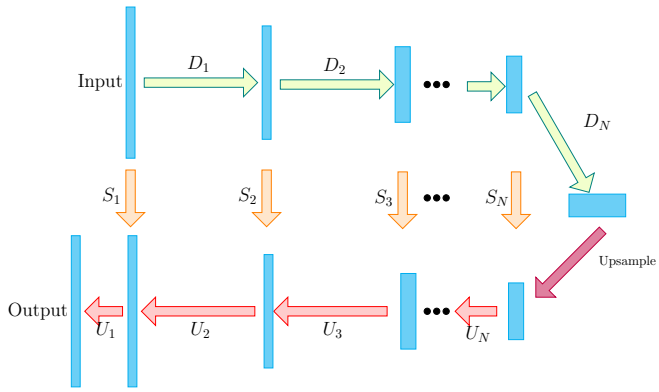
2.4.2 Implementation

For the implementation of the network, the number of layer N is set to 5. For the layers from $i = 1, \cdots, N$, the filter numbers are set as $c_d[i] = c_u[i] = 128$ and $c_s[i] = 4$. All Conv layers are with kernel size of $k_d[i] = k_u[i] = 3$ and $k_d[i] = 1$, strides of 1, and reflection padding of length 2 with $i = 1, \cdots, N$. LeakyReLU [20] is used as each non-linear activation unit where the slop is set to 0.1. Max pooling is used for downsampling, and bi-linear interpolation is used for upsampling. There is no dropout in downsampling and upsampling blocks, *i.e.* the dropout probability of D_i and U_i are set to 0. For the other blocks, the dropout is conducted element-wisely with dropout probability set to $p_s[i] = 0.3$.

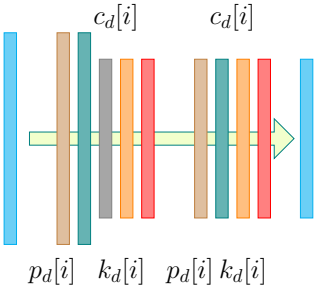
For the initial value $\boldsymbol{x}_0$, we adopt the $\mathcal{J}$-invariant transform of the FBP reconstructed image $\boldsymbol{x}_{\mathbf{FBP}}$ as [4],

$$\boldsymbol{x}_0 = \boldsymbol{b} \odot \boldsymbol{x}_{\mathbf{FBP}} + (1 - \boldsymbol{b}) \odot \boldsymbol{s}(\boldsymbol{x}_{\mathbf{FBP}}),$$

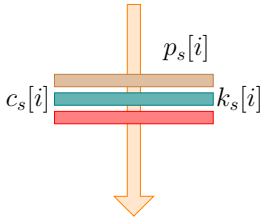
where $\odot$ denotes the element-wise multiplication, $\boldsymbol{b}$ denotes a binary Bernoulli vector whose entries are independently sampled from a Bernoulli distribution with probability $p = 0.3$, and the function $\boldsymbol{s}(\cdot)$ is set to the convolution with kernel $\frac{1}{6}[\frac{1}{2}, 1, \frac{1}{2}; 1, 0, 1; \frac{1}{2}, 1, \frac{1}{2}]$. Note that such an initialization can be implemented by adding a Conv Layer with enabled dropout and a pre-defined low-pass filter. For the MC sampling in testing, we set $K = 50$.



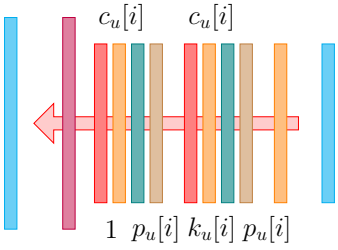
(a) Overall architecture used for performance evaluation.



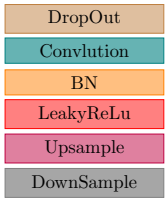
(b) Downsampling block D_i



(c) Skip-connection block S_i



(d) Upsampling Block U_i



(e) Legend

Figure 1: Diagram of the network used for evaluating the proposed method.

3 Experiments

The proposed method is implemented in PyTorch [31] interface on a NVIDIA Titan GPU. Adam optimizer is used with the momentum parameter $\beta = 0.9$, and the learning rate is set as 10^{-5}.

3.1 Methods for comparison

To evaluate its performance on LDCT image reconstruction, the proposed method is compared to several representative non-learning methods, including TV-based penalized weighed least squares method (PWLS-TV), KSVD [46] and BM3D [9]. The PWLS-TV uses the following regularization model for reconstructing the image from the measurement:

$$\arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \alpha\|\nabla\boldsymbol{x}\|_1, \tag{13}$$

where α is the regularization parameter. It is solved by the ADMM method in our experiments.

Also, the proposed method is compared to two recent unsupervised learning methods and a supervised learning method. Two unsupervised learning methods include S2S for denoising-based post-process and DIP+TV for direct image reconstruction. S2S is a recent unsupervised learning method which also uses a dropout-based network for denoising an image. We used it as a post-process to denoise the image reconstructed by the FBP method, where the denoising NN is trained using the S2S method with the same configuration as the proposed method. The DIP+TV [2] combines the DIP approach and TV-based regularization for CT, whose loss function is defined by

$$\mathcal{L}(\boldsymbol{\theta}) = \|\boldsymbol{A}f(\boldsymbol{z};\boldsymbol{\theta}) - \boldsymbol{y}\|_2^2 + \alpha\|\nabla f(\boldsymbol{z};\boldsymbol{\theta})\|_1. \tag{14}$$

The DIP+TV is related to the proposed method. Indeed, the proposed method is degenerated to the DIP+TV by setting the dropout probability p to 1 and using the same input $\boldsymbol{z}$. In the DIP-based denoising network, a random noise is used as the initial seed. In this paper, the initial seed for the DIP+TV is also set to $\boldsymbol{x}_0$, the same as the proposed method. DIP+TV is trained using the Adam optimizer with the momentum parameter set as 0.9. The learning rate is set as 10^{-2}.

In addition, for the dataset of prostate images, we also compare the proposed method with a supervised learning method, FBPConvNet [25]. FBPConvNet is one of a representative deep learning methods for CT reconstruction that uses the deep NN as a post-processing technique. In FBPConvNet, U-net architecture is trained with low-dose and normal-dose image pairs to directly denoise the image reconstructed by the FBP method.

3.2 Data simulation

We adopted the proposed method to LDCT reconstruction, in which $\boldsymbol{A}$ is set as the projection matrix. By using a monoenergetic source in CT imaging, the

measurements from CT scan follow Poisson distribution which can be expressed as [12]:

$$\bar{\boldsymbol{y}}_i \sim \text{Poisson}\{I_i \exp(-[\boldsymbol{A}\boldsymbol{x}]_i)\} + N(0, \sigma_e^2), \tag{15}$$

where N refers to normal distribution, $\boldsymbol{x}$ denotes the attenuation map with x_j being the linear attenuation coefficient in the j-th pixel for $j = 1, \cdots, n$ and n denotes the total number of pixels; $\bar{\boldsymbol{y}}$ represents the measured projection. The matrix $\boldsymbol{A}$ is the $m \times n$ system matrix with entries a_{ij}, and $[\boldsymbol{A}\boldsymbol{x}]_i = \sum_{j=1}^{n} a_{ij} x_j$ denotes the line integral of the attenuation map $\boldsymbol{x}$ along the i-th X-ray with $i = 1, \cdots, m$. I_i is the incident X-ray intensity incorporating X-ray source illumination and the detector efficiency. The noise level is controlled by I_i, *i.e.*, the noise of measures data becomes larger when the dose level I_i decreases. The scalar σ_e^2 denotes the variance of the background electronic noise. To reconstruct the attenuation map $\boldsymbol{x}$, we take the logarithm transform on the noisy measurements $\bar{\boldsymbol{y}}$ to generate the noisy sinogram $\boldsymbol{y}$.

3.3 LDCT reconstruction result

For quantitative analysis of image quality, three indices: peak signal to noise ratio (PSNR), root mean square error (RMSE) and structural similarity index measure (SSIM) [37] are compared for different reconstruction methods.

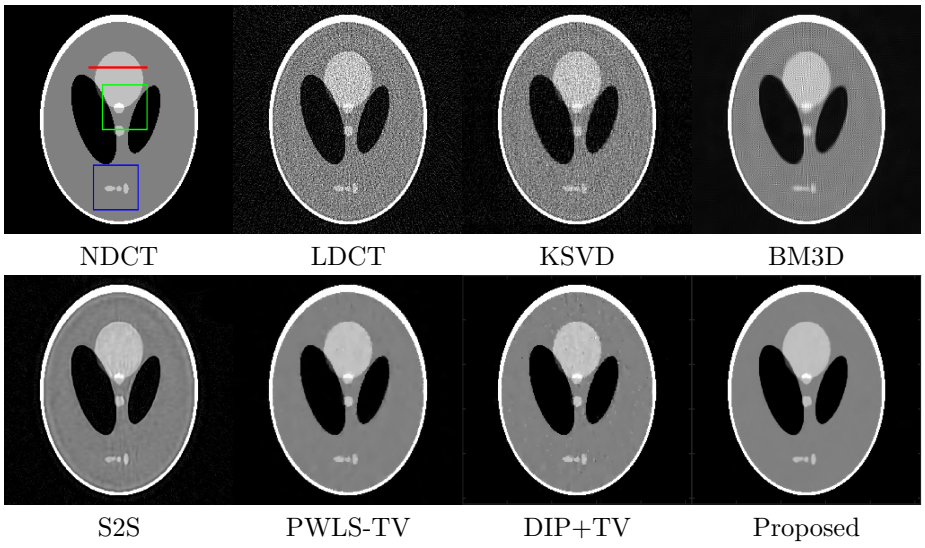


Figure 2: Reconstruction results of phantom.

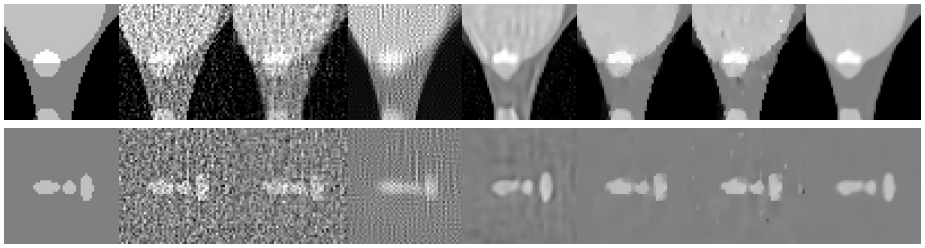


Figure 3: Zoom-in results of Phantom in Fig. 2.

	LDCT(FBP)	KSVD	BM3D	S2S	PWLS-TV	DIP+TV	Proposed
PSNR	25.2472	26.9100	29.8907	34.0764	35.4817	40.5554	42.3255
RMSE	0.0273	0.0226	0.0160	0.0099	0.0084	0.0047	0.0038
SSIM	0.3604	0.4642	0.4453	0.8343	0.9805	0.9861	0.9949
Reconstruction Time	0.01s	145.80s	23.61s	40.07min	3.24s	13.52min	55.22min

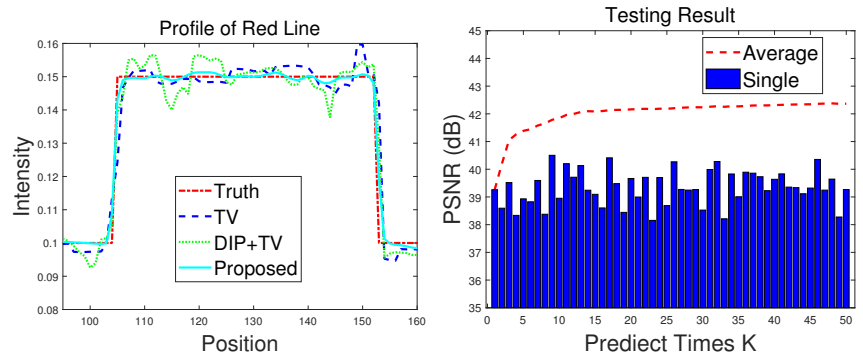
Table 1: Quantitative reconstruction results of phantom in Fig. 2.

3.3.1 Phantom image reconstruction

To evaluate the effectiveness of the proposed method, we simulated the digital phantom of size 256×256 and the corresponding noisy sinogram. The LDCT projection data was simulated by adding Poisson noise and the background electronic onto the normal-dose projection data with $I_i = 1 \times 10^3$ and $\sigma_e^2 = 10$. The simulated geometry for projection data is as follows: fan-beam CT scanner, flat-panel detector of 0.388 mm $\times$ 0.388 mm pixel size, 600 projection views evenly spanning a 360° circular orbit, 512 detector bins for each projection with 1mm pixel size, 100.0cm source to detector distance and 50.0 cm source to isocenter distance. The hyper-parameter α was set to $0.02, 0.2$ and 0.2 for PWLS-TV, DIP+TV and the proposed respectively.

Fig.2 shows the images reconstructed by different methods, and their zoomed-in images of boxes in Fig. 2 are displayed in Fig. 3. With low-dose measurements, LDCT image reconstructed by FBP present large noisy and streaky artifacts. In comparison to the zoomed-in NDCT image, the results of KSVD, BM3D and S2S have more streaky artifacts than that of the proposed method. In the proposed method, PWLS-TV and DIP+TV, TV regularizer can help suppress the noise and remove the artifact in the reconstructed image. Furthermore, the proposed one has a better performance in structure preservation and noise suppression.

Table 1 shows quantitative comparison of the results shown in Fig. 2. The proposed method has the best performance in terms of three metrics PSNR, RMSE and SSIM. Both PWLS-TV and DIP+TV improves upon the conventional LDCT result as expected, and the proposed outperformed DIP+TV by a noticeable margin, *i.e.* 1.8dB advantage in PSNR. Table 1 shows the compari-



(a) Profile (b) PSNR vs. K

(c) Single Prediction 1 (d) Single Prediction 2 (e) Average

Figure 4: Fig.4(a): Profile of the line labeled in red in Fig.2 NDCT. Fig.4(b): PSNR versus prediction times K. Fig.4(c) and Fig.4(d) are the two single predictions of the image. Fig.4(e): the average of multiple predictions.

son of running time for the proposed method and the other methods. With an NVIDIA A100 graphics card, the reconstruction time of the proposed method is comparable with S2S. For fewer iterations, DIP+TV has the advantage over the proposed method.

Fig. 4(a) shows the profile outlined in red in Fig. 2, where the results of PWLS-TV, DIP+TV and the proposed method are compared. It can be easily seen that the cyan line is more close to the ground truth (red line). See Fig. 4(b) for the illustration of how the value of K, the number of predictions for averaging, impacts the performance of the simulated phantom. It shows that the PSNR value steadily increases with the value of K until it hits 15. Afterward, the improvement brought by more predictions is rather small. See Fig. 4(c) and Fig. 4(d) for the visualization of two single predictions of the image. It can be seen that there is a noticeable difference between two single predictions in certain regions. Fig. 4(e) visualizes the average of multiple predictions, which contains fewer artifacts than the two predictions shown in Fig. 4(c) and Fig. 4(d). In other words, the average of multiple predictions from the proposed dropout-

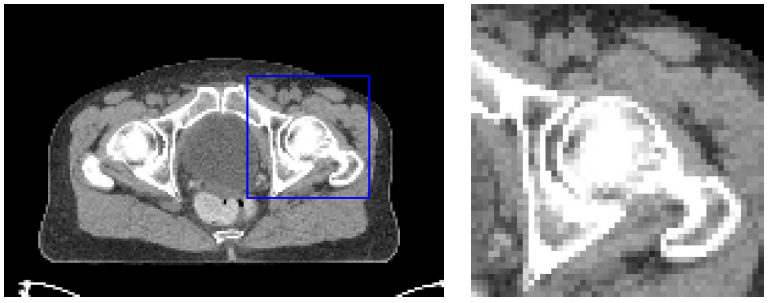
based method indeed can reduce the artifacts in the reconstructed image.

3.3.2 Reconstruction from different noise levels

We evaluated the proposed method with different noise levels. For patient's normal dose prostate image with size 256×256, we simulated the low-dose measurement with dose levels $I_i = 1 \times 10^3$, 5×10^3, 1×10^4, 5×10^4 and $\sigma_e^2 = 10$. Then, the sinograms of different noise levels were obtained by taking logarithm on projection data. The simulated geometry for projection data is the same as that of phantom data simulation.

Fig. 5 demonstrates the NDCT image and the zoomed region for comparison. Fig. 6 shows the images reconstructed by different methods of different dose levels, and their zoomed-in images of boxes in Fig. 5 are displayed in Fig. 7. The displayed window is set to $[-150, 200]$HU for all figures with $\mu_{air} = -1000$HU. For all the reconstruction methods, the image quality decreases with the lower dose level. The recovered images by BM3D are not visually satisfactory. The images by S2S are blurry that some image details are missing. The proposed, PWLS-TV and DIP+TV are the three best performers among all methods without dataset. For the three methods with TV regularizer, the proposed method achieved the be image quality with preserved image structure and less noise. In this experiment, the values of α are adjusted to the noise level of the data. For PWLS-TV, DIP+TV and the proposed method, $\alpha = 0.1,\ 0.1,\ 0.1$ with $I = 1 \times 10^3$; $\alpha = 0.05,\ 0.05,\ 0.05$ with $I = 5 \times 10^3$; $\alpha = 0.02,\ 0.03,\ 0.01$ with $I = 1 \times 10^4$ and $\alpha = 0.01,\ 0.01,\ 0.001$ with $I = 5 \times 10^4$.

In this prostate dataset, there are 6400 normal-dose prostate CT images. We adopted FBPConvNet with 80% of low-dose and normal dose image pairs. It is shown in Table 2 and Fig. 6 that supervised method has the best performance in comparison with non-learning methods and unsupervised DL methods. With the dose $I = 5 \times 10^3$ and $I = 1 \times 10^3$, FBPConvNet gained 1-2dB advantage over the proposed method. The proposed method and DIP+TV achieved higher PSNR and SSIM, and smaller RMSE among all unsupervised methods. Moreover, the proposed method outperformed DIP+TV by 1.0-1.3dB in PSNR.



NDCT ROI

Figure 5: Normal dose CT image and a zoomed region (ROI).

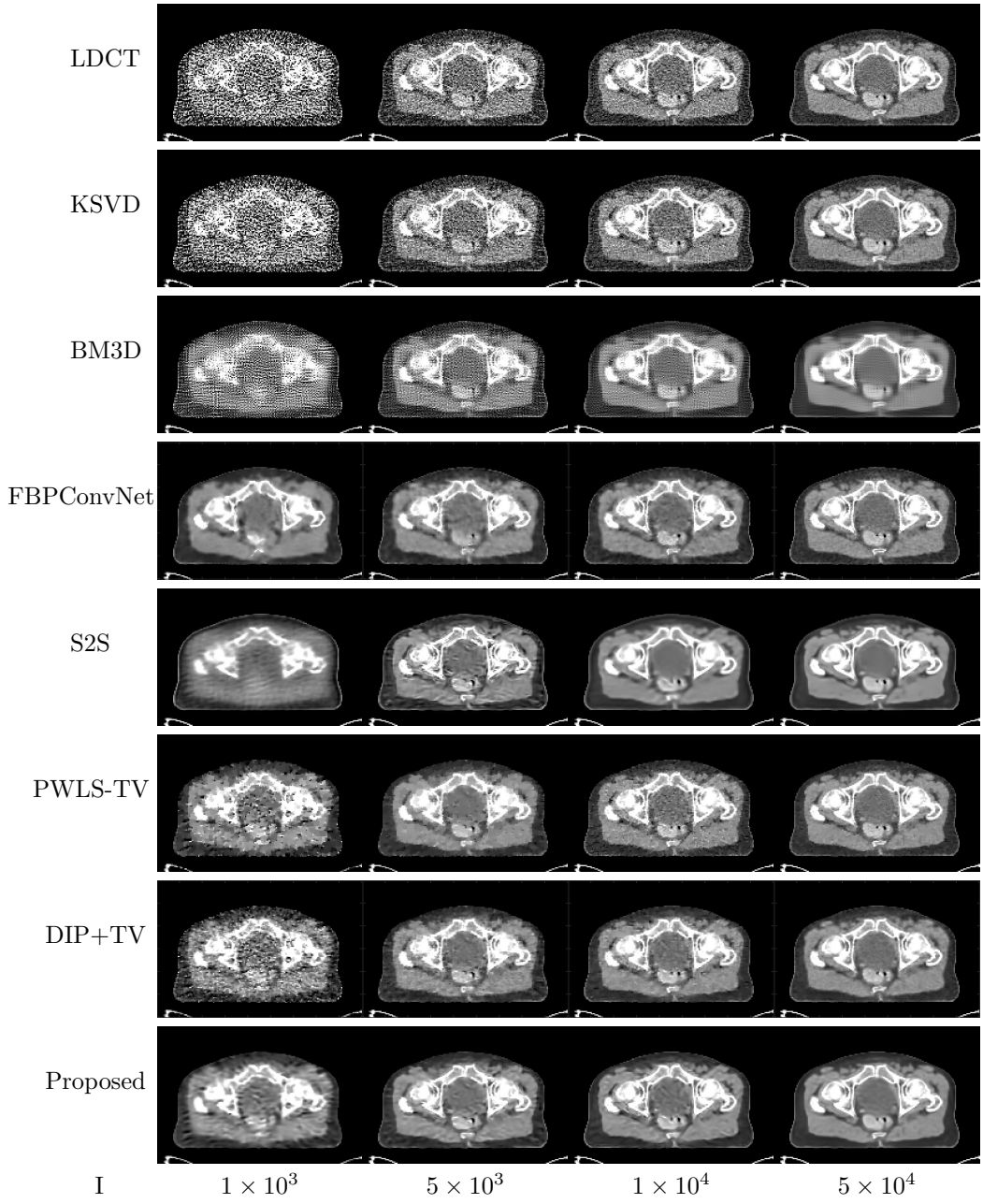


Figure 6: Reconstruction results at different dose levels by different methods with $\sigma_e^2 = 10$

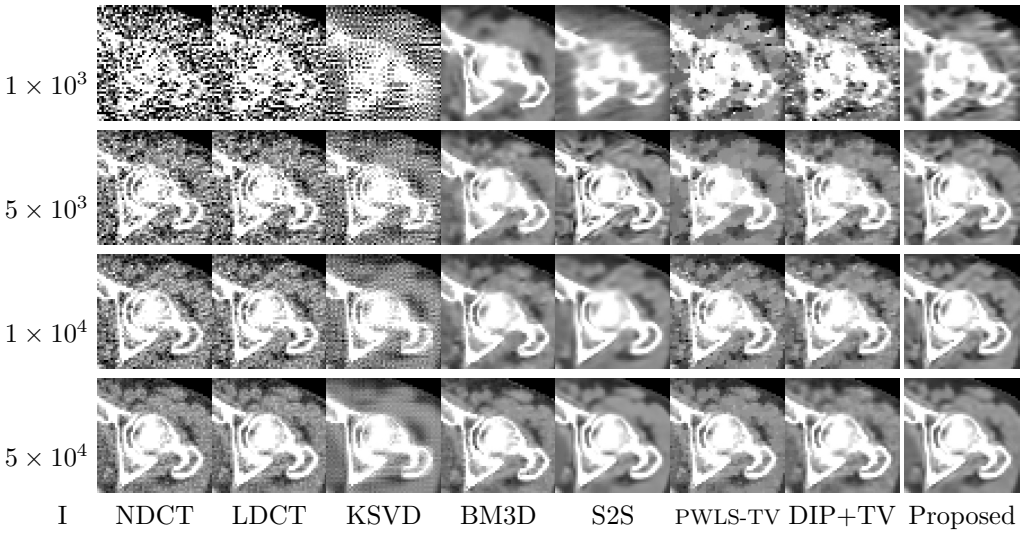
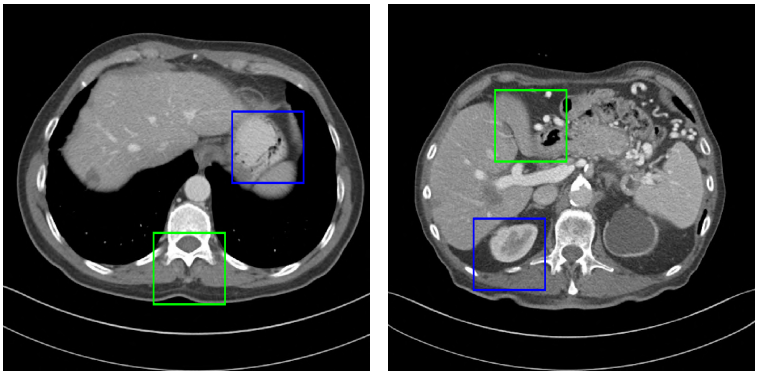


Figure 7: Zoomed-in results corresponding to Fig. 6 at different dose levels by different methods with $\sigma_e^2 = 10$.

Dose Level	Index	LDCT	KSVD	BM3D	FBPConvNet	S2S	PWLS-TV	DIP+TV	Proposed
1×10^3	PSNR	16.0859	16.7163	19.9995	29.6054	25.7518	25.8050	25.3503	27.5996
	RMSE	162.8839	151.4813	103.8002	34.3481	53.5282	53.2013	56.0607	43.2707
	SSIM	0.2744	0.2841	0.4379	0.8159	0.5893	0.7001	0.6739	0.7675
5×10^3	PSNR	24.3676	24.9922	26.2975	33.2609	30.3793	30.0352	30.8064	31.9237
	RMSE	62.7758	58.4203	50.2687	26.3339	31.4200	32.6897	29.9125	26.3018
	SSIM	0.5510	0.5805	0.6622	0.8762	0.8086	0.8535	0.8544	0.8722
1×10^4	PSNR	27.4944	27.5416	29.1881	33.3785	31.3774	31.3655	32.4104	33.5989
	RMSE	43.7979	43.5605	36.0386	22.2457	28.0094	28.0477	24.8687	21.6883
	SSIM	0.6834	0.7024	0.7808	0.8943	0.8702	0.8612	0.8894	0.9070
5×10^4	PSNR	33.1157	32.5643	33.6708	36.9016	34.3365	35.0872	35.4009	36.1700
	RMSE	22.9291	24.4320	21.5097	14.8283	19.9226	18.2731	17.3623	16.1314
	SSIM	0.8861	0.8894	0.9166	0.9518	0.9301	0.9234	0.9220	0.9408

Table 2: Quantitative reconstruction results of Fig. 6.

3.3.3 Clinical Data Reconstruction



(a) (b)

Figure 8: Two normal dose CT images from mayo dataset.

To evaluate the performance of the proposed method under realistic conditions, a clinical image was used, which was established by Mayo Clinics for "the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge". We aimed to reconstruct the image with size 512×512 from the projection data of full dose CT images of 3mm thickness. The projection data is composed of projection data from 600 projection views evenly spanning a 360° circular orbit, 768 detector bins for each projection, 100.0 cm source to detector distance and 50.0 cm source to isocenter distance.

Fig. 8 shows two normal dose mayo image slices and two ROIs of each slice are labeled by blue and green boxes. Fig. 9 and Fig. 10 demonstrate the zoomed region of interest corresponding to Fig. 8(a) and Fig. 8(b) respectively. The displayed window is set to $[-200, 400]$HU for all figures with $\mu_{air} = -1000$HU. LDCT image is reconstructed by FBP with heavy noise and artifacts. Both post-processing-based type methods BM3D and S2S are not able to remove the streaky artifacts. TV and DIP+TV could reduce noise and remove artifacts, but the image details are smoothed out. The proposed method achieves the image result with better structure preservation and noise suppression. For the image in Fig. 8(a), the hyper-parameter α was set to 0.5, 0.3 and 0.3 for PWLS-TV, DIP+TV and the proposed method respectively for optimal performance. For the image in Fig. 8(b), the hyper-parameter α was set to 0.3 for all the TV-based methods.

Quantitative reconstruction results corresponding to the image in Fig. 8(a) and Fig. 8(b) are given in Table 3. The proposed method has the best performance in terms of the metrics than the other methods. In comparison with DIP+TV, the proposed method outperforms DIP+TV in PSNR about 1dB.

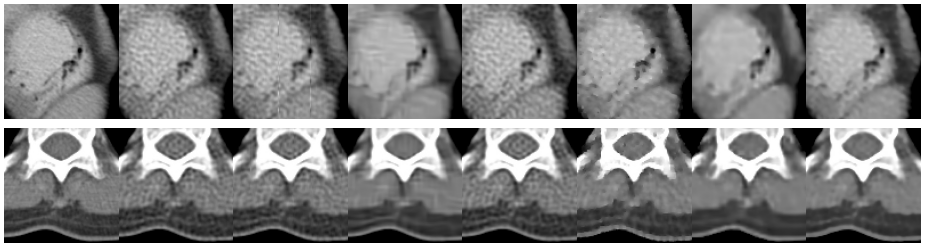


Figure 9: Zoomed-in results of Mayo image corresponding to Fig. 8(a) .

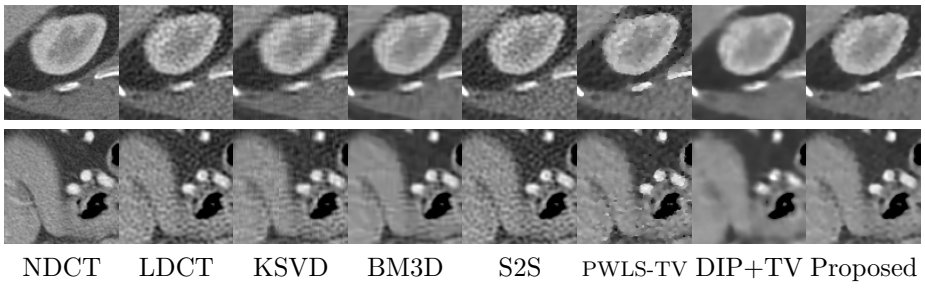


Figure 10: Zoomed-in results of Mayo image corresponding to Fig. 8(b) .

	Index	LDCT	KSVD	BM3D	S2S	PWLS-TV	DIP+TV	Proposed
Fig. 8(a)	PSNR	30.0174	29.9941	30.4400	30.1597	34.3136	34.7112	36.4061
	RMSE	42.4475	42.5617	40.4317	41.7577	25.8847	24.7265	20.3432
	SSIM	0.9098	0.9113	0.9398	0.9138	0.9475	0.9518	0.9577
Fig. 8(b)	PSNR	31.011	31.0529	31.6613	31.3406	33.9875	33.7172	35.2124
	RMSE	41.9409	41.7388	38.9155	40.3789	29.7721	30.7134	25.8562
	SSIM	0.8921	0.8950	0.9273	0.9095	0.9276	0.9192	0.9461

Table 3: Quantitative reconstruction results of Fig. 8(a) and Fig. 8(b).

4 Discussion and Conclusion

In this paper, we proposed an unsupervised learning method for LDCT image reconstruction, which is based on a re-parametrization via the network with random weights for Bayesian inference. The proposed method does not require any external training samples, which is flexible and easy to use in practice. The experiments showed that the proposed method outperformed the representative non-learning-based methods and two recent unsupervised deep learning methods. The proposed method can be potentially adapted to other medical image reconstruction problems, where the training samples are costly or difficult to collect, *e.g.* sparse-view CT reconstruction and image reconstruction from sparse samples in MRI. In the future, we would like to investigate better data-driven regularization for avoiding the possible over-fitting caused by the absence of ground-truth. In this paper, the values of the regularization parameter and dropout probability are manually tuned up for optimal performance. How to automate the setting of these parameters remains a question to be studied in future work. In addition, unsupervised deep learning methods canot pre-train a model such that it can be called to process test images without training. Computational efficiency is important for practical usage. It will be our future work on how to address such an issue. One direction is to study a light-weight network with few parameters for LDCT. Another direction is to study test-time adaption which uses an unsupervised method to quickly adapt a pre-trained model to process test data.

References

[1] Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE transactions on medical imaging*, 37(6):1322–1332, 2018.

[2] Daniel Otero Baguer, Johannes Leuschner, and Maximilian Schmidt. Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Problems*, 36(9):094004, 2020.

[3] Michael Balda, Joachim Hornegger, and Bjoern Heismann. Ray contribution masks for structure adaptive sinogram filtering. *IEEE transactions on medical imaging*, 31(6):1228–1239, 2012.

[4] Joshua Batson and Loic Royer. Noise2Self: Blind Denoising by Self-Supervision. In *International Conference on Machine Learning*, pages 524–533, 2019.

[5] Jianfeng Cai, Xun Jia, Hao Gao, Steve B Jiang, Zuowei Shen, and Hongkai Zhao. Cine cone beam CT reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study. *IEEE transactions on medical imaging*, 33(8):1581–1591, 2014.

[6] Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36(12):2524–2535, 2017.

[7] Hu Chen, Yi Zhang, Weihua Zhang, Peixi Liao, Ke Li, Jiliu Zhou, and Ge Wang. Low-dose CT via convolutional neural network. *Biomedical optics express*, 8(2):679–694, 2017.

[8] Mingqin Chen, Yuhui Quan, Tongyao Pang, and Hui Ji. Nonblind image deconvolution via leveraging model uncertainty in an untrained deep neural network. *International Journal of Computer Vision*, pages 1–20, 2022.

[9] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.

[10] Qiaoqiao Ding, Gaoyu Chen, Xiaoqun Zhang, Qiu Huang, Hui Ji, and Hao Gao. Low-dose CT with deep learning regularization via proximal forward–backward splitting. *Physics in Medicine & Biology*, 65(12):125009, 2020.

[11] Qiaoqiao Ding, Hui Ji, Hao Gao, and Xiaoqun Zhang. Learnable multi-scale fourier interpolation for sparse view ct image reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 286–295. Springer, 2021.

[12] Qiaoqiao Ding, Yong Long, Xiaoqun Zhang, and Jeffrey A Fessler. Modeling mixed Poisson-Gaussian noise in statistical image reconstruction for X-ray CT. *Proc. 4th Intl. Mtg. on image formation in X-ray CT*, pages 399–402, 2016.

[13] Qiaoqiao Ding, Yuesong Nan, Hao Gao, and Hui Ji. Deep Learning with Adaptive Hyper-parameters for Low-Dose CT Image Reconstruction. *IEEE Transactions on Computational Imaging*, 7:648–660, 2021.

[14] Lee A Feldkamp, Lloyd C Davis, and James W Kress. Practical cone-beam algorithm. *Josa a*, 1(6):612–619, 1984.

[15] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[16] Kuang Gong, Ciprian Catana, Jinyi Qi, and Quanzheng Li. PET image reconstruction using deep image prior. *IEEE transactions on medical imaging*, 38(7):1655–1665, 2018.

[17] Harshit Gupta, Kyong Hwan Jin, Ha Q Nguyen, Michael T McCann, and Michael Unser. CNN-based projected gradient descent for consistent CT image reconstruction. *IEEE transactions on medical imaging*, 37(6):1440–1453, 2018.

[18] Ahmed M Hasan, Mohammad Reza Mohebbian, Khan A Wahid, and Paul Babyn. Hybrid Collaborative Noise2Noise Denoiser for Low-dose CT Images. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2020.

[19] Ji He, Yan Yang, Yongbo Wang, Dong Zeng, Zhaoying Bian, Hao Zhang, Jian Sun, Zongben Xu, and Jianhua Ma. Optimizing a parameterized plug-and-play ADMM for iterative low-dose CT reconstruction. *IEEE transactions on medical imaging*, 38(2):371–382, 2018.

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[21] Allard Adriaan Hendriksen, Daniël Maria Pelt, and K Joost Batenburg. Noise2Inverse: Self-Supervised Deep Convolutional Denoising for Tomography. *IEEE Transactions on Computational Imaging*, 6:1320–1335, 2020.

[22] Jiang Hsieh. *Computed tomography: principles, design, artifacts, and recent advances*, volume 114. SPIE press, 2003.

[23] Xun Jia, Bin Dong, Yifei Lou, and Steve B Jiang. GPU-based iterative cone-beam CT reconstruction using tight frame regularization. *Physics in Medicine & Biology*, 56(13):3787, 2011.

[24] Xun Jia, Yifei Lou, Bin Dong, Zhen Tian, and Steve Jiang. 4D computed tomography reconstruction from few-projection data via temporal non-local regularization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 143–150. Springer, 2010.

[25] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[26] Avinash C Kak, Malcolm Slaney, and Ge Wang. Principles of computerized tomographic imaging, 2002.

[27] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning Image Restoration without Clean Data. In *ICML*, 2018.

[28] Heyi Li and Klaus Mueller. Low-dose CT streak artifacts removal using deep residual neural network. In *Proc. Fully Three-Dimensional Image Reconstruction Radiol. Nucl. Med.(Fully3D)*, pages 191–194, 2017.

[29] Ji Li, Yuesong Nan, and Hui Ji. Un-supervised learning for blind image deconvolution via monte-carlo sampling. *Inverse Problems*, 38(3):035012, 2022.

[30] Armando Manduca, Lifeng Yu, Joshua D Trzasko, Natalia Khaylova, James M Kofler, Cynthia M McCollough, and Joel G Fletcher. Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Medical physics*, 36(11):4911–4919, 2009.

[31] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. *Neural Information Processing Systems*, 2017.

[32] Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2Self With Dropout: Learning Self-Supervised Denoising From Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1890–1898, 2020.

[33] Emil Y Sidky, Chien-Min Kao, and Xiaochuan Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-ray Science and Technology*, 14(2):119–139, 2006.

[34] Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 53(17):4777, 2008.

[35] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[36] Dave Van Veen, Ajil Jalal, Mahdi Soltanolkotabi, Eric Price, Sriram Vishwanath, and Alexandros G Dimakis. Compressed sensing with deep image prior and learned regularization. *arXiv preprint arXiv:1806.06438*, 2018.

[37] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[38] Bruce R Whiting, Parinaz Massoumzadeh, Orville A Earl, Joseph A O Sullivan, Donald L Snyder, and Jeffrey F Williamson. Properties of preprocessed sinogram data in X-ray computed tomography. *Medical physics*, 33(9):3290–3303, 2006.

[39] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. Generative adversarial networks for noise reduction in low-dose CT. *IEEE transactions on medical imaging*, 36(12):2536–2545, 2017.

[40] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. 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, 2018.

[41] Dong Hye Ye Ye, Somesh Srivastava, Jean-Baptiste Thibault, Jiang Hsieh, Ken Sauer, and Charles Bouman. Deep residual learning for model-based iterative ct reconstruction using plug-and-play framework. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6668–6672, 2018.

[42] Jong Chul Ye, Yoseob Han, and Eunju Cha. Deep convolutional framelets: A general deep learning framework for inverse problems. *SIAM Journal on Imaging Sciences*, 11(2):991–1048, 2018.

[43] Tatsuya Yokota, Kazuya Kawai, Muneyuki Sakata, Yuichi Kimura, and Hidekata Hontani. Dynamic PET Image Reconstruction Using Nonnegative Matrix Factorization Incorporated with Deep Image Prior. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3126–3135, 2019.

[44] Jaejun Yoo, Kyong Hwan Jin, Harshit Gupta, Jerome Yerly, Matthias Stuber, and Michael Unser. Time-dependent deep image prior for dynamic mri. *IEEE Transactions on Medical Imaging*, 2021.

[45] Nimu Yuan, Jian Zhou, and Jinyi Qi. Half2Half: deep neural network based CT image denoising without independent reference data. *Physics in Medicine & Biology*, 2020.

[46] Qiang Zhang and Baoxin Li. Discriminative k-svd for dictionary learning in face recognition. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 2691–2698. IEEE, 2010.

[47] Xiaoqun Zhang and Jacques Froment. Total variation based fourier reconstruction and regularization for computer tomography. In *Nuclear Science Symposium Conference Record, 2005 IEEE*, volume 4, pages 2332–2336. IEEE, 2005.

[48] Kevin C Zhou and Roarke Horstmeyer. Diffraction tomography with a deep image prior. *Optics Express*, 28(9):12872–12896, 2020.