Efficient deep unrolling networks for image reconstruction of X-ray CT scans

Project-II report submitted to

Indian Institute of Technology Kharagpur

in partial fulfilment for the award of the degree of

Bachelor of Technology

in

Electronics & Electrical Communication Engineering

by

Kadagala Sai Siva Sankar

(20EC39017)

Under the supervision of

Professor Subhadip Mukherjee

Department of E&ECE

Indian Institute of Technology Kharagpur

Spring Semester, 2023-2024

April 29, 2024

DECLARATION

I certify that

(a) The work contained in this report has been done by me under the guidance of my supervisor.

(b) The work has not been submitted to any other Institute for any degree or diploma.

(c) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

(d) Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Date: April 29, 2024 (Kadagala Sai Siva Sankar)
Place: Kharagpur (20EC39017)

DEPARTMENT OF E&ECE

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

KHARAGPUR - 721302, INDIA

CERTIFICATE

This is to certify that the project report entitled "**Efficient deep unrolling networks for image reconstruction of X-ray CT scans**" submitted by **Kadagala Sai Siva Sankar** (Roll No. 20EC39017) to Indian Institute of Technology Kharagpur towards partial fulfilment of requirements for the award of degree of Bachelor of Technology in Electronics & Electrical Communication Engineering is a record of bona fide work carried out by him under my supervision and guidance during Spring Semester, 2023-2024.

Date: April 29, 2024
Place: Kharagpur

Professor Subhadip Mukherjee
Department of E&ECE
Indian Institute of Technology Kharagpur
Kharagpur - 721302, India

Abstract

Name of the student: **Kadagala Sai Siva Sankar** Roll No: **20EC39017**

Degree for which submitted: **Bachelor of Technology**

Department: **Department of E&ECE**

Thesis title: **Efficient deep unrolling networks for image reconstruction of X-ray CT scans**

Thesis supervisor: **Professor Subhadip Mukherjee**

Month and year of thesis submission: **April 29, 2024**

X-ray computed tomography involves reconstructing images from incomplete and noisy observations known as sinograms. Traditional methods like filtered back projection (FBP) have been widely used but have limitations, particularly in sparse or noisy data scenarios. The emergence of deep learning has offered promising alternatives, with deep unrolling networks being the current state-of-the-art technique. However, they face efficiency, reliability, and scalability challenges, especially in high-dimensional tasks. Conventional deep-unrolling methods are computationally intensive due to their reliance on full forward and adjoint operators' usage at every layer, which is particularly noticeable in 3D image reconstruction tasks. From the following set of experiments, leveraging stochastic primal-dual and sketched learned stochastic primal-dual networks (Junqi Tang, Subhadip Mukherjee, and Carola-Bibiane Schönlieb, 2022), we show that they achieve comparable results with the state-of-the-art Learned Primal-Dual network (Adler & Öktem, 2018) while being computationally less expensive.

Acknowledgements

I wish to express my heartfelt gratitude to my project advisor, Prof. Subhadip Mukherjee, whose unwavering support and guidance have been instrumental in my research. Professor Subhadip's expertise, mentorship, and willingness to engage in constructive discussions have greatly enriched my academic journey. His willingness to offer guidance and insights, while still allowing me the autonomy to explore and learn independently have been invaluable.

Contents

Chapter 1

Introduction

1.1 Background

In recent years, there have been significant advancements in deep learning for image reconstruction, particularly in fields like medical imaging. Image reconstruction is a common task in medical imaging where an image of interest needs to be recovered from its incomplete and noisy observation. In many computational imaging applications, image reconstruction serves as a vital tool for scientific exploration, often without a definitive ground truth available. Consequently, it's crucial to have confidence in the accuracy of the reconstructed images to ensure reliable scientific discoveries. Traditional methods like filtered back projection (FBP) have long been the cornerstone, but their limitations in scenarios with sparse or noisy data have prompted the exploration of alternative approaches. With the emergence of deep learning and availability of high-quality data, it is now possible to explore deep learning approaches in medical image reconstruction tasks. Deep unrolling networks are the state-of-the-art method in image inverse problems, utilizing neural networks' ability to learn complex patterns from data directly. These networks effectively unroll iterative algorithms into neural architectures, offering a blend of flexibility and performance improvement. However, their efficacy, especially in high-dimensional imaging tasks like X-ray CT and MRI, faces significant computational challenges, impeding their practical application.

1.2 Problem statement and research question

The image reconstruction task in CT entails recovering $x^\dagger$ from noisy measurement given by $y = Ax^\dagger + w$, where $\mathbf{x}^\dagger$ denotes the ground truth image, and $\mathbf{A}$ denotes the forward measurement operator. $\mathbf{w}$ represents the measurement noise, while $\mathbf{b}$ denotes the measurement data. A classical way to obtain a good estimate of $\mathbf{x}^\dagger$ is to solve the optimization problem:

$$\mathbf{x}^* \in \arg\min_{\mathbf{x}\in\mathbb{R}^d} f_b(\mathbf{A}\mathbf{x}) + r(\mathbf{x}),$$

where the data fidelity term $f_b(\mathbf{A}\mathbf{x})$ is typically a convex function, while $r(\mathbf{x})$ is a regularization term.

Despite the progress in deep learning-based image reconstruction, concerns about their reliability and stability linger. The primary problem lies in ensuring the trustworthiness and consistency of deep learning methods, particularly in the absence of definitive ground truth, which is common in many computational imaging applications. The current state-of-the-art method, Learned Primal Dual addresses these concerns. But, in real-world clinical CT applications, the reconstruction of a 3D volume is essential, yet the computational expense of the forward operator poses a significant challenge. This complexity renders state-of-the-art unrolling approaches inapplicable in the 3D setting. Addressing these concerns, the research question at hand is: How can we enhance the efficiency of deep unrolling networks for image reconstruction, especially in high-dimensional imaging tasks like X-ray CT in 3D setting.

1.3 Objectives and challenges

Efficiency: Deep unrolling networks often struggle with efficiency, especially in tasks with high dimensions. This is due to the computational burden of repeatedly calculating high-dimensional forward and adjoint operations.

Reliability: Ensuring the reliability and stability of deep learning methods in image reconstruction is crucial for their practical use. This is particularly important in critical applications like medical imaging, where accuracy is essential.

Noise Robustness: In real-world clinical settings, medical imaging data are commonly corrupted with noise. Therefore, robust reconstruction methods are needed to effectively handle noisy measurements.

Scalability: The proposed methodologies must be scalable to handle 3D volumes and high-resolution images commonly found in medical imaging. This scalability should be achieved without sacrificing performance or computational efficiency.

1.4 Literature review

1.4.1 Filtered backprojection

Filtered back projection (FBP) is a fundamental technique used extensively in medical imaging, especially in X-ray computed tomography (CT). It operates by first acquiring raw projection data from multiple angles around the patient. These data are then processed by applying a filter, typically a high-pass filter, to remove low-frequency noise and artifacts while preserving important image details. After filtering, the processed data are back-projected onto a two-dimensional image grid, where each point in the grid represents the attenuation properties of the imaged object along a specific line of sight. This process is repeated for each angle of projection, and the results are summed to produce the final reconstructed image. It is computationally efficient, allowing for rapid reconstruction of images, which is crucial in clinical settings where quick results are needed. Despite its widespread use and effectiveness, FBP has limitations, particularly in scenarios with limited data or when dealing with highly attenuating objects.

1.4.2 Deep unrolling networks

Unrolling arises from variational methods, such as total variation regularization. Also, variational regularization, solved using iterative gradient-based solvers has been the most successful approach classically. Deep unrolling networks are a class of neural networks used primarily for inverse problems in image reconstruction and signal processing. They are designed to address complex optimization tasks by unrolling iterative algorithms into neural network architectures. These networks leverage the concept of "unrolling" iterative algorithms, such as iterative optimization methods like gradient descent or alternating minimization, into a fixed number of steps. Each step of the algorithm is represented by a layer in the neural network, and the parameters of these layers are learned during training. By unrolling the iterative process into a neural network, deep unrolling networks can learn to solve inverse problems directly from data, without requiring explicit knowledge of the forward model or the inverse problem itself. This makes them highly flexible and adaptable to various tasks, including denoising, and medical image reconstruction. One significant advantage of deep unrolling networks is their ability to incorporate prior knowledge about the problem domain into the network architecture, allowing for improved performance and generalization. Additionally, they can handle ill-posed inverse problems more effectively compared to traditional optimization methods.

1.4.3 Learned primal dual

The saddle-point problem can be efficiently solved by the primal-dual hybrid gradient (PDHG) method, which is also known as the Chambolle-Pock algorithm in the optimization literature. The PDHG method for solving the saddle-point problem obeys the following updating rule:

Primal-Dual Hybrid Gradient (PDHG)

Initialize $x_0, \bar{x}_0 \in \mathbb{R}^d, \quad y_0 \in \mathbb{R}^p$

For $k = 0, 1, 2, \ldots, K$

$$y_{k+1} = \text{prox}_{\sigma f^*}(y_k + \sigma A\bar{x}_k);$$

$$x_{k+1} = \text{prox}_{\tau g}(x_k - \tau A^\top y_{k+1});$$

$$\bar{x}_{k+1} = x_{k+1} + \beta(x_{k+1} - x_k);$$

The PDHG algorithm takes alternatively the gradients regarding the primal variable x and dual variable y and performs the updates. The state-of-the-art unrolling scheme – learned primal-dual network is based on unfolding the iteration of PDHG by replacing the proximal operators with multilayer convolutional neural networks applied on the both primal and dual spaces. The step sizes at each steps are also set to be trainable.

Learned Primal-Dual (LPD)

Initialize $x_0 \in \mathbb{R}^d, \quad y_0 \in \mathbb{R}^p$

For $k = 0, 1, 2, \ldots, K-1$

$$y_{k+1} = D_\theta^k(y_k, \sigma_k, Ax_k, b);$$

$$x_{k+1} = P_\theta^k(x_k, \tau_k, A^\top y_{k+1});$$

where x is the primal variable, y is the dual variable, D_θ^k is the kth dualNet and P_θ^k is the kth PrimalNet. A, A^T are the forward and adjoint operators and σ_k, τ_k are the trainable step sizes for kth Dual and Primal networks respectively

1.4.4 Learned stochastic primal dual and Sketched learned stochastic primal dual

In Learned stochastic primal dual , we replace the forward and adjoint operators in the full-batch LPD network with only subsets of it. We partition the forward and adjoint operators into m subsets, and also the corresponding measurement data. In each layer, we use only one of the subsets, in a cycling order. The general framework of LSPD is:

Learned Stochastic Primal-Dual (LSPD)

Initialize $x_0 \in \mathbb{R}^d, \quad y_0 \in \mathbb{R}^{p/m}$

For $k = 0, 1, 2, \ldots, K-1$

$i = \mathrm{mod}(k, m);$

(or pick i from $[0, m-1]$ uniformly at random)

$y_{k+1} = D_\theta^k(y_k, \sigma_k, (S_iA)x_k, S_ib);$

$x_{k+1} = P_\theta^k(x_k, T_k, (S_iA)^\top y_{k+1});$

where S $:= [S_0, S_1, S_2, ..., S_{m-1}]$ are the set of sub-sampling operators. For the same number of layers, the LSPD network is approximately m-time more efficient than the full-batch LPD network in terms of computational complexity.

In the accelerated variant of LSPD, the main idea is to speedily approximate the products Ax_k, A^Ty_{k+1} :

$$Ax_k \approx A_{s_k}S_{\theta_s^k}(x_k), A^Ty_{k+1} \approx U_{\theta_u^k}(A^Ty_{k+1})$$

The Sketched LPD network is written as:

Sketched-LPD

Initialize $x_0 \in \mathbb{R}^d, y_0 \in \mathbb{R}^{p/m}$

For $k = 0, 1, 2, \ldots, K-1$

$i = \quad \mathrm{mod}(k, m);$

(or pick i from $[0, m-1]$ uniformly at random)

$y_{k+1} = D_{\theta_d^k}(y_k, \sigma_k, A_{s_k}S_{\theta_s^k}(x_k), b);$

$x_{k+1} = P_{\theta_p^k}(x_k, \tau_k, U_{\theta_u^k}(A_{s_k}^T y_{k+1}));$

Again, we can use the same approximation for stochastic gradient steps:

$$(S_iA)x_k \approx (S_iA_{s_k})S_{\theta_u^k}(x_k),$$
$$(S_iA)^Ty_{k+1} \approx U_{\theta_u^k}((S_iA_{s_k})^Ty_{k+1}),$$

and hence we can write Sketched LSPD (SkLSPD) network as:

Sk-LSPD(option-2)

Initialize $x_0 \in \mathbb{R}^d, y_0 \in \mathbb{R}^{p/m}$

For $k = 0, 1, 2, \ldots, K-1$

$i = \mod(k, m)$;

(or pick i from $[0, m-1]$ uniformly at random)

$y_{k+1} = D_{\theta_d^k}(y_k, \sigma_k, (S_i A_{s_k}) S_{\theta_s^k}(x_k), S_i b)$;

$x_{k+1} = U_{\theta_u^k}(P_{\theta_p^k}(S_{\theta_s^k}(x_k), T_k, (S_i A_{s_k})^T y_{k+1}))$;

In practice, we use the most simple off-the-shelf up/down-sampling operators in Pytorch for example the bilinear interpolation delivers excellent performance for the sketched unrolling networks. We use a "coarse-to-fine" strategy for skLPD and skLSPD. We use more aggressive sketch at the beginning for efficiency, while conservative sketch or non-sketch at latter iterations for accuracy. One possible choice is: for the last few unrolling layers of SkLPD and SkLSPD, we switch to usual LPD/LSPD (say if the number of unrolling layers is 12, we can choose last 4 unrolling layers to be unsketched, such that the reconstruction accuracy is best preserved.

Chapter 2

Methodology and Preliminary Results

2.1 Learned Primal Dual (LPD) on MNIST dataset

For this experiment, the MNIST dataset, consisting of 28x28 grayscale handwritten digit images, is used. A subset of 51,200 samples is randomly selected and split into training (80%) and testing (20%) sets. The ODL toolbox simulates forward and adjoint operators, while the ASTRA toolbox aids GPU implementation for faster training and inference. We set the number of angles to 24, uniformly sampled in the range (0, π). Ground truth images are forward-projected to generate sinograms using the ODL library. A dataset class pairs ground truths and sinograms of each sample. Using PyTorch, we implement the Learned Primal-Dual algorithm, combining deep learning with model-based reconstruction. This involves CNNs in both reconstruction and data space, connected by the forward operator and its adjoint. Networks are trained to minimize mean squared error. Key hyperparameters include 5 DualNet and PrimalNet blocks, a batch size of 1024, and an initial learning rate of 0.001. Each dual/primal network has 3 convolutional layers followed by a PReLU activation (with a skip connection between the first input channel and output) and 32 channels, with kernel size 3, without momentum for simplicity and memory efficiency. The dual variable initializes to zeros. During training, the model's performance is periodically evaluated on the test set, and the

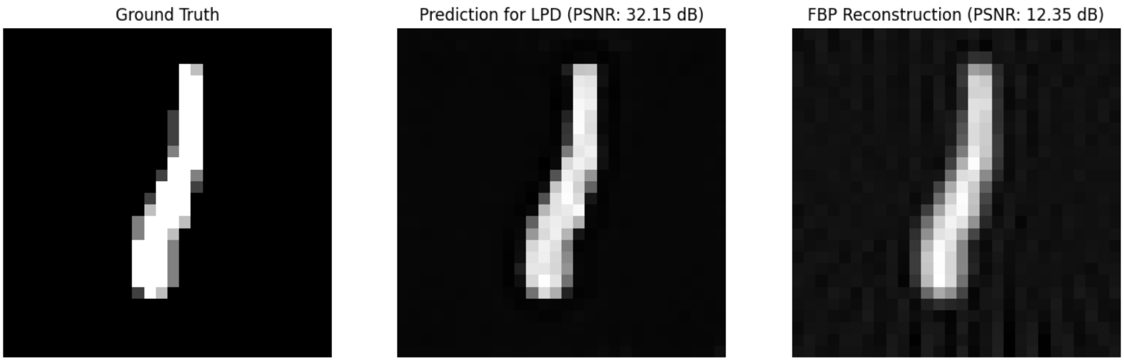

FIGURE 2.1: Example of LPD network performance on MNIST dataset

model with the lowest test loss is saved for further evaluation and inference. Peak Signal-to-Noise Ratio (PSNR) metric calculates differences between predicted and ground truth images, as well as between filtered backprojection (FBP) reconstructions and ground truth images. Average PSNR values for both model predictions and FBP reconstructions are reported, providing a quantitative measure of model reconstruction quality compared to traditional FBP methods. An average PSNR of 30.18 dB is obtained when the model is tested on 5 random test samples while the filtered backprojection achieves an average PSNR of 10.23 dB.

2.2 LPD on Mayo-Clinic dataset

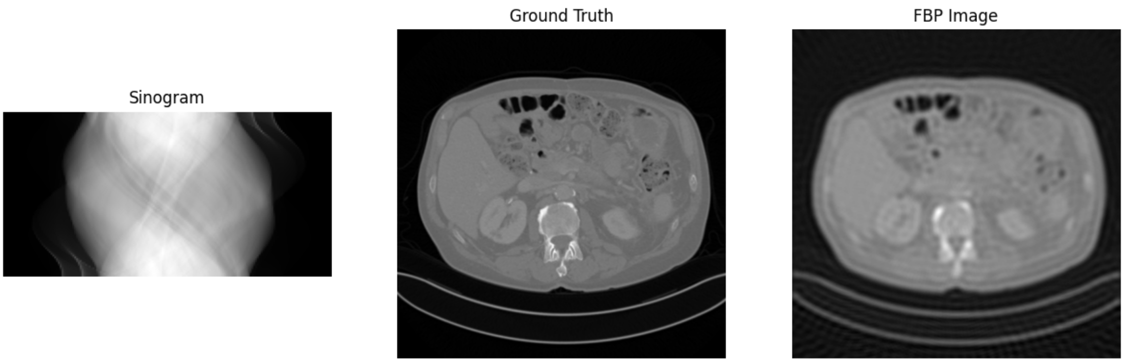

FIGURE 2.2: A sample from the sparse-view dataset

After ensuring the model's functionality with the MNIST dataset, we now utilize the Mayo-Clinic dataset for training. The Mayo-Clinic dataset comprises 2111 slices of 2D images sized 512 x 512, with an 80-20 train-test split. We employ the ODL toolbox with Astra CUDA implementation to simulate parallel beam projection with 90 equally spaced angles (sparse-view dataset). A dataset class is created to pair sinograms, filtered backprojected images, and ground truths. For training,

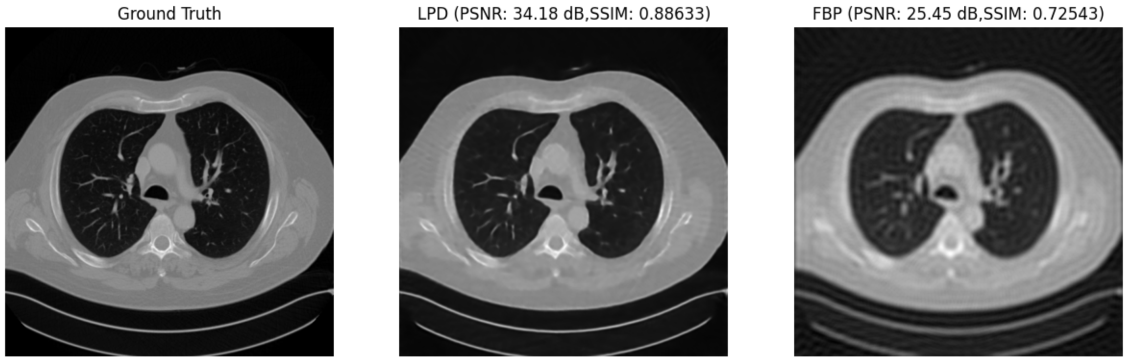

Figure 2.3: Comparing the results of LPD network with FBP

we employ a similar model architecture to that used in Experiment 1, with some modifications. We utilize 12 DualNet and PrimalNet blocks. The kernel size is set to 5, and the step sizes tau and sigma are initialized to 0.01. The batch size is set to 1. We use MSE loss with the Adam optimizer, with an initial learning rate of 5e-5 and betas set to (0.5, 0.99). The network is trained for 15 epochs. After training, on testing the network on 100 samples from the test data, we obtained an average PSNR value of 38.58 dB. The Fbp images achieved an average of 28.06 dB.

2.3 Learned Stochastic Primal Dual (LSPD) and Sketched-LSPD on Mayo-Clinic dataset

Now that we have a working LPD network model, in this experiment, we implemented two architectures proposed in the paper (LSPD and sk-LSPD). For the LSPD network, we replace the forward and adjoint operators in the full-batch LPD network of (Adler & Oktem, 2018) with subsets of it. We partition the forward and adjoint operators into 4 subsets, along with the corresponding measurement data. In each layer, we use only one of the subsets in a cycling order. We create four operators using these geometries and replace the forward and adjoint operators with a list of these sampled forward and adjoint operators. We utilize 12 DualNet and PrimalNet blocks, with a batch size of 1, and maintain the same hyperparameters,

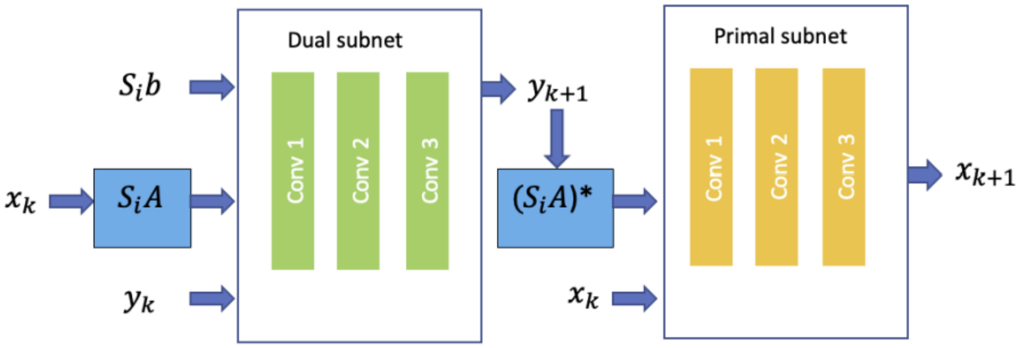

FIGURE 2.4: An example of a layer of LSPD network.
source: https://arxiv.org/pdf/2208.14784.pdf

loss function, and optimizer as in Experiment 2. The LSPD network is trained for 15 epochs. For implementing the Sk-LSPD, we use sampled forward and adjoint

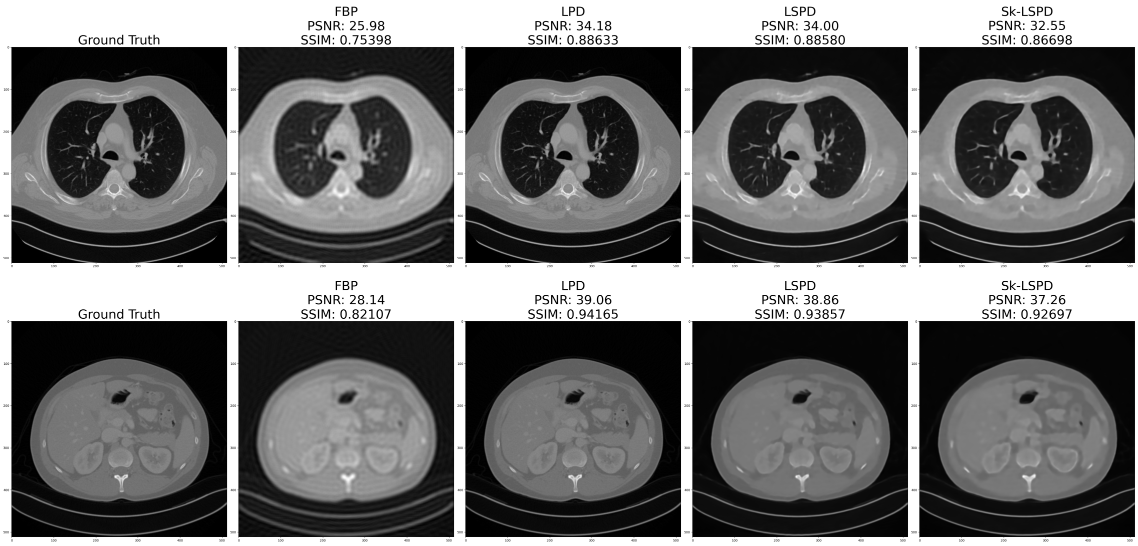

FIGURE 2.5: Examples for Low-dose CT. We can observe that LSPD networks achieves almost the same reconstruction performance as the full-batch LPD .

operators from the LSPD network, with the modification that the input dimensions are reduced to 256x256 instead of 512x512. We use sk-LSPD (option 2) with the coarse-to-fine sketch size. We utilize 12 DualNet and PrimalNet blocks, out of which the last four layers use unsketched LSPD forward and adjoint operators. For the upsampling/downsampling in our Sketched LSPD, we choose the bilinear upsample and downsample functions in PyTorch. While the full forward operator A is defined on the grid of 512x512, the sketched operator As is defined on the grid of 256x256, requiring only quarter of the computation in this setting. The initial guess x_0 is set to be the standard filtered backprojection for all the unrolling networks. We train the network for 15 epochs with a batch size equal to 1.

	PSNR	SSIM	Training time per epoch	Inference time per sample
FBP	28.06 dB	0.82329	-	-
LPD	38.58 dB	0.94332	16.77 min	0.1821 s
LSPD	38.53 dB	0.94281	16.33 min	0.1703 s
Sk-LSPD	36.93 dB	0.93183	9.03 min	0.1178 s

Table 2.1: Low-dose CT testing results for LPD, LSPD and SkLSPD networks on Mayo dataset

2.4 Sk-LSPD on Mayo Clinic CT dataset corrupted with Poisson noise

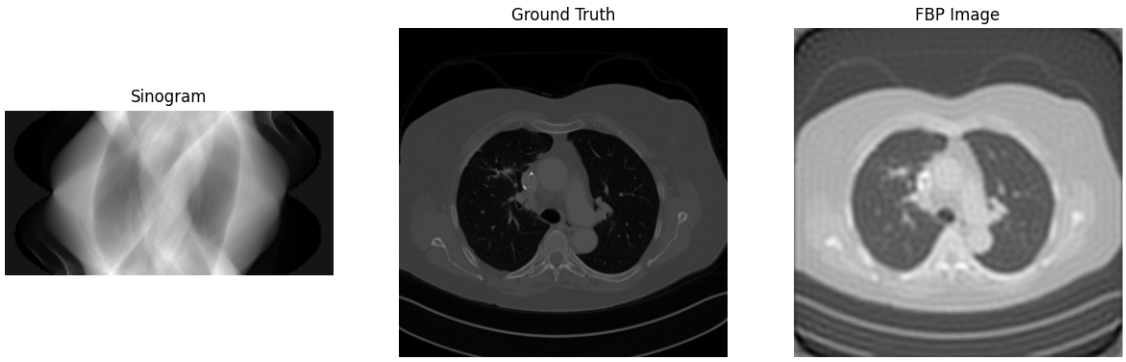

Figure 2.6: A sample from the low-dose, sparse view Dataset

In real-world clinical practice, low-dose CT is widely employed and strongly advocated due to the potential risks associated with high exposure to X-rays, which could substantially elevate the risk of cancer induction. Low-dose CT involves acquiring a substantial number of low-energy X-ray views, resulting in vast volumes of noisy measurements. This poses challenges for reconstruction schemes, making it difficult to achieve efficient and accurate estimations. To simulate such CT scans, we corrupt

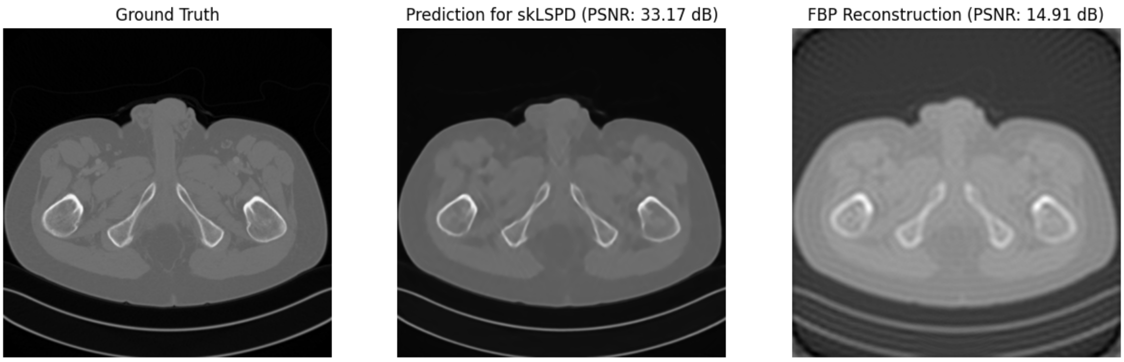

FIGURE 2.7: An example of Sk-LSPD network performance in sparse and low dose CT setting

the parallel-beam CT measurement with Poisson noise, $b : Poisson(I_0 exp(-Ax^*)$, where we make a low-dose choice of $I_0 = 7 * 10^4$. We used sparse-view dataset used in experiment 3 and corrupted it with poisson noise to simulate low-dose CT images. We train the sk-LSPD network on this dataset for 15 epochs with MSE loss.

Chapter 3

Conclusions

Through our experiments, we have demonstrated the competitiveness of the sketched and stochastic variants of LPD in terms of reconstructed image quality when compared to the full-batch version. While we were able to achieve comparable results, our focus was primarily on 2D images, limiting our ability to showcase the potential reduction in forward pass time. In the 2D setting, the convolutional networks dominate the overall computation, yet we anticipate that these variants will significantly decrease forward pass time in the computationally intensive 3D setting.

Looking ahead, our future objectives will center on building a 3D proof of concept for stochastic unrolling and sketching. We aim to extend this approach to more general problems, particularly focusing on designing ML-based solvers for large-scale optimization problems. By exploring these avenues, we anticipate unlocking the full potential of these techniques in addressing the challenges of 3D reconstruction and beyond.

Bibliography

Adler, J. and Oktem, O. (2018). Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332.

Mukherjee, S., Hauptmann, A., Öktem, O., Pereyra, M., and Schönlieb, C.-B. (2023). Learned reconstruction methods with convergence guarantees: A survey of concepts and applications. *IEEE Signal Processing Magazine*, 40(1):164–182.

Tang, J., Mukherjee, S., and Schönlieb, C.-B. (2022a). Accelerating deep unrolling networks via dimensionality reduction.

Tang, J., Mukherjee, S., and Schönlieb, C.-B. (2022b). Stochastic primal-dual deep unrolling.