Proximal Neural Networks based reconstruction for few-view CT applications

H. T. V. Le[1], C. Bossuyt[2], J. Escoda[1], M. Costin[1], H. Wang[1], J. De Beenhouwer[2], J. Sijbers[2]

[1]Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France, e-mail: hoang-trieuvy.le@cea.fr

[2]IMEC - Vision Lab, Department of Physics, University of Antwerp, 2610 Antwerpen, Belgium

Abstract

This paper addresses the challenge of tomographic reconstruction from a limited number of views by using learning-based approaches. Recent advancements in Plug-and-Play (PnP) algorithms have shown promise for solving imaging inverse problems by utilizing the capabilities of Gaussian denoising algorithms to handle complex optimization tasks. Traditional denoising hand-crafted methods produce images with predictable features but require intricate parameter tuning and suffer from slow convergence. In contrast, learning-based models offer faster performance and higher reconstruction quality, although they lack interpretability. In this work, we propose training proximal neural networks (PNNs) to eliminate arbitrary artifacts and improve the performances of PnP algorithms. These networks are obtained by unrolling a proximal algorithm designed to find a maximum a posteriori (MAP) estimate, but within a fixed number of iterations, using learned linear operators. PNNs, grounded in optimization theory, offer flexibility and can be adapted to any image restoration task solvable by a proximal algorithm. Additionally, they feature much simpler architectures compared to traditional neural networks.

Keywords: inverse problem, PNNs, Plug-and-Play, computed tomography, sparse view reconstruction.

1 Introduction

X-ray computed tomography (XCT) is now a well established technology with multiple uses in various domains. The typical implementation and usage require a relatively large input number of projections, for recent experimental systems up to few thousands of projections. However, for some particular applications, such as robotic CT [1] or static CT setups for characterization or security inspection [2], the available data can be as low as few tens of projections.

Traditionally, conventional reconstruction methods such as Filtered BackProjection (FBP) can produce high-quality CT images given a complete set of projections. However, in the framework of sparse-view CT reconstruction, FBP reconstruction leads to strong artifacts. Standard approach to this ill-posed problem is model-based optimization that integrates a forward model characterizing the image system and a regularizer imposing priors on the image. From this, we can solve the problem by using various splitting algorithms such as Alternating Direction Method of Multipliers (ADMM) [24], Forward-Backward Splitting [25] and recent studies demonstrated that Plug-and-Play methods effectively leverage the performance and convergence.

Initially, the main idea of PnP methods was introduced by [13], which implicitly handle the image priors by a powerful Gaussian denoiser such as BM3D [12]. In the recent years, although there are many progresses in algorithms that leverage sophisticated image priors such as transform-domain sparsity, learned dictionaries, more powerful DL denoisers have been introduced in numerous studies [14]. More recently, Deep generative models such as Generative Adversarial Networks (GANs) [22] or diffusion models [23] have also demonstrated the ability to handle those ill-posed inverse problem due to their ability to model complex distribution but at the cost of high volume requirement of available datasets. Unlike traditional denoisers, convolutional neural networks based Gaussian denoisers can handle various noise with a better computation time and better performance but the architectures usually have high cost in training phase. Furthermore, using these denoisers can make PnP lose interpretability as optimization problems as most image denoisers do not have optimization formulation.

Recently, PNNs have been also explored in this context of PnP [17], PnP methods using PNNs have been extensively evaluated and applied in image restoration (e.g. [5, 11]) and in CT image reconstruction [19]. In [17], the authors have presented a unified framework for building denoising PNNs with learned linear operators. PNNs have similar denoising performances to DL denoisers, although being much lighter ($\sim$ 1000 less parameters) and they are generally more robust in the PnP approaches.

Inspired by the ability of PnP methods to utilize off-the-shelf denoisers as an implicit prior, in this work, we study PNNs in the context of CT reconstruction problem. While the data term can be solved independently, we improve the ability of denoiser priors by training PNNs to eliminate arbitrary artifacts present in CT images.

Notation – In the remainder of this paper we will use the following notations. An element of $\mathbb{R}^N$ is denoted by x. For every $n \in \{1, \ldots, N\}$, the n-th coefficient of x is denoted by $\mathrm{x}^{(n)}$. The spectral norm is denoted $\|\cdot\|_S$. Let $C \subset \mathbb{R}^N$ be a closed, non-empty, convex set. The indicator function of C is denoted by ι_C, and is equal to 0 if its argument belongs to C, and $+\infty$ otherwise. Let $\mathrm{x} \in \mathbb{R}^N$. The Euclidean projection of x onto C is denoted by $\mathrm{P}_C(\mathrm{x}) = \operatorname*{argmin}_{\mathrm{v} \in C} \|\mathrm{v} - \mathrm{x}\|^2$. Let $\psi \colon \mathbb{R}^N \to (-\infty, +\infty]$ be a convex, lower semicontinuous, proper function. The proximity operator of ψ at x is given by $\mathrm{prox}_\psi(\mathrm{x}) = \operatorname*{argmin}_{\mathrm{v} \in \mathbb{R}^N} \psi(\mathrm{v}) + \frac{1}{2}\|\mathrm{v} - \mathrm{x}\|^2$.

The Fenchel-Legendre conjugate function of ψ is given by $\psi^*(\mathrm{x}) = \sup_{\mathrm{v} \in \mathbb{R}^N} \mathrm{v}^\top \mathrm{x} - \psi(\mathrm{v})$. When $\psi = \lambda \|\cdot\|_1$ is the ℓ_1 norm, then

$\psi^* = \iota_{\mathscr{B}_\infty(0,\lambda)}$ corresponds to the indicator function of the ℓ_∞-ball centred in 0 with radius $\lambda > 0$, i.e., $\mathscr{B}_\infty(0,\lambda) = \{x \in \mathbb{R}^N \mid (\forall n \in \{1,\dots,N\}) -\lambda \leq x^{(n)} \leq \lambda\}$.

2 Background and inverse problem formulation

CT reconstruction aims to find an estimate of an image $\bar{x}$, from noisy measurements $y \in \mathbb{R}^M$. The present contribution focuses on the the following general inverse problem:

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

where $A \in \mathbb{R}^{M \times N}$ is the forward operator and n is the measurement noise. Since Sparse-View CT reconstruction is an ill-posed inverse problem, the prior which is also called regularisation need to be adopted to constrain the solution space. From Bayesian perspective, a common method to reconstruct x is to rely on a maximum *a posteriori* (MAP) approach,

$$\hat{x}_{MAP} = \underset{x}{\text{argmax}} \ \log p(y \mid x) + \log p(x),$$

where $\log p(y \mid x)$ represents the log-likelihood of the observation y and $\log p(x)$ delivers the priors of x. More formally, we can define the estimate $\hat{x}_{MAP} \in \mathbb{R}^N$ as a minimizer of a penalized least-squares objective function. A general formulation of this problem is to find

$$\hat{x}_{MAP} = \underset{x \in \mathbb{R}^N}{\text{argmin}} \ F(x), \tag{2}$$

where

$$F(x) := \frac{1}{2}\|Ax - y\|_2^2 + \lambda g(Dx) + \iota_C(x), \tag{3}$$

$C \subset \mathbb{R}^N$ is a closed, convex, non-empty constraint set, $\lambda > 0$ is a regularisation parameter, $D : \mathbb{R}^N \mapsto \mathbb{R}^{|\mathbb{F}|}$ is a linear operator mapping an image from $\mathbb{R}^N$ to a feature space $\mathbb{R}^{|\mathbb{F}|}$, and $g : \mathbb{R}^{|\mathbb{F}|} \to (-\infty, +\infty]$ denotes a proper, lower-semicontinous, convex function. To be specific, the data term ensures that the solution adheres to the degradation process, while the prior term enforces the solution to follow the desired data distribution. The function g and the operator D are chosen according to the type of images of interest. For instance, functions of choice for piece-wise constant images are those in the family of total variation (TV) regularizations [8], which can be expressed as an ℓ_1 (or an $\ell_{1,2}$) norm composed with a linear operator performing horizontal and vertical finite differences of the image. More generally, D can be chosen as a sparsifying operator (e.g., wavelet transform [6, 7, 16]), and g as a function promoting sparsity (e.g., ℓ_1). For any choice of g and D, the parameter $\lambda > 0$ is used to balance the penalization term (i.e., function $g \circ D$) with the data-fidelity term (i.e., least-squares function).

Model-based methods – Traditionally, (3) can be solved by using model-based optimization methods with proximal splitting methods. These methods are suitable for handling composite objectives but usually involve a time-consuming iterative inference. As example, in this work, we focus on the forward-backward splitting method which is given by:

$$\begin{aligned}
&\text{Let } x_0 \in \mathbb{R}^N, \\
&\text{For } t = 0, 1, \dots \\
&\quad \left\lfloor \begin{aligned}
&x_{t+\frac{1}{2}} = x_t - \gamma P_C\left(A^\top(Ax_t - y)\right), \\
&x_{t+1} = \text{prox}_{\gamma \lambda g(Dx)}(x_{t+\frac{1}{2}}),
\end{aligned} \right.
\end{aligned} \tag{4}$$

Plug-and-Play Priors – The idea of PnP algorithms is to replace the penalization term (often handled by a proximity operator) by a powerful denoiser. There are multiple choices of denoisers, that can be classified into two main categories: hand-crafted denoisers (e.g. BM3D [12]) and learning-based denoisers (e.g., DnCNN [14] and UNet [15]). In this work, we focus on the widely adopted approach of denoising a noisy image z through a maximum *a posteriori* (MAP) estimation framework. The MAP estimate $x^\dagger_{MAP} \in \mathbb{R}^N$ is defined as the solution to a penalized least-squares optimization problem. Specifically, it is formulated as:

$$x^\dagger_{MAP} = \underset{x \in \mathbb{R}^N}{\text{argmin}} \ F(x), \tag{5}$$

where

$$F(x) := \frac{1}{2}\|x - z\|_2^2 + \nu g(Dx) + \iota_C(x), \tag{6}$$

$C \subset \mathbb{R}^N$ is a closed, convex, non-empty constraint set, $\nu > 0$ is a regularization parameter proportional to δ^2, $\mathrm{D}\colon \mathbb{R}^N \to \mathbb{R}^{|\mathbb{F}|}$ is a linear operator mapping an image from $\mathbb{R}^N$ to a feature space $\mathbb{R}^{|\mathbb{F}|}$, and $g\colon \mathbb{R}^{|\mathbb{F}|} \to (-\infty, +\infty]$ denotes a proper, lower-semicontinous, convex function.

Hence the objective of this section is to follow a similar approach as in [11], and to plug the proposed unfolded PNNs in a FB algorithm to solve (3). Following the approach proposed in [11], the PnP-FB algorithm is given by

$$
\begin{aligned}
&\text{Let } \mathrm{x}_0 \in \mathbb{R}^N, \mathrm{u}_0 \in \mathbb{R}^{|\mathbb{F}|} \\
&\text{For } t = 0, 1, \ldots \\
&\left|
\begin{aligned}
&\mathrm{y}_t = \mathrm{x}_t - \gamma \mathrm{A}^\top (\mathrm{A}\mathrm{x}_t - \mathrm{y}), \\
&(\mathrm{x}_{t+1}, \mathrm{u}_{t+1}) = f^K_{\mathrm{y}_t, \lambda\gamma, \Theta}(\mathrm{y}_t, \mathrm{u}_t),
\end{aligned}
\right.
\end{aligned}
\tag{7}
$$

where, for every $t \in \mathbb{N}$, $f^K_{\mathrm{y}_t, \lambda\gamma, \Theta}$ is the denoising model which is designed to approximate the proximity operator associated with the regularization term and will be discussed in the next section. In algorithm (7), parameters (λ, γ) are given as inputs of $f^K_{\mathrm{y}_t, \lambda\gamma, \Theta}$. Precisely, the regularization parameter ν for the denoising problem (5) is chosen to be the product between the regularization parameter λ for the restoration problem (3) and the stepsize of the algorithm γ, i.e., $\nu = \lambda\gamma$.

3 Methodology

In this section, we describe the chosen architecture and the training procedure of our Learned Sparse-View Proximal Operator approach and compare it with the state-of-the-art methods on the task of 2D sparse-view fan-beam computed tomography. Our goal is to adopt the unfolded denoising Proximal Neural Networks then fine tuning it to adapt with sparse-view data.

3.1 Deep Strongly Convex Chamboll-Pock with Learned Normalized Operator (DScCP-LNO)

A strategy to solve (5)-(6) consists in applying the Strongly Convex Chambolle-Pock (ScCP) algorithm to problem (5). The associated iterations can be reformulated as:

$$
\begin{aligned}
&\text{for } k = 0, 1, \ldots \\
&\left|
\begin{aligned}
&\mathrm{x}_{k+1} = \mathrm{P}_C\left(\frac{\mu_k}{1+\mu_k}(\mathrm{z} - \mathrm{D}^\top \mathrm{u}_k) + \frac{1}{1+\mu_k}\mathrm{x}_k\right), \\
&\mathrm{u}_{k+1} = \mathrm{prox}_{\tau_k(\nu g)^*}\left(\mathrm{u}_k + \tau_k \mathrm{D}\left((1+\alpha_k)\mathrm{x}_{k+1} - \alpha_k \mathrm{x}_k\right)\right),
\end{aligned}
\right.
\end{aligned}
\tag{8}
$$

where $\mathrm{x}_0 \in \mathbb{R}^N$ and $\mathrm{u}_0 \in \mathbb{R}^{|\mathbb{F}|}$, P_C is the projection function onto constraint set C. Note that when, for every $k \in \mathbb{N}$, $\alpha_k = 1$, then algorithm (8) reduces to standard iterations of the primal-dual Chambolle-Pock algorithm [9]. When $\alpha_k = (1 + 2\zeta\mu_k)^{-1/2}$ it leads to ScCP and when $\alpha_k = 0$, it leads to the classical Arrow-Hurwicz algorithm [10].

Deep Strongly convex Chambolle-Pock denoising model – The unfolding DScCP building block is then given below:

$$
f^{K, \mathrm{DScCP}}_{\mathrm{z}, \nu, \Theta}(\mathrm{x}_0, \mathrm{u}_0) = \mathrm{L}^{\mathrm{DScCP}}_{\mathrm{z}, \nu, \Theta_K} \circ \cdots \circ \mathrm{L}^{\mathrm{DScCP}}_{\mathrm{z}, \nu, \Theta_1}(\mathrm{x}_0, \mathrm{u}_0),
\tag{9}
$$

where, for every $k \in \{1, \ldots, K\}$,

$$
\begin{aligned}
\mathrm{u}_k &= \mathrm{L}_{\nu, \Theta_{k,\mathscr{D}}, \mathscr{D}}(\mathrm{x}_{k-1}, \mathrm{u}_{k-1}) = \eta_{\nu,k,\mathscr{D}}\left(\mathrm{W}_{k,\mathscr{D}}\mathrm{x} + \mathrm{V}_{k,\mathscr{D}}\mathrm{u} + \mathrm{b}_{k,\mathscr{D}}\right) \\
\mathrm{x}_k &= \mathrm{L}_{\mathrm{z}, \Theta_{k,\mathscr{P}}, \mathscr{P}}(\mathrm{x}_{k-1}, \mathrm{u}_k) = \eta_{k,\mathscr{P}}\left(\mathrm{W}_{k,\mathscr{P}}\mathrm{x} + \mathrm{V}_{k,\mathscr{P}}\mathrm{u} + \mathrm{b}_{k,\mathscr{P}}\right) \\
(\widetilde{\mathrm{x}}_k, \mathrm{u}_k) &= \mathrm{L}^{\mathrm{DScCP}}_{\mathrm{z}, \nu, \Theta_k}(\mathrm{x}_{k-1}, \mathrm{u}_{k-1}) \\
\mathrm{x}_k &= (1 + \alpha_k)\widetilde{\mathrm{x}}_k - \alpha_k \mathrm{x}_{k-1}.
\end{aligned}
$$

with

$$
\begin{cases}
\mathrm{W}_{k,\mathscr{D}} = \tau_k \mathrm{D}_{k,\mathscr{D}}, \\
\mathrm{V}_{k,\mathscr{D}} = \mathrm{Id}, \\
\mathrm{b}_{k,\mathscr{D}} = 0, \\
\eta_{\nu,k,\mathscr{D}} = \mathrm{prox}_{\tau_k(\nu g)^*},
\end{cases}
\quad \text{and} \quad
\begin{cases}
\mathrm{W}_{k,\mathscr{P}} = \frac{1}{1+\mu_k}, \\
\mathrm{V}_{k,\mathscr{P}} = -\frac{\mu_k}{1+\mu_k}\mathrm{D}_{k,\mathscr{P}}, \\
\mathrm{b}_{k,\mathscr{P}} = \frac{\mu_k}{1+\mu_k}\mathrm{z}, \\
\eta_{k,\mathscr{P}} = \mathrm{P}_C.
\end{cases}
$$

In this work, we employed the **Learned Normalized Operators (LNO)** framework, as detailed in [5], due to its better performance. Specifically, for each $k \in \{1, \ldots, K\}$, the operators $\mathrm{D}_{k,\mathscr{P}}$ were defined to be equal to the adjoints $\mathrm{D}^\top_{k,\mathscr{D}}$ of $\mathrm{D}_{k,\mathscr{D}}$. These operators, along with τ_k and μ_k, were included as learnable parameters in $\Theta_{k,\mathscr{D}}$ and $\Theta_{k,\mathscr{P}}$. Notably, $\mathrm{D}_{k,\mathscr{D}}$ are allowed to differ across the layers $k \in \{1, \ldots, K\}$. In pratice, the operators $D_{k,\mathscr{P}}\colon \mathbb{R}^N \to \mathbb{R}^{|\mathbb{F}|}$ are equivalent to 2D convolution operator with kernel size 3×3 over an input image for each layer.

3.2 Pre-trained Gaussian denoising model

We consider two sets of images: the *training set* $(\overline{x}_s, z_s)_{s \in \mathbb{I}}$ of size $|\mathbb{I}|$ and the *test set* $(\overline{x}_s, z_s)_{s \in \mathbb{J}}$ of size $|\mathbb{J}|$. For both sets, each couple $(\overline{x}_s, z_s)$ consists of a clean image $\overline{x}_s$ of size N_s, and a noisy version of this image given by $z_s = \overline{x}_s + \varepsilon_s$ with $\varepsilon_s \sim \mathcal{N}(0, \delta^2 \mathrm{Id})$ for $\delta > 0$.

Training strategy for unfolded denoising networks – The network parameters are optimized by minimizing the ℓ^2 empirical loss between noisy and ground-truth images:

$$\widehat{\Theta} \in \underset{\Theta}{\mathrm{Argmin}} \; \frac{1}{|\mathbb{I}|} \sum_{s \in \mathbb{I}} \mathscr{L}(\overline{x}_s, z_s; \Theta) \tag{10}$$

where

$$\mathscr{L}(\overline{x}_s, z_s; \Theta) := \frac{1}{2} \| \overline{x}_s - f^K_{z_s, \delta^2, \Theta} (z_s, D_{1, \mathscr{D}}(z_s)) \|^2,$$

and $f^K_{z_s, \delta^2, \Theta}$ the unfolded networks described in Section 3. The loss (10) will be optimized in Pytorch with Adam algorithm [18] and the standard DScCP is trained on variable noise, considering random values of $\delta_i \sim \mathscr{U}([0, 0.1])$ during the training procedure.

3.3 Leveraged by Sparse-View training

To leverage the ability of denoiser priors, we propose to apply another strategy of training process on the architecture on a sparse-view dataset. The learning set include reconstructions performed both in full-view and sparse-view configurations, all using the standard FBP method. Full-view slices use 1200 projections and they account as ground truth, while sparse-view cases are generated with a reduced subset of projections. More specifically, for each sample we generated sparse-view reconstructions by varying the number of projection views between 10 and 600, thus producing a range of under-sampled reconstructions. This selection process generated a training database consisting of sparse-view and full-view pairs. Our implementation uses tomography operators from ASTRA Toolbox [21].

We then consider two sets of images: the *training set* $(\overline{x}_s, z_s)_{s \in \mathbb{I}}$ of size $|\mathbb{I}|$ and the *test set* $(\overline{x}_s, z_s)_{s \in \mathbb{J}}$ of size $|\mathbb{J}|$. For both sets, each couple $(\overline{x}_s, z_s)$ consists of a full-view image reconstruction $\overline{x}_s$ of size N_s obtained from simulated projections without noise, and a sparse-view version of this image z_s obtained from a reduced number of simulated projections including acquisition noise.

Sparse-view training strategy for unfolded denoising networks – The network parameters are optimized by minimizing the ℓ^2 empirical loss between noisy and ground-truth images:

$$\widehat{\Theta} \in \underset{\Theta}{\mathrm{Argmin}} \; \frac{1}{|\mathbb{I}|} \sum_{s \in \mathbb{I}} \mathscr{L}(\overline{x}_s, z_s; \Theta) \tag{11}$$

where

$$\mathscr{L}(\overline{x}_s, y_s; \Theta) := \frac{1}{2} \| \overline{x}_s - f^K_{y_s, \beta, \Theta} (y_s, D_{1, \mathscr{D}}(z_s)) \|^2,$$

The loss (11) will be optimized in Pytorch with Adam algorithm [18] and hyperparameter β is chosen to be adapted with the number of projection views.
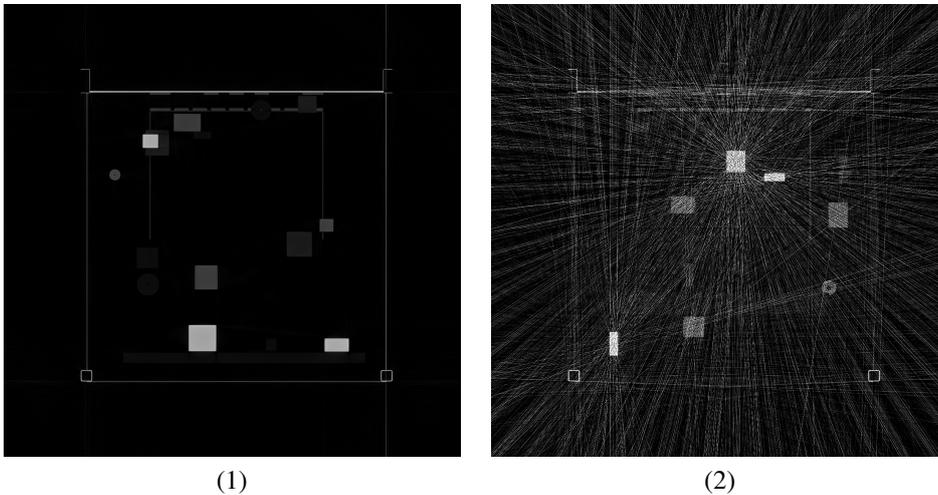
Figure 1: (1) The groundtruth $\overline{x}_s$ is computed using the FBP algorithm from the dense-view project data, (2) sparse-view reconstruction z_s with 30 projections is computed using FBP algorithm.

4 Experiments

4.1 Datasets

Walnut-CBCT dataset – Our first tests with the proposed algorithm were on a public data collection of Cone-Beam Computed Tomography (CB-CT) scans of walnuts, for which we gracefully thank the authors [20]. Each walnut is of size about 30mm. A source of 40 kV is used for acquisition. The detector is binned to 768×972 pixels, each of side 0.1496mm. During acquisition, source-object distance is of 66mm and that of object-detector is of 199mm. For each walnut, 1201 images are acquired. Since our aim is to develop a reconstruction method for fan-beam acquisition, we only consider the central slice of each scan. We split the dataset into a training set of 40 walnuts and a test set of 2 remaining samples.

W3D dataset – The cargo dataset is simulated using CIVA software [26] developed by CEA List (see Fig. 2). The dataset contains 13 cargos scans and their corresponding sinogram. Each cargo are identical (and of side around 2300mm) and includes 16 objects of varying size and shapes (based on parallelepipeds and hollow cylinders). Cargos are made of iron and each object inside is made of one of the following material: carbon, aluminum, iron, wood and water. Fan beam acquisition is simulated, with a source-axe distance of 9831mm and an axe-detector distance of 1907mm. A 6MeV source is considered, and the detector is of length 1200 pixels and pixel size of 3.2mm. For each cargo, 1200 angles views are simulated for the full views reconstruction, and the corresponding 2D scan (of size 1200×1200 pixels) is obtained using FBP algorithm. In this works, we split the cargos into a training set of 10 cargos and a test set of remaining cargos.
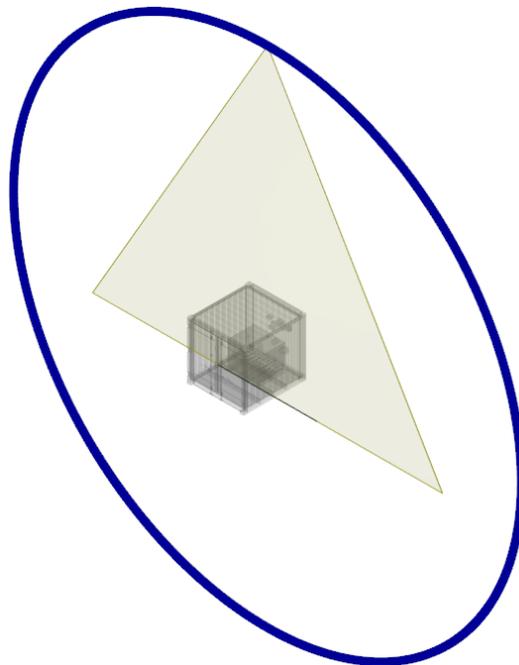
Figure 2: Visualization of cargo scanning simulation using CIVA software.

4.2 Training details and parameters

We train and evaluate the models with Pytorch using AD L4 GPUs using 5 Gb memory. To show the effectiveness of the method, we train our model with a batch-size of 3 in case of W3D dataset and batch-size of 4 in Walnut Dataset, we use the Adam optimizer with a fixed learning rate of 10^{-3}. We do not have any checkpoint selection strategy on our model and it takes about 2 hours to train a DSCCP-model with $K = 10$, $F = 12$ for both training phases.

4.3 Results

Walnut-CBCT Dataset – In this experiment, we evaluate the model's performance when trained exclusively on Gaussian noise perturbations applied to ground truth images. The test was conducted using the 42nd sample of the Walnut dataset. The model's performance was assessed by comparison to a TV-based reconstruction method [3], the ground truth, and a sparse-view FBP reconstruction. Quantitative comparisons are provided through the Peak Signal-to-Noise Ratio (PSNR) computed for the entire image. The results demonstrate that DScCP-LNO outperforms the TV-based method both qualitatively and quantitatively, with parameters carefully chosen to preserve as many critical image features as possible.
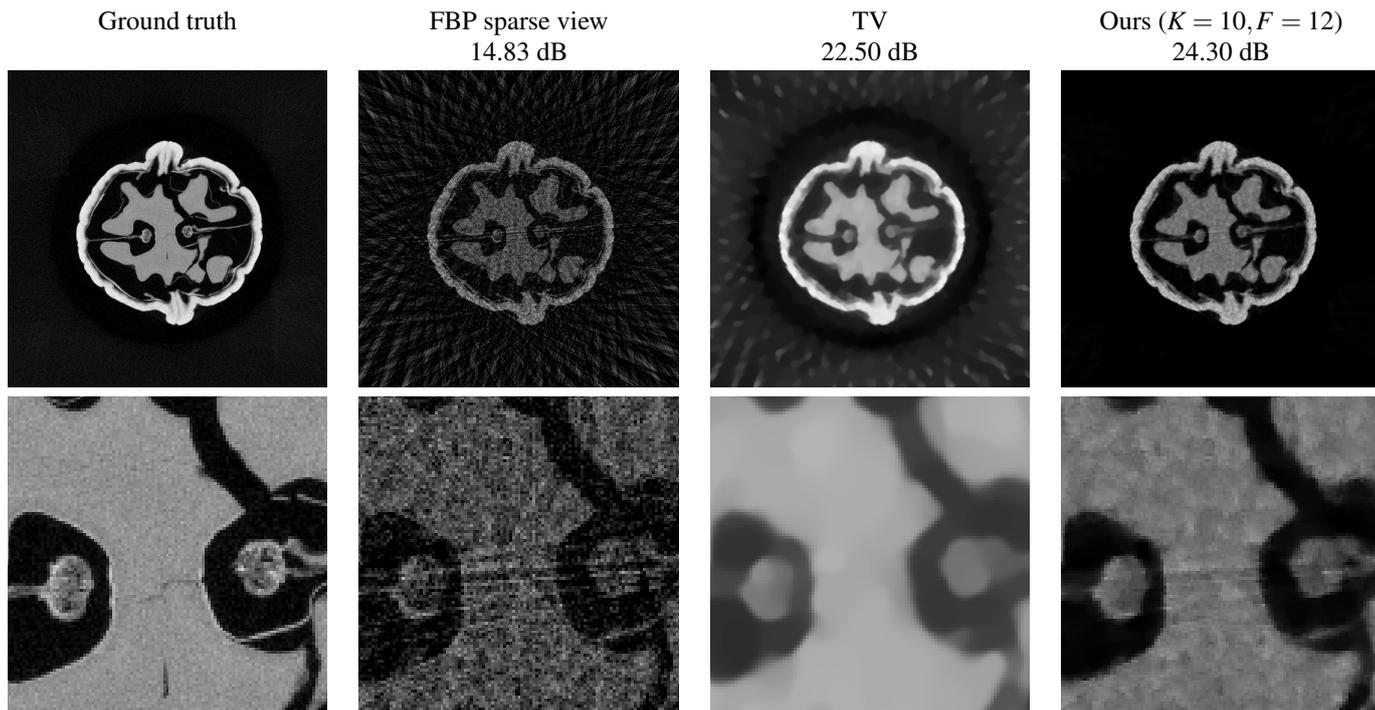
Figure 3: Qualitative and PSNR comparison for the central slice of 42th Walnut test sample with 30 projections.

W3D Dataset – This experiment focuses on a more realistic application scenario where artifacts are more prominent, posing greater challenges for accurate reconstruction. The test was conducted on the 11th sample of the dataset. The performance of the model was compared to a TV-based reconstruction method [3], the ground truth, and a sparse-view FBP reconstruction. Figure 4 illustrates the results for a sparse-view scenario with 30 projections and a reconstruction grid of 1200×1200 pixels. Zoomed-in regions from the image center, shown in the second row, highlight the qualitative differences. The model successfully suppresses artifacts while preserving essential features, surpassing alternative methods. Quantitative comparisons underscores the robustness of the proposed approach in challenging scenarios.
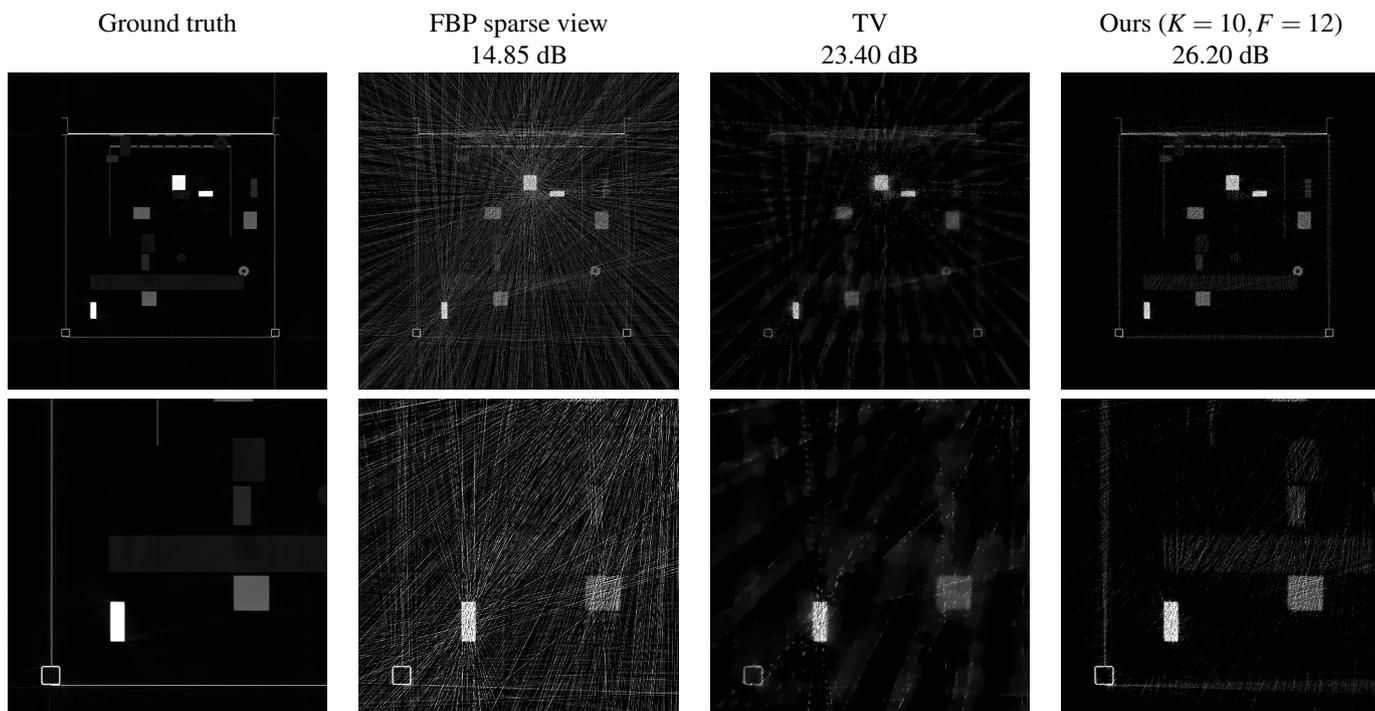
Figure 4: Qualitative and PSNR comparison for the 11th W3D test sample with 30 projections.

5 Discussion and conclusions

In this work, we propose training a tailored Proximal Neural Network (PNN) model, specifically the DScCP-LNO model [5], on sparse-view CT datasets. Our findings reveal that post-training the model on sparse-view datasets significantly enhances reconstruction quality by effectively reducing streak artifacts. Notably, the experiments demonstrate the model's ability to achieve robust CT reconstructions using minimal training data, as evidenced by its success with the limited W3D dataset (10 cargos). This underscores the potential of PNNs to deliver high-quality results even in data-constrained scenarios.

Compared to traditional Total Variation (TV) methods, our approach not only mitigates artifacts but also better preserves critical image details often lost with TV. This highlights the advantages of PNNs in balancing computational efficiency with superior reconstruction performance. The next phase of this research will extend the algorithm to static CT setups, where the source points are distributed across a rectangular support, further exploring its applicability in practical scenarios.

References

[1] M. Costin, J. Escoda, V. Bussy, J.S. Rathore, Robotized X-ray inspection of large industrial parts: non-standard trajectories and 3D reconstruction, WCNDT 2024, Seoul, South Korea (2024).

[2] C. Bossuyt, J. De Beenhouwer, J. Sijbers, Sparse view rectangular X-ray CT for cargo inspection using the ASTRA toolbox, 5th international Conference on Tomography of Materials and Structures, Grenoble, France (2022).

[3] E. Boigné, D. Y. Parkinson, M. Ihme. Towards data-informed motion artifact reduction in quantitative CT using piecewise linear interpolation. IEEE Transactions on Computational Imaging, 8 (2022) 917-932.

[4] R. Vo, J. Escoda, C. Vienne, E. Decencière, Neural Field Regularization by Denoising for 3D Sparse-View X-Ray Computed Tomography, 2024 International Conference on 3D Vision (3DV), Davos, Switzerland, (2024) 1166-1176.

[5] H. T. V. Le, A. Repetti, N. Pustelnik, Unfolded Proximal Neural Networks for Robust Image Gaussian Denoising, IEEE Transactions on Image Processing, 33 (2024) 4475-4487.

[6] Jacques, L., Duval, L., Chaux, C. & Peyré, G. A Panorama on Multiscale Geometric Representations, Intertwining Spatial, Directional and Frequency Selectivity. *Signal Processing*. **91**, 2699-2730 (2011)

[7] Pustelnik, N., Benazza-Benhayia, A., Zheng, Y. & Pesquet, J. Wavelet-based image deconvolution and reconstruction. *Wiley Encyclopedia Of EEE*. (2016)

[8] Rudin, L., Osher, S. & Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*. **60**, 259-268 (1992)

[9] Chambolle, A. & Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vis.*. **40**, 120-145 (2011)

[10] Arrow, K., Hurwicz, L. & Uzawa, H. Studies in linear and non-linear programming. *In: Chenery, H.B., Johnson, S.M., Karlin, S., Marschak, T., Solow, R.M. (eds.) Stanford Mathematical Studies In The Social Sciences, Vol. II.*. (1958)

[11] A. Repetti, M. Terris, Y. Wiaux, J. C. Pesquet, Dual forward-backward unfolded network for flexible plug-and-play, 2022 30th European Signal Processing Conference (EUSIPCO) IEEE, (2022) pp. 957-961.

[12] Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions On Image Processing*. **16**, 2080-2095 (2007)

[13] Venkatakrishnan, S., Bouman, C. & Wohlberg, B. Plug-and-play priors for model based reconstruction. *2013 IEEE Global Conference On Signal And Information Processing*. pp. 945-948 (2013)

[14] Zhang, K., Li, Y., Zuo, W., Zhang, L., Van Gool, L. & Timofte, R. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. **44**, 6360-6376 (2021)

[15] Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing And Computer-assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. pp. 234-241 (2015)

[16] Mallat, S. A wavelet tour of signal processing. (Academic Press,1997)

[17] Le, H., Repetti, A. & Pustelnik, N. Unfolded proximal neural networks for robust image Gaussian denoising. *IEEE Transactions On Image Processing*. (2024)

[18] Kingma, D. & Ba, J. Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*. (2014)

[19] M. Savanier, E. Chouzenoux, J. C. Pesquet, C. Riddell, Deep unfolding of the DBFB algorithm with application to ROI CT imaging with limited angular density, IEEE Transactions on Computational Imaging, 9 (2023) 502-516.

[20] H. Der Sarkissian, F. Lucka, M. van Eijnatten, G. Colacicco, S. B. Coban, K. J. Batenburg, A cone-beam X-ray computed tomography data collection designed for machine learning. Scientific data, 6(1) (2019) 215.

[21] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, J. Sijbers, Fast and Flexible X-ray Tomography Using the ASTRA Toolbox, Optics Express, 24(22) (2016) 25129-25147.

[22] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. Generative adversarial networks. *Communications Of The ACM*. **63**, 139-144 (2020)

[23] Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Advances In Neural Information Processing Systems*. **33** pp. 6840-6851 (2020)

[24] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J. & Others Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations And Trends® In Machine Learning*. **3**, 1-122 (2011)

[25] Combettes, P. & Wajs, V. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*. **4**, 1168-1200 (2005)

[26] M. Costin, D. Tisseur, C. Vienne, R. Guillamet, H. Banjak, N. Bhatia and R. Fernandez, "CIVA CT, an advanced simulation platform for NDT," 6th Conference on Industrial Computed Tomography, Wels, Austria, 2016