QN-Mixer: A Quasi-Newton MLP-Mixer Model for Sparse-View CT Reconstruction

Ishak Ayad[1,2*†] Nicolas Larue[1,3†] Maï K. Nguyen[1]
[1]ETIS (UMR 8051), CY Cergy Paris University, ENSEA, CNRS, France
[2]AGM (UMR 8088), CY Cergy Paris University, CNRS, France
[3]University of Ljubljana, Slovenia
ishak.ayad@cyu.fr

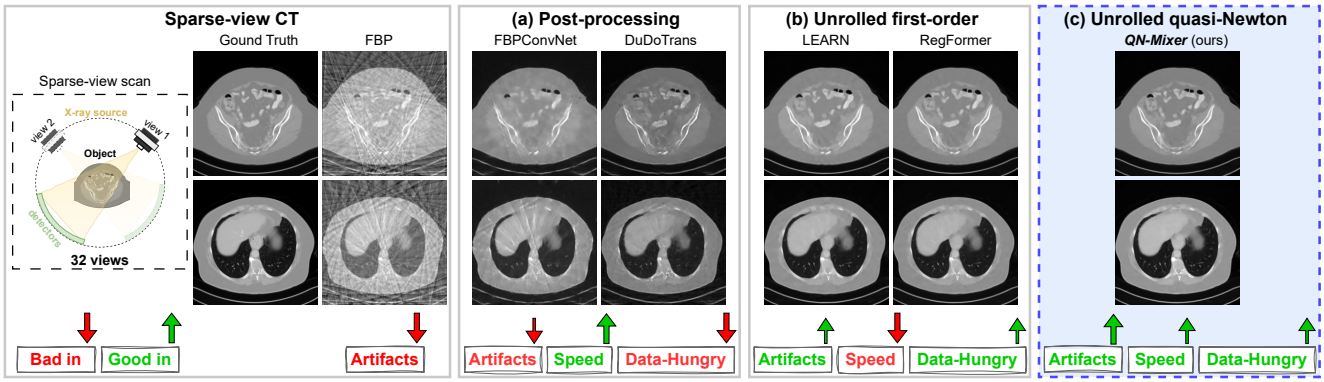

Figure 1. **CT Reconstruction with 32 views of State-of-the-Art Methods**. Comparative analysis with post-processing and first-order unrolling networks highlights QN-Mixer's superiority in artifact removal, training time, and data efficiency.

Abstract

Inverse problems span across diverse fields. In medical contexts, computed tomography (CT) plays a crucial role in reconstructing a patient's internal structure, presenting challenges due to artifacts caused by inherently ill-posed inverse problems. Previous research advanced image quality via post-processing and deep unrolling algorithms but faces challenges, such as extended convergence times with ultra-sparse data. Despite enhancements, resulting images often show significant artifacts, limiting their effectiveness for real-world diagnostic applications. We aim to explore deep second-order unrolling algorithms for solving imaging inverse problems, emphasizing their faster convergence and lower time complexity compared to common first-order methods like gradient descent. In this paper, we introduce QN-Mixer, an algorithm based on the quasi-Newton approach. We use learned parameters through the BFGS algorithm and introduce Incept-Mixer, an efficient neural architecture that serves as a non-local regularization term, capturing long-range dependencies within images. To address the computational demands typically associated with quasi-Newton algorithms that require full Hessian matrix computations, we present a memory-efficient alternative. Our approach intelligently downsamples gradient information, significantly reducing computational requirements while maintaining performance. The approach is validated through experiments on the sparse-view CT problem, involving various datasets and scanning protocols, and is compared with post-processing and deep unrolling state-of-the-art approaches. Our method outperforms existing approaches and achieves state-of-the-art performance in terms of SSIM and PSNR, all while reducing the number of unrolling iterations required.

1. Introduction

Computed tomography (CT) is a widely used imaging modality in medical diagnosis and treatment planning, delivering intricate anatomical details of the human body with precision. Despite its success, CT is associated with high radiation doses, which can increase the risk of cancer induction [50]. Adhering to the ALARA principle (As Low As Reasonably Achievable) [37], the medical community emphasizes minimizing radiation exposure to the

*Corresponding author. † Equal contribution.

lowest level necessary for accurate diagnosis. Numerous approaches have been proposed to reduce radiation doses while maintaining image quality. Among these, sparse-view CT emerges as a promising solution, effectively lowering radiation doses by subsampling the projection data, often referred to as the sinogram. Nonetheless, reconstructed images using the well-known Filtered Back Projection (FBP) algorithm [34], suffer from pronounced streaking artifacts (see Fig. 1), which can lead to misdiagnosis. The challenge of effectively reconstructing high-quality CT images from sparse-view data is gaining increasing attention in both the computer vision and medical imaging communities.

With the success of deep learning spanning diverse domains, initial image-domain techniques [6, 19, 25, 28, 59] have been introduced as post-processing tasks on the FBP reconstructed images, exhibiting notable accomplishments in artifact removal and structure preservation. However, the inherent limitations of these methods arise from their constrained receptive fields, leading to challenges in effectively capturing global information and, consequently, suboptimal results.

To address this limitation, recent advances have seen a shift toward a dual-domain approach [18, 27, 29, 49], where post-processing methods turn to the sinogram domain. In this dual-domain paradigm, deep neural networks are employed to perform interpolation tasks on the sinogram data [15, 24], facilitating more accurate image reconstruction. Despite the significant achievements of post-processing and dual-domain methods, they confront issues of interpretability and performance limitations, especially when working with small datasets and ultra-sparse-view data, as shown in Fig. 1. To tackle these challenges, deep unrolling networks have been introduced [1, 7, 8, 11, 16, 20, 51, 54]. Unrolling networks treat the sparse-view CT reconstruction problem as an optimization task, resulting in a first-order iterative algorithm like gradient descent, which is subsequently unrolled into a deep recurrent neural network in order to learn the optimization parameters and the regularization term. Like post-processing techniques, unrolling networks have been extended to the sinogram domain [52, 56] to perform interpolation task.

Unrolling networks, as referenced in [12, 36, 44], exhibit remarkable performance across diverse domains. However, they suffer from slow convergence and high computational costs, as illustrated in Fig. 1, necessitating the development of more efficient alternatives [14]. More specifically, they confront two main issues: *Firstly*, they frequently grapple with capturing long-range dependencies due to their dependence on locally-focused regularization terms using CNNs. This limitation results in suboptimal outcomes, particularly evident in tasks such as image reconstruction. *Secondly*, the escalating computational costs of unrolling methods align with the general trend of increased complexity in modern neural networks. This escalation not only amplifies the required number of iterations due to the algorithm's iterative nature but also contributes to their high computational demand.

To tackle the aforementioned issues, we introduce a novel second-order unrolling network for sparse-view CT reconstruction. *In particular*, to enable the learnable regularization term to apprehend long-range interactions within the image, we propose a non-local regularization block termed **Incept-Mixer**. Drawing inspiration from the multi-layer perceptron mixer [46] and the inception architecture [45], it is created to combine the best features from both sides: capturing long-range interactions from the attention-like mechanism of MLP-Mixer and extracting local invariant features from the inception block. This block facilitates a more precise image reconstruction. *Second*, to cut down on the computational costs associated with unrolling networks, we propose to decrease the required iterations for convergence by employing second-order optimization methods such as [21, 30]. We introduce a novel unrolling framework named **QN-Mixer**. Our approach is based on the quasi-Newton method that approximate the Hessian matrix using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update [10, 13, 57]. Furthermore, we reduce memory usage by working on a projected gradient (latent gradient), preserving performance while reducing the computational cost tied to Hessian matrix approximation. This adaptation enables the construction of a deep unrolling network, showcasing superlinear convergence. Our contributions are summarized as follows:

- We introduce a novel second-order unrolling network coined **QN-Mixer** where the Hessian matrix is approximated using a latent BFGS algorithm with a deep-net learned regularization term.

- We propose **Incept-Mixer**, a neural architecture acting as a non-local regularization term. Incept-Mixer integrates deep features from inception blocks with MLP-Mixer, enhancing multi-scale information usage and capturing long-range dependencies.

- We demonstrate the effectiveness of our proposed method when applied to the sparse-view CT reconstruction problem on an extensive set of experiments and datasets. We show that our method outperforms state-of-the-art methods in terms of quantitative metrics while requiring less iterations than first-order unrolling networks.

2. Related Works

In this section, we present prior work closely related to our paper. We begin by discussing the general framework for unrolling networks in Sec. 2.1, which is based on the

gradient descent algorithm. Subsequently, in Sec. 2.2 and Sec. 2.3, we delve into state-of-the-art methods in post-processing and unrolling networks, respectively.

2.1. Background

Inverse Problem Formulation for CT. Image reconstruction problem in CT can be mathematically formalized as the solution to a linear equation in the form of:

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

where $\boldsymbol{x} \in \mathbb{R}^n$ is the (unknown) object to reconstruct with $n = h \times w$, $\boldsymbol{y} \in \mathbb{R}^m$ is the data (i.e. sinogram), where $m = n_v \times n_d$, n_v and n_d denote the number of projection views and detectors, respectively. $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ is the forward model (i.e. discrete Radon transform [40]). The goal of CT image reconstruction is to recover the (unknown) object, $\boldsymbol{x}$, from the observed data $\boldsymbol{y}$. As the problem is ill-posed due to the missing data, the linear system in Eq. (1) becomes underdetermined and may have infinite solutions. Hence, reconstructed images suffer from artifacts, blurring, and noise. To address this issue, iterative reconstruction algorithms are utilized to minimize a regularized objective function with a L^2 norm constraint:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} J(\boldsymbol{x}) = \frac{\lambda}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \mathcal{R}(\boldsymbol{x}), \tag{2}$$

where $\mathcal{R}(\boldsymbol{x})$ is the regularization term, balanced with the weight λ. Those ill-posed problems were initially addressed using optimization techniques, such as the truncated singular value decomposition (SVD) algorithm [42], or iterative approaches like the algebraic reconstruction technique (ART) [4], simultaneous ART (SART) [2], conjugate gradient for least squares (CGLS) [22], and total generalized variation regularization (TGV) [43]. Additionally, techniques such as total variation [47] and Tikhonov regularization [9] can be employed to enhance reconstruction results.

Deep Unrolling Networks. By assuming that the regularization term in Eq. (2) (i.e. $\mathcal{R}$) is differentiable and convex, a simple gradient descent scheme can be applied to solve the optimization problem:

$$\begin{aligned} &\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \alpha \nabla_{\boldsymbol{x}} J(\boldsymbol{x}_t), \\ &\text{where} \quad \nabla_{\boldsymbol{x}} J(\boldsymbol{x}_t) = \lambda \boldsymbol{A}^\dagger (\boldsymbol{A}\boldsymbol{x}_t - \boldsymbol{y}) + \nabla_{\boldsymbol{x}} \mathcal{R}(\boldsymbol{x}_t). \end{aligned} \tag{3}$$

Here, α represents the step size (i.e. search step), and $\boldsymbol{A}^\dagger$ is the pseudo-inverse of $\boldsymbol{A}$.

Previous research [16, 53] has emphasized the limitations of optimization algorithms, such as the manual selection of the regularization term and the optimization hyperparameters, which can negatively impact their performance, limiting their clinical application. Recent advancements in deep learning techniques have enabled automated parameter selection directly from the data, as demonstrated in [7, 11, 23, 33, 38, 56]. By allowing the terms in Eq. (3) to be dependent on the iteration, the gradient descent iteration becomes:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \lambda_t \boldsymbol{A}^\dagger (\boldsymbol{A}\boldsymbol{x}_t - \boldsymbol{y}) + \mathcal{G}(\boldsymbol{x}_t), \tag{4}$$

where $\mathcal{G}$ is a learned mapping representing the gradient of the regularization term. It is worth noting that the step size α in Eq. (3) is omitted as it is redundant when considering the learned components of the regularization term. Finally, Eq. (4) is unrolled into a deep recurrent neural network in order to learn the optimization parameters.

2.2. Post-processing Methods

Recent advances in sparse-view CT reconstruction leverage two main categories of deep learning methods: post-processing and dual-domain approaches. Post-processing methods, including RedCNN [6], FBPConvNet [19], and DDNet [59], treat sparse-view reconstruction as a denoising step using FBP reconstructions as input. While effective in addressing artifacts and reducing noise, they often struggle with recovering global information from extremely sparse data. To overcome this limitation, dual-domain methods integrate sinograms into neural networks for an interpolation task, recovering missing data [15, 24]. Dual-domain methods, surpassing post-processing ones, combine information from both domains. DuDoNet [29], an initial dual-domain method, connects image and sinogram domains through a Radon inversion layer. Recent Transformer-based dual-domain methods, such as DuDoTrans [49] and DDPTransformer [27], aim to capture long-range dependencies in the sinogram domain, demonstrating superior performance to CNN-based methods.

Self-supervised learning. SSL methods [5, 17, 26, 48, 58], have been applied for CT reconstruction. For instance, [5] proposed an equivariant imaging paradigm through a training strategy that enforces measurement consistency and equivariance conditions. To ensure equitable comparisons, we focus on supervised methods in this work.

2.3. Advancements in Deep Unrolling Networks

Unrolling networks constitute a line of work inspired by popular optimization algorithms used to solve Eq. (2). Leveraging the iterative nature of optimization algorithms, as presented in Eq. (4), unrolling networks aim to directly learn optimization parameters from data. These methods have found success in various inverse problems, including sparse-view CT [7, 20, 52, 54, 56], limited-angle CT [8, 11, 51], low-dose CT [1, 16], and compressed sensing MRI [12, 44].

First-order. One pioneering unrolling network, Learned Primal-Dual reconstruction [1], replaces traditional proximal operators with CNNs. In contrast, LEARN [7] and LEARN++ [56] directly unroll the optimization algorithm

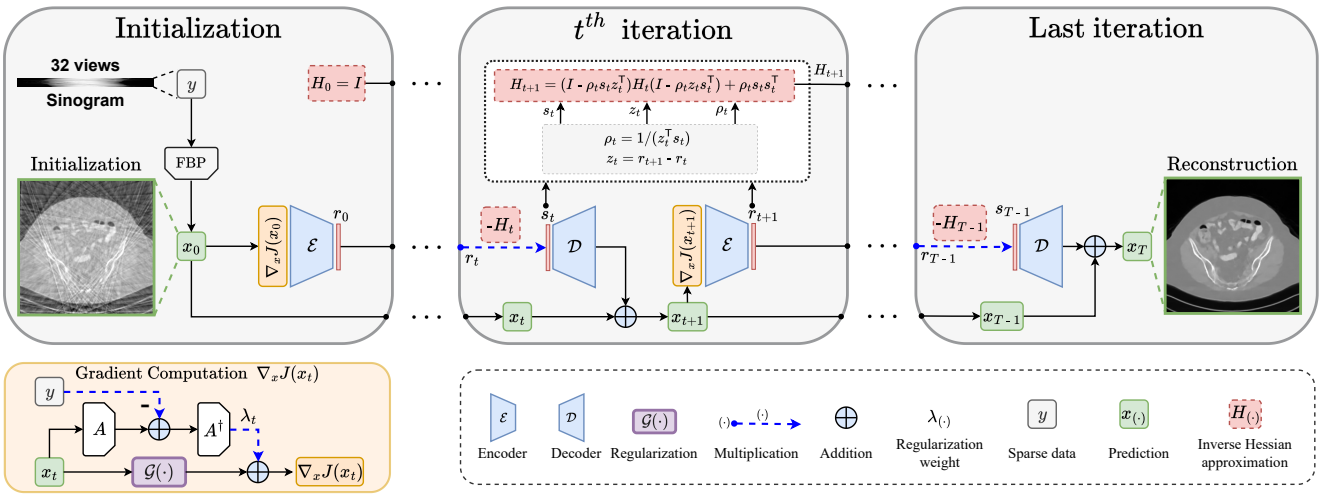

Figure 2. **Overall structure of the proposed QN-Mixer** for sparse-view CT reconstruction, unrolled from Algorithm 2. The method leverages the advantages of the quasi-Newton method for faster convergence while incorporating a latent BFGS update.

from Eq. (4) into a deep recurrent neural network. More recently, Transformers [3, 31] have been introduced into unrolling networks, such as RegFormer [54] and HUMUS-Net [12]. While achieving commendable performance, these methods require more computational resources than traditional CNN-based unrolling networks and incur a significant memory footprint due to linear scaling with the number of unrolling iterations.

Second-order. To address this, a new category of unrolling optimization methods has emerged [14], leveraging second-order techniques like the quasi-Newton method [10, 13, 21]. These methods converge faster, reducing computational demands, but struggle with increased memory usage due to Hessian matrix approximation and their application is limited to small-scale problems [30, 57]. In contrast our method propose a memory-efficient approach by operating within the latent space of gradient information (i.e. $\nabla_{\boldsymbol{x}} J(\boldsymbol{x})$ in Eq. (3)).

Algorithm 1: Quasi-Newton for sparse-view CT

Data: $\boldsymbol{y}$ (sparse sinogram)
Manual choice of the regularization term $\mathcal{R}$;
$\boldsymbol{H}_0 \leftarrow \boldsymbol{I}^{n\times n}$;
$\boldsymbol{x}_0 \leftarrow \boldsymbol{A}^\dagger \boldsymbol{y}$;
for $t \in \{0, \dots, T-1\}$ **do**
  $\boldsymbol{s}_t \leftarrow -\boldsymbol{H}_t \nabla_{\boldsymbol{x}} J(\boldsymbol{x}_t)$
  $\boldsymbol{x}_{t+1} \leftarrow \boldsymbol{x}_t + \boldsymbol{s}_t$
  $\boldsymbol{z}_t \leftarrow \nabla_{\boldsymbol{x}} J(\boldsymbol{x}_{t+1}) - \nabla_{\boldsymbol{x}} J(\boldsymbol{x}_t)$
  $\rho_t \leftarrow 1/(\boldsymbol{z}_t^\top \boldsymbol{s}_t)$
  $\boldsymbol{H}_{t+1} \leftarrow (\boldsymbol{I} - \rho_t \boldsymbol{s}_t \boldsymbol{z}_t^\top) \boldsymbol{H}_t (\boldsymbol{I} - \rho_t \boldsymbol{z}_t \boldsymbol{s}_t^\top) + \rho_t \boldsymbol{s}_t \boldsymbol{s}_t^\top$

3. Methodology

QN-Mixer is a novel second-order unrolling network inspired by the quasi-Newton (Sec. 3.1) method. It approximates the inverse Hessian matrix with a latent BFGS algorithm and includes a non-local regularization term, Incept-Mixer, designed to capture non-local relationships (Sec. 3.2). To cope with the significant computational burden associated with the full approximation of the inverse Hessian matrix, we use a latent BFGS algorithm (Sec. 3.3). An overview of the proposed method is depicted in Fig. 2, and the complete algorithm is presented in Sec. 3.4.

3.1. Quasi-Newton method

The quasi-Newton method can be applied to solve Eq. (2) and the iterative optimization solution is expressed as:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \alpha_t \boldsymbol{H}_t \nabla_{\boldsymbol{x}} J(\boldsymbol{x}_t), \tag{5}$$

where $\boldsymbol{H}_t \in \mathbb{R}^{n\times n}$ represents the inverse Hessian matrix approximation at iteration t, and α_t is the step size. The BFGS method updates the Hessian matrix approximation in each iteration. This matrix is crucial for understanding the curvature of the objective function around the current point, guiding us to take more efficient steps and avoiding unnecessary zigzagging. In the classical BFGS approach, the line search adheres to Wolfe conditions [10, 13]. A step size of $\alpha_t = 1$ is attempted first, ensuring eventual acceptance for superlinear convergence [21]. In our approach, we adopt a fixed step size of $\alpha_t = 1$. The algorithm is illustrated in Algorithm 1.

3.2. Regularization term: Incept-Mixer

Recent research on unrolling networks has often focused on selecting the representation of the regularization term gradi-

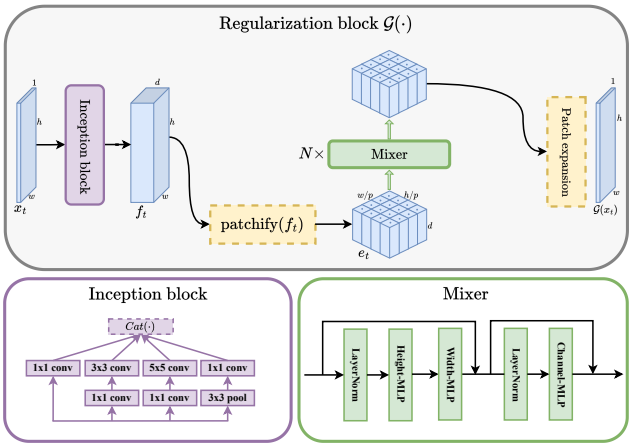

Figure 3. **Architecture of our regularization block**. It is referred to as "**Incept-Mixer**" and denoted as $\mathcal{G}$ in Eq. (4)

ent (i.e. $\mathcal{G}$ in Eq. (4)), ranging from conv-nets [7, 44, 56] to more recent attention-based nets [12, 54]. In alignment with this trend, we introduce a non-local regularization block named **Incept-Mixer** and depicted in, Fig. 3. This block is crafted by drawing inspiration from both the multi-layer perceptron mixer [46] and the inception architecture [45], leveraging the strengths of each: capturing long-range interactions through the attention-like mechanism of MLP-Mixer and extracting local invariant features from the inception block. This design choice is evident in the ablation study (see Tab. 6) where Incept-Mixer outperforms both alternatives.

Starting from an image $\boldsymbol{x}_t \in \mathbb{R}^{h\times w\times c}$ at iteration t, we pass it through an Inception block to create a feature map $\boldsymbol{f}_t \in \mathbb{R}^{h\times w\times d}$, where d is the depth of features. Subsequently, $\boldsymbol{f}_t$ undergoes patchification using a CNN with a kernel size and stride of p, representing the patch size. This process yields patch embeddings, $\boldsymbol{e}_t = \text{patchify}(\boldsymbol{f}_t) \in \mathbb{R}^{\frac{h}{p}\times\frac{w}{p}\times d}$. These embeddings are then processed through a **Mixer Layer** with token and channel MLPs, layer normalization, and skip connections for inter-layer information flow, following [46]:

$$\text{Mixer}(\boldsymbol{e}_t) = \text{Mix}(\text{MLP}_{\text{c}}, \text{Mix}([\text{MLP}_{\text{h}}, \text{MLP}_{\text{w}}], \boldsymbol{e}_t), \quad (6)$$

where $\text{Mix}(\text{Layer}, \boldsymbol{e}_t) = \text{Layer}(\text{LN}(\boldsymbol{e}_t)) + \boldsymbol{e}_t$, with LN as layer normalization. MLP_{h}, MLP_{w} are applied to height and width features, respectively, and MLP_{c} to rows and shared. Finally, after N such mixer layers, the regularized sample is transformed back to an image through a patch expansion step to obtain $\mathcal{G}(\boldsymbol{x}_t)$. Consequently, the iterative optimization solution is as follows:

$$\begin{aligned} &\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \boldsymbol{H}_t \nabla_{\boldsymbol{x}} J(\boldsymbol{x}_t), \\ &\text{where} \quad \nabla_{\boldsymbol{x}} J(\boldsymbol{x}_t) = \lambda_t \boldsymbol{A}^{\dagger}\left(\boldsymbol{A}\boldsymbol{x}_t - \boldsymbol{y}\right) + \mathcal{G}(\boldsymbol{x}_t). \end{aligned} \quad (7)$$

Here, $\mathcal{G}(\boldsymbol{x}_t)$ denotes the Incept-Mixer model, representing the learned gradient of the regularization term.

3.3. Latent BFGS update

We propose a memory-efficient latent BFGS update. Drawing inspiration from LDMs [41], at step t, given the gradient value $\nabla_{\boldsymbol{x}} J(\boldsymbol{x}_t) \in \mathbb{R}^{h\times w\times c}$, the encoder $\mathcal{E}$ encodes it into a latent representation $\boldsymbol{r}_t = \mathcal{E}(\nabla_{\boldsymbol{x}} J(\boldsymbol{x}_t)) \in \mathbb{R}^{l_h\cdot l_w}$. Importantly, the encoder downsamples the gradient by a factor $\boldsymbol{f}_{\mathcal{E}} = \frac{h}{h_l} = \frac{w}{w_l}$. Throughout the paper, we explore different downsampling factors (see Tab. 5) $\boldsymbol{f}_{\mathcal{E}} = 2^k$, where $k \in \mathbb{N}$ is the number of downsampling stacks. Encoding the gradient reduces the optimization variable size of BFGS (i.e. $\boldsymbol{H}_t \in \mathbb{R}^{(l_h\cdot l_w)\times(l_h\cdot l_w)}$), thereby decreasing the computational cost associated with high memory demand. The direction is then computed in the latent space $s_t = -\boldsymbol{H}_t\boldsymbol{r}_t$, and finally, the decoder $\mathcal{D}$ reconstructs the update from the latent direction, giving $\mathcal{D}(s_t) = \mathcal{D}(-\boldsymbol{H}_t\mathcal{E}(\nabla_{\boldsymbol{x}} J(\boldsymbol{x}_t))) \in \mathbb{R}^{h\times w\times c}$. It is noteworthy that $\mathcal{E}$ and $\mathcal{D}$ are shared across the algorithm iterations, as shown in Fig. 2.

3.4. Proposed algorithm of QN-Mixer

Algorithm 2: QN-Mixer (latent BFGS update)

Data: $\boldsymbol{y}$ (sparse sinogram)
$\boldsymbol{H}_0 \leftarrow \boldsymbol{I}^{(l_h\cdot l_w)\times(l_h\cdot l_w)}$;
$\boldsymbol{x}_0 \leftarrow \boldsymbol{A}^{\dagger}\boldsymbol{y}$;
$\boldsymbol{r}_0 \leftarrow \mathcal{E}(\nabla_{\boldsymbol{x}} J(\boldsymbol{x}_0))$;
for $t \in \{0, \ldots, T-1\}$ **do**
  $\boldsymbol{s}_t \leftarrow -\boldsymbol{H}_t\boldsymbol{r}_t$
  $\boldsymbol{x}_{t+1} \leftarrow \boldsymbol{x}_t + \mathcal{D}(\boldsymbol{s}_t)$
  $\boldsymbol{r}_{t+1} \leftarrow \mathcal{E}(\nabla_{\boldsymbol{x}} J(\boldsymbol{x}_{t+1}))$
  $\boldsymbol{z}_t \leftarrow \boldsymbol{r}_{t+1} - \boldsymbol{r}_t$
  $\rho_t \leftarrow 1/(\boldsymbol{z}_t^{\mathsf{T}}\boldsymbol{s}_t)$
  $\boldsymbol{H}_{t+1} \leftarrow (\boldsymbol{I} - \rho_t\boldsymbol{s}_t\boldsymbol{z}_t^{\mathsf{T}})\boldsymbol{H}_t(\boldsymbol{I} - \rho_t\boldsymbol{z}_t\boldsymbol{s}_t^{\mathsf{T}}) + \rho_t\boldsymbol{s}_t\boldsymbol{s}_t^{\mathsf{T}}$

Our method, builds on the BFGS update [10, 13] rank-one approximation for the inverse Hessian. This approximation serves as a preconditioning matrix, guiding the descent direction. In contrast to [14], which directly learns the inverse Hessian approximation from data, our approach incorporates the mathematical equations of the BFGS algorithm for more accurate approximations. The full QN-Mixer algorithm is illustrated in Algorithm 2.

4. Experiments

In this section, we initially present our experimental settings, followed by a comparison of our approach with other state-of-the-art CT reconstruction methods. Finally, we delve into the contribution analysis of each component in our model.

Method	No noise ($N_0 = 0$)						Low noise ($N_1 = 10^6$)						High noise ($N_2 = 5 \times 10^5$)					
	$n_v = 32$		$n_v = 64$		$n_v = 128$		$n_v = 32$		$n_v = 64$		$n_v = 128$		$n_v = 32$		$n_v = 64$		$n_v = 128$	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
FBP	22.65	40.49	27.29	57.94	33.04	79.50	22.09	32.73	26.51	49.56	31.69	71.09	19.05	15.56	22.71	25.74	26.52	40.87
FBPConvNet [19]	30.32	85.11	35.42	90.15	39.71	94.64	30.20	84.46	35.09	89.72	39.06	94.08	29.91	82.52	34.13	87.85	36.89	91.28
DuDoTrans [49]	30.48	84.70	35.37	91.87	40.62	96.41	30.34	83.72	35.36	91.42	39.75	95.49	30.09	81.83	34.09	88.67	37.08	93.44
Learned PD [1]	35.88	92.09	41.03	96.28	43.33	97.31	35.78	92.21	39.03	94.79	41.65	96.44	33.80	89.23	37.34	93.23	39.17	94.69
LEARN [7]	37.58	94.65	42.26	97.25	43.11	97.57	36.95	93.63	39.91	95.82	42.17	97.11	34.38	90.51	37.15	93.53	39.38	95.18
RegFormer [54]	38.71	95.42	43.56	97.76	47.95	98.98	37.21	94.73	41.65	96.92	**44.38**	98.02	**35.93**	**92.78**	38.53	94.84	**40.52**	96.19
QN-Mixer (ours)	**39.51**	**96.11**	**45.57**	**98.48**	**50.09**	**99.32**	**37.50**	**94.92**	**42.46**	**97.70**	44.27	**98.11**	35.91	92.49	**38.73**	**94.92**	40.51	**96.27**

Table 1. **Quantitative evaluation on AAPM** of state-of-the-art methods (PSNR in dB and SSIM in %). **Bold**: Best, under: second best.

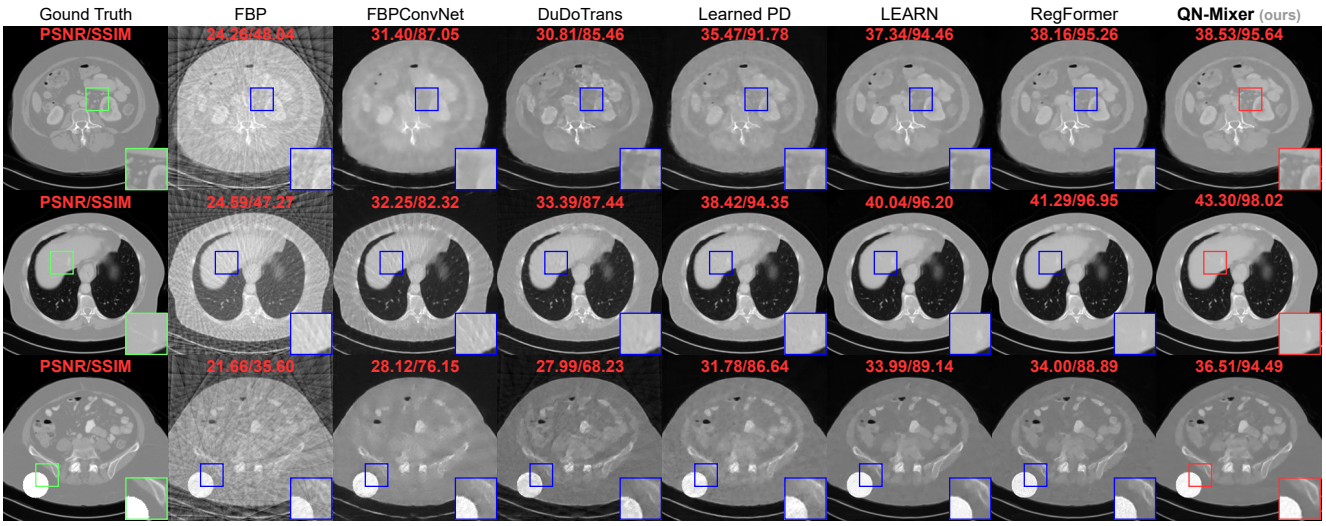

Figure 4. **Visual comparison on AAPM**. From top to bottom: the results under the following conditions: first (n_v=32, N_1), second (n_v=64, N_1), third (n_v=32, N_0). The last row presents out-of-distribution (OOD) results with a randomly overlaid circle on a test image. The display window is set to $[-1000, 800]$ HU.

4.1. Experimental Setup

Datasets. We evaluate our method on two widely used datasets: the "2016 NIH-AAPM-Mayo Clinic Low-Dose CT Grand Challenge" dataset (AAPM) [35] and the DeepLesion dataset [55]. The AAPM dataset comprises 2378 full-dose CT images from 10 patients, while DeepLesion is the largest publicly accessible multi-lesion real-world CT dataset, including 4427 unique patients.

Implementation details. For AAPM, we select 1920 training images from 8 patients, 244 validation images from 1 patient, and 214 testing images from the last patient. For DeepLesion, we select a subset of 2000 training images and 300 testing images randomly from the official splits. All images are resized to 256×256 pixels. To simulate the forward and backprojection operators, we use the Operator Discretization Library (ODL) [39] with a 2D fan-beam geometry (512 detector pixels, source-to-axis distance of 600 mm, axis-to-detector distance of 290 mm). Sparse-view CT images are generated with $n_v \in \{32, 64, 128\}$ projection views, uniformly sampled from a full set of 512 views covering $[0, 2\pi]$. To mimic real-world CT images, we introduce mixed noise to the sinograms, combining 5% Gaussian noise and Poisson noise with an intensity of 1×10^6.

Training details. For each set of n_v views, we train our model for 50 epochs using 4 Nvidia Tesla V100 (32GB RAM). We employ the AdamW optimizer [32] with a learning rate of 1×10^{-4}, weight decay 1×10^{-2}, and utilize the mean squared error loss with a batch size of 1. Additionally, we incorporate a learning rate decay factor of 0.1 after 40 epochs. Unrolling iterations for QN-Mixer are set to $T = 14$. Incept-Mixer uses a patch size of $p = 4$, $d = 96$ embedding dimension, and $N = 2$ mixer layers. The inverse Hessian size is $64^2 \times 64^2$ with $k = 2$ downsampling blocks. $\mathcal{E}$ comprises cascading 3x3 CNNs with max-pooling for downsampling, culminating in a 1x1 CNN layer for a one-channel latent gradient. $\mathcal{D}$ utilizes 2x2 ConvTranspose operations. Both $\mathcal{E}$ and $\mathcal{D}$ layers incorporate instance normalization and PReLU activation. Following [54], $\boldsymbol{A}^\dagger$ is implemented using the FBP algorithm for the pseudo-inverse of $\boldsymbol{A}$.

Evaluation metrics. Following established evaluation protocols [1, 49, 54], we employ the structural similarity index

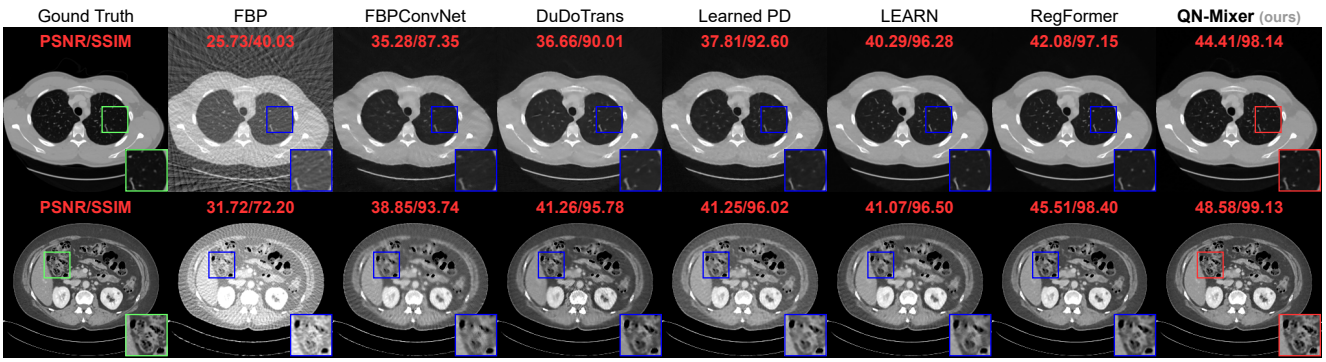

Figure 5. **Visual comparison on DeepLesion** of state-of-the-art methods. Rows display results under different conditions: (n_v=64, N_1) and (n_v=128, N_1). Display windows are set to $[-1000, 800]$ HU for the first row and $[-200, 300]$ HU for the second row.

Method	$n_v = 32$		$n_v = 64$		$n_v = 128$	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
FBP	21.55	31.65	26.07	47.17	31.49	69.63
FBPConvNet [19]	30.74	80.41	34.64	87.36	38.69	92.94
DuDoTrans [49]	32.11	79.86	36.02	88.14	40.47	93.81
Learned PD [1]	34.02	88.44	37.56	92.46	40.79	95.32
LEARN [7]	35.76	92.12	39.83	95.66	41.34	96.21
RegFormer [54]	37.38	93.89	41.70	96.78	46.10	98.39
QN-Mixer (ours)	**39.39**	**95.67**	**43.75**	**97.73**	**48.62**	**98.64**

Table 2. **Quantitative evaluation on DeepLesion** for state-of-the-art methods (PSNR in dB and SSIM in %). With Poisson noise level of $N_1 = 10^6$. **Bold**: Best, under: second best.

Method	$n_v = 32$		$n_v = 64$		$n_v = 128$	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
FBP	21.38	33.36	26.08	50.29	31.43	73.06
FBPConvNet [19]	28.05	75.96	32.50	82.90	35.45	88.14
DuDoTrans [49]	28.11	68.17	32.71	83.26	36.41	90.36
Learned PD [1]	31.96	87.10	36.40	92.57	37.63	93.17
LEARN [7]	34.48	90.15	36.89	91.85	38.32	94.67
RegFormer [54]	34.49	89.98	36.95	91.48	38.02	92.44
QN-Mixer (ours)	**36.84**	**94.84**	**42.11**	**97.78**	**45.69**	**98.82**

Table 3. **Quantitative evaluation on out-of-distribution (OOD) AAPM test dataset** of state-of-the-art methods (PSNR in dB and SSIM in %). **Bold**: Best, under: second best.

measure (SSIM) with parameters: level 5, a Gaussian kernel of size 11, and standard deviation 1.5, as our primary performance metric. Furthermore, we supplement our assessment with the peak signal-to-noise ratio (PSNR).

State-of-the-art baselines. We compare QN-Mixer to multiple state-of-the-art competitors: (1) *post-processing* based denoising methods, i.e., FBPConvNet [19], and DuDoTrans [49]; (2) *first-order unrolling* reconstruction networks, i.e., Learned Primal-Dual [1], LEARN [7], and RegFormer [54]. Note that we replace the pseudo-inverse operator used by LEARN with the FBP algorithm, as it has been demonstrated to be more effective according to [54]. To ensure a fair comparison, we utilize the code-base released by the authors when possible or meticulously implement the methods based on the details provided in their papers. All approaches undergo training and testing on the same datasets, as elaborated in implementation details.

4.2. Comparison with state-of-the-art methods

Quantitative comparison. We compared our model with state-of-the-art baselines on two public datasets. For AAPM, models were trained and tested across three projection views ($n_v \in \{32, 64, 128\}$) and three noise levels, namely no noise $N_0 = 0$, low noise $N_1 = 10^6$, and high noise $N_2 = 5 \times 10^5$ (see Tab. 1). For DeepLesion, models were trained and tested on the same three projection views and a noise level of $N_1 = 10^6$ (see Tab. 2). Visual results are provided in Fig. 4 (AAPM) and Fig. 5 (DeepLesion). Impressively, our method achieves state-of-the-art results on DeepLesion across all projection views. It outperforms the second-best baseline, RegFormer, with an average improvements of $+2.23$ dB in PSNR and $+1.02\%$ in SSIM. On AAPM without noise, we achieve state-of-the-art results across all projection views and improve the second best by an average $+1.65$ dB and $+0.58\%$. In the presence of low noise, QN-Mixer achieves state-of-the-art results performance in all cases except n_v=128 with -0.11 dB and shows an average improvements of $+0.33$ dB and $+0.35\%$ over RegFormer. With high noise, our method performs nearly on par in n_v=32 (-0.02 dB and -0.29%), achieves state-of-the-art in n_v=64 ($+0.2$ dB and $+0.08\%$), and competes closely in n_v=128 (-0.01 dB and $+0.08\%$). As noise increases, we attribute the decline in improvement to the compressed gradient information in the latent BFGS, influenced by sinogram changes, and the utilization of the FBP algorithm instead of the pseudo-inverse.

Performance comparison on OOD textures. We evaluate frozen model performance on CT images featuring a randomly positioned white circle with noise-free sinograms, as illustrated in the third row of Fig. 4. The rationale and details are provided in the supplementary material. In Tab. 3, QN-Mixer attains state-of-the-art results across all n_v views. First-order unrolling networks such as LEARN and RegFormer exhibit significant PSNR degradation of

-3.1 dB and -4.22 dB, respectively, for n_v=32, while our method demonstrates a milder degradation of -2.67 dB.

Visual comparison. As it can be seen on Fig. 4 and Fig. 5, FBPConvNet and DuDoTrans show significant blurring and pronounced artifacts when n_v=32. While Learned PD and LEARN show satisfactory performance, they struggle with intricate details, like in the liver and spine. In contrast, RegFormer produces high-quality images but faces challenges in generalizing to OOD data. QN-Mixer excels in producing high-quality images with fine details, even under challenging conditions such as n_v=32 views and OOD data.

Method	#Iters	Epoch time (s)	Time (ms)	#Params (M)	Memory (GB)
Post-processing based denoising					
FBPConvNet [19]	-	68	12.4	31.1	1.30
DuDoTrans [49]	-	92	60.1	15.0	1.38
First-order unrolling reconstruction networks					
Learned PD [1]	10	82	47.2	0.25	0.81
LEARN [7]	30	780	679.8	4.50	1.85
RegFormer [54]	18	700	598.9	5.00	10.19
Second-order unrolling Quasi-Newton					
QN-Mixer (ours)	14	594	610.2	8.50	7.83

Table 4. **Comparison of computational efficiency**. Training epoch time is reported in seconds, #Params in M and memory costs for state-of-the-art methods on AAPM with $n_v = 32$ views.

Efficiency comparison. The results in Tab. 4 show that QN-Mixer is more computationally efficient than RegFormer, with a $1.3\times$ reduction in memory usage. Furthermore, our training time demonstrates a significant enhancement, realizing a speed improvement of 106 seconds per epoch compared to first-order unrolling methods like LEARN and RegFormer. Additionally, our method requires only 14 iterations, in contrast to the 30 and 18 iterations needed by LEARN and RegFormer, respectively.

Hessian size	PSNR ↑	SSIM ↑
$8^2 \times 8^2$	35.69	93.71
$16^2 \times 16^2$	38.11	95.31
$32^2 \times 32^2$	39.37	96.01
$64^2 \times 64^2$	**39.51**	**96.11**

Table 5. **Ablation on the inverse Hessian approximation size**.

Method	PSNR ↑	SSIM ↑
QN with different learned regularization		
Inception	31.65	85.28
U-Net	34.29	92.92
MLP-Mixer	36.89	93.87
Incept-Mixer	39.51	96.11
Pseudo-inverse $A^\dagger$ vs Filtered Back Projection (FBP)		
QN-Mixer+$A^\dagger$	38.94	95.83
QN-Mixer+FBP	39.51	96.11
First vs second order Quasi-Newton (QN)		
Incept-Mixer+first-order	37.45	94.25
Incept-Mixer+QN	39.51	96.11

Table 6. **QN-Mixer ablation**.

4.3. Ablation Study

In this section, we leverage the AAPM dataset with n_v=32 views by default, and no noise is introduced to the sinogram.

Inverse Hessian approximation size. The results in Tab. 5 emphasize the significant impact of the inverse Hessian approximation size on our performance. When too small, a notable degradation is observed (e.g., $8^2 \times 8^2$), while larger sizes result in performance improvements as the approximation approaches the full inverse Hessian. Exceeding $64^2 \times 64^2$ was unfeasible in our experiments due to memory constraints.

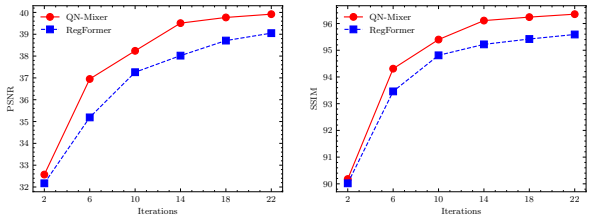

Figure 6. **Ablation on the number of unrolling iterations**. Left: PSNR (dB); Right: SSIM (%)

Number of unrolling iterations. In Fig. 6, we visually depict the influence of the number of unrolling iterations on the performance of QN-Mixer and RegFormer. Notably, the performance of both methods shows improvement with an increase in the number of iterations. When subjected to an equal number of iterations, our method consistently surpasses RegFormer in performance. Remarkably, we achieve comparable results to RegFormer even with only 10 iterations, demonstrating the efficiency of our approach.

Regularization term. In Tab. 6, we evaluate the impact of the regularization term in our framework. Our Incept-Mixer is compared against various learned alternatives, including the Inception block [45] and MLP-Mixer block [46]. Additionally, employing the pseudo-inverse $\boldsymbol{A}^\dagger$ instead of the FBP results in a less pronounced degradation (-0.57 dB and -0.28%), enhancing the interpretability of QN-Mixer. Finally, we test our Incept-Mixer in the first-order framework, highlighting the significance of the second-order latent BFGS approximation with a significant improvement ($+2.06$ dB and $+1.86\%$).

5. Conclusion

In this paper, we investigate the application of deep second-order unrolling networks for tackling imaging inverse problems. To this end, we introduce QN-Mixer, a quasi-Newton inspired algorithm where a latent BFGS method approximates the inverse Hessian, and our Incept-Mixer serves as the non-local learnable regularization term. Extensive experiments confirm the successful sparse-view CT reconstruction by our model, showcasing superior performance with fewer iterations than state of-the-art methods. In summary, this research offers a fresh perspective that can be applied to any iterative reconstruction algorithm. A limitation of our work is the memory requirements associated with quasi-Newton algorithm. We introduced a memory efficient alternative by projecting the gradient to a lower dimension, successfully addressing the CT reconstruction problem. However, its applicability to other inverse problems may be limited. In future work, we aim to extend our approach to handle larger Hessian sizes, broadening its application to a range of problems.

References

[1] Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE TMI*, 37:1322–1332, 2018. 2, 3, 6, 7, 8

[2] Andersen Arie and Kak Avinash. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the art algorithm. *Ultrasonic imaging*, 6:81–94, 1984. 3

[3] Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N, Kaiser Lukasz, and Polosukhin Illia. Attention is all you need. In *NeurIPS*, 2017. 4

[4] Kak Avinash and Slaney Malcolm. *Principles of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics, 2001. 3

[5] Dongdong Chen, Julián Tachella, and Mike E. Davies. Equivariant imaging: Learning beyond the range space. In *ICCV*, pages 4359–4368, 2021. 3

[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 TMI*, 36:2524–2535, 2017. 2, 3

[7] Hu Chen, Yi Zhang, Yunjin Chen, Junfeng Zhang, Weihua Zhang, Huaiqiang Sun, Yang Lv, Peixi Liao, Jiliu Zhou, and Ge Wang. LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT. *IEEE TMI*, 37: 1333–1347, 2018. 2, 3, 5, 6, 7, 8

[8] Weilin Cheng, Yu Wang, Hongwei Li, and Yuping Duan. Learned full-sampling reconstruction from incomplete data. *IEEE TCI*, pages 945–957, 2020. 2, 3

[9] Peng Chengbin, Rodi William L., and M. Nafi Toksöz. *A Tikhonov Regularization Method for Image Reconstruction*, pages 153–164. Springer US, 1993. 3

[10] William C. Davidon. Variable metric method for minimization. *SIAM Journal on Optimization*, 1:1–17, 1991. 2, 4, 5

[11] Hu Dianlin, Zhang Yikun, Liu Jin, Luo Shouhua, and Chen Yang. DIOR: Deep iterative optimization-based residual-learning for limited-angle CT reconstruction. *IEEE TMI*, pages 1778–1790, 2022. 2, 3

[12] Zalan Fabian, Berk Tinaz, and Mahdi Soltanolkotabi. HUMUS-Net: Hybrid unrolled multi-scale network architecture for accelerated MRI reconstruction. In *NeurIPS*, 2022. 2, 3, 4, 5

[13] Roger Fletcher. *Practical Methods of Optimization*. John Wiley & Sons, New York, NY, USA, 1987. 2, 4, 5

[14] Erik Gartner, Luke Metz, Mykhaylo Andriluka, C. Daniel Freeman, and Cristian Sminchisescu. Transformer-based Learned Optimization. In *CVPR*, pages 11970–11979, 2023. 2, 4, 5

[15] Muhammad Usman Ghani and W. Clem Karl. Deep learning-based sinogram completion for low-dose CT. In *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop*, pages 1–5, 2018. 2, 3

[16] Gupta Harshit, Jin Kyong Hwan, Nguyen Ha Q., McCann Michael T., and Unser Michael. CNN-based projected gradient descent for consistent CT image reconstruction. *IEEE TMI*, 37:1440–1453, 2018. 2, 3

[17] Allard Adriaan Hendriksen, Daniël Maria Pelt, and K. Joost Batenburg. Noise2inverse: Self-supervised deep convolutional denoising for tomography. *IEEE TCI*, pages 1320–1335, 2020. 3

[18] Dianlin Hu, Jin Liu, Tianling Lv, Qianlong Zhao, Yikun Zhang, Guotao Quan, Juan Feng, Yang Chen, and Limin Luo. Hybrid-domain neural network processing for sparse-view CT reconstruction. *IEEE TRPMS*, 5:88–98, 2020. 2

[19] Kyong Hwan Jin, Michael T. McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE TIP*, 26:4509–4522, 2017. 2, 3, 6, 7, 8

[20] Xiang Jinxi, Dong Yonggui, and Yang Yunjie. FISTA-Net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE TMI*, 40:1329–1339, 2021. 2, 3

[21] Nocedal Jorge and Wright Stephen J. Quasi-Newton methods. *Numerical optimization*, 75:135–163, 2006. 2, 4

[22] Satoshi Kawata and Orhan Nalcioglu. Constrained iterative reconstruction by the conjugate gradient method. *IEEE TMI*, 4:65–71, 1985. 3

[23] Erich Kobler, Alexander Effland, Karl Kunisch, and Thomas Pock. Total deep variation for linear inverse problems. In *CVPR*, pages 7546–7555, 2020. 3

[24] Hoyeon Lee, Jongha Lee, Hyeongseok Kim, Byungchul Cho, and Seungryong Cho. Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction. *IEEE TRPMS*, 3:109–119, 2018. 2, 3

[25] Minjae Lee, Hyemi Kim, and Hee-Joung Kim. Sparse-view CT reconstruction based on multi-level wavelet convolution neural network. *Physica Medica*, 80:352–362, 2020. 2

[26] Meng Li, William Hsu, Xiaodong Xie, Jason Cong, and Wen Gao. Sacnn: Self-attention convolutional neural network for low-dose ct denoising with self-supervised perceptual loss network. *IEEE TMI*, pages 2289–2301, 2020. 3

[27] Runrui Li, Qing Li, Hexi Wang, Saize Li, Juanjuan Zhao, Yan Qiang, and Long Wang. DDPTransformer: Dual-domain with parallel transformer network for sparse view CT image reconstruction. *IEEE TCI*, pages 1–15, 2022. 2, 3

[28] Zilong Li, Chenglong Ma, Jie Chen, Junping Zhang, and Hongming Shan. Learning to distill global representation for sparse-view ct. In *ICCV*, pages 21196–21207, 2023. 2

[29] Wei-An Lin, Cheng Liao, Haofu Peng, Xiaohang Sun, Jingdan Zhang, Jiebo Luo, Rama Chellappa, and Zhou S. Kevin. DuDoNet: Dual domain network for CT metal artifact reduction. In *CVPR*, pages 10512–10521, 2019. 2, 3

[30] Chengchang Liu and Luo Luo. Quasi-newton methods for saddle point problems. In *NeurIPS*, 2022. 2, 4

[31] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, and Stephen Lin Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, pages 9992–10002, 2021. 4

[32] Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6

[33] Sebastian Lunz, Ozan Öktem, and Carola-Bibiane Schönlieb. Adversarial regularizers in inverse problems. In *NeurIPS*, 2018. 3

[34] Cormack Allan Macleod. Representation of a function by its line integrals, with some radiological applications. *Journal of Applied Physics*, 34:2722–2727, 1963. 2

[35] C. McCollough. TU-FG-207A-04: Overview of the low dose CT grand challenge. *Medical Physics*, 43:3759–3760, 2016. 6

[36] Luke Metz, C. Daniel Freeman, James Harrison, Niru Maheswaranathan, and Jascha Sohl-Dickstein. Practical tradeoffs between memory, compute, and performance in learned optimizers. *arXiv preprint arXiv:2203.11860*, 2022. 2

[37] Donald L. Miller and David Schauer. The alara principle in medical imaging. *Philosophy*, 44:595–600, 1983. 1

[38] Subhadip Mukherjee, Marcello Carioni, Ozan Öktem, and Carola-Bibiane Schönlieb. End-to-end reconstruction meets data-driven regularization for inverse problems. In *NeurIPS*, 2021. 3

[39] Ozan Öktem, Jonas Adler, Holger Kohr, and The ODL Team. Operator discretization library (ODL), 2014. 6

[40] Johann Radon. über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Berichte über die Verhandlungen der Königlich-Sächsischen Akademie der Wissenschaften zu Leipzig*, 69:262–277, 1917. 3

[41] Rombach Robin, Blattmann Andreas, Lorenz Dominik, Esser Patrick, and Ommer Björn. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022. 5

[42] Liu Rui, He Lu, Luo Yan, and Yu Hengyong. Singular value decomposition-based 2D image reconstruction for computed tomography. *Journal of X-ray science and technology*, 25: 113–134, 2017. 3

[43] Niu Shaohua, Gao Yan, Bian Zhaoying, Huang Jing, Chen Wufan, Yu Hengyong, Liang Zhengrong, and Ma Jianhua. Sparse-view x-ray CT reconstruction via total generalized variation regularization. *PMB*, 59:2997–3017, 2014. 3

[44] Anuroop Sriram, Jure Zbontar, Tullie Murrell, Aaron Defazio, C. Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-End variational networks for accelerated MRI reconstruction. In *MICCAI*, pages 64–73, 2020. 2, 3, 5

[45] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, and Vincent Vanhoucke Andrew Rabinovich. Going deeper with convolutions. In *CVPR*, pages 1–9, 2015. 2, 5, 8

[46] Ilya Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Peter Steiner, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, and Alexey Dosovitskiy. MLP-mixer: An all-MLP architecture for vision. In *NeurIPS*, 2021. 2, 5, 8

[47] Gomi Tsutomu and Koibuchi Yukio. Use of a Total Variation minimization iterative reconstruction algorithm to evaluate reduced projections during digital breast tomosynthesis. *BioMed Research International*, 2018:1–14, 2018. 3

[48] Mehmet Ozan Unal, Metin Ertas, and Isa Yildirim. Self-supervised training for low-dose ct reconstruction. In *ISBI*, pages 69–72, 2021. 3

[49] Ce Wang, Kun Shang, Haimiao Zhang, Qian Li, and S. Kevin Zhou. DuDoTrans: Dual-domain transformer for sparse-view CT reconstruction. In *Machine Learning for Medical Image Reconstruction*, pages 84–94. Springer International Publishing, 2022. 2, 3, 6, 7, 8

[50] Ge Wang, Hengyong Yu, and Bruno De Man. An outlook on X-ray CT research and development. *Medical Physics*, 35: 1051–1064, 2008. 1

[51] Jiaxi Wang, Li Zeng, Chengxiang Wang, and Yumeng Guo. ADMM-based deep reconstruction for limited-angle CT. *PMB*, 64, 2019. 2, 3

[52] Wu Weiwen, Hu Dianlin, Niu Chuang, Yu Hengyong, Vardhanabhuti Varut, and Wang Ge. DRONE: Dual-domain residual-based optimization network for sparse-view CT reconstruction. *IEEE TMI*, 40:3002–3014, 2021. 2, 3

[53] Dufan Wu, Kyungsang Kim, Georges El Fakhri, and Quanzheng Li. Iterative low-dose ct reconstruction with priors trained by artificial neural network. *IEEE TMI*, 36:2479–2486, 2017. 3

[54] Wenjun Xia, Ziyuan Yang, Qizheng Zhou, Zexin Lu, Zhongxian Wang, and Yi Zhang. Transformer-based iterative reconstruction model for sparse-view CT reconstruction. In *MICCAI*, 2022. 2, 3, 4, 5, 6, 7, 8

[55] Ke Yan, Xiaosong Wang, Le Lu, and Ronald M. Summers. DeepLesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of Medical Imaging*, 5:036501, 2018. 6

[56] Zhang Yi, Chen Hu, Xia Wenjun, Chen Yang, Liu Baodong, Liu Yan, Sun Huaiqiang, and Zhou Jiliu. LEARN++: Recurrent dual-domain reconstruction network for compressed sensing CT. *IEEE TRPMS*, 7:132–142, 2023. 2, 3, 5

[57] Tsai Yu-Jung, Bousse Alexandre, Ehrhardt Matthias J., Stearns Charles W., Ahn Sangtae, Hutton Brian F., Arridge Simon, and Thielemans Kris. Fast quasi-newton algorithms for penalized reconstruction in emission tomography and further improvements via preconditioning. *IEEE TMI*, 37: 1000–1010, 2018. 2, 4

[58] Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: Self-supervised learning-based tomography via sinogram synthesis and prediction. In *ICCV*, pages 1940–1950, 2021. 3

[59] Zhicheng Zhang, Xiaokun Liang, Xu Dong, Yaoqin Xie, and Guohua Cao. A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE TMI*, 37:1407–1417, 2018. 2, 3