

Learned Full-sampling Reconstruction

Weilin Cheng¹, Yu Wang², Tong Xin², and Yuping Duan¹[0000–0003–2433–3056]

¹ Center for Applied Mathematics, Tianjin University, Tianjin 300072, China
yuping.duan@tju.edu.cn

² Alibaba Group
tonggou.wangyu@alibaba-inc.com

Abstract. X-ray computed tomography (CT) reconstruction with sparse projection views was proposed to reduce both the radiation dose and scan time. However, lacking of sufficient projection views may lead to severe artifacts for analytical reconstruction method such as the filtered back projection (FBP). Although the projection data is incomplete, we can generate the full-sampling system matrices according to the sufficient-sampling conditions [5]. Thus, we propose a novel iterative reconstruction model by fitting the CT images and their corresponding measurements in Radon domain through the full-sampling system matrices to fully make use of the prior information contained in observed data. Our proposed model is solved by the learned alternating minimization method, which accounts for a forward operator in deep neural network by the unrolling strategy. Numerical results demonstrate that the proposed approach outperforms some latest learning based reconstruction methods for the sparse-view CT problems.

Keywords: Sparse-view · CT reconstruction · full sampling · CNN.

1 Introduction

Image reconstruction from down-sampled or limited measurements, e.g., low dose and limited angle CT, are examples of ill-posed inverse problems, which can be formulated as estimating the image $u \in X$ from the measurement $g \in Y$,

$$g = Au + n, \quad (1)$$

where the reconstruction space X and data space Y are typically Hilbert space, $A : X \rightarrow Y$ is the projection matrix for sparse CT, and $n \in Y$ is the random noises generated during the imaging processes. The goal of CT reconstruction is to recover the image u from the set of acquired projection data g . For the sparse-view CT, the system matrix, denoted by A_S , has fewer rows than columns so that there is a nontrivial nullspace and has infinity many solutions. Even if the solution of the inverse problem exists and is unique, the linear operator A_S may still be ill-conditioned such that the condition number $\|A\|\|A^{-1}\|$ is large and the linear system (1) is sensitive to the perturbations in data.

One way for the ill-posed inverse problem is to introduce certain regularity into the problem to guarantee the existence, uniqueness and stability of the solution. The general regularization method gives the following energy minimization problem

$$\min_{u \in X} \mathcal{D}(Au, g) + \mathcal{R}(u), \quad (2)$$

where $\mathcal{D}(Au, g)$ is the data fidelity term and $\mathcal{R}(u)$ is the regularization term. Thus, the task of solving (2) mainly includes: 1) how to define the data fidelity to describe the interrelationship between g and u ; and 2) how to model the regularization according to the prior information of u . In case of additive Gaussian noise and the unknown u being piecewise constant, we can obtain the well-known total variation minimization model for CT reconstruction [7] as follows

$$\min_{u \in X} \frac{1}{2} \|Au - g\|_X^2 + \lambda \|\nabla u\|_1, \quad (3)$$

where λ acts as a trade-off parameter between the data discrepancy and regularization. Although TV regularization improves the reconstruction quality compared to analytical reconstruction such as filtered back-projection (FBP) method, it is still not judicious to choose the data fidelity and regularization in such a sophisticated way.

Due to the development of deep convolutional neural networks (CNN) in a broad range of computer vision tasks, deep learning techniques are being actively used in medical imaging community. The pioneer work of Yang *et al.* [8] reformulated an ADMM algorithm for compressive sensing MR imaging into a deep network by learning the parameters end-to-end in the training phase. Jin *et al.* [4] used the deep CNN as a post-processing step after the reconstruction of FBP to mitigate noises and artifacts. Adler and Öktem [1] proposed the learned primal dual algorithm for CT reconstruction by unrolling the proximal primal-dual optimization method and replacing the proximal operators with convolutional neural networks. Liu, Kuang and Zhang [6] used a deep learning regularization structure to learn the data consistence from the observed data. Dong, Li and Shen [3] proposed a joint spatial-Radon domain reconstruction (JSR) model for sparse view CT imaging, and was recently reformulated into the feed-forward deep network [9]. Learning-based models have been already proven efficient for image reconstruction problems.

In this work, we aim to reconstruct the sparse-view CT by making using of the full-sampling system matrix, which is called as the learned full sampling reconstruction (FSR). Instead of modeling the data fidelity term according to the noise distribution and the regularization term based on the prior information, we take the advantages of deep CNN to learn the interrelationship between observed data and reconstruction image and the prior information directly from the data. As we can obtain the full sampling system-matrix according to the sufficient sampling conditions in [5], we introduce another fidelity term to enforce the closeness of the reconstructed image and the full sampling projection data. In this way, we can learn the prior information of the completed Radon domain data from the training data, which is then applied to approximate the full-sampling projection in the testing. We use the alternating direction method to achieve an iterative scheme, and find the best update in each iteration using the CNN. Numerical experiments demonstrate that the proposed FSR-net achieves better performance in sparse-view CT reconstruction.

2 Our Approach

In CT reconstruction, the system matrix A reflects the relationship between the projections on detector and the reconstructed objects. For the circular fan-beam CT, the

dimensions of the system matrix A are $M \times N_{\text{pix}}$, where N_{pix} denotes the total number of pixels and M is the number of ray integrations defined by

$$M = N_{\text{views}} \times N_{\text{bins}}$$

with N_{views} being the number of views (i.e., 2π arc is divided into N_{views} equally spaced angular intervals) and N_{bins} being the number of bins on the detectors (i.e., the detector is equally divided into N_{bins}). Before we discuss the sparse-view CT, we define the full sampling based on the four sufficient-sampling conditions (SSCs) in [5], which is obtained by setting the sampling parameters N_{views} and N_{bins} for given N_{pix} to characterize the invertibility and stability of the system matrix. The first pair of the SSCs characterizes invertibility of A , that is

$$\text{SSC1} : M \geq N_{\text{pix}} \quad \text{and} \quad \text{SSC2} : \sigma_{\min} \neq 0,$$

where σ_{\min} is the smallest singular value of A . The other pair of the SSCs characterizes the numerical stability for inversion of A , which is defined as

$$\text{SSC3} : \frac{\kappa(A)}{\kappa_{DC}} < r_{\text{samp}} \quad \text{and} \quad \text{SSC4} : N_{\text{views}} = N_{\text{bins}} = 2N,$$

where $\kappa(A) = \frac{\sigma_{\max}}{\sigma_{\min}}$, $\kappa_{DC} = \lim_{\substack{N_{\text{bins}} \rightarrow \infty \\ N_{\text{views}} \rightarrow \infty}} \kappa(A)$, r_{samp} is a finite ratio parameter greater than 1, and N is the length of the field-of-view (ROV) of the detector ($N_{\text{pix}} \approx \frac{\pi}{4} N^2$). As shown, both the SSC1 and SSC4 are simple to evaluate, which will be used to define the full-sampling system matrix in our work.

In contrast, when the N_{views} is not large enough to meet the SSCs for the fixed N_{bins} , it can be regarded as the sparse-view CT problem. Our goal is to develop efficient reconstruction methods for such ill-posed inverse problem. Since the full-sampling system matrix can be constructed according to the SSCs, we directly bridge the completed Radon domain data $f \in Z$ and the reconstructed image $u \in X$ through a full-sampling system matrix such that

$$f = A_F u,$$

where $A_F : X \rightarrow Z$ is the full-sampling projection matrix and Z is a Hilbert space. Therefore, we propose the following minimization model to jointly reconstruct the spatial and Radon domain data for sparse-view CT

$$\min_{u \in X, f \in Z} \mathcal{D}(A_S u, g) + \mathcal{R}(u) + \mathcal{F}(A_F u, f), \quad (4)$$

where $\mathcal{F}(A_F u, f)$ is used to measure the distance between $A_F u$ and f . Since the unknown u and f are coupled together in (4), we introduce a new variable \tilde{u} and rewrite (4) by adding a fitting term $\|\tilde{u} - u\|^2$ as follows

$$\min_{u \in X, f \in Z, \tilde{u} \in X} \mathcal{D}(A_S \tilde{u}, g) + \mathcal{R}(\tilde{u}) + \mathcal{F}(A_F u, f) + \frac{1}{2r} \|\tilde{u} - u\|_X^2, \quad (5)$$

where r is a positive parameter used to measure the trade-off between the under-sampling data g and a full-sampling projected data f . The first term in (5) contains the linear operator A_S , which can be reformulated based on the Legendre-Fenchel conjugate [2]

$$\min_{u \in X, f \in Z, \tilde{u} \in X} \max_{p \in Y} \langle A_S \tilde{u}, p \rangle - \mathcal{D}^*(p, g) + \mathcal{R}(\tilde{u}) + \mathcal{F}(A_F u, f) + \frac{1}{2r} \|\tilde{u} - u\|_X^2, \quad (6)$$

where \mathcal{D}^* denotes the conjugate of \mathcal{D} . The classical alternating direction method can be used to obtain an efficient algorithm for the multiple variable minimization problem (6), which gives

$$\begin{cases} p^{k+1} = \arg \min_{p \in Y} \mathcal{D}^*(p, g) - \langle A_S \tilde{u}^k, p \rangle + \frac{1}{2\tau} \|p - p^k\|_Y^2, \\ \tilde{u}^{k+1} = \arg \min_{\tilde{u} \in X} \mathcal{R}(\tilde{u}) + \langle A_S \tilde{u}, p^{k+1} \rangle + \frac{1}{2r} \|\tilde{u} - u^k\|_X^2, \\ f^{k+1} = \arg \min_{f \in Z} \mathcal{F}(A_F u^k, f) + \frac{1}{2\sigma} \|f - f^k\|_Z^2, \\ u^{k+1} = \arg \min_{u \in X} \mathcal{F}(A_F u, f) + \frac{1}{2r} \|u - \tilde{u}^{k+1}\|_X^2, \end{cases} \quad (7)$$

where τ and σ are positive parameters. As shown, the proximal method is adopted for the subproblem with respect to p and f in case the likelihood functional $\mathcal{D}(\cdot, \cdot)$ and $\mathcal{F}(\cdot, \cdot)$ are non-smooth. The solutions to each subproblem can be expressed as follows

$$\begin{cases} p^{k+1} = (\mathcal{I} + \tau \partial \mathcal{D}^*)^{-1}(p^k, \tau A_S \tilde{u}, g), \\ \tilde{u}^{k+1} = (\mathcal{I} + r \partial \mathcal{R})^{-1}(u^k, r A_S^* p^{k+1}), \\ f^{k+1} = (\mathcal{I} + \sigma \partial \mathcal{F})^{-1}(f^k, \sigma A_F u^k), \\ u^{k+1} = (\mathcal{I} + r \partial \mathcal{F})^{-1}(\tilde{u}^{k+1}, r A_F^* f^{k+1}). \end{cases} \quad (8)$$

Guided by the success of deep learning, we use CNN for unrolled iterative scheme such that the network can learn how to combine the variables in the object functional, which accounts for a deep feed-forward neural network by using CNNs to approximate the inverse operators in (8). The alternating direction algorithm with I iterations is outlined as Algorithm 1.

Algorithm 1 Learned Full Sampling Reconstruction (FSR)

- 1: Initialize $u^0, f^0, p^0, \tilde{u}^0$
 - 2: **for** $k = 1, \dots, I$, **do**
 - 3: $p^{k+1} \leftarrow \Gamma_{\theta^p}(p^k, A_S u^k, g)$
 - 4: $\tilde{u}^{k+1} \leftarrow \Lambda_{\theta^{\tilde{u}}}(u^k, A_S^* p^{k+1})$
 - 5: $f^{k+1} \leftarrow \Pi_{\theta^f}(f^k, A_F \tilde{u}^{k+1})$
 - 6: $u^{k+1} \leftarrow \Theta_{\theta^u}(\tilde{u}^{k+1}, A_F^* f^{k+1})$
 - 7: **return** u^I, f^I
-

Remark 1. In the algorithm, we assume the constraint $\tilde{u} = u$ holds unconditionally. Therefore, f^{k+1} is calculated based on $A_F \tilde{u}^{k+1}$ rather than $A_F u^k$ as \tilde{u}^{k+1} was already updated in the previous step. Besides, instead of selecting specific values for τ, σ and r , we let the network learn the appropriate value by itself.

3 Experiments and Results

In this section, we evaluate the proposed algorithm on both ellipse data [1] and piglet data³ by comparing with the state-of-the-art work, i.e., FBP-Unet denoising [4] and Leaned Primal-Dual [1].

3.1 Implementation

The methods are implemented in Python using Operator Discretization Library (ODL) and TensorFlow. We let the number of data that persists between the iterates be $N_u = N_{\tilde{u}} = 6$ and $N_p = N_f = 7$. The convolution are all 3×3 pixel size, and the numbers of channels in each iteration are p of $9 \rightarrow 32 \rightarrow 32 \rightarrow 7$, \tilde{u} of $7 \rightarrow 32 \rightarrow 32 \rightarrow 6$, f of $8 \rightarrow 32 \rightarrow 32 \rightarrow 7$ and u of $7 \rightarrow 32 \rightarrow 32 \rightarrow 6$. The network structure of one iteration is illustrated in Fig. 1, where totally 10 iterations are contained in our network. As shown, each iteration involves four 3-layer that is the depth of network is 120 layers.

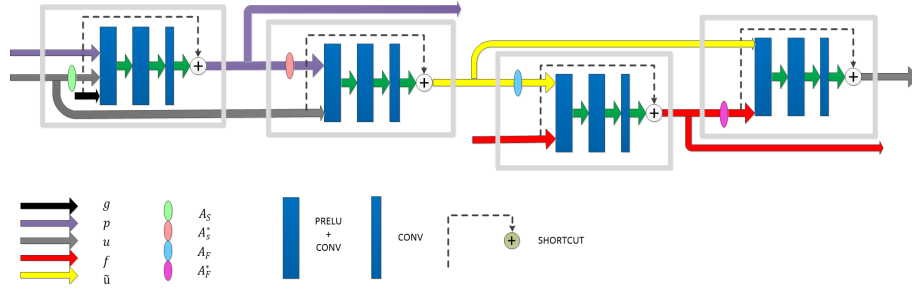


Fig. 1. Network architecture to solve the tomography problem. Each box corresponds to one variable, which are all of the same architecture.

We use the Xavier initialization scheme to initialize the convolution parameters, and initialize all biases to zero. Let $\Theta = \{\theta^p, \theta^{\tilde{u}}, \theta^f, \theta^u\}$ and \mathcal{T}^\dagger be the *pseudo-inverse* of the minimization process (4) defined as

$$\mathcal{T}_\Theta^\dagger(g) \approx (u_{\text{true}}, f_{\text{true}}) \quad \text{for data } g \text{ satisfying (1),}$$

Suppose $(\mathcal{T}_\Theta(u), \mathcal{T}_\Theta(f)) = \mathcal{T}_\Theta^\dagger(g)$ and $(g_1, u_1^*), (g_2, u_2^*), \dots, (g_L, u_L^*)$ be L training samples. We apply the ADAM optimizer in TensorFlow to minimize the following empirical loss function

$$\mathcal{L}(\Theta) = \frac{1}{2L} \sum_{i=1}^L \left(\|\mathcal{T}_\Theta(u_i) - u_i^*\|_X^2 + \|\mathcal{T}_\Theta(f_i) - A_F u_i^*\|_Z^2 \right). \quad (9)$$

Most parameters are set the same as the PD-net in [1]. We use 2×10^5 batches on each problem and a learning rate schedule according to cosine annealing, i.e., the learning

³ https://1drv.ms/u/s!Aj4lQl4ug0_9hEb1c6UzqHKxA4nZ

rate at step t is $\eta^t = \frac{\eta^0}{2}(1 + \cos(\pi \frac{t}{t_{max}}))$, where the initial learning rate is set as $\eta^0 = 10^{-3}$ for the ellipse data and $\eta^0 = 10^{-4}$ for the piglet phantom. We let the parameter β_2 of the ADAM optimizer to 0.99 and limit the gradient norms to 1 to improve training stability. The batch size is set as 5 and 1 for the ellipse data and piglet phantom, respectively.

3.2 Results on Ellipse Phantoms

Similar to [1], we randomly generate ellipses on a 128×128 pixel domain by parallel beam projection geometry with $N_{bins} = 128$ and $N_{views} = 15, 30$. Both 5% and 10% additive white Gaussian noises are added to the projection data. We use the full sampling system matrix provided by ODL for parallel beam CT as A_F in our model (4). Table 1 presents the PSNR and SSIM obtained by the CNN based models. It is obviously shown that the best PSNR values are always achieved by our FSR-net and PD-net ranks the second position, both of which are significantly better than the Unet based post-processing method. Especially, the advantage of our FSR-net over PD-net becomes more convincing for $N_{views} = 15$ and 5% Gaussian noise, giving an improvement exceeding 1 dB, which demonstrates the effectiveness of our model in sparse-view reconstruction. The comparison of the PSNR and SSIM between the FBP with g and FBP using the reconstructed projection data f from our model in Table 1 also demonstrates that our model can recover the Radon domain data to certain qualities. We display the reconstruction results of the sparse 15 views with 5% Gaussian noise in Fig. 2, which shows that our reconstruction preserves the geometry and details better than the other two methods.

N_{views}	Noises	Methods	PSNR	SSIM	time	FBP(PSNR)	FBP(SSIM)
30	5%	FBP-Unet	28.1693	0.9346	1.5	19.7216	0.5938
		PD-net	39.2301	0.9860	4.6		
		FSR-net	39.6588	0.9897	6.6	21.5168	0.8930
	10%	FBP-Unet	26.2985	0.9250	1.5	19.1199	0.5073
		PD-net	31.9033	0.9707	4.6		
		FSR-net	32.1213	0.9695	6.6	21.2971	0.8881
15	5%	FBP-Unet	19.8411	0.7224	1.4	16.1378	0.4217
		PD-net	30.3615	0.9719	4.4		
		FSR-net	32.0468	0.9707	6.4	21.3333	0.8607
	10%	FBP-Unet	19.4893	0.7370	1.4	15.5722	0.3394
		PD-net	25.1603	0.9301	4.4		
		FSR-net	25.9528	0.9238	6.4	20.8019	0.8470

Table 1. Comparison of reconstruction methods for the ellipses in terms of PSNR, SSIM and Runtime.

In addition, we apply the parameters trained by the sparse 30 views and 5% Gaussian noise to testing data with different sparsity, i.e., g obtained by $N_{views} = 30, 25, 20$. We compare the results with the learned PDHG net, learned primal net and learned primal-dual net from [1] in terms of PSNR and SSIM in Table 2. As shown, our model

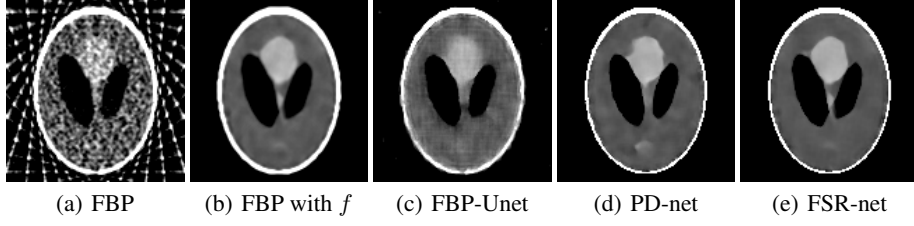


Fig. 2. Reconstruction comparison on the ellipse data, where the window is set to $[0.1, 0.4]$.

performs more stable in adapting with different testing data, which is because our model tries to minimize the distance between the reconstructed image and the full-sampling projection data.

N_{views}	30		25		20	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PDHG-net	29.6762	0.9111	27.4642	0.8602	22.8205	0.6866
Primal-net	37.2040	0.9848	35.4175	0.9812	33.1559	0.9655
PD-net	39.2301	0.9860	36.4508	0.9757	34.6969	0.9365
FSR-net	39.6588	0.9897	38.6696	0.9853	35.8236	0.9769

Table 2. Reconstruction comparison on the ellipse data of learning-based models with different sparsities.

3.3 Results on piglet phantom data

We test the proposed model on simulated CT data of a deceased piglet, which is scanned from a 64-slice multi-detector CT scanner (Discovery CT750 HD, GE Healthcare) using 100kV and 0.625mm slice thickness. We use 896 images of size 512×512 as the ground truth for training and 10 for evaluation. We adopt the fan-beam geometry with $N_{\text{bins}} = 512$ and $N_{\text{bins}} = 1024$, source to axis distance 500mm and axis to detector distance 500mm. The number of views is set as follows

- For $N_{\text{bins}} = 512$, the observed data g is generated by 64 uniformly distributed views over 2π arc with two different Poisson noises of 10^4 and 5×10^5 incident photons per pixel before attenuation. The full-sampling system matrix A_F is constructed according to SSC1, i.e., $N_{\text{views}} = 512$;
- For $N_{\text{bins}} = 1024$, the observed data g is generated with either 120 views or 60 views and Poisson noise of 10^4 incident photons. The full-sampling system matrix A_F is defined according to SSC4, i.e., $N_{\text{views}} \approx 1024$. Because $N_{\text{views}} \approx 1024$ gives too much computational burden, we use $N_{\text{views}} = 720$ in practice.

As shown in Table 3 and 4, our model outperforms other methods in reconstruction quality. Especially when we use the parameters trained by $N_{\text{views}} = 64$ to reconstruct the sparse data such as $N_{\text{views}} = 32, 28, 24$, our model achieves a PSNR 0.5 ~ 3dB higher than PD-net. Both the reconstructed images and the error maps are displayed in Fig. 3, the first column displays the FBP reconstruction of observed data g (row

one) and our estimated full-sampling measurement f (row two). It is obviously shown that our model can well inpaint the Radon domain data and improve the reconstruction quality.

N_{bins}	N_{views}	Noises	Methods	PSNR	SSIM	time	FBP/PSNR	FBP/SSIM
512	64	5×10^5	FBP-Unet	32.165	0.992	5.2	27.02	0.832
			PD-net	36.202	0.997	5.3		
			FSR-net	36.875	0.997	6.6	27.423	0.9222
	64	10^4	FBP-Unet	28.917	0.9853	5.2	22.7613	0.4196
			PD-net	29.999	0.9887	5.3		
			FSR-net	30.564	0.9903	6.6	27.3457	0.9182
1024	120	10^4	FBP-Unet	30.8926	0.9898	5.3	20.5832	0.3023
			PD-net	32.382	0.9933	5.4		
			FSR-net	33.084	0.9941	6.7	27.3967	0.921
	60	10^4	FBP-Unet	28.192	0.9793	5.5	18.0077	0.1658
			PD-net	31.263	0.9913	5.6		
			FSR-net	31.763	0.9921	7	27.3917	0.9198

Table 3. Comparison of reconstruction methods for a piglet phantom in terms of PSNR, SSIM and Runtime (s).

N_{views}	32		28		24	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FBP-Unet	23.44	0.9513	20.1502	0.9008	18.9658	0.855
PD-net	25.27	0.9761	21.0392	0.9453	18.333	0.906
FSR-net	25.86	0.973	23.0581	0.9534	21.1436	0.9304

Table 4. Reconstruction comparison on the piglet dataset of learning-based models with different sparsities.

4 Conclusion

We proposed a novel iterative reconstruction model by fitting the reconstructed image with its corresponding measurements in Radon domain through the full-sampling system matrices. This new algorithm is in the family of deep learning based iterative reconstruction schemes. The application on sparse-view CT image reconstruction demonstrates the effectiveness of the proposed model and it is also clearly shown that the proposed method can be applied to other applications such as limited-angle CT reconstruction and compressed-sensing MR reconstruction.

References

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

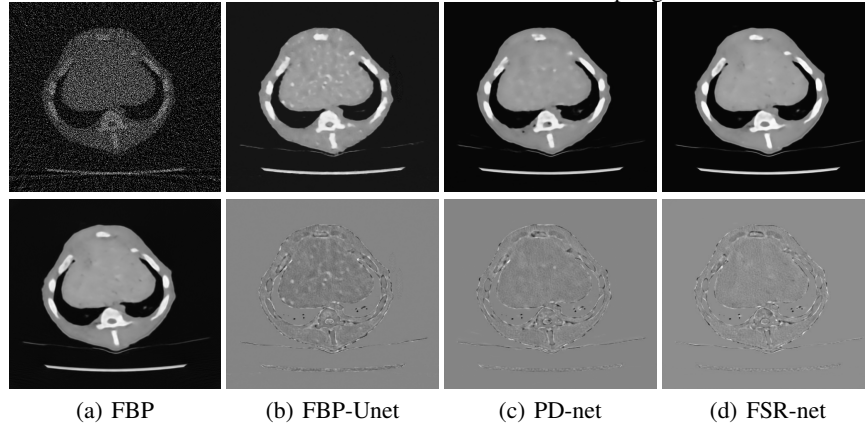


Fig. 3. Reconstruction comparison of a piglet phantom with $N_{\text{views}} = 64$ and $N_{\text{bins}} = 512$.

2. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision* **40**(1), 120–145 (2011)
3. Dong, B., Li, J., Shen, Z.: X-ray ct image reconstruction via wavelet frame based regularization and radon domain inpainting. *Journal of Scientific Computing* **54**(2-3), 333–349 (2013)
4. Jin, K.H., McCann, M.T., Froustey, E., Unser, M.: Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing* **26**(9), 4509–4522 (2017)
5. Jorgensen, J.S., Sidky, E.Y., Pan, X.: Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in X-ray CT. *IEEE Transactions on Medical Imaging* **32**(2), 460–473 (2013)
6. Liu, J., Kuang, T., Zhang, X.: Image reconstruction by splitting deep learning regularization from iterative inversion. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 224–231 (2018)
7. Sidky, E.Y., Jørgensen, J.H., Pan, X.: Convex optimization problem prototyping for image reconstruction in computed tomography with the chambolle–pock algorithm. *Physics in Medicine & Biology* **57**(10), 3065 (2012)
8. Yang, Y., Sun, J., Li, H., Xu, Z.: Deep ADMM-Net for compressive sensing MRI. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*. pp. 10–18 (2016)
9. Zhang, H., Dong, B., Liu, B.: Jsr-net: A deep network for joint spatial-radon domain CT reconstruction from incomplete data. *arXiv preprint arXiv:1812.00510* (2018)