Supplementary Material

A. ADMM-TV

In this section, we derive the ADMM-TV optimization framework for completeness. We are interested in solving the problem of TV-regularized WLS of the following form:

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{D}_z\boldsymbol{x}\|_1, \tag{19}$$

where $\boldsymbol{D}_z$ takes the finite difference across the $z-$dimension. In order to solve the problem in an alternating fashion, we split the variables

$$\min_{\boldsymbol{x},\boldsymbol{z}} \quad \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{z}\|_1 \tag{20}$$

$$\text{s.t.} \quad \boldsymbol{z} = \boldsymbol{D}_z\boldsymbol{x}. \tag{21}$$

The scaled formulation of ADMM [3] is then given by

$$\boldsymbol{x}^+ = \operatorname*{argmin}_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2 + \frac{\rho}{2}\|\boldsymbol{D}_z\boldsymbol{x}-\boldsymbol{z}+\boldsymbol{w}\|_2^2 \tag{22}$$

$$\boldsymbol{z}^+ = \operatorname*{argmin}_{\boldsymbol{z}} \lambda\|\boldsymbol{z}\|_1 + \frac{\rho}{2}\|\boldsymbol{D}_z\boldsymbol{x}^+-\boldsymbol{z}+\boldsymbol{w}\|_2^2 \tag{23}$$

$$\boldsymbol{w}^+ = \boldsymbol{w} + \boldsymbol{D}_z\boldsymbol{x}^+ - \boldsymbol{z}^+. \tag{24}$$

(22) is convex and smooth, and thus has a closed form solution

$$\boldsymbol{x}^+ = (\boldsymbol{A}^T\boldsymbol{A} + \rho\boldsymbol{D}_z^T\boldsymbol{D}_z)^{-1}(\boldsymbol{A}^T\boldsymbol{y} + \rho\boldsymbol{D}_z^T(\boldsymbol{z}-\boldsymbol{w})), \tag{25}$$

where one can perform CG rather than computing the matrix inverse directly. In order to solve, (23), we define the proximal operator [24] as

$$\text{prox}_{f,\eta}(\boldsymbol{z}) \triangleq \operatorname*{argmin}_{\boldsymbol{x}} f(\boldsymbol{x}) + \frac{1}{2\eta}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2. \tag{26}$$

By inspecting (23), we know that it is in the form of proximal mapping

$$\boldsymbol{z}^+ = \text{prox}_{\|\cdot\|_1,\lambda/\rho}(\boldsymbol{D}_z\boldsymbol{x}+\boldsymbol{w}) \tag{27}$$

$$= \mathcal{S}_{\lambda/\rho}(\boldsymbol{D}_z\boldsymbol{x}+\boldsymbol{w}), \tag{28}$$

where we have leveraged the fact that the proximal mapping of the ℓ_1 norm is given as the soft thresholding operator $\mathcal{S}$. In summary, we have

$$\begin{aligned}
\boldsymbol{x}^+ &= (\boldsymbol{A}^T\boldsymbol{A} + \rho\boldsymbol{D}_z^T\boldsymbol{D}_z)^{-1}(\boldsymbol{A}^T\boldsymbol{y} + \rho\boldsymbol{D}_z^T(\boldsymbol{z}-\boldsymbol{w})) \\
\boldsymbol{z}^+ &= \mathcal{S}_{\lambda/\rho}(\boldsymbol{D}_z\boldsymbol{x}^+ + \boldsymbol{w}) \\
\boldsymbol{w}^+ &= \boldsymbol{w} + \boldsymbol{D}_z\boldsymbol{x}^+ - \boldsymbol{z}^+.
\end{aligned}$$

The algorithmic detail can be found in Algorithm 2.

Algorithm 2 DiffusionMBIR (slow)

Require: $s_\theta, N, M, K, \lambda, \rho, \{\sigma_i\}$
$\boldsymbol{x}_N \sim \mathcal{N}(\boldsymbol{0}, \sigma_T^2\boldsymbol{I})$
for $i = N-1:0$ **do** ▷ SDE iteration
  $\bar{\boldsymbol{x}}_i \leftarrow \text{Solve}(\boldsymbol{x}_{i+1}, \boldsymbol{s}_{\theta^*})$
  $\boldsymbol{A}_{\text{CG}} \leftarrow \boldsymbol{A}^T\boldsymbol{A} + \rho\boldsymbol{D}_z^T\boldsymbol{D}_z$
  $\boldsymbol{z}^{(1)} \leftarrow \texttt{torch.zeros\_like}(\bar{\boldsymbol{x}}_i)$
  $\boldsymbol{w}^{(1)} \leftarrow \texttt{torch.zeros\_like}(\bar{\boldsymbol{x}}_i)$
  for $j = 1:M$ **do** ▷ ADMM iteration
    $\boldsymbol{b}_{\text{CG}}^{(j)} \leftarrow \boldsymbol{A}^T\boldsymbol{y} + \rho\boldsymbol{D}^T(\boldsymbol{z}^{(j)} - \boldsymbol{w}^{(j)})$
    $\bar{\boldsymbol{x}}_i^{(j+1)} \leftarrow \text{CG}(\boldsymbol{A}_{\text{CG}}, \boldsymbol{b}_{\text{CG}}^{(j)}, K)$ ▷ CG iteration
    $\boldsymbol{z}^{(j+1)} \leftarrow \mathcal{S}_{\lambda/\rho}(\boldsymbol{D}_z\bar{\boldsymbol{x}}_i^{(j+1)} + \boldsymbol{w}^{(j)})$
    $\boldsymbol{w}^{(j+1)} \leftarrow \boldsymbol{w}^{(j)} + \boldsymbol{D}_z\bar{\boldsymbol{x}}_i^{(j+1)} - \boldsymbol{z}^{(j+1)}$
  end for
  $\boldsymbol{x}_i \leftarrow \bar{\boldsymbol{x}}_i^{(M+1)}$
end for
return $\boldsymbol{x}_0$

Patient ID	# slices	FOV (mm)	KVP	Exposure time (ms)	X-ray tube current (mA)
L097	500	430	120	500	327.6
L109	254	400	100	500	322.3
L143	418	440	120	500	416.9
L192	370	380	100	500	431.6
L286	300	380	120	500	328.9
L291	450	380	120	500	322.7
L310	340	380	120	500	300.0
L333	400	400	100	500	348.7
L506	300	380	100	500	277.7
L067 (test)	448	370	100	500	234.1

Table 2. AAPM dataset specification. L067 volume is used for testing while the other volumes are used for training.

B. Details of experiment

B.1. Dataset

AAPM. We take the dataset from the AAPM 2016 CT low-dose grand challenge, where the data are acquired in a fan-beam geometry with varying parameters, as presented in Table 2. The data preparation steps follow that of [13]. From the helical cone beam projections, approximation to fanbeam geometry is performed via single-slice rebinning technique [23]. Reconstruction is then performed via standard filtered backprojection (FBP), where the reconstructed axial images have the matrix size of 512×512. We resize the axial slices to have the size 256×256, and use these slices to train the score function. The whole dataset consists of 9 volumes (3332 slices) of training data, and 1 volume (448 slices) of testing data. To generate sparse-view measurements, we retrospectively employ the parallel-view geometry for simplicity.

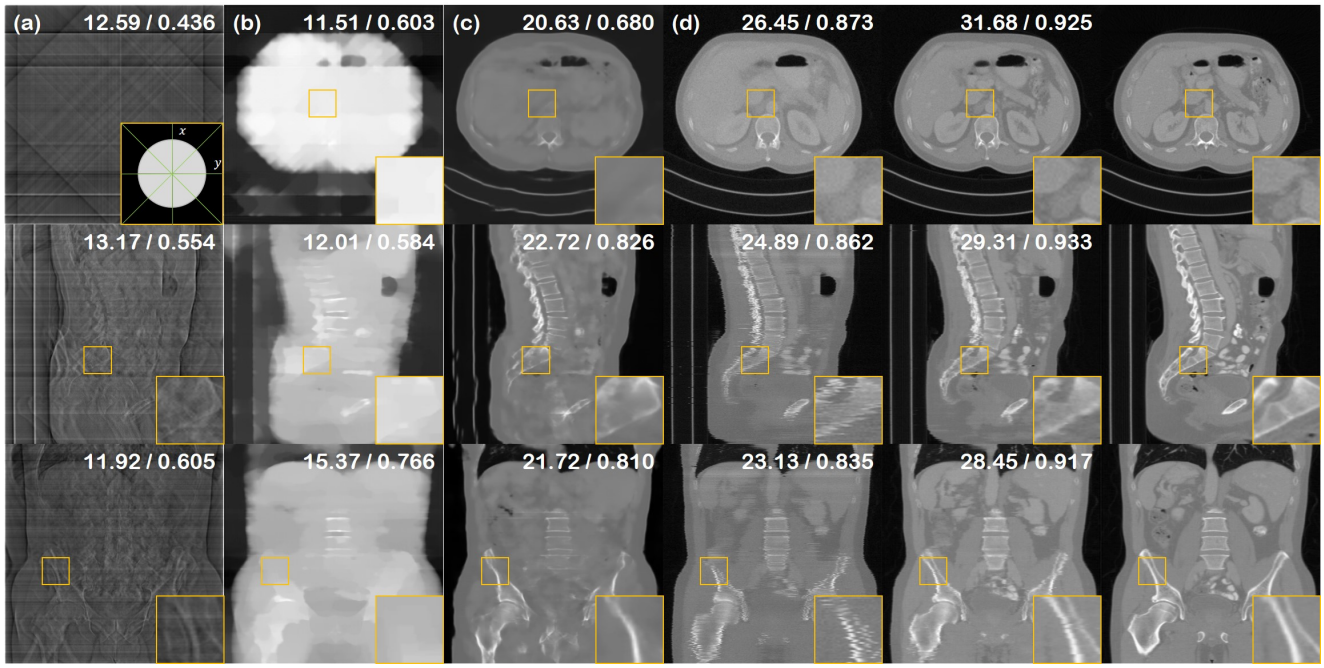

Figure 6. 4-view SV-CT reconstruction results of the test data (First row: axial slice, second row: sagittal slice, third row: coronal slice). (a) FBP, (b) ADMM-TV, (c) Lahiri *et al.* [16], (d) Chung *et al.* [5], (e) proposed method, (f) ground truth. Green lines in the inset of first row (a): measured angles.

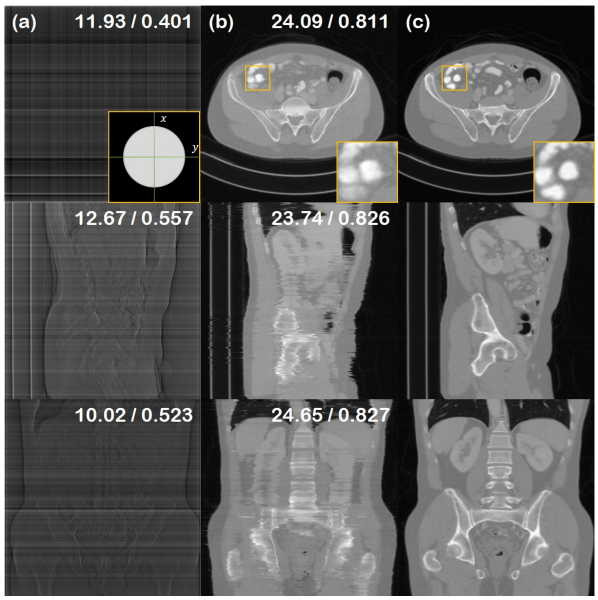

Figure 7. 2-view SV-CT reconstruction results of the test data (First row: axial slice, second row: sagittal slice, third row: coronal slice). (a) FBP, (b) proposed method, (f) ground truth. Green lines in the inset of first row (a): measured angles.

BRATS. We take the dataset from the multimodal brain tumor segmentation BRATS 2018 challenge [22], where we select the test data as the first FLAIR volume, which has the matrix size of $240 \times 240 \times 154$. As stated in the main text, all the methods were trained with the separate fastMRI 2019 knee database [36].

B.2. Details of network training

For the CT score function, we train the `ncsnpp` network [32] without modifications with (6) by setting $\lambda = \sigma^2(t)$ [29], and $\varepsilon = 10^{-5}$. Our network is trained using the Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$) with a linear warm-up schedule, reaching 2×10^{-4} at the 5000^{th} step, and with a batch size of 2 using a single RTX 3090 GPU for 200 epochs. Training took about a week and a half.

B.3. Comparison methods

Chung *et al.* [5] As the method is based on diffusion models, we use the same pre-trained score function, and use the reconstruction scheme of [5], which amounts to applying ART after every PC update steps.

Lahiri *et al.* [16] We use 2-stage reconstructing 3D CNNs, where we train the networks with slabs, not taking patches in the xy dimension, as the paper suggests. The architecture of CNN was taken as a standard U-Net [36] architecture rather than stack of single-resolution CNNs, as we achieved better performance with U-Nets. Furthermore, we drop the adversarial loss and only use the standard reconstruction loss, as we found the training to be more stable. CG was

Method	Axial*		Coronal		Sagittal	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DiffusionMBIR (ours)	**34.92**	**0.956**	**32.48**	**0.947**	**28.82**	0.832
Chung *et al.* [5]	26.01	0.838	24.55	0.823	21.59	0.706
Lahiri *et al.* [16]	28.08	0.931	26.02	0.856	23.24	0.812
Zhang *et al.* [38]	26.76	0.879	25.77	0.874	22.92	**0.841**
ADMM TV	23.19	0.793	22.96	0.758	19.95	0.782

Table 3. Quantitative evaluation of LA-CT (90°) (PSNR, SSIM) on the AAPM 256×256 test set. **Bold**: Best, under: second best.

Method	Axial*		Coronal		Sagittal	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DiffusionMBIR (ours)	**41.49**	**0.974**	**37.36**	**0.942**	**37.18**	**0.953**
Score-MRI [7]	40.38	0.968	33.97	0.925	34.02	0.928
DuDoRNet [16]	39.78	**0.974**	33.56	0.927	33.48	0.927
Unet [11]	37.15	0.929	31.56	0.899	30.90	0.816
Zero-filled	34.18	0.923	29.53	0.897	27.82	0.903

Table 4. Quantitative evaluation of CS-MRI (acc. ×2) (PSNR, SSIM) on the BRATS data. **Bold**: Best, under: second best.

applied with 30 iterations.

FBPConvNet [11], Zhang *et al.* [37] We use the same U-Net architecture that was used to train Lahiri *et al.* [16], but only on 2D images. Note that the original work of Zhang *et al.* [37] uses a much simpler CNN architecture, which leads to degraded performance.

ADMM-TV. We minimize the following objective

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{D}\boldsymbol{x}\|_{2,1}, \tag{29}$$

where $\boldsymbol{D} = [\boldsymbol{D}_x, \boldsymbol{D}_y, \boldsymbol{D}_z]$, which corresponds to the isotropic TV. The outer iterations are solved with ADMM (30 iterations), while the inner iterations are solved with CG (20 iterations). We perform coarse grid search to find the parameter values that produce low MSE values, then perform another grid search to find the most visually pleasing solution — images with salient edges. The parameter is set to $(\lambda, \rho) = (0.5, 50)$ for SV-CT, and $(\lambda, \rho) = (0.15, 40)$ for LA-CT.

Score-MRI. [7] The method utilizes the same score function as the proposed method, but relies on iterated projections onto the measurement subspace after every iteration of the PC update. Every slices in the xy dimension are reconstructed separately, then stacked to form the volume.

DuDoRNet. [39] We train DuDoRNet with 4 recurrent blocks and the default parameters, following the configuration of [7]. The network is trained on the fastMRI knee dataset, with the proton density (PD) / proton density fat suppressed (PDFS) image for the prior information.

U-Net. [36] We leverage the pre-trained U-Net on the fastMRI dataset, trained only with the L1 loss in the image domain.

C. Further experiments

C.1. Additional experimental results

4-view sparse view tomographic reconstruction is presented in Fig. 6. Furthermore, we demonstrate that we can even perform *2-view* reconstruction, as can be seen in Fig. 7. In this regime, the information contained in the measurement is very few and sparse — clearly not sufficient for achieving an *accurate* reconstruction. As we are leveraging the generative prior however, we can sample multiple reconstructions that are 1) perfectly measurement feasible, and 2) looks realistic. Although this might not be of significant importance in the medical imaging field, it could greatly impact fields where approximate reconstructions from very limited acquisitions are necessary.

C.2. Number of views vs. performance

We have verified that we can perform extreme sparse-view reconstruction with the proposed method. One can naturally ask the limit of the proposed method, and the trend between the measured number of views versus the reconstruction performance, which we show in the plot shown in Fig. 9. Note that the given plot is a *log* plot in the x-axis (i.e. # views). We can easily see that the performance caps if we increase the number of measurements to higher than 16, and we also see that down to 8-views, we can acquire reconstructions with only a modest drop in the performance. The performance starts to heavily degrade as we drop down the number of views below 4. We conclude that there is a singular point, where the information in the measurement is just not enough, even when we have a very strong reconstruction algorithm.

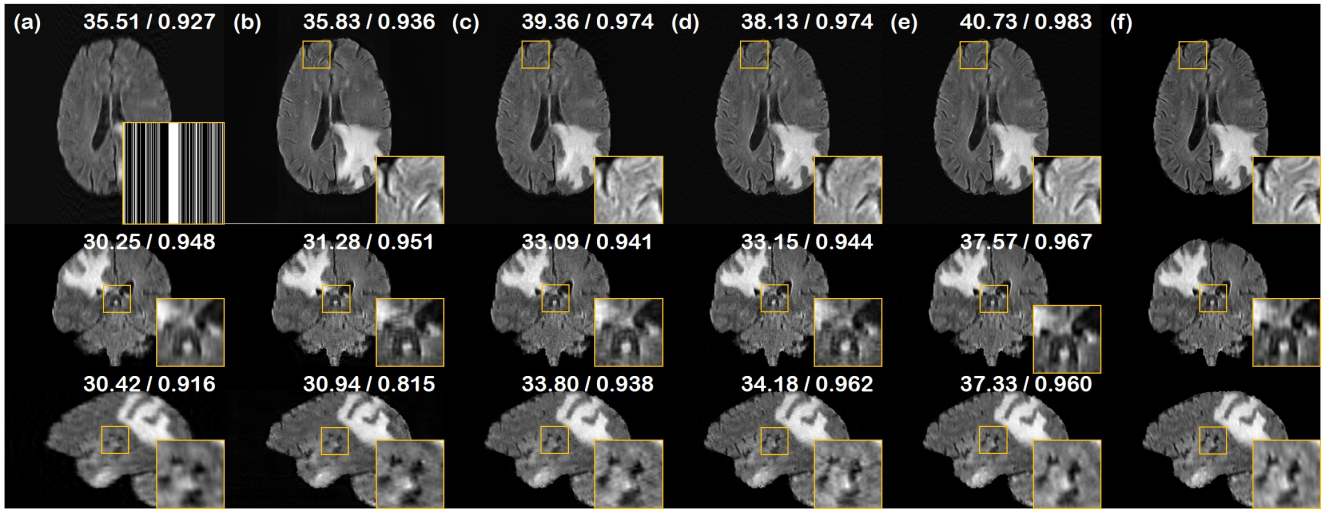

Figure 8. CS-MRI results of the test data (First row: axial slice, second row: sagittal slice, third row: coronal slice). (a) zero-filled, (b) U-Net [36], (c) DuDoRNet [39], (d) Score-MRI [7], (e) proposed method, (f) ground truth. PSNR/SSIM values presented in the upper right corner. Mask presented in the first row of (a): Sub-sampling mask applied to all slices.

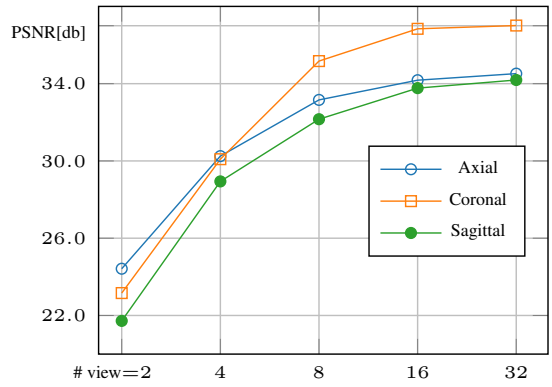

Figure 9. Number of measured views vs. PSNR[db]

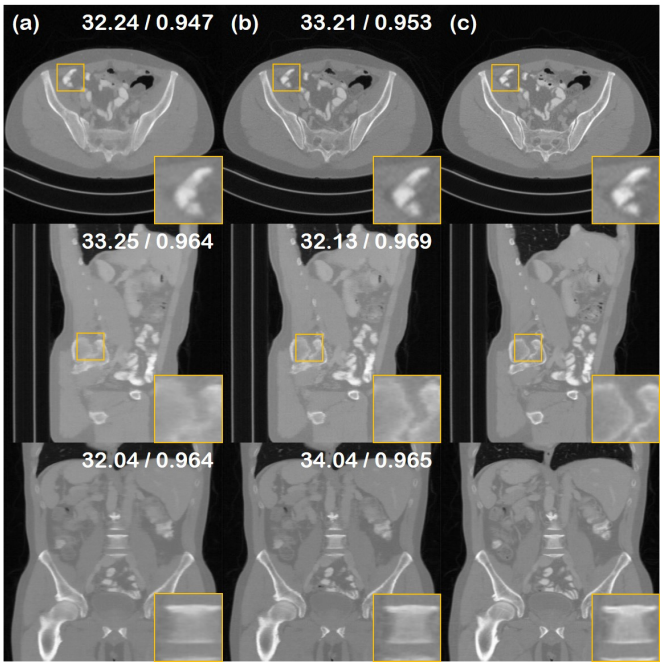

Figure 10. Ablation study for the choice of augmented prior. (a) TV (xyz) prior, (b) TV (z) prior; proposed method, (c) Ground truth.