Supporting Information

A scalable neural network architecture for self-supervised tomographic image reconstruction

Hongyang Dong[1], Simon D. M. Jacques[2], Winfried Kockelmann[3], Stephen W. T. Price[2], Robert Emberson[4], Dorota Matras[5,6], Yaroslav Odarchenko[2], Vesna Middelkoop[7], Athanasios Giokaris[2], Olof Gutowski[8], Ann-Christin Dippel[8], Martin von Zimmermann[8], Andrew M. Beale[1,2,9], Keith T. Butler[10,11*], Antonis Vamvakeros[2*]

1. Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, United Kingdom
2. Finden Ltd, Building R71, Rutherford Appleton Laboratory, Harwell, Oxford, OX11 0QX, United Kingdom
3. STFC, Rutherford Appleton Laboratory, ISIS Facility, Harwell, OX11 0QX, United Kingdom
4. Department of Mathematics & Statistics, Lancaster University, Bailrigg, Lancaster, LA1 4YW, United Kingdom
5. Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, United Kingdom
6. The Faraday Institution, Quad One, Harwell Science and Innovation Campus, Didcot, OX11 0RA, United Kingdom
7. Flemish Institute for Technological Research (VITO), B-2400 Mol, Belgium.
8. Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany
9. Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, Oxon OX11 0FA, United Kingdom
10. SciML, Scientific Computing Department, STFC Rutherford Appleton Laboratory, Harwell Campus, Didcot, OX11 0QX, United Kingdom
11. School of Engineering and Materials Science, Queen Mary University of London, Mile End Rd, Bethnal Green, London E1 4NS, United Kingdom

Self-supervised learning: SD2I

SD2I architecture

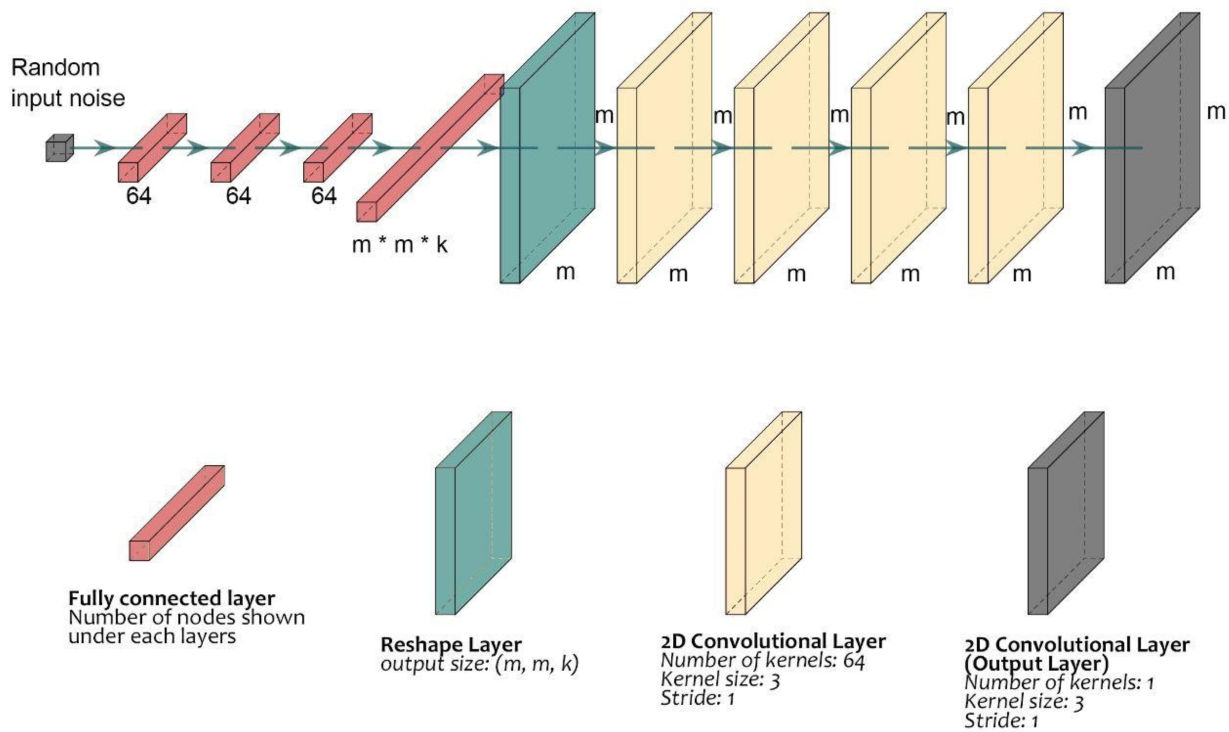


Figure S1: A representation of the CNN reconstruction SD2I architecture. The kernel types and parameter settings are shown in the figure. The final fully connected layer size is adjusted by an integer k, which adjusts the number of kernels used as the input of the following reshape and convolutional layers.

SD2I: the impact of k factor

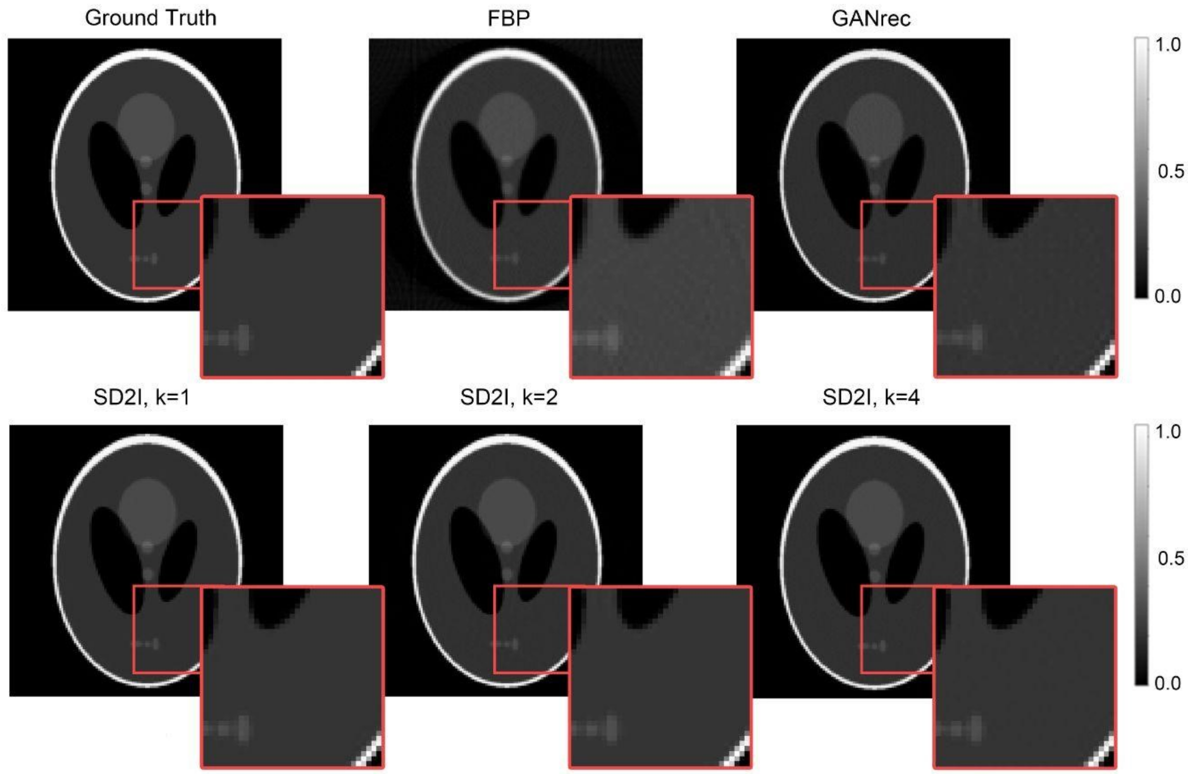


Figure S2. TaShepp-Logan, image size: 128x128, sinogram size: 128x200.

Table S1. Shepp-Logan, image size: 128x128, sinogram size: 128x200.

	GANrec	SD2I - no upsampling (k = 1)	SD2I - no upsampling (k = 2)	SD2I - no upsampling (k = 4)	FBP
Number of parameters	8,698,242	2,337,154	4,418,947	8,582,533	\
MAE	0.00384	0.00168	0.00150	0.00148	0.01088
MSE	5.55772×10^{-5}	1.03833×10^{-5}	1.02758×10^{-5}	0.957322×10^{-5}	0.001000
SSIM	0.9844	0.9913	0.9937	0.9979	0.9599
PSNR	42.55	49.79	49.88	50.19	30.00

Architecture: SD2Iu
Image size: 256 x 256
Sinogram size: 256 x 64

Number of epochs: 6000
Loss function: SSIM + MAE
Start learning rate: 0.0005
Ground truth type: Clean Shepp-Logan image

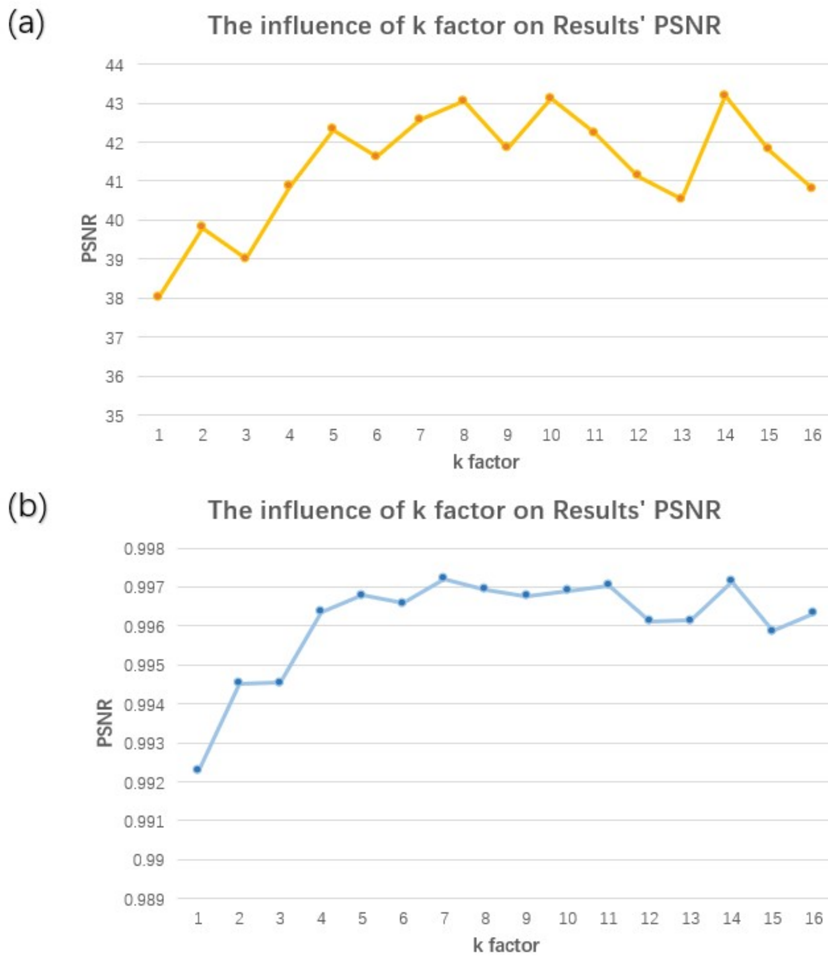


Figure S3. The influence of different k factors used in SD2I on the result's (a) PSNR and (b) SSIM. Using larger k factors can improve the quality of reconstructed results on both metrics. In practice, using a k factor between 4 and 8 is more appropriate for achieving a good balance between model size and accuracy.

SD2I: the impact of loss function

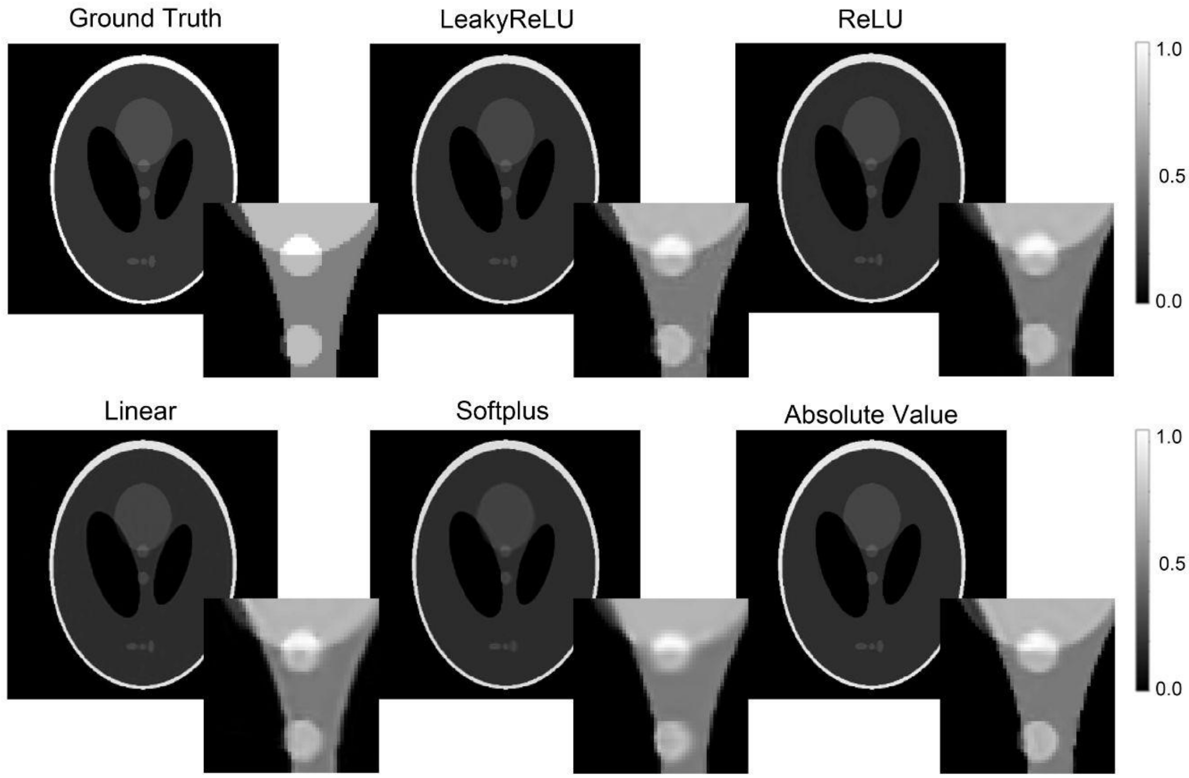


Figure S4. Impact of different choice of activation functions on the final layer on the Shepp-Logan phantom image with image size 256x256, sinogram size: 256x64.

Table S2. Metrics on the impact of different choice of activation functions.

	LeakyReLU	ReLU	Linear	Softplus	Absolute Value
MAE	0.002774	0.002715	0.003756	0.003533	0.002513
MSE	1.280×10^{-3}	1.168×10^{-3}	1.900×10^{-3}	1.604×10^{-3}	1.149×10^{-3}
SSIM	0.9957	0.9965	0.9920	0.9943	0.9967
PSNR	38.93	39.30	37.21	37.95	39.40

Vector input

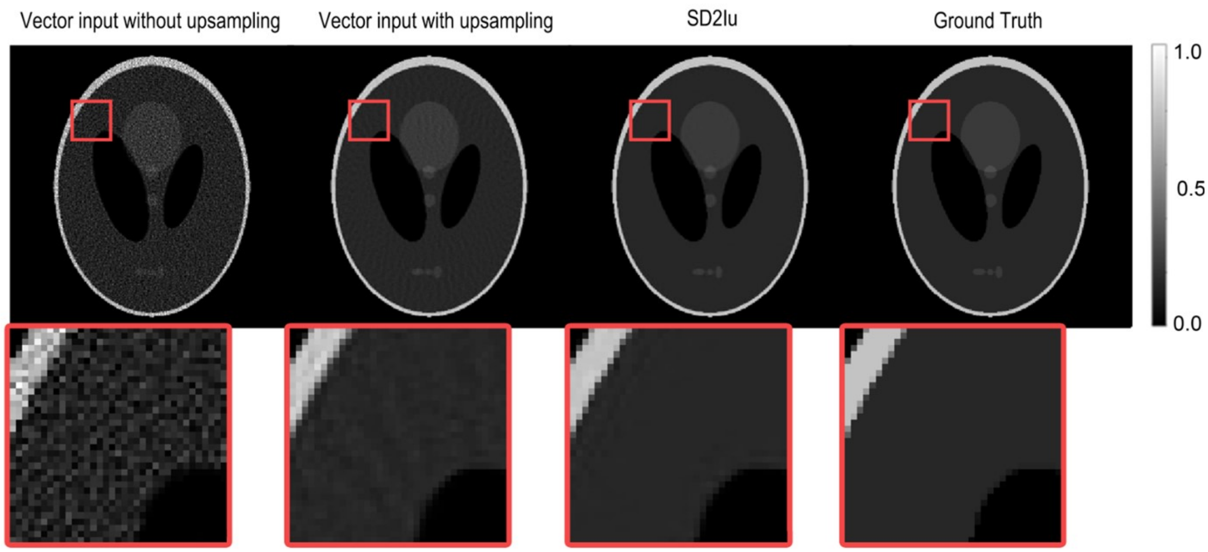


Figure S5. The result when using a vector of ones as input. The Vector input without/with upsampling represents the networks that started from the last fully connected layer on the SD2I/SD2Iu network respectively. Both networks receive a vector of 64 ones as input which has the same size as the fully connected layer before the last fully connected layer of both SD2I and SD2Iu networks. All images are 256x256 large and reconstructed from the 256x64 Shepp-Logan sinogram.

.

Pixel learning network

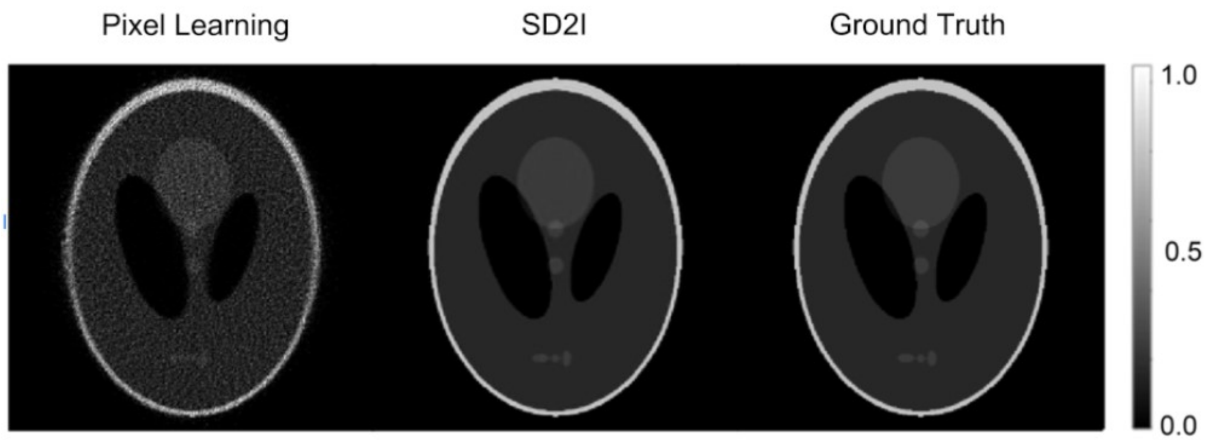


Figure S6. The Pixel learning result compared with the SD2I result on the reconstruction of the 256x256 Shepp-Logan image with 400 projections. The pixel learning network only has one large fully connected layer that receives a single digit of one as input.

SD2I scalability

Table S3. Scalability. Number of parameters (nop) per network as a function of image size

Sinogram/ Reconstructed Image size (pixels)	Automap (nop)	GANrec (nop)	SD2I (factor 8) (nop)	SD2I (factor 4) (nop)	SD2Iu (factor 8) (nop)	SD2Iu (factor 4) (nop)
64 x 64	79,356,424	2,443,778	4,360,585	2,307,973	331,457	262,593
128 x 128	1,304,960,136	8,698,242	16,909,705	8,582,533	730,817	462,273
256 x 256	-	33,814,658	67,370,377	33,812,869	2,328,257	1,260,993
512 x 512	-	134,477,442	269,741,449	134,998,405	8,718,017	4,455,873
1024 x 1024	-	537,522,818	1,080,282,505	540,268,933	34,277,057	17,235,393

SD2I reconstruction times

Table S4. Reconstruction times (time per epoch)

Sinogram/ Reconstructed Image size (pixels)	Automap (s)	GANrec (s)	SD2I (k = 8) (s)	SD2I (k = 4) (s)	SD2Iu (k = 8) (s)	SD2Iu (k = 4) (s)
64 x 64	0.0131	0.0118	0.0081	0.0079	0.0110	0.0092
128 x 128	0.1285	0.0121	0.0103	0.0105	0.0136	0.0093
256 x 256	-	0.0220	0.0206	0.0176	0.0144	0.0137
512 x 512	-	0.1491	0.0940	0.0798	0.0653	0.06558
1024 x 1024	-	0.5835	0.6625	0.5962	0.5373	0.53529

TableS5. Reconstruction times for the SD2Iu (factor 8).

Sinogram size (pixels)	Number of parameters	Time per epoch (s)	Number of epochs	Total reconstruction time (s)
64 x 64	331,457	0.0110	4000	44
128 x 128	730,817	0.0136	4000	54.4
256 x 256	2,328,257	0.0144	4000	57.6
512 x 512	8,718,017	0.0653	4000	261.2
1024 x 1024	34,277,057	0.5373	4000	2149.2

Table S6. Impact of sinogram size (number of projections) for the SD2Iu (factor 8).

Sinogram size (pixels)	Number of parameters	Time per epoch (s)	Total reconstruction time (s)
512 x 64	4,455,873	0.03270	130.8
512 x 128	4,455,873	0.03518	140.72
512 x 256	4,455,873	0.04335	173.4

512 x 512	4,455,873	0.06534	261.36
512 x 1024	4,455,873	0.12548	501.92

Shepp-Logan phantom: SD2I performance

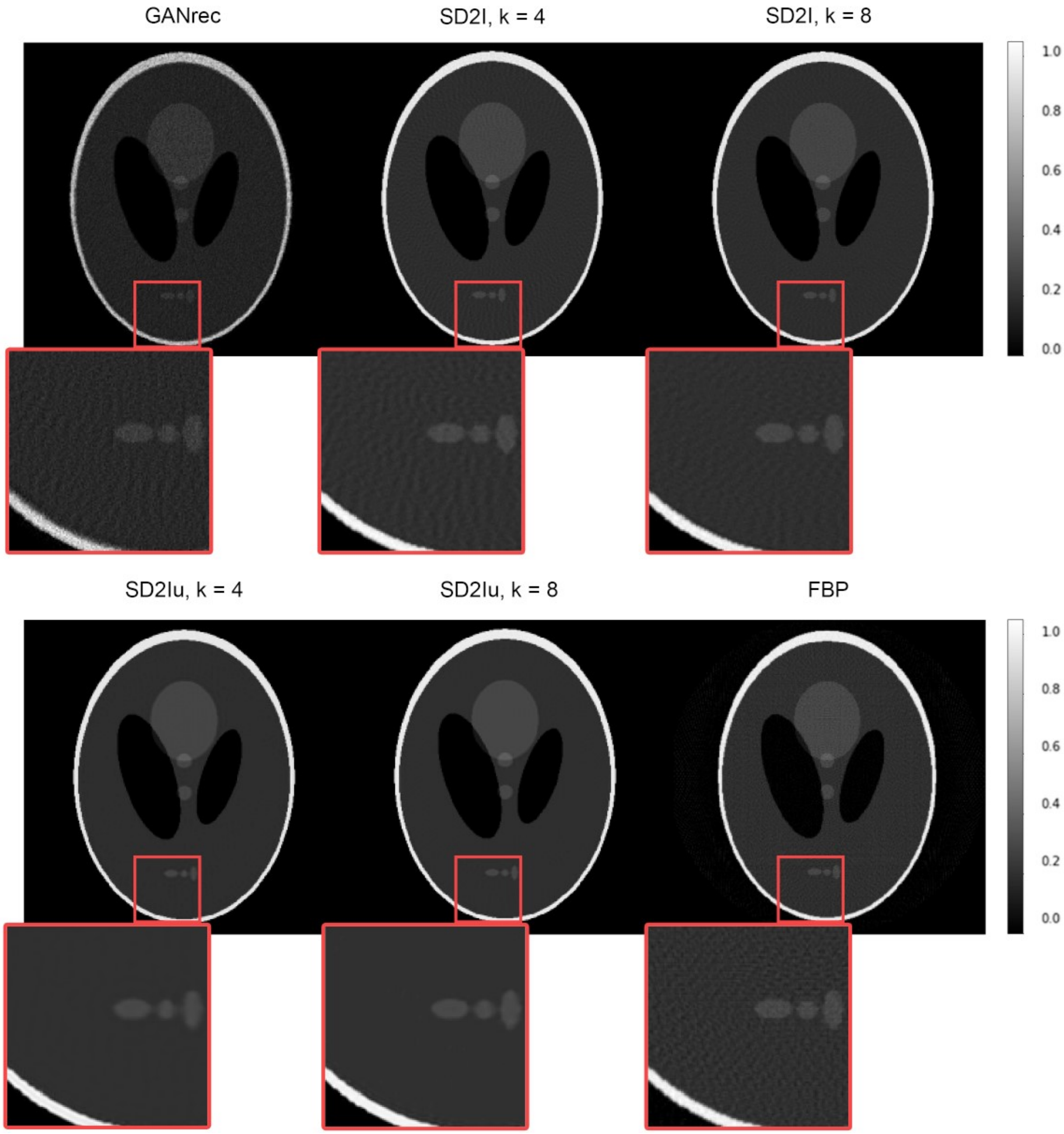


Figure S7. Comparison between the SD2I result and conventional reconstruction methods. The image size is 512x512 and reconstructed from the 512x128 Shepp-Logan sinogram.

Table S7. Accuracy. Comparison of approaches for a 512x128 Shepp-Logan sinogram

	FBP	GANrec	SD2I (k = 8)	SD2I (k = 4)	SD2Iu (k = 8)	SD2Iu (k = 4)
MAE	0.01230	0.02712	0.002925	0.002127	0.001315	0.001387
MSE	0.0005125	0.003924	$4.890\text{x}10^{5}$	$3.916\text{x}10^{5}$	$3.645\text{x}10^{5}$	$3.912\text{x}10^{5}$
SSIM	0.6855	0.6772	0.9850	0.9895	0.9980	0.9977
PSNR	32.90	24.06	43.10	43.95	44.38	44.08

SD2I loss functions

Table S8. Accuracy. Comparison of approaches for a 512x128 Shepp-Logan sinogram with different loss functions.

	MSE	MAE	MSE + SSIM	MAE +SSIM
MAE	0.00346	0.00291	0.00337	0.00246
MSE	0.000241	0.000202	0.000212	0.000116
SSIM	0.9899	0.9937	0.9912	0.9952
PSNR	36.1844	36.9427	36.7385	39.3724

SD2I: impact of different ground truth choices on accuracy metrics

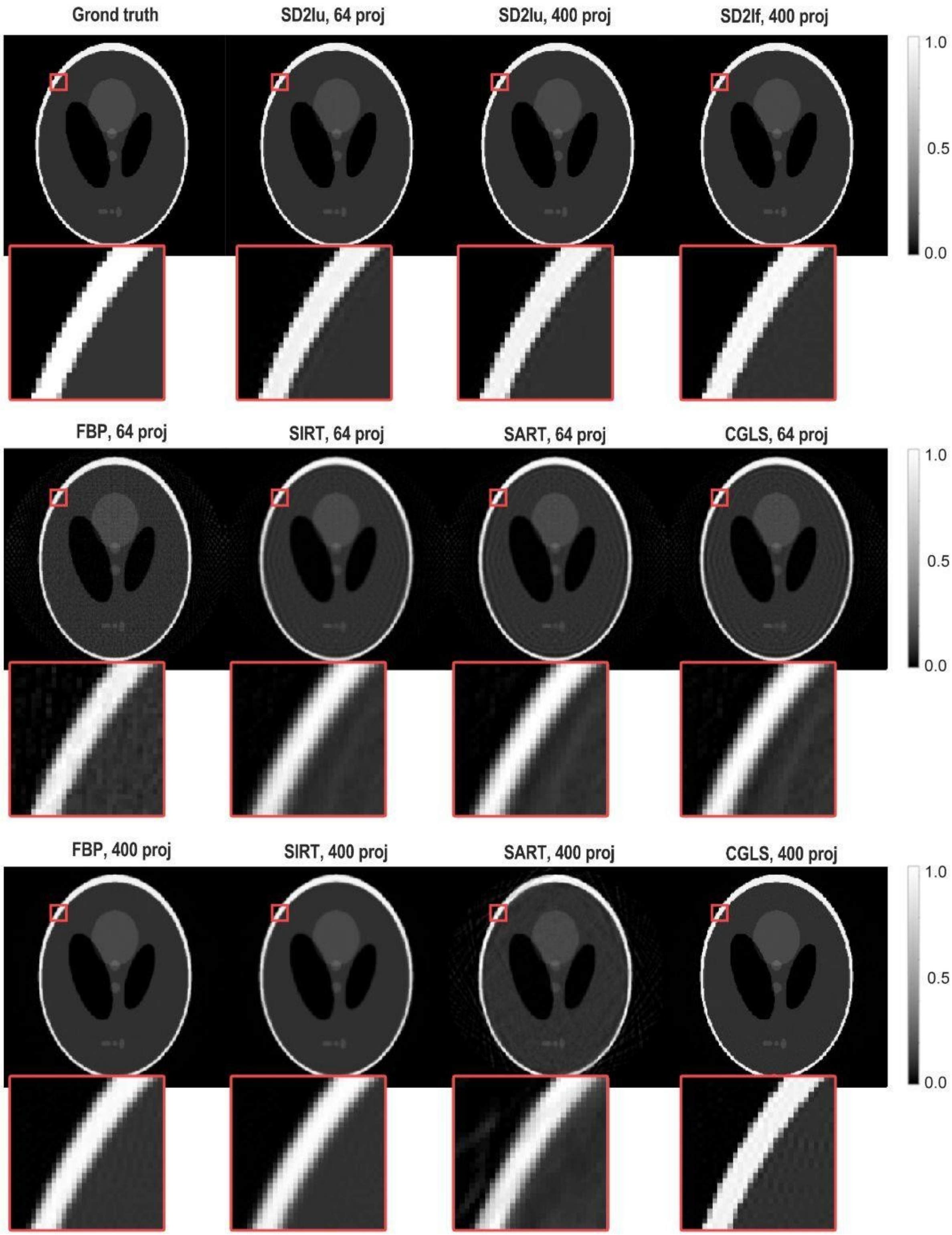


Figure S8. Comparison between the SD2I results and conventional reconstruction methods on 256x256 Shepp-Logan images with either 64 or 400 projections. All SD2I-based methods were using k = 8.

Table S9. Accuracy. Comparison of the different reconstruction methods' performance with the reference of the clean Shepp-Logan image.

		MAE	MSE	SSIM	PSNR
Reconstructed from 64 proj	SD2Iu	**0.002881**	**0.00009763**	**0.9931**	**40.10**
	FBP	0.01906	0.001405	0.6129	28.52
	SART	0.01702	0.001851	0.7572	27.33
	CGLS	0.01722	0.001717	0.7329	27.65
	SIRT	0.01768	0.002327	0.7984	26.33
Reconstructed from 400 proj	SD2Iu	0.002600	5.827x105	0.9950	42.35
	SD2I	**0.0005762**	**2.8541x106**	**0.99965**	**55.44**
	FBP	0.007819	0.0007783	0.9565	31.09
	SART	0.01505	0.0015	0.7907	28.24
	CGLS	0.005191	0.0001365	0.9401	38.65
	SIRT	0.01036	0.001504	0.9665	28.23

Table S10. Accuracy. Comparison of the different reconstruction methods' performance with the reference of the FBP image reconstructed from 400 projections.

		MAE	MSE	SSIM	PSNR
Reconstructed from 64 proj	SD2Iu	**0.007863**	0.0006805	**0.9639**	31.98
	FBP	0.01448	0.0006236	0.6776	32.36
	SIRT	0.01214	0.0006918	0.8692	31.90
	SART	0.01154	0.0004871	0.8306	33.43
	CGLS	0.01189	**0.0004715**	0.8064	**33.57**
Reconstructed from 400 proj	SD2Iu	0.007807	0.0007706	0.9595	31.44
	SD2I	0.007741	0.0007555	0.9613	31.52
	SIRT	**0.005085**	**0.0002037**	**0.9828**	**37.22**
	SART	0.01341	0.0007248	0.8055	31.70
	CGLS	0.007319	0.0004429	0.9733	33.84
	Clean image	0.007819	0.0007783	0.9585	31.39

Table S11. Accuracy. Comparison of the different reconstruction methods' performance with the reference of the CGLS image reconstructed from 400 projections.

		MAE	MSE	SSIM	PSNR
Reconstructed from 64 proj	SD2Iu	**0.005695**	**0.0001475**	**0.9483**	**38.88**
	FBP	0.01794	0.001040	0.6807	30.40

	SIRT	0.01720	0.001855	0.8549	27.89
	SART	0.01639	0.001409	0.8212	29.08
	CGLS	0.01652	0.001293	0.8009	29.45
Reconstructed from 400 proj	SD2Iu	0.005196	0.0001344	0.9470	39.29
	SD2I	**0.005105**	**0.0001285**	**0.9492**	**39.48**
	FBP	0.007319	0.0004429	0.9744	34.11
	SIRT	0.01102	0.001093	0.9488	30.18
	SART	0.01774	0.001572	0.7947	28.61
	Clean image	0.005191	0.0001365	0.9462	39.22

SD2I: More XRD-CT Image

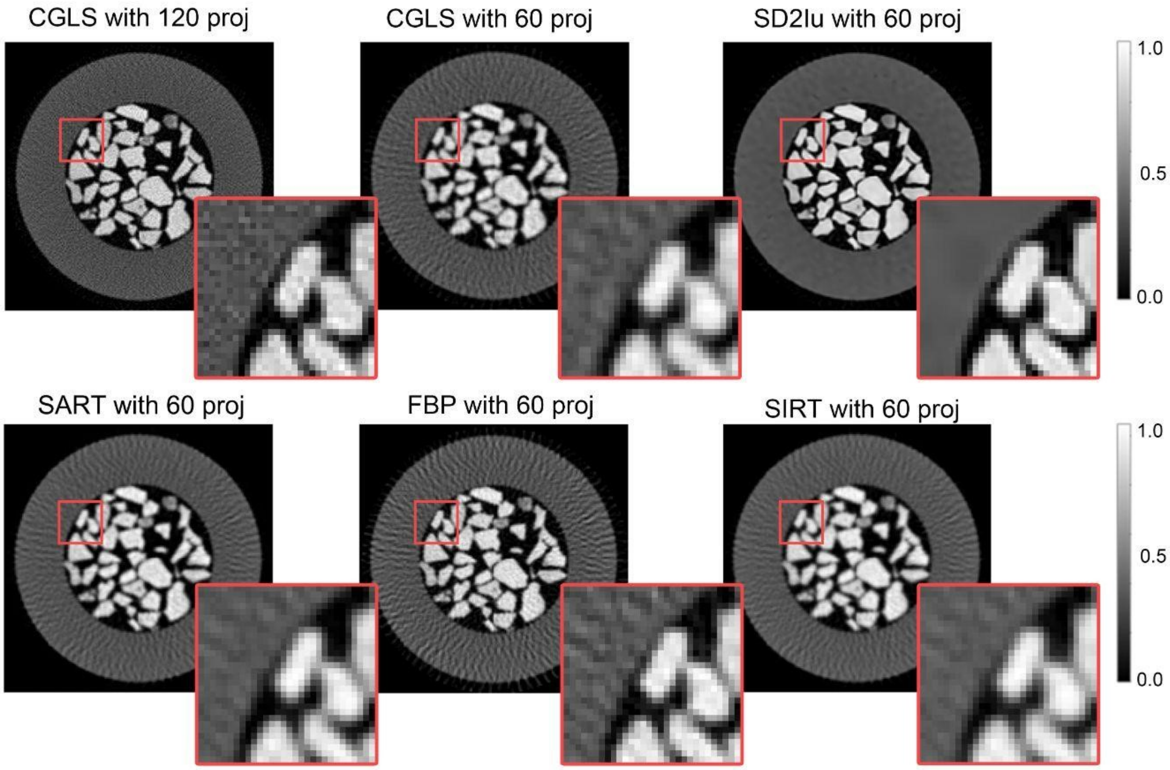


Figure S9. OCM catalyst, Image size: 179x179.

Table S12. OCM catalyst.

	FBP	SART	CGLS	SIRT	SD2lu
MAE	0.1479	0.1279	0.1403	0.1291	**0.07403**
MSE	0.04819	0.03994	0.04601	0.04249	**0.01392**
SSIM	0.6970	0.7854	0.7435	0.7946	**0.8768**
PSNR	27.23	28.05	27.43	27.78	**32.63**

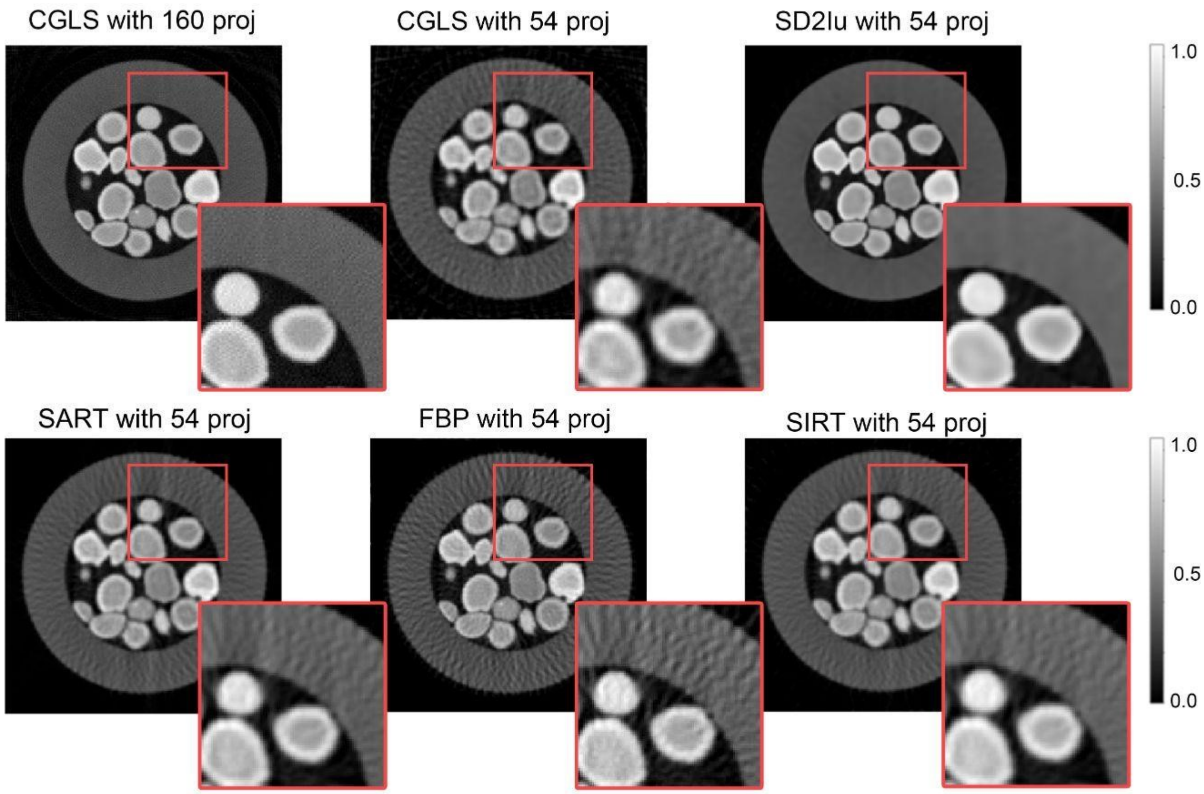


Figure S10. POX, Image size: 223x223.

Table S13. POX, photocatalyst.

	FBP	SART	CGLS	SIRT	SD2lu
MAE	0.09090	0.07544	0.8665	0.07444	**0.0241**
MSE	0.2348	0.01924	0.02338	0.01927	**0.00148**
SSIM	0.6859	0.7993	0.7299	0.8108	**0.9617**
PSNR	27.01	27.88	27.22	27.87	**39.02**

Micro-CT images: Positions in radiograph

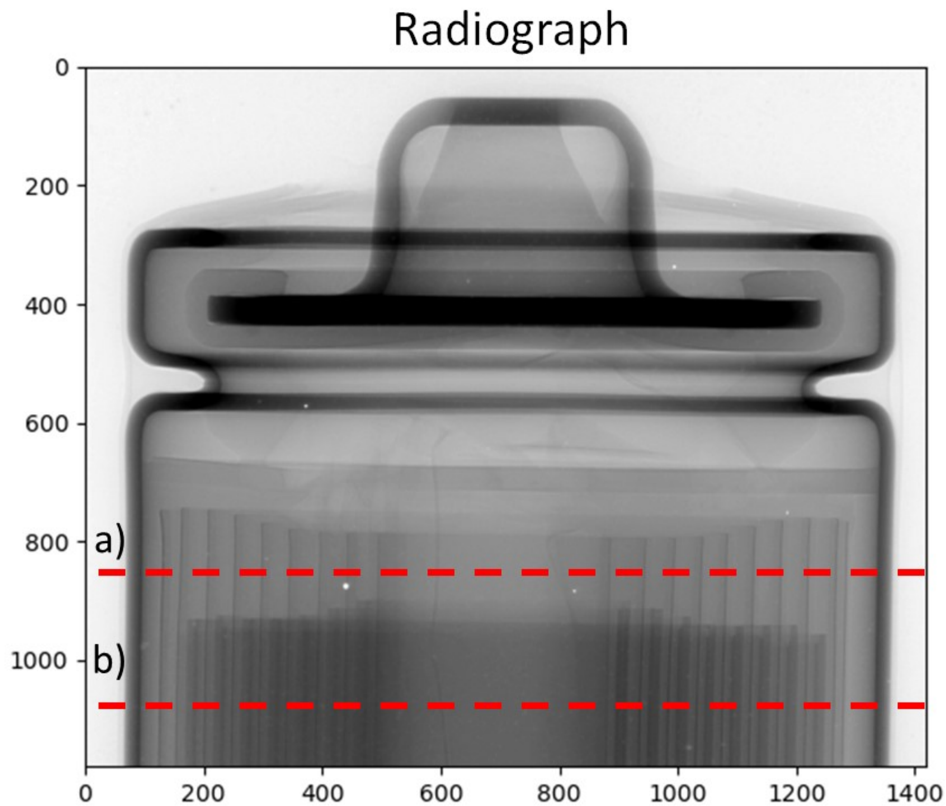


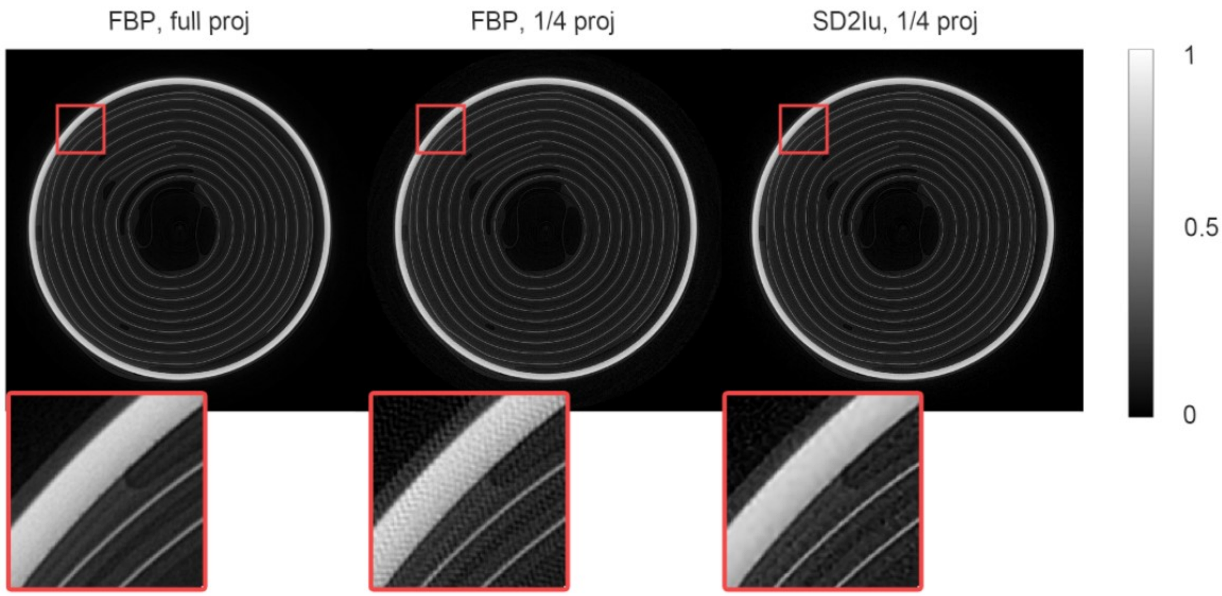
Figure S11. Normalised (dark current and flat field corrected) radiograph of the NMC532 Li-ion battery. Positions a) and b) correspond to the two sinograms (battery cross-sections) used in this work.

SD2I: Summary of angular undersampling ratio

Table S14: Comparison of ideal and angular undersampled data used in this work.

Dataset	No. translation steps	Nyquist No. of projections	Angular undersampling projections
Photocatalyst	331	520	60 (**11.5%**)
NMC532 (xrd-ct)	547	859	100 (**11.6%**)
NMC532 (micro-ct)	779	1224	261 (**21.3%**)
OCM catalyst	179	282	60 (**21.3%**)
POX catalyst	223	351	54 (**15.4%**)

Large micro-CT images: SD2I performance



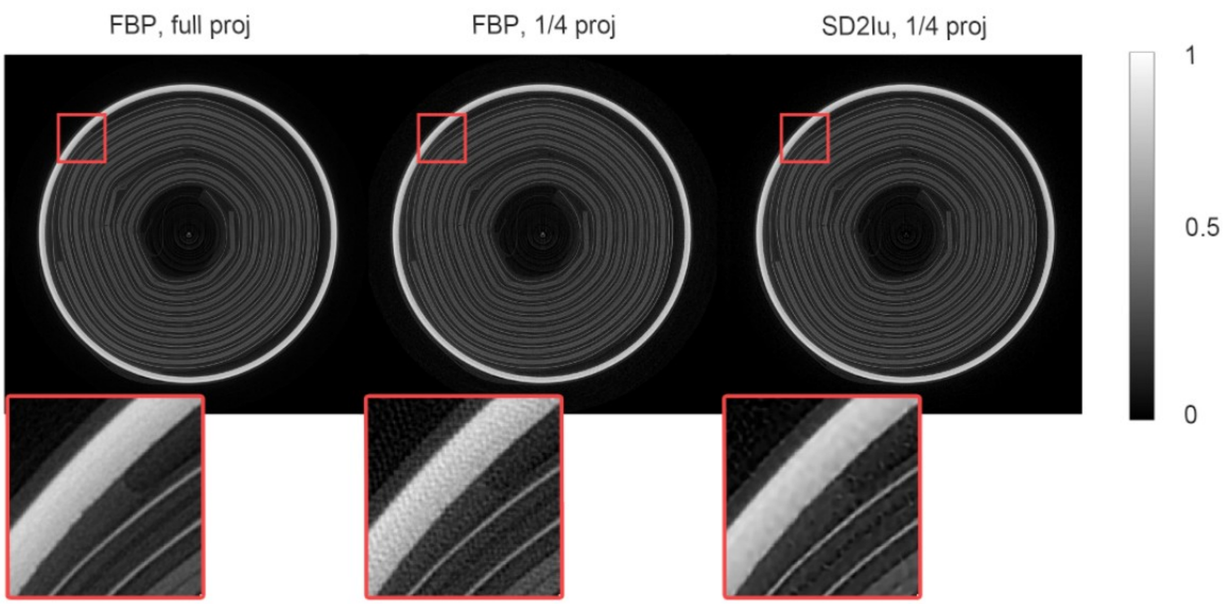


Figure S12. Two example micro-CT reconstruction images. SD2I results are using k factors equal to 8. The image sizes are 1559 x 1559 . The SD2I and FBP results are reconstructed from the sinogram size as 1559 x 391. The ground truth is obtained by the FBP reconstruction of the 1559 x 1561 sinogram. Each sinogram was acquired after taking the mean of five neighbouring sinograms

Table S15. Accuracy. Comparison of approaches for reconstructing two example full-size micro-CT images shown in figure S6 using FBP of the full projection set as the ground truth image.

	(a)		(b)	
	FBP 1/4 proj	SD2Iu	FBP 1/4 proj	SD2Iu
MAE	0.01307	**0.00968**	0.02922	**0.02137**
MSE	0.000405	**0.000228**	0.001893	**0.001072**
SSIM	0.6977	**0.8213**	0.7063	**0.8210**
PSNR	30.6278	**33.1142**	29.9556	**32.42354**

SD2I: impact of discriminator

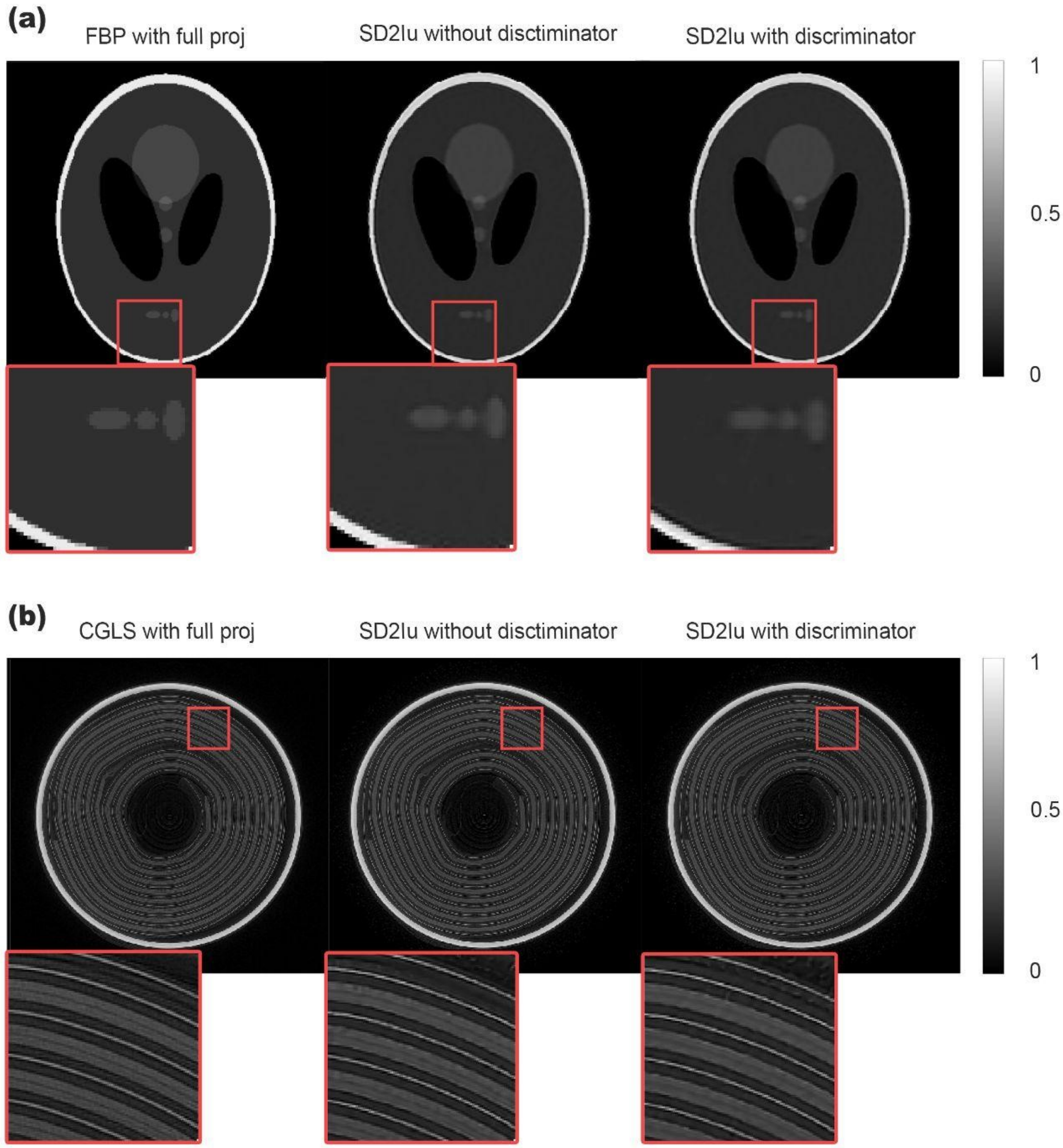


Figure S13. The images show the impact of the discriminator in the training loop. (a) The image is reconstructed from a Shepp-Logan simulated sinogram with 256x64 projections as input.(b) The micro-CT experimental image with 779x261 sinogram as input.

Table S16. Accuracy. Comparison of the training loop with and without discriminator for both experimental (using CGLS of the full projection set as the ground truth image) and simulated data. The generator architecture is the SD2Iu.

	(a) Shepp-Logan		(b) Micro-CT	
	Without discriminator	With discriminator	Without discriminator	With discriminator
MAE	**0.00315**	0.00647	**0.03191**	0.03220
MSE	**0.000125**	0.000542	**0.002306**	0.002372
SSIM	**0.9941**	0.9853	**0.7897**	0.7887
PSNR	**39.74**	33.39	**30.95**	30.82

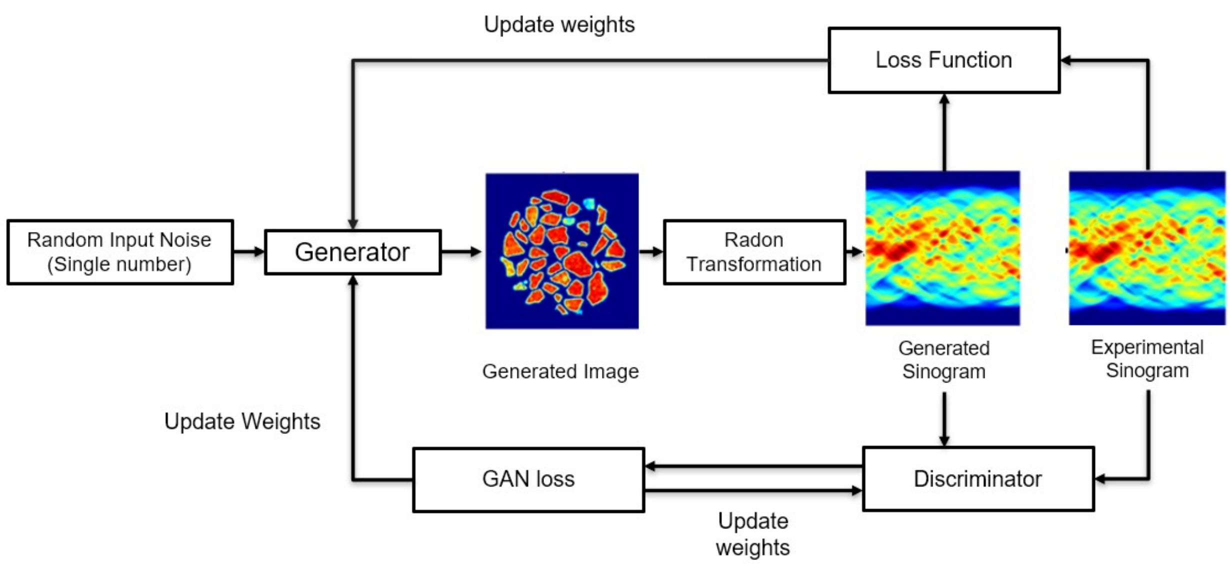


Figure S14. The flowchart of the SD2I training algorithm with a discriminator. The input of the SD2I is a random constant which should ideally have a similar order of magnitude as the reconstructed image's signal. The generator creates an image based on the single input. The generated image is converted into a sinogram by the forward operator, and then both the generated sinogram and the original experimental sinogram are sent into the discriminator for calculating the GAN loss. The weights of the generator are then updated by minimising the joint loss function with the GAN loss, MSE and SSIM while the discriminator is updated by the GAN loss only.

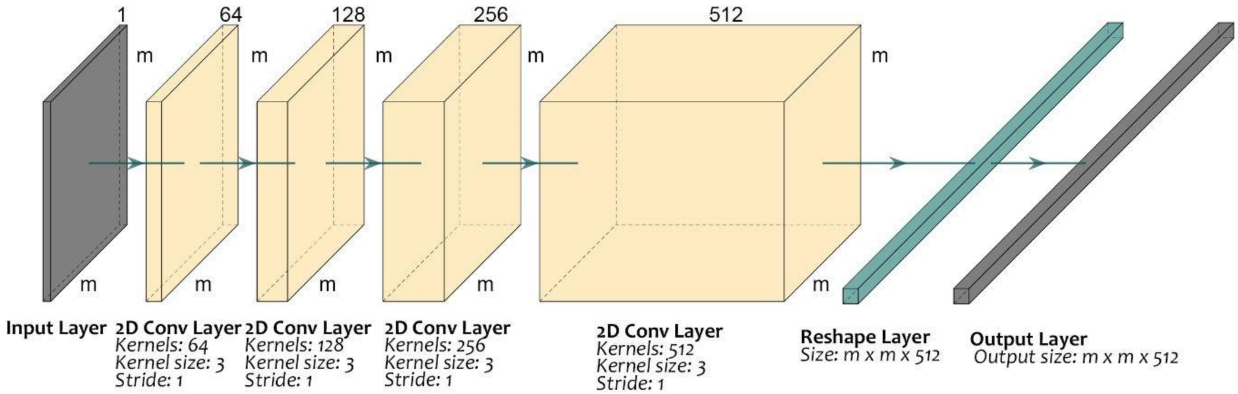


Figure S15. A representation of the Discriminator network used in figure S14. The kernel types and parameter settings are shown in the figure. There are no fully connected (dense) layers in the discriminator so the number of parameters is very low compared to the generator networks used in this work.