Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Sparse-View CT Reconstruction Using Deep Learning

Suthar Het Nileshbhai1
Indian Institute of Technology Roorkee
Mayank Goswami
mayankfph@iitr.ac.in

Indian Institute of Technology Roorkee https://orcid.org/0000-0001-8829-9980

Research Article

Keywords: sparse-view CT, streak artefacts, filtered back-projection, SART, U-Net, FBPConvNet, deep learning, low-dose CT, industrial CT

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11733389/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Sparse-View CT Reconstruction Using Deep Learning

Suthar Het Nileshbhai[1], and Mayank Goswami [1, 2, #]
[1] Department of Physics, IIT Roorkee
[2]*Mehta Family School of Data Science and Artificial Intelligence*, IIT Roorkee,
Roorkee, Uttarakhand-247667, India.
[#]mayank.goswami@ph.iitr.ac.in

Abstract

Sparse-view computed tomography (CT) reduces radiation dose and scan time by acquiring few projection angles but filtered back-projection (FBP) of such data suffers from severe streak artefacts because the angular sampling falls far below the Nyquist requirement. We study the post-processing approach in which a residual U-Net removes these artefacts from the FBP image, and test it on five synthetic datasets that differ in object class and physics: head phantoms at normal dose, chest phantoms at ten-fold lower dose, aluminium machined parts, photon-starved titanium components, and aluminium resolution targets. Sinograms at 30, 60 and 90 views are simulated with the Beer-Lambert law, Poisson photon statistics and Gaussian read noise, and the network is compared against FBP and the Simultaneous Algebraic Reconstruction Technique (SART), both with settings tuned on validation data. On 150 held-out test images per dataset, the U-Net achieves the highest peak signal-to-noise ratio (PSNR) in all 15 dataset and view settings, improving on the best classical method by 1.9 to 5.8 dB (paired Wilcoxon test, $p < 10^{-20}$ in every setting), and at 30 views it exceeds the PSNR of 180-view FBP on all five datasets, while running in about 20 ms per image on a CPU. Stress tests show the limits of the approach: networks transfer between similar industrial datasets but degrade or harm images of other object classes, fail when the test data are more sparsely sampled or receive ten times fewer photons than in training, and are outperformed by SART on the unseen Shepp-Logan phantom. A network trained on resolution targets does generalise to a resolution plate taken from laboratory data, improving on SART by 2.7 to 3.4 dB. The code, data and trained models are released for full reproducibility.

Keywords: sparse-view CT, streak artefacts, filtered back-projection, SART, U-Net, FBPConvNet, deep learning, low-dose CT, industrial CT

1. Introduction

X-ray computed tomography reconstructs a map of the linear attenuation coefficient inside an object from many X-ray projections taken at different angles. It is indispensable in medicine and in industrial non-destructive testing, but each projection costs radiation dose and time. In medicine, CT is one of the largest contributors to population radiation exposure, which matters most for children, for patients who need repeated scans and for screening programmes such as low-dose lung cancer screening. In industry, scan time limits throughput on inspection lines, and some systems physically cannot reach a full set of angles.

Acquiring fewer projections, known as *sparse-view* CT, addresses both concerns directly: with V views instead of 180, the number of exposures falls by a factor 180/V. The price is paid in reconstruction. The standard analytic method, filtered back-projection (FBP), assumes dense angular sampling; with too few views it produces characteristic streak artefacts that hide small features and create false ones. Iterative algebraic methods such as SART reduce the streaks by enforcing consistency with the measured data and non-negativity, but they are slow and tend to smooth edges.

Since 2016, deep convolutional neural networks have been used to remove sparse-view artefacts. The simplest and most widely used approach, introduced as FBPConvNet by Jin et al. [1], applies a residual

U-Net [2] to the FBP image. Reported gains are large, but evaluations often use a single object class, compare against untuned baselines, or do not probe when the network fails; Sidky et al. [5] showed a case where a network did not actually solve the inverse problem.

This work re-implements the FBPConvNet approach and evaluates it carefully on five physically distinct synthetic datasets with exact ground truth, against classical baselines whose parameters are tuned on separate validation data, and with explicit stress tests: transfer across datasets, mismatched view counts and doses, and objects never seen in training.

2. Problem statement and objectives

Given a noisy sinogram measured at V equally spaced angles over 180°, with $V \in \{30, 60, 90\}$, the task is to reconstruct the attenuation image as accurately as possible. Because V is far below the number of angles needed for alias-free sampling (Section 3.4), the problem is under-determined: many images are consistent with the data, and additional prior knowledge is needed to choose among them. A trained network supplies such a prior implicitly, learned from examples.

The objectives were:

- **Realistic simulation with exact ground truth.** Generate five datasets of random phantoms that differ in object class, field of view, material and dose, and simulate their sparse-view sinograms with the Beer-Lambert law, Poisson photon statistics and detector read noise.
- **Strong classical baselines.** Compare against FBP and SART with their free parameters (filter; number of iterations) chosen on validation data.
- **A trained artefact-removal network.** Train one residual U-Net per dataset and view count (15 networks) on FBP inputs.
- **Honest evaluation.** Report PSNR, SSIM and RMSE on held-out test images with paired significance tests, and probe failure modes with cross-dataset, cross-view, dose and unseen-object tests.

3. Theory

3.1 Beer-Lambert law

A monochromatic X-ray beam traversing an object along a line L is attenuated according to the Beer-Lambert law,

$$I = I_0 \exp(-\int_L \mu(x, y)\, dl), \tag{1}$$

where I_0 is the number of incident photons, I the number reaching the detector and $\mu(x, y)$ the linear attenuation coefficient, which is the quantity a CT image displays. Taking the logarithm gives the line integral

$$p = -\ln(I / I_0) = \int_L \mu(x, y)\, dl, \tag{2}$$

which is linear in μ and is the input to every reconstruction method. In this work μ values come from NIST mass-attenuation tables: cortical bone at 70 keV (0.0385 mm^{-1}), aluminium at 100 keV (0.046 mm^{-1}) and titanium at 100 keV (0.122 mm^{-1}).

3.2 Radon transform and sinogram

Parametrising each line by its angle θ and signed distance s from the origin, the set of all line integrals p(s, θ) is the Radon transform of μ. Displayed as an image over (s, θ) it is called the sinogram, because a point in the object traces a sinusoid. A sparse-view scan measures only V columns of the sinogram (Figure 3 shows examples).

3.3 Fourier slice theorem and filtered back-projection

The Fourier slice theorem states that the one-dimensional Fourier transform of the projection p(·, θ) equals the two-dimensional Fourier transform of μ along the radial line through the origin at angle θ. Each view therefore samples one spoke of Fourier space. Inverting the two-dimensional Fourier transform in polar coordinates gives filtered back-projection,

$$\mu(x, y) = \int_0^{\pi} (p(\cdot, \theta) * h)(x \cos\theta + y \sin\theta) \, d\theta, \quad \hat{h}(\omega) = |\omega|, \tag{3}$$

in which each projection is convolved with the ramp filter h (whose Fourier transform is |ω|, compensating the higher density of spokes near the origin) and smeared back across the image. Apodised variants (Shepp-Logan, cosine, Hann) taper |ω| at high frequencies to suppress noise. FBP is fast and exact for complete, noise-free data.

3.4 Angular sampling and aliasing

For an image of N × N pixels, the spokes at frequency radius k are separated by about kπ/V. Requiring this gap not to exceed the frequency sampling interval at the Nyquist radius gives the angular sampling condition (Crowther criterion)

$$V \geq (\pi / 2)\, N \approx 201 \text{ views for } N = 128. \tag{4}$$

With V = 30, 60 and 90 only the inner 15 %, 30 % and 45 % of the Nyquist radius are adequately sampled (Figure 1). FBP implicitly sets the unmeasured Fourier coefficients between spokes to zero; the resulting aliasing appears in the image as streaks along the measured directions, strongest around sharp, high-contrast edges.

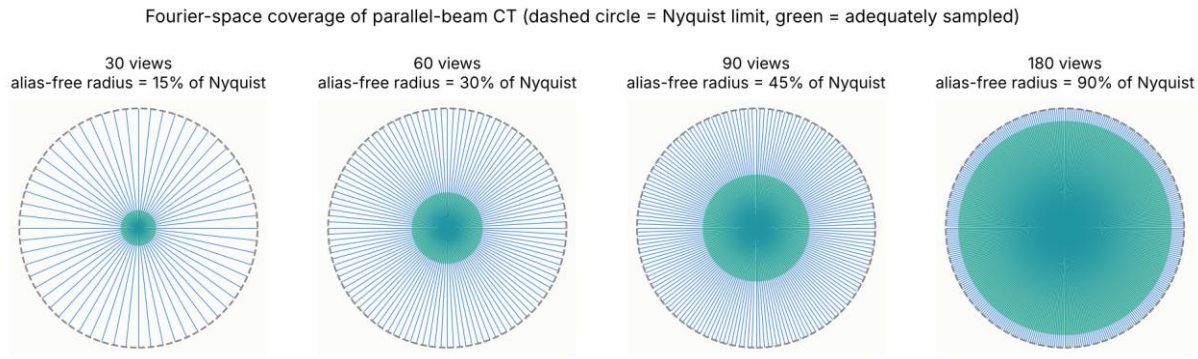


Figure 1. Fourier-space coverage of parallel-beam CT. Each view fills one radial line (Fourier slice theorem). The shaded disc marks the region sampled densely enough to avoid aliasing; the dashed circle is the Nyquist limit for 128 × 128 images.

3.5 Photon statistics

Photon detection is a counting process, so the measured count follows a Poisson distribution with mean $I_0 e^{-p}$, for which the variance equals the mean and the relative noise scales as $1/\sqrt{I}$. Detector electronics add approximately Gaussian read noise. The measurement model used here is therefore

$$I_m = \text{Poisson}(\, I_0\, e\text{^}(-p)\,) + \mathcal{N}(\,0, \sigma_e{}^2\,), \quad \hat{p} = -\ln(\max(I_m, 1) / I_0), \tag{5}$$

with $\sigma_e = 10$ counts. Rays through strongly attenuating material receive few photons and give very noisy line integrals (photon starvation); the incident count I_0 sets the dose.

3.6 Algebraic reconstruction (SART)

Discretising the image into a vector x and the projector into a sparse system matrix A gives the linear system Ax = p. SART [3] updates the estimate one view at a time,

$$x \leftarrow x + \lambda\, A^{T}_\theta\, [\,(\,p_\theta - A_\theta\, x\,) / A_\theta\, 1\,] / (\,A^{T}_\theta\, 1), \tag{6}$$

with relaxation $\lambda = 0.15$, followed here by clipping to $x \geq 0$. Non-negativity acts as prior knowledge, so SART suppresses streaks better than FBP, while the number of passes over all views controls a trade-off between streak removal and noise amplification.

3.7 Convolutional networks, U-Net and residual learning

A convolutional layer applies learned 3 × 3 filters across the image followed by a non-linearity (ReLU). The U-Net [2] arranges such layers in an encoder that repeatedly halves the resolution and doubles the number of channels, a decoder that restores the resolution with transposed convolutions, and skip connections that pass encoder features to the decoder to preserve detail (Figure 2). The large receptive field at the bottleneck is important because streaks extend across the whole image. Following [1], the network is residual: it predicts a correction that is added to the FBP input, so it only needs to learn the artefact.

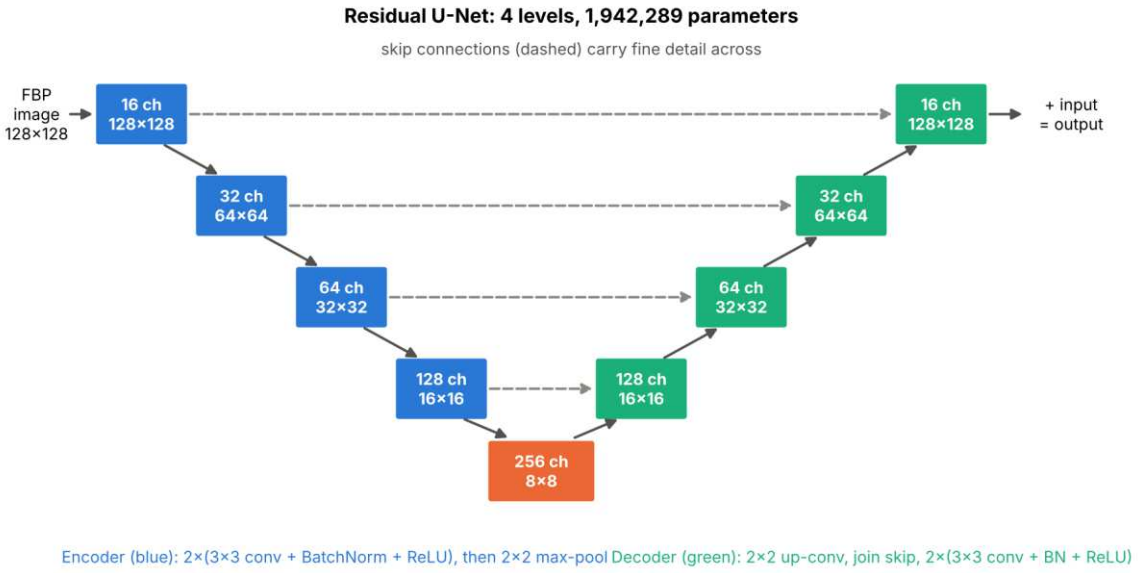


Figure 2. The residual U-Net used in this work: four resolution levels (16 to 128 channels, 256 at the 8 × 8 bottleneck), 1,942,289 trainable parameters. The output is added to the FBP input image.

The network is trained by supervised learning: the parameters are adjusted by the Adam optimiser [12] to minimise the mean squared error (MSE) between the output and the ground-truth image, which is equivalent to maximising PSNR and to a Gaussian likelihood for the residual error.

4. Previous work

Jin, McCann, Froustey and Unser [1] proposed FBPConvNet, a residual U-Net applied after FBP, and showed large gains for sparse-view CT on synthetic ellipse phantoms and on real data; it is the method studied here. Han and Ye [4] analysed the U-Net through deep convolutional framelets, explaining why a plain U-Net can blur high-frequency detail, and proposed dual-frame and tight-frame variants. Adler and Öktem [6] unrolled a primal-dual optimisation algorithm and replaced its proximal steps with small networks (learned primal-dual), so that the forward operator is used inside the network. Song, Shen, Xing and Ermon [7] used score-based generative models as priors for inverse problems in medical imaging, which require no paired training data. Sidky, Lorente, Brankov and Pan [5] tested a published CNN for sparse-view breast CT and found substantial errors even on training images, whereas constrained total-variation minimisation recovered the images, concluding that the network had not solved the inverse problem in that setting. More recently, Dorosti et al. [8] compared U-Net inputs for clinical sparse-view CT and found that 2-D networks performed best across parallel-, fan- and cone-beam geometries.

Relative to these studies, the contribution of this work is a controlled, fully reproducible evaluation of the post-processing approach across five datasets with different physics, with tuned baselines and explicit failure tests in the spirit of [5].

5. Method

5.1 Datasets

Five datasets were generated (Figure 3, Table 1). Each contains 1000 training, 100 validation and 150 test phantoms of 128 × 128 pixels, 6,250 phantoms in total. Shapes are drawn on a grid twice as fine and block-averaged, which reproduces partial-volume edges, and everything outside the inscribed circle is set to zero. Phantom values are normalised to [0, 1] and multiplied by the dataset's μ for value 1.0 to obtain physical attenuation.

- **DS1 Head:** elliptical skull (value 0.85-1.0) around brain tissue (≈ 0.5, close to water), one or two CSF-filled ventricles, 2-7 low-contrast lesions (±0.02-0.12) and small calcifications.
- **DS2 Chest (low dose):** body with a fat layer, lungs (≈ 0.12) containing vessels and 0-3 nodules, heart, aorta, vertebra with cortical and cancellous bone, ribs and sternum. Ten times fewer incident photons than DS1.
- **DS3 Aluminium parts:** rotated blocks or disks with drilled holes, dense inclusions, thin cracks and groups of slots 1-4 pixels wide.
- **DS4 Titanium parts:** one to three gears, hex nuts, washers or bolt sections with keyways, hexagonal sockets and pores of 1-3 pixels.
- **DS5 Resolution targets:** aluminium plates with a grid of USAF-style three-bar groups 1-6 pixels wide (horizontal, vertical or both) and rows of contrast-detail holes of decreasing size and depth.

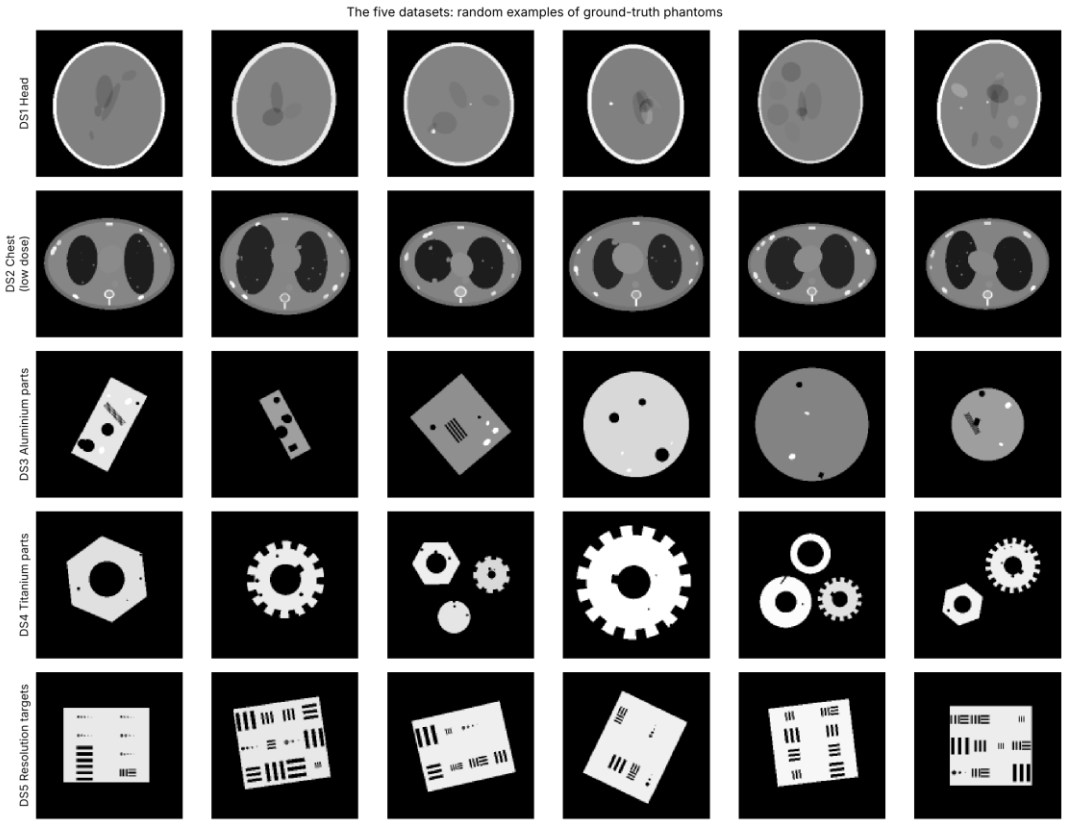


Figure 3. Random ground-truth examples from the five datasets (rows: DS1 Head, DS2 Chest, DS3 Aluminium parts, DS4 Titanium parts, DS5 Resolution targets).

Table 1. Physical parameters of the five datasets. All images are 128 × 128 pixels; read noise $\sigma_e = 10$ counts for all.

Dataset	Field of view (pixel)	Value 1.0 corresponds to	I_0 (photons)
DS1 Head	205 mm (1.6 mm)	cortical bone, 70 keV (0.0385 mm^{-1})	1×10^5
DS2 Chest (low dose)	358 mm (2.8 mm)	cortical bone, 70 keV (0.0385 mm^{-1})	1×10^4
DS3 Aluminium parts	64 mm (0.5 mm)	aluminium, 100 keV (0.046 mm^{-1})	2×10^4
DS4 Titanium parts	38 mm (0.3 mm)	titanium, 100 keV (0.122 mm^{-1})	2×10^4
DS5 Resolution targets	32 mm (0.25 mm)	aluminium, 100 keV (0.046 mm^{-1})	2×10^4

5.2 Scan simulation

For every phantom, parallel-beam sinograms with 128 detector bins were computed with the Radon transform of scikit-image [10] at V = 30, 60 and 90 views equally spaced over [0°, 180°), plus V = 180 for validation and test images as a dense reference. Line integrals were converted to expected counts, corrupted according to Eq. (5) and log-transformed back. Each phantom has its own random seed derived from a fixed master seed, so the data are identical on every run. Sinograms are stored in half precision, whose rounding error ($\approx 10^{-3}$ relative) is far below the noise.

5.3 Classical baselines

FBP was computed with the ramp, Shepp-Logan, cosine and Hann filters, and SART with 1, 2, 3, 5, 8, 12, 20, 30, 50 and 80 passes. For every dataset and view count, the filter and the number of SART passes with the highest mean PSNR on 20 validation images were selected. The chosen SART passes ranged from 2 (DS2 at 60 and 90 views, where noise dominates) to 80 (DS1 and DS4 at 30 views); in the latter two cases the validation curve had flattened (+0.17 dB and +0.06 dB from 50 to 80 passes), so the baselines are close to their optimum.

5.4 Network training

For each dataset and view count, a U-Net was trained on pairs of (tuned-filter FBP image, phantom) from the training set for 30 epochs with batches of 16, Adam with learning rate 10^{-3} decaying to 10^{-5} on a cosine schedule, MSE loss and random horizontal and vertical flips. Flips are valid augmentations because, for equally spaced views over 180°, they map the set of angles onto itself. After each epoch the mean validation PSNR was computed and the best epoch kept; the test set was not used for any decision. Training used bfloat16 mixed precision with channels-last memory layout, taking about 5.5 minutes per network on a two-core CPU; the complete study (data, tuning, 15 networks, evaluation) runs in about 2.2 hours with a single command.

5.5 Evaluation

All reconstructions were clipped to [0, 1] and compared with the ground truth using RMSE, PSNR = $10 \log_{10}(1/\mathrm{MSE})$ and SSIM [9] with data range 1. Methods were compared on the same 150 test images per dataset with the paired Wilcoxon signed-rank test. Additional experiments used the test sets as follows: (i) cross-dataset transfer at 60 views, (ii) cross-view transfer within each dataset, (iii) dose robustness at 60 views with I_0 scaled by 0.1-50 on 50 test images, (iv) two objects never seen in training, the Shepp-Logan phantom [11] and an aluminium resolution plate extracted from a laboratory FDK reconstruction (360 cone-beam projections, 100 kV, source-object distance 300 mm, source-detector distance 1100 mm), each simulated with every dataset's physics, and (v) run time per image.

6. Results

6.1 Test-set image quality

Table 2 and Figure 4 summarise the test-set results. The U-Net achieved the highest mean PSNR in all 15 dataset and view settings, by 1.9 dB (DS4, 30 views) to 5.8 dB (DS3, 90 views) over the better of FBP and SART. The paired Wilcoxon test gave $p < 10^{-20}$ against SART in every setting; the U-Net improved PSNR on all 150 test images in 12 of the 15 settings and on at least 129 of 150 in all. SSIM improved similarly, from 0.40-0.82 for FBP and 0.76-0.95 for SART to 0.90-0.98 for the U-Net, with the exception of DS4 at 30 views, where U-Net and SART have equal SSIM (0.921 vs 0.920) despite a 1.9 dB PSNR advantage. At 30 views the U-Net exceeded the PSNR of 180-view FBP on every dataset.

Table 2. Mean test-set PSNR (dB) / SSIM over 150 images. Bold: best method. Last column: U-Net PSNR gain over the better baseline.

Dataset	Views	FBP	SART	U-Net	Gain
DS1 Head	30	27.28 / 0.595	33.02 / 0.861	36.76 / 0.966	+3.74
	60	29.56 / 0.734	32.94 / 0.854	37.26 / 0.970	+4.33
	90	30.14 / 0.823	33.62 / 0.870	37.81 / 0.963	+4.19
DS2 Chest (low dose)	30	27.47 / 0.601	28.61 / 0.829	31.77 / 0.899	+3.16
	60	29.16 / 0.707	29.78 / 0.864	34.02 / 0.942	+4.23
	90	29.91 / 0.769	30.57 / 0.869	34.95 / 0.955	+4.37
DS3 Aluminium parts	30	26.65 / 0.517	29.19 / 0.832	32.65 / 0.938	+3.46
	60	28.89 / 0.544	30.90 / 0.859	35.80 / 0.967	+4.90
	90	29.81 / 0.648	31.85 / 0.873	37.68 / 0.973	+5.83
DS4 Titanium parts	30	24.60 / 0.401	30.19 / 0.920	32.07 / 0.921	+1.87
	60	27.93 / 0.496	32.27 / 0.939	36.60 / 0.980	+4.33
	90	29.11 / 0.636	33.43 / 0.946	36.21 / 0.964	+2.78
DS5 Resolution targets	30	20.59 / 0.419	22.76 / 0.756	28.23 / 0.916	+5.47
	60	23.26 / 0.439	25.51 / 0.804	30.54 / 0.952	+5.02
	90	24.42 / 0.462	26.55 / 0.816	31.94 / 0.967	+5.39

FBP with 180 views reached 30.60, 30.78, 31.02, 29.80 and 25.66 dB on DS1-DS5 respectively.

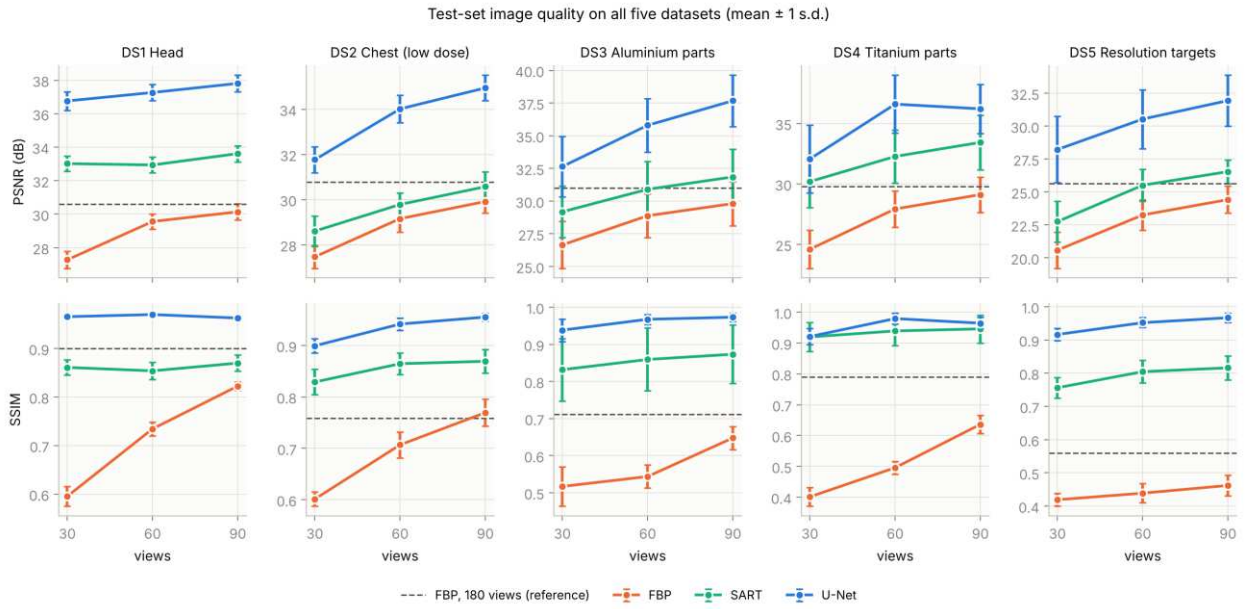


Figure 4. Test-set PSNR (top) and SSIM (bottom) versus number of views for each dataset; mean ± 1 standard deviation over 150 images. Orange: FBP; green: SART; blue: U-Net; dashed: FBP with 180 views.

Figure 5 shows one test image per dataset at 30 views. FBP images contain radial streaks and grainy interiors; SART removes most streaks but smooths edges, blurring gear teeth and merging fine bars; the U-Net output has clean backgrounds and sharp edges, but the finest bar groups and the smallest lung vessels are still not resolved.

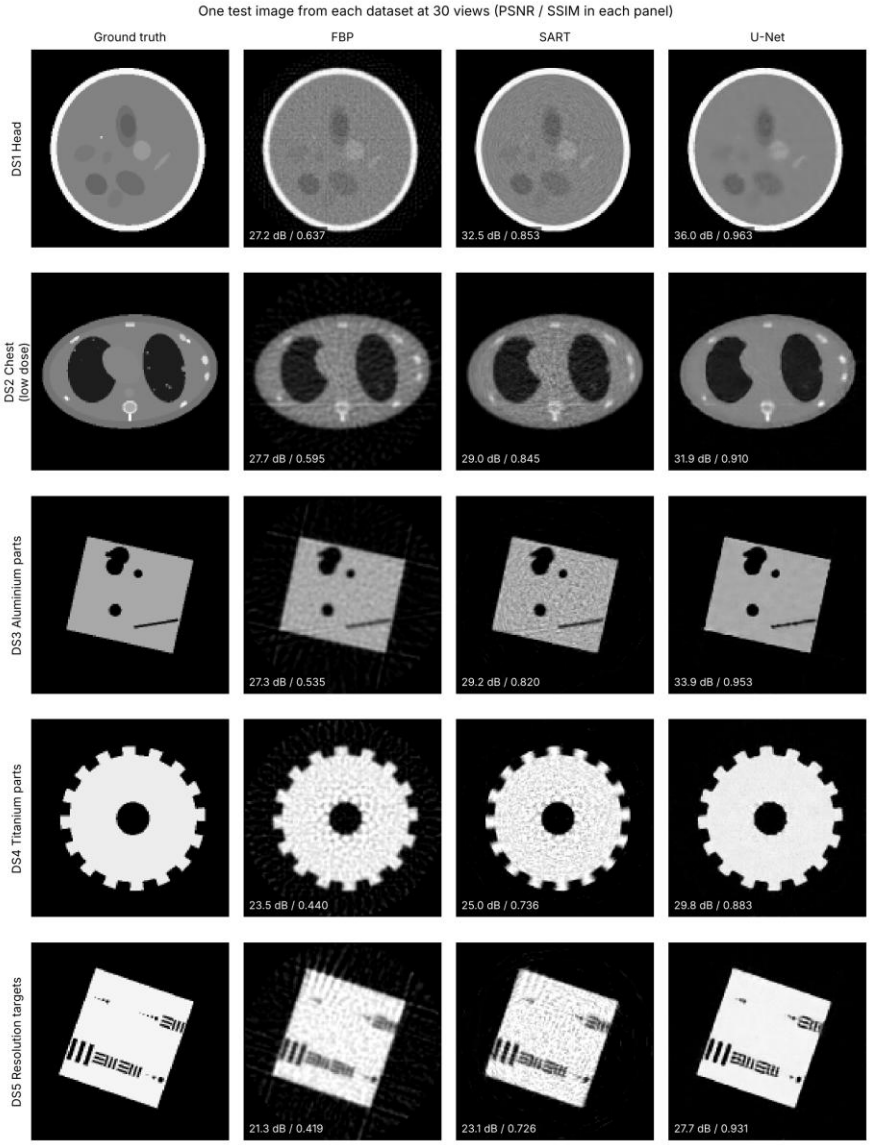


Figure 5. One test image per dataset at 30 views (columns: ground truth, FBP, SART, U-Net; PSNR / SSIM printed in each panel).

6.2 Cross-dataset transfer

Applying each 60-view network to every dataset's test set (Figure 6) shows that the matched network gives the largest gain over FBP in every column (+7.7, +4.9, +6.9, +8.7 and +7.3 dB for DS1-DS5). Networks transfer well among the three industrial datasets, which share piecewise-constant, sharp-edged objects (e.g. the DS3 network gains +7.3 dB on DS4), but transfer between medical and industrial data is weak or harmful; the DS5 network lowers PSNR on DS2 by 7.0 dB relative to FBP.

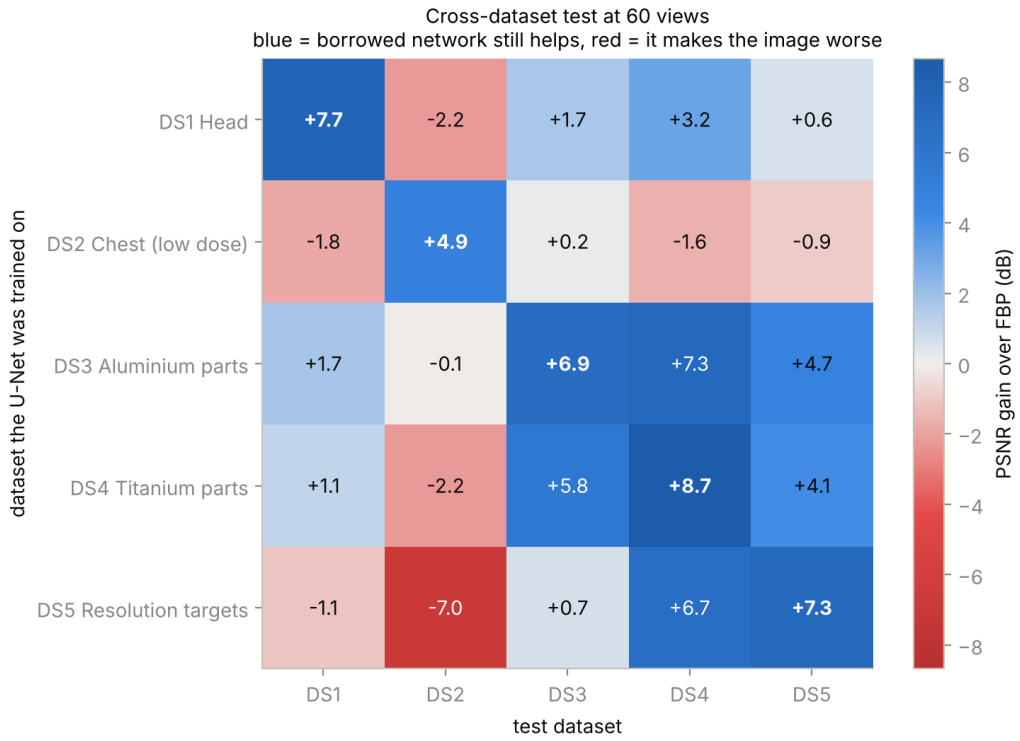


Figure 6. Cross-dataset test at 60 views: PSNR gain over the test set's own tuned FBP. Rows: training dataset; columns: test dataset. Blue: the borrowed network helps; red: it degrades the image.

6.3 View-count and dose mismatch

A network trained on sparser data copes with denser test data, but not the reverse. For DS5, the 30-view network reaches 30.1 dB on 90-view data, whereas the 90-view network falls from its matched 31.9 dB to 20.6 dB on 30-view data, no better than FBP (20.6 dB). Similar but smaller drops occur for all datasets (e.g. DS4: 28.0 dB vs 32.1 dB matched). Under dose mismatch at 60 views, the U-Net remained the best method down to one tenth of the training dose for DS1, DS3, DS4 and DS5, but for the low-dose chest (DS2) at $I_0 = 1000$ photons (0.1 × training dose) its PSNR fell to 16.6 dB, below SART (18.4 dB); at half the training dose it was still clearly best (31.1 vs 27.8 dB) (Figure 7).

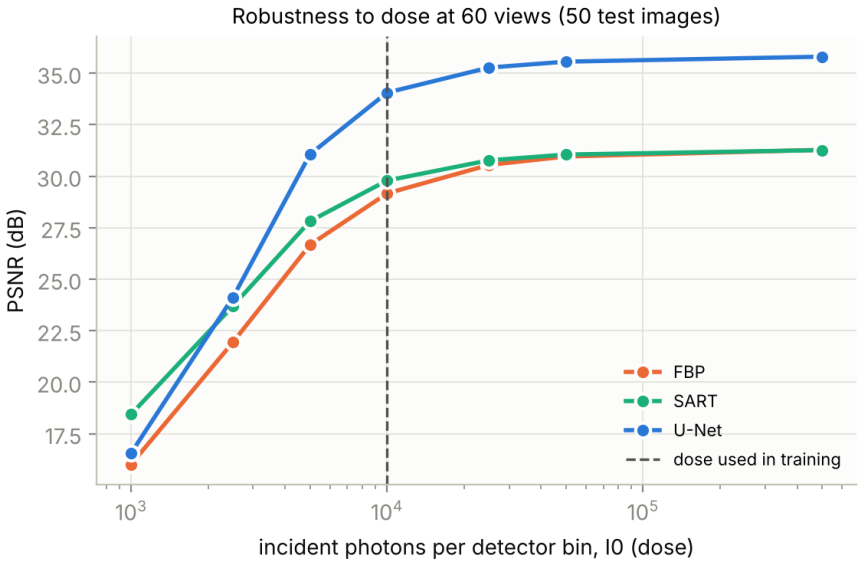


Figure 7. Dose robustness for DS2 (low-dose chest) at 60 views: mean PSNR over 50 test images as the incident photon count I_0 varies. Dashed line: dose used in training.

6.4 Objects never seen in training

On the Shepp-Logan phantom, SART achieved a higher PSNR than the U-Net in all 15 network and view combinations, including the networks trained on head phantoms (e.g. DS1 network at 30 views: 27.1 dB vs 35.6 dB for SART), although the U-Net's SSIM was sometimes higher. On the laboratory resolution plate, the network trained on resolution targets (DS5) outperformed SART by +3.4, +2.7 and +2.7 dB at 30, 60 and 90 views (Figure 8), the industrial networks DS3 and DS4 were roughly level with SART at 60-90 views, and the medical networks were worse than SART. Under DS2's physics the plate corresponds to more than 200 mm of bone-equivalent material, so all methods fail on it (≈ 12-14 dB) because of photon starvation.

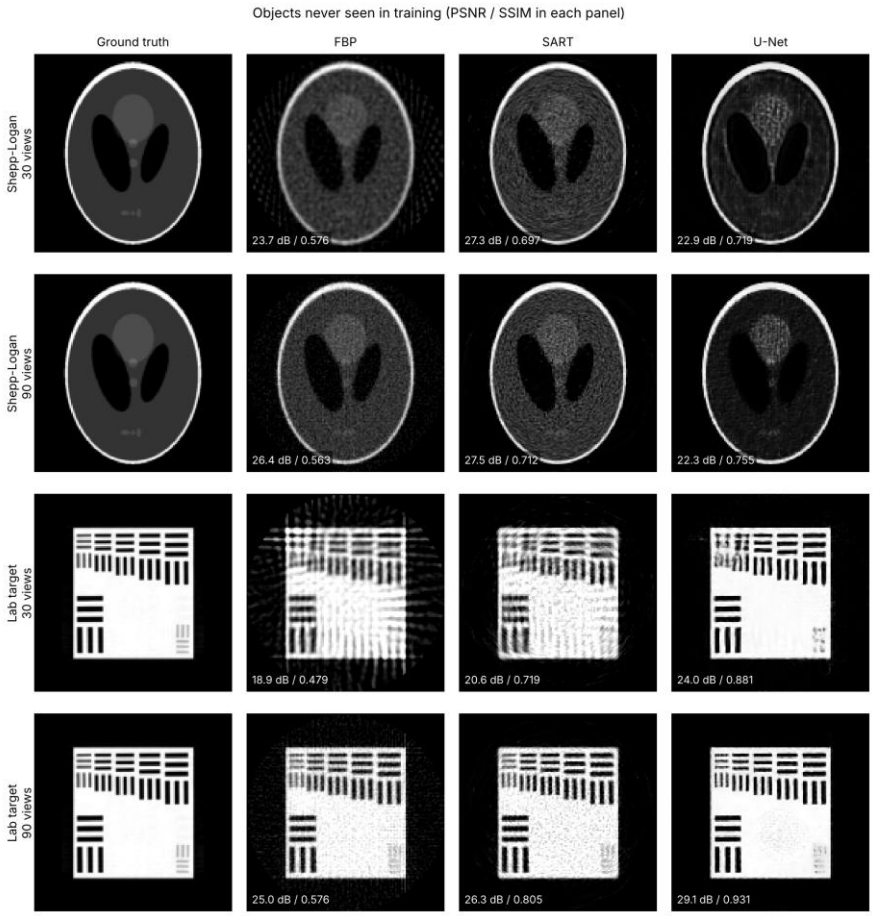


Figure 8. Objects never seen in training, reconstructed with the DS5 (resolution-target) networks. Top two rows: Shepp-Logan phantom at 30 and 90 views; bottom two rows: laboratory aluminium resolution plate at 30 and 90 views (columns: ground truth, FBP, SART, U-Net).

6.5 Run time

Per image on a two-core CPU, FBP took 4-13 ms, FBP followed by the U-Net 16-38 ms, and SART 81-1661 ms depending on the number of passes and views. The network therefore adds little to the cost of FBP and is one to two orders of magnitude faster than the iterative baseline.

7. Conclusion

Within the conditions it was trained for, the post-processing U-Net solved the sparse-view artefact problem considerably better than tuned classical methods: across five datasets with different objects, materials and doses it improved PSNR by 1.9-5.8 dB over the best of FBP and SART in every setting, raised SSIM to 0.90-0.98, and at 30 views surpassed FBP with six times as many views, at a computational cost of about 20 ms per image. These results support the use of sparse-view acquisition with learned post-processing to reduce dose or scan time.

The stress tests show that this success is conditional. The network learns a prior for the class of objects, the sampling and the noise level it was trained on. It transfers among similar object classes, but degrades or damages images of different objects, of sparser acquisitions and of much lower doses, and it does not match SART on the Shepp-Logan phantom. In agreement with Sidky et al. [5], the network does not solve the CT inverse problem in general; its training data must represent the intended application, and its outputs should be verified against the measurements before clinical or safety-critical use.

8. Limitations and future work

- All data are simulated in 2-D with parallel-beam geometry and monochromatic X-rays; real scanners use fan or cone beams and polychromatic spectra, with beam hardening and scatter.
- Images are 128 × 128 pixels and the network has 1.9 million parameters, chosen so the study runs on a CPU; clinical images are typically 512 × 512.
- Each network was trained once with a single random seed; differences below about 0.5 dB between networks (e.g. DS4 at 60 vs 90 views) are within run-to-run variation.
- PSNR, SSIM and RMSE do not measure diagnostic or inspection performance; no reader study was performed.
- The network is a pure post-processor and does not enforce consistency with the measured projections.
- The laboratory data were available only as a rendered preview of a reconstruction, not as raw projections.

Future work will apply the pipeline to raw laboratory cone-beam projections and to fan- and cone-beam simulations with beam hardening, add a data-consistency step or a learned primal-dual network [6], train on mixtures of view counts, doses and object classes with several random seeds, and explore score-based generative priors [7] that require no paired data and provide uncertainty estimates.

Code and data availability

All code (Python, NumPy, scikit-image, PyTorch), the five datasets with their training, validation and test splits, the 15 trained networks and every result table and figure are released with this article. The command *python run_all.py* regenerates the complete study from fixed random seeds.

Video: https://youtu.be/vrY9rs8ixTc

Acknowledgements

X-Ray CT SMILE Facility IITR

References

[1] K. H. Jin, M. T. McCann, E. Froustey, M. Unser, "Deep convolutional neural network for inverse problems in imaging," IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4509-4522, 2017. arXiv:1611.03679.

[2] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), LNCS 9351, pp. 234-241, 2015.

[3] A. H. Andersen, A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," Ultrasonic Imaging, vol. 6, no. 1, pp. 81-94, 1984.

[4] Y. Han, J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," IEEE Transactions on Medical Imaging, vol. 37, no. 6, pp. 1418-1429, 2018. arXiv:1708.08333.

[5] E. Y. Sidky, I. Lorente, J. G. Brankov, X. Pan, "Do CNNs solve the CT inverse problem?" arXiv:2005.10755, 2020.

[6] J. Adler, O. Öktem, "Learned primal-dual reconstruction," IEEE Transactions on Medical Imaging, vol. 37, no. 6, pp. 1322-1332, 2018. arXiv:1707.06474.

[7] Y. Song, L. Shen, L. Xing, S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," International Conference on Learning Representations (ICLR), 2022. arXiv:2111.08005.

[8] T. Dorosti, J. Thalhammer, S. Peterhansl, D. Pfeiffer, F. Pfeiffer, F. Schaff, "Beam geometry and input dimensionality: Impact on sparse-sampling artifact correction for clinical CT with U-Nets," arXiv:2508.17961, 2025.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.

[10] S. van der Walt et al., "scikit-image: image processing in Python," PeerJ, vol. 2, e453, 2014.

[11] L. A. Shepp, B. F. Logan, "The Fourier reconstruction of a head section," IEEE Transactions on Nuclear Science, vol. 21, no. 3, pp. 21-43, 1974.

[12] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," International Conference on Learning Representations (ICLR), 2015. arXiv:1412.6980.

[13] A. C. Kak, M. Slaney, Principles of Computerized Tomographic Imaging, IEEE Press, 1988 (reprinted SIAM, 2001).

[14] J. H. Hubbell, S. M. Seltzer, "Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients," NIST Standard Reference Database 126, 2004.