



Assessment of Dose-Reduction Strategies in Wavelength-Selective Neutron Tomography

M. Cyrus Daugherty^{1,2} · Victoria H. DiStefano¹ · Jacob M. LaManna¹ · David L. Jacobson¹ · Paul A. Kienzle¹ · Youngju Kim^{1,3} · Daniel S. Hussey¹ · Peter Bajcsy¹

Received: 7 October 2022 / Accepted: 15 June 2023

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023

Abstract

This study aims to determine an acquisitional and computational workflow that yields the highest quality spatio-spectral reconstructions in four-dimensional neutron tomography studies. The properties of neutrons enable unique image contrast modes, but accessing these modes requires defining the energy of the neutron beam, resulting in long acquisition times. We seek sparse angular tomography approaches to collect of order 100 tomograms at different neutron wavelengths using the minimum number of input projection images. In these computational image workflows, we identified and evaluated the main factors affecting the quality of the tomographic reconstruction such as the projection number, the reconstruction method, and the post-processing method and we report relationships between 3D reconstruction quality metrics and acquisition time. Based on these relationships, the performance of seeded simultaneous iterative reconstruction-based techniques (SIRT and SIRT with total variation regularization) yielded improved image quality and more accurate estimates of the reconstructed attenuation values compared to other methods, which included convolutional neural networks. The methods were then applied to a dose-reduced monochromatic dataset and characterized via signal-to-noise ratio (SNR) and single-voxel resolution.

Keywords 3D reconstruction image quality · Dose reduction · Neutron imaging · Bragg-edge tomography

Introduction

The properties of the neutron, a massive, neutral, spin-1/2 particle that interacts primarily through the strong nuclear force, enables one to create images with unique sources of contrast compared to other penetrating probes [1]. The diverse set of neutron image contrasts include quantitative

imaging of magnetic and electric fields with polarized neutron imaging [2, 3], characterizing the porosity with sub-pixel resolution through dark-field or phase imaging [4, 5], and crystal phase mapping with Bragg-edge imaging [6, 7], and can be thought of as hyperspectral images. Common to these measurements is the desire to form full tomographic datasets across a range of experimental conditions for the same sample, repeating the tomographic acquisition at many different neutron wavelengths or other instrument settings. However, with few neutron facilities, intrinsically weak neutron sources (especially compared to synchrotron facilities), and limited beam time per experiment, it is not possible to measure a fully sampled tomography dataset in each configuration. This work examines strategies for reconstructing neutron tomograms from sparse angular samples and employing computational imaging methods to minimize noise and artifacts.

In neutron tomography, the beam direction is fixed and the sample is rotated through a series of angles with a projection image taken at each angle. Similar to X-ray tomography, reconstructing the neutron attenuation of the sample can be realized via the Radon transform or iteratively via algebraic

This article is part of the topical collection “Advances on Image Processing and Vision Engineering” guest edited by Sebastiano Battiato, Francisco Imai and Cosimo Distanto.

✉ Daniel S. Hussey
daniel.hussey@nist.gov

✉ Peter Bajcsy
peter.bajcsy@nist.gov

¹ National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

² Department of Chemical and Biomolecular Engineering, University of Maryland, Gaithersburg, MD 20742, USA

³ Department of Chemistry and Biochemistry, University of Maryland, Gaithersburg, MD 20742, USA

and statistical methods [8]. The underlying measurement uncertainty follows Poisson counting statistics, with the signal-to-noise ratio (SNR) proportional to the square root of measurement time [9]. Further, neutron imaging studies are conducted on a wide array of samples of widely varying structure. This is a subtly different problem than presented in medical imaging, wherein anatomical structures share similarities between patients. Thus, we seek methods that require limited high-quality data for either training supervised models or forming *a priori* estimates for parameterized models.

A property of the wavelength-selective image contrasts is that the image contrast varies slowly with wavelength. We postulate that we can obtain quantitative data for the entire wavelength series by measuring one tomogram with a large number of projections (which we refer to as the “high-quality tomogram”) and using it to aid the reconstruction of a series of sparsely sampled measurements at the other wavelengths. The high-quality tomogram serves either as training data for machine learning, selecting subsets of the rotation angles to generate sparsely sampled training data or as an *a priori* estimate (i.e., seed) for iterative algorithms. To test this hypothesis, we used Bragg-edge imaging to identify the crystal phases in samples of well-known composition in a scene also comprised of unknown materials. We compared several different reconstruction and de-noising techniques including simultaneous iterative reconstruction technique (SIRT) as implemented in the ASTRA Toolbox [10, 11], a regularized iterative reconstruction method that imposes smoothness constraints (total variation minimization with SIRT: TV-SIRT, based on the work of Sidky et al. [12]), the stochastic primal dual hybrid gradient method (SPDHG) [13] as implemented via the core imaging library [14, 15] and three machine learning algorithms: a mixed-scale dense convolutional neural network (MS-D Net) [16, 17], the filtered back projection convolutional neural network (FBP-ConvNet) [18] and the hybrid-domain network (HDNet) [19]. Throughout the manuscript, these are referred to as the convolutional neural network (CNN) methods.

Our objective is to establish models and rankings among the factors that affect 3D reconstruction image quality and acquisition time. These factors can guide neutron imaging experimentalists in (a) maximizing image quality and minimizing acquisition time, and (b) quantifying the trade-offs between several image quality metrics and different dose-reduction approaches with a factorial design. In our experimental design, we varied the number of projections (60, 80, 360, 600, and 800) for each workflow (e.g., SIRT, SIRT+seed, TV-SIRT+seed) with the polychromatic dataset to determine baseline metrics. For the SIRT reconstructions, the FBPConvNet and MS-D Nets were trained as a post-processing step with the input training sets consisting of dose-reduced images paired with fully sampled images. The HDNet operates partially as a pre-processing step (projection

domain) and a post-processing step (image domain), with the reconstruction operation occurring between. The SPDHG method is included here as a method that incorporates smoothing along the energy channel, it is, therefore, exclusively analyzed in the monochromatic analysis and omitted from the polychromatic comparisons.

First, accuracy of reconstructed tomographic volumes is related to the number of acquired 2D projections via a theoretical relationship. This was followed by characterizing and contrasting the image quality of seeded SIRT reconstructions as a function of iterations. Next, all the reconstruction and CNN methods were compared across key image quality metrics (root mean square error (RMSE), a no-reference perceptual blur metric (NRPBM) and SNR). Finally, the performance of the different methods on a dose-reduced monochromatic scan was reported.

The contributions of our work lie in:

1. Designing a factorial experimental design to understand trade-offs between acquisition time and image quality of 3D tomographic reconstructions from neutron imaging data,
2. Evaluating (a) reference material-based image quality, such as SNR, (b) imaging quality focused metrics, such as blur, (c) reference 3D reconstruction acquired for over-sampled 2D projections, such as RMSE, and (d) theory for circularly symmetric objects and the relationship between intensity variance and the number of 2D projections.
3. Including the CNN-based methods as both post-processing (MS-D Net and FBPCovNet) and hybrid-domain (HDNet) de-noising methods to leverage high-quality *a priori* information.
4. Reporting single-voxel resolution and SNR as a function of wavelength for the different dose-reduction approaches.

The novelty of this work is in establishing model-based and ranking relationships between 3D reconstruction accuracy and acquisition time represented by intensity variance, SNR, RMSE, blur, number of 2D projections, number of iterations and seeding of 3D tomographic reconstruction (SIRT), and supervised and de-noising models (HDNet, MS-D Net and FBPCovNet).

Related Work

Tomographic Reconstruction Algorithms

The two most common computed tomography (CT) algorithms that reconstruct the raw 2D projections into 3D space are filtered back projection (FBP) and iterative algebraic

reconstruction methods. The mathematical theory for these algorithms is beyond the scope of this work but are detailed in [8]. In back projection (BP), a slice is reconstructed by “smearing out” the line integrals for each angle and summing them together. The FBP improves this process by applying a spatial frequency filter to account for the over-sampling of certain areas [20].

In recent years, improvements in computer processing have made iterative reconstruction (IR) techniques popular for dose and noise reduction. There are several types of IR algorithms, but they generally follow a procedure of, first, forward-projecting a reconstruction image (either initialized with a blank image or a reconstruction image) and create a simulated sinogram [21, 22]. Next, this simulated sinogram is compared to the sinogram of the raw data and corrections to the reconstruction image are made. The algorithm iterates through this process a set number of times or until some convergence criteria is met. In SIRT, the projection differences and sinogram differences are weighted. Additional details can be found in [8, 10, 21]. For both FBP and SIRT, the image quality and accuracy increase with an increasing number of projections.

While SIRT and other algebraic reconstruction methods are well suited for incorporating magnitude constraints and morphological *a priori* information, regularized methods can extend this to impose other desired features onto the reconstruction such as smoothness. Regularized methods are particularly beneficial to dose-reduced reconstructions because they can suppress some of the noise and artifacts that occur from sparse sampling (aliasing distortions). The adaptive steepest descent projection onto convex subsets (ASD-POCS) algorithm [12, 23] is one such regularized method that balances sensor-based image fidelity with image smoothness.

Improvements to spectral X-ray detectors and multi-modal imaging methods such as positron emission tomography (PET) have motivated the incorporation of smoothing approaches that leverage energy content to improve de-noising [24]. These methods naturally extend the regularization approaches by adding smoothness constraints across an energy dimension. Chambolle et al. developed both the primal dual hybrid gradient algorithm [25] and the more recent stochastic primal dual hybrid gradient (SPDHG) algorithm [13], which have seen high utility in optimizing regularized problems of this type for PET [24, 26] and neutron Bragg-edge tomography [27]. Recently, Yu et al. proposed a novel regularized method that adaptively filters reconstructions with global, local, and non-local image components for multi-energy CT images [28].

Image De-noising

With the orders of magnitude increase in performance available from graphics processing units (GPUs) and new techniques for training deep convolutional neural networks, the last decade has seen a tremendous increase in research on artificial intelligence and machine learning (ML) across many fields, particularly image processing. Foundational to the dose reduction-related ML work of the past decade is the theory of compressed sensing that subverts the requirements of the Nyquist sampling theory. Xu et al. employed compressed sensing techniques via global and adaptive dictionary-based statistical iterative reconstruction methods to improve low dose CT reconstructions [29]. These dictionary-based methods are effectively single-layer neural networks that learn a sparse representation of the CT images and preceded the large multi-layer convolutional networks that followed (e.g., U-net [30]). FBPCONVNet [18] is a U-net architecture consisting of a series of contractions, forcing generalization by limiting the size of the representation, paired with a series of expansions to generate the target image. An additional skip connection was added, reducing the vanishing gradient problem exhibited by deep networks early in training. The mixed-scale dense network (MS-D Net) [16] is an architecture that maintains the image size throughout, replacing the contractions with dilations to spread information across the image. Both of these architectures are trained using high-quality tomograms as the target, with tomograms from subsets of the projections as the source. The hybrid-domain network (HDNet) [19] operates in both the projection image domain (PID) and the reconstruction image domain (RID). The HDNet uses a modified U-net architecture for both PID and RID which replace the max pooling operations with strided convolutions that allow the networks to weight the down-sampling along the network contraction path. The HDNet is trained as follows: a weighted interpolation scales the sinograms up to the fully sampled size, which is the input to the PID U-net. The PID U-net is trained on these interpolated sinograms and, following training, its inferences are reconstructed and used as inputs to train the RID U-net. An extension of this hybrid-domain approach is the dual-domain attention guided convolutional neural network, which employs the same encoder–decoder-type architecture as the U-net, but uses spatial attention mechanism blocks to improve reconstruction fidelity [31]. Fu et al. revisited non-convolutional networks with a hierarchical domain decomposition that directly maps sinograms to the reconstruction image space [32]. Recently, a CNN architecture specifically tailored for hyperspectral images was proposed that dramatically reduces

the reconstruction time over iterative optimization-based methods [33]. Generative adversarial networks (GANs) have also seen utility in tomographic image de-noising as Liu *et al.* employed an architecture that uses a U-net generator with a typical convolutional discriminator [34]. Work on hyperspectral image de-noising is still a highly active area of research with recent work by Gkillas leveraging regularized methods and neural networks to improve image quality [35]. We point readers to reference [36] as a valuable resource for insights into the of deep learning for tomographic imaging from a medical perspective.

Image Quality Metrics

Comparing different image processing methods necessitates objective quality metrics to determine whether one approach is superior to another. We can evaluate (a) the reconstruction against a priori known data using the RMSE, (b) foreground vs. background discrimination using signal-to-noise ratio (SNR) over calibration regions, (c) conservation of high-frequency content with blurring metrics, and (d) the reconstruction accuracy as a function of the number of 2D projections following a theoretical model. Each quality evaluation requires some assumptions about a priori knowledge. RMSE assumes co-registered ground truth 3D reconstruction. SNR quantification requires known foreground and background masks. Blur metrics are derived from intensity histograms using multiple mathematical models that must be empirically chosen. In our work, the ground truth 3D reconstruction was established from over-sampled angular 2D projections (2400 projections). Next, foreground and background masks were created manually for two reference cylindrical objects filled with known material. Finally, a blur model was selected that aligned with human visual perception of blur [37, 38].

Materials and Methods

Figure 1 shows an overview of the key components in evaluating the trade-offs between acquisition time (dose reduction) and 3D reconstruction quality. These key components define the relationships among variables, such as number of 2D projections (or acquisition time), number of iterations during 3D reconstruction, variance of intensities in 3D reconstructed datasets, SNR, RMSE, blur, and availability of highly accurate seed for a 3D reconstruction algorithm and supervised post-processing model. Following Fig. 1, we describe each component in our assessment of dose-reduction strategies.

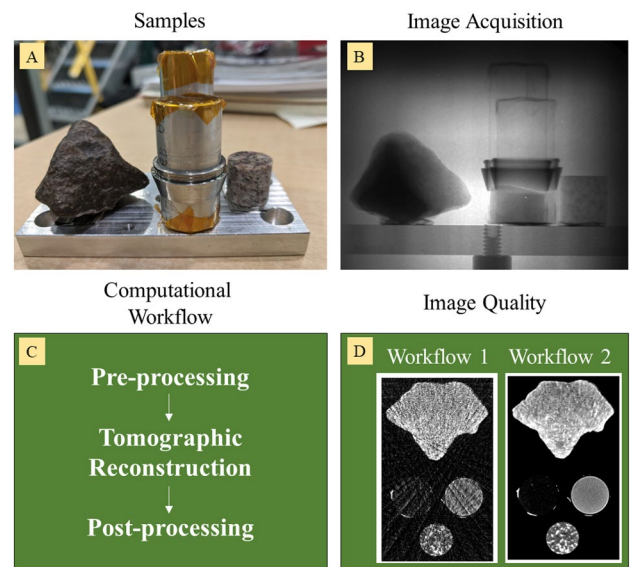


Fig. 1 An overview of processes for assessing the dose-reduction strategies

Samples

The test sample set consisted of four geological samples to analyze. The first two samples were a meteorite (LMT204) and a 1 cm diameter core of Westerly Granite, which has been extensively analyzed in [39, 40]. The other two samples were standard reference powders obtained from National Institute of Standards and Technology (NIST) Standard Reference Material (SRM) collection. Several grams of these powders were placed in separate 6061-aluminum tubes, with 316 stainless steel ferrules around them, and sealed on both ends with polyimide tape. The powders were not compacted or leveled off and aluminum tape was used to secure all the samples in place. Figure 1A shows an image of the samples before they were placed in the beam.

The SRM powders were used as reference objects for all the subsequent metric evaluations. The first powder, SRM 691—Reduced Iron Oxide, was an iron powder consisting of 90% by mass of iron and trace amounts of oxides and other metals. The second powder, SRM 70b—Potassium Feldspar, was prepared from a high-purity feldspar obtained from pegmatite deposits in the Black Hills of South Dakota. The material is a mixture of alkali feldspar, plagioclase feldspar, quartz, and a small amount of mica. The SRMs were blended and bottled at NIST.

Beam, Detector and Image Acquisition

Neutron tomography datasets were measured at the NG-6 Cold Neutron Imaging Instrument at the NIST Center for Neutron Research (NCNR) [41]. A densely sampled dataset

was collected, with 2400 evenly spaced projections over 360 degrees (2399 unique projections). This dataset is referred to herein as the “high-quality” dataset and serves as ground truth. Figure 1B shows an example projection image. The high-quality dataset was collected using a polychromatic neutron beam that can be approximated from a kinetic molecular theory as a Maxwell–Boltzmann distribution with characteristic temperature of about 50 K [42]. An Andor NEO scientific complementary metal oxide semiconductor (sCMOS) camera operating in 12-bit mode [43] was used to collect images from a P43 scintillator detector (i.e., gadolinium oxysulfide doped with terbium, $\text{Gd}_2\text{O}_2\text{S:Tb}$ also known as GdOx) with a Nikon Nikkor 50 mm f1.2 lens.

Each image was acquired over 4 s and the median of 5 images was taken for each projection, leading to maximum intensity around 3500 gray levels. The pixel pitch of the images was $51.35\mu\text{m}$ (resolution about $100\mu\text{m}$) and the field of view was 2560 pixels by 2160 pixels (~ 13 cm by 11 cm). This scan took approximately 16.5 h to complete. From this original dataset, several subsets were taken to simulate smaller projection numbers: 60, 80, 360, 600, and 800. Angles from these sets were evenly taken throughout the 2400 projections in order to span the same 360-degree sample space. If these datasets were collected on the polychromatic beam line, they would take 0.4 h, 0.6 h, 2.5 h, 4.2 h, and 5.5 h, respectively to acquire.

Additionally, monochromatic scans were obtained over a series of wavelengths from 0.200 nm to 0.588 nm in 0.002 nm increments (195 tomograms). Each of these sparsely sampled datasets are comprised of 80 evenly spaced projections collected over 360 degrees. Two highly oriented pyrolytic graphite crystal monochromators were used to select the wavelength and the $\Delta\lambda/\lambda$ was about 1%. The same camera and lens as above were used to collect images from a zinc sulfide/lithium fluoride (ZnS:LiF) scintillator. For these data to be consistent with the polychromatic datasets in terms of maximum intensity, the lower incident beam intensity required each projection image to be acquired over 10 s with a median of 3 images (maximum intensity 3500 counts). The pixel pitch of the images was $51.35\mu\text{m}$ (resolution about $250\mu\text{m}$) and the field of view was 2560 pixels by 2160 pixels (~ 13 cm by 11 cm). Each of these scans took approximately 0.8 h to complete. Using this same set up (monochromator, scintillator and camera), a scan was taken with 720 projections at a wavelength of 0.37 nm, which took approximately 7 h to complete. This wavelength was chosen arbitrarily to generate a data point for characterizing and validating the signal fidelity of the different de-noising methods applied to the dose-reduced (80 projection) scan at 0.37 nm. Note that it would have required over 55 continuous days to complete such a validation scan at all the wavelengths, hence, we only validated at one wavelength.

Computational Workflow

Image Pre-processing

As stated, the images from each projection angle were combined by taking the median to remove non-statistical noise such as gamma streaks and hot spots. An image with the beam off was used to subtract the additive noise of the detector background. An image of the open beam was used to normalize the projections, and a region of the image that did not contain a sample was used to correct for any small drift in beam intensity. The normalized projections were formed into sinograms, correcting for rotation axis tilt. Ring artifacts were removed from the sinograms by the algorithm developed by Vo et al [44]. As mentioned in the introduction, the HDNet differs from the other neural networks by operating in both the PID and the RID and is, therefore, an additional pre-processing step in its respective workflow.

Tomographic Reconstruction

For this study, we are using the Simultaneous Iterative Reconstruction Technique (SIRT) as implemented in the ASTRA toolbox [10, 11, 45]. A basic description of algebraic IR techniques was given in Sect. “Related Work”. Rigorously, SIRT can be written as in [10]

$$v' = v + CW^T R(p - Wv) \quad (1)$$

where v is the seed volume, p is the set of projections and W is the sparse projection matrix indicating which points in the volume contribute to which pixels in the projection. C and R are for bookkeeping, with $r_{ii} = 1/\sum_j a_{ij}$ and $c_{jj} = 1/\sum_i a_{ij}$, to avoid over-counting back-projected pixels inside the volume.

Two regularized reconstruction methods were also used in this work, the first of which is based on the ASD-POCS algorithm, which is referred to herein as TV-SIRT. The primary differences between this algorithm and ASD-POCS are that it employs SIRT instead of algebraic reconstruction technique (ART) as the reconstruction method, and it uses a prior image estimate (seed data) instead of being initialized to zeros. The pseudo-code for the ASD-POCS algorithm can be found on page 4788 of reference [12] and the implementation details of our TV-SIRT variant can be found in Appendix A. The general form of these regularized optimization problems can be stated as in [26]

$$f^* = \arg \min \{D(Wf - g) + \tau R(f)\} \quad (2)$$

where f^* is the optimized image, f is the current estimate of the image, W is the system matrix that describes the transformation between the reconstructed image and the sensor information, g is the sensor data. The operator D

is an operator which constrains the image fidelity within some norm (e.g., l^2) and the R operator penalizes solutions that contain undesirable features. The τ term balances these two operations. TV is the most common regularizing operator which is the $l^{2,1}$ norm of the gradient image ($\text{TV}(f) = \|\nabla f\|_{2,1}$) [23]. This is the method used herein for TV-SIRT.

The second regularized method incorporates both the 3D information and the spectral information. The formulation of this optimization problem is similar to Eq. 2, however, it includes a term for the energy channel that has its own regularization parameter (β) [14]:

$$f^* = \arg \min \frac{1}{2} \|Wf - g\|_2^2 + \beta \|f_e\|_1 + \tau \text{TV}(f_{x,y,z}) \quad (3)$$

where subscripts the e , x , y , z correspond with the energy channel and x , y and z directions, respectively, and β is the regularization parameter for the energy channel. We employ the stochastic primal dual-hybrid gradient (SPDHG) algorithm [13] as implemented via the core imaging library to optimize Eq. 3.

Image Post-processing

The MS-D Net and FBPCConvNet post-processing methods were applied to the reconstructions generated with SIRT, while HDNet also pre-processed sinograms, and, therefore, had its own reconstruction input that differed from the inputs to MS-D Net and FBPCConvNet. Each neural network was trained and validated with high/low image quality pairs using the Enki high-performance computing (HPC) cluster at NIST, which is equipped with NVIDIA Tesla V100 SXM2 GPUs. Throughout the manuscript, we use the letter M to indicate the number of projections and the subscripts Tr and Inf (M_{Tr} and M_{Inf}) to specify the number of projections a CNN was trained with and the number of projections upon which a CNN made inferences respectively. A total of 5 networks were trained, varying the low image quality input datasets ($M_{Tr} = x$ where $x \in \{800, 600, 360, 80, 60\}$) and using the 2400 projection SIRT reconstruction for the high-quality dataset throughout. The MS-D Networks were each trained for 100 epochs. Each network was then applied to all the low-quality datasets ($M_{Inf} = x$ where $x \in \{800, 600, 360, 80, 60\}$). Likewise, the FBPCConvNet was trained on each high/low image quality pairing, however, the conditions from the original manuscript [18] were used: 101 epochs, logarithmically decreasing learning rate from 0.01 to 0.001 with a tiling crop size of 128×128 pixels. The HDNet was likewise trained for 100 epochs in both the projection image space and the reconstruction image space with tiles of size 128×128 .

3D Reconstruction Quality Versus Acquisition Time

Accuracy and Number of 2D Projections

To assess the accuracy of each reconstruction method as a function of projection number (focusing on aliasing distortions [8]), we evaluated the standard deviation of the reconstructed values of the SRM powders, as shown in Fig. 4. As discussed in Kak and Slaney [8], for a circularly symmetric object, the variance of the reconstruction is proportional to the inverse of the number of projections ($\text{var} \propto M^{-1}$) and is approximated as

$$\text{var}\{\hat{f}(0, 0)\} = \frac{\pi^2 \tau}{MN_0} \int_{-\infty}^{\infty} h^2(t) dw \quad (4)$$

where M the number of projections, N_0 the number of neutrons detected in the center of the object, τ is the sampling width, and $h(t)$ is the filtering window, which is the ramp function for the SIRT algorithm used in this work.

SNR and RMSE

RMSE and SNR metrics were computed by leveraging reference data constructed via a tomographic reconstruction from 2400 projections and manual segmentation of image regions delineating standard reference materials in the field of view—see Fig. 2. We calculated RMSE and SNR using the definitions shown as follows:

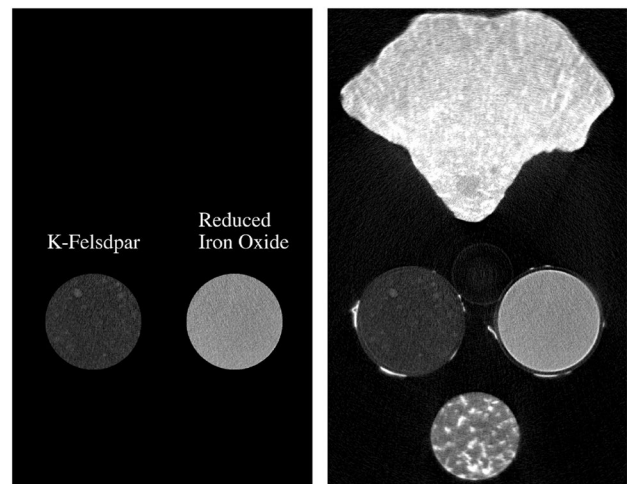


Fig. 2 Left—a cross-section of the reference powders with the mask overlain on top. The powder regions are labeled. Right—a cross-section of 3D tomographic reconstruction from 2400 projections which was considered as a reference

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n |y_i - \hat{y}_i|^2}{n}} \quad (5)$$

$$\text{SNR} = \frac{\bar{\mu}}{s} \quad (6)$$

where y_i is the reference value, \hat{y}_i is the measured value, n is the number of values being compared, $\bar{\mu}$ is the average intensity and s is the sample standard deviation of the attenuation.

Blur Metric

Blur mostly affects structure and detail (high-frequency content), which are largely absent in the reference powders. For this reason, the blur metrics were calculated over entire reconstruction images and not just the masked SRM powders. To determine the optimal blur metric for our datasets, we evaluated different blur metrics from [37, 38]. The no-reference perceptual blur metric (NRPBM) developed by Crete *et al.* [37] was chosen for this work as it aligns with human perception of blur. The NRPBM varies between 0 and 1 with higher NRPBM values correspond with higher perceptual blur and conversely low NRPBM values corresponding with low perceptual blur.

Experimental Results

To compare the under-sampled datasets with ground truth (2400 projections), we took samples at a regular interval from the polychromatic dataset to form datasets with 60, 80, 360, 600, and 800 projections, we also refer to these datasets as “low-quality”. From each of these under-sampled datasets, we computed five reconstructed volumes. We formed the SIRT estimate of the volume using the corresponding low-quality FBP as a prior estimate, and we designate these as SIRT. The second set of volumes seed the SIRT with a prior estimate from the high-quality (2400 projection SIRT dataset) reconstruction, and we designate these as SIRT+seed. For all the polychromatic datasets, 3 iterations were used for the SIRT and SIRT+seed datasets. Likewise, the TV-SIRT+seed datasets were calculated at the same down-sampling conditions and seeded with the 2400 projection SIRT dataset. The CNN methods were used to create remaining volumes by de-noising the unseeded SIRT volume (FBP-ConvNet and MS-D Net) and processing hybrid-domain data (HDNet). The 2400 projection dataset was used to determine the best approaches to evaluate the blur metrics for image quality. Polychromatic reconstructions and

de-noised images from the different methods are shown in Fig. 3.

Method-Based Image Quality Comparisons

Acquisition Time and Estimated Accuracy of 3D reconstruction

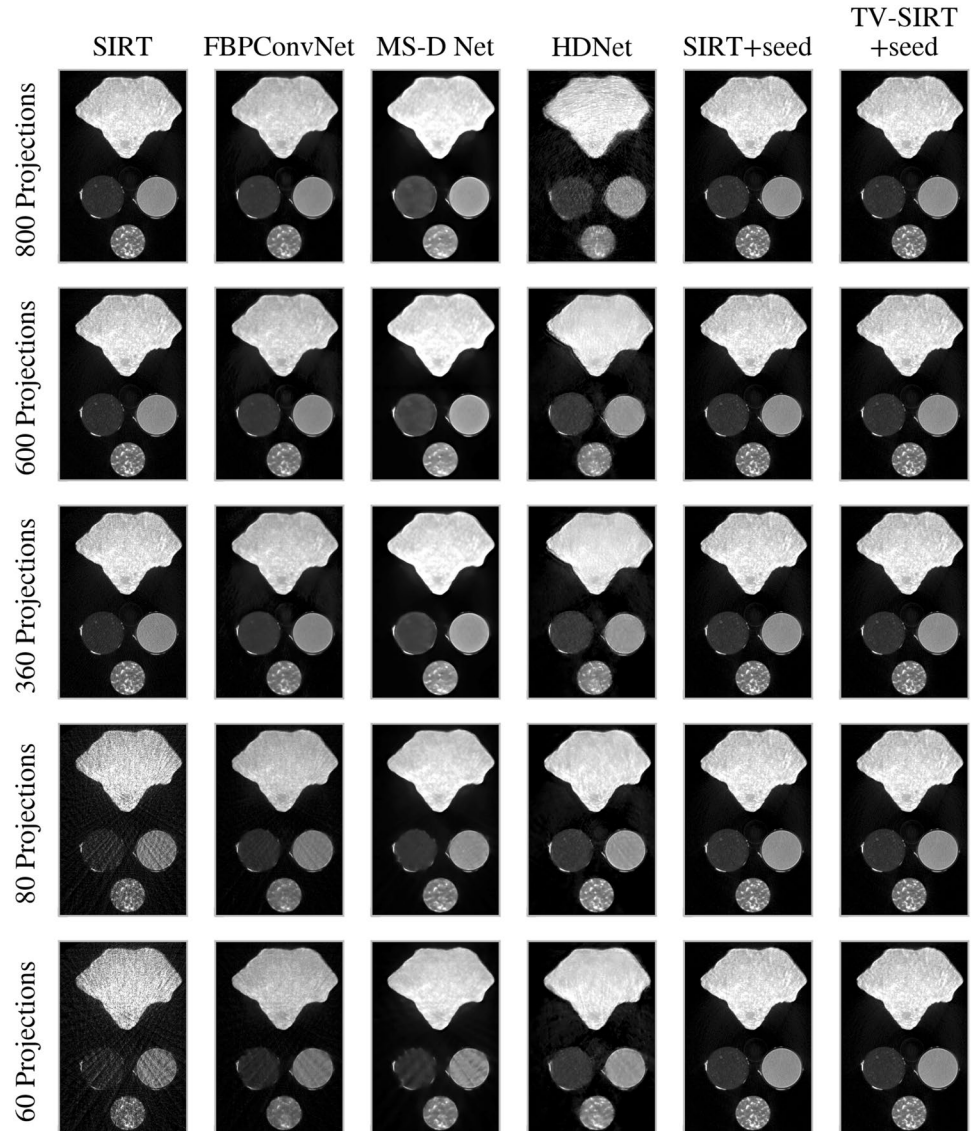
Acquisition time is directly proportional to the number of acquired 2D projections M . Following Eq. 4, the number of 2D projections (M) and the number of neutrons (N_0) influence the standard deviation. This is shown in Fig. 4 by different slopes and intercept of the fit of the standard deviation as a function of $M^{-1/2}$. The standard deviation for SIRT+seed, TV-SIRT+seed, MS-D Net and HDNet reconstructions do not possess the standard deviation dependence on projection number, but instead are approximately that of the 2400 projection SIRT dataset, which is used as the seed or ground truth. The FBPCovNet does exhibit some dependence on $M^{-1/2}$, however, it is non-linear with $M^{-1/2}$.

RMSE, NRPBM and SNR

The polychromatic reconstructions for each method are compared across RMSE and NRPBM at all down-sampling levels in Fig. 5. Note the box-plots at each down-sampling level (M) are arranged left to right in the same order as the legend, enabling identification of the smaller distributions (e.g., SIRT+seed’s RMSE). The reconstructions that comprise these distributions are those that were used to calculate the SNR in the SRM powders (115 frames total). The RMSE value for SIRT is observed deviating from the other methods as it becomes highly corrupted by aliasing distortions as M decreases. SIRT+seed conversely shows superior RMSE as it has only undergone 3 iterations (see Fig. 9) and is still highly biased toward its seed. TV-SIRT+seed, likewise, has low RMSE values; however, it exhibits no statistically meaningful difference in median as a function of M (overlapping box-plot notches [46]). The CNNs had differing trends with respect to M , where the FBPCovNet increased RMSE with decreased projections, and the MS-D Net decreased its RMSE. The HDNet exhibited the lowest RMSE for the case that it was trained on (80 projections), which suggests that this larger network generalizes more poorly than the MS-D net and is likely over-fitting the training data.

For the NRPBM results, SIRT exhibits the lowest perceptual blur at all down-sampling conditions. SIRT+seed shows very little response to number of projections as it does not change significantly over the M values. The FBPCovNet is

Fig. 3 Reconstruction and de-noising methods at different angular down-sampling conditions for polychromatic dataset. MS-D Net, FBPCConvNet and HDNet results are shown for networks trained on 80 projections



initially blurrier than SIRT+seed, but at the 60 and 80 projection conditions it becomes sharper. Likewise, the MS-D Net is initially blurrier than TV-SIRT+seed, but becomes sharper for the 60 and 80 projection cases. The HDNet produced the highest perceptual blur for the case that the network was trained on, indicating a high level of smoothing. TV-SIRT+seed generally has high perceptual blur as it iteratively reduces its total variation.

The results for the SNR in the two SRM regions for each method are shown in Fig. 6. SIRT+seed and TV-SIRT+seed, like in Fig. 5, exhibit virtually no change in SNR as a function of M . The CNNs, however, exhibit high SNRs for $M \geq 360$. This is likely due to high smoothing as seen in the NRPBM plots (Fig. 5). The increase in aliasing distortions

begins to corrupt the CNNs at the 60 and 80 projection conditions.

Method-Specific Trends

SIRT Dataset

The SNR masking method described in section “SNR and RMSE” was applied to the polychromatic datasets reconstructed with SIRT. The differences in SNR values calculated for the reduced iron oxide and K-feldspar shown in Fig. 2 (left) are due to the difference in the average attenuation intensity of the regions, a function of the properties of the reference powders, and will vary depending on the

Fig. 4 Standard deviation for the two regions of interest (SRMs) as a function of inverse square root of number of projections $M^{-1/2}$ ($var \propto M^{-1} \rightarrow std \propto M^{-1/2}$). We include the dashed lines to highlight the relationship of the SIRT dataset standard deviation to $M^{-1/2}$. In contrast to SIRT datasets, the SIRT+seed, TV-SIRT, MS-D Net and HDNet datasets show approximately constant standard deviations as a function of number of projections. The FBPConvNet does not provide as much smoothing and exhibits a non-linear dependence on $M^{-1/2}$

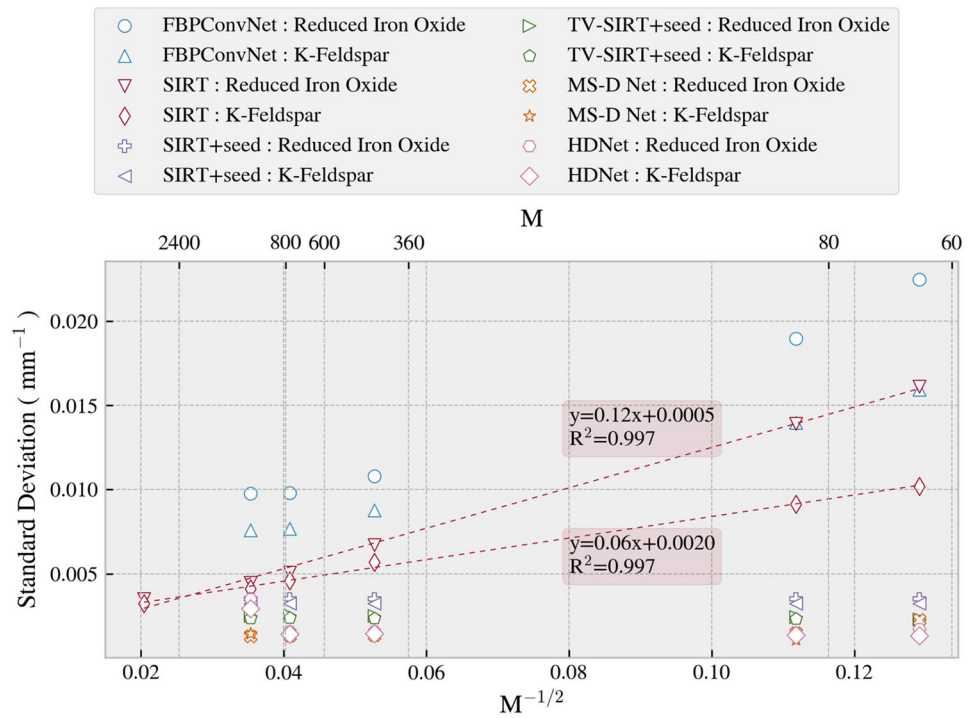
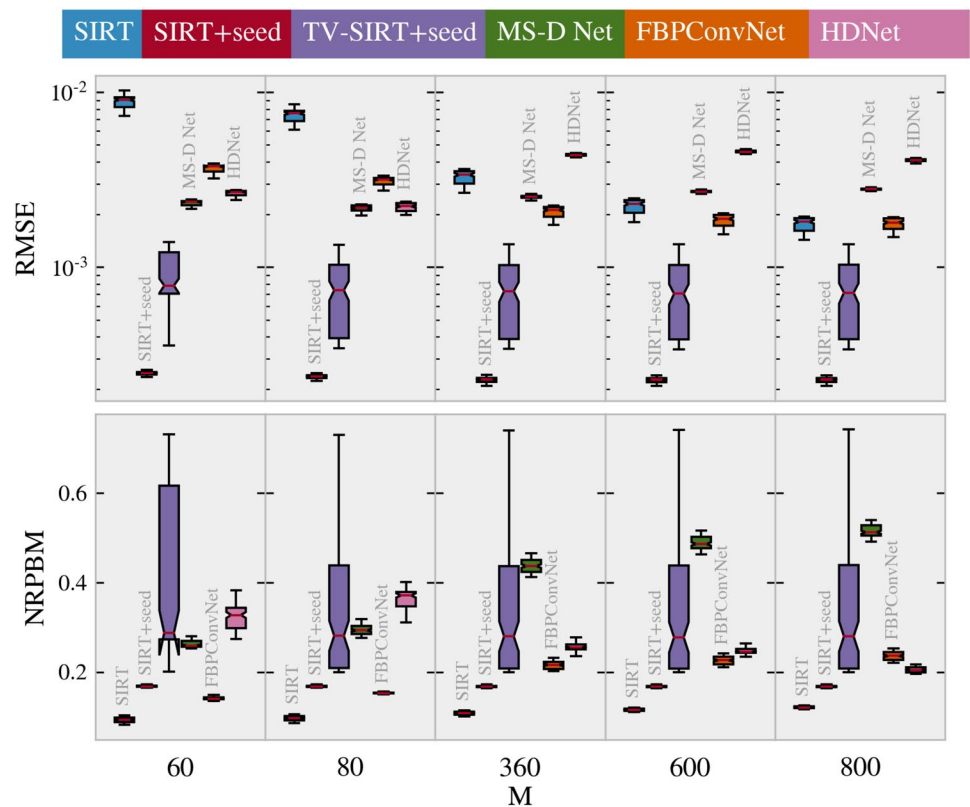


Fig. 5 RMSE and NRPBM for all methods as a function of number of projections (M). The MS-D Net, FBPConvNet and HDNet distributions shown here are for the 80 projection condition ($M_{Tr} = 80$). Note these data are all polychromatic. Recall lower and higher NRPBM values indicate lower and higher perceptual blur respectively. The SNR results that correspond with these volumes are shown in Fig. 6. These distributions are composed of the values of reconstruction slices that were used to compute the SNRs (115 reconstructions)



homogeneous material being analyzed. As with most neutron imaging datasets, the noise in the data is dominated by Poisson counting statistics with the SNR is given by

$$SNR = \frac{N_{events}}{\sqrt{N_{events}}} \tag{7}$$

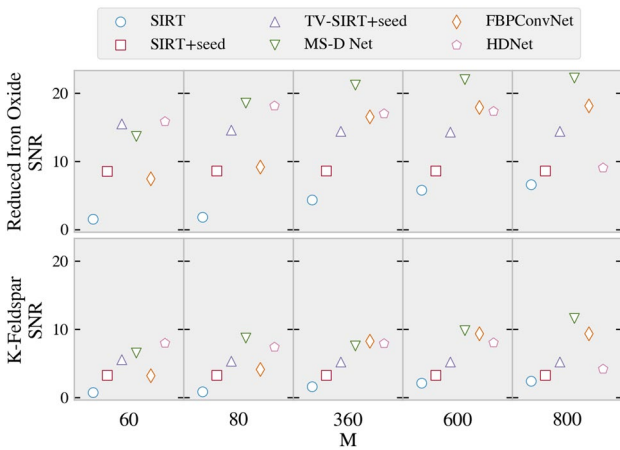


Fig. 6 SNR for the different methods as a function of number of projections (M). These data correspond to the same volumes in Fig. 5. Again, the CNNs shown here were trained on (M_T) 80 projections

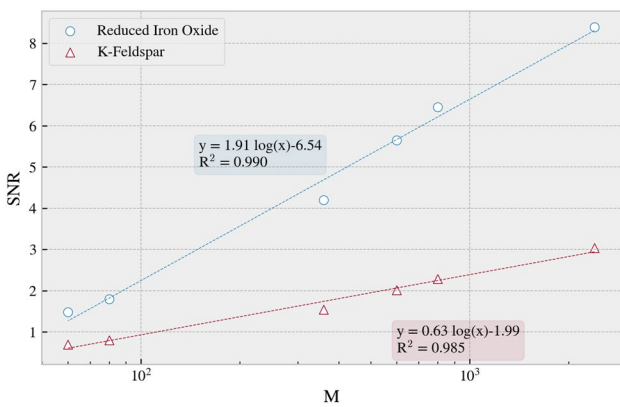


Fig. 7 The SNR as a function of the number of projections M in the reference powder regions (method = SIRT with no seed; data are polychromatic)

where N_{events} is the number of detected events (incident neutrons) [9]. Thus, a factor of four increase in counting statistics via increased exposure time or increased projections produces a factor of two increase in SNR. The application

of the Beer–Lambert Law, shown in the following equation, transforms this into a logarithmic relationship:

$$T = \frac{I}{I_0} = e^{-\mu t} \tag{8}$$

where I is the measured intensity, I_0 the incident intensity, T is the transmission, t is the thickness, and μ is the attenuation, a product of the neutron cross-section and the atom density (dependent on the material). The expected logarithmic relationship between SNR and the number of 2D projections is observed for the data shown in Fig. 7.

SIRT+seed Dataset

Applying the SIRT+seed reconstruction method requires trade-offs between accuracy, compute time, and image quality as illustrated in Figs. 8 and 9, which show sample reconstructions at 3, 20, 100 and 1000 iterations (Fig. 8) and the dependence of SNR, RMSE, and NRPBM on the SIRT iterations (N) for both a polychromatic reconstruction (left column) and a monochromatic reconstruction (right column) (Fig. 9). We include the monochromatic reconstruction here to highlight the dependence of the image quality on the input sinogram. Note that the RMSE in Fig. 9 for the polychromatic is computed with respect to the 2400 projection polychromatic scan and the monochromatic is computed with respect to the 720 projection monochromatic scan (0.37 nm). The polychromatic reconstruction maintains generally high SNR and low RMSE in both regions (iron and K-feldspar) up to around 100 iterations. The local peak for RMSE corresponds with a dip in both SNR and NRPBM, which indicates the initial influence of the low-quality sinogram as the reconstruction moves away from the seed. The subsequent decrease in RMSE (from $N=3$ to $N=30$) corresponds with increased smoothing from SIRT that is evidenced in the increasing NRPBM in that region. Above 100 iterations, the polychromatic reconstruction begins to be dominated by the aliasing distortions in the low-quality sinogram which

Fig. 8 SIRT+seed reconstructed slices using 3, 20, 100, and 1000 iterations (N). The top row shows the monochromatic reconstructions at 0.37 nm and the bottom row shows the polychromatic. These data correspond with the image quality metrics shown in Fig. 9. The images are cropped to the reduced iron oxide powder region (see Fig. 2)

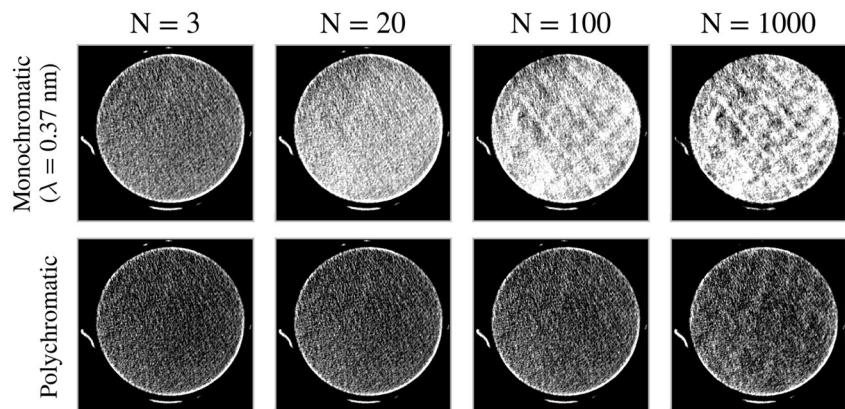
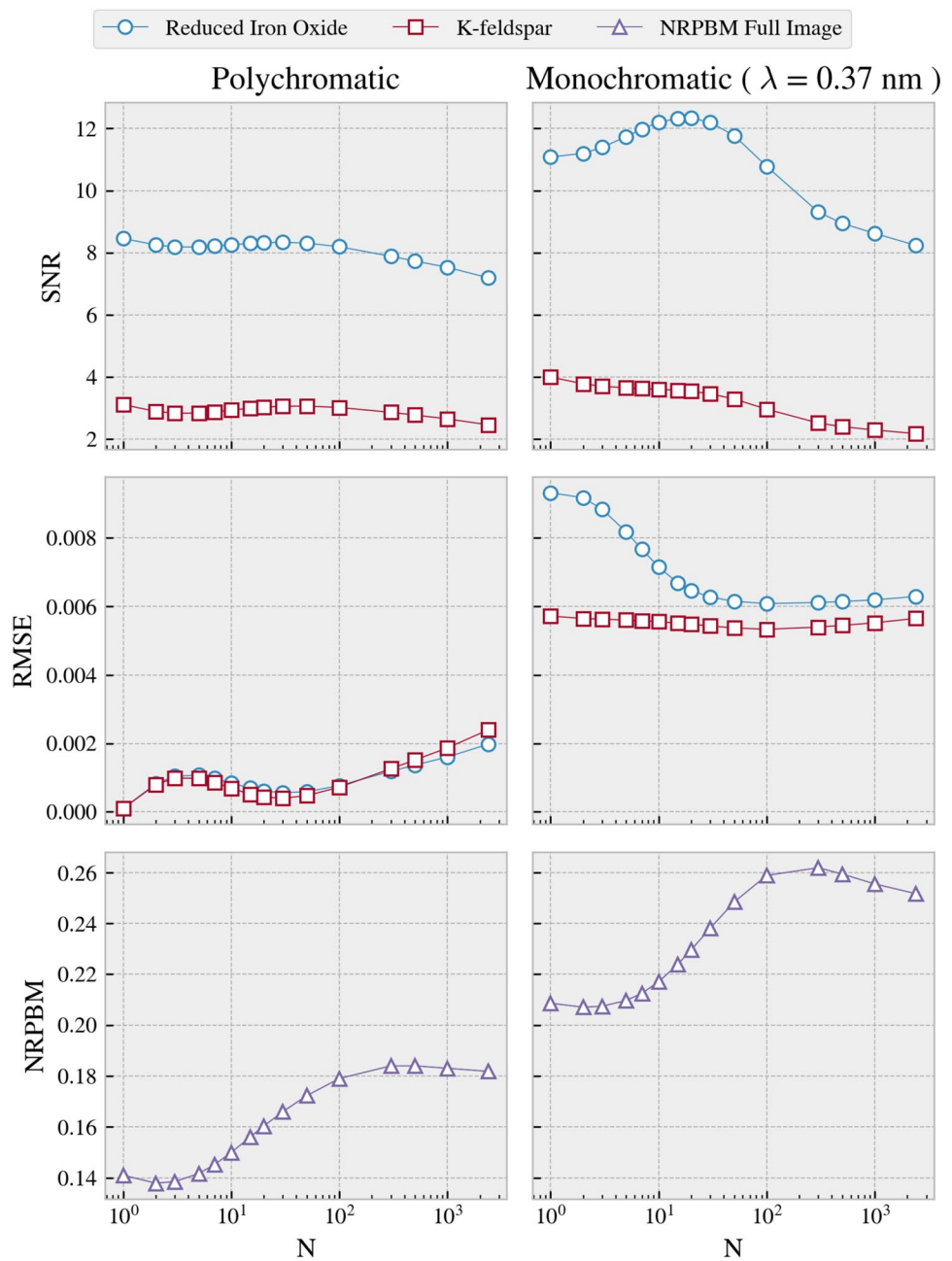


Fig. 9 SIRT+seed image quality metrics as a function of iterations (N). The left column shows the metrics for a polychromatic reconstruction image (frame 121). The right column shows the metrics for the monochromatic 0.37 nm reconstruction of the same frame. Figure 8 shows reconstructions for sample N values (3, 20, 100 and 1000). Both reconstructions were seeded with the 2400 projection polychromatic unseeded SIRT. Note the RMSE for the monochromatic reconstruction is with respect to the 720 projection monochromatic reconstruction



lowers the SNR and increases RMSE. The monochromatic reconstruction exhibits increasing SNR up to 20 iterations for the iron region and slightly decreasing SNR for the K-feldspar region while the RMSE for both regions passes through a local minimum at 100 iterations. The NRPBM for the monochromatic reconstruction, like the polychromatic, increases up to around 300 iterations then begins to decrease. These data illustrate that the polychromatic image quality metrics effectively diverge from higher quality to lower quality as the iterations incorporate increasingly more sensor data. This contrasts with the monochromatic reconstruction that passes through local optima for

the SNR in the reduced iron oxide region and the RMSE for both regions. The decrease in SNR for both monochromatic and polychromatic as the iterations increase (above 100 for polychromatic, above 20 for monochromatic) is due to over-fitting the noise, which is a known feature of SIRT [47].

CNN Datasets

Next, we analyzed the machine learning post-processing method for each trained network on the polychromatic datasets. We calculated the SNR for each calibration region and

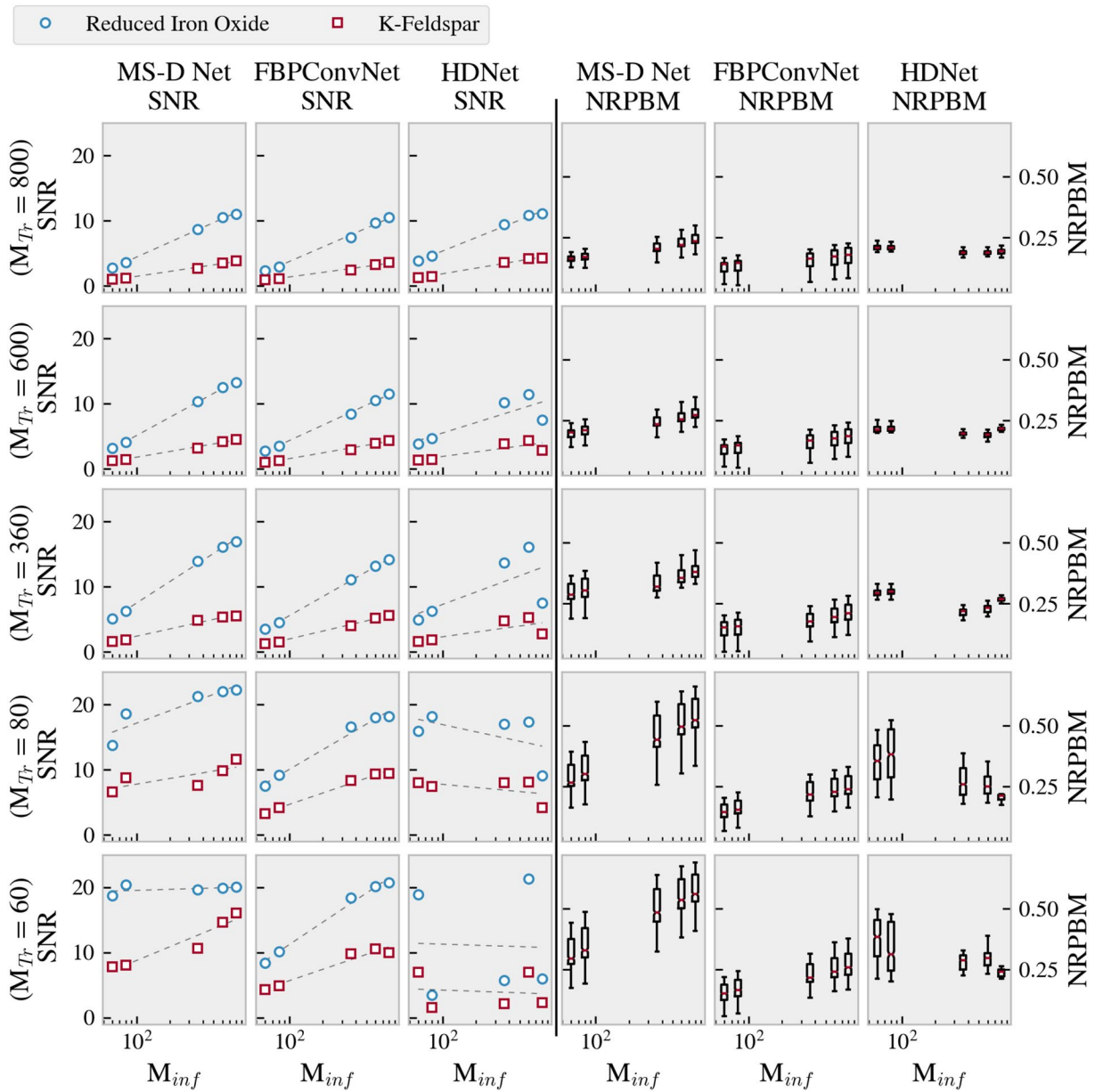


Fig. 10 The SNR and NRPBM values for the trained CNNs making inferences on the 3D polychromatic datasets. The row designations: $M_{Tr} = x$, indicate the number of projections that the network was trained on for that row where $x \in \{800, 600, 360, 80, 60\}$. The left three columns show the results for the MS-D Net, FBPCConvNet, and

HDNet, respectively, for the SNR in the SRM powders, and, the right three columns show the respective results for the NRPBM (for the entire volume). The x-axis for all plots is labeled M_{inf} indicating the number of projections for the dataset on which the inferences were made

for each network as a function of the number of projections using the same procedure as before and the NRPBM for the whole volume (see Fig. 10).

The top three rows of Fig. 10 show the results for the networks trained on ($M_{Tr} =$) 800, 600, and 360 projections respectively. The FBPCConvNet and MS-D Net increase the SNR at all conditions over the unseeded SIRT reconstructions (compare with Figure 7) and, SNR magnitudes

generally follow the same logarithmic increase that was observed for the SIRT reconstructions that are highlighted in Fig. 10. As the number of training projections (M_{Tr}) decreases, we likewise observe an increase in NRPBM at all inference conditions.

The fourth and fifth rows of Fig. 10 show the results for the networks trained on ($M_{Tr} =$) 80 and 60 projections, respectively. The FBPCConvNet still follows the general

Table 1 Image quality metrics for monochromatic 80 projection \rightarrow 720 projection ($\lambda = 0.37$ nm). The best method for each criterion is shown in bold (highest SNR; lowest RMSE)

Method	SNR		Median RMSE \pm std ($\times 10^{-3}$)		
	Iron oxide	K-Feldspar	Iron oxide	K-Feldspar	Full image
SIRT	1.29	0.68	25.9 \pm 0.54	14.7 \pm 0.41	13.26 \pm 0.90
SIRT+seed (N=3)	12.20	4.09	9.34 \pm 0.14	6.08 \pm 0.08	4.07 \pm 0.23
SIRT+seed (N=20)	12.67	3.82	7.57 \pm 0.09	6.10 \pm 0.08	4.02 \pm 0.25
TV-SIRT+seed	32.53	9.25	7.87 \pm 0.12	5.66 \pm 0.09	3.66 \pm 0.26
MS-D Net	14.80	5.32	7.42 \pm 0.11	5.97 \pm 0.15	4.61 \pm 0.33
FBPConvNet	4.22	1.24	11.0 \pm 0.67	8.93 \pm 0.29	7.01 \pm 0.62
HDNet	11.13	5.61	7.79 \pm 0.19	6.10 \pm 0.12	5.44 \pm 0.26
SPDHG	17.39	5.61	7.18 \pm 0.12	5.73 \pm 0.10	3.26 \pm 0.32
Ground truth (SIRT—720 Pro)	5.40	1.74	—	—	—

trend of increasing SNR with the log of number of projections, however, the MS-D Net begins to diverge from this trend, particularly for the $M_{Tr} = 60$ condition. We see, again, the relationship of increasing SNR at a given down-sampling condition (M_{Inf}) that is coupled with increasing levels of perceptual blur (increased SNR at constant $M_{Inf} \rightarrow$ increased NRPBM). For this reason, we attribute the ability for the networks to produce excessively high SNR values to over-smoothing the reconstructions, which is an undesirable image feature for our use-cases. The HDNet behaved differently from the other networks as its output across different down-sampling conditions did not vary smoothly (see column 3). Likewise, the levels of perceptual blur varied non-linearly as seen in column 6.

Monochromatic Datasets

Table 1 summarizes the performance of the different methods on the monochromatic dataset at 0.37 nm, which was scanned at both 80 and 720 projections as described in section “SNR and RMSE”. We show two different SIRT+seed reconstructions here to illustrate the difference between 3 iterations and 20 iterations (maximum SNR in the reduced iron oxide powder at $N = 20$ —see Fig. 9). TV-SIRT+seed produced the highest SNR for both SRM powders and the lowest RMSE in the K-feldspar SRM while SPDHG yielded the lowest RMSE both the iron oxide powder and the full image. Note these RMSE values were calculated with respect to the fully sampled 0.37 nm 720 projection tomogram. The MS-D Net was ranked as the best of the CNNs across these image quality metrics, except for the HDNet having slightly higher SNR in the K-Feldspar SRM. The 720 projection reconstruction was not noise-free as evidenced by its SNR values which are higher than only FBPConvNet and unseeded SIRT.

Performance as a Function of Wavelength

Lastly, we assess the performance of the different dose-reduction methodologies over the spectrum of monochromatic scans as described in section “SNR and RMSE”. Figure 11 shows the results for the resolution of a single voxel across all the scanned wavelengths. The image arrays show the reconstruction and inset of the reduced iron oxide powder, with the voxel probe position highlighted with the red \times . Using the crystallography database software GSAS [48], the location and relative intensities of Bragg peaks were calculated as shown in the dashed vertical lines in the bottom row of Fig. 11. Two different wavelengths are shown (sharing a constant greyscale range) to highlight the decrease in attenuation at 0.42 nm. Both the FBPConvNet and SIRT produce large variations in attenuation across the scanned wavelengths as seen in the bottom row of Fig. 11. The MS-D Net, HDNet and SPDHG improve over these methods, but still do not readily resolve all the expected Bragg edges. The 20 iteration SIRT+seed case shows higher attenuation than the 3 iteration case and more noise across wavelengths. SIRT+seed with 3 iterations and TV-SIRT+seed both resolve the voxel smoothly over all the wavelengths and align closely with the modeled data.

Taking a large statistical sample of (over 4 million) voxels from the SRM powders (via the masks), Fig. 12 illustrates the mean, standard deviation and SNR for the different methods across all the wavelengths. The mean for each method captures the expected Bragg-edge position, however, the standard deviations vary strongly which impacts the SNR. SIRT and FBPConvNet both show large standard deviations which lower their respective SNR values significantly. Both the MS-D Net and HDNet exhibit relatively high SNR (> 10) in the 0.25–0.4 nm range but have larger standard deviations at the extrema of the scanned wavelengths. Likewise,

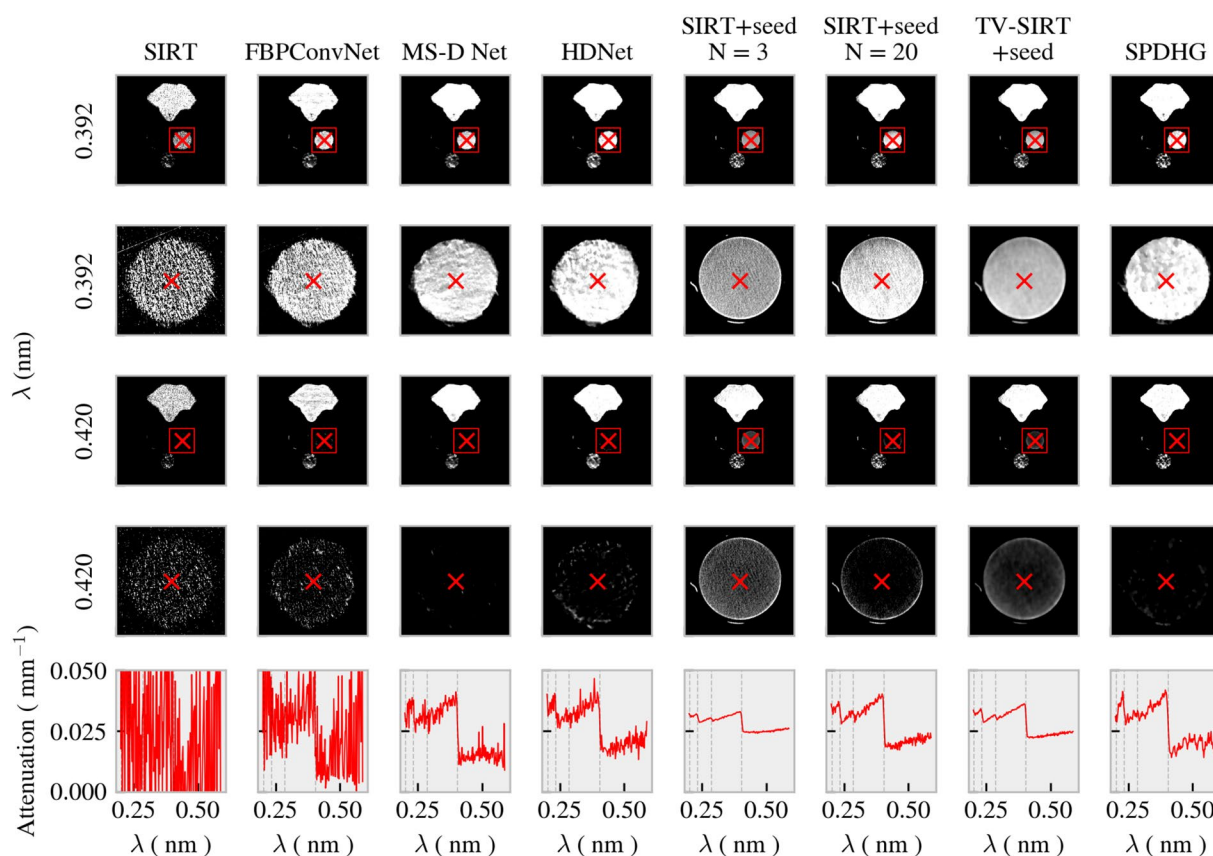


Fig. 11 Single-voxel resolution of dose-reduction methods as a function of wavelength. The top two rows show a reconstruction at the 0.392 nm wavelength with the inset box on the first row expanded on the second row. Likewise, the third and fourth rows show the recon-

struction and its respective inset at 0.42 nm. The bottom row shows the attenuation of the voxel highlighted with the red \times at all the measured wavelengths with the vertical lines designating the GSAS [48] predicted edge positions

TV-SIRT+seed exhibits high SNR for both SRMs but also has wavelength dependence. Both the 3 and 20 iteration SIRT+seed reconstructions exhibit a larger standard deviation than TV-SIRT, however, this value is nearly constant over the scanned spectrum, producing SNR values that are effectively a scaled version of the mean. SPDHG, which exhibited high SNR for 0.37 nm (Table 1), shows wavelength-related SNR dependence qualitatively similar to the CNN methods.

Discussion

This work examined seven separate methods of improving energy-resolved 3D tomographic reconstructions at neutron imaging beam lines: (1) one baseline reconstruction method as a function of varying input numbers of projections, (2) two methods as a function of incorporated seeds into iterative 3D reconstruction algorithms, (3) two post-processing methods as a function of incorporated non-linear mappings derived from existing datasets, (4) one hybrid-domain

method that operated in both the projection and reconstruction domains and (5) one method that incorporated smoothing along the energy-channel of the data without a high-resolution seed. The first reconstruction algorithm (SIRT) established baseline metrics for analyzing neutron tomograms, including SNR and the NRPBM. The second two reconstruction algorithms (SIRT+seed and TV-SIRT+seed) used a high-quality dataset to initialize the reconstruction. The third approaches applied machine learning algorithms (MS-D net and FBPCConvNet) to sharpen and de-noise the reconstruction images. The hybrid-domain smoothed both the interpolated sinogram data in the projection domain and the subsequent reconstructions in the image domain. Last, the SPDHG algorithm incorporated spatial and spectral information to iteratively smooth the monochromatic reconstructions.

Using the metrics determined when analyzing the SIRT datasets, we found that the SIRT+seed method could utilize a high-quality dataset of similar attenuation values and the same shape to reconstruct unknown datasets with trade-offs between accuracy, time, and image quality (see Fig. 9). The

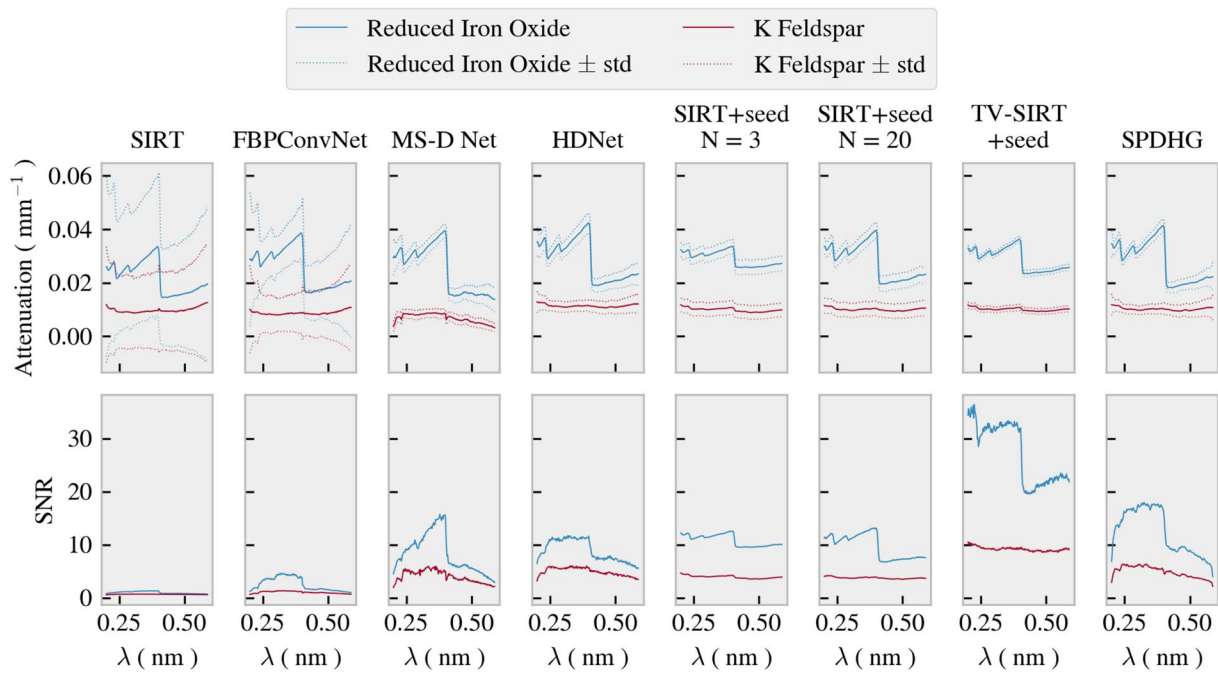


Fig. 12 SNR of reference powders as a function of wavelength. Each wavelength comprises 4,060,122 voxels

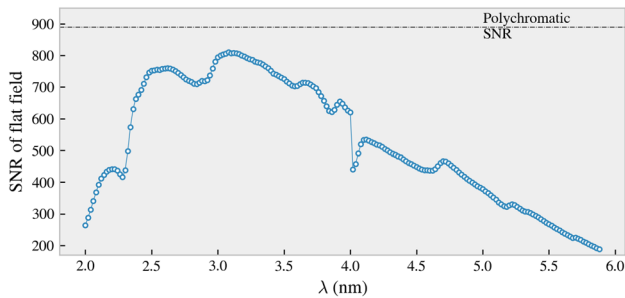


Fig. 13 Variation in SNR of the flat field (open beam) across the spectrum (polychromatic is included for reference). Noise from this source (lower neutron counts) manifests differently from computing a reconstruction with sparse projections

success of this method is contingent upon having a seed dataset of high-quality that is co-registered with the monochromatic datasets. Experiments that involve significant stochastic or morphological changes to the internal structure of the sample violate this condition and may be better served by different approaches. Employing SIRT in this context can dramatically decrease the time required to reconstruct, denoise and collect datasets, allowing more advanced neutron imaging methods to be more widely adopted.

The post-processing methods using the CNNs demonstrated the potential of these methods for low-projection datasets, especially if the algorithm is trained on a dataset with the same number of projections. These networks showed improvements in SNR values at the expense of

higher NRPBM which indicates a trade-off in image quality between fidelity and smoothness. As shown in the monochromatic wavelength sweep, care must be taken when applying machine learning models across multiple configurations on the neutron imaging conditions.

This work is not intended to be dismissive of CNN-based methods for de-noising neutron tomograms on account of two considerations: network optimization and experimental configuration. Each of the CNN-based methods was trained with hyper parameters and out-of-the-box architectures that could be sub-optimal for the size of these datasets. Indeed, there are many other architectures for tomogram de-noising that were not tested in this work that could prove to be superior to seeded TV-SIRT. Also, the ability for iterative tomographic reconstruction methods to incorporate explicit a priori information proved to produce superior reconstructions in this context where the CNNs were trained on data that imply the mapping from low-quality image to high. Figure 13 illustrates a key challenge in applying the networks trained on polychromatic data to the monochromatic data: the underlying projection images contain wavelength-dependent noise that is larger at the extrema of the surveyed wavelengths. We expect this variation in input noise, which roughly coincides with low SNRs for the CNN methods in Fig. 12, to account for this phenomenon. Future work applying neural networks to de-noise wavelength-dependent neutron tomograms may benefit from modeling the noise process and augmenting NN training with it, particularly with implementations, such as the HDNet that can operate

directly on the noisier projection domain. Measuring the wavelength-dependent attenuation of a registered static sample lends itself to seeded iterative reconstruction methods by having essentially no morphological variability across scans. Bragg-edge imaging requires such a configuration. However, there are other cases of neutron tomography where the sample undergoes internal morphological changes and where the seeding geometry is not as accurate and can require many iterations to reconstruct accurately. Instances such as this could see high utility in employing these CNN-based methods.

The performance of the methods on the dose-reduced monochromatic data demonstrates their ability to improve image quality of dose-reduced data with different attributes than the polychromatic (see Figs. 11 and 12). The seeded SIRT-based methods (SIRT+seed N=3, SIRT+seed N=20, and TV-SIRT+seed) both produce superior reconstructions as a function of wavelength and allow a simpler computational pipeline that can be optimized and tested more rapidly than the CNN-based approaches (see Fig. 9). While the SPDHG algorithm produced very promising results (see Table 1 and Fig. 12), it exhibits the same dependence on wavelength that the CNN methods showed. Further work can incorporate a high-resolution seed into this spatio-spectral smoothing to produce superior results where Poisson noise dominates.

In terms of modeling approaches, the conventional approaches to solving the de-noising and artifact removal problems use mathematical and statistical models. These approaches have limitations when mathematical and statistical models do not capture real data complexity. The deep learning models are supervised data-driven methods that can increase the accuracy of conventional approaches by adapting the data-driven models to specific training datasets and hence capturing real data complexities. The aim of our work was to explore whether data-driven models were increasing accuracy and by how much with respect to conventional models. While the data-driven models have been improving in their model capacity and complexity over time, the used CNNs did not lack model capacity and/or model complexity to encode the neutron imaging datasets (Note: model capacity for memorization of each input is highly undesirable for general purpose models). Thus, comparing performances of the presented neural networks to the higher capacity neural network architectures would not change the conclusions about the utility of neural networks for addressing the de-noising and artifact removal problems in neutron Bragg-edge tomography.

Future work will continue to look at computationally efficient regularized methods as they have shown high utility for this problem, but are still very time-intensive. Likewise, further CNN investigations will look at novel architectures and variations of segmentation-task neural networks, which

have seen high portability to tomographic de-noising problems. The transformer/V-Net architecture employed by Zhao *et al.* is one such network that employs a novel architecture to segmentation [49]. Likewise, MRI-based techniques will continue to be investigated for these problems [50, 51].

Conclusions

This work presented (1) an experimental design to understand the trade-offs between acquisition time and image quality of 3D tomographic reconstructions from neutron imaging data, (2) evaluations of SNR, RMSE, blur metrics, and intensity variance as measurements of image quality and their relationships to acquisition parameters, (3) integration of the CNN model-based de-noising to leverage previously acquired high-quality dataset and (4) characterized the methods on a dose-reduced monochromatic dataset. The family of seeded SIRT-based methods, including the regularized TV-SIRT, exhibited superior SNR and single-voxel resolution over all the wavelengths, by comparison to the unseeded SIRT and CNN methods.

A Appendix

The modified ASD-POCS algorithm (TV-SIRT) is stated here (Algorithm 1). The original ASD-POCS algorithm can be found on page 4788 of reference [12]. The differences between TV-SIRT and ASD-POCS are:

1. We use 10 gradient descent steps (ng in line 2) instead of 20
2. We use a seed dataset as input (line 4) instead of initializing \vec{f} with zeros
3. We use SIRT instead of algebraic reconstruction technique (ART) in (line 7)
4. We set the stopping criteria to an image residual lower than 0.005: $|\vec{f} - \vec{f}_0| < 0.005 \rightarrow$ Stopping Criteria = True

The gradient of the TV norm with respect to each pixel is approximated by

$$\nabla_{\vec{f}} \|\vec{f}\|_{TV} = \frac{(f_{s,t} - f_{s-1,t}) + (f_{s,t} - f_{s,t-1})}{\sqrt{\epsilon + (f_{s,t} - f_{s-1,t})^2 + (f_{s,t} - f_{s,t-1})^2}} - \frac{(f_{s+1,t} - f_{s,t})}{\sqrt{\epsilon + (f_{s+1,t} - f_{s,t})^2 + (f_{s+1,t} - f_{s+1,t-1})^2}} - \frac{(f_{s,t+1} - f_{s,t})}{\sqrt{\epsilon + (f_{s,t+1} - f_{s,t})^2 + (f_{s,t+1} - f_{s-1,t+1})^2}} \quad (9)$$

where s and t subscripts are for two-dimensional x and y pixel coordinates and ϵ is 10^{-8} , which keeps the denominators from being zero.

Algorithm 1 TV-SIRT Algorithm

```

1:  $\beta \leftarrow 1.0$ ;  $\beta_{\text{red}} \leftarrow 0.995$ 
2:  $ng \leftarrow 10$ ;  $\alpha \leftarrow 0.2$ 
3:  $r_{\text{max}} \leftarrow 0.95$ ;  $\alpha_{\text{red}} \leftarrow 0.95$ 
4:  $\vec{f} \leftarrow \text{Seed}$ 
5: while Stopping Criteria = False do
6:    $\vec{f}_0 \leftarrow \vec{f}$ 
7:    $\vec{f} \leftarrow \vec{f} + \beta * \text{SIRT}(\vec{g} - M \cdot \vec{f})$ 
8:   for  $i = 1, N_i$  do: if  $\vec{f}_i < 0$  then  $\vec{f}_i = 0$ 
9:    $\vec{f}_{\text{res}} \leftarrow \vec{f}$ 
10:   $\vec{g} \leftarrow M\vec{f}$ 
11:   $dd \leftarrow |\vec{g} - \vec{g}_0|$ 
12:   $dp \leftarrow |\vec{f} - \vec{f}_0|$ 
13:  if {first iteration} then  $dtvg \leftarrow \alpha * dp$ 
14:   $\vec{f}_0 \leftarrow \vec{f}$ 
15:  for  $i = 1, ng$  do
16:     $\vec{d}f \leftarrow \nabla_{\vec{f}} \|\vec{f}\|_{\text{TV}}$ 
17:     $\hat{d}f \leftarrow \vec{d}f / |\vec{d}f|$ 
18:     $\vec{f} \leftarrow \vec{f} - dtvg * \hat{d}f$ 
19:  end for
20:   $dg \leftarrow |\vec{f} - \vec{f}_0|$ 
21:  if  $dg > r_{\text{max}} * dp$  and  $dd > \epsilon$  then  $dtvg \leftarrow dtvg * \alpha_{\text{red}}$ 
22:   $\beta \leftarrow \beta * \beta_{\text{red}}$ 
23:  update Stopping Criteria
24: end while
25: return  $\vec{f}_{\text{res}}$ 

```

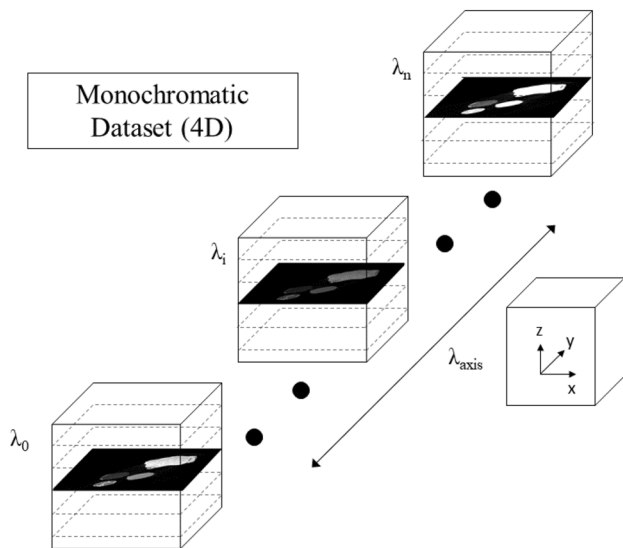


Fig. 14 Schematic of 4D dataset acquired by monochromatic neutron imaging

Dataset Structure

The monochromatic data were represented as a 4D array with x, y, and z coordinates accompanied by a wavelength coordinate (see Fig. 14).

Acknowledgements This work was partially funded by the National Institute of Standards and Technology and the University of Maryland.

Data availability The data is not available to the public.

Declarations

Conflict of interest On behalf of all the authors, the corresponding authors state that there is no conflict of interest.

Disclaimer No approval or endorsement of any commercial product by NIST is intended or implied. Certain commercial software, products, and systems are identified in this report to facilitate better understanding. Such identification does not imply recommendations or endorsement by NIST, nor does it imply that the software and products identified are necessarily the best available for the purpose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source,

provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Helmut Rauch; Samuel A. Werner: Neutron Interferometry : Lessons in Experimental Quantum Mechanics, Wave-particle Duality, and Entanglement, 2nd edn. Oxford University Press, Oxford, UK (2015)
2. Hilger A, Manke I, Kardjilov N, Osenberg M, Markötter H, Banhart J. Tensorial neutron tomography of three-dimensional magnetic vector fields in bulk materials. *Nature Communications* 9(1) (2018). <https://doi.org/10.1038/s41467-018-06593-4>
3. Jau Y.Y, Hussey D.S, Gentile T.R, Chen W. Electric Field Imaging Using Polarized Neutrons. *Physical Review Letters* 125(11) (2020). <https://doi.org/10.1103/PHYSREVLETT.125.110801>
4. Strobl M. General solution for quantitative dark-field contrast imaging with grating interferometers. *Scientific Reports* 4 (2014). <https://doi.org/10.1038/srep07243>
5. Brooks A.J, Knapp G.L, Yuan J, Lowery C.G, Pan M, Cadigan B.E, Guo S, Hussey D.S, Butler L.G. Neutron imaging of laser melted SS316 test objects with spatially resolved small angle neutron scattering. *Journal of Imaging* 3(4) (2017). <https://doi.org/10.3390/jimaging3040058>
6. Woracek, R., Penumadu, D., Kardjilov, N., Hilger, A., Boin, M., Banhart, J., Manke, I.: 3D mapping of crystallographic phase distribution using energy-selective neutron tomography. *Advanced Materials* 26(24), 4069–4073 (2014). DOI: 10.1002/adma.201400192.
7. Vitucci, G., Minniti, T., Di Martino, D., Musa, M., Gori, L., Micieli, D., Kockelmann, W., Watanabe, K., Tremsin, A.S., Gorini, G.: Energy-resolved neutron tomography of an unconventional cultured pearl at a pulsed spallation source using a microchannel plate camera. *Microchemical Journal* 137, 473–479 (2018). DOI: 10.1016/j.microc.2017.12.002.
8. Kak AC, Slaney M. Principles of Computerized Tomographic Imaging. New York: Society for Industrial and Applied Mathematics; 2001.
9. Lewandowski, R., Cao, L., Turkoglu, D.: Noise evaluation of a digital neutron imaging device. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 674, 46–50 (2012). DOI: 10.1016/j.nima.2012.01.025.
10. van Aarle, W., Palenstijn, W.J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K.J., Sijbers, J.: The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy* 157, 35–47 (2015). DOI: 10.1016/j.ultramic.2015.05.002.
11. van Aarle, W., Jan Palenstijn, W., Cant, J., Janssens, E., Bleichrodt, F., Dabrovolski, A., Beenhouwer, J.D., Joost Batenburg, K., Sijbers, J.: Fast and Flexible X-ray tomography using the ASTRA toolbox. *Optics Express* 24(22), 25129–25147 (2016).
12. Sidky, E.Y., Pan, X.: Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine and Biology* 53(17), 4777–4807 (2008). DOI: 10.1088/0031-9155/53/17/021.
13. Chambolle, A., Ehrhardt, M.J., Richtarik, P., Schonlieb, C.B.: Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications. *SIAM Journal on Optimization* 28(4), 2783–2808 (2018). DOI: 10.1137/17M1134834.
14. Papoutsellis E, Ametova E, Delplancke C, Fardell G, Jørgensen J.S, Pasca E, Turner M, Warr R, Lionheart W.R.B, Withers P.J. Core Imaging Library - Part II: Multichannel reconstruction for dynamic and spectral tomography. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379(2204) (2021). <https://doi.org/10.1098/rsta.2020.0193>
15. Jørgensen J.S, Ametova E, Burca G, Fardell G, Papoutsellis E, Pasca E, Thielemans K, Turner M, Warr R, Lionheart W.R.B, Withers P.J. Core Imaging Library - Part I : a versatile Python framework for tomographic imaging (2020). <https://www.ccpi.ac.uk/CIL>
16. Pelt, D.M., Sethian, J.A.: A mixed-scale dense convolutional neural network for image analysis. *Proceedings of the National Academy of Sciences* 115(2), 254–259 (2018). DOI: 10.1073/pnas.1715832114.
17. Pelt, D., Batenburg, K., Sethian, J.: Improving Tomographic Reconstruction from Limited Data Using Mixed-Scale Dense Convolutional Neural Networks. *Journal of Imaging* 4(11), 128 (2018). DOI: 10.3390/jimaging4110128.
18. 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). DOI: 10.1109/TIP.2017.2713099.
19. Hu, D., Liu, J., Lv, T., Zhao, Q., Zhang, Y., Quan, G., Feng, J., Chen, Y., Luo, L.: Hybrid-Domain Neural Network Processing for Sparse-View CT Reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences* 5(1), 88–98 (2021). DOI: 10.1109/TRPMS.2020.3011413.
20. Schofield, R., King, L., Tayal, U., Castellano, I., Stirrup, J., Pontana, F., Earls, J., Nicol, E.: Image reconstruction: Part 1 - understanding filtered back projection, noise and image acquisition. Elsevier (2020). DOI: 10.1016/j.jcct.2019.04.008.
21. Tayal, U., King, L., Schofield, R., Castellano, I., Stirrup, J., Pontana, F., Earls, J., Nicol, E.: Image reconstruction in cardiovascular CT: Part 2 - Iterative reconstruction; potential and pitfalls. Elsevier (2019). DOI: 10.1016/j.jcct.2019.04.009.
22. Stiller, W.: Basics of iterative reconstruction methods in computed tomography: A vendor-independent overview. *European Journal of Radiology* 109, 147–154 (2018). DOI: 10.1016/j.ejrad.2018.10.025.
23. Sidky, E.Y., Kao, C.M., Pan, X.: Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-Ray Science and Technology* 14(2), 119–139 (2006).
24. Warr, R., Ametova, E., Cernik, R.J., Fardell, G., Handschuh, S., Jørgensen, J.S., Papoutsellis, E., Pasca, E., Withers, P.J.: Enhanced hyperspectral tomography for bioimaging by spatio-spectral reconstruction. *Scientific Reports* 11(1), 20818 (2021). DOI: 10.1038/s41598-021-00146-4.
25. 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). DOI: 10.1007/s10851-010-0251-1.
26. Ehrhardt, M.J., Markiewicz, P., Schönlieb, C.B.: Faster PET reconstruction with non-smooth priors by randomization and preconditioning. *Physics in Medicine and Biology* 64(22), 225019 (2019). DOI: 10.1088/1361-6560/ab3d07.
27. Ametova E, Burca G, Chilingaryan S, Fardell G, Jørgensen J.S, Papoutsellis E, Pasca E, Warr R, Turner M, Lionheart W.R.B, Withers P.J. Crystalline phase discriminating neutron tomography using advanced reconstruction methods. *Journal of Physics D:*

- Applied Physics 54(32), 325502 (2021). <https://doi.org/10.1088/1361-6463/ac02f9>
28. Yu X, Cai A, Li L, Jiao Z, Yan B. Low-dose spectral reconstruction with global, local, and nonlocal priors based on subspace decomposition. *Quantitative Imaging in Medicine and Surgery* 13(2), 889–911 (2023). <https://doi.org/10.21037/qims-22-647>
 29. Xu, Q., Yu, H.Y., Mou, X.Q., Zhang, L., Hsieh, J., Wang, G.: Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Transactions on Medical Imaging* 31(9), 1682–1697 (2012). DOI: 10.1109/TMI.2012.2195669.
 30. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28. <http://lmb.informatik.uni-freiburg.de/>
 31. Chao L, Zhang P, Wang Y, Wang Z, Xu W, Li Q. Dual-domain attention-guided convolutional neural network for low-dose cone-beam computed tomography reconstruction. *Knowledge-Based Systems* 251 (2022). <https://doi.org/10.1016/j.knsys.2022.109295>
 32. Fu L, De Man B. Deep learning tomographic reconstruction through hierarchical decomposition of domain transforms. *Visual Computing for Industry, Biomedicine, and Art* 5(1) (2022). <https://doi.org/10.1186/s42492-022-00127-y>
 33. Huang, W.C., Peters, M.S., Ahlebæk, M.J., Frandsen, M.T., Eriksen, R.L., Jørgensen, B.: The application of convolutional neural networks for tomographic reconstruction of hyperspectral images. *Displays* 74, 102218 (2022). DOI: 10.1016/j.displa.2022.102218.
 34. Liu, Z., Bicer, T., Kettimuthu, R., Gursoy, D., De Carlo, F., Foster, I.: TomoGAN: low-dose synchrotron x-ray tomography with generative adversarial networks: discussion. *Journal of the Optical Society of America A* 37(3), 422–434 (2020). DOI: 10.1364/josaa.375595.
 35. Gkillas, A., Ampeliotis, D., Berberidis, K.: Connections between Deep Equilibrium and Sparse Representation models with Application to Hyperspectral Imaging. *IEEE Transactions on Image Processing* 32, 1513–1528 (2023). DOI: 10.1109/TIP.2023.3245323.
 36. Wang, G., Ye, J.C., De Man, B.: Deep learning for tomographic image reconstruction. *Nature Machine Intelligence* 2(12), 737–748 (2020). DOI: 10.1038/s42256-020-00273-z.
 37. Crete F, Dolmiere T, Ladret P, Nicolas M. The blur effect: perception and estimation with a new no-reference perceptual blur metric. In: *Human Vision and Electronic Imaging XII*, vol. 6492, p. 64920 (2007). <https://doi.org/10.1117/12.702790>
 38. Pertuz, S., Puig, D., Garcia, M.A.: Analysis of focus measure operators for shape-from-focus. *Pattern Recognition* 46(5), 1415–1432 (2013). DOI: 10.1016/j.patcog.2012.11.011.
 39. Bingham P, Polsky Y, Anovitz L. Neutron imaging for geothermal energy systems. In: Bingham, P.R., Lam, E.Y. (eds.) *Image Processing: Machine Vision Applications VI*, vol. 8661, p. 88610 (2013). International Society for Optics and Photonics. <https://doi.org/10.1117/12.2004617>
 40. Gates, C.H., Perfect, E., Lokitz, B.S., Brabazon, J.W., McKay, L.D., Tyner, J.S.: Transient analysis of advancing contact angle measurements on polished rock surfaces. *Advances in Water Resources* 119, 142–149 (2018). DOI: 10.1016/j.advwatres.2018.03.017.
 41. Hussey D.S, Brocker C, Cook J.C, Jacobson D.L, Gentile T.R, Chen W.C, Baltic E, Baxter D.V, Doskow J, Arif M. A New Cold Neutron Imaging Instrument at NIST. In: *Physics Procedia*, vol. 69, pp. 48–54 (2015). <https://doi.org/10.1016/j.phpro.2015.07.006>
 42. Peckham, G.D., McNaught, I.J.: Applications of Maxwell-Boltzmann distribution diagrams. *Journal of Chemical Education* 69(7), 554 (1992). DOI: 10.1021/ed069p554.
 43. Andor's fast and sensitive sCMOS cameras. Oxford Instruments. <https://andor.oxinst.com/products/fast-and-sensitive-scmos-cameras>
 44. Vo, N.T., Atwood, R.C., Drakopoulos, M.: Superior techniques for eliminating ring artifacts in X-ray micro-tomography. *Optics Express* 26(22), 28396 (2018). DOI: 10.1364/oe.26.028396.
 45. Palenstijn, W.J., Batenburg, K.J., Sijbers, J.: Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs). *Journal of Structural Biology* 176(2), 250–253 (2011). DOI: 10.1016/j.jsb.2011.07.017.
 46. McGill, R., Tukey, J.W., Larsen, W.A.: Variations of Box Plots. *The American Statistician* 32(1), 12–16 (1978).
 47. Chen, Y., Zhang, Y., Zhang, K., Deng, Y., Wang, S., Zhang, F., Sun, F.: FIRT: Filtered iterative reconstruction technique with information restoration. *Journal of Structural Biology* 195(1), 49–61 (2016). DOI: 10.1016/j.jsb.2016.04.015.
 48. Larson A.C, Dreele R.B.V. General Structure Analysis System (GSAS). Los Alamos National Laboratory Report LAUR 86-748 (2004)
 49. Zhao C, Xiang S, Wang Y, Cai Z, Shen J, Zhou S, Zhao D, Su W, Guo S, Li S. Context-aware network fusing transformer and V-Net for semi-supervised segmentation of 3D left atrium. *Expert Systems with Applications* 214(October 2022), 119105 (2023). <https://doi.org/10.1016/j.eswa.2022.119105>
 50. Chen, Y., Schonlieb, C.B., Lio, P., Leiner, T., Dragotti, P.L., Wang, G., Rueckert, D., Firmin, D., Yang, G.: AI-Based Reconstruction for Fast MRI-A Systematic Review and Meta-Analysis. *Proceedings of the IEEE* 110(2), 224–245 (2022). DOI: 10.1109/JPROC.2022.3141367.
 51. Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D.: A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging* 37(2), 491–503 (2018). DOI: 10.1109/TMI.2017.2760978.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.