

PAPER • OPEN ACCESS

Convolutional neural network based non-iterative reconstruction for accelerating neutron tomography*

To cite this article: Singanallur Venkatakrishnan *et al* 2021 *Mach. Learn.: Sci. Technol.* **2** 025031

View the [article online](#) for updates and enhancements.

You may also like

- [Cost-effective electric vehicle charging infrastructure siting for Delhi](#)
Colin J R Sheppard, Anand R Gopal, Andrew Harris et al.
- [Strength and durability studies of waste nylon cable ties concrete](#)
Virabhadra Tiwari, Harshad R Parate and Nayana N Patil
- [A Novel Coating Technology for Preparation of Cathodes in Li-Ion Batteries](#)
Robert Dominko, Miran Gaberšek, Jernej Drogenik et al.



PAPER

OPEN ACCESS

RECEIVED
15 July 2020REVISED
20 October 2020ACCEPTED FOR PUBLICATION
21 January 2021PUBLISHED
14 April 2021

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Convolutional neural network based non-iterative reconstruction for accelerating neutron tomography*

Singanallur Venkatakrishnan^{1,*} , Amirkoushyar Ziabari¹ , Jacob Hinkle² , Andrew W Needham³,
Jeffrey M Warren⁴  and Hassina Z Bilheux⁵ 

¹ Multimodal Sensor Analytics Group, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America

² Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America

³ Jacobs/NASA Johnson Space Center, Houston, TX 77058, United States of America

⁴ Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America

⁵ Neutron Scattering Division, Neutron Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America

* Author to whom any correspondence should be addressed.

E-mail: venkatakrishv@ornl.gov

Keywords: deep-learning, neutron tomography, sparse view reconstruction

Abstract

Neutron computed tomography (NCT), a 3D non-destructive characterization technique, is carried out at nuclear reactor or spallation neutron source-based user facilities. Because neutrons are not severely attenuated by heavy elements and are sensitive to light elements like hydrogen, neutron radiography and computed tomography offer a complementary contrast to x-ray CT conducted at a synchrotron user facility. However, compared to synchrotron x-ray CT, the acquisition time for an NCT scan can be orders of magnitude higher due to lower source flux, low detector efficiency and the need to collect a large number of projection images for a high-quality reconstruction when using conventional algorithms. As a result of the long scan times for NCT, the number and type of experiments that can be conducted at a user facility is severely restricted. Recently, several deep convolutional neural network (DCNN) based algorithms have been introduced in the context of accelerating CT scans that can enable high quality reconstructions from sparse-view data. In this paper, we introduce DCNN algorithms to obtain high-quality reconstructions from sparse-view and low signal-to-noise ratio NCT data-sets thereby enabling accelerated scans. Our method is based on the supervised learning strategy of training a DCNN to map a low-quality reconstruction from sparse-view data to a higher quality reconstruction. Specifically, we evaluate the performance of two popular DCNN architectures—one based on using patches for training and the other on using the full images for training. We observe that both the DCNN architectures offer improvements in performance over classical multi-layer perceptron as well as conventional CT reconstruction algorithms. Our results illustrate that the DCNN can be a powerful tool to obtain high-quality NCT reconstructions from sparse-view data thereby enabling accelerated NCT scans for increasing user-facility throughput or enabling high-resolution time-resolved NCT scans.

* This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

1. Introduction

Neutron computed tomography (NCT) is a powerful tool for non-destructive 3D characterization of samples relevant to various sciences [1, 2]. In order to conduct an NCT scan, the sample is ‘illuminated’ with an almost parallel beam of neutrons and a collection of projection images are measured by rotating the sample about a single axis. After standard pre-processing, the collection of projection images are processed using an algorithm in order to obtain the 3D reconstruction. NCT is usually conducted at specialized *scientific user facilities* that rely on a nuclear reactor or an accelerator-based pulsed neutron source to generate the neutron beam. While synchrotron-based x-rays are the most widely used sources for CT scans at various user facilities, neutrons can penetrate heavy elements such as metals and are sensitive to certain light elements such as hydrogen, thereby offering a complementary 3D imaging capability to x-rays. However, due to the limitations of the source flux, detection efficiency of the detectors for certain neutron energy ranges [3], or the need to scan large parts, a typical NCT scan take a significantly longer time to measure compared to a typical synchrotron-based x-ray CT scan. For example, an attenuation-based CT scan using current limited flux source and detector technology can take on the order of several hours (at reactors) or even days (at spallation sources, multiple CT scans are collected at different neutron energies simultaneously) to complete in order to produce reconstructions of sufficient quality when conventional tomographic reconstruction algorithms are used. In summary, while NCT is a unique technique offered at neutron user facilities, it is time-consuming to conduct and hence limits the number of samples (overall throughput) that can be measured in a given time frame.

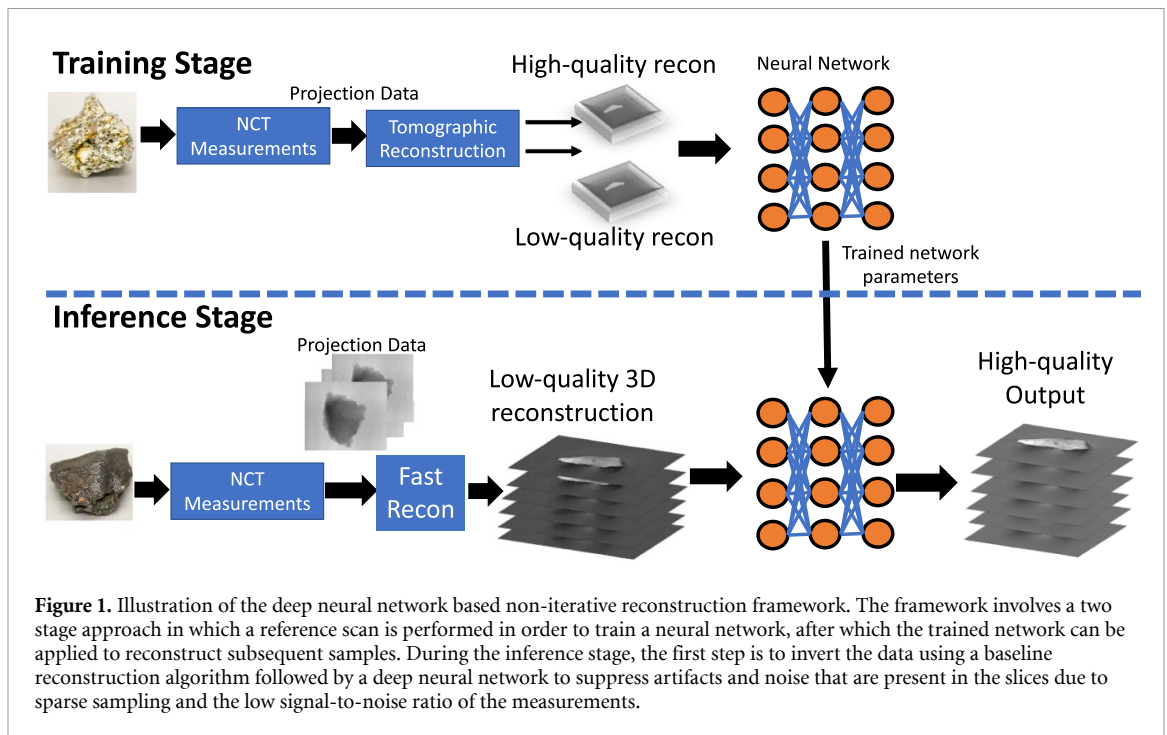
Over the last few decades, the topic of accelerating CT scans for various applications has been widely researched especially from the point of view of designing new algorithms. The most straight-forward approaches to accelerate the acquisitions are to either decrease the exposure time per projection image (resulting in low signal-to-noise ratio (SNR) data) or to decrease the number of projection images (sparse-view data) acquired in the course of a scan. However, the first generation of CT reconstructions relied almost exclusively on the filtered back-projection (FBP) algorithm and its Fourier space variants [4]; these can produce reconstructions with significant artifacts from low SNR or sparse-view data. The advantage of the FBP algorithm is that it can rapidly produce a reconstruction from large data sets encountered at user facilities and is easy to implement. In order to overcome the drawbacks of the FBP method, regularized inversion methods such as model-based image reconstruction (MBIR) [5] algorithms have been developed over the past few decades. MBIR algorithms involve casting the reconstruction as minimizing a cost function that balances a data-fidelity term (that accounts for a physics-based model for the imaging system, and a noise model for the detector) and a regularization term (that is based on a ‘prior model’ for the underlying sample to be imaged). MBIR techniques have demonstrated that it is possible to obtain high-quality CT reconstructions from noisy and/or sparse-view data for a wide range of applications including medical x-ray CT [6], ultrasound CT [7], magnetic resonance imaging (MRI) [8], synchrotron based x-ray CT [9, 10], electron-tomography [11] and NCT [12–14]. Despite of the commercial success of MBIR type techniques in medical imaging [15], they have not been adopted in the user-facilities community for routine experiments. Indeed, a cursory literature survey of research publications from CT beam-lines over the last five years indicates that beam-lines are still collecting a large number of projection images and processing their data using the FBP (and related) algorithms despite of several research publications developing MBIR for user facilities applications including for NCT [9, 13, 16]. We surmise that this is because of a mixture of factors including—familiarity with the FBP algorithm (inertia to change); the high computational cost of running MBIR methods (that may require investing in significant computing resources to obtain reconstructions is reasonable amounts of time) and the challenge of choosing appropriate regularization/prior-model parameters for each experiment. In summary, while new CT reconstruction algorithms have been developed to enable potential acceleration of the measurement process, these techniques have not been widely adopted for NCT.

The past few years have seen an explosion of research on using deep neural-networks (DNNs) for improving CT systems by allowing faster scans (thus reducing dose), and improving the noise-resolution performance [17, 18] by enabling reconstructions from challenging sparse-view and noisy data sets. These methods typically rely on the availability of a sufficient amount of training data; and use this data to train a neural network to produce high-quality 3D reconstructions from sparse-view and noisy measurements. Broadly, the DNN-based algorithms can be classified into iterative and non-iterative techniques. The non-iterative techniques are designed so that they can obtain a reconstruction quality comparable to or better than baseline MBIR algorithms but with the computational complexity that is closer to the FBP algorithm; and hence, can be computed rapidly using modern parallel computing platforms like graphics processing units (GPUs). The iterative reconstruction techniques that rely on DNNs are more robust and are typically

geared towards improving image quality over baseline MBIR methods while the computational complexity may be of the same order. For the purposes of this manuscript, we focus on the non-iterative DNN based reconstruction algorithms because of their simplicity, appealing computational complexity and promising performance in several applications. The neural network—filtered back projection (NN-FBP) [19], a *shallow* neural network, was developed to non-linearly combine several baseline low-quality FBP reconstructions obtained by adjusting the filter parameters and map it to a high-quality reconstruction. In contrast to the NN-FBP technique, a host of DNN methods (that have dramatically larger numbers of parameters) have been developed that effectively map a low quality image/reconstruction to a higher quality target. A U-Net [20] type architecture was developed in [21] to map from a low-quality sparse-view 2D tomographic reconstruction to a high-quality reconstruction. The filtered back projection convolutional neural network (FBPConvNet) [22] also uses a similar architecture to map between a low-quality sparse-view tomographic reconstruction to a full-view reconstruction and was applied to medical x-ray CT and MRI data. The work in [23] makes use of a network to de-noise the acquired low-dose data and map it on to an image that resembles a high-dose acquisition; followed by the use of the FBP algorithm to produce the 3D reconstruction for synchrotron-based x-ray CT (SXCT). The 2.5D artificial intelligence CT (AI-CT) [24–27], a neural network that exploits the ‘3D’ structure of the problem, was developed to approximate a high-quality reconstruction (e.g. MBIR) starting from a noisy FBP reconstruction for medical and industrial x-ray tomography. TomoGAN [28], another neural network architecture for 3D tomographic denoising, uses a generative adversarial network (GAN) in order to obtain a high-quality output from a low-quality input image and was applied to SXCT. Another popular network to obtain high-quality images from sparse-view and noisy data is the mixed-scale dense network (MSD-Net) [29, 30] that is based on learning various dilated convolutional filters that are then combined to map from a low-quality to a high-quality 3D volume. One of the attractive features of MSD-Net compared to most other DNNs, is that it has dramatically fewer parameters to train and has been shown to be effective for SXCT when we do not have large training databases; a scenario typical at other user facility based CT applications. Despite the rapid development of neural network-based algorithms for accelerating CT acquisition, there have been very few efforts adapting them to NCT. One note-worthy work is that of [31], which uses the NN-FBP network to map a sparse-view data set to a higher-quality reconstruction. However, that method is based on a simple multi-layer perceptron and does not exploit the 3D structure of the tomographic reconstructions leaving open the scope to further improve image quality by the use of DNNs. In summary, there are several promising neural network-based approaches developed to accelerate tomographic acquisitions mainly at synchrotron-based user facilities and they have not been adapted to NCT. Furthermore, there has not been a study of the performance of different DNN architectures on the reconstruction quality that can be obtained for such large-scale tomographic applications.

In this paper, we propose to use deep convolutional neural network (DCNN)-based non-iterative algorithms to obtain high-quality reconstructions from sparse-view and noisy NCT data sets in order to accelerate the acquisition and enable more efficient use of the allocated beam time. Our method uses the popular approach of training a neural-network (NN) with pairs of low-quality and high-quality reconstructions from a reference sample, and then applying the trained network to suppress artifacts in subsequent reconstructions obtained from sparse-view data sets. Hence the method effectively involves the design of a data-driven artifact-suppression NN that is specifically adapted to the type of samples being scanned for the experiment. Specifically, we explore two powerful ‘3D’ artifact removal networks—(a) the 2.5D AI-CT [24–26] network that is trained on image/volumetric patches; and (b) the MSD-Net [29, 30] that is trained on whole images—in order to study the impact of different networks on the reconstruction quality. We empirically evaluate (a) how the networks perform when the projection data is significantly sub-sampled and (b) the ability of these trained networks to generalize to different measurement scenarios, i.e. how the image quality is impacted when the network trained for one sub-sampling factor is applied to measurements from a different sub-sampling factor. We apply the DNN-based algorithms to two scenarios: (a) a NCT study of a collection of four meteorite rocks, (b) a time-resolved CT scan of a plant-root system; and demonstrate that it is possible to obtain high-quality reconstructions using the DCNN-based algorithm while being able to accelerate acquisitions time by about a factor of 4 compared to the traditional approaches for our data sets. We emphasize that the possible acceleration factor depends on the sample, the SNR ratio and the end goal of the experiment. We also compare the DCNN-based approach to the simpler NN-FBP[31] algorithm that has been used for NCT, observing that the proposed approaches improve reconstruction quality and generalize better compared to NN-FBP, at the cost of a higher amount of training time for the same amount of training data.

The rest of this paper is organized as follows. In section 2, we describe the proposed approach including the details of the two DCNN architectures used. In section 3, we present extensive results from experimental data followed by conclusions in section 4.

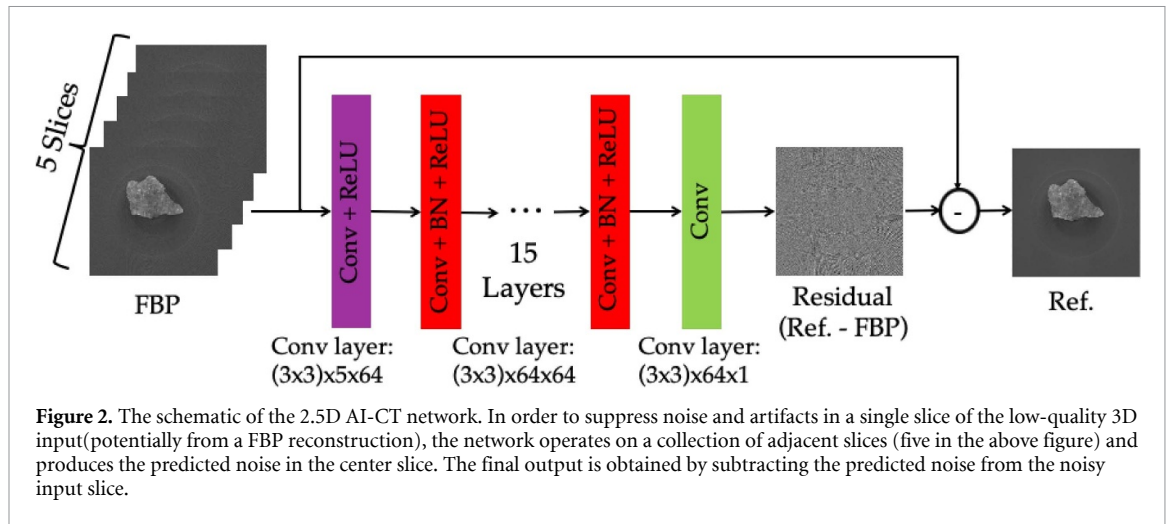


2. Deep-neural networks for accelerated neutron CT

The conventional approach for an NCT experiment at a user-facility involves determining the number of CT scans to be performed, finding the number of projection images to be measured for each scan, and then performing the measurements. The number of projection images and exposure settings for each scan are chosen so that when conventional reconstruction algorithms are applied to the data, they result in a sufficiently high-quality result to meet the end user's need. Often, the number of projection images is determined based on the available beam time and so that the number is close to the Nyquist-rate for the given sample [32]. The Nyquist-rate for high-resolution detectors is large (typically over 1000) and using this criterion results in only a few CT scans being performed in a given amount of time because of the need to collect a large number of projection images for each scan. In contrast to the conventional approach, we propose a measurement framework as shown in figure 1 that is applicable when users want to measure a collection of similar samples or are performing time-resolved NCT studies. Specifically, we propose to make a conventional measurement on a sub-set of the samples, and using this data to obtain pairs of low-quality and high-quality reconstructions which are then used to train a neural network that learns a mapping between the low-quality and high-quality reconstructions (training stage in figure 1). The reference pairs needed to train the neural networks can be obtained either by sub-sampling the acquired projection data or acquiring the data at different dose rates (exposure times). The extent to which the data can be sub-sampled/acquired at a low dose depends on the sample itself and the end-task the user desires from the NCT study. Once the neural network has been trained, the rest of the CT scans can be made at the sub-sampled rate thereby significantly reducing the overall measurement time for the experiment while ensuring similar outcomes for the end user. In summary, the proposed machine-learning driven approach offers a fundamentally different framework to how experiments are planned and performed at NCT facilities by tightly integrating the details of the specific measurement along with algorithms to reduce the overall measurement time leading to a more productive beam time by enabling more samples to be measured.

Specifically, we use the supervised learning approach of training deep neural networks to suppress artifacts that occur when applying standard reconstruction algorithms to sparse-view tomographic data. Once the training data is available, the neural network is trained using the pairs of low-quality/high-quality 3D reconstructions to determine the network parameters by minimizing a loss-function of the form

$$c(\theta; y, x) = \frac{1}{N} \sum_{i=1}^N l(y_i, f_{\theta}(x_i)), \quad (1)$$



with respect to θ , where y_i is the high-quality target, x_i is the noisy input, f_θ represents the neural network with parameters θ , N is the total number of training examples and $l(\cdot)$ is a penalty function on the difference between the output of the neural network and the reference. Once the optimal θ is determined, the reconstructed output from a new data set is determined by

$$\hat{x} \leftarrow f_{\hat{\theta}}(x_{\text{in}}), \quad (2)$$

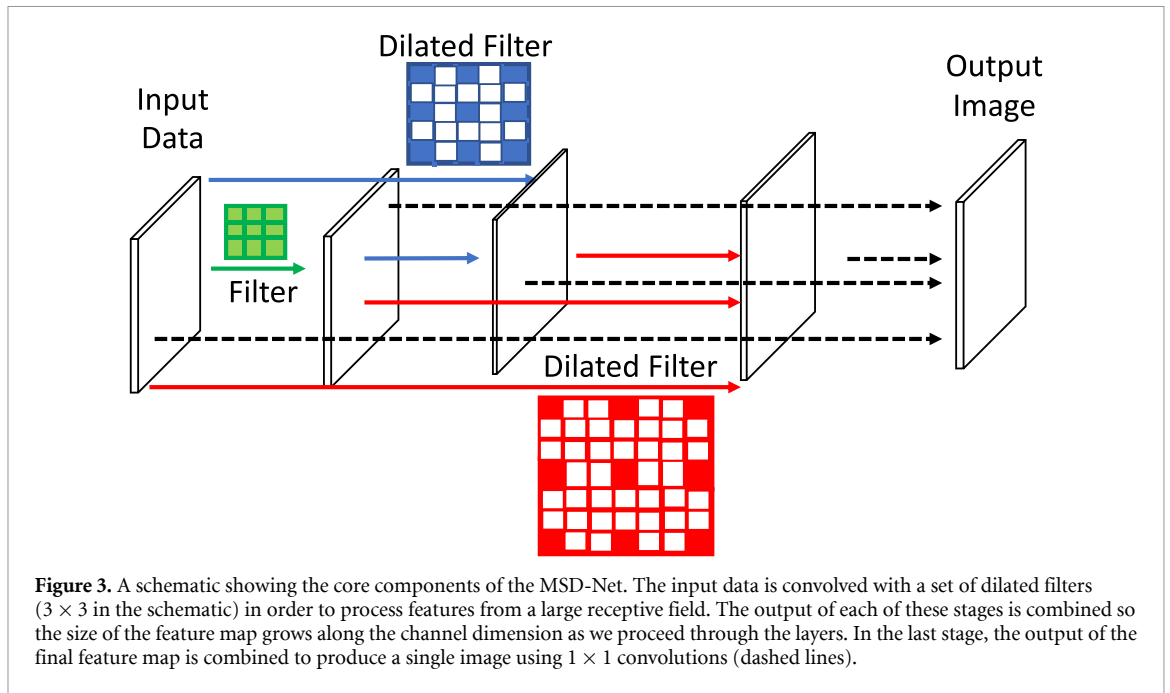
where x_{in} is the low-quality input image and $\hat{\theta}$ is the optimal parameter set from the neural network training.

Contrary to instruments designed for very specific applications like medical x-ray CT, MRI, medical ultrasound, etc user facilities scientific tomography instruments are used to scan a wide variety of samples. As a result there are not sufficient ‘previous’ scans that can be used to train a machine learning algorithm and merely apply the trained model to every new sample scanned at these facilities to produce a reliable reconstruction. Instead, we foresee DNN based methods being useful when the users are interested in scanning a collection of similar samples, or to measure gradual changes in a sample using time-resolved CT. In such a context, the core method involves scanning one representative sample and preparing a pair of low-quality and high-quality reconstructions which can potentially take a long duration. Once such a pair(s) of 3D volumes are available, we can train a ‘3D’ DCNN to suppress artifacts and be able to map from the low-quality to high-quality output. Because we may only have a single low-quality/high-quality pair of 3D reconstructions available in a typical user-facility based CT application, it is not feasible to train a fully 3D CNN that would exploit the non-local correlations across the entire volume. Such a network would require several examples of 3D volumetric pairs, along with a non-local network with a large number of parameters. Instead, we use a technique of training ‘2.5D’ deep neural networks that are trained to remove noise and artifacts from a single slice of the 3D volume by using a few adjacent slices in 3D (see figure 2). The trained network can then be applied to the rest of the sample which are acquired using an accelerated sparse-view scanning protocol. Thus the overall time required to scan the collection will be significantly decreased especially if the users are interested in scanning a large collection of samples. Alternately, in the case of time-resolved NCT, this can translate to users being able to rapidly acquire data and see a high-quality reconstruction in near-real time because of DCNNs can be easily implemented using commodity GPU platforms.

In this paper we focus on using two popular deep convolutional neural network architectures for f_θ in equation (1). The first is the 2.5D AI-CT network [24] and the second is the mixed-scale dense network (MSD-Net) [30]. We choose these two networks for NCT because they have been demonstrated to perform well for other applications, have a smaller number of training parameters compared to large networks like the U-Net, generative adversarial networks (GAN), etc and hence, are easier to train. Next, we summarize the key components of these two networks.

2.1. 2.5D AI-CT

The 2.5D AI-CT network that we use in this work, was developed by Ziabari *et al* [24–26] inspired by deep residual denoiser [33]. This deep neural network is designed to learn the non-linear mapping between pairs of low/high-quality reconstruction of 3D volumes by exploiting the correlations that exist in volumetric data. The term 2.5D was used in [24] because unlike traditional image de-noisers based on deep neural networks,



this network considers neighboring slices from a 3D volume in-order to de-noise a single slice. The experiments in [24] using this architecture showed that this type of a network had significantly reduced computational complexity compared to a fully 3D convolutional neural network while enabling similar performance.

Figure 2 shows a schematic of the 2.5D AI-CT network. The aim of this network is to find a non-linear mapping that transforms a low-quality input volume (say from the FBP algorithm) into an accurate approximation of the higher quality reconstruction of the volume by processing the input volume one chunk (collection of slices) at a time. In order to train the network, a higher quality CT reconstruction for training the network can be obtained from a FBP reconstruction of data measured using a longer scan time (high SNR), or a dense set of projection images, or by using a sophisticated reconstruction algorithm such as model-based iterative reconstruction (MBIR) on a sparse-view data set.

The 2.5D AI-CT network we propose to use in this work consists of 17 layers. The convolution kernel in each layer has the form $(3 \times 3) \times N_i \times N_o$, corresponding to a (3×3) convolution kernel with N_i input and N_o output channels. The number of input channels in the first layer was set to 5, which means each slice and 4 neighboring slices are the input to the network. The first layer applies 64 convolutional filter kernels followed by rectified linear units (ReLU) to form 64 output channels. Therefore, layer one's kernel has the form $(3 \times 3) \times 5 \times 64$. Layers 2 to 16 apply a kernel of size $(3 \times 3) \times 64 \times 64$ and generate 64 output channels from the 64 input channels. Each convolution is followed by a batch normalization and ReLU [34, 35]. The last layer applies a $(3 \times 3) \times 64 \times 1$ kernel to generate a single output residual image. For this network architecture, the 2.5D AI-CT network has a total of 559 361 trainable parameters. In conclusion, the y_i in equation (1) is a vector of representing the residual output error for the i th training sample, x_i is a vector representing the noisy input accounting for the corresponding input slice along with a set of adjacent slices and θ corresponds to all the trainable parameters of the 2.5D AI-CT network.

2.2. MSD-Net

The MSD-Net, developed by Pelt and Sethian [30], is an image-to-image convolutional network that has exhibited good performance across a variety of imaging tasks including noise and artifact suppression for accelerated synchrotron x-ray CT [29]. The MSD-Net architecture repeatedly filters the inputs using dilated convolutions, i.e. filters that have a large receptive field and hence are able to capture multi-scale and non-local structures effectively (see figure 3). The size of the 'feature maps' through the network is the same as that of the input image along the row and column dimension, with the size growing along the channel dimension. In each layer, the input is filtered by a set of dilated convolutions and the output is concatenated with its input and fed to the next layer, as in a conventional DenseNet [36]. Thus, the tuneable parameters of this network are only the coefficients of the dilated convolutions and the number of layers. The central advantage of the MSD-Net is that it is easy to train since it has a very small number of parameters compared to several popular deep learning architectures with similar receptive fields. For example, an MSD-Net with

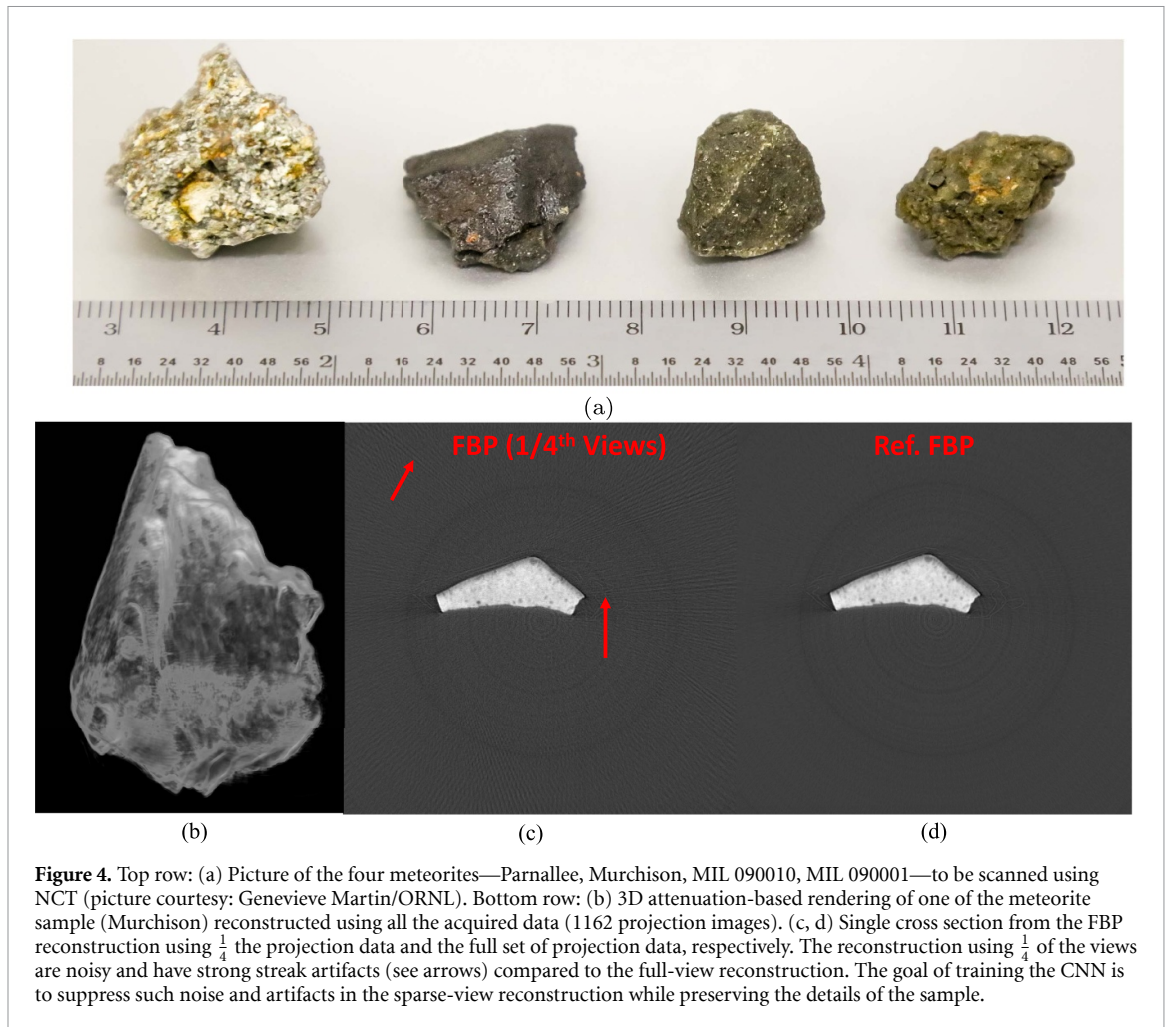
100 layers, 10 dilations and 3×3 convolutions has approximately 50000 parameters, about an order of magnitude smaller than the 2.5D AI-CT network. In contrast to the 2.5D AI-CT network, the MSD-Net can be trained on the entire image thereby allowing the network to learn global features like streaks that occur due to sparse-view reconstructions. We use standard rotation and flipping operations in order to augment the limited training data available in the NCT application. Similarly to the 2.5D AI-CT network, we set the number of input slices to 5, which means each slice and 4 neighboring slices are the input to the network. In conclusion, the y_i in equation (1) is a vector representing the entire target image, x_i is a vector representing the noisy input for the corresponding slice along with a set of adjacent slices and θ corresponds to all the parameters of the MSD-Net.

3. Experimental results

We compare the performance of the DCNN-based algorithms for two sets of NCT experiments. In the first scenario, the goal is to image a collection of similar samples (meteorite rocks) and to do so by reducing the total acquisition time. In the second scenario, the goal is to improve the reconstruction quality of time-resolved NCT data sets which have been acquired using a sparse set of projection images in order to visualize the changes in the sample as a function of time. We compare the results of the proposed DCNN techniques to a standard post-processing algorithm based on total-variation (TV) denoising [37–39] as well as a simpler neural network approach, the NN-FBP [19], that has been proposed for accelerated neutron tomography. The regularization parameter of the TV algorithm is manually adjusted so that the residual noise in the final reconstruction is similar to those of the DCNN approaches. Specifically, we adjust the regularization values so that the standard deviation of the noise in a uniform region of the sample is similar to that from the deep neural network approach. The NN-FBP method is trained in a ‘point-wise’ fashion, i.e. x_i in equation (1) corresponds to a vector of reconstructed voxel values using a collection of filter parameters of the FBP algorithm. For this paper, we set the number of hidden layers of NN-FBP to 4. The number of pixels used to train the NN-FBP algorithm is set to 100 000. In contrast to NN-FBP, the DCNN methods take into account a large neighborhood around the voxel to be de-noised because of the presence of convolution operators. For each of the experiments, the data was pre-processed using a median filter to suppress ‘gamma hits’, followed by tilt-axis correction and a stripe-removal filter was applied to the data in order to suppress ring artifacts in the reconstruction. All pre-processing routines were performed using the TomoPy tool-box [40] and the reconstructions were performed using the ASTRA [41, 42] and pyMBIR library [43]. For the NN-FBP and MSD-Net we use the publicly available software from the first author’s GitHub page (<https://github.com/dmpelt>). The images in the results section are best viewed on a screen with the ability to zoom in to best evaluate the qualitative improvements in image performance.

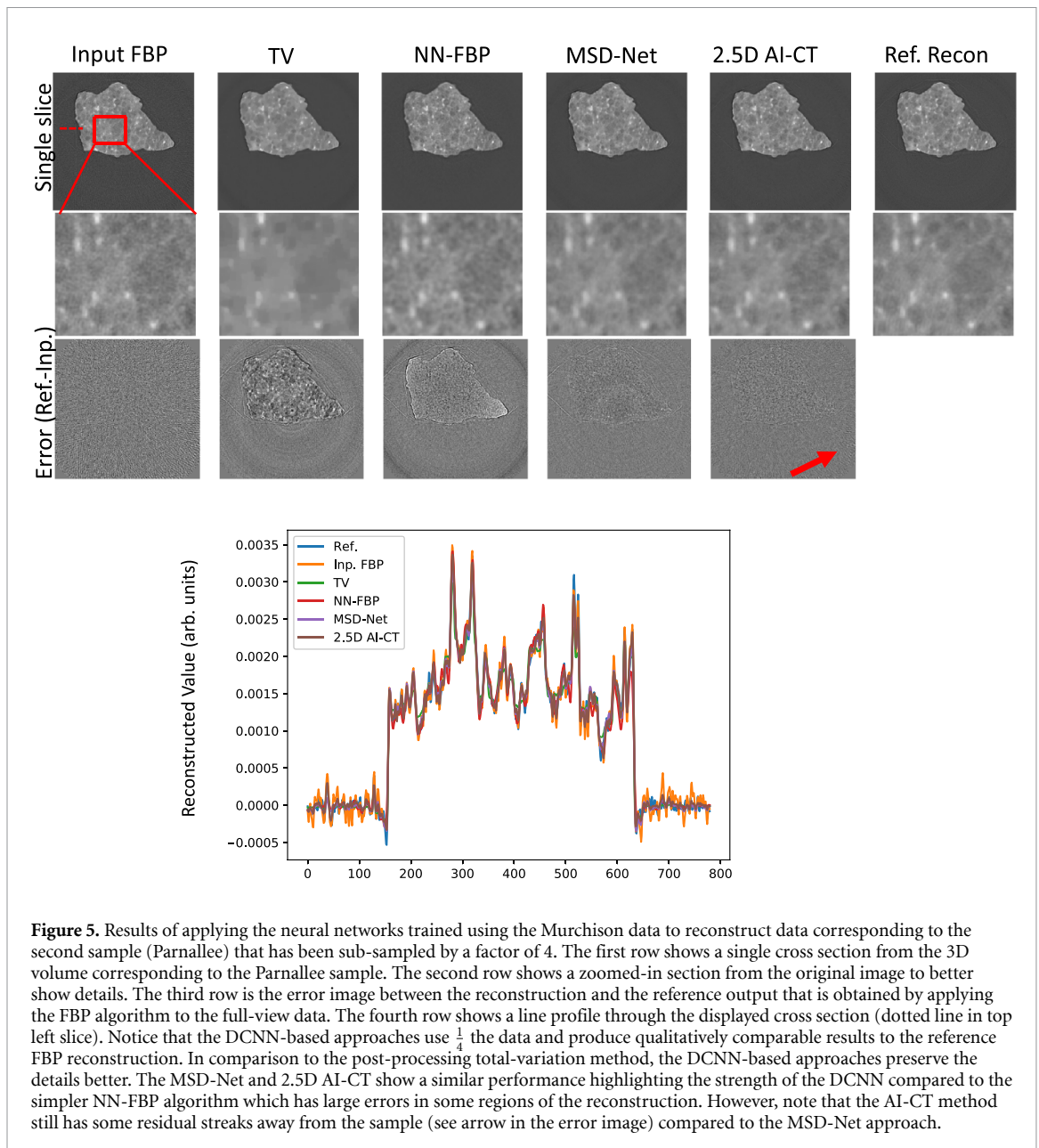
In the case of 2.5D AI-CT, training on large volumetric data sets encountered in user-facility applications (of the order of $1000 \times 1000 \times 1000$ voxels or larger) can be computationally expensive if we provide the entire image as input to the network. Furthermore, such strategy results in a severe scarcity of training data since we typically only have a single reference volume available and a large number of parameters to train. Therefore, we divide the input images to smaller patches and train the network to learn the features from those patches by typically augmenting the training set using standard techniques such as flipping and rotation. In this study, we use a patch size of 256×256 with a stride of 64. In addition, in certain samples measured at user-facilities, the reconstruction can be unbalanced, i.e. the object only fills a fraction of the entire volume and most of the voxels are from the background. Therefore, several generated patches will include only background pixels and have no information about the objects of interest. Such patches may decrease the training accuracy of the 2.5D AI-CT network; and hence, in order to avoid them, we remove all the patches that have a mean value smaller than a threshold value determined by using the mean value of all the patches. The generated patches are then randomly augmented during the training process using rotation/flip operations in order to increase the number of training samples for the neural network. We use the ADAM algorithm [44] with a learning rate of 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.99$. The learning rate is decreased by a factor of 2X every 70 epochs, or if the validation loss increases for three consecutive epochs, whichever happens first.

For training the MSD-Net we use a 100 layer network with 10 dilations with filters of size 3×3 . In contrast to the 2.5D AI-CT network, we do not have to generate or prune patches for training the MSD-Net since the entire high-resolution image is provided as an input. However, we augment the training data sets by rotating and flipping the image chunks after normalizing the inputs to have zero mean and unit standard deviation. The network is trained using a batch size of 1, by using the ADAM method [44] with parameters $\beta_1 = 0.9$, $\beta_2 = 0.99$ and learning rate set to 0.001.



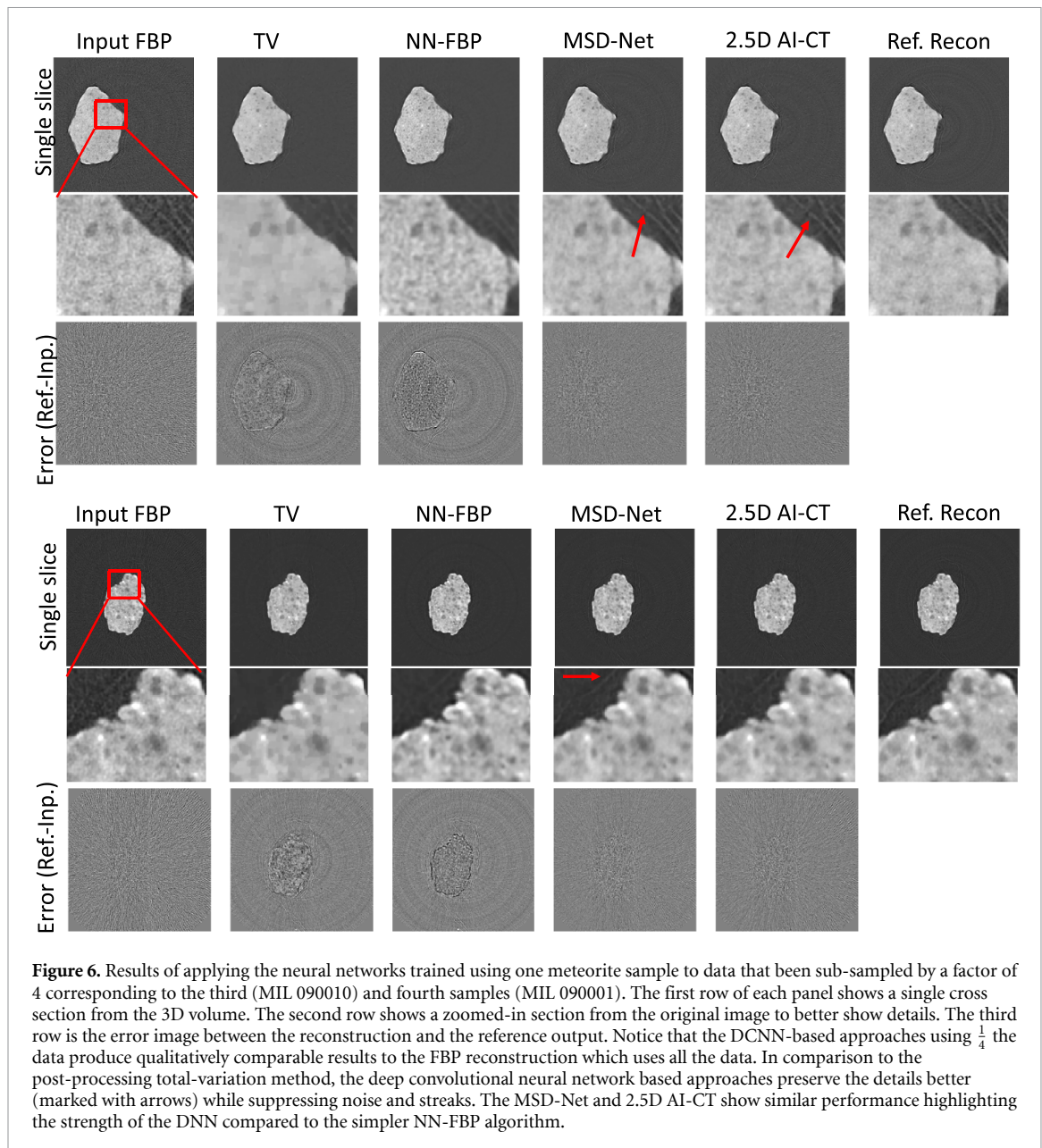
3.1. Collection of similar samples

A collection of four meteorite samples—three carbonaceous chondrites (two Miller Range meteorites MIL090001 and MIL090010, and Murchison) and one ordinary chondrite (Parnallee)—were measured at the CG1-D beam-line at the High Flux Isotope Reactor (HFIR) at ORNL over the course of two sessions separated by a couple of weeks (see figure 4). For the meteorite CT scans, the aperture was set to 1.6 mm (for the Murchison and Parnallee meteorites) and 8.2 mm (for the MIL 090010 and 090001 meteorites), yielding a collimation ratio of 400 and 800, respectively. Each CT scan was performed using a 16-bit Andor iKon-L 936 charge-couple device (CCD) model with a 2048 pixels \times 2048 pixels chip, equipped with a 100 μ m thick $^6\text{LiF/ZnS}$ scintillator. For the MIL meteorites, a total of 1162 projections were measured by rotating the sample from 0 to 360 degrees, with an angular step of 0.31 degrees (to ensure unique projections were acquired after reaching 180 degrees). The acquisition time per projection image was approximately 63 s, resulting in a total scan time of nearly 23 h including the measurement of open-beam, dark-current images for normalization, and counting for the rotation stage movements and the transfer of each radiograph from the CCD to the data server via USB 2.0. For the Murchison and the Parnallee meteorites, the CT scan was also performed over an angular range of 360 degrees, with a rotation angle of 0.31 degrees and each radiograph took 40 s to measure. By opening up the aperture, the CT scan was measured in about 14 h. The difference in acquisition parameters was motivated by the available time at the beam line, which is often the case at neutron user facilities and reinforces the importance of the availability of advanced reconstruction algorithms. In each case, the width of the sample on the detector was about 800 pixels in the horizontal direction, for which the Nyquist view sampling rate is $800 \times \pi/2 \approx 1256$ projection images. Thus the acquired number of projection images is close to the Nyquist rate for this sample. In order to train the neural networks, we use the data from one meteorite (Murchison) and retro-actively sub-sample it in order to obtain the sparse-view data set. We cropped the original data to use only 1280 pixels along the horizontal dimension for reconstruction. The sparse-view and full-view data are then reconstructed using the FBP algorithm and serve as the training pairs for the neural networks (see figure 4). The parameters for the FBP algorithm are adjusted manually to attain a reasonable visual quality of the reconstruction. We use 400 slices



to train the networks and 112 slices for validation. Once the networks are trained, we use them to reconstruct the data from the other three meteorites by retro-actively sub-sampling the acquired data. In each case, we use the full-set of data in order to produce the reference FBP reconstruction that serves as the ‘ground-truth’ to which we compare the output of our proposed approach. In order to evaluate the output of the different algorithms we use the visual quality from a representative cross-section and the normalized root-mean squared error (NRMSE) to gauge quantitative trends across different algorithms.

Our first set of experiments are done in order to evaluate the performance of the different algorithms by training them to suppress artifacts from sparse-view FBP reconstructions obtained from different sub-sampling factors. In each case, we train the networks to map between an FBP reconstruction using a fraction of the data to the FBP from the full set of projection images. Once the networks have been trained, we evaluate their performance on multiple test data-sets that have been retro-actively sub-sampled at the same rate as the training data. We choose sub-sampling factors of 2, 4, 8 and 16 in order to evaluate the algorithms. Figures 5 and 6 show a single cross-section from the 3D reconstruction of the meteorites using different algorithms for a sub-sampling factor of 4. Notice that the MSD-Net and the 2.5D AI-CT are able to significantly suppress streak artifacts and noise which are present in the FBP reconstruction of the sub-sampled data while preserving the features in the reconstruction. In comparison to the post-processing TV method (which is ‘waxy’ looking as also evidenced in the line profile in figure 5), the deep neural network approaches are able to better preserve the texture and details while suppressing artifacts in the



reconstructions. This type of artifact suppression is also highlighted by the error images of the difference between the output of the network and the reference FBP that was obtained by using all the data from the scans. We were not able to observe any significant visual differences between the MSD-Net and the 2.5D AI-CT output in the central region containing the meteorites despite of the limited availability of training data. However, the MSD-Net was better able to suppress some of the larger streaks (see arrow in figure 5) which we attribute to the fact that the AI-CT is trained in a patch-wise manner as compared to the MSD-Net which is trained using entire images. We also observed that the DCNN methods (MSD-Net and AI-CT) improved the reconstructions compared to the simpler NN-FBP algorithm (see also table 1). While the NN-FBP method resulted in a higher NRMSE than the post-processing TV algorithm, we visually observe that it preserves the texture better than the TV-based method. We carried out similar experiments for sub-sampling factors of 8 and 16 and observed similar trends (see figures 7 and 8). However, we also noticed that all the methods started suppressing some of the finer details (marked with arrows in figures 7 and 8) in the samples compared to the ground-truth reconstructions (i.e. FBP applied to the full set of projection images).

Next, we evaluate the generalization performance of the trained neural network algorithms on the same data sets but by using different sub-sampling factor. In each case, we use the networks that have been trained to effectively reconstruct data that has been sub-sampled by a factor of 4. This was done in order to evaluate the performance of the neural network based approaches when the test data does not strictly adhere to the

Table 1. Comparison of the normalized root mean squared error (NRMSE) as a percentage of maximum value of the reconstruction with respect to the reference reconstruction for various scenarios for the three test samples. In each case, the neural networks were trained using the same sub-sampling factor as that used for the test data sets.

Sub-samp.	Paranallee				
	Inp.	TV	NN-FBP	MSD-Net	AI-CT
16	2.67	1.04	1.20	0.91	0.93
8	1.59	0.81	0.93	0.71	0.71
4	1.05	0.72	0.79	0.56	0.58
2	0.51	0.58	0.69	0.51	0.49

Sub-samp.	MIL090010				
	Inp.	TV	NN-FBP	MSD-Net	AI-CT
16	2.95	0.93	1.41	0.90	0.91
8	1.88	0.81	0.92	0.81	.80
4	1.22	0.74	0.82	0.68	.71
2	0.64	0.63	0.68	0.61	0.57

Sub-samp.	MIL090001				
	Inp.	TV	NN-FBP	MSD-Net	AI-CT
16	2.58	0.85	1.36	0.73	0.76
8	1.56	0.64	0.68	0.64	0.64
4	1.00	0.57	0.61	0.53	0.61
2	0.49	0.47	0.53	0.46	0.45

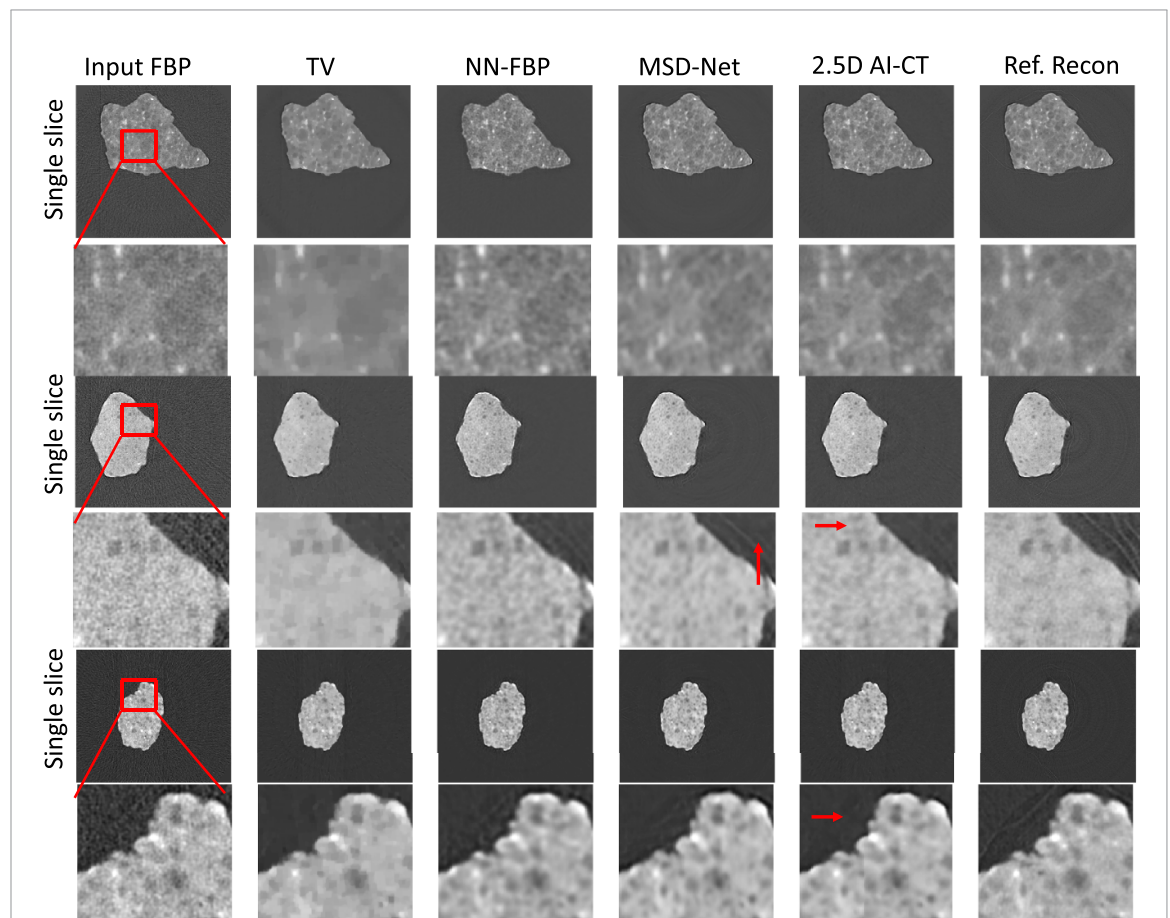
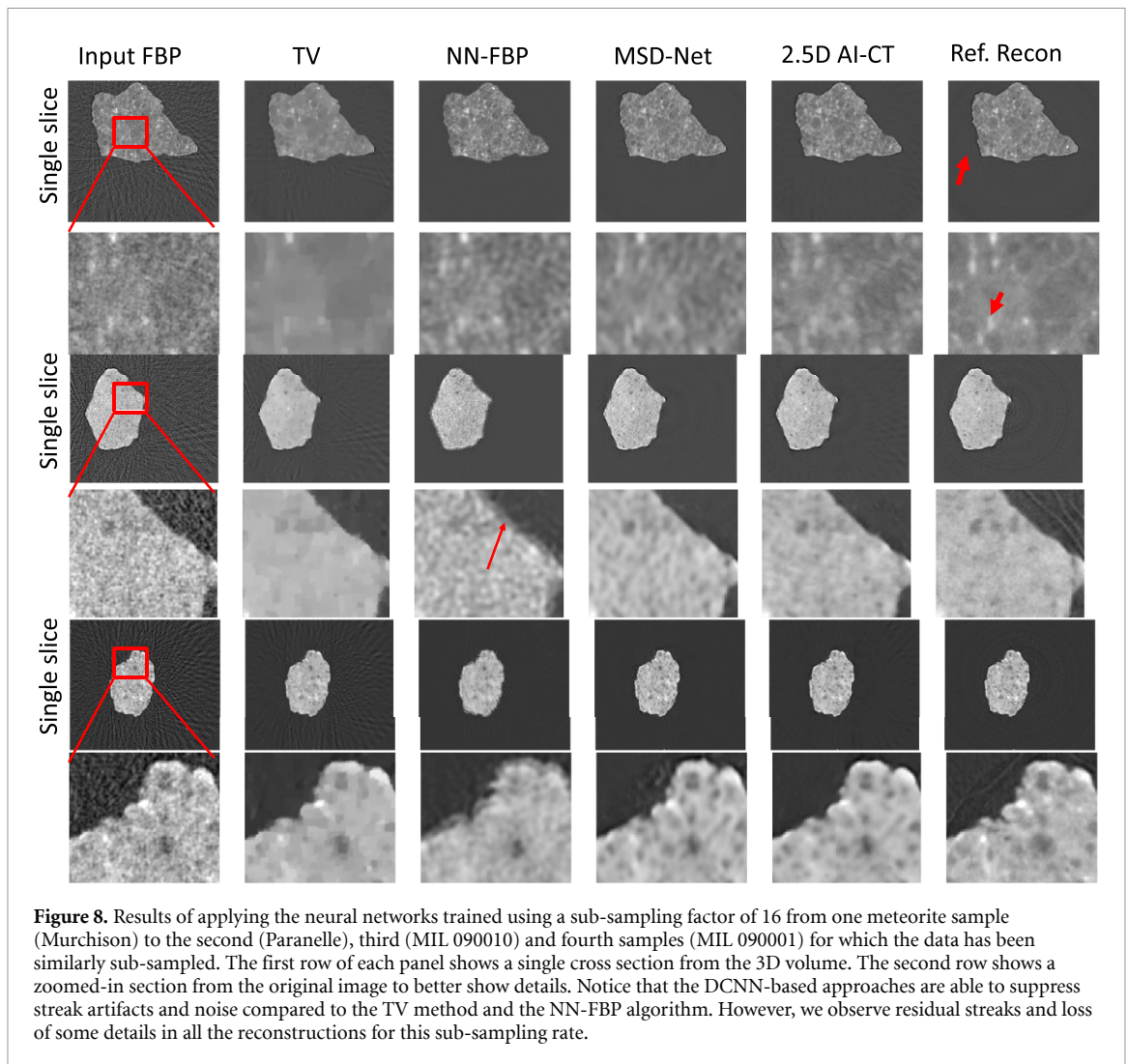
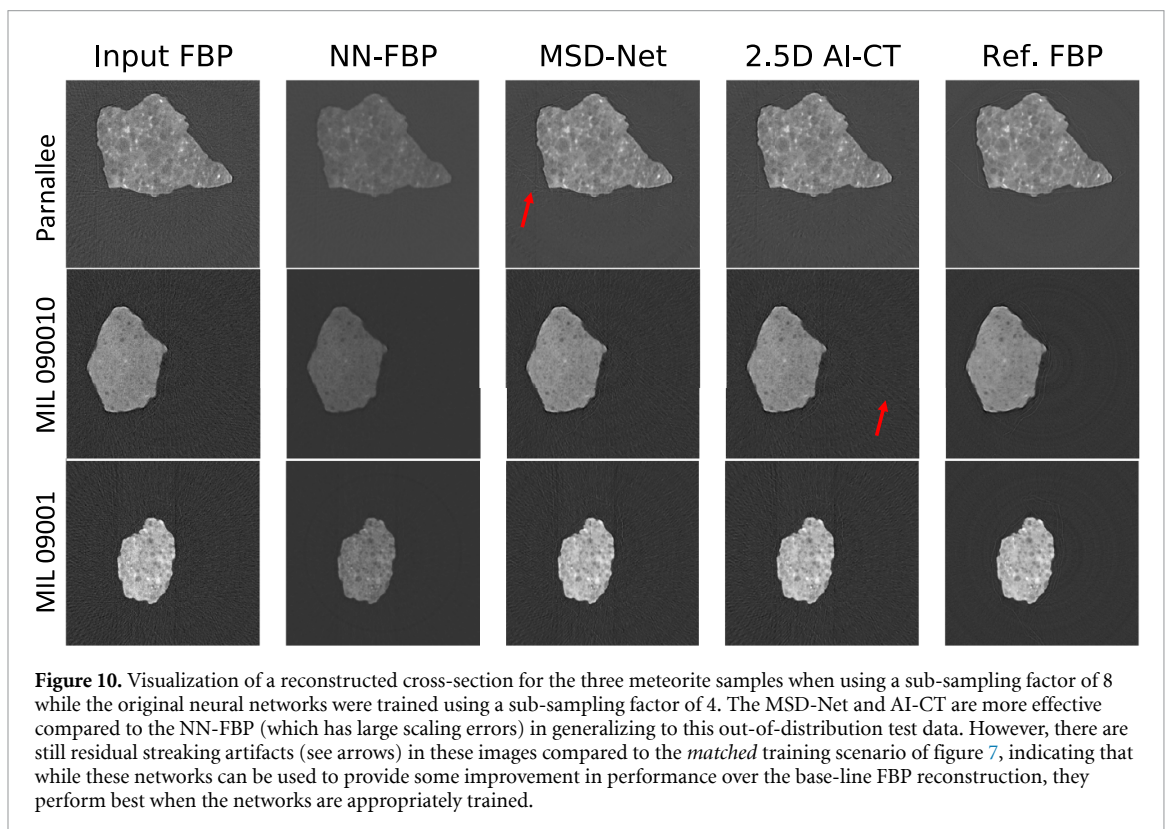
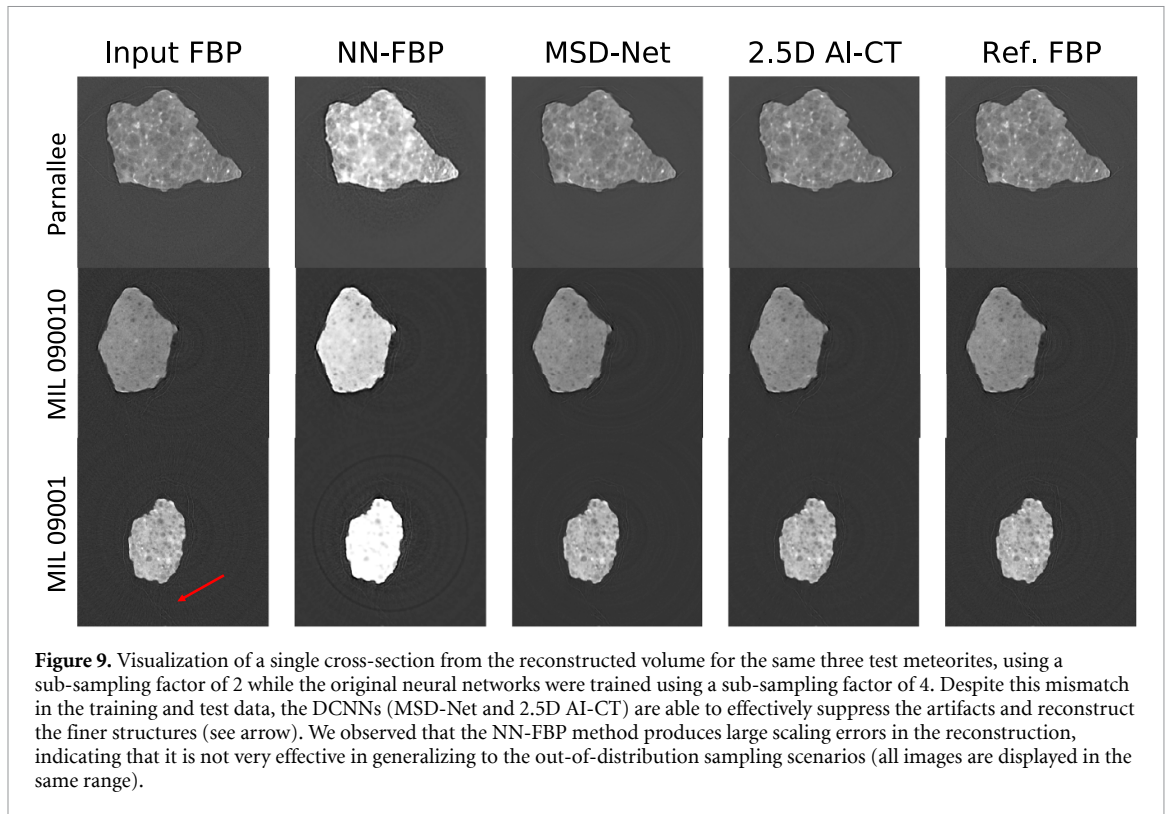


Figure 7. Results of applying the trained neural networks using a sub-sampling factor of 8 from one meteorite sample (Murchison) to the second (Paranalle), third (MIL 090010) and fourth (MIL 090001) meteorite for which the data has been similarly sub-sampled. The first row of each panel shows a single cross section from the 3D volume. The second row of each panel shows a zoomed-in section from the original image to better show details. Notice that the DCNN-based approaches produce qualitatively superior results to the other approaches by suppressing noise and streak artifacts. Specifically, the MSD-Net and 2.5D AI-CT show a similar performance highlighting the strength of the DNN compared to the simpler NN-FBP algorithm. However, some of the finer structures in the sample are smoothed out at due to the large sub-sampling factor used (see arrows).



accelerated scanning protocols of the training set. Figures 9 and 10 show a single reconstructed cross section from the three meteorite samples when using sub-sampling factors of 2 and 8. We observe that while the MSD-Net and AI-CT approaches are able to suppress the streaks and noise in the reconstructions, especially when the sub-sampling factor is smaller than the training set. They are not as effective (see table 2) when the sub-sampling factors are higher than what the original networks were trained for. This trend can also be inferred from the quantitative results in table 2 and by comparing them with the results in table 1. Finally, we also observed that the DCNN-based approaches are significantly superior to the NN-FBP method, which produces strong-artifacts (exacerbated due to scaling errors) when applied to out-of-distribution data.

Finally, we provide details of the approximate run time of the different methods presented here. We emphasize that the goal is not to make a precise comparison of the speed of each method, because each of the algorithms/software are not optimized for a given compute platform (for example: the publicly available implementation of MSD-Net does not run on multiple GPUs and loads the data one set of slices at a time from the disk). Our goal is to provide the reader with an estimate of the time required to train and test the data for a particular compute platform. Table 3 shows the run time for the different neural network-based approaches for the training scenario where we use a sub-sampling factor of 4. Notice that the NN-FBP algorithm can be rapidly trained compared to the DCNN approaches because of the simplicity of its architecture and the small number of trainable parameters. In contrast, the DCNN methods (MSD-Net and AI-CT) took approximately 1 day to train which is comparable to the measurement time for the training data. However, the inference time of all the methods for the reconstruction of a single meteorite is dramatically faster than the training time. Thus the DCNN methods trade off performance, and generalization with training and inference speed compared to the NN-FBP algorithm. While we have not used the MBIR method as a comparison for this data set because of the availability of a high-quality FBP reference, our implementation of MBIR [12] for a $512 \times 1280 \times 1280$ size volume took approximately 6.7 h using 150 iterations, which is significantly higher than those of the neural network based approaches if we



only compare the inference time. While there have been several efforts in accelerating MBIR techniques using modern parallel computing platforms [45–47], their computational complexity (number of forward and back-projections, terms associated with regularization parameters etc) is much higher than the proposed neural network approaches (which consists of a single FBP operation followed by a forward-pass through the network) when considering a sequence of similar samples to be measured.

Table 2. Comparison of the normalized root mean squared error (NRMSE) as a percentage of maximum value of the reconstruction with respect to the reference reconstruction for various scenarios for the three test samples. In each case, the neural networks were trained using one-fourth of the total number of views and then applied to different sparse sampling scenarios for the test data sets in order to evaluate generalization of the trained networks to different measurement scenarios.

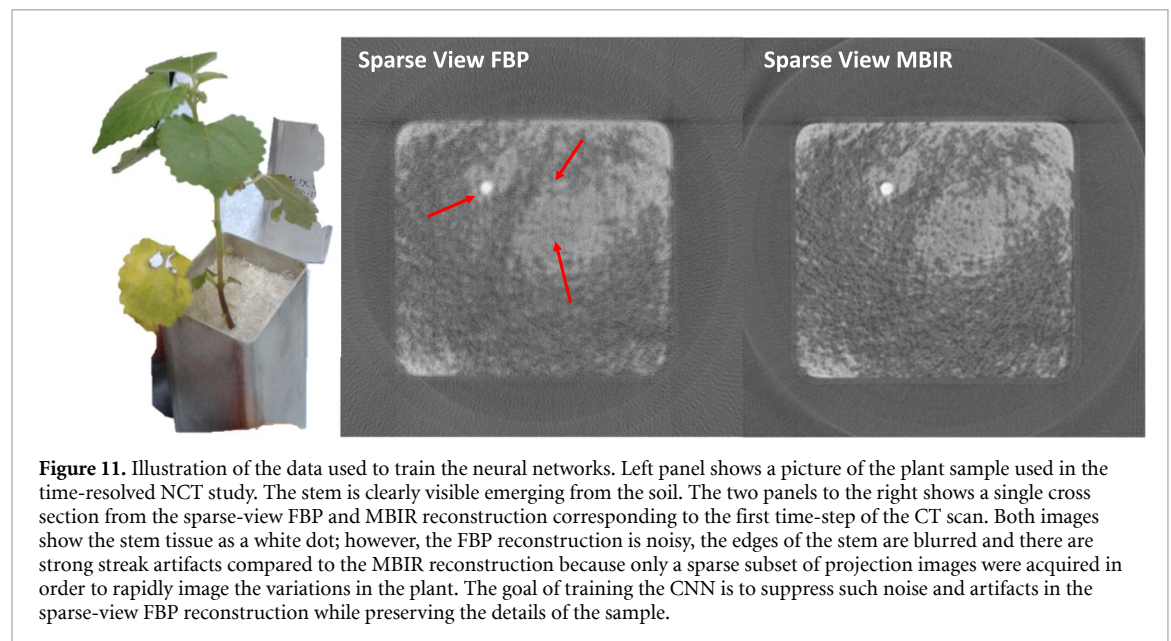
Parnallee				
Sub-samp.	Inp.	NN-FBP	MSD-Net	AI-CT
16	2.67	4.39	1.85	1.95
8	1.59	3.19	0.77	0.87
2	0.51	5.43	0.47	0.47
MIL 090010				
Sub-samp.	Inp.	NN-FBP	MSD-Net	AI-CT
16	2.96	4.27	2.22	2.30
8	1.88	3.03	1.21	1.22
2	0.64	4.97	0.55	0.52
MIL 090001				
Sub-samp.	Inp.	NN-FBP	MSD-Net	AI-CT
16	2.58	4.50	2.05	2.13
8	1.56	2.96	1.01	1.08
2	0.49	3.81	0.42	0.40

Table 3. Approximate time for training and inference for the different algorithms on volumes of size $512 \times 1280 \times 1280$ voxels.

Algorithm	Train time (h)	Compute	Test time (min)	Compute
NN-FBP	4	CPU only ^a	7	CPU only ^a
MSD-Net	21	CPU + 1 GPU ^a	21	CPU + 1 GPU ^a
AI-CT	24	CPU + 4 GPU ^b	7	CPU + 1 GPU ^a

^a Intel(R) Xeon(R) CPUE5-2640 v4 @ 2.40 GHz, 40 core, 2X Nvidia P100, 24 GB GPU.

^b 2X 20-Core Intel Xeon E5-2698 v4 2.2 GHz, 6X Tesla V100, 16 GB GPU.



3.2. Time-resolved neutron CT

The goal of a time-resolved CT scan is to study changes in the sample in 3D as a function of time. Because the sample is changing, typically only a sparse-set of CT measurements can be made corresponding to a ‘time-step’ before the reconstructions start suffering from significant blur. The state-of-the-art algorithms for time-resolved CT reconstruction are typically MBIR methods which exploit the spatio-temporal correlations in the data but can be computationally very expensive [10]. In the context of NCT, there has been recent

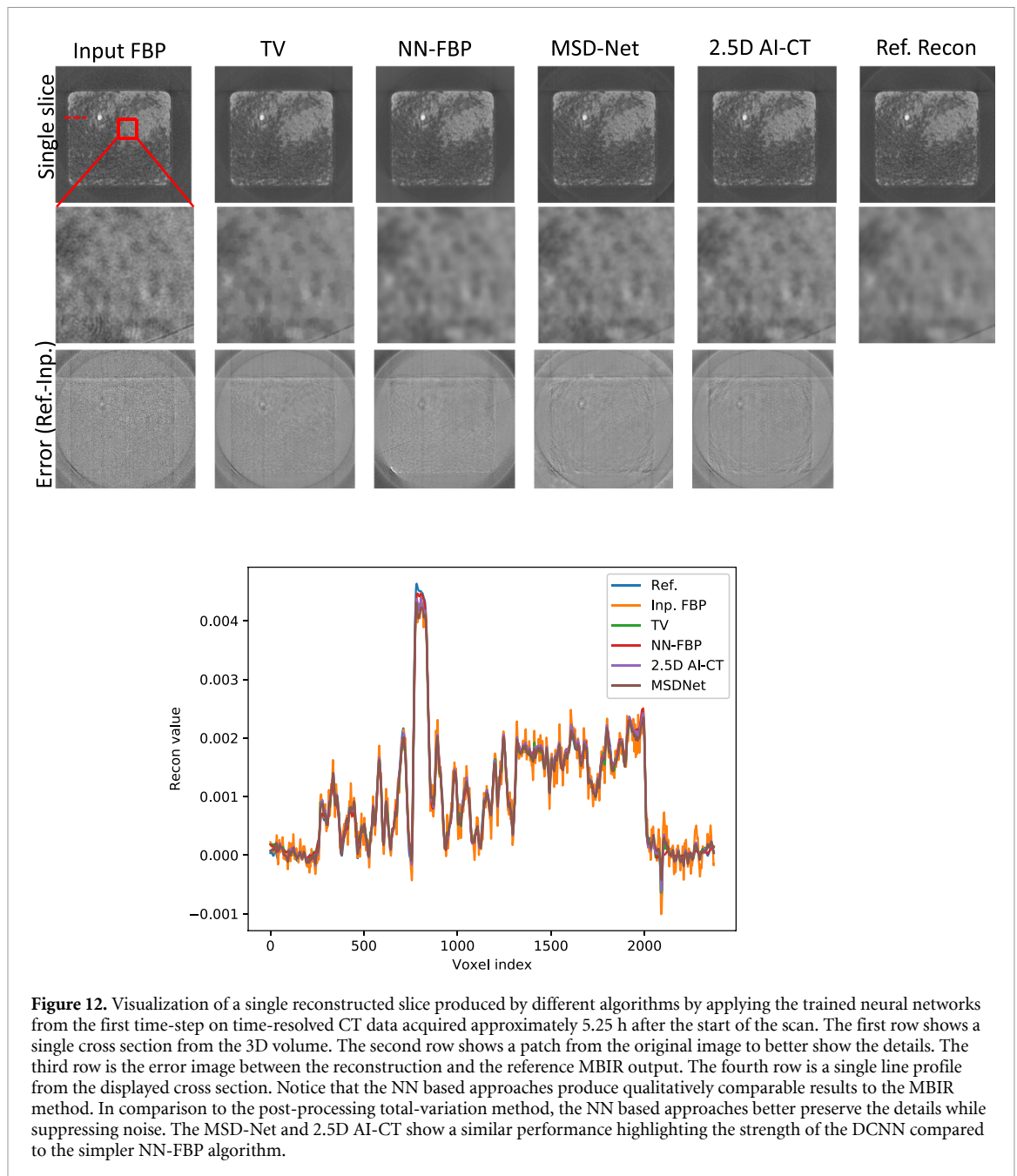
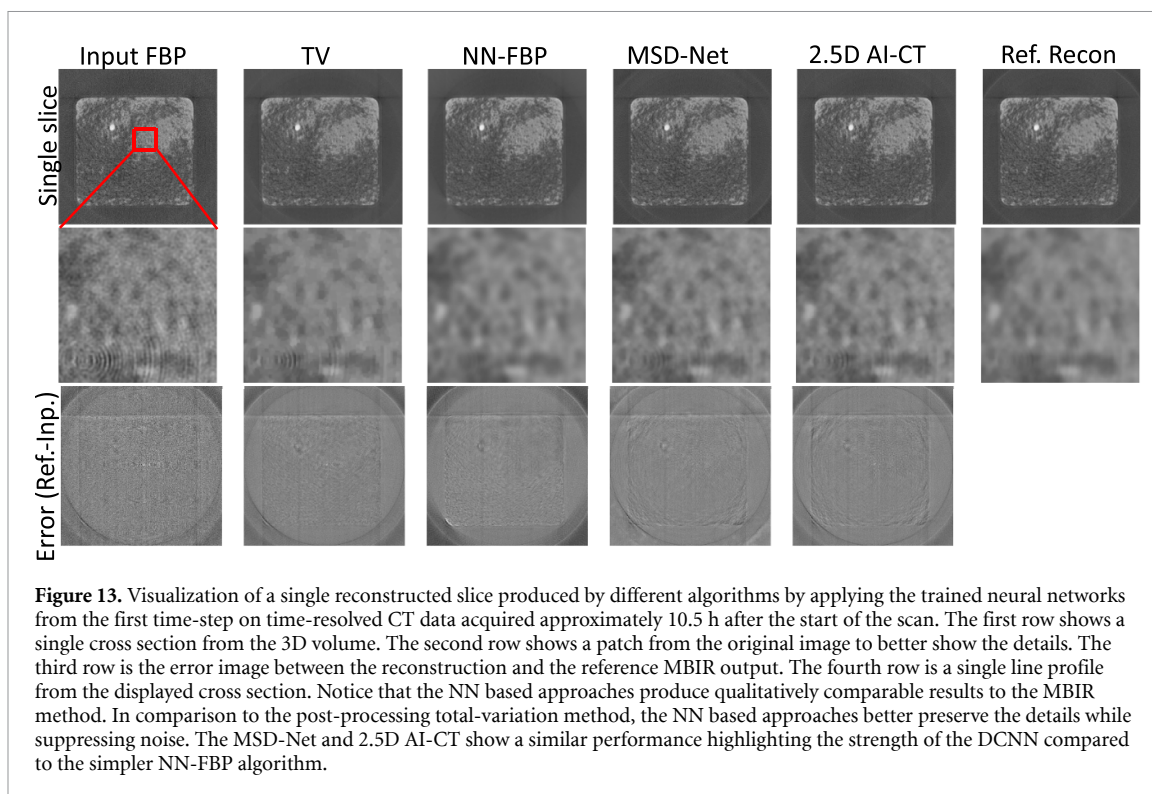


Figure 12. Visualization of a single reconstructed slice produced by different algorithms by applying the trained neural networks from the first time-step on time-resolved CT data acquired approximately 5.25 h after the start of the scan. The first row shows a single cross section from the 3D volume. The second row shows a patch from the original image to better show the details. The third row is the error image between the reconstruction and the reference MBIR output. The fourth row is a single line profile from the displayed cross section. Notice that the NN based approaches produce qualitatively comparable results to the MBIR method. In comparison to the post-processing total-variation method, the NN based approaches better preserve the details while suppressing noise. The MSD-Net and 2.5D AI-CT show a similar performance highlighting the strength of the DCNN compared to the simpler NN-FBP algorithm.

interest to improve the resolution of time-resolved CT mainly by using new detectors and sources [48, 49] with standard reconstruction algorithms. Here, we instead propose to acquire the sparse projection data and use it from one (or a small number) of the time-stepped CT scans and obtain a high-quality reference reconstruction by using an MBIR method. We can then train a DCNN to map between the low-quality FBP reconstruction from the sparse-view and low SNR data to the high-quality MBIR reconstruction for that single/few time-steps. Once this network has been trained, we can then rapidly apply it to all the time-steps in the scan thereby enabling high-quality real-time feedback to the end-user. The reference scan can also be obtained ‘offline’ prior to the experiment if the sample can be scanned prior to the time-resolved CT study.

Our experiment involved studying the water uptake through the roots of a mulberry weed plant (*Fatoua villosa* (Thunb.) Nakai) using time-resolved NCT (see figure 11). Since neutrons are heavily attenuated by water, they are an ideal tool to conduct such studies. Prior NCT of plant systems over 13 h followed by FBP reconstruction clearly revealed plant structure above and below-ground, but such long acquisition time limits assessment of sub-daily 3D water uptake dynamics, instead relying on 2D radiography [50, 51]. The plant-soil system consisted of a single plant seedling propagated in pure silica sand within a 2 cm wide square aluminum cylinder (see figure 11). NCT was done at the HFIR CG-1D neutron imaging beamline using the ANDOR Zyla scientific Complementary Metal-Oxide Semiconductor (sCMOS) detector with 2560 pixels \times



2160 pixels equipped with a 100 μm thick $^6\text{LiF/ZnS}$ scintillator. A collection of 388 projections were measured by rotating the sample from 0 to 360 degrees. Each projection was acquired in 10 s and each CT scan took 1.75 h. This process was repeated for approximately 2 days to obtain a large collection of projections corresponding to water uptake through the sample. The data from the first ‘time-step’ corresponding to 388 views (about 1.75 h) was used to train the DCNNs by reconstructing the data using FBP and a base-line MBIR algorithm [12] with the parameters chosen to produce a visually high reconstruction quality. We use a total of 512 slices splitting them into 400 slices for training and 112 slices for validation. The trained network is then used to reconstruct the data from other time steps by first obtaining and low-quality FBP reconstruction and then post-processing the results using the DCNN approach. We compare the results of the proposed DCNN techniques to a standard post-processing algorithm based on total-variation denoising as well as a simpler neural network approach, the NN-FBP, that has been proposed for accelerated neutron tomography.

Figures 12 and 13 shows the output of the different reconstruction algorithms on data from time-step 3 (approximately 5.25 h from the start) and 6 (approximately 10.5 h from the start) from the data set. We note that all the DCNN approaches are able to suppress artifacts in the base-line FBP method while preserving details in the reconstruction. The pores in the sample are clearly reconstructed using the proposed DCNN algorithms compared to the base-line FBP and the simple NN-FBP algorithm, resulting in a quality similar to the MBIR method but with significantly reduced computational complexity. In this case, we do not again observe a discernible difference in the qualitative performance between the MSD-Net and the 2.5D AI-CT networks, suggesting that in despite of having limited training data, the 2.5D AI-CT algorithm can be trained using standard augmentation techniques to have a similar performance to the MSD-Net. While the DCNNs are better at preserving details compared the NN-FBP method, the performance of NN-FBP is not significantly degraded compared to the other algorithms as in the case of the meteorite rocks because of the higher similarity of the acquisition conditions and the CT reconstructions to the training data. In summary, we illustrate that it is possible to obtain high-quality reconstructions for high-speed (sparse-view) time resolved NCT by effectively training DCNNs to remove artifacts from noisy FBP reconstructions.

4. Conclusion

In this paper, we use deep convolutional neural network-based tomographic reconstruction algorithms for obtaining high-quality reconstructions from sparse-view and low SNR neutron computed tomography data thereby enabling accelerated scans. In our experiments, we demonstrate that two popular network architectures—2.5D AI-CT and MSD-Net—could be effectively trained to obtain high-quality

reconstruction from sparse-view and noisy data. Despite the limited amount of training data available in NCT applications, we observed that the 2.5D AI-CT method was able to perform similarly to the MSD-Net that is designed to work well with limited training data. We also observed that compared to a simpler neural network technique, the use of the convolutional neural networks can result in significant improvements in performance while also being better at generalization. We demonstrated the utility of our method on two sets of experiments—one for accelerated neutron tomography of a collection of meteorite samples and the other for a high-speed time-resolved tomography of a plant system. Our results illustrate that non-iterative deep convolutional neural network based reconstruction algorithms can lead to an efficient use of precious beam time at NCT facilities by decreasing measurement time and enabling high-speed experiments leading to an overall throughput close up to a factor of four for the samples considered in the manuscript. Finally, we caution that as with any reconstruction algorithm, DCNN-based techniques also have known drawbacks such as poor generalization and the potential to create/blur features [52]. Hence, while using these techniques it is important to account for these risk factors that might affect downstream analysis.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

A portion of this research used resources at the High Flux Isotope Reactor, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory. US Antarctic meteorite samples are recovered by the Antarctic Search for Meteorites (ANSMET) program which has been funded by NSF and NASA, and characterized and curated by the Department of Mineral Sciences of the Smithsonian Institution and Astromaterials Curation Office at NASA Johnson Space Center. This research was also supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. Amir Ziabari acknowledges that this research used resources of the Compute and Data Environment for Science (CADES) at Oak Ridge National Laboratory, which is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC05-00OR22725. This work has been supported in part by the Artificial Intelligence Initiative at Oak Ridge National Laboratory.

ORCID iDs

Singanallur Venkatakrishnan  <https://orcid.org/0000-0002-7637-4987>

Amirkoushyar Ziabari  <https://orcid.org/0000-0003-4776-457X>

Jacob Hinkle  <https://orcid.org/0000-0002-7751-1760>

Jeffrey M Warren  <https://orcid.org/0000-0002-0680-4697>

Hassina Z Bilheux  <https://orcid.org/0000-0001-8574-2449>

References

- [1] Anderson I S, McGreevy R L and Bilheux H Z 2009 *Neutron Imaging and Applications* vol 2209 (New York: Springer Science+ Business Media) (<https://doi.org/10.1007/978-0-387-78693-3>)
- [2] Kardjilov N, Manke I, Hilger A, Strobl M and Banhart J 2011 *Mater. Today* **14** 248–56
- [3] Zhang Y, Chandran K R and Bilheux H Z 2018 Imaging of the Li spatial distribution within V2O5 cathode in a coin cell by neutron computed tomography *J. Power Sources* **376** 125–30
- [4] Marone F and Stampanoni M 2012 *J. Synchrotron Radiat.* **19** 1029–37
- [5] Bouman C A 2013 *Model Based Image Processing* (available at: <https://engineering.purdue.edu/bouman/publications/pdf/MBIP-book.pdf>)
- [6] Yu Z, Thibault J B, Bouman C A, Sauer K D and Hsieh J 2010 *IEEE Trans. Image Process.* **20** 161–75
- [7] van Sloun R, Pandharipande A, Mischi M and Demi L 2015 *IEEE Trans. Biomed. Eng.* **62** 1660–4
- [8] Lustig M, Donoho D and Pauly J M 2007 *Magn. Reson. Med.* **58** 1182–95
- [9] Mohan K A, Venkatakrishnan S V, Drummy L F, Simmons J, Parkinson D Y and Bouman C A 2014 Model-based iterative reconstruction for synchrotron x-ray tomography 2014 *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) (Florence, Italy 4–9 May 2014)* (IEEE) (<https://doi.org/10.1109/ICASSP.2014.6854939>)
- [10] Mohan K, Venkatakrishnan S, Gibbs J, Gulsoy E, Xiao X, De Graef M, Voorhees P and Bouman C 2015 *IEEE Trans. Comput. Imaging* **1** 96–111
- [11] Venkatakrishnan S, Drummy L, Jackson M, De Graef M, Simmons J and Bouman C 2013 *IEEE Trans. on Computational Imaging* **22** 4532–44
- [12] Venkatakrishnan S, Cakmak E, Billheux H, Bingham P and Archibald R K 2017 Model-based iterative reconstruction for neutron laminography 51st *Asilomar Conf. on Signals, Systems, and Computers (Pacific Grove, CA, 29 October–1 November 2017)* (IEEE) pp 1864–9
- [13] Barnard R C, Bilheux H, Toops T, Nafziger E, Finney C, Splitter D and Archibald R 2018 *Rev. Sci. Instrum.* **89** 053704

- [14] Micieli D, Minniti T, Formoso V, Kockelmann W and Gorini G 2018 *J. Instrum.* **13** C06006
- [15] Forman C, Wetzl J, Hayes C and Schmidt M 2016 *MAGNETOM Flash* **66** 8–13 (http://clinical-mri.com/wp-content/uploads/2016/11/Forman_CompressedSensing_RSNA_2016.pdf)
- [16] Gürsoy D, Biçer T, Lanzirotti A, Newville M G and De Carlo F 2015 *Opt. Express* **23** 9014–23
- [17] McCann M T, Jin K H and Unser M 2017 *IEEE Signal Process. Mag.* **34** 85–95
- [18] Arridge S, de Hoop M, Maass P, Öktem O, Schönlieb C and Unser M 2019 Deep learning and inverse problems *Snapshots of Modern Mathematics from Oberwolfach* (<https://doi.org/10.14760/SNAP-2019-015-EN>)
- [19] Pelt D M and Batenburg K J 2013 *IEEE Trans. Image Process.* **22** 5238–51
- [20] Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 234–41
- [21] Han Y S, Yoo J and Ye J C 2016 (arXiv:1611.06391)
- [22] Jin K H, McCann M T, Froustey E and Unser M 2017 *IEEE Trans. Image Process.* **26** 4509–22
- [23] Yang X, De Andrade V, Scullin W, Dyer E L, Kasthuri N, De Carlo F and Gürsoy D 2018 *Sci. Rep.* **8** 1–13
- [24] Ziabari A, Ye D H, Srivastava S, Sauer K D, Thibault J B and Bouman C A 2018 2.5D deep learning for CT image reconstruction using a multi-GPU implementation *2018 52nd Conf. Signals, Systems and Computers* (IEEE) pp 2044–9
- [25] Ziabari A, Kirka M, Paquit V, Bingham P and Venkatakrishnan S 2019 *Microsc. Microanal.* **25** 376–7
- [26] Ziabari A, Venkatakrishnan S, Kirka M, Dehoff R, Bingham P and Paquit V 2020 Beam hardening artifact reduction in x-ray CT reconstruction of 3D printed metal parts leveraging deep learning and CAD models *Proc. ASME 2020 Int. Mechanical Engineering Congress and Exposition IMECE2020 (IMECE2020–23766) (16–19 November Virtual, Online)* **84492** V02BT02A043
- [27] Majee S, Balke T, Kemp C A, Buzzard G T and Bouman C A 2019 4D x-ray CT reconstruction using multi-slice fusion *2019 IEEE Int. Conf. Computational Photography (ICCP)* (IEEE)
- [28] Liu Z, Bicer T, Kettimuthu R, Gürsoy D, De Carlo F and Foster I 2020 *J. Opt. Soc. Am. A* **37** 422–34
- [29] Pelt D M, Batenburg K J and Sethian J A 2018 *J. Imaging* **4** 128
- [30] Pelt D M and Sethian J A 2018 *Proc. Natl Acad. Sci.* **115** 254–9
- [31] Micieli D, Minniti T, Evans L M and Gorini G 2019 *Sci. Rep.* **9** 1–12
- [32] Kak A C and Slaney M 2001 *Principles of Computerized Tomographic Imaging (Classics in Applied Mathematics)* (SIAM) (<https://doi.org/10.1137/1.9780898719277>)
- [33] Zhang K, Zuo W, Chen Y, Meng D and Zhang L 2017 *IEEE Trans. Image Process.* **26** 3142–55
- [34] Ye D H, Srivastava S, Thibault J B, Sauer K and Bouman C A 2018 Deep residual learning for model-based iterative CT reconstruction using plug-and-play framework *2018 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)* (IEEE) pp 6668–72
- [35] Ioffe S and Szegedy C 2015 *Proc. Machine Learning Research* **37** 448–56
- [36] Huang G, Liu Z, Van Der Maaten L and Weinberger K Q 2017 Densely connected convolutional networks *Proc. IEEE Conf. Computer Vision and Pattern Recognition* pp 4700–8
- [37] Rudin L I, Osher S and Fatemi E 1992 *Phys. D: Nonlinear Phenom.* **60** 259–68
- [38] Barbero A and Sra S 2011 Fast newton-type methods for total variation regularization *ICML L Getoor and T Scheffer* (Omnipress) pp 313–20 (available at: <http://dblp.uni-trier.de/db/conf/icml/icml2011.html#JimenezS11>)
- [39] Barbero A and Sra S 2018 *J. Mach. Learn. Res.* **19** 1–82 (<https://www.jmlr.org/papers/volume19/13-538/13-538.pdf>)
- [40] Gürsoy D, De Carlo F, Xiao X and Jacobsen C 2014 *J. Synchrotron Radiat.* **21** 1188–93
- [41] van Aarle W, Palenstijn W J, Beenhouwer J D, Altantzis T, Bals S, oost Batenburg K J and Sijbers J 2015 The ASTRA Toolbox: a platform for advanced algorithm development in electron tomography *Ultramicroscopy* **157** 35–47
- [42] van Aarle W, Palenstijn W J, Cant J, Janssens E, Bleichrodt F, rei Dabravolski A, Beenhouwer J D, Batenburg K J and Sijbers J 2016 *Opt. Express* **24** 25129–47
- [43] Venkatakrishnan S V 2019 *pyMBIR Technical Report* Oak Ridge National Lab. (ORNL), Oak Ridge, TN
- [44] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization *3rd Int. Conf. for Learning Representations* (arXiv:1412.6980)
- [45] Wang X, Sabne A, Sakdhnagool P, Kisner S J, Bouman C A and Midkiff S P 2017 Massively parallel 3D image reconstruction *Proc. Int. Conf. High Performance Computing, Networking, Storage and Analysis* pp 1–12
- [46] Marchesini S, Trivedi A, Enfedaque P, Perciano T and Parkinson D 2020 Sparse matrix-based HPC tomography *Int. Conf. on Computational Science* (https://doi.org/10.1007/978-3-030-50371-0_18)
- [47] Hidayetoglu M, Bicer T, de Gonzalo S G, Ren B, De Andrade V, Gürsoy D, Kettimuthu R, Foster I T and Hwu W m W 2020 Petascale XCT: 3D image reconstruction with hierarchical communications on multi-GPU nodes pp 1–13
- [48] Tötze C, Kardjilov N, Lenoir N, Manke I, Oswald S E and Tengattini A 2019 *Opt. Express* **27** 28640–8
- [49] Tudisco E, Etxegarai M, Hall S A, Charalampidou E M, Couples G D, Lewis H, Tengattini A and Kardjilov N 2019 *J. Geophys. Res. Solid Earth* **124** 3557–69
- [50] Warren J M, Bilheux H, Kang M, Voisin S, Cheng C L, Horita J and Perfect E 2013 *Plant Soil* **366** 683–93
- [51] Dhiman I, Bilheux H, DeCarlo K, Painter S L, Santodonato L and Warren J M 2018 *Plant Soil* **424** 73–89
- [52] Antun V, Renka F, Poon C, Adcock B and Hansen A C 2020 *Proc. Natl Acad. Sci.* **117** 48