
TOMOGAN: LOW-DOSE SYNCHROTRON X-RAY TOMOGRAPHY WITH GENERATIVE ADVERSARIAL NETWORKS

Zhengchun Liu

Data Science and Learning Division
Argonne National Laboratory
Lemont, IL 60439, USA
`zhengchun.liu@anl.gov`

Tekin Bicer

Data Science and Learning Division
X-ray Science Division
Argonne National Laboratory
Lemont, IL 60439, USA

Rajkumar Kettimuthu

Data Science and Learning Division
Argonne National Laboratory
Lemont, IL 60439, USA

Doga Gursoy

X-ray Science Division
Argonne National Laboratory
Lemont, IL 60439, USA

Francesco De Carlo

X-ray Science Division
Argonne National Laboratory
Lemont, IL 60439, USA

Ian Foster

Data Science and Learning Division
Argonne National Laboratory
Lemont, IL 60439, USA

January 1, 2020

ABSTRACT

Synchrotron-based x-ray tomography is a noninvasive imaging technique that allows for reconstructing the internal structure of materials at high spatial resolutions from tens of micrometers to a few nanometers. In order to resolve sample features at smaller length scales, however, a higher radiation dose is required. Therefore, the limitation on the achievable resolution is set primarily by noise at these length scales. We present TomoGAN, a denoising technique based on generative adversarial networks, for improving the quality of reconstructed images for low-dose imaging conditions. We evaluate our approach in two photon-budget-limited experimental conditions: (1) sufficient number of low-dose projections (based on Nyquist sampling), and (2) insufficient or limited number of high-dose projections. In both cases the angular sampling is assumed to be isotropic, and the photon budget throughout the experiment is fixed based on the maximum allowable radiation dose on the sample. Evaluation with both simulated and experimental datasets shows that our approach can significantly reduce noise in reconstructed images, improving the structural similarity score of simulation and experimental data from 0.18 to 0.9 and from 0.18 to 0.41, respectively. Furthermore, the quality of the reconstructed images with filtered back projection followed by our denoising approach exceeds that of reconstructions with the simultaneous iterative reconstruction technique, showing the computational superiority of our approach.

1 Introduction

X-ray computed tomography (CT) is a common noninvasive imaging modality for resolving the internal structure of materials at synchrotrons [1]. In CT, 2D projection images of an object are collected at different views of the object around a common axis, and a numerical reconstruction process is applied afterwards to recover the object's morphology in 3D. Although CT experiments at synchrotrons can collect data at high spatial and temporal resolution, however, *in situ* or dose-sensitive experiments require short exposure times to capture relevant dynamic phenomena or to avoid sample damage. These low-dose (LD) imaging conditions yield noisy measurements that significantly impact the quality

of the resulting 3D reconstructions. Similar concerns arise when the number of projections is limited to meet speed and/or dose requirements, such as in conventional lab-based micro-CT systems, where measurements are collected at discrete rotations of the object. Thus, we want techniques that can map noisy reconstructions (due to fewer projections, lower resolutions, and/or shorter imaging times with noisy measurements) to an approximation of the ideal image.

Much research has been conducted on methods for improving the quality of noisy low-dose images. Broadly, these approaches fall into three categories: (i) methods for denoising measurements/raw data, for example, sinograms or projections [2, 3, 4, 5]; (ii) advanced iterative reconstruction algorithms, for example, model-based approaches [6, 7]; and (iii) methods, especially deep convolution neural network based [8, 9, 10], for denoising reconstructed images [11]. For (i) and (iii), various deep learning (DL) approaches have shown great promise in the context of medical CT imaging [12, 13, 14, 15, 16]. In the context of synchrotron-based CT, for (i), Yang et al. [17] used a deep convolutional neural network (CNN) to denoise prereconstruction short-exposure-time projection images from a network trained by a few long-exposure-time (i.e., high-dose) and short-exposure-time (i.e., low-dose) projection pairs. They achieved a 10-fold increase in signal-to-noise ratio, enabling the reliable tracing of brain structures in low-dose datasets. For (iii), Pelt et al. [15] trained a mixed-scale dense convolutional neural network[16] in a supervised fashion to learn a mapping from low-dose to normal-dose reconstructions. They achieved impressive results on simulation datasets; but unfortunately the performance of proposed model is not thoroughly evaluated on different experimental datasets.

DL techniques use multi-layer (“deep”) neural networks (DNNs) to learn representations of data with multiple levels of abstraction. These techniques can discover intricate structure in a dataset by using a back-propagation algorithm to set the internal parameters that are used to transform data as they flow between network layers. Recent advances in DL, such as convolutional networks [18], rectifier linear units (RELUs) [19], batch normalization [20], dropout [21], and residual learning [22], have enabled exciting new applications in many areas. DL techniques have been applied successfully to a range of scientific imaging problems, such as denoising, super-resolution, and image enhancement and restoration [23, 24, 25, 26].

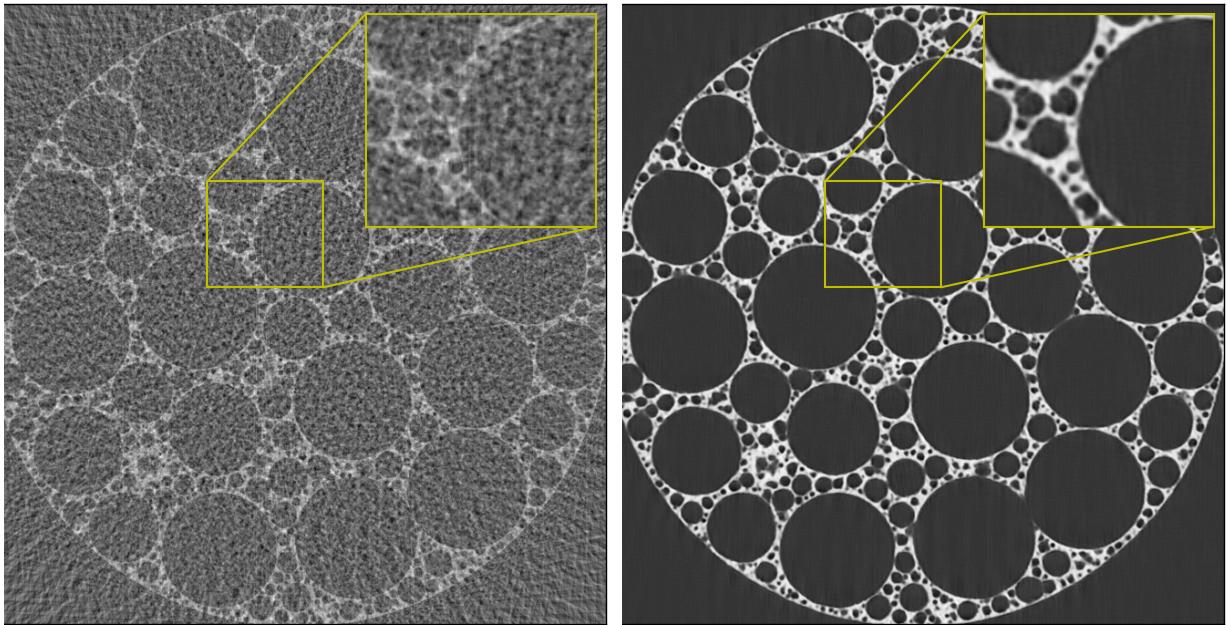


Figure 1: Two different reconstructions of a noisy simulated dataset, constructed by subsampling 64 projections from a 1024-projection simulated dataset containing foam features in a 3D volume. On the left, the results of conventional reconstruction, which are highly noisy. On the right, those same results after denoising with TomoGAN; the features are much more visible. In these images and others that follow, an inset shows details of a representative feature.

In this article we explore an alternative DL approach to image enhancement, namely, the use of generative adversarial networks (GANs). In general, a GAN involves two neural networks, a generator G and a discriminator D , that contest with each other in a zero-sum-game framework[27]. Training a GAN model involves a minimax game between the G that mimics the true distribution and the D that distinguishes samples produced by the G from the real samples. Compare with the DNN based methods, the discriminator, provides adversarial loss, is a helper for training the generator

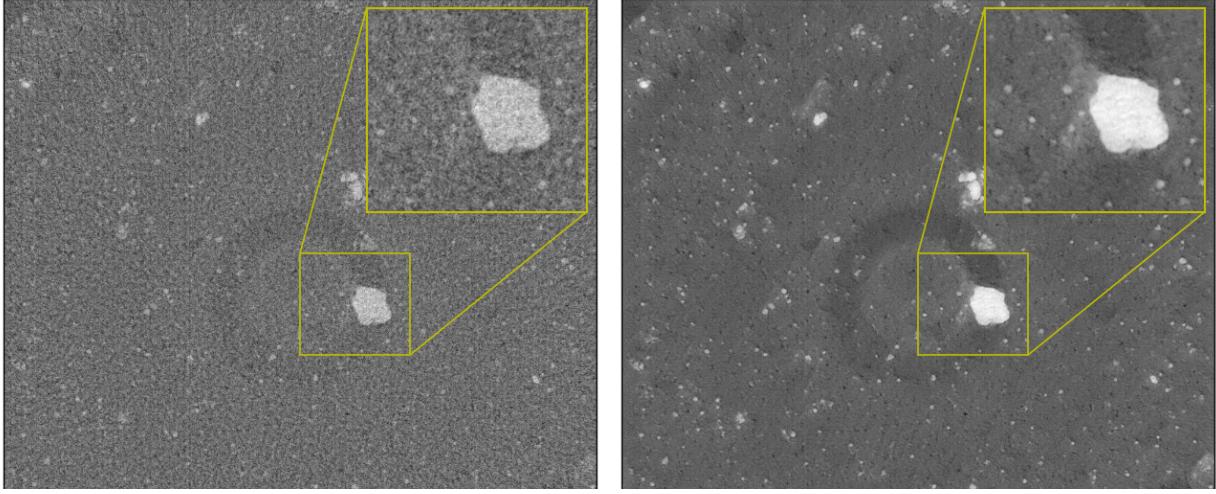


Figure 2: Two different reconstructions of a noisy experimental dataset, constructed by subsampling 64 projections from a 1024-projection shale sample dataset. On the left, the results of conventional reconstruction, which are highly noisy. On the right, those same results after denoising with TomoGAN; the features are much more visible.

to generate better perceptual quality images. GANs have been applied successfully in medical imaging [28, 29, 30] but have not previously been used with high-resolution imaging techniques at synchrotrons. The challenge in the synchrotron context is that the high-resolution images produced include finely detailed features with high-frequency content. Approaches developed for medical images are typically insufficient since they are tailored to easily recognizable features with low-frequency content and, when applied to high-resolution images, can introduce undesired artifacts such as nonexistent features.

Our GAN-based method, TomoGAN, adapts the U-Net network architecture[28] to meet the specialized requirements of improving the quality of images generated by high-resolution tomography experiments at synchrotron light sources. We demonstrate that the TomoGAN model can be trained with limited data, performs well with high-resolution datasets, and generates greatly improved reconstructions of low-dose and noisy data, as shown in Figure 1 and Figure 2 (with high-frequency content). We also show that our model can be applied to a variety of experimental datasets from different instruments, showing that it is resilient to overfitting and has wider applicability in practice.

We extensively evaluate our approach with real-world tomography datasets in order to prove the applicability of the proposed method in practice. These experimental datasets are from different types of shale samples collected at different facilities by using the same technique but different imaging conditions (different x-ray sources and detectors). We simulate two low-dose scenarios[31]: (1) picking a subset of the x-ray projections, to simulate reduced number of projections as in the case of a lab-based CT system, and (2) applying synthetic noise to the x-ray projections, to simulate short exposure times. Both scenarios lead to noisy reconstructed images. We use one dataset to train TomoGAN and then evaluate the trained model on others. We compare the denoised (DN) images with ground truth and measure the quality of denoised images using (1) the structural similarity (SSIM) [32, 33, 34] index and (2) image pixel value plots. Our evaluation results show that our approach can significantly improve image quality by reducing the noise in reconstructed images. We believe that this approach will also be effective for improving reconstruction quality when the same sample structure is imaged with different techniques with different imaging contrasts, for example, in multimodal imaging systems.

2 Methods

We describe in turn the TomoGAN model architecture, the process by which we train a TomoGAN model, and the datasets and experimental setup used for evaluation.

2.1 Model architecture

Generally, the task of denoising a reconstructed image can be posed as that of translating the noisy image into a corresponding output image that represents exactly the same features, with the features in the enhanced image indistinguishable from those in a ground truth version. Machine learning models learn to minimize a loss function—an objective that scores the quality of results—and although the learning process is automatic, the model still must be *told* what needs to be minimized.

A GAN combines two neural networks, a generator (G) and a discriminator (D), which compete in a zero-sum game: G generates candidates that D evaluates; those evaluations serve as feedback to G . Thus, GANs are designed to reach a Nash equilibrium at which neither of the two networks can reduce its costs without changing the other player's parameters. In this paper, we train a conditional GAN in which the generator model G maps a noisy reconstruction (i.e., conditionally use the noisy reconstruction as input to the G instead of a random value, as it in a standard GAN [27]) into a form that is indistinguishable from the adversarial model D that is trained to distinguish reconstructions of noisy projections from the enhanced noisy reconstructions created by G . Thus, we use G to enhance images; D simply works as a helper to train the G . A classic GAN generates samples from random noise inputs [27]; In contrast, our TomoGAN network creates samples from closely related noisy inputs.

2.1.1 Generator

The TomoGAN generator network architecture, shown in Figure 3, is a variation of the U-Net architecture proposed for biomedical image segmentation by Shan et al.[28]. It comprises a down-sampling network (left) followed by an up-sampling network (right). It adapts the U-Net architecture in three main ways: (1) there are three (instead of four) down-sampling layers and up-sampling layers; (2) all convolution layers have zero padding in order to keep the same image size; and (3) the input is a stack of d adjacent images (discussed in the next section), and eight 1×1 convolution kernels are applied to the input.

In the down-sampling process, three sets of two convolution kernels (the three boxes) extract feature maps. Then, followed by a pooling layer, the feature map projections are distilled to the most essential elements by using a signal maximizing process. Ultimately, the feature maps are 1/8 of the original size: 128×128 in Figure 3. Successful training should result in 128 channels in this feature map, retaining important features, which in Figure 19 seems to be the case.

In the up-sampling process, bi-linear interpolation is used to expand feature maps. At each layer, high-resolution features from the down-sampling path (transmitted via Copy operations) are concatenated to the up-sampled output from the layer below to form a large number of feature channels. This structure allows the network to propagate context information to higher-resolution layers, so that the following convolution layer can learn to assemble a more precise output based on this information.

2.1.2 Discriminator

The TomoGAN discriminator has six 2D 3×3 CNN layers and two fully connected layers. Each CNN layer is followed by a leaky rectified linear unit as the activation function. Following the same logic as in G , all convolutional layers in D have the same small 3×3 kernel size. Let C_{kSs-n} denote a convolution layer with a kernel size of $k \times k$, a stride of $s \times s$, n output channels, and leaky ReLU activation function. The discriminator network consists of C_{3S1-64} , C_{3S2-64} , $C_{3S1-128}$, $C_{3S2-128}$, $C_{3S1-256}$, C_{3S2-4} , one hidden fully connected layer with 64 neurons and leaky ReLU activation, and an output layer with one neuron and linear activation. There is no sigmoid cross-entropy layer at the end of the discriminator, and the Wasserstein distance is used for calculating generator loss in order to improve the training stability[35].

2.2 Model training

We present the loss functions used in the TomoGAN discriminator and generator.

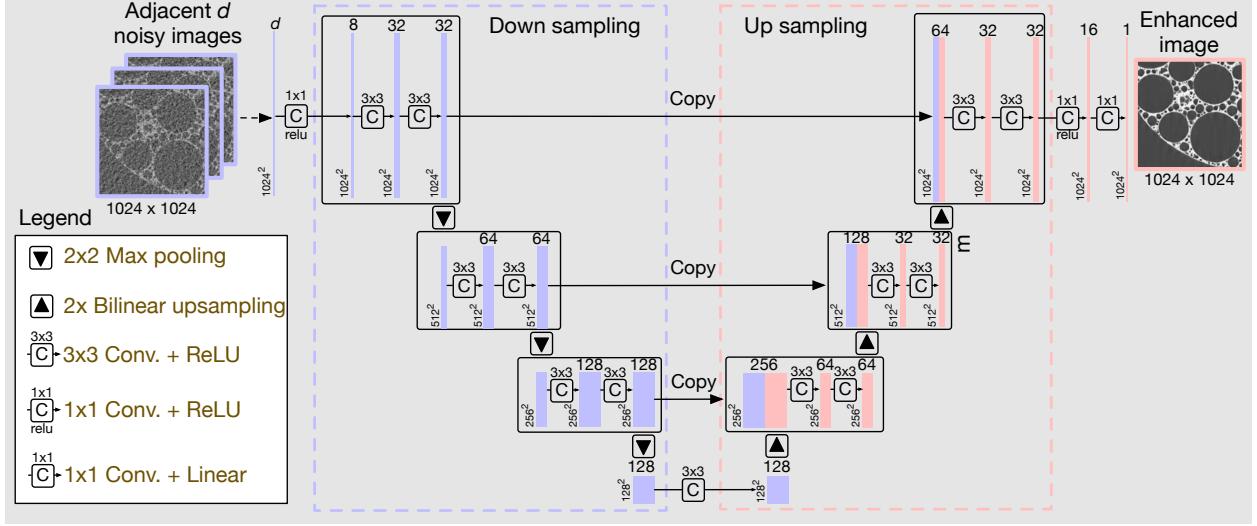


Figure 3: The TomoGAN generator architecture is applied to images of size $m \times n$. Here, we show its operation when $m = n = 1024$ as an example. Each bar corresponds to a multi-channel feature map, with the number of channels shown on the top of the bar and the height and width at the lower left edge. The symbols (see legend) show the operations used to transform one feature map to the next. Zero padding is used for all convolution layers to keep the output size the same as the input size. The Copy operations allow low-level information to shortcut across the network, thus improving the spatial resolution of the corresponding features.

2.2.1 Discriminator loss

In addition to the original critical loss presented by Arjovsky et al.[35], we add a gradient penalty as suggested by Gulrajani et al.[36] to Equation 1 for better training stability. Thus, the discriminator is trained to minimize

$$L(\theta_D) = \frac{1}{m} \sum_{i=1}^m [D(G(I_{LD}^i)) - D(I_{ND}^i)] + \lambda_D \frac{1}{m} \sum_{i=1}^m \left[\left(\left\| \nabla_{\bar{I}} D(\bar{I}^i) \right\|_2 - 1 \right)^2 \right], \quad (1)$$

where θ_D is the trainable weight/parameter of D , I_{LD}^i is one noisy image (i th in the minibatch), m is the training minibatch size, and I_{ND}^i is the corresponding ground truth image.

$\bar{I}^i = \epsilon^i G(I_{LD}^i) + (1 - \epsilon^i) I_{ND}^i$, where ϵ is a random number between 0 and 1. As in Gulrajani et al.[36], we use $\lambda_D = 10$ to balance the trade-off between the Wasserstein distance and the gradient penalty.

2.2.2 Generator loss

An important requirement while denoising images is that **the generator not introduce artificial features**. Thus, previous approaches have found it beneficial to extend the loss function to be minimized by the GAN generator with a more traditional loss, such as the **ℓ_2 -norm distance** (i.e., mean squared error)[37]. ℓ_2 -norm distance is one of the most commonly used loss metrics for image restoration problems [23, 8] although it is well known to produce blurry results on image generation problems in classical approaches [38, 39]. Thus, we also add **ℓ_2 -norm loss** for the generator to enforce correctness of low-frequency structures [26, 40, 41]. The discriminator's job remains unchanged, but the generator is tasked with not only generating adversarial samples but also being near the ground truth output in an **ℓ_2 -norm sense**. Moreover, perceptual losses are also used to penalize any structure that differs between output and target. Thus, the generator loss is a weighted average of three losses: the original GAN adversarial loss ℓ_{adv} , perceptual loss ℓ_{VGG} , and pixelwise mean ℓ_2 -norm ℓ_{mse} .

Adversarial loss. As in the Wasserstein GAN [35], we compute the adversarial loss as

$$\ell_{adv}(\theta_G) = -\frac{1}{m} \sum_{i=1}^m D(G(I_{LD}^i)). \quad (2)$$

Perceptual loss. To allow the generator to retain a visually desirable feature representation, we also use the mean squared error (MSE) of features extracted by a ImageNet pre-trained VGG-19 network for the perceptual loss. Specifically, we use the first 16 layers of the pre-trained VGG[42] network to extract the feature representation of a given image. We then define the perceptual loss as the Euclidean distance between the feature representations of a ground truth image $V_{\theta_{vgg}}(I^{ND})$ and the corresponding denoised image $V_{\theta_{vgg}}(G_{\theta_G}(I^{LD}))$. Thus, the perceptual loss is calculated by

$$\ell_{vgg} = \sum_{i=1}^{W_f} \sum_{j=1}^{H_f} \left(V_{\theta_{vgg}}(I^{ND})_{i,j} - V_{\theta_{vgg}}(G_{\theta_G}(I^{LD}))_{i,j} \right)^2, \quad (3)$$

where W_f and H_f denote the dimensions of the feature maps extracted by the pre-trained VGG network.

Since the VGG network is trained with natural images, namely, ImageNet, one may have concerns about how well it can perform on light source images. Yang et al.[29] demonstrated that the VGG network, when pre-trained with the ImageNet dataset, can also serve as a good feature extractor for CT images.

Pixel-wise MSE. The pixel-wise MSE loss is calculated as

$$\ell_{mse} = \sum_{c=1}^W \sum_{r=1}^H \left(I_{c,r}^{ND} - G_{\theta_G}(I^{LD})_{c,r} \right)^2, \quad (4)$$

where W and H denote the width and height of the image, respectively. The final loss function that the generator must minimize is thus

$$\ell^G = \lambda_g \ell_{adv} + \lambda_p \ell_{mse} + \lambda_v \ell_{vgg}. \quad (5)$$

To reduce the chance of “mode collapse”[27], we train G once every four training steps of D . Figure 4 shows the training pipeline of the model. Once the GAN model is trained, we can input a noisy image to the generator and it outputs the corresponding enhanced image. In practice, we save the generator and publish it on the DLHub [43] to serve for users.

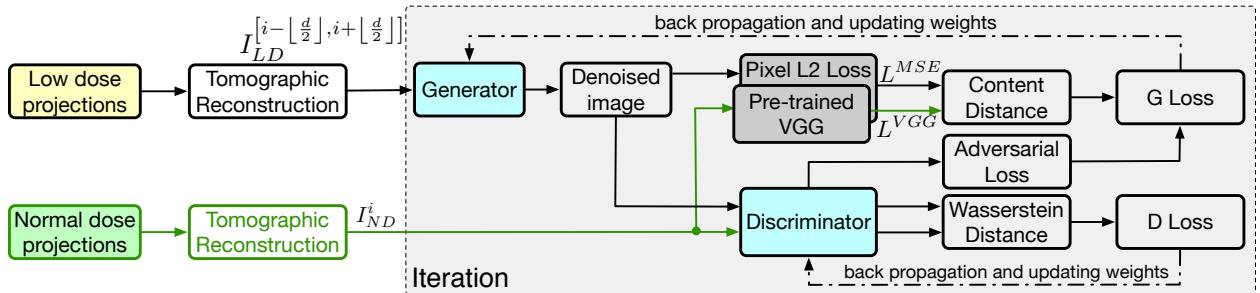


Figure 4: Model training pipeline. The parameter i is chosen randomly for each training iteration, and the corresponding slices of the reconstructed low-dose (I_{LD}) and normal-dose (I_{ND}) projections are used as input. Once the model is trained, only the generator, G , is used to advance the tomographic reconstructions.

2.3 Datasets and experimental setup

We used two simulated datasets and four experimental datasets to evaluate the proposed TomoGAN. The experimental datasets are two shale samples that are imaged at two different facilities[44, 45]). The datasets are provided for benchmarking purposes and retrieved from TomoBank[46].

2.3.1 Simulated datasets

Each simulated dataset[15] has 100 000 different-sized foam features distributed randomly in the 3D volume. We use the ASTRA toolbox[47] to generate the corresponding x-ray projections, with the total number of simulated projections (i.e., the total number of angles) set to 1024 and each projection consisting of 1024×1024 pixels. We reconstruct the 3D volumes, size $1024 \times 1024 \times 1024$, with the filtered back projection algorithm in the TomoPy and ASTRA toolkits[48, 49], and we use the generated 2D slices in the 3D volumes for training and testing. Two different random seeds are used to generate two different datasets, DS_{simu}^1 and DS_{simu}^2 . We thus train G with DS_{simu}^1 and evaluate the trained model on DS_{simu}^2 .

To simulate the case with a limited number of x-ray projections, we subsample projections in DS_{simu}^1 and DS_{simu}^2 to 1/2, 1/4, 1/8, and 1/16 of the total 1024 projections. For the case with a shorter exposure time, we simulate the measured photon counts simulated by using the Beer-Lambert law [31]. Specifically, background per-pixel photon counts are initially set to 100, 500, 1000, and 10 000, to simulate x-ray intensity. Then, the photon counts are re-sampled by using Poisson distribution [50], and the new noisy measurements are used to generate the line integrals and noisy image reconstructions. The photon absorption of the samples is set to 2.5% for all datasets.

2.3.2 Experimental datasets

Shale is a challenging material because of its multi-phase composition, small grain size, low but significant amount of porosity, and strong shape- and lattice-preferred orientation. In this work, we use two shale sample datasets obtained from the North Sea (sample *N1*) and the Upper Barnett Formation in Texas (sample *B1*)[45]. Each sample has been imaged at two different light source facilities: the Advanced Photon Source (APS) at Argonne National Laboratory and the Swiss Light Source (SLS) at the Paul Scherrer Institut. For ease of reference, we name the resulting four datasets as DS_{APS}^{B1} , DS_{APS}^{N1} , DS_{SLS}^{B1} , and DS_{SLS}^{N1} .

When evaluating performance with a limited number of projections, we subsample to 1/2, 1/4, 1/8, and 1/16 of the total number of projections in each full dataset. Similar to simulated datasets, we apply Poisson noise [50, 51] to simulate the impact of shorter exposure times on projection quality, and we add this simulated noise to the normal exposure time projections.

We arbitrarily select the dataset DS_{APS}^{B1} to train TomoGAN and use the other three datasets to evaluate its effectiveness and performance. To train or enhance the i^{th} slice, we use slices from $i - \lfloor d/2 \rfloor$ to $i + \lfloor d/2 \rfloor$ as input to the generator G , where d is a tunable parameter. The generated (enhanced) images as well as the corresponding i^{th} slice of the normal dose are then input to the discriminator to compute the loss and update θ_D . We implemented¹ TomoGAN with Tensorflow[52] and used one NVIDIA Tesla V100 GPU cards for training. We used the ADAM algorithm[53] to train both the generator and discriminator, with a batch size of four images and a depth of three for each image.

3 Experimental Results

In this section, we evaluate the performance of TomoGAN with different configurations and various datasets. We first show the effectiveness of the depth parameter in TomoGAN and importance of the various loss terms. Next, we present the performance of TomoGAN with both simulation and experimental datasets. We then compare the computational requirements and image quality of FBP followed by TomoGAN against a simultaneous iterative reconstruction technique. At the end, we compare TomoGAN with other representative solutions on experimental dataset.

3.1 Effectiveness of using adjacent slices in image enhancement

Features observed in adjacent slices tend to be highly correlated, usually with similar shape but different size. Stacking several neighboring slices to form multiple channels for the first CNN layer is equivalent to using a multiple linear regressor as a filter to reduce noise. Thus we use slices from $i - \lfloor d/2 \rfloor$ to $i + \lfloor d/2 \rfloor$ as input to enhance the i^{th} slice, where d is a hyperparameter. To explore the effectiveness of d , we used DS_{simu}^1 to train models with different d (i.e., $d = 1$, $d = 3$, $d = 5$, and $d = 7$) and test their performance on DS_{simu}^2 . Figure 5, which compares the quality of an arbitrarily chosen image enhanced by model with different d , shows the effectiveness of using adjacent slices. For all of the slices in DS_{simu}^2 , Figure 6 compares their SSIM and PSNR. One can see from Figure 6 that d has big influence on mode performance, and that $d = 3$ gets the best quality, especially when the original feature edge is not sharp (e.g., the center circle in Figure 5). Using adjacent slices is one of the difference between our method and WGAN-VGG [29].

¹Our open source implementation is available at <git@github.com:ramsesproject/TomoGAN.git>

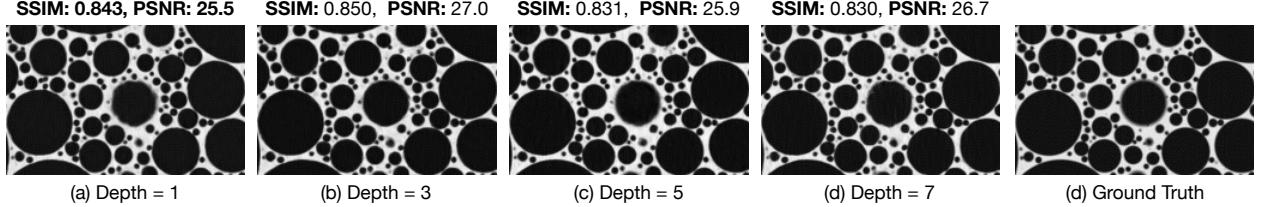


Figure 5: Effectiveness of using varying numbers of adjacent slices when enhancing reconstructed images.

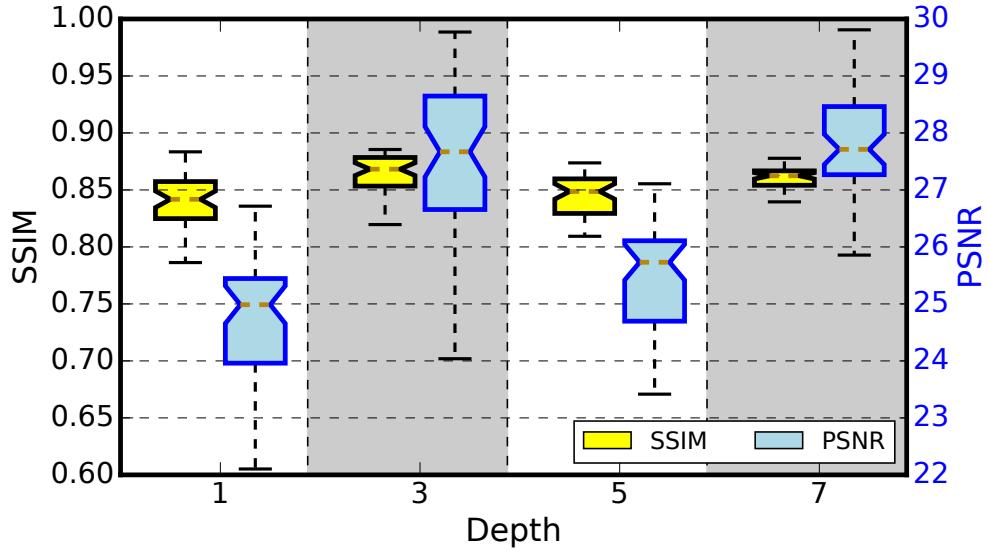


Figure 6: Effectiveness of using varying numbers of adjacent slices when enhancing reconstructed images.

We note that the best depth d depends on dataset characteristics such as feature resolution. $d = 3$ may not be the best for other datasets where feature sizes change slowly across slices.

3.2 Importance of the various loss terms

Previous works have used supervised machine learning techniques[17, 15, 54] to learn the mapping between reconstructions from noisy and noise-free images. Our approach of training a GAN model and using its generator to improve reconstructions is intended to make TomoGAN more resilient to overfitting and the creation of artifacts. In order to evaluate how sensitive results are to the chosen loss terms, we conducted experiments that selectively disable ℓ_{adv} and ℓ_{vgg} . Figure 7 shows the results achieved on a representative slice of a 1024-slice dataset. Notice how, in the three images to the left, some small features (e.g., those highlighted by the red rectangle) are merged and others (e.g., those highlighted by the red ellipse) are difficult to see. Such artifacts are not tolerable in practice. ℓ_{adv} and ℓ_{vgg} help to avoid such artifacts.

Figure 8 provides another perspective on the effectiveness of adversarial and perceptual loss, showing the SSIM and PSNR statistics for all slices of the same dataset. We see that the adversarial and perceptual loss terms each provide considerable improvements when used in isolation. The two together are only slightly better than adversarial loss alone.

3.3 Simulated datasets

Here, we show the effectiveness of TomoGAN in improving noisy reconstructed images with simulated datasets. We show both fewer projections and shorter exposure time cases that result in noisy measurement data and reconstructed images.

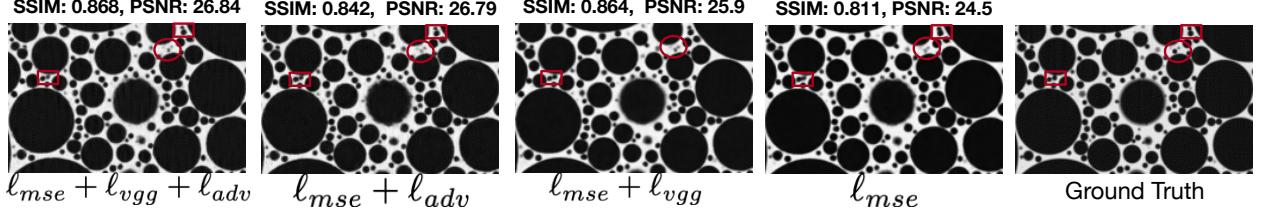


Figure 7: Effectiveness of adversarial and perceptual loss. From left to right with ℓ_2 -norm, reconstruction with both adversarial and perceptual loss; with just adversarial loss; with just perceptual loss; and with neither adversarial nor perceptual loss.

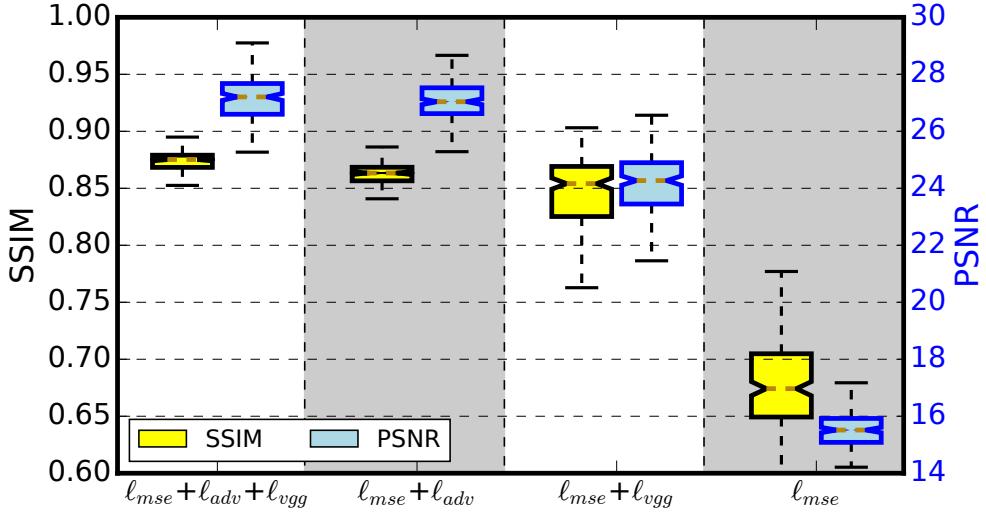


Figure 8: Effectiveness of adversarial loss and perceptual loss: SSIM and PSNR statistics.

3.3.1 Fewer projections

In these experiments, we subsample the 1024 projections to 512, 256, 128, and 64 projections to simulate a scenario for the step scan CT systems where a limited number of projections are collected to reduce either the total experiment time (e.g., to capture dynamic features) or the x-ray dose (e.g., for x-ray dose sensitive sample). We perform the same subsampling for both DS_{simu}^1 and DS_{simu}^2 . Then, once TomoGAN is trained with DS_{simu}^1 , we use its generator to enhance reconstructions based on subsampled projections of DS_{simu}^2 . In Figure 9, we zoom in on a randomly selected area and plot the pixel values of a single horizontal line to show the improvement in image quality. We see that the reconstruction quality with reduced numbers of projections is poor in the absence of enhancement; however, the enhancement with TomoGAN yields images with quality comparable to that of the reconstructions based on 1024 projections.

We use SSIM [32] to perform quantitative comparisons of the quality of TomoGAN-enhanced reconstructions. Using a reconstruction of a full 1024-projection dataset as ground truth, we calculate for each noisy (fewer projections or fewer photons) dataset both (a) SSIM between ground truth and the subsampled reconstruction, $SSIM_{GT}^{LD}$, and (b) SSIM between ground truth and the TomoGAN-enhanced version of that subsampled reconstruction, $SSIM_{GT}^{DN}$. The difference between these two quantities is the image quality improvement that results from the use of TomoGAN.

We further calculated the two values $SSIM_{GT}^{LD}$ and $SSIM_{GT}^{DN}$ for each slice in the reconstructed image that has features, thus allowing us to plot in Figure 11 the distribution of SSIM image quality scores for both the conventional ($SSIM_{GT}^{LD}$) and TomoGAN-enhanced ($SSIM_{GT}^{DN}$) images. We see that while image quality declines significantly with subsampling in the absence of TomoGAN enhancement, it remains high when enhancement is applied, even with high degrees of subsampling. SSIM is improved significantly by TomoGAN and is improved consistently across all the slices, suggesting that the method is reliable. The evaluation of this low-dose scenario can also be used for streaming tomography where the views are very limited at the beginning of experimentation (e.g., Liu et al. 2019 [55]).

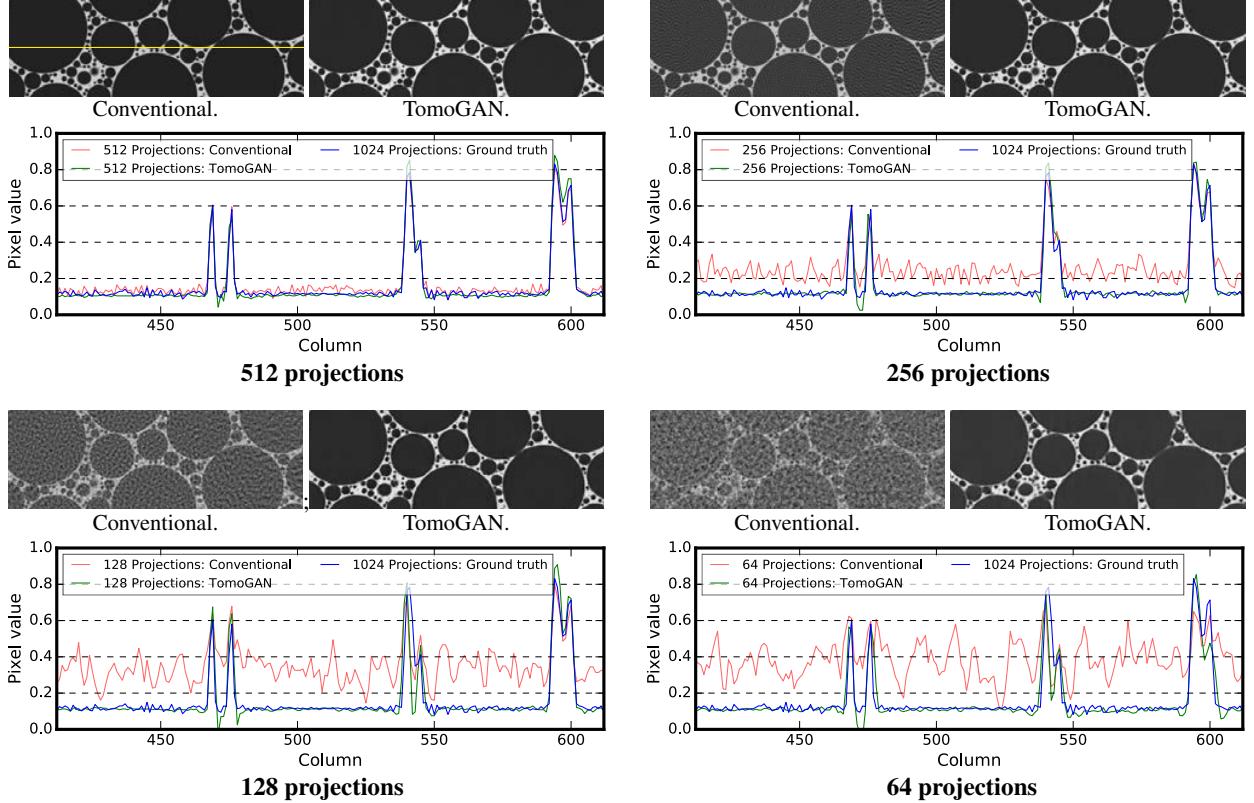


Figure 9: Conventional vs. TomoGAN-enhanced reconstructions of simulated data, subsampled to (512, 256, 128, 64) projections. In each group of three elements, the two images show conventional and TomoGAN reconstructions, while the plot shows conventional, TomoGAN, and ground truth values for the 200 pixels on the horizontal line in the top left image.

3.3.2 Shorter exposure times

Another way of reducing dose (or, equivalently, accelerating experiments) is to reduce x-ray exposure times. The use of shorter exposure times has attracted major attention for synchrotron-based tomography systems (e.g., Advanced Photon Source), because it can reduce data collection times significantly, as needed both to capture dynamical features and to reduce damage to organic samples from light source radiation. Reduced exposure time, however, compromises the signal-to-noise ratio, which affects final reconstruction image quality and may affect scientific findings.

In this experiment, we simulate x-ray projections with different photon intensities to simulate the effect of different exposure times. Specifically, we simulate with intensities of 100, 500, 1000, and 10 000 photons per pixel and, for each, compute conventional and TomoGAN-enhanced reconstructions. Figure 10 shows representative results, while Figure 11b quantifies the quality of the different reconstructed images via SSIM scores computed relative to ground truth. Ground truth here is a reconstruction based on the noise-free simulation data. We see that the trained TomoGAN model can enhance the reconstructions of noisy data to a quality comparable with that of reconstructions based on ground truth. Comparison of Figure 10 with Figure 9 shows that reducing exposure time leads to a different kind of noise from that when using fewer projections.

3.4 Experimental datasets

A key issue to address when dealing with experimental datasets is whether we can train TomoGAN on data collected from one sample and then use the trained model on other samples and/or at other facilities. Thus, we train TomoGAN with data collected on one sample at one facility, DS_{APS}^{BI} , and then evaluate its performance on noisy versions (both fewer projections and shorter exposure times) of three other datasets, DS_{APS}^{NI} , DS_{SLS}^{BI} , and DS_{SLS}^{NI} . Then, we use the trained model to enhance noisy images of the other three datasets. Since these other datasets include different samples

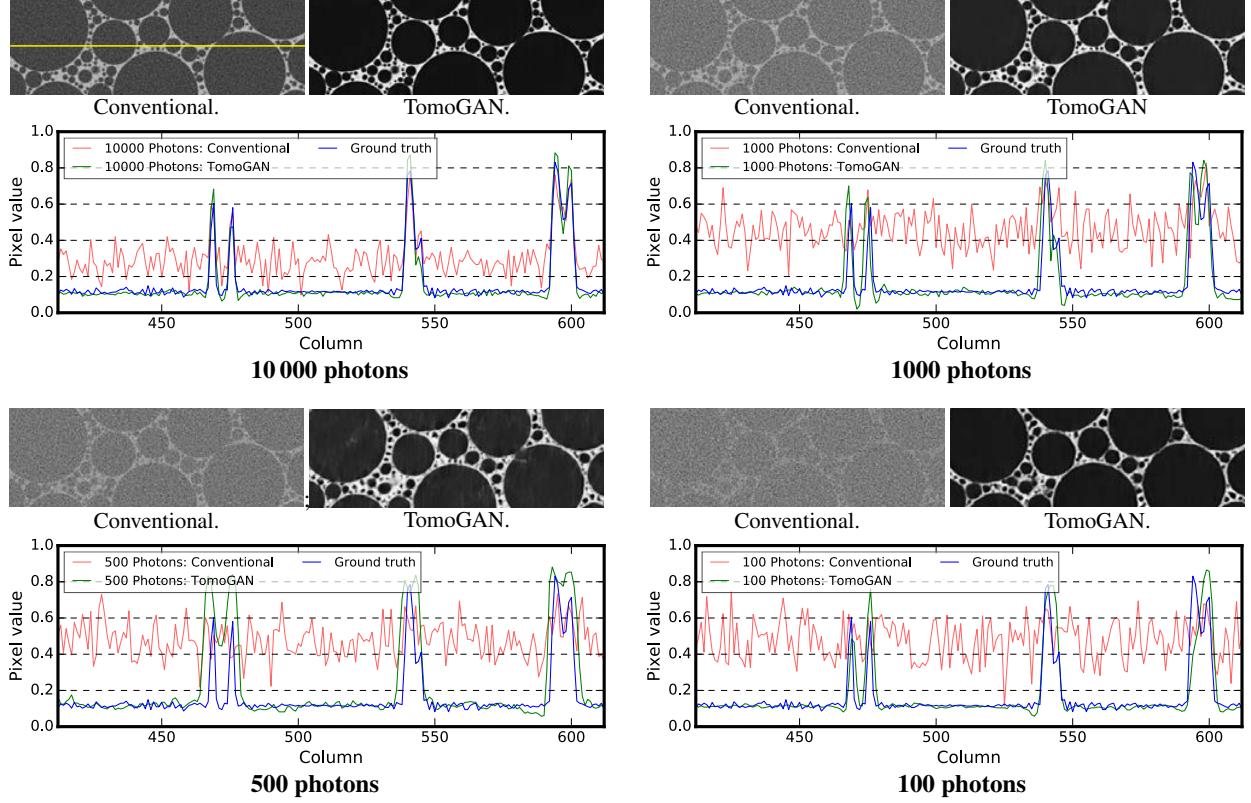


Figure 10: Conventional vs. TomoGAN-enhanced reconstructions of simulated data with intensity limited to {10 000, 1000, 500, 100} photons per pixel. Figure elements are as in Figure 9.

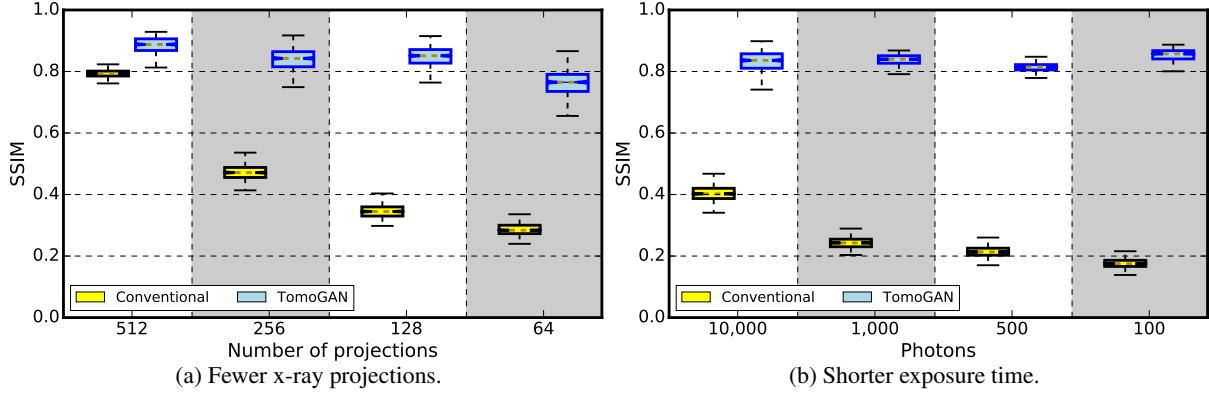


Figure 11: Per-slice SSIM similarities with ground truth simulated data, for both conventional and TomoGAN-enhanced reconstructions and for different degrees of subsampling.

(DS_*^{NI} and DS_*^{BI}) and x-ray projections collected at different facilities (DS_{APS}^* and DS_{SLS}^*), this configuration mimics a practical use case.

Evaluation of TomoGAN is more difficult for the experimental datasets since there is no ground truth. Therefore, we considered the images that are reconstructed using full datasets (1024-projection) as their corresponding ground truths.

3.4.1 Fewer projections

As we did for the simulated datasets, we subsample the full-resolution (1024-projection) experimental datasets and then use TomoGAN to enhance the reconstructions of their subsampled datasets. In Figure 12, we show conventional and

TomoGAN-enhanced reconstructions (for varying degrees of subsampling) of a randomly selected area of a slice from DS_{SLS}^{BI} . The corresponding pixels of a randomly selected horizontal line within the selected areas are also presented for comparison. We see that the image qualities are improved significantly: even the small features are clearly visible in the enhanced images. The pixel value plots also show that the TomoGAN-enhanced images are comparable with the ground truth.

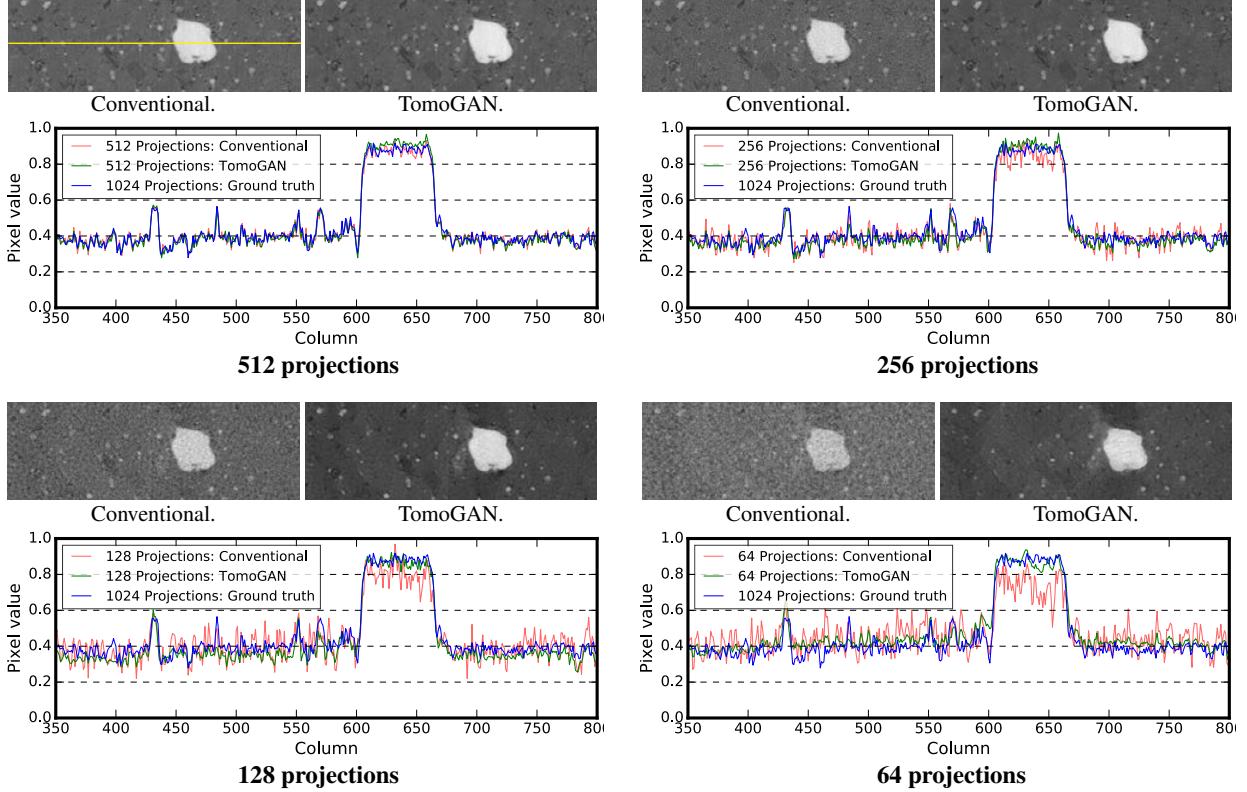


Figure 12: Conventional vs. TomoGAN-enhanced reconstructions of experimental dataset DS_{SLS}^{BI} , subsampled to (512, 256, 128, 64) projections. Figure elements are as in Figure 9.

For the most challenging cases, in which only 64 x-ray projections are available for reconstruction, Figure 13 shows the pixel values of a horizontal line that crosses an arbitrarily chosen feature in each of the four datasets. Comparing the enhanced (green) and the ground truth (blue) lines in each case, we see that TomoGAN yields reconstructions comparable to the ground truth. The contrast of the enhanced images is clearly higher than that of the noisy reconstructions; and the variance of the feature pixel values (corresponding to the noise level) has been reduced considerably from the noisy reconstruction cases. Moreover, not only is the experiment time reduced when using fewer projections (by a factor corresponding to the subsampling ratio), but the dataset size and computation cost for reconstructions are also reduced by the same fraction.

We again use SSIM to quantify the qualities of the conventional and TomoGAN-enhanced reconstructions. Recall that our model was trained with DS_{APS}^{BI} . Here we evaluate the trained model on a different shale sample imaged at a different facility: DS_{SLS}^{NI} . The SSIM metric scores, shown in Figure 14a, indicate that TomoGAN consistently improves image quality for all slices at each subsampling level. However, the overall quality scores for the TomoGAN-enhanced reconstructions are clearly not as good as those for the simulated data in Figure 11a. We attribute this difference to the fact that the simulated dataset has a much better training dataset (ground truth) than does the experimental dataset and that the features in the simulation dataset are much simpler (only circles) than the features in the experimental dataset.

The results in Figure 14 show metric scores that are consistent across slices, suggesting that our model behaves well for all slices. We claim that this property is important for a black-box predictor.

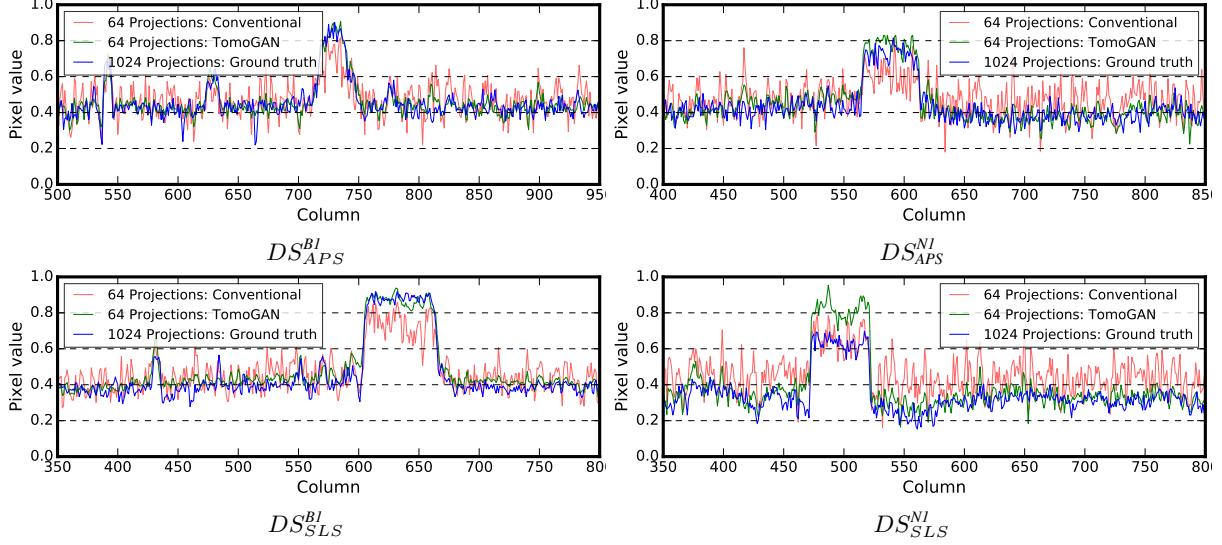


Figure 13: Pixel values of an arbitrarily chosen line in each of the four experimental datasets, subsampled to 64 projections.

3.4.2 Shorter exposure time

We also trained TomoGAN to enhance the quality of images reconstructed from short-exposure-time projections. First we applied Poisson noise [51] to the four experimental datasets to create new (simulated) short-exposure-time datasets. Next, we used one of the experimental datasets, DS_{APS}^{BI} , plus its associated short exposure datasets, to train TomoGAN. Then, we used the trained model to enhance the noisy images of the other three datasets.

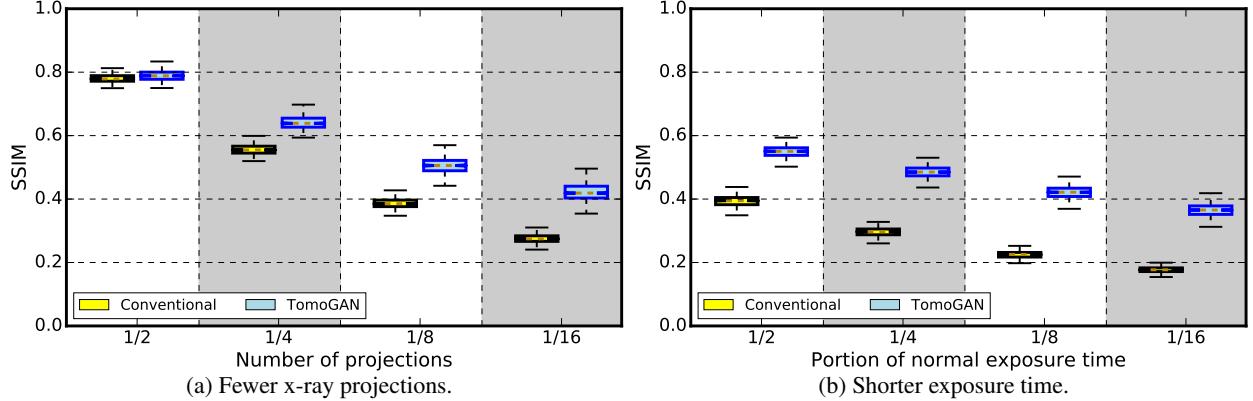


Figure 14: Per-slice SSIM similarities with a reconstruction of 1024-projection experimental data, for both conventional and TomoGAN-enhanced reconstructions and for different degrees of subsampling.

The qualities of the conventional and TomoGAN-enhanced reconstructions using SSIM are shown in Figure 14b. Comparing Figure 14 with Figure 11, we see that the improvement obtained for the experimental dataset is less than that obtained for the simulated dataset. The features in the experimental dataset are much more diverse (different shapes, sizes, and contrast) compared with those of the simulated dataset (only circles), and therefore the denoising performance is less for the experimental datasets.

We also show in Figure 15, for the most challenging cases that use only 1/16 of normal exposure time to reconstruct, the pixel value of a horizontal line that cross an arbitrarily chosen feature in each of the four datasets.

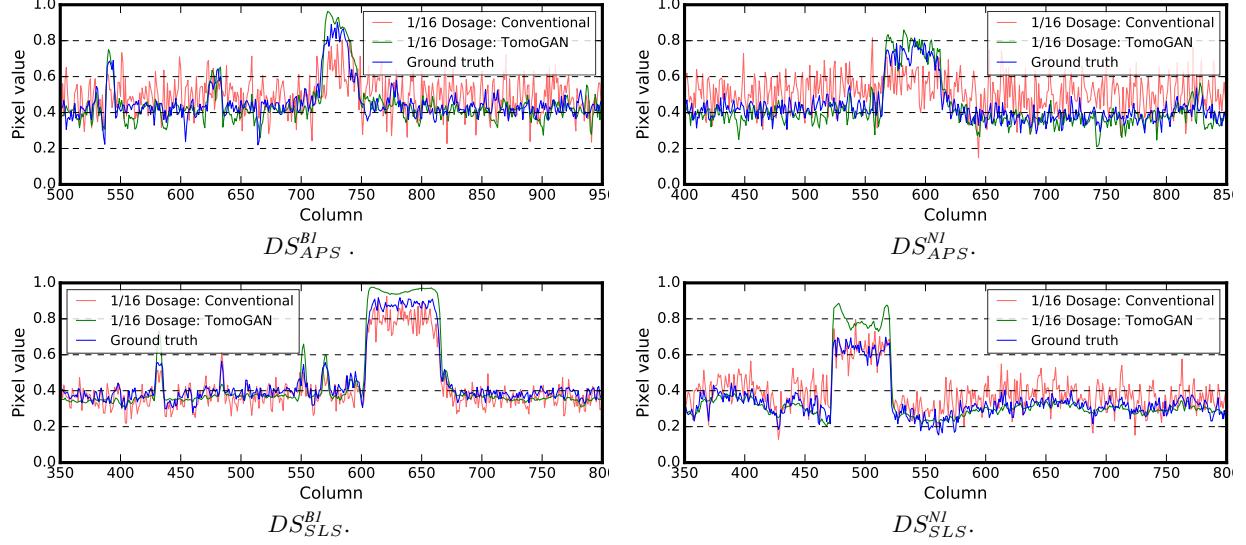


Figure 15: Pixel values of an arbitrarily chosen feature in each of the four experimental datasets, with projections generated by using 1/16 of the normal exposure time. Feature shapes are different for each dataset.

3.5 Comparison with iterative methods

Iterative methods are more resilient to noise in projections [56], but they are computationally demanding and prohibitively expensive for large data [57, 58, 59, 60]. The filtered back-projection algorithm takes 40 ms to reconstruct one slice (using the TomoPy [48, 49] toolkit), and TomoGAN takes 30 ms to enhance the reconstruction, for a total of about 70 ms per slice. In contrast, the simultaneous iterative reconstruction technique (SIRT) takes 550 ms to reconstruct one slice with 400 iterations. These times are all measured on one NVIDIA Tesla V100-SXM2 graphic card. Moreover, as illustrated in Figure 16, iterative reconstruction does not provide better image quality than that of our method.

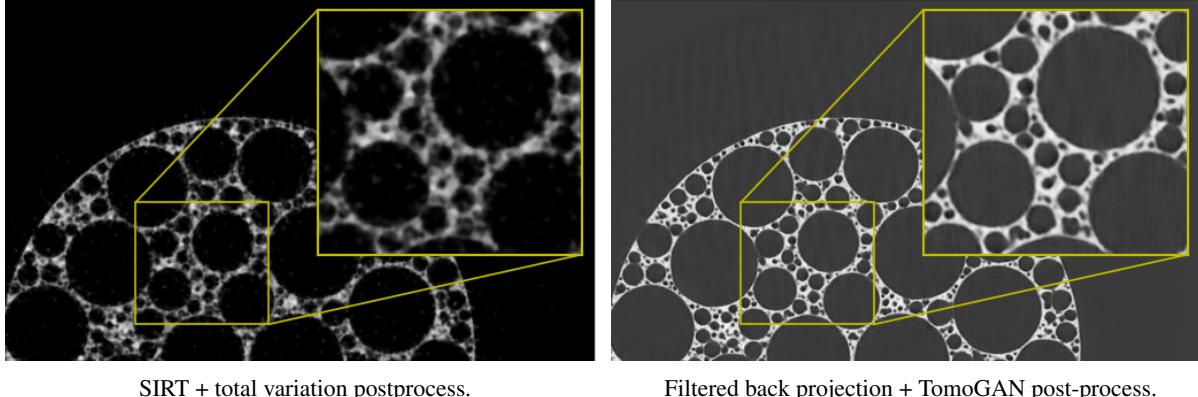


Figure 16: SIRT + total variation vs. TomoGAN: an image reconstructed from 64 simulated projections.

3.6 Comparison with other solutions on experimental datasets

Jin et al. [54] proposed a deep convolutional neural network (a variation of U-Net [61] named FBPConvNet) for the sparse view computed tomography problem. Their network outperforms total variation-regularized iterative reconstruction for more realistic phantoms, and also runs faster. FBPConvNet, like TomoGAN’s generator, uses the FBP algorithm to reconstruct sparse X-ray projections, after which the CNN combines multi-resolution decomposition and residual learning in order to remove artifacts while preserving image structure.

FBPConvNet and TomoGAN differ in two key areas. First, TomoGAN uses a GAN architecture, with adversarial loss, to help train the generator, and a pre-trained VGG [42] network; results presented below suggest that the adversarial

loss avoids artifacts. Second, our generator, although also based on U-Net, has three U-Net boxes, as shown in Figure 3 instead of the four in FBPCConvNet, and no batch normalization layer [20], two factors that reduce computation and memory needs for inference (e.g., TomoGAN can efficiently run on Google edge TPU [62]). For example, with one NVIDIA Tesla V100 card and a batch size of eight (minimizing DL framework overheads), TomoGAN and FBPCConvNet take an average of 30ms and 90ms to process one 1024×1024 image, respectively: TomoGAN is three times faster.

Yang et al. [29] use Wasserstein GAN [35] plus perceptual loss to denoise low-dose medical CT images, with perceptual loss calculated with a pre-trained VGG [42] neural network. Their model, which we refer to as WGAN-VGG here, differs from TomoGAN in two major respects. First, WGAN-VGG uses different generator and discriminator architectures: while TomoGAN uses a variant of the commonly used U-net architecture as its generator, WGAN-VGG's is handcrafted. Second, TomoGAN incorporates information from adjacent slices to enhance images. Results such as those shown in Figure 6 suggest that this use of multiple layers is beneficial in practice.

To permit quantitative comparison of TomoGAN with FBPCConvNet [54] and WGAN-VGG [29], we downloaded the open source implementations of both the latter models and trained them with our experimental datasets. Specifically, we trained both models with DS_{APS}^{BI} and tested on DS_{SLS}^{NI} . Figure 17 compares the SSIM and PSNR of FBPCConvNet, TomoGAN, and WGAN-VGG. TomoGAN outperforms FBPCConvNet on PSNR (higher median and smaller variance means more stable performance) but has a worse SSIM score. However, when we examine an arbitrarily chosen denoised

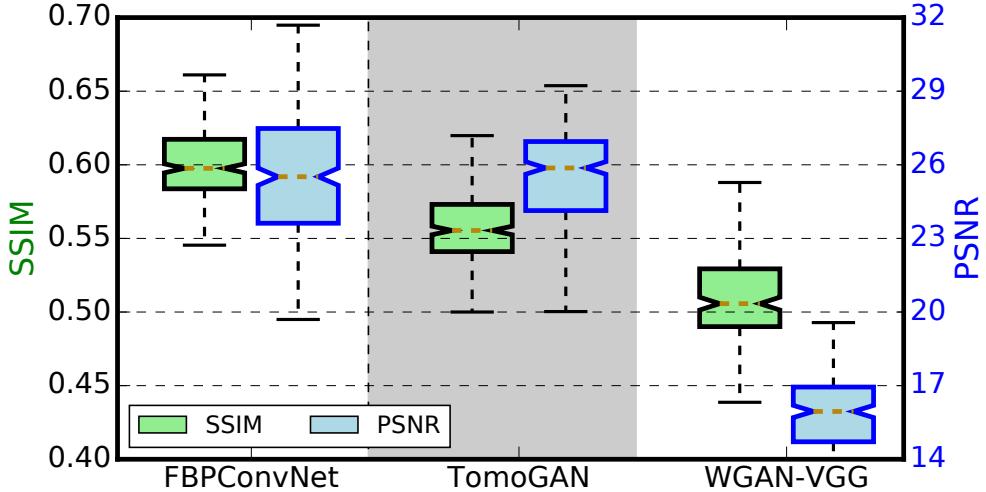


Figure 17: SSIM and PSNR comparison between FBPCConvNet, TomoGAN, and WGAN-VGG for a shale dataset.

image as shown in Figure 18, we see that FBPCConvNet introduces low-contrast artifacts as marked in Figure 18(a) (there are visibly more white dots in Figure 18(a) than in Figure 18(d)), likely because: (1) FBPCConvNet only used MSE loss and those artifacts were not significant to the MSE, and/or (2) as observed by the FBPCConvNet authors[15], FBPCConvNet has a high risk of overfitting. But these artifacts are significant to TomoGAN's adversarial loss and perceptual loss. WGAN-VGG has less good SSIM and PSNR values and, as shown in Figure 18(c), does not work well

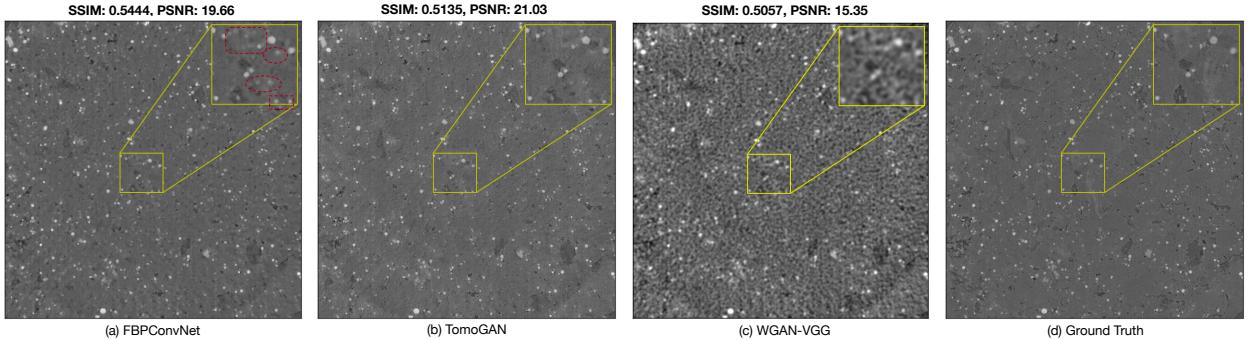


Figure 18: Denoising performance of FBPCConvNet, TomoGAN, and WGAN-VGG on an arbitrary slice of a shale sample.

for this dataset. We attribute this failure to the fact that the model and parameters are tuned for brain image datasets, which have quite different characteristics to those of our hard X-ray shale sample dataset, e.g., synchrotron hard x-ray images have much more high-frequency content. A model designed for one typically does not work well for the other. So, we note that results in Figure 18(c) is not a fair comparison to WGAN-VGG. However, we believe that this is not a well-understood fact in the literature, which furthermore includes much more work on deep learning for medical images. Thus, we view this comparison as a useful contribution, in that it shows that the medical image method cannot be used unchanged.

3.7 Model interpretation

A straightforward technique for understanding the working of a deep convolutional neural network is to examine the outputs of its various layers—its *feature maps*[63]—during the forward pass. Recall from Figure 3 that down-sampling reduces the image size to 1/8 of the original and increases the number of channels from d to 128. Successful training requires that the convolutional kernel retain the important features as the image size decreases. To qualitatively understand how the trained generator improves image quality, we show in Figure 19 the feature maps for an input (noisy) image, six representative channels from the bottom-most layer (the last layer of down sampling; red here indicates larger values and thus greater feature importance), and the final processed image. We focus on the bottom-most layer because we expect it to show the most refined or important features selected by the preceding convolution kernels.

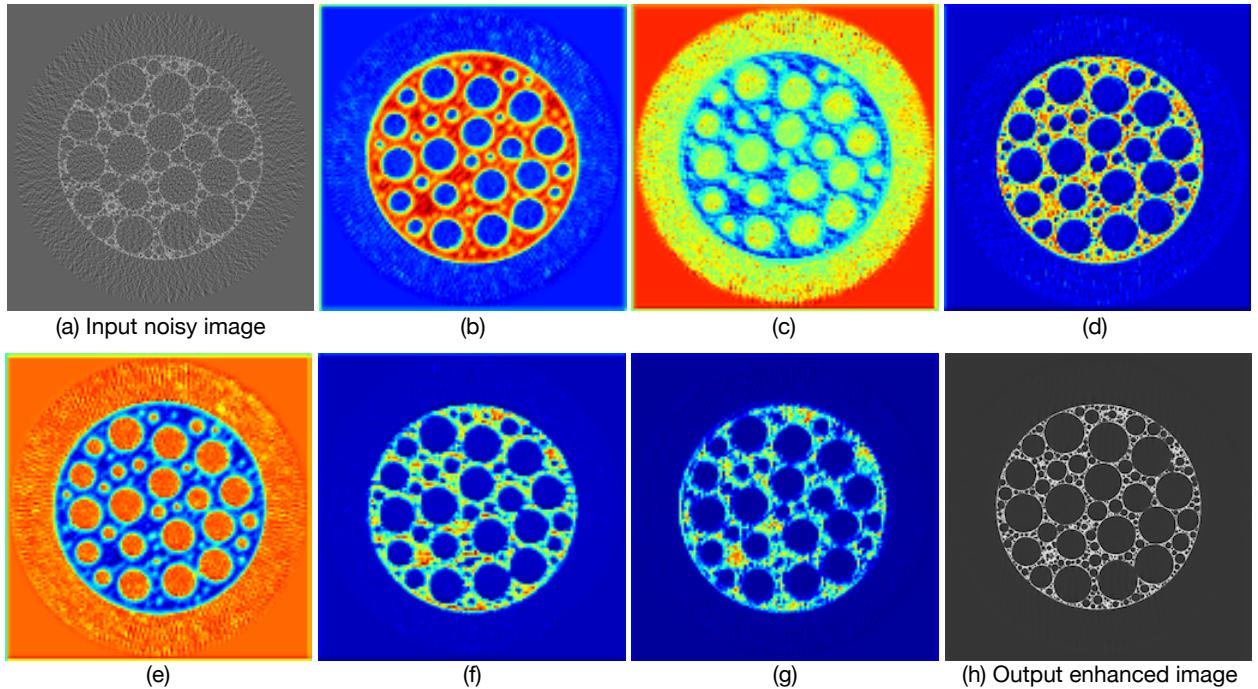


Figure 19: Feature maps from the TomoGAN network (Figure 3): (a) input (noisy) image; (b)–(g) six representative channels from the bottom-most layer; (h) final processed image.

Examining Figure 19(b)–(g) in detail, we see that (b) tends to keep dense regions of the sample (white in (h)), while (e) detects non-dense regions (air inside sample, black area in (h)); (c) detects the outer area in which there is no sample (should ideally be 0 in the output image); and (d) is similar to (b) but pays more attention to the details. Studying (h) closely, or (even better) examining Figure 1, which shows a zoomed-in version of the same image, we see that some areas contain many small features in the form of small holes. These small features could easily be lost in noise; interestingly, (f) and (g) seem to pay special attention to them. Intuitively, then, we may conclude that different channels (i.e., different convolution kernels) are responsible for capturing different types of features.

4 Discussion

We applied a generative adversarial network to improve the quality of images reconstructed from noisy x-ray tomographic projections. Experiments show that our model, TomoGAN, once trained on one sample, can then be applied effectively

to other similar samples, even if those samples are collected at a different facility and show different noise characteristics. Our results show that TomoGAN is general enough to provide visible improvement to images reconstructed from noisy projection datasets. We claim that this method has great potential for studying samples with dynamic features because it allows for good-quality reconstructions from experiments that run much faster than conventional experiments, either by collecting fewer projections or by using shorter exposure times per projection.

Acknowledgements

This work was supported in part by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. We thank Daniel M. Pelt at Centrum Wiskunde & Informatica for sharing the simulated x-ray projection datasets. We gratefully acknowledge the computing resources provided and operated by the Joint Laboratory for System Evaluation at Argonne National Laboratory.

References

- [1] U. Bonse and F. Busch, “X-ray computed microtomography (μ CT) using synchrotron radiation (SR),” *Progress in Biophysics and Molecular Biology*, vol. 65, no. 1-2, pp. 133–169, 1996.
- [2] J. Wang, H. Lu, T. Li, and Z. Liang, “Sinogram noise reduction for low-dose ct by statistics-based nonlinear filters,” in *Medical Imaging 2005: Image Processing*, vol. 5747. International Society for Optics and Photonics, 2005, pp. 2058–2067.
- [3] J. Wang, T. Li, H. Lu, and Z. Liang, “Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography,” *IEEE transactions on medical imaging*, vol. 25, no. 10, pp. 1272–1283, 2006.
- [4] A. Manduca, L. Yu, J. D. Trzasko, N. Khaylova, J. M. Kofler, C. M. McCollough, and J. G. Fletcher, “Projection space denoising with bilateral filtering and ct noise modeling for dose reduction in ct,” *Medical physics*, vol. 36, no. 11, pp. 4911–4919, 2009.
- [5] R. Anirudh, H. Kim, J. J. Thiagarajan, K. Aditya Mohan, K. Champliy, and T. Bremer, “Lose the views: Limited angle ct reconstruction via implicit sinogram completion,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [6] C. R. Vogel and M. E. Oman, “Iterative methods for total variation denoising,” *SIAM Journal on Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996.
- [7] M. Beister, D. Kolditz, and W. A. Kalender, “Iterative reconstruction methods in x-ray ct,” *Physica medica*, vol. 28, no. 2, pp. 94–108, 2012.
- [8] S. Bazrafkan, V. V. Nieuwenhove, J. Soons, J. D. Beenhouwer, and J. Sijbers, “Deep learning based computed tomography whys and wherefores,” *ArXiv*, vol. abs/1904.03908, 2019.
- [9] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, and S. Arridge, “Model-based learning for accelerated, limited-view 3-d photoacoustic tomography,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1382–1393, June 2018.
- [10] J. Adler and O. Öktem, “Deep bayesian inversion,” *arXiv preprint arXiv:1811.05910*, 2018.
- [11] J. Ma, J. Huang, Q. Feng, H. Zhang, H. Lu, Z. Liang, and W. Chen, “Low-dose computed tomography image restoration using previous normal-dose scan,” *Medical physics*, vol. 38, no. 10, pp. 5713–5731, 2011.
- [12] E. Kang, W. Chang, J. Yoo, and J. C. Ye, “Deep convolutional framelet denoising for low-dose CT via wavelet residual network,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1358–1369, June 2018.
- [13] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Işgum, “Generative adversarial networks for noise reduction in low-dose ct,” *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.
- [14] G. Wang, “A perspective on deep imaging,” *IEEE Access*, vol. 4, pp. 8914–8924, 2016.
- [15] D. M. Pelt, K. J. Batenburg, and J. A. Sethian, “Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks,” *Journal of Imaging*, vol. 4, no. 11, 2018. [Online]. Available: <http://www.mdpi.com/2313-433X/4/11/128>
- [16] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” *ArXiv e-prints*, Aug. 2016.

- [17] X. Yang, V. De Andrade, W. Scullin, E. L. Dyer, N. Kasthuri, F. De Carlo, and D. Gürsoy, “Low-dose x-ray tomography through a deep convolutional neural network,” *Scientific Reports*, vol. 8, no. 1, p. 2575, 2018. [Online]. Available: <https://doi.org/10.1038/s41598-018-19426-7>
- [18] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, p. 436, may 2015.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017. [Online]. Available: <http://doi.acm.org/10.1145/3065386>
- [20] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *32nd International Conference on Machine Learning*, vol. 37. JMLR.org, 2015, pp. 448–456. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3045118.3045167>
- [21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2627435.2670313>
- [22] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2016, pp. 770–778.
- [23] S. Anwar, S. Khan, and N. Barnes, “A deep journey into super-resolution: A survey,” *CoRR*, vol. abs/1904.07523, 2019. [Online]. Available: <http://arxiv.org/abs/1904.07523>
- [24] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” *ArXiv e-prints*, Nov. 2017.
- [25] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” in *IEEE Conference on Computer Vision and Pattern Recognition*, July 2017, pp. 105–114.
- [26] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5967–5976.
- [27] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.
- [28] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, “3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1522–1534, June 2018.
- [29] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, “Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, June 2018.
- [30] Y. Han and J. C. Ye, “Framing u-net via deep convolutional framelets: Application to sparse-view CT,” *CoRR*, vol. abs/1708.08333, 2017. [Online]. Available: <http://arxiv.org/abs/1708.08333>
- [31] M. Du, R. Vescovi, K. Fezzaa, C. Jacobsen, and D. Gürsoy, “X-ray tomography of extended objects: a comparison of data acquisition approaches,” *J. Opt. Soc. Am. A*, vol. 35, no. 11, pp. 1871–1879, Nov 2018. [Online]. Available: <http://josaa.osa.org/abstract.cfm?URI=josaa-35-11-1871>
- [32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.
- [33] D. J. Ching and D. Gürsoy, “XDesign: an open-source software package for designing X-ray imaging phantoms and experiments,” *Journal of Synchrotron Radiation*, vol. 24, no. 2, pp. 537–544, Mar 2017. [Online]. Available: <https://doi.org/10.1107/S1600577517001928>
- [34] S. van der Walt, J. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. Warner, N. Yager, E. Gouillart, T. Yu, and t. scikit-image contributors, “scikit-image: Image processing in python,” *PeerJ*, vol. 2, 07 2014.
- [35] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” *ArXiv e-prints*, Jan. 2017.
- [36] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of Wasserstein GANs,” *CoRR*, 2017. [Online]. Available: <http://arxiv.org/abs/1704.00028>
- [37] K. Nasrollahi and T. B. Moeslund, “Super-resolution: A comprehensive survey,” *Machine Vision and Applications*, vol. 25, no. 6, pp. 1423–1468, Aug 2014. [Online]. Available: <https://doi.org/10.1007/s00138-014-0623-4>
- [38] A. B. Tibbs, I. M. Daly, N. W. Roberts, and D. R. Bull, “Denoising imaging polarimetry by adapted bm3d method,” *J. Opt. Soc. Am. A*, vol. 35, no. 4, pp. 690–701, Apr 2018. [Online]. Available: <http://josaa.osa.org/abstract.cfm?URI=josaa-35-4-690>

- [39] C. Li, Y. Ma, J. Huang, X. Mei, and J. Ma, "Hyperspectral image denoising using the robust low-rank tensor recovery," *J. Opt. Soc. Am. A*, vol. 32, no. 9, pp. 1604–1612, Sep 2015. [Online]. Available: <http://josaa.osa.org/abstract.cfm?URI=josaa-32-9-1604>
- [40] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2016, pp. 1637–1645.
- [41] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for neural networks for image processing," *CoRR*, vol. abs/1511.08861, 2015. [Online]. Available: <http://arxiv.org/abs/1511.08861>
- [42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [43] R. Chard, Z. Li, K. Chard, L. T. Ward, Y. N. Babuji, A. Woodard, S. Tuecke, B. Blaiszik, M. J. Franklin, and I. T. Foster, "Dlhub: Model and data serving for science," *CoRR*, vol. abs/1811.11213, 2018. [Online]. Available: <http://arxiv.org/abs/1811.11213>
- [44] W. Kanitpanyacharoen, D. Parkinson, F. De Carlo, F. Marone, M. Stampanoni, R. Mokso, A. MacDowell, and H.-R. Wenk, *The Tomography Round-Robin datasets*.
- [45] W. Kanitpanyacharoen, D. Y. Parkinson, F. De Carlo, F. Marone, M. Stampanoni, R. Mokso, A. MacDowell, and H.-R. Wenk, "A comparative study of X-ray tomographic microscopy on shales at different synchrotron facilities: ALS, APS and SLS," *Journal of Synchrotron Radiation*, vol. 20, no. 1, pp. 172–180, Jan 2013. [Online]. Available: <https://doi.org/10.1107/S0909049512044354>
- [46] F. D. Carlo, D. Gürsoy, D. J. Ching, K. J. Batenburg, W. Ludwig, L. Mancini, F. Marone, R. Mokso, D. M. Pelt, J. Sijbers, and M. Rivers, "TomoBank: a tomographic data repository for computational x-ray science," *Measurement Science and Technology*, vol. 29, no. 3, p. 034004, feb 2018. [Online]. Available: <https://doi.org/10.1088%2F1361-6501%2Faa9c19>
- [47] W. van Aarle, W. J. Palenstijn, J. D. Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35 – 47, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0304399115001060>
- [48] D. Gürsoy, F. De Carlo, X. Xiao, and C. Jacobsen, "Tomopy: a framework for the analysis of synchrotron tomographic data," *Journal of synchrotron radiation*, vol. 21, no. 5, pp. 1188–1193, 2014.
- [49] D. M. Pelt, D. Gürsoy, W. J. Palenstijn, J. Sijbers, F. De Carlo, and K. J. Batenburg, "Integration of TomoPy and the ASTRA toolbox for advanced processing and reconstruction of tomographic synchrotron data," *Journal of Synchrotron Radiation*, vol. 23, no. 3, pp. 842–849, May 2016. [Online]. Available: <https://doi.org/10.1107/S1600577516005658>
- [50] B. Reiffen, "A note on "very noisy" channels," *Information and Control*, vol. 6, pp. 126–130, 1963.
- [51] F. E. Boas and D. Fleischmann, "CT Artifacts: Causes and Reduction Techniques," *Imaging Med*, vol. 4, no. 2, pp. 229–240, 2012.
- [52] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: A system for large-scale machine learning," in *12th USENIX Conference on Operating Systems Design and Implementation*, ser. OSDI'16. Berkeley, CA, USA: USENIX Association, 2016, pp. 265–283. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3026877.3026899>
- [53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, 2014. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [54] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.
- [55] Z. Liu, T. Bicer, R. Kettimuthu, and I. Foster, "Deep learning accelerated light source experiments," *arXiv preprint arXiv:1910.04081*, 2019.
- [56] L. L. Geyer, U. J. Schoepf, F. G. Meinel, J. W. Nance Jr, G. Bastarrika, J. A. Leipsic, N. S. Paul, M. Rengo, A. Laghi, and C. N. De Cecco, "State of the art: iterative ct reconstruction techniques," *Radiology*, vol. 276, no. 2, pp. 339–357, 2015.
- [57] S. T. Schindera, D. Odedra, S. A. Raza, T. K. Kim, H.-J. Jang, Z. Szucs-Farkas, and P. Rogalla, "Iterative reconstruction algorithm for ct: Can radiation dose be decreased while low-contrast detectability is preserved?" *Radiology*, vol. 269, no. 2, pp. 511–518, 2013, PMID: 23788715.

- [58] T. Bicer, D. Gürsoy, V. D. Andrade, R. Kettimuthu, W. Scullin, F. D. Carlo, and I. T. Foster, “Trace: a high-throughput tomographic reconstruction engine for large-scale datasets,” *Advanced Structural and Chemical Imaging*, vol. 3, no. 1, p. 6, Jan 2017.
- [59] X. Wang, A. Sabne, P. Sakdhnagool, S. J. Kisner, C. A. Bouman, and S. P. Midkiff, “Massively parallel 3d image reconstruction,” in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC ’17. New York, NY, USA: ACM, 2017, pp. 3:1–3:12.
- [60] T. Bicer, D. Gursoy, R. Kettimuthu, F. De Carlo, G. Agrawal, and I. T. Foster, “Rapid tomographic image reconstruction via large-scale parallelization,” in *European Conference on Parallel Processing*. Springer, 2015, pp. 289–302.
- [61] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [62] V. Abeykoon, Z. Liu, R. Kettimuthu, G. Fox, and I. Foster, “Scientific image restoration anywhere,” *arXiv preprint arXiv:1911.05878*, 2019.
- [63] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.

License

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy 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>.