

IntraTomo: Self-supervised Learning-based Tomography via Sinogram Synthesis and Prediction

Guangming Zang Ramzi Idoughi Rui Li Peter Wonka Wolfgang Heidrich
KAUST

Abstract

We propose IntraTomo, a powerful framework that combines the benefits of learning-based and model-based approaches for solving highly ill-posed inverse problems in the Computed Tomography (CT) context. IntraTomo is composed of two core modules: a novel sinogram prediction module, and a geometry refinement module, which are applied iteratively. In the first module, the unknown density field is represented as a continuous and differentiable function, parameterized by a deep neural network. This network is learned, in a self-supervised fashion, from the incomplete or/and degraded input sinogram. After getting estimated through the sinogram prediction module, the density field is consistently refined in the second module using local and non-local geometrical priors. With these two core modules, we show that IntraTomo significantly outperforms existing approaches on several ill-posed inverse problems, such as limited angle tomography with a range of 45 degrees, sparse view tomographic reconstruction with as few as eight views, or super-resolution tomography with eight times increased resolution. The experiments on simulated and real data show that our approach can achieve results of unprecedented quality.

1. Introduction

Computed tomography (CT) is a powerful imaging tool that enables the examination of internal structures of the scanned object in a non-destructive and non-invasive manner. The computed tomography technique is often associated with medical X-ray tomography. However, it has numerous other applications at different scales and involving various techniques to produce the projections such as: industrial X-ray CT [17, 69], synchrotron X-ray tomographic microscopy [19, 56], electron tomography microscopy [46] and visible light tomography [31, 70]. Tomography techniques operate on a set of projections of the scanned object from different angles. Reconstruction algorithms are applied to retrieve a density field representing the object from these projections. Mathematically, the reconstruction process is an

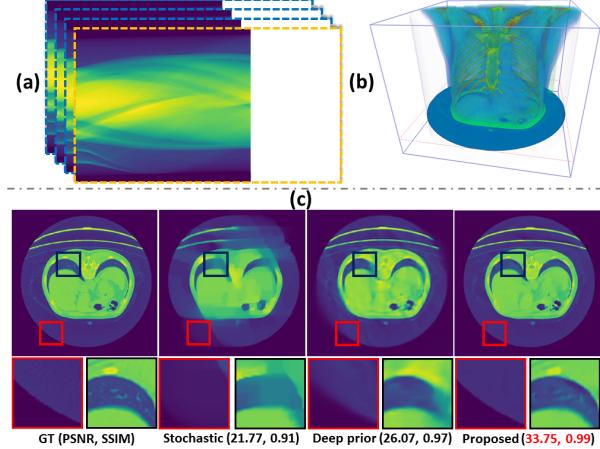


Figure 1. (a) Taking a limited angles CT measurements as input, (b) our **proposed** method successfully restores the 3D density field with both high-frequency details and geometric structures for a Covid19 patient (covid19-A-0070) [30]. (c) A comparison with other SOTA approaches using PSNR and SSIM metrics. A stochastic model-based approach [13] (**Stochastic**) fails to retrieve the field’s boundary (red box). Deep shape image prior [23, 62] (**Deep prior**) struggles to reconstruct fine details (black box).

inverse problem with N_v unknowns (number of pixels/voxels representing the object) calculated using the $N_\theta \times N_p$ known parameters (number of projections times the number of pixels for each projection). Ideally, the acquisition process provides a large number of projections that are uniformly sampled over the angular space.

However, in several applications, the data is captured for a limited number of projections or using a non-uniform angular sampling. For example, in medical CT, reducing the number of projections is employed as means to reduce the dosage of the X-ray radiation. On the other hand, in tilt-series electron tomography microscopy, the scanned sample can only be rotated for a limited range of angles, which results in a non-uniform angular sampling and a corresponding loss of resolution in the axial direction, also known as the “missing wedge problem”. The former acquisition scenario is known as *sparse-view* tomography, while the latter is known as *limited-*

angle tomography. These two scenarios illustrate examples where the tomography problem becomes highly ill-posed and underdetermined. In such scenarios, the commonly used analytical reconstruction Filtered Back-Projection (FBP) and its 3D variant FDK [21] become obsolete. Two families of approaches have been proposed in the literature to deal with the ill-posedness in those scenarios. The first one is the model-based optimization approaches that introduce non-linear regularizers (priors) [42, 73, 67, 20] to compensate for the missing information. These methods typically employ iterative solvers like the Algebraic Reconstruction Technique (ART) and its variants. They are flexible and achieve good results. However, they are memory-hungry and require substantial computational time. More recently, a second family of approaches emerged with the rise of deep learning techniques. For specific applications and scenarios, these approaches achieve outstanding performance [4, 45, 35]. Nevertheless, for supervised learning methods [2], the time for training and the difficulty in collecting enough labeled training data are issues in most applications.

In conventional deep learning applications to image analysis, the neural networks (NNs) are trained and tuned using arrays of pixels to represent the images. However, recent studies [54, 50] show that NN based approaches fail to recover the high frequency details of the targets [48, 50, 53, 60], since a rapid decay can be observed for the eigenvalue spectrum of these networks, which is called “spectral bias” [60, 54, 6]. To overcome the limitation of these biases, some coordinate-based neural approaches were proposed [50, 18, 7, 58]. By such powerful representations, a natural continuous version of an unknown target can be provided.

We incorporate recent coordinate-based neural approaches [50, 60], which act as a sinogram prediction prior, into our forward-backward optimization framework [8] called IntraTomo. By combining the sinogram prediction prior and geometric refinement priors, we successfully combine the benefits of model-based and data-driven techniques for tackling various tomographic imaging tasks.

In summary, the main contributions of our proposed approach are:

- We propose a novel self-supervised tomographic reconstruction framework that only uses the object’s own projections as training data. We are the first to successfully combine the benefits of learning-based and model-based approaches in a self-supervised fashion.
- We show state-of-the-art results on several challenging ill-posed tomography problems: limited angle tomography, sparse view tomography, and super-resolution tomography. We validate our improvements comparing to multiple state-of-the-art baseline methods using quantitative and qualitative metrics on a wide range of 2D and 3D, simulated and real CT data.

- We demonstrate our framework’s ability to solve inverse problems in the projections space, such as sinogram extrapolation, inpainting, super-resolution and denoising.

- The code and data are publicly available¹.

2. Related work

Tomographic reconstruction. Since its invention in the early 1970s, the Computed Tomography (CT) technique has been established as a non-invasive and non-destructive tool for inspecting scanned objects’ internal structures. It has applications in many varied fields, from medical diagnosis [64, 41], material and geosciences [10, 22, 72], industrial inspection [55] to security [47]. Computed tomography encompasses several approaches for reconstructing a density field representing the scanned object from the acquired projections (also called sinogram). The commonly used approach is Filtered Back-Projection (FBP) and its 3D cone beam variant, introduced by Feldkamp, Davis, and Kress (FDK) [21]. This method provides an analytical reconstruction based on the Radon transform and its inverse. The success of these transform-based approaches rests on their fast reconstruction times, as well as reasonable quality results when a sufficient number of projections is available. However, when the tomographic problem is highly ill-posed like in the sparse-view scenarios, these methods’ reconstruction quality is inferior. The second family of tomographic reconstruction methods is known as the iterative methods. These approaches, like The Algebraic Reconstruction Technique (ART) [28] and the Simultaneous Algebraic Reconstruction Technique (SART) [3], solve a discrete formulation of the reconstruction problem. These approaches can be combined with regularization terms in an optimization framework. This makes them well suited to the ill-posed tomographic problems like sparse views [36, 51], limited angles [37, 65], superresolution tomography [63, 67], or dynamic tomography [68, 69, 34]. With the rise of AI, several deep-learning approaches have been introduced to the computed tomography field. These methods follow three different strategies: (1) Preprocessing networks applied on the projections to reduce the ill-posedness of the reconstruction problem [4, 61]. (2) Postprocessing networks that improve the quality of the reconstructed object: [45, 26]. (3) Operator-based learning, where a differentiable forward model is included in the network [2, 40, 14]. One of the main downsides of these deep learning approaches is the large number of datasets needed for network training, which is very hard to obtain in practice. Moreover, it is difficult to transfer the network learning from one application domain to another since the CT data statistics are domain-specific

¹<https://github.com/vccimaging/IntraTomo>

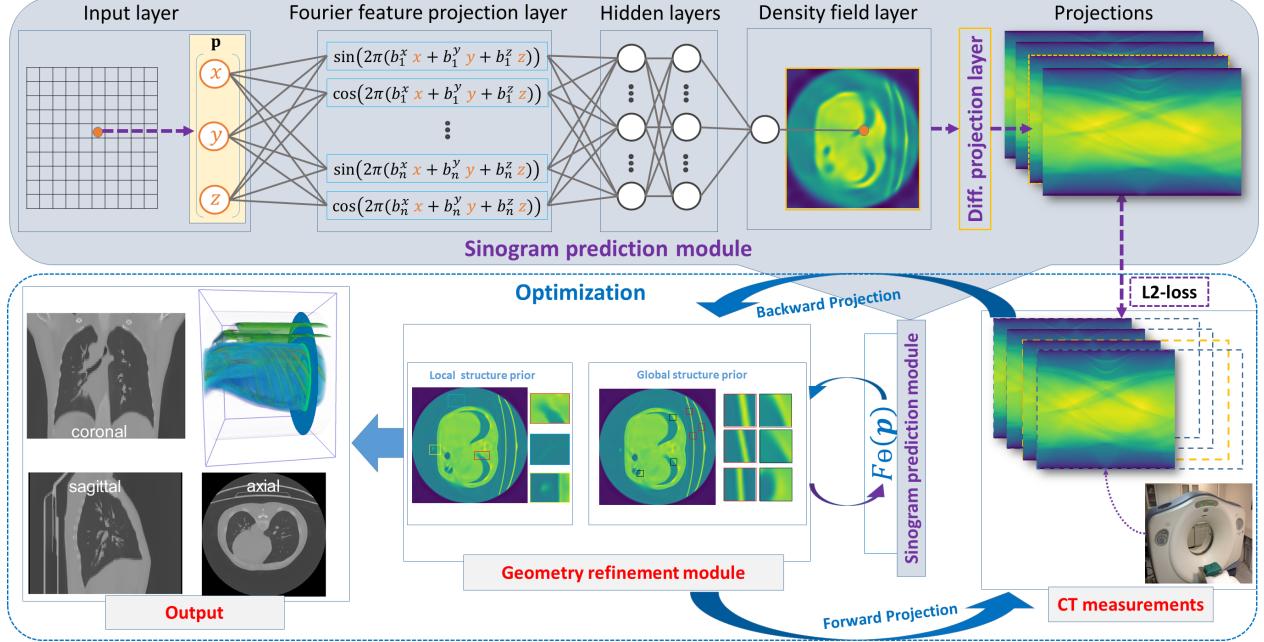


Figure 2. Overview of the architecture of our framework.

(for example, please see the Supplement for generalization experiments for FBPConvNet [39]). We propose a self-supervised learning-based (SSL) reconstruction approach that can handle most ill-posed scenarios without training on additional datasets.

Coordinate-based representations. In traditional computer vision tasks, the images and volumes are usually represented as arrays of pixels (2D) or voxels (3D). An emerging paradigm called coordinate-based representation (CBR) is proposed as an alternative to the discrete representation. CBR represents the signal of interest (image or volume) as a function of the coordinates, parameterized using a multiple-layer perceptron (MLP). In the literature, this implicit representation was already introduced to interesting applications, such as 3D shape reconstruction [25, 49], 3D scene modeling [12, 24, 66], and 2D vision applications like super-resolution [15], segmentation [71], and pixel synthesis [5] to name a few. Mildenhall et al. [50] propose the neural radiance fields (NeRF) based on a continuous implicit representation for view synthesis tasks. By leveraging a so-called positional encoding, and re-direction conditioning, they successfully reconstruct scenes with high-frequency details. Tancik et al. [59, 60] generalize the idea of positional encoding to more general Fourier features. We refer the readers to [18] for more details about coordinate-based representations.

3. Optimization framework

3.1. System overview

In different applications, hardware constraints or other considerations may lead to challenging CT scenarios like sparse views scanning, limited angle tomography, or super-resolution reconstruction. In the sparse views scanning, only few projections are captured. This can be motivated by reducing the acquisition time or limiting the radiation dosage received by a patient. In the limited angle scenario, the projections are captured only from a limited range of angles. This scenario occurs when there are restrictions on the relative rotation between the beam source and the scanned object. This scenario is typical for the Cryo-Electron Microscopy (Cryo-EM) setting. Finally, in the super-resolution problem, the desired resolution is higher than the resolution of the captured projection. Our proposed framework, entitled IntraTomo, tackles such highly ill-posed inverse CT problems flexibly. IntraTomo is composed of two complementary modules, which are iteratively applied: the sinogram prediction module and the geometry refinement stage (see Figure 2). The first module receives input CT measurements acquired according to one of the challenging CT scenarios described previously. Using a coordinate-based representation and a differentiable projection model, we train a fully connected multi-layer perceptron (MLP) in the sinogram prediction module. Trained in a self-supervised end-to-end fashion, this network allows for the estimation of a coordinate-based representation of the scanned object's

density field. Thus, new CT measurements can be synthesized in order to make the inverse problem well-posed. After this first step, the estimated density field is used to initialize the second module’s forward-backward optimization. This geometry refinement module aims to improve the density field’s quality iteratively by incorporating different geometry priors. We introduce a total variation and a non-local means regularizers to act respectively on a local and a global level in the current work. The obtained density field is used to acquire new synthetic projections. By adding these projections to the captured one, we get a new sinogram that we pass as an input to the sinogram prediction module, in order to improve the predicted result. After few iterations between the two modules of our framework, the result improves significantly. In the following, we present our framework in more detail.

3.2. Framework details

3.2.1 Sinogram prediction module.

The sinogram prediction module is a self-supervised end-to-end network for synthesizing the missing information from the acquired CT projections. It is comprised of four different (groups of) layers: the input layer, the Fourier feature projection layer, the hidden layers, and the differentiable projection layer.

Input layer. A given scanned object is usually represented by a 2D/3D discretized density field $\mathbf{V} \in \mathbb{R}^{N_v}$. Where N_v is the number of pixels/voxels that discretize the space. Note that, in the following, we will use the notations for the 3D case. Nevertheless, it is straightforward to apply our framework to reconstruct 2D density fields. We define $\mathbf{p}_i = (x_i, y_i, z_i)$ as the point in the coordinate domain associated to the i^{th} pixel/voxel. The main task of our sinogram prediction module is to learn a mapping function F_Θ between the coordinates $(\mathbf{p}_i)_{i=1..N_v}$ and the density field \mathbf{V} (see Equation 1). This is done by finding the best set of parameters Θ . Thus, our system’s input layer is the three-dimensional coordinate point \mathbf{p}_i .

$$F_\Theta(\mathbf{p}_i) = \mathbf{V}_i, \quad i = 1..N_v \quad (1)$$

Fourier feature projection layer. Fourier feature projection (FFP) is a technique firstly used in natural language processing and time series analysis, and recently introduced to the implicit representations [50, 60]. It consists of mapping low dimensional input point \mathbf{p}_i to a higher dimension vector $\mathcal{M}(\mathbf{p}_i)$, that encodes sinusoidal and cosinusoidal features (also called Gaussian random Fourier features). The use of this encoding layer is essential to retrieve high-frequency details for the signal to reconstruct (see Sec. 4.1). This mapping is given by the function $\mathcal{M}(.)$ as follows:

$$\mathcal{M}(\mathbf{p}_i) = [\sin(2\pi \cdot \mathbf{B} \cdot \mathbf{p}_i), \cos(2\pi \cdot \mathbf{B} \cdot \mathbf{p}_i)] \quad (2)$$

Here, $\mathbf{B} \in \mathbb{R}^{n \times 3}$ is a vector of frequencies sampled randomly from the Gaussian distribution $\mathcal{N}(0, \sigma^2)$. The standard deviation σ is a hyperparameter that controls which frequencies the network can represent. The reconstructed density fields may be over-smoothed if a lower value of σ is chosen, while a higher σ value may yield over-fitted and noisy results. In practice, this hyperparameter should be tuned for each different task.

Hidden layers. The hidden layers consist of a seven-layer MLP. Each of the six first layers has 256 neurons, while the last layer has only one neuron. A Swish activation function is applied between each two successive layers of this network. Indeed, we notice experimentally that in the highly under-determined inverse problems, such as the CT applications in this paper, a Swish non-linearity performs better than the ReLU function (please refer to the ablation study in Section 4). Finally, a Sigmoid activation is applied to the output of the last hidden layer. The inputs of this MLP are the Gaussian random Fourier features from the previous layer. The MLP predicts the corresponding density field \mathbf{V}_i at the location of the coordinate point \mathbf{p}_i from the input layer.

Differentiable CT projection layer. In this layer, we implement a differentiable projection model. Since the CT measurements that we have in the tomography problem are the projections (i.e., the sinogram), the main objective of this layer is to map the density field obtained from previous layers to the sinogram domain. We apply a differentiable CT projection operator in this layer that computes the Radon transform of the density field. In this layer, we project the density field according to a uniform and dense sampling of the angular range: $\theta \in [0, \pi[$ in parallel beam acquisition scenarios or $\theta \in [0, 2\pi[$ in the fan/cone beam scenarios. After training, this layer gives us complete control of the number and the sampling of the projections in the sinogram, making the synthesis of new projections possible to augment the captured CT measurements. Then the CT reconstruction problem can be well-posed. Furthermore, with this layer, we also have control over the projections’ resolution, which will be used in the super-resolution tomography scenarios.

Loss function. We use an L2-loss between the projections obtained after the differentiable CT projection layer and those acquired during the scan to train in the sinogram prediction module.

3.2.2 Geometry refinement module

By synthesizing new CT measurements, thanks to the sinogram prediction module, our framework restores to some extent the damaged or missing CT information in the angular/frequency domain. This leads to an improved recon-

structed density field. However, there is still space for refining the reconstruction in the spatial domain. To this end, we propose an optimization-based geometric refinement module. This module combines local and global structure priors into an optimization framework to refine the spatial domain's reconstructed density. This geometry refinement module can be expressed as the following optimization problem:

$$\mathbf{V}^* = \underset{\mathbf{V}}{\operatorname{argmin}} \underbrace{f(\mathbf{V})}_{\text{DFT}} + k_1 \cdot \mathbf{TV}(\mathbf{V}) + k_2 \cdot \mathbf{NLM}(\mathbf{V}) \quad (3)$$

local prior non-local prior

where $\mathbf{V} \in \mathbb{R}^{N_v}$ is the density field. $f(\cdot)$ is an L2 norm data fidelity term (DFT), which is composed of two parts:

$$f(\mathbf{V}) = f_{\text{captured}}(\mathbf{V}) + \mathbf{W} \cdot f_{\text{novel}}(\mathbf{V}) \quad (4)$$

These two parts of DFT measure respectively the discrepancy between the projections obtained from the density field \mathbf{V} and those captured by the CT scanner ($f_{\text{captured}}(\cdot)$), and obtained from the sinogram prediction module ($f_{\text{novel}}(\cdot)$). \mathbf{W} is a confidence-based weight matrix. Each of its entries is in the range $[0, 1]$ and corresponds to the weight associated with a given synthesized view from the sinogram prediction module. $\mathbf{TV}(\cdot)$ in Equation 3 corresponds to the total variation regularizer, chosen as the local structure prior. While $\mathbf{NLM}(\cdot)$ is the non-local means prior [11, 33, 1] which aims to regularize the geometry information at a global level to enhance global geometric information of the density, as illustrated in Figure 2. k_1 and k_2 are respectively the relative weights for the TV and the NLM regularizers [1] in the optimization. The reconstruction process is first initialized using the density field obtained as the output of the hidden layers (see Figure 2). Then, we iteratively update it using a forward-backward splitting solver [8]. The main advantage of using this optimization algorithm is its flexibility in incorporating different regularizers into the optimization framework. Thus, it is straightforward to later add other task-oriented regularizers, if needed.

4. Experimental results

Baselines. We compare our proposed method with six baseline techniques. The first method is the **FBP** reconstruction approach (i.e. FDK in 3D), since it is still the most popular tomographic reconstruction algorithm [52]. The second baseline is **SART**, a robust iterative reconstruction algorithm [38, 3], which is flexible and can be adapted to different camera models and applications [29, 32, 70]. A TV regularized algorithm [27] is used as the third baseline (**TV**). The stochastic primal-dual algorithm (**Stochastic**) [13] is chosen as the fourth baseline since it has decent performance on arbitrarily sampled data. The deep shape prior [23, 62] (**Deep prior**) is selected as the unsupervised state-of-the-art learning technique². Our method without geometry re-

finement optimization (**Ours (no g.r.)**) is included in the comparison as an ablation study to illustrate the impact of each module of IntraTomo. Finally, **Ours (full)** refers to our complete framework IntraTomo.

Implementation details. We implemented our framework in JAX [9] using the Operator Discretization Library (ODL) [1], to speed up the computing and to optimize the designed differentiable network. All experiments are conducted on a single RTX 2080Ti GPU. Six layers of 256 neurons each and a Swish activation between each layer are applied in the self-training process. We use the Adam optimizer with a learning rate of 10^{-3} in all experiments. The sizes of the evaluated density fields are 256×256 by default if no other specification is given. For additional experimental details and results, please refer to the supplementary material.

4.1. Ablation studies

Influence of Fourier feature projection. To demonstrate the importance of using the FFP layer in the sinogram prediction module, we run our framework with (W/) and without (W/o) the FFP layer. The data used in this experiment correspond to the sinogram of the PET phantom [13]. Only the first 120 angles are used from the sinogram in order to simulate the limited angle scenario. We also add a white Gaussian noise with a standard deviation of 5 to the sinogram. The obtained results provided in Figure 3 (top) show clearly the impact of the FFP layer. This layer allows a better reconstruction of the high-frequency features both for the density field and the sinogram. This results in a 6.38 dB PSNR improvement.

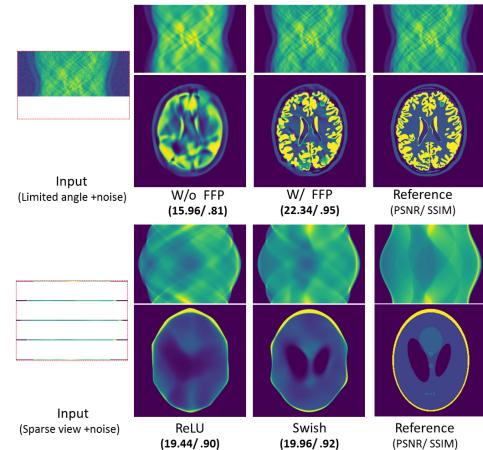


Figure 3. Ablation studies. On top we compare the network with and without Fourier features. The PSNR and SSIM values indicate a significant improvement when using Fourier features. On the bottom we compare the use of ReLU vs. Swish activation functions. The improvement in PSNR and SSIM values is more modest, but the visual difference is still very significant.

²<https://github.com/matheusgadelha/ShapeRecDeepPrior>

Influence of Swish activations. For highly ill-posed problems, the Swish activation function outperforms the ReLU function. We visually compare these two activation functions using a sparse view scenario with a sinogram of the Shepp-Logan phantom (SL) with only five captured projections and white Gaussian noise ($\text{std} = 1$) added to the data (see Fig. 3 bottom). Even though the use of Swish improves the PSNR only by 0.52 dB, we can visually notice a much better reconstruction quality compared to the reconstruction using ReLU.

Influence of geometric refinement priors. We demonstrate the advantages of applying geometric refinement priors in Figure 4 and Figure 5 (last three columns). Additional results and extensive analysis are provided in the supplement.

4.2. Applications

Sparse view tomography. We first evaluate IntraTomo on sparse view settings using the Shepp-Logan phantom [57] (SL), as well as data from the BrainWeb database [16], the ATLAS dataset [44], and Covid-19 dataset [30]. For the first dataset, only 8 projections are used as input. While for the two other datasets, we use 20 projections. As shown in Figure 4, the **FBP** method performs worst, as this analytical approach relies heavily on a large number of input measurements to retrieve good reconstruction results. The **TV**-regularized method outperforms the iterative **SART** algorithm. A sharper result with reasonable boundaries is obtained from the **deep prior** approach. However, reasonable geometrical details are missing with the **deep prior** approach. IntraTomo (**Ours (full)**) produces qualitatively better reconstructions, with a significant increase of the PSNR and the SSIM values as shown in Table 1 (left). Additional comparisons with the learning-based approach FBPCovNet [39] are reported in the supplemental materials. This technique consists in learning a network to improve the quality of an FBP sparse reconstruction. The network was trained using images of randomly generated ellipses and uses 50 projections. The comparisons in the supplement show that our method produces excellent quality results over a large set of different scene types, while FBPCovNet struggles to generalize to more diverse scenes and even has trouble with the SL dataset, which is comprised of ellipses at a different scale compared to their training data.

Limited angle tomography. We also evaluate our approach on limited angle tomography, which is even more challenging than the sparse view setting. To the best of our knowledge, we are the first to tackle the challenging case of angles being limited to the range of only 45 degrees with satisfying reconstruction results. Moreover, our framework does not need any pre-trained or additional labeled data. The reconstructed results are shown in Figure 5. The chosen

	Sparse view tomography		Limited angle tomography	
	Covid-19 [30]	Average-S	Pepper	Average-L
FBP [52]	11.92/.68	11.55/.42	12.63/.35	11.79/.42
SART [3]	18.48/.80	15.30/.52	12.55/.36	12.39/.46
TV [27]	21.61/.88	21.59/.83	15.40/.67	16.23/.70
Stochastic [13]	22.50/.90	22.16/.86	15.59/.69	16.69/.71
Deep prior [23]	23.55/.90	24.06/.87	18.12/.83	19.53/.80
Ours (no g.r.)	24.54/.92	24.53/.88	25.82/.97	25.53/.89
Ours (full)	27.19/.95	27.13/.93	28.40/.98	26.74/.91

Table 1. PSNR/SSIM measurements for each approach. Higher values are better for PSNR and SSIM. Besides the measurements for Covid-19 [30] and Pepper dataset, average measurements for all sparse view experiments (Average-S) in Figure 4 and all limited angle experiments (Average-L) in Figure 5 are also presented. Refer to supplementary for more numerical analysis.

limited angle settings are more ill-posed and challenging than the sparse view scenarios. For example, reconstructing the SL using 45 projection angles uniformly sampled in the range $[0, \frac{\pi}{4}]$ (see the first row of Figure 5) yields worse results than the case of 8 uniformly sampled projections in the range $[0, \pi[$ (see the first row in Figure 4). In Figure 5, for SL, LoDoPaB-CT data [43], Pepper, and Rose dataset, the iterative approaches (**SART** and **TV**) outperform the analytical solution (**FBP**), while the **deep prior** achieves the best performance among all baselines. IntraTomo, even without geometric refinement (**Ours (no g.r.)**), already reconstructs the density field in unprecedented quality. This result is further improved by the addition of the geometry refinement regularization for **Ours (full)**. The Table 1 shows that the highest PSNR and SSIM values are obtained with our reconstruction. Some of our experiments are done on 3D data. A corresponding 3D rendering is shown in Figure 6. We can see clearly that our reconstruction enables a good retrieving of the high-frequency features, like for the rose dataset, even though the input sinogram is incomplete.

Super-resolution tomography. Our coordinate-based representation can be used to reconstruct volumes in varying resolutions. We analyze this property by solving tomography reconstruction problems in the super-resolution setting. In the following, we generate a higher resolution density field using a network that is trained only on the low-resolution projections. We run this experiment on a super-resolution phantom [13], and a real scanned pepper (see Figure 6) captured from a Nikon XT H 255 scanner. This second dataset was captured in a cone-beam configuration and then reprojected according to a parallel-beam configuration. The sinogram prediction module is first trained using projections with fixed resolution. Then, the evaluation is started at eight times higher resolution. The results of these experiments are shown in Figures 7 and 8. As baseline methods, different methods are used to provide a super-resolution version of the sinogram. Then, the Rudin-Osher-Fatemi (ROF) approach

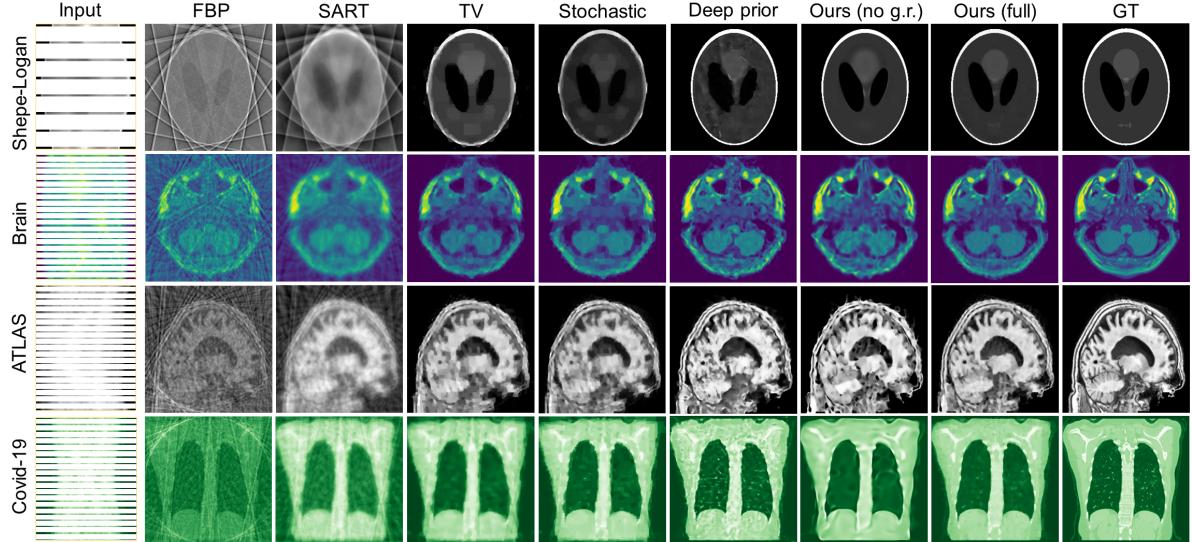


Figure 4. Sparse view tomographic reconstruction. From top to bottom: reconstruction with 8, 20, 20, and 20 sparse views over 180 degrees, respectively. PSNR(dB) and SSIM are shown for each approach in Table 1. For visibility, we rescaled the input sinogram and assigned different colormaps to different datasets.

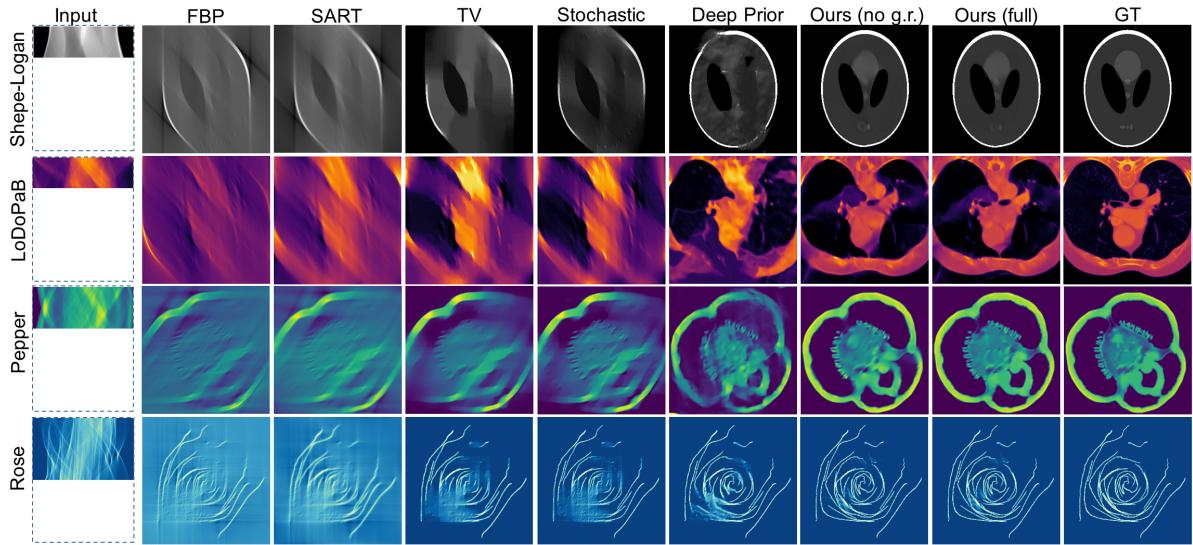


Figure 5. Limited angle tomographic reconstruction, where the input measurements are captured respectively from an angular coverage of 45°, 45°, 60°, and 90° for Shepe-Logan (first row), the LoDoPaB-CT dataset (second row), the Pepper dataset (third row), and the rose dataset (last row). PSNR(dB) and SSIM are shown in Table 1 for each approach. For better visualization, we rescaled the input sinogram and assigned different colormaps to different datasets.

[27] is applied to these super-resolved projections to reconstruct higher resolution density fields. Surprisingly, even if only 64 measurements from each angle are used in the training process, the Figure 8) shows that the high-resolution sinogram (512 measurements for each angle) evaluated on the model is very close to the ground truth sinogram. Furthermore, the Figure 7 shows clearly that our reconstruction is much sharper than the other baseline methods, with this

challenging super-resolution factor ($\times 8$).

Applications in the projection domain. Our framework's high performance in solving ill-posed problems can also be observed in the sinogram (projection) domain. Indeed, each of the challenging CT scenarios can be interpreted in the sinogram domain as a classical vision task. The sparse view setting can be seen as sinogram inpainting or interpola-



Figure 6. Reconstruction with CT scans. From left to right: CT reconstruction of a Pepper with angles limited to 60° , a rose flower with 90° , and a Covid19 patient (covid19-A-0000) [30] with 20 sparse views.

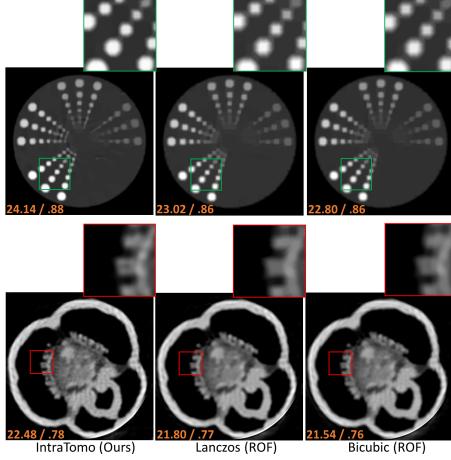


Figure 7. Examples of super-resolution tomography. From left to right: a IntraTomo trained on low-resolution sinogram with 64 measurements per view is evaluated in high-resolution with 512 measurements ($8 \times$ upsampling), a Lanczos upsampling, and a Bicubic upsampling of low-resolution sinogram with ROF reconstruction [27]. The (PSNR(dB)/ SSIM) measurements are given for each approach on the image.

tion problem. The limited angles tomography corresponds to a sinogram extrapolation task. Finally, super-resolution tomography coincides with the super-resolution problem on the projections. In Figure 8, we illustrate our framework’s results showing the sinograms and comparing them with other baseline methods. For these baseline methods, the full sinogram of 180 projections uniformly sampled in $[0, \pi]$ is obtained after reprojection of the reconstructed density field. The experiments in Figure 8 correspond to: sinogram inpainting with only 20 projections as input (uniformly sampled between $[0, \pi]$), a sinogram extrapolation using as input 45 projection in the range $[0, \frac{\pi}{4}]$, a super-resolution task with a factor of 8, and sinogram denoising from corrupted projections by white Gaussian noise (std = 15). More results can be found in the supplementary.

5. Conclusion and future work

We presented IntraTomo, a self-supervised learning-based framework for highly ill-posed CT imaging scenarios. We

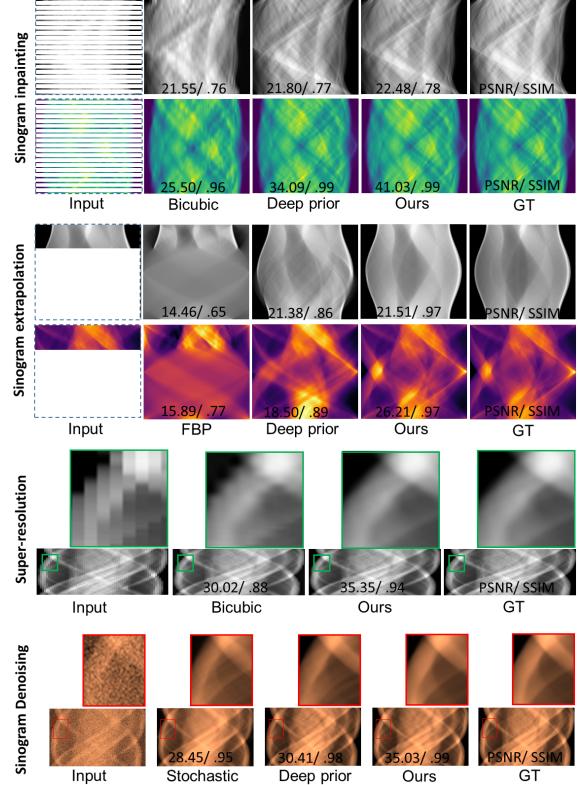


Figure 8. Visualizing the results in the projection domain. The (PSNR(dB)/ SSIM) measurements are given for each approach on the image.

demonstrated our framework on a wide variety of synthetic and real data, and we compared to multiple state-of-the-art approaches quantitatively and qualitatively. Our framework’s success relies on two novel modules: First, a sinogram prediction module that synthesizes and predicts missing projection images by representing the density fields with continuous and differentiable functions implicitly rather than discrete voxel grids. Second, we use a geometry refinement module, which includes local and global geometric priors, to improve the tomographic reconstruction in a forward-backward fashion. Since our IntraTomo framework has the flexibility of incorporating additional spatial and/or temporal priors, we propose to extend the framework to dynamic tomography problems in future work. Example applications are dynamic medical imaging and fluid capture. Another avenue of future work is to extend IntraTomo to the Cryo-EM setting.

6. Acknowledgements

This work was supported by KAUST as part of the VCC Competitive Funding. The authors would like to thank the anonymous reviewers for their insightful comments.

References

- [1] J. Adler, H. Kohr, and O. Oktem. Operator discretization library (odl). *Software available from https://github.com/odlgroup/odl*, 2017. 5
- [2] J. Adler and O. Öktem. Learned primal-dual reconstruction. *IEEE Trans. Med. Img.*, 37(6):1322–1332, 2018. 2
- [3] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic imaging*, 6(1):81–94, 1984. 2, 5, 6
- [4] R. Anirudh, H. Kim, J. J. Thiagarajan, K. A. Mohan, K. Champie, and T. Bremer. Lose the views: Limited angle CT reconstruction via implicit sinogram completion. In *Proc. CVPR*, pages 6343–6352, 2018. 2
- [5] I. Anokhin, K. Demochkin, T. Khakhulin, G. Sterkin, V. Lempitsky, and D. Korzhenkov. Image generators with conditionally-independent pixel synthesis. *arXiv preprint arXiv:2011.13775*, 2020. 3
- [6] R. Basri, M. Galun, A. Geifman, D. Jacobs, Y. Kasten, and S. Krithman. Frequency bias in neural networks for input of non-uniform density. In *International Conference on Machine Learning*, pages 685–694. PMLR, 2020. 2
- [7] S. Bi, Z. Xu, P. Srinivasan, B. Mildenhall, K. Sulkavalli, M. Hašan, Y. Hold-Geoffroy, D. Kriegman, and R. Ramamoorthi. Neural reflectance fields for appearance acquisition. <https://arxiv.org/abs/2008.03824>, 2020. 2
- [8] R. I. Boț and E. R. Csetnek. On the convergence rate of a forward-backward type primal-dual splitting algorithm for convex optimization problems. *Optimization*, 64(1):5–23, 2015. 2, 5
- [9] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. Vander-Plas, S. Wanderman-Milne, and Q. Zhang. JAX: composable transformations of Python+NumPy programs, 2018. 5
- [10] S. Brisard, M. Serdar, and P. J. Monteiro. Multiscale X-ray tomography of cementitious materials: A review. *Cement and Concrete Research*, 128:105824, 2020. 2
- [11] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *Proc. CVPR*, volume 2, pages 60–65. IEEE, 2005. 5
- [12] R. Chabra, J. Lenssen, E. Ilg, T. Schmidt, J. Straub, S. Lovegrove, and R. Newcombe. Deep local shapes: Learning local SDF priors for detailed 3D reconstruction. In *Proc. ECCV*, 2020. 3
- [13] A. Chambolle, M. J. Ehrhardt, P. Richtárik, and C.-B. Schönlieb. Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications. *SIAM Journal on Optimization*, 28(4):2783–2808, 2018. 1, 5, 6
- [14] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang. LEARN: Learned experts assessment-based reconstruction network for sparse-data CT. *IEEE Trans. Med. Img.*, 37(6):1333–1347, 2018. 2
- [15] Y. Chen, S. Liu, and X. Wang. Learning continuous image representation with local implicit image function. *arXiv preprint arXiv:2012.09161*, 2020. 3
- [16] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, G. B. Pike, and A. C. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. In *NeuroImage*. Citeseer, 1997. 6
- [17] L. De Chiffre, S. Carmignato, J.-P. Kruth, R. Schmitt, and A. Weckenmann. Industrial applications of computed tomography. *CIRP annals*, 63(2):655–677, 2014. 1
- [18] F. Dellaert. NeRF Explosion 2020. <https://dellaert.github.io/NeRF/>, 2020. 2, 3
- [19] P. C. Donoghue, S. Bengtson, X.-p. Dong, N. J. Gostling, T. Hultgren, J. A. Cunningham, C. Yin, Z. Yue, F. Peng, and M. Stampanoni. Synchrotron X-ray tomographic microscopy of fossil embryos. *Nature*, 442(7103):680–683, 2006. 1
- [20] J. Du, G. Zang, B. Mohan, R. Idoughi, J. Sim, T. Fang, P. Wonka, W. Heidrich, and W. L. Roberts. Study of spray structure from non-flash to flash boiling conditions with space-time tomography. *Proceedings of the Combustion Institute*, 38(2):3223–3231, 2021. 2
- [21] L. Feldkamp, L. Davis, and J. Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America*, 1(6):612–619, 1984. 2
- [22] A. Fly, Q. Meyer, M. Whiteley, F. Iacoviello, T. Neville, P. Shearing, D. J. Brett, C. Kim, and R. Chen. X-ray tomography and modelling study on the mechanical behaviour and performance of metal foam flow-fields for polymer electrolyte fuel cells. *international journal of hydrogen energy*, 44(14):7583–7595, 2019. 2
- [23] M. Gadelha, R. Wang, and S. Maji. Shape reconstruction using differentiable projections and deep priors. In *Proc. ICCV*, pages 22–30, 2019. 1, 5, 6
- [24] C. Gao, Y. Shih, W.-S. Lai, C.-K. Liang, and J.-B. Huang. Portrait neural radiance fields from a single image. <https://arxiv.org/abs/2012.05903>, 2020. 3
- [25] K. Genova, F. Cole, D. Vlasic, A. Sarna, W. Freeman, and T. Funkhouser. Learning shape templates with structured implicit functions. In *Proc. ICCV*, pages 7154–7164, 2019. 3
- [26] M.-I. Georgescu, R. T. Ionescu, and N. Verga. Convolutional neural networks with intermediate loss for 3D super-resolution of CT and MRI scans. *IEEE Access*, 8:49112–49124, 2020. 2
- [27] P. Getreuer. Rudin-Osher-Fatemi total variation denoising using split Bregman. *Image Processing On Line*, 2:74–95, 2012. 5, 6, 7, 8
- [28] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical Biology*, 29(3):471–481, 1970. 2
- [29] J. Gregson, M. Krimerman, M. B. Hullin, and W. Heidrich. Stochastic tomography and its applications in 3d imaging of mixing fluids. *ACM Trans. Graph.*, 31(4):1–10, 2012. 5
- [30] S. A. Harmon, T. H. Sanford, S. Xu, E. B. Turkbey, H. Roth, Z. Xu, D. Yang, A. Myronenko, V. Anderson, A. Amalou, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nature communications*, 11(1):1–7, 2020. 1, 6, 8
- [31] S. W. Hasinoff and K. N. Kutulakos. Photo-consistent reconstruction of semitransparent scenes by density-sheet decomposition. *IEEE Trans. PAMI*, 29(5):870–885, 2007. 1

- [32] F. Heide, J. Gregson, G. Wetzstein, R. Raskar, and W. Heidrich. Compressive multi-mode superresolution display. *Optics express*, 22(12):14981–14992, 2014. 5
- [33] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, et al. Flexisp: A flexible camera image processing framework. *ACM Trans. Graph.*, 33(6):1–13, 2014. 5
- [34] R. Heylen, G. Schramm, P. Suetens, and J. Nuyts. 4D CBCT reconstruction with TV regularization on a dynamic software phantom. In *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–3. IEEE, 2019. 2
- [35] D. Hu, J. Liu, T. Lv, Q. Zhao, Y. Zhang, G. Quan, J. Feng, Y. Chen, and L. Luo. Hybrid-domain neural network processing for sparse-view CT reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(1):88–98, 2020. 2
- [36] J. Huang, Y. Zhang, J. Ma, D. Zeng, Z. Bian, S. Niu, Q. Feng, Z. Liang, and W. Chen. Iterative image reconstruction for sparse-view CT using normal-dose image induced total variation prior. *PloS one*, 8(11):e79709, 2013. 2
- [37] Y. Huang, O. Taubmann, X. Huang, V. Haase, G. Lauritsch, and A. Maier. Scale-space anisotropic total variation for limited angle tomography. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2(4):307–314, 2018. 2
- [38] M. Jiang and G. Wang. Convergence of the simultaneous algebraic reconstruction technique (SART). *IEEE Trans. Image Proc.*, 12(8):957–961, 2003. 5
- [39] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Proc.*, 26(9):4509–4522, 2017. 3, 6
- [40] E. Kang, W. Chang, J. Yoo, and J. C. Ye. Deep convolutional framelet denosing for low-dose CT via wavelet residual network. *IEEE Trans. Med. Img.*, 37(6):1358–1369, 2018. 2
- [41] T. Kiljunen, T. Kaasalainen, A. Suomalainen, and M. Kortesniemi. Dental cone beam CT: A review. *Physica Medica*, 31(8):844–860, 2015. 2
- [42] H. Kudo, T. Suzuki, and E. A. Rashed. Image reconstruction for sparse-view CT and interior CT—introduction to compressed sensing and differentiated backprojection. *Quantitative imaging in medicine and surgery*, 3(3):147, 2013. 2
- [43] J. Leuschner, M. Schmidt, D. O. Baguer, and P. Maaß. The lodopab-ct dataset: A benchmark dataset for low-dose ct reconstruction methods. *arXiv preprint arXiv:1910.01113*, 2019. 6
- [44] S.-L. Liew, J. M. Anglin, N. W. Banks, M. Sondag, K. L. Ito, H. Kim, J. Chan, J. Ito, C. Jung, N. Khoshab, et al. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Scientific data*, 5(1):1–11, 2018. 6
- [45] Z. Liu, T. Bicer, R. Kettimuthu, D. Gursoy, F. De Carlo, and I. Foster. TomoGAN: low-dose synchrotron x-ray tomography with generative adversarial networks: discussion. *JOSA A*, 37(3):422–434, 2020. 2
- [46] V. Lučić, F. Förster, and W. Baumeister. Structural studies by electron tomography: from cells to molecules. *Annu. Rev. Biochem.*, 74:833–865, 2005. 1
- [47] D. Mery, D. Saavedra, and M. Prasad. X-ray baggage inspection with computer vision: A survey. *IEEE Access*, 8:145620–145633, 2020. 2
- [48] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger. Occupancy Networks: Learning 3D reconstruction in function space. In *Proc. CVPR*, June 2019. 2
- [49] M. Michalkiewicz, J. K. Pontes, D. Jack, M. Baktashmotlagh, and A. Eriksson. Implicit surface representations as layers in neural networks. In *Proc. ICCV*, pages 4743–4752, 2019. 3
- [50] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *Proc. ECCV*, 2020. 2, 3, 4
- [51] S. Niu, Y. Gao, Z. Bian, J. Huang, W. Chen, G. Yu, Z. Liang, and J. Ma. Sparse-view x-ray CT reconstruction via total generalized variation regularization. *Physics in Medicine & Biology*, 59(12):2997, 2014. 2
- [52] X. Pan, E. Y. Sidky, and M. Vannier. Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction? *Inverse problems*, 25(12):123009, 2009. 5, 6
- [53] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *Proc. CVPR*, pages 165–174, 2019. 2
- [54] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, pages 5301–5310. PMLR, 2019. 2
- [55] B. Sagbas and M. N. Durakbasa. Industrial computed tomography for nondestructive inspection of additive manufactured parts. In *Proceedings of the International Symposium for Production Research 2019*, pages 481–490. Springer, 2019. 2
- [56] A. Sakdinawat and D. Attwood. Nanoscale X-ray imaging. *Nature photonics*, 4(12):840, 2010. 1
- [57] L. A. Shepp and B. F. Logan. The Fourier reconstruction of a head section. *IEEE Transactions on nuclear science*, 21(3):21–43, 1974. 6
- [58] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. In *Proc. NeurIPS*, 2020. 2
- [59] M. Tancik, B. Mildenhall, T. Wang, D. Schmidt, P. Srinivasan, J. T. Barron, and R. Ng. Learned initializations for optimizing coordinate-based neural representations. <https://arxiv.org/abs/2012.02189>, 2020. 3
- [60] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 2020. 2, 3, 4
- [61] C. Tang, W. Zhang, Z. Li, A. Cai, L. Wang, L. Li, N. Liang, and B. Yan. Projection super-resolution based on convolutional neural network for computed tomography. In *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, volume 11072, page 1107233. International Society for Optics and Photonics, 2019. 2

- [62] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *Proc. CVPR*, pages 9446–9454, 2018. [1](#), [5](#)
- [63] W. van Aarle, K. J. Batenburg, G. Van Gompel, E. Van de Castelee, and J. Sijbers. Super-resolution for computed tomography based on discrete tomography. *IEEE Trans. Image Proc.*, 23(3):1181–1193, 2014. [2](#)
- [64] B. Van Ginneken, B. T. H. Romeny, and M. A. Viergever. Computer-aided diagnosis in chest radiography: a survey. *IEEE Trans. Med. Img.*, 20(12):1228–1241, 2001. [2](#)
- [65] M. Xu, D. Hu, F. Luo, F. Liu, S. Wang, and W. Wu. Limited angle X ray CT reconstruction using image gradient L0 norm with dictionary learning. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2020. [2](#)
- [66] A. Yu, V. Ye, M. Tancik, and A. Kanazawa. pixel-NeRF: Neural radiance fields from one or few images. <https://arxiv.org/abs/2012.02190>, 2020. [3](#)
- [67] G. Zang, M. Aly, R. Idoughi, P. Wonka, and W. Heidrich. Super-resolution and sparse view CT reconstruction. In *Proc. ECCV*, pages 137–153, 2018. [2](#)
- [68] G. Zang, R. Idoughi, R. Tao, G. Lubineau, P. Wonka, and W. Heidrich. Space-time tomography for continuously deforming objects. *ACM Trans. Graph.*, 2018. [2](#)
- [69] G. Zang, R. Idoughi, R. Tao, G. Lubineau, P. Wonka, and W. Heidrich. Warp-and-project tomography for rapidly deforming objects. *ACM Trans. Graph.*, 38(4):1–13, 2019. [1](#), [2](#)
- [70] G. Zang, R. Idoughi, C. Wang, A. Bennett, J. Du, S. Skeen, W. L. Roberts, P. Wonka, and W. Heidrich. TomoFluid: Reconstructing dynamic fluid from sparse view videos. In *Proc. CVPR*, pages 1870–1879, 2020. [1](#), [5](#)
- [71] H. Zhang, R. Wang, J. Zhang, C. Li, G. Yang, P. Spincemaille, T. Nguyen, and Y. Wang. NeRD: Neural Representation of Distribution for Medical Image Segmentation. *arXiv preprint arXiv:2103.04020*, 2021. [3](#)
- [72] P. Zhang, Y. I. Lee, and J. Zhang. A review of high-resolution X-ray computed tomography applied to petroleum geology and a case study. *Micron*, 124:102702, 2019. [2](#)
- [73] R. Zhang, J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh. Model-based iterative reconstruction for dual-energy X-ray CT using a joint quadratic likelihood model. *IEEE Trans. Med. Img.*, 33(1):117–134, 2014. [2](#)