ONIX: an X-ray deep-learning tool for 3D reconstructions from sparse views

YUHE ZHANG,1,* ZISHENG YAO,1, TOBIAS RITSCHEL,2, AND PABLO VILLANUEVA-PEREZ1

1Synchrotron Radiation Research and NanoLund, Lund University, Box 118, 221 00, Lund, Sweden
2University College London, WC1E 6BT London, UK
*yuhe.zhang@sljus.lu.se

Abstract: Three-dimensional (3D) X-ray imaging techniques like tomography and confocal microscopy are crucial for academic and industrial applications. These approaches access 3D information by scanning the sample with respect to the X-ray source. However, the scanning process limits the temporal resolution when studying dynamics and is not feasible for some applications, such as surgical guidance in medical applications. Alternatives to obtaining 3D information when scanning is not possible are X-ray stereoscopy and multi-projection imaging. However, these approaches suffer from limited volumetric information as they only acquire a small number of views or projections compared to traditional 3D scanning techniques. Here, we present ONIX (Optimized Neural Implicit X-ray imaging), a deep-learning algorithm capable of retrieving 3D objects with arbitrary large resolution from only a set of sparse projections. ONIX, although it does not have access to any volumetric information, outperforms current 3D reconstruction approaches because it includes the physics of image formation with X-rays, and it generalizes across different experiments over similar samples to overcome the limited volumetric information provided by sparse views. We demonstrate the capabilities of ONIX compared to state-of-the-art tomographic reconstruction algorithms by applying it to simulated and experimental datasets, where a maximum of eight projections are acquired. We anticipate that ONIX will become a crucial tool for the X-ray community by i) enabling the study of fast dynamics not possible today when implemented together with X-ray multi-projection imaging, and ii) enhancing the volumetric information and capabilities of X-ray stereoscopic imaging in medical applications.

© 2022 Optica Publishing Group

1. Introduction

As discovered by Roentgen in 1895, the high penetration power of X-rays makes them an excellent probe to access volumetric and internal information for a plethora of materials and systems. This property led to the development of three-dimensional (3D) techniques like tomography [1, 2] and confocal microscopy [3, 4], which have become standard probes to access structural information in a non-destructive manner for many and diverse disciplines [5–8], e.g., medicine, biology, chemistry, physics, cultural heritage, materials science, geoscience, and industrial processes. These state-of-the-art 3D X-ray imaging techniques access volumetric information by scanning the sample over different exposures, which is a time-consuming process. Thus, the scanning process may hamper the applicability of these 3D imaging approaches to i) 3D temporally-resolved studies [9, 10], where the scanning process takes longer than the studied dynamics or the scanning may induce forces that alter the studied dynamics, ii) experiments where the setup or configuration makes the scanning process difficult or impossible [11, 12], e.g., in medical surgeries and guidance approaches, and iii) dose-limited experiments [13, 14], for example medical applications or radiosensitive samples.

An alternative to avoid the scanning process is to use techniques that illuminate the sample simultaneously from different views, like X-ray stereographic methods [15], from only two
views, or X-ray multi-projection methods, using multiple views [16, 17]. These methods have been demonstrated and applied using diverse sources for different purposes. For example, they have been used together with laboratory or conventional X-ray sources [18] for positioning and guidance of medical surgeries or to capture fast dynamics [11, 12]. Lately, their application in conjunction with X-ray high-brilliance sources, i.e., X-ray sources with a high-coherent flux, such as diffraction-limited storage rings [19] and X-ray Free-electron Lasers (XFELs) [20–24] has been proposed and demonstrated [17, 25–28]. When applied at such facilities, these non-scanning techniques open up the possibility to access temporal and spatial resolutions not possible before. Moreover, they can obtain, for the first time, 3D information from single intense pulses from XFELs for stochastic and non-reproducible samples when operated in diffraction-before-destruction mode [29, 30], i.e., the sample is destroyed after being illuminated with a single XFEL pulse. Although stereographic and multi-projection approaches can overcome the aforementioned limitations of state-of-the-art 3D X-ray imaging approaches, they can only access limited volumetric via a sparse number of views or projections.

A solution to overcome the limited 3D information obtained by stereo or multi-projection imaging techniques is to add prior knowledge about our sample and system. Many frameworks have tried to address this issue, like compressed-sensing [31, 32], and Deep learning (DL) approaches [33]. Specifically, the latter has demonstrated its ability to generate complex 3D scenes or volumes from sparse views in computer vision [34]. Some of the most successful DL approaches to this problem rely on convolutional neural networks and representation of the 3D volume using regular grids [35] or irregular point clouds [36]. However, these representations and approaches have a large memory consumption limiting their applicability to small volumes, and so are not suitable for the aforementioned applications. Recently, neural implicit 3D representations [37, 38] have been developed to overcome this limitation by modeling 3D volumes at arbitrary resolution through a continuous space function. Although these implicit 3D representations have been a paradigm shift in computer vision and enable 3D reconstructions with only two-dimensional (2D) supervision [39–41], they have not made an impact on the X-ray community yet.

This paper presents ONIX (Optimized Neural Implicit X-ray Imaging) an implicit 3D DL approach for X-ray imaging that has the potential to retrieve arbitrary large resolution over 3D volumes from only a set of sparse 2D projections. ONIX uses only one or a small set of 2D radiographs or projections to achieve 3D volume reconstructions extending the capabilities of current implicit representations by: i) including an accurate physical description of the interaction of X-rays with matter via the projection approximation to impose self-consistency between the 3D reconstructed volume and the recorded radiographs, and ii) using a convolutional layer that enables transferring knowledge between similar X-ray multi-projection experiments to enhance the 3D reconstructions from limited views. The interaction of X-rays with matter includes not only the attenuation term but also the phase term making ONIX compatible not only with standard attenuation radiographs but also with coherent [42] and phase-sensitive techniques [43]. We demonstrate the capabilities of ONIX by exploiting experimental and simulated data using up to a maximum of eight views in realistic configurations for X-ray multi-projection imaging. The obtained results are compared to two grid approaches i) Simultaneous Algebraic Reconstruction Technique (SART) a traditional iterative reconstruction approach from sparse views [44], and ii) a state-of-the-art 3D supervised machine learning approach [45] (see subsection 2.5). Figure 1 depicts a schematic comparison between the three approaches that highlights the unique advantages of ONIX in terms of the inputs and outputs. From this comparison, we conclude that ONIX will enable 3D reconstructions from sparse views at a quality not possible today without any 3D supervision. Thus, it will become a tool that will expand the capabilities of current stereoscopic imaging techniques for medical applications and time-resolved experiments with laboratory sources, XFELs, and diffraction-limited storage rings.
This paper is structured as follows: First, we describe the ONIX approach and how the physics of the X-ray interaction with matter is accounted for. We also introduce a supervised 3D DL approach for comparison purposes. Second, we validate ONIX with synthetic and experimental data using up to a maximum of eight projections, and we compare its performance to state-of-the-art 3D reconstruction algorithms and 3D supervised machine learning. Finally, we discuss the results and future application of ONIX to enhance and enable 3D reconstructions from sparse projections coming from current X-ray stereoscopic and multi-projection approaches.

Fig. 1. Comparison and two main advantages of our approach over previous work. First, a) shows three 3D experiments with their corresponding 2D projections and how different approaches access this information. SART (blue) and other conventional 3D methods only access 2D projections from a single 3D experiment at a time. Typical DL methods access all the projections from different 3D experiments but require 3D supervision (red), i.e., access the 3D volume itself. ONIX (green) access all the projections from all 3D experiments, but without requiring 3D supervision at any point. b) The second unique point of our approach is that it does not require any grid but works directly on implicit functions.

2. Methods

2.1. X-ray propagation and interaction with matter

The interaction between matter and electromagnetic radiation is usually described via the refractive index ($n$). Unlike visible light, the refractive index for X-rays is very close to unity, and is usually expressed as a complex refractive index:

$$n = 1 - \delta + i\beta,$$

where $\delta$ and $\beta$ are positive and real quantities that denote the phase-shifting and the absorptive part of the X-ray refractive index. Due to the high penetration power and weak interaction of X-rays with matter, the values of $\delta$ and $\beta$ are small. This enables the projection approximation [46], which neglects diffraction within the sample volume. Under this approximation, the transmission of X-rays through a sample can be approximated by the exit wave immediately after the object:

$$\psi_{z_{\text{exit}}} = \psi_{z_0} \exp \left( -ik \int_{z=z_0}^{z=z_{\text{exit}}} (\delta[x, y, z] - i\beta[x, y, z])dz \right),$$

where $\psi_{z_0}$ is the incoming wave on the sample, $k = 2\pi/\lambda$ is the wavenumber which is inversely proportional to the wavelength ($\lambda$), $\delta$ and $\beta$ are a function of the transverse coordinates $x$ and $y$ orthogonal to the propagation direction $z$.

In the context of this work, we assume the projection approximation to describe the interaction of X-rays with matter. This approximation is used to transfer information between the 3D reconstructed volumes and the 2D recorded projections. It therefore enforces consistency between the 3D and 2D domains.
2.2. Simulated data

To demonstrate the capabilities and feasibility of ONIX, we generated a synthetic 3D dataset mimicking tomographic experiments and their measured projections. Each of the 3D objects consists of a collection of five to ten ellipsoids randomly positioned inside a cylinder. In our simulations, the ellipsoids were void, while the cylinder was made out of pure aluminum. We assumed the index of refraction of aluminum at 18 keV, where $\delta = 1.6741 \times 10^{-7}$ and $\beta = 6.4088 \times 10^{-9}$. The void ellipsoids were simulated assuming $n = 1$. The 3D volumes had $256 \times 256 \times 256$ voxels of $3.2 \, \mu m$. Each of the volumes contained the aluminum cylinder with a diameter of 100 voxels and a height of 256 voxels. Inside the cylinder, the ellipsoids were positioned with semi-axes randomly generated between 20 to 80 voxels. This synthetic dataset contains a total of 1000 of the previously described 3D objects. For each object, we simulated the absorption and phase contrast projections from eight equally-spaced angles between 0° and 140°. These projections were used as the input for the training of the DL methods.

2.3. Experimental data: metallic foams

We also validated the performance of ONIX using experimental data coming from X-ray tomography or 3D imaging on metallic foams. The X-ray tomographic experiments of metallic foams [47] made out of thixo, a well-established compound for metallic foams, were performed at the TOMCAT beamline of the Swiss Light Source [48]. Each of these acquisitions used the polychromatic spectrum provided by the TOMCAT’s bending magnet source. This means that each radiograph only contains attenuation contrast as the coherence was not sufficient to access the phase information. The radiographs were acquired with the GigaFRoST camera system that enabled 3D time-resolved acquisitions of the foaming process [49]. The camera system provided an effective pixel size of $2.75 \, \mu m$, and the radiographs had dimensions of $180 \times 960$ pixels. To speed up the training, we downsized the projections into $180 \times 256$ frames, reducing the number of voxels. A total of 168 tomographic experiments were used to train our DL approaches. From each of these experiments, we retrieved a 3D volume from a total of 96 radiographs with projection angles homogeneously distributed between 0 and 180 degrees. Given the limited number of projections (below the Crowther criterion), we used the SART [44] to retrieve the 3D volumes. SART provides high-quality tomographic reconstructions from limited views using an iterative approach based on linear algebra methods. To train our DL methods, we selected eight radiographs per volume with projection angles equally spaced between 0° and 131°, from all the acquired projections.

2.4. 3D unsupervised learning: ONIX

This subsection describes ONIX, our DL approach capable of reconstructing 3D volumes from only a set of sparse projections without accessing 3D information and with theoretically infinite resolution. First, we discuss the main features of ONIX. Then, we discuss the architecture and implementation details of ONIX and how it includes the physics of X-ray imaging and can learn across different samples and experiments to overcome the limitations of state-of-the-art 3D reconstruction approaches like SART. Finally, we describe the optimization process and specific parameters used for the aforementioned datasets.

Most 3D X-ray image reconstruction methods work on discrete representations, defined by grids of pixels and voxels. This limits the achievable resolution, tends to be computationally demanding, and consumes large amounts of memory. Though we are used to representing spatial information as grids, the physical world is primarily continuous, especially on the microscale. Thus, one may consider using continuous and differentiable functions instead of discrete representations to describe nature. This functional representation is called implicit representation [50]. Although implicit representations have an excellent potential for 3D X-ray imaging approaches, they have not been used, as finding a function that approximates a 3D sample of interest is a cumbersome
The advent of DL methods, such as deep neural networks, has opened new opportunities to address this problem by exploiting their potential as universal approximators. To approximate a 3D sample by a DL approach, we need to optimize (train) its parameters. This optimization is done by minimizing a cost function, which is usually called the "loss function" or "objective" of the neural network. By finding a proper neural network and optimization approach, we can use DL methods to infinitely approximate the shape or distribution of an object to overcome the aforementioned limitations of current discrete models.

ONIX is an implicit neural representation that retrieves the 3D complex index of refraction ($n$) as a function of the 3D spatial coordinate $x$ from a sparse number of views. ONIX has two main components, and its implementation is depicted in Fig. 2.

![Fig. 2. Overall scheme of ONIX. Four training views ($c_1,...,c_4$) are shown in the figure, where the first three are selected as constraints, i.e., views to condition the 3D model. The views used as constraints are encoded into latent space via a encoder ($E$). For each 3D spatial coordinate $x$, the predicted complex index of refraction ($n$) can be related to the local coordinates ($y_1, y_2, y_3$) of the encoded views and the corresponding latent vectors ($p_1, p_2, p_3$) via a fully connected neural network ($F$). The function between the 3D complex index of refraction ($n$) and the 3D spatial coordinate ($x$) is learned by optimizing the difference between the pixel value of the input image ($c_4(r)$) and the predicted line-integral of $n$ ($\tilde{c}(r)$) along each ray ($r$).]

The first part is an encoder ($E$) based on a 2D Convolutional Neural Network (CNN) that captures the information across the measured 2D views for different stereoscopic or X-ray multi-projection experiments. Thus, $E$ enables learning 3D objects across experiments in order to enhance the 3D information retrieved from sparse-view approaches. In each training iteration, we randomly encode a subset of the projections ($V$) from all measured ones ($V$). We call the subset of encoded views constraints, as they are used to constrain our 3D volume. $E$ takes as input the attenuation and phase-contrast images of each of the $m$-th views in that subset ($c_m$) and encodes them into a high-dimensional space also known as latent space ($E(c_m)$). The latent space provided by the $E(c_m)$ can be understood as a pixel image where every pixel contains a vector, also known as the latent vector, instead of a scalar as in a conventional single-channel image. To obtain the encoded information for any given 3D point $x$, we use the affine coordinate transformation $G$ to transfer from the global 3D coordinate system to the local coordinate system of each view, where the point coordinate of $x$ in the view $m$ local system is denoted by $y_m = G_m(x)$. In the local coordinate system for each view or detector position, the 3D volume is parametrized in terms of i) the distance between any point and the detector plane along the propagation direction of the X-rays that generated that view and ii) the intersection coordinates of that ray with the camera plane. This coordinate system simplifies the projection of
information between the encoded images and the object information in 3D. Our encoder $E$ is based on ResNet34 [51], which has three convolutional operations to extract key features and three pooling layers to reduce the dimension of the convolution results to keep the translation invariance of the image. The second component uses the information provided by each $E(c_m)$ and the geometrical information of the experimental arrangement.

The second component ($F$) is the main neural network. It reconstructs an object ($n(x)$) from the given constraints and prior information using a fully-connected neural network. Specifically, we use a fully-connected neural network with five layers, width 128, and residual connections [51], similar to the ones in Ref. [40]. This implementation is composed of three parts. The first part of $F$ has $v$ parallel weight-sharing ResNet 3× blocks. Each of these parallel blocks takes, for a spatial point $x$, the converted local coordinates $y_m$ from all constraints, and the corresponding latent vectors $p_m[y_m]$ that contain the projected information of $x$ into that view. Note that for each 3D spatial point, we only need the pixel-aligned latent vector from each constraint $p_m[y_m]$, not all the encoded information per view $E(c_m)$. The second part is just an average operator, which takes the mean of the processed information from the different views. The third part is made out of two ResNet blocks, introducing new learnable parameters to the averaged information.

Implementations, like the one used by ONIX, are known to be biased towards learning low-resolution or low-frequency functions [52]. It has also been demonstrated that mapping the inputs to a higher dimensional space allows the networks to learn better high-frequency functions [53]. For this reason, the coordinates are mapped from $\mathbb{R}$ to a high-dimensional space $\mathbb{R}^{2L}$ using positional encoding [39,54]. The positional encoding can be expressed in terms of a Fourier basis as:

$$\gamma(x) = \left(\sin\left(2^0 \pi x\right), \cos\left(2^0 \pi x\right), \ldots, \sin\left(2^{L-1} \pi x\right), \cos\left(2^{L-1} \pi x\right)\right),$$

where $L$ is a positive integer. In this work, we set $L = 10$. As a result, we express for each $x$ a generated $n$ or output of ONIX as:

$$n(x) = F(\gamma(y_1), p_1[y_1], \ldots, \gamma(y_m), p_m[y_m], \ldots, \gamma(y_v), p_v[y_v]).$$

With the 3D information of the refractive index $n(x)$, we can calculate the 2D projection images at any viewpoint. This transfer between 3D and 2D is done via ray tracing and the projection approximation as described in subsection 2.1. In this work, we deal with a fixed distance between the cameras and the sample and a parallel beam geometry, i.e., all the X-rays that generate a view have the same direction. However, this approach can be adapted quickly to other experimental setups, such as the cone-beam or fan-beam configurations, by modifying the geometric expressions that described the experimental configuration in our code. We parametrize the rays as $r(i) = o + rd$, where $i$ denotes the coordinate along the ray, $o$ and $d$ stand for ray origin and ray direction, respectively. Under the parallel-beam geometry, the origins correspond to the position of sensor pixels, and the ray directions are the same for each individual view. The 2D calculated rays and projections from the constraints are used to optimize the parameters of ONIX using the measured views $(c_v)$. This is done by optimizing the following loss function:

$$L = \sum_{r_i \in R} \|c_v(r) - \hat{c}(r)\|^2,$$

where $R$ is the set of all rays and $\hat{c}$ denotes the estimated phase or attenuation contrast, respectively. Due to computational limitations, we do not calculate the loss function on all of the pixel points or all possible rays of an image in the actual training. Instead, we select a set of ray sampling points or pixels from all input views and calculate the loss function on these sampling points. The rays are selected randomly over a probability density function based on the magnitude of the
spatial gradient of the input images. The calculated pixel values for phase and attenuation are obtained using a discrete version of the projection approximation:

$$\hat{c} = \log(\psi_z/\psi_{z0}) = \sum_{j=1}^{N} \Delta_j (-i k \delta_j - k \beta_j),$$

where $\Delta_j = t_{j+1} - t_j$ is the distance between two adjacent sampling points. The sampling points used along the ray are selected randomly. By calculating the squared $L_2$ norm between the line integral of the predicted $n$ and the pixel value from the inputs, ONIX can be trained to learn 3D self-consistency between different projections.

After the training, ONIX can generate 3D reconstructions using only the subset of constraints ($v$). As shown in Fig. 2, if we had four known views for each object, we could choose three of them (denoted by indices 1-3 in Fig. 2) as constraints ($v$) to encode their information. Please note that one can use an arbitrary number (smaller than the number of total input views) of views, as few as one, as the subset of constraints, depending on the actual situation. Having more constraints in training provides better results and increases computational demand. In this work, given that we had eight known views in total, we constrained our 3D reconstructions with four and six encoded views for the simulated and experimental datasets, respectively.

ONIX was implemented in PyTorch 1.6.0 and Python 3.8.8. The training was performed on an NVIDIA A100 GPU with 40 GB of RAM. For training of the simulation dataset, we chose 1024 ray sampling points from all input views of an object to calculate the loss along each ray. For each ray, 256 depth points were sampled. The number of sampling points can be flexibly adjusted based on hardware conditions. The Adam optimization algorithm [55] with a mini-batch size of two was used for the learning. We divided the training dataset into mini-batches because the dataset was too large to be loaded into the memory all at a time. We used a batch size of two, meaning that each mini-batch contained all known views of two objects or independent experiments, and the networks needed 500 training iterations to load all 1000 objects and finish an epoch, i.e., one complete training cycle. The step size of the optimizer, or the learning rate, was set to be 0.005 for the networks. We stopped the training after 500 epochs. This took ~24 hours. For the training of the experimental dataset, we selected 3096 ray sampling points from all input views, and 64 depth samples were chosen along each ray. As we only had attenuation contrast, the phase-contrast channel was removed from the output of the $F$, i.e., we had a single output channel for the attenuation reconstruction. We used the same initial learning rate as for the simulation dataset but reduced the learning rate by 0.1 after 1000 epochs. The training was stopped after 1500 epochs, which took ~12 hours.

### 2.5. 3D supervised learning

This subsection introduces a 3D-supervised baseline DL approach, which is used to assess the capabilities of ONIX, our non-3D unsupervised approach, in comparison to a DL approach that has access to 3D information. The 3D-supervised approach retrieved volumetric information by training and reconstructing stacks of individual 2D slices [45]. The paired dataset used for this supervised approach consisted of: i) input: 2D slices reconstructed from only eight views using SART, and ii) ground truth: 2D slices obtained either from the simulated volume or reconstructed from 96 views for the experimental data. Further details of the datasets and the data preparation can be found in subsections 2.2 and 2.3. The architecture of the DL was based on CNNs given their unique capabilities for image reconstruction and pattern recognition. Specifically, we used a U-NET [56], a CNN architecture that includes an image contracting path consisting of several downblocks, and an image expanding path consisting of the same amount of upblocks. In our task, the size of the input and output datasets was the same ($256 \times 256$). The optimization or loss function for this approach was an Mean-square-error (MSE) or $L_2$ norm...
between the ground-truth and the output of the U-NET. In order to minimize the overfitting, a $L_2$ regularization is implemented during the minimization process to penalize large values of the weights or fitted parameters of the U-NET architecture. The training of the U-NET was also performed on PyTorch 1.6.0 and Python 3.8.8 using an NVIDIA A100 GPU with 40 GB of RAM.

We should highlight that there were some slight differences between the training of the simulated and experimental data. First, the number of channels was different. The U-NET had two input channels for the simulated data as the attenuation and phase-information were available. For the experimental data, one channel was used, as only the attenuation channel was available. Second, the architecture or structure of the U-NET was different. For the simulated data, the two-channel U-NET started with 16 filters and contained a total of five downblocks and five upblocks. For the experimental data, the one-channel U-NET started with 32 filters and contained a total of six downblocks and six upblocks. Finally, the hyperparameters used in training were different. For the simulated data, the $L_2$-regularization was set as 0.001. For experimental data, the $L_2$-regularization was set as 0.005.

2.6. Data analysis

This subsection describes the approaches used to evaluate the performance of the different 3D reconstruction algorithms from a sparse number of projections: i) supervised 3D learning, ii) SART, and iii) our approach ONIX. We used two datasets to validate and analyze the capabilities of the three different approaches: i) simulated tomographic data using ellipsoid objects, described in subsection 2.2, and ii) experimental data using metallic foams as described in subsection 2.3. In this work, we used only eight projections to train and retrieve our 3D reconstructed volumes for all the reconstruction approaches presented. The reconstructed 3D volumes were compared to either the simulated 3D volume or the reconstructed volume from a large number of projections for the simulated and experimental datasets, respectively. This comparison was done using two different image quality metrics: $L_2$ norm and Dissimilarity Structure Similarity Index Metric (DSSIM), as shown in Table 1. These two metrics calculate the image correlation in the object space and better quality corresponds to smaller values of these two metrics. We applied a ring mask to remove the borders and keep only the relevant information for comparison purposes to all the reconstructed volumes and slices. This is a common practice to evaluate the performance of tomographic and 3D reconstruction algorithms. The same mask was used to produce the ground truth and the DL methods results. Furthermore, for the experimental dataset, we set a cutoff value for the masked outputs to select the region of interest for validation and improved visualization in the reference volumes. No threshold was applied to the SART reconstruction. For the 3D rendering, we applied a common threshold to the 3D reconstructions to visualize the metallic foams. Different thresholds were applied to the SART reconstructions to show the feature of interest for both simulation and experimental dataset.

3. Results and discussion

This section presents the results of ONIX when applied to simulated (subsection 3.1) and experimental data (subsection 3.2). For comparison purposes, we studied the reconstructions using state-of-the-art iterative reconstructions based on SART and a supervised 3D approach described in subsection 2.5. For simplicity, we refer in this section to the three methods as 3D supervised, SART, and ONIX. Finally, we discussed all the obtained results in the last subsection (subsection 3.3).

3.1. Simulation results

In this subsection, we applied ONIX on the simulated tomographic data using ellipsoid objects and compared the performance of ONIX with SART and 3D supervised learning.
Fig. 3. Tomographic reconstruction results for the simulation dataset. (a) The eight input projections that were used for the reconstructions. (b) The simulated ground truth. (c-e) Reconstruction results from (c) 3D supervised learning, (d) SART, and (e) ONIX. The input projections and the output results have both absorption- and phase-contrast components. We show only the absorption contrast results overlaid with the phase. For the ground truth and the three reconstruction methods, we depict the projections along an orthogonal system from the (i) top view, (ii) front view, and (iii) and left view, as well as the (iv) 3D rendered reconstructions.
Fig. 3 shows the inputs and outputs of the compared tomographic reconstruction approaches of a validation object, together with its ground truth. The validation object was not in the training dataset of the supervised learning, while its projections were in the training dataset of ONIX. As discussed in subsection 2.2, we used eight projection angles for the training of the supervised, and ONIX approaches. On the one hand, the supervised approach also accessed the simulated 3D volume in the training process as shown in Fig 1. On the other hand, ONIX and SART are fully unsupervised, i.e., no 3D information was accessed in any stage of learning or reconstruction, see Fig. 1. The ONIX results shown here were reconstructed using four views as constraints of the eight recorded projections. For the specific case of the SART reconstructions, we used the same eight projections used for the training and reconstructions. The projected absorption- and phase-contrast images used to obtain the depicted results are shown in Fig. 3 (a). To visually evaluate the reconstructions from the three different algorithms, we plotted for each reconstruction algorithm the projections of the 3D volume through three orthogonal axes denoted by (i) top, (ii) front, and (iii) left. We also rendered the 3D reconstructed volumes for all the algorithms and ground truth (iv). The results of these four evaluation images (i-iv) for the ground truth, 3D supervised, SART, and ONIX are shown in Fig 3(b-e), respectively. The absorption and phase reconstructions were very similar as the ratio of the real and imaginary parts of the index of refraction for a single material simulation were fixed, see subsection 2.2. For simplicity, we showed only the absorption contrast in Fig. 3. These results can be compared with the ground truth: simulated 3D objects. To quantitatively evaluate the 3D reconstructions from the three methods, we computed the $L_2$ norm and DSSIM of the results in Table 1. We validated the absorption- and phase-contrast volumes separately and then took the average of the two. A reader is referred to the supplementary material to compare the results over slices.

| Reconstructed results | Simulation dataset | Experimental dataset |
|-----------------------|--------------------|----------------------|
|                       | $L_2 \times 10^{-2}$ | DSSIM $\times 10^{-2}$ | $L_2\times 10^{-2}$ | DSSIM $\times 10^{-2}$ |
| 3D supervised         | 1.1                | 0.86                 | 5.5                | 3.3                |
| SART                  | 21                 | 7.5                  | 21                 | 16                 |
| ONIX                  | 5.1                | 1.6                  | 5.4                | 3.3                |

### 3.2. Experimental result

To demonstrate the capability of ONIX, we also applied it to the experimental metallic foam dataset collected from TOMCAT, an X-ray imaging instrument at the Swiss Light Source, as described in 2.3. The collected images in this experiment only contained absorption-contrast as the coherence of the experiment was not sufficient to retrieve phase-contrast information.

Fig. 4 shows (a) measured absorption-contrast images at eight equally-spaced projection angles, between 0° and 131° of the metallic foam, (b) SART-retrieved ground truth from 96 projections, and the corresponding reconstruction results from (c) 3D supervised learning, (d) SART, and (e) ONIX using six encoded views. Similarly, we plotted in different columns the (i) top, (ii) front, (iii) left projections through orthogonal axes, and (iv) side-view rendering of the 3D reconstructions. The results were rendered from the projection angle of 169°, which was not used for our reconstructions. The reader is referred to the supplementary material that visualizes the results of the reconstructions over orthogonal slices and provides a movie showing the comparison between the ground truth and the three reconstruction results. It can be seen from Fig. 4 that under our experimental conditions, SART with 96 projections can provide clean and
sharp 3D reconstructions, which is reliable enough for the supervised learning to learn from and for the other methods to be compared. The quantitative assessment for the experimental dataset is also shown in Table 1. In the supplementary material, we also compared the ONIX reconstruction results using four and six views to reconstruct our 3D volume and the results from all the different combinations of them. The results shown in Fig. 4(e) were the best reconstructions we found using six views among our investigation.

---

**Fig. 4.** Tomographic reconstruction results from different methods. (a) The eight input projections used for training and reconstructions. (b) The ground truth obtained from SART using 96 projections. (c-e) Reconstruction results from (c) 3D supervised learning, (d) SART, and (e) ONIX. The (i) top view, (ii) front view, (iii) left view, and (iv) 3D rendering images are plotted for the ground truth volume and the 3D reconstructions. The size of the top views was adjusted for better arrangement. The scale bar shown in (e) corresponds to 200 μm.
3.3. Discussion

This subsection studies the results presented in 3.1 and 3.2 to analyze the capabilities of ONIX by comparing it with state-of-the-art SART approach and a 3D supervised learning approach. It can be seen from Table 1 and the figures that, generally speaking, 3D supervised learning provides the most competitive results among the three experimented approaches. However, this approach requires accessing the 3D ground truth during training, limiting its applicability to techniques and scenarios where it is possible to acquire 3D information. Thus, this approach can be conceived as a simple denoising approach rather than an enabling technique for sparse acquisition approaches. On the other hand, ONIX taking no 3D supervision i) clearly outperforms state-of-the-art iterative reconstruction like SART using the same information, and ii) performs almost at the same level as the 3D supervised approach. If one compares the simulated and experimental results, it is noticeable that the 3D supervised learning results for the experimental dataset have inferior quality to the results of the simulations. This difference may arise from two factors. First, this could partially be because the ground truth is downsized due to the memory limitation of the computing resources. Thus, the benefit of ONIX having infinite resolution becomes apparent, as the downsizing of the input projections has less impact than it does for 3D supervised learning. Second, the noise contribution of the simulated and experimental data are unrelated. The performance of ONIX depends on the number of available views. Although ONIX implementation can use a random number of input views and as low as one, we have focused on eight views for this article based on current designs for X-ray multi-projection imaging techniques [17, 27]. We have observed that even using the same number of projections, ONIX provides better results for the experimental data when six views are used instead of four, as shown in the supplementary material. Thus, the performance of ONIX can be further improved by encoding with more views given the same amount of projections. We have also found that the performance of ONIX is independent of the permutation of views used as constraints but dependent on the combination of them. In general, the reconstruction obtained from closely spaced projection angles is not as good as the ones obtained from widely spaced angles. The best reconstructions were obtained when choosing widely spaced projections as constraints, which is the case for results shown in 3.1 and 3.2. Although we have only applied ONIX on objects with known 3D information, we have demonstrated its capabilities and potential when using a subset of sparse projections that emulates realistic conditions for X-ray multi-projection imaging experiments. The unique capabilities of ONIX to enhance 3D reconstructions to levels previously impossible from only a sparse set of views are the result of including the physics of image formation with X-rays and transferring prior knowledge between different objects and experiments [40, 41]. We envision that ONIX will become an enabling tool for 3D sparse X-ray imaging techniques like time-resolved X-ray multi-projection imaging experiments and medical stereoscopy.

4. Conclusion

To conclude, we have presented ONIX, a 3D unsupervised DL approach for X-ray imaging that can retrieve high-quality 3D reconstructions from an arbitrary set of sparse projections. ONIX extends the capabilities of current 3D volume reconstruction approaches from sparse techniques by i) offering an implicit representation with a theoretically infinite resolution that overcomes the memory demands of approaches like conventional voxel representations, ii) including the physical description of the interaction of X-rays with matter, enables an accurate transfer of the information between the 3D model and the recorded 2D projections, and iii) including a convolutional neural network that transfers prior knowledge between experiments and samples to enhance the performance of this approach. ONIX can retrieve, in general, the complex index of refraction, making our approach not only compatible with traditional absorption-contrast radiographs but also phase-contrast or coherent imaging techniques. To demonstrate the capabilities of ONIX,
we have compared it to state-of-the-art iterative reconstructions (SART) and 3D supervised machine learning approaches using only eight views under realistic conditions for X-ray multi-projection imaging experiments. We have observed with simulated and experimental data that ONIX outperforms current iterative approaches and performs at the level of 3D supervised approaches. However, 3D supervised approaches are only presented for comparison purposes as these methods cannot be applied to experiments with sparse 2D views as the 3D information required for the training is not available. Therefore, we envision ONIX as a tool that will enable 3D reconstructions for X-ray imaging at arbitrary resolutions from a sparse set of views with a quality that was not previously possible. Specifically, we envision its application to expand i) the capabilities of time-resolved experiments not possible today that exploit the unique properties and high-flux of XFELs and diffraction-limited storage rings, and ii) the information provided by stereoscopic techniques for different fields such as medical applications.

The ONIX code is available at GitHub.

Acknowledgments. We thank F. Garcia-Moreno, Mike Noack, and Paul H. Kamm from the Helmholtz-Zentrum Berlin for providing the metallic foam data used for the experimental validation and insightful discussions about this data. We wish to acknowledge the insightful comments of Elizabeth Blackburn and Gary Harlow. We are grateful to Z. Matej for his support and access to the GPU-computing cluster at MAX IV. We are also grateful to the European XFEL for its support, especially the computing and data analysis groups, for granting access to the GPU-computing cluster. We also gratefully acknowledge the support of NVIDIA Corporation with the donation of a Quadro P4000 GPU used for this research.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References
1. A. M. Cormack, “Representation of a function by its line integrals, with some radiological applications. II,” J. Appl. Phys. 35, 2908–2913 (1964).
2. G. N. Hounsfield, “Computerized transverse axial scanning (tomography): I. Description of system,” Br. J. Radiol. 46, 1016–1022 (1973).
3. P. Davidovits and M. D. Egger, “Scanning laser microscope,” Nature 221, 831 (1969).
4. A. Ale, V. Ermolayev, E. Herzog, C. Cohrs, M. H. de Angelis, and V. Ntziachristos, “FMT-XCT: in vivo animal studies with hybrid fluorescence molecular tomography–X-ray computed tomography,” Nat. Methods 9, 615–620 (2012).
5. L. Salvo, P. Cloetens, E. Maire, S. Zabler, J. J. Blandin, J. Y. Buffière, W. Ludwig, E. Bollier, D. Bellet, and C. Josserond, “X-ray micro-tomography an attractive characterisation technique in materials science,” in Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms, vol. 200 (2003), pp. 273–286.
6. C. J. Galbán, M. K. Han, J. L. Boes, K. A. Chughtai, C. R. Meyer, T. D. Johnson, S. Galbán, A. Rehemtulla, E. A. Kazerooni, F. J. Martínez, and B. D. Ross, “Computed tomography–based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression,” Nat. Medicine 18, 1711–1715 (2012).
7. A. Ale, V. Ermolayev, E. Herzog, C. Cohrs, M. H. de Angelis, and V. Ntziachristos, “FMT-XCT: in vivo animal studies with hybrid fluorescence molecular tomography–X-ray computed tomography,” Nat. Methods 9, 615–620 (2012).
8. P. J. Withers, C. Bouman, S. Carmignato, V. Cnudde, D. Grimaldi, C. K. Hagen, E. Maire, M. Manley, A. Du Plessis, and S. R. Stock, “X-ray computed tomography,” Nat. Rev. Methods Primers 1, 18 (2021).
9. F. García-Moreno, P. H. Kamm, T. R. Neu, F. Bühl, R. Mokso, C. M. Schlepütz, M. Stampanoni, and J. Banhart, “Using X-ray tomoscopy to explore the dynamics of foaming metal,” Nat. Commun. 10, 3762 (2019).
10. R. F. Ziesche, T. Aftl, D. P. Finegan, T. M. M. Heenan, A. Tengattini, D. Baum, N. Kardjilov, H. Markötter, I. Manke, W. Kockelmann, D. J. L. Brett, and P. R. Shearing, “4D imaging of lithium-batteries using correlative neutron and X-ray tomography with a virtual unrolling technique,” Nat. Commun. 11, 777 (2020).
11. J. Evans and H. Hon, “Dynamic stereoscopic x-ray imaging,” NDT & E Int. 35, 337–345 (2002).
12. D. Verellen, G. Soetens, N. Linthout, S. Van Acker, P. De Roover, V. Vinh-Hung, J. Van de Steene, and G. Storme, “Quality assurance of a system for improved target localization and patient set-up that combines real-time infrared tracking and stereoscopic x-ray imaging,” Radiother. Oncol. 67, 129–141 (2003).
13. 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 Med. Imaging 25, 1272–1283 (2006).

14. E. Zanette, M. Bech, A. Rack, G. Le Duc, P. Tailfoure, C. David, J. Mofr, F. Pfeiffer, and T. Weitkamp, “Trimodal low-dose x-ray tomography,” Proc. Natl. Acad. Sci. 109, 10199–10204 (2012).

15. E. Thomson, “Stereoscopic roentgen pictures,” Elec Eng 2 (1896).

16. M. R. Howells and C. J. Jacobsen, “Possibility for one-shot tomography using a high-gain free-electron laser,” Work. on scientific applications coherent X-rays SLAC-R-437, 159–162 (1994).

17. P. Villanueva-Perez, B. Pedrini, R. Mokso, P. Vagovic, V. A. Guzenko, S. J. Leake, P. R. Willmott, P. Oberta, C. David, H. N. Chapman et al., “Hard x-ray multi-projection imaging for single-shot approaches,” Optica 5, 1521–1524 (2018).

18. R. D. Moseley and J. R. Williams, “Apparatus for stereoscopic chest radiography,” Radiology 78, 103–108 (1962). PMID: 14476439.

19. M. Eriksson, J. F. van der Veen, and C. Quitmann, “Diffraction-limited storage rings – a window to the science of tomorrow,” J. Synchrotron Radiat. 21, 837–842 (2014).

20. B. W. J. McNeil and N. R. Thompson, “X-ray free-electron lasers,” Nat. Photonics 4, 814–821 (2010).

21. P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F. J. Decker, Y. Ding, D. Dowell, S. Edstrom, A. Fisher, J. Frisch, S. Gilevich, J. Hastings, G. Hays, P. Hering, Z. Huang, R. Iverson, H. Loos, M. Messerschmidt, A. Miahnahri, S. Moeller, H. D. Nuhn, G. Pile, D. Ratner, J. Rzepiela, D. Schultz, T. Smith, P. Stefan, H. Tompkins, J. Turner, J. Welch, W. White, J. Wu, G. Yocky, and J. Galadz, “First lasing and operation of an ångstrom-wavelength free-electron laser,” Nat. Photonics 4, 641–647 (2010).

22. Z. Huang and I. Lindau, “SACLA hard-X-ray compact FEL,” Nat. Photonics 6, 505–506 (2012).

23. H.-S. Kang, C.-K. Min, H. Heo, C. Kim, H. Yang, G. Kim, I. Nam, S. Y. Baek, H.-J. Choi, G. Min, B. R. Park, Y.-J. Suh, D. C. Shin, J. Hu, J. Hong, S.-H. Kim, K. Kim, D. Na, S. S. Park, Y. J. Park, J.-H. Hyeon, Y. G. Jung, S. H. Jeong, H.-G. Lee, S. Lee, S. Lee, W.-W. Lee, B. Oh, H. S. Suh, Y. W. Park, J.-S. Park, M. H. Kim, N.-S. Jung, Y.-C. Kim, M.-S. Lee, B.-H. Lee, C.-W. Sung, I.-S. Mok, J.-M. Yang, C.-S. Lee, H. Shin, J. H. Kim, Y. Kim, J. H. Lee, S.-Y. Park, J. Kim, J. Park, I. Eom, S. Rahn, S. Kim, K. H. Nam, J. Park, J. Park, S. Kim, S. Kwon, S. H. Park, K. S. Kim, H. Hyn, S. N. Kim, S. Kim, S.-m. Hwang, M. J. Kim, C.-y. Lim, C.-J. Yu, B.-S. Kim, T.-H. Kang, K.-W. Kim, S.-H. Kim, H.-S. Lee, H.-S. Lee, K.-H. Kim, T.-Y. Koo, D.-E. Kim, and I. S. Ko, “Hard X-ray free-electron laser with femtosecond-scale timing jitter,” Nat. Photonics 11, 708–713 (2017).

24. E. Prat, R. Abela, M. Aiba, A. Alarcon, J. Alex, Y. Arbeo, C. Arrell, C. Beard, P. Beaud, S. Bettoni, R. Bitfiger, M. Bopp, H.-H. Braun, M. Calvi, A. Cassar, T. Celcer, M. Cherqui, P. Chevtsov, C. Cirelli, A. Citterio, P. Craievich, M. C. Divall, A. Dax, M. Dehler, Y. Deng, A. Dietrich, P. Dijkstal, R. Dinapoli, S. Dordovic, S. Ebner, D. Engeler, C. Ery, V. Exposito, E. Ferrari, U. Flechsig, R. Follaith, F. Frei, R. Ganter, T. Garvey, Z. Geng, A. Gobbo, C. Gough, A. Hauff, C. P. Hauri, N. Hiller, S. Hunziker, M. Huppert, G. Ingold, R. Ischebeck, M. Janousch, P. J. M. Johnson, S. L. Johnson, P. Juranic, M. Jurevic, M. Kaiser, R. Kalt, B. Keil, D. Kiselev, C. Kittel, G. Knopp, W. Koprek, M. Laznovsky, H. T. Lemke, D. L. Sancho, F. Lohl, A. Malyshenkov, G. F. Mancini, R. Mankowsky, F. Marcellini, G. Marinkovic, I. Martiel, F. Märki, C. M. Milne, A. Mozzanica, K. Nass, G. L. Orlandi, C. O. Loch, M. Paraliev, B. Patterson, L. Patthey, B. Pedrini, M. Pedrotti, C. Pradervand, P. Radi, J.-Y. Raguin, S. Redford, J. Rehanek, S. Reiche, L. Rivkin, A. Romann, L. Sala, M. Sander, T. Schietinger, T. Schilcher, V. Schlott, T. Schmidt, M. Seidel, M. Stadler, L. Stingelin, C. Svetina, D. M. Treyer, A. Trisorio, C. Vicario, D. Voulot, A. Wrulich, S. Zerdane, and E. Zimoch, “A compact and cost-effective hard X-ray free-electron laser driven by a high-brightness and low-energy electron beam,” Nat. Photonics 14, 748–754 (2020).

25. J. Duarte, R. Cassin, J. Huijs, B. Iwan, F. Fortuna, L. Delbecq, H. Chapman, M. Fajardo, M. Kovacev, W. Boutu, and H. Hird, “Computed stereo lensless X-ray imaging,” Nat. Photonics 13, 449–453 (2019).

26. K. M. Sowa, M. P. Kujda, and P. Korecki, “Plenoptic x-ray microscopy,” Appl. Phys. Lett. 116, 014103 (2020).

27. W. Voegeli, K. Kajiwara, H. Kudo, T. Shirasawa, X. Liang, and W. Yoshih, “Multibeam x-ray optical system for high-speed tomography,” Optica 7, 514–517 (2020).

28. D. Goldberger, J. Barolak, C. G. Durfee, and D. E. Adams, “Three-dimensional single-shot ptychography,” Opt. Express 28, 18887–18898 (2020).

29. R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, “Potential for biomolecular imaging with femtosecond X-ray pulses.” Nature 406, 752–757 (2000).

30. H. N. Chapman, A. Barty, M. J. Bogan, S. Boutet, M. Frank, S. P. Hau-Riege, S. Marchesini, B. W. Woods, S. Bajt, W. H. Benner, R. A. London, E. Plönjes, M. Kuhlmann, R. Treusch, S. Düsterer, T. Schietinger, J. R. Schneider, E. Spiller, T. Möller, C. Bostedt, M. Hoener, D. A. Shapiro, K. O. Hodgson, D. van der Spoel, F. Burmeister, M. Bergh, C. Caleman, G. Huldt, M. M. Seibert, F. R. N. C. Maia, R. W. Lee, A. Szöke, N. Timneanu, and J. Hajdu, “Femtosecond diffraction imaging with a soft-X-ray free-electron laser,” Nat. Phys. 2, 839–843 (2006).

31. E. J. Candés, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” Commun. on Pure Appl. Math. 59, 1207–1223 (2006).

32. D. Donoho, “Compressed sensing,” IEEE Transactions on Inf. Theory 52, 1289–1306 (2006).

33. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature 521, 436–444 (2015).

34. A. Kar, C. Hâne, and J. Malik, “Learning a multi-view stereo machine,” Adv. neural information processing systems 30 (2017).

35. J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum, “Learning a probabilistic latent space of object shapes via...
3d generative-adversarial modeling,” Adv. neural information processing systems 29 (2016).

36. C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3d classification and segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, (2017), pp. 652–660.

37. Z. Chen and H. Zhang, “Learning implicit fields for generative shape modeling,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2019), pp. 5939–5948.

38. I. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, “DeepSDF: Learning continuous signed distance functions for shape representation,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2019), pp. 165–174.

39. 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 European conference on computer vision, (Springer, 2020), pp. 405–421.

40. A. Yu, V. Ye, M. Tancik, and A. Kanazawa, “pixelnerf: Neural radiance fields from one or few images,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2021), pp. 4578–4587.

41. P. Henzler, J. Reizenstein, P. Labatut, R. Shapovalov, T. Ritschel, A. Vedaldi, and D. Novotny, “Unsupervised learning of 3d object categories from videos in the wild,” in Proc. CVPR, (2021).

42. H. N. Chapman and K. A. Nugent, “Coherent lensless X-ray imaging,” Nat. Photonics 4, 833–839 (2010).

43. A. Bravin, P. Coan, and P. Suortti, “X-ray phase-contrast imaging: from pre-clinical applications towards clinics.” Phys. medicine biology 58, R1–35 (2013).

44. A. Andersen and A. Kak, “Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm,” Ultrason. Imaging 6, 81–94 (1984).

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,” J. Opt. Soc. Am. A 37, 422–434 (2020).

46. D. Paganin, Coherent X-Ray Optics (Oxford University Press, United Kingdom, 2006), 1st ed.

47. F. García-Moreno, P. H. Kamm, T. R. Neu, and J. Banhart, “Time-resolved in situ tomography for the analysis of evolving metal-foam granulates,” J. Synchrotron Radiat. 25, 1505–1508 (2018).

48. M. Stampanoni, A. Grosso, A. Isenegger, G. Mikuljan, Q. Chen, A. Bertrand, S. Henein, R. Betemps, U. Frommer, P. Böhler, D. Meister, M. Lange, and R. Abela, “Trends in synchrotron-based tomographic imaging: the SLS experience,” in Developments in X-Ray Tomography V, vol. 6318 U. Bonse, ed., International Society for Optics and Photonics (SPIE, 2006), pp. 193 – 206.

49. R. Mokso, C. M. Schlepütz, G. Theidel, H. Billich, E. Schmid, T. Celcer, G. Mikuljan, L. Sala, F. Marone, N. Schlumpf, and M. Stampanoni, “GigaFRoST: the gigabit fast readout system for tomography,” J. Synchrotron Radiat. 24, 1250–1259 (2017).

50. Y. Xie, T. Takikawa, S. Saito, O. Litany, S. Yan, N. Khan, F. Tombari, J. Tompkin, V. Sitzmann, and S. Sridhar, “Neural fields in visual computing and beyond,” (2021).

51. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, (2016), pp. 770–778.

52. 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, (PMLR, 2019), pp. 5301–5310.

53. 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,” arXiv preprint arXiv:2006.10739 (2020).

54. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in Advances in neural information processing systems, (2017), pp. 5998–6008.

55. I. D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980 (2014).

56. O. Ronneberger, F. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical image computing and computer-assisted intervention, (Springer, 2015), pp. 234–241.
ONIX: an X-ray deep-learning tool for 3D reconstructions from sparse views: supplemental document

This document provides supplementary information to “ONIX: an X-ray deep-learning tool for 3D reconstructions from sparse views.” In this material, we report supplementary figures of the results of ONIX for both simulation and experimental datasets.

1. SUPPLEMENTARY FIGURES

This section presents two figures with slices over the 3D volumes used and retrieved from the simulation and experimental dataset. Finally, we present a study over the ONIX reconstructions obtained by different combinations of four and six encoded views using the experimental dataset.

First, we present the results of ONIX applying to the simulation dataset. The Deep learning (DL) methods were trained using 1000 simulated 3D objects, where we simulated randomly placed ellipsoids inside a cylinder. We compared ONIX results to the ground truth, the 3D supervised learning results, and the Simultaneous Algebraic Reconstruction Technique (SART) [1] results. The results are shown in Fig. S1, where we plotted the three slices at the center of the three orthogonal axes, denoted by (i) slice 1, (ii) slice 2, and (iii) slice 3, for the (a) ground truth, and the results from (b) 3D supervised learning, (c) SART, and (d) ONIX. From Fig. S1 we can see that the best reconstructions were achieved by the 3D supervised learning approach. Nonetheless, 3D supervised learning requires the 3D information (ground truth) of the training data, limiting its application. On the other hand, it can be seen from the figure that, ONIX, outperformed SART method and provided decent reconstructions with no 3D supervision.

Next, we show the reconstruction results of ONIX applying on the experimental metallic foam dataset [2] collected from the TOMCAT beamline at the Swiss Light Source [3]. Eight projections were used in the training of the DL methods and also for the SART Reconstruction. We used the reconstructions from SART using 96 projections as the ground truth. Fig. S2 shows the (a) ground truth and the reconstructed results from (b) 3D supervised learning, (c) SART, and (d) ONIX. We plotted the three center slices of the three orthogonal axes (i-iii), the same as for the simulation dataset. It can be seen from the figure that the experimental dataset was more complicated and contained more details compared to the simulation dataset. The reconstructions from both DL methods outperformed the SART results, providing clear features in the slices. The 3D supervised results, as trained on slices, failed to capture the circular shape of the metallic foam in the direction perpendicular to the training slices, as shown in Fig. S2 b (iii). ONIX, on the other hand, learned better the structure of the foam even without accessing the 3D information provided by the ground truth.

Finally, we show the dependency of ONIX on using different encoded views to retrieve the 3D volume from eight views studying the experimental dataset. The eight projections used to train ONIX for each tomographic experiment were equally spaced between 0° to 131°. Although eight views were used for training, only a subset of them was used for the reconstructions to constrain the retrieved 3D volume at a time. Specifically, we evaluated the quality of the reconstructions using four and six encoded views to constrain the 3D reconstructions, and the results are summarized in Fig. S3. The ONIX reconstructions were compared to the ground truth via two evaluation metrics: $L_2$ norm and Dissimilarity Structure Similarity Index Metric (DSSIM). Fig. S3 shows the distribution of the reconstructions results from all the different combinations of the encoded views. Note that the six-view and four-view results came from two different trained ONIX models, each of them trained using six or four random encoded views in each training iteration. The different view combinations were only used in the reconstruction stage, where we applied different encoded views to the trained model to generate 3D reconstructions.

It can be seen that ONIX is capable of performing tomographic reconstruction for all of the experiments we did. For the reconstructions of four encoded views, the values of the $L_2$ norm are similar to the one from six views, but the deviation of DSSIM is more significant for the four-view case. We see that by adding more encoded views in training, the distribution of the reconstructions shifted towards a Gaussian distribution, implying more reliable reconstructions.
Fig. S1. Reconstructed slices from different methods for the simulation dataset. (a) The simulated ground truth. (b-d) Reconstructed results from (b) 3D supervised learning, (c) SART, and (d) ONIX. The three columns show the three slices at the center of the three orthogonal axes, denoted by (i)slice1, (ii)slice2, and (iii)slice3. We plotted only the absorption images overlaid with the phase for the reconstructions of the three methods.
Fig. S2. Reconstructed slices from different methods. (a) Ground truth from SART using 96 projections. (b-d) Reconstructed results from (b) 3D supervised learning, (c) SART, and (d) ONIX. The three slices (i-iii) at the center of the three orthogonal axes are plotted for the ground truth and the reconstructed results. The size between slices was adjusted for better arrangement.
Fig. S3. Distribution of ONIX results reconstructed using all possible combinations of four and six views over eight measured projections. (a,c) The distribution of $L_2$ norm for the six-view and four-view results. (b, d) The distribution of DSSIM for the six-view and four-view results.
REFERENCES

1. A. Andersen and A. Kak, “Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm,” Ultrason. Imaging 6, 81–94 (1984).

2. F. García-Moreno, P. H. Kamm, T. R. Neu, and J. Banhart, “Time-resolved in situ tomography for the analysis of evolving metal-foam granulates,” J. Synchrotron Radiat. 25, 1505–1508 (2018).

3. M. Stampanoni, A. Groso, A. Isenegger, G. Mikuljan, Q. Chen, A. Bertrand, S. Henein, R. Betemps, U. Frommherz, P. Böhler, D. Meister, M. Lange, and R. Abela, “Trends in synchrotron-based tomographic imaging: the SLS experience,” in Developments in X-Ray Tomography V, vol. 6318 U. Bonse, ed., International Society for Optics and Photonics (SPIE, 2006), pp. 193 – 206.