Missing Wedge Completion via Unsupervised Learning with Coordinate Networks

Dave Van Veen[1,†], Jesús G. Galaz-Montoya[2], Liyue Shen[3], Philip Baldwin[4,5], Akshay S. Chaudhari[6], Dmitry Lyumkis[5,7], Michael F. Schmid[8], Wah Chiu[2,8,9] and John Pauly[1]

1: Dept. of Electrical Engineering, Stanford University 2: Dept. of Bioengineering, Stanford University 3: Dept. of Electrical and Computer Engineering, University of Michigan 4: Dept. of Biochemistry and Molecular Pharmacology, Baylor College of Medicine 5: Dept. of Genetics, The Salk Institute for Biological Sciences 6: Dept. of Radiology, Stanford University 7: Graduate School of Biological Sciences, University of California San Diego 8: Division of CryoEM and Bioimaging, SSRL, SLAC National Accelerator Laboratory 9: Dept. of Microbiology and Immunology, Stanford University

Cryogenic electron tomography (cryoET) is a powerful tool in structural biology, enabling detailed 3D imaging of biological specimens at a resolution of nanometers. Despite its potential, cryoET faces challenges such as the missing wedge problem, which limits reconstruction quality due to incomplete data collection angles. Recently, supervised deep learning methods leveraging convolutional neural networks (CNNs) have considerably addressed this issue; however, their pretraining requirements render them susceptible to inaccuracies and artifacts, particularly when representative training data is scarce. To overcome these limitations, we introduce a proof-of-concept unsupervised learning approach using coordinate networks (CNs) that optimizes network weights directly against input projections. This eliminates the need for pretraining, reducing reconstruction runtime by $3-20\times$ compared to supervised methods. Our *in silico* results show improved shape completion and reduction of missing wedge artifacts, assessed through several voxel-based image quality metrics in real space and a novel directional Fourier Shell Correlation (FSC) metric. Our study illuminates benefits and considerations of both supervised and unsupervised approaches, guiding the development of improved reconstruction strategies.

1 Introduction

Recent advancements in cryogenic electron microscopy (cryoEM) [1, 2] have elevated it from a specialized technique to a cornerstone of structural biology [3, 4] and molecular sciences [5]. Cryogenic electron tomography (cryoET), an extension of cryoEM, offers detailed three-dimensional (3D) representations of macromolecules, cells, and tissues in states close to their natural environment at nanometer-scale resolution [6]. This technique is versatile, allowing for the examination of a wide array of macromolecular complexes *in vitro* [7] and *in situ* [8], including amyloid filaments [9–14] and enveloped viruses such as SARS-CoV-2 [15]. CryoET can also probe the structure of other clinically-relevant samples ranging from individual organelles [16, 17] and cells [18, 19] to complex tissue sections [20–24]. Insights from cryoET data analysis can facilitate understanding dynamic molecular processes such as viral infections [25–30], enable pathology diagnosis [31, 32], and reveal the impacts of potential therapeutic interventions [19], among other biomedical applications. A critical downstream technique, subtomogram averaging (STA) [33, 34], further refines cryoET data to achieve subnanometer [35] to near-atomic [36] resolution of repeated structures within tomograms [37, 38], underscoring the importance of precise cryoET reconstructions for accurate particle localization and structural analysis.

CryoET involves the rapid freezing of biological specimens followed by their examination with a transmission electron microscope (TEM) [39]. This process entails capturing a series of two-dimensional (2D) projection

† *Corresponding author: vanveen@stanford.edu* *Code and data:* github.com/davevanveen/unsup_cryoet

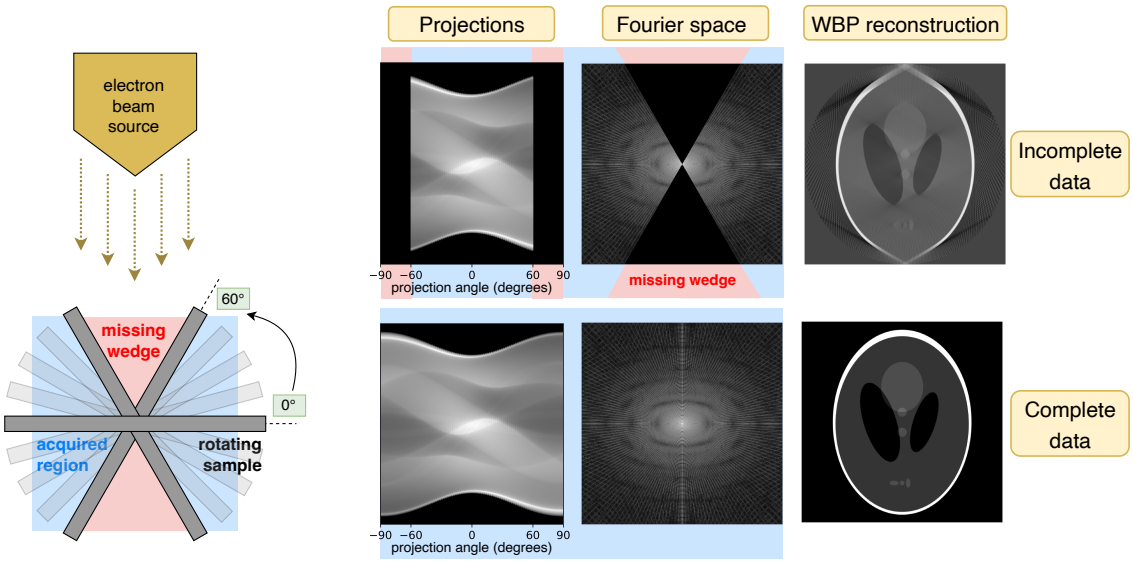

Figure 1 | **Left:** CryoET acquires tomograms by collecting a tilt series of 2D projection images of vitrified biological samples over a limited angular range. **Right:** Simulations of the widely used Shepp-Logan phantom model show that restricted projection angles result in a missing wedge in Fourier space, leading to distortions in the reconstructed image (top row). The objective of this study is to reconstruct the uncollected data in this region, effectively completing the wedge (bottom row).

images as the specimen is incrementally tilted, compiling what is known as a tilt series [40]. These images are then aligned and combined to produce a 3D reconstruction, or tomogram, typically through weighted-back projection (WBP) methods [41], as enabled by software such as IMOD [42]. Despite cryoET's ability to capture intricate structural details, the technique is limited by the specimens' susceptibility to radiation damage [43] and the inherent mechanical constraints of TEM, which restrict the tilt series to angles between -60° and +60°. This limitation results in the "missing wedge" phenomenon [44], where the lack of data at experimentally inaccessible angles leads to artifacts that distort tomogram quality. This affects the resolution and density accuracy of visualized features, hence complicating 3D analyses. Such distortions are particularly problematic for structures perpendicular to the electron beam, often resulting in the omission of critical top and bottom details in images of spherical, oblong, or elongated biological features (*e.g.* cell membranes, organelles, vesicles, microtubules, and virions).

Efforts to mitigate the missing wedge's impact have ranged from introducing new data collection techniques, such as dual-axis [45] and conical [46] tomography, to novel applications of statistical and iterative data processing methods [47–49], including total variation minimization [50] and compressed sensing [51]. In the realm of supervised deep learning, IsoNet stands out by employing a convolutional neural network (CNN) U-Net [52] trained on subtomograms extracted from tomograms reconstructed via weighted-back projection (WBP), intentionally adding missing wedge artifacts to create a paired training set. IsoNet, alongside other supervised methods [53–55], has shown significant success in addressing the missing wedge problem. However, these data-driven approaches face limitations: first, they require computationally intensive supervised pretraining; second, they rely on WBP reconstructions that already exhibit missing wedge artifacts; and third, supervised learning techniques can be prone to generating fictitious densities and inaccurately positioning structures within the reconstruction [56, 57]. Supervised learning applications in tomography, including those using the U-Net architecture [52], have been documented to result in such feature hallucinations and misplacements [58, 59]. These problems are exacerbated when the training data is limited; thus, in cryoET, it could lead to misinterpreting rare events [60, 61] that are scantly represented in large datasets.

As an alternative, we explore a novel unsupervised learning strategy [62–64] that bypasses the limitations associated with supervised learning [56–59] and the reliance on artifact-prone WBP reconstructions. Our approach starts with a randomly initialized network, optimizing its weights so that the generated image agrees with the experimentally captured projections, thus avoiding the need for pretraining on compromised

WBP reconstructions with missing-wedge-induced artifacts. We employ coordinate networks (CNs) [65] to reconstruct this unsupervised representation of the tomogram. The CN determines 3D voxel values in the reconstruction volume by relating them to the corresponding 2D pixels in the projection images. Unlike conventional kernel-based methods such as CNNs, CNs offer a continuous representation by mapping coordinates to their corresponding values through a network-embedded continuous function—this allows CNs to capture image details without being constrained by a fixed grid resolution. Given their growing application in computationally intensive tasks in computer graphics [66, 67] and demonstrated potential in various biomedical imaging applications [68–73], CNs present a powerful solution for accurately representing extensive cryoET volumes, addressing the challenges of high computational costs and fixed-resolution limitations associated with CNNs.

Our study reveals that unsupervised learning with CNs can enhance shape fidelity and diminish the impact of the missing wedge on *in silico* data compared to traditional and CNN-based methods. Furthermore, bypassing the pretraining step allows CNs to produce reconstructions between three to over twenty times more rapidly than pretrained CNN methods. To rigorously assess image quality, we employed various voxel-based metrics and introduced a novel directional Fourier Shell Correlation (FSC) metric. This new metric is tailored to specifically quantify the restoration of information within the regions affected by the missing wedge.

While our findings highlight certain advantages of unsupervised learning for cryoET reconstruction, they are preliminary and not intended to establish superiority over other methods. Instead, we compare traditional WBP and Fourier inversion reconstructions against supervised and unsupervised machine learning frameworks, shedding light on their respective benefits and limitations within the broader context of structural biology and molecular imaging. Through this comparison, we aim to contribute valuable insights into the ongoing discourse of cryoET reconstruction techniques.

This manuscript is organized as follows: Section 2 describes the cryoET forward model (Figure 1), our reconstruction algorithm (Figure 2), data, experimental setup, and evaluation methods. Qualitative and quantitative results (Figures 3, 4, 5) in Section 3 are followed by a discussion in Section 4. Additional results can be found in the Appendix.

2 Materials and Methods

2.1 Forward Model

Let $\mathbf{v}^* \in \mathbb{R}^{x \times y \times z}$ be the true image volume (tomogram) we wish to reconstruct given access to projections $\mathbf{p} \in \mathbb{R}^{l \times x \times y}$, *i.e.* $\mathbf{p} = \mathbf{P}\mathbf{v}^*$. Here $\mathbf{P}$ denotes the projection operator, which projects electrons through the volume $\mathbf{v}^*$ in a parallel beam at l different tilt angles. This process provides $\mathbf{p}$, a tilt series of l projection images, each size $x \times y$.

2.2 Reconstruction Algorithm

We employ a CN $G_{\boldsymbol{\theta}} : \mathbb{R}^3 \to \mathbb{R}$ with trainable parameters $\boldsymbol{\theta}$ that maps an individual 3D coordinate $\mathbf{c} \in \mathbb{R}^3$ in the reconstruction volume to a pixel value at the corresponding locations in the 2D projection images. Evaluating this network over the entire set of coordinates $\mathbf{C} = \{\mathbf{c}_q\}^{x \times y \times z}$ produces a reconstruction volume $G_{\boldsymbol{\theta}}(\mathbf{C}) \in \mathbb{R}^{x \times y \times z}$.

Our goal is to find a set of parameters for the CN such that its reprojections—the projector applied to the network output, *i.e.* $\mathbf{P}G_{\boldsymbol{\theta}}(\mathbf{C})$—matches the experimentally given projections $\mathbf{p}$. Hence we randomly initialize parameters $\boldsymbol{\theta}$ and solve the following:

$$\boldsymbol{\theta}^* = \text{argmin}_{\boldsymbol{\theta}} \|\mathbf{p} - \mathbf{P}G_{\boldsymbol{\theta}}(\mathbf{C})\| + \lambda R(G_{\boldsymbol{\theta}}(\mathbf{C})), \tag{1}$$

where R is a regularization term applied to the estimated image with strength λ. Because the projection operator $\mathbf{P}$ is differentiable, we can use gradient-based backpropagation to solve this equation. Figure 2 depicts this training process. The resulting network then produces $\hat{\mathbf{v}}$, an estimate of the image volume we wish to reconstruct, *i.e.* $\mathbf{v}^* \approx \hat{\mathbf{v}} = G_{\boldsymbol{\theta}^*}(\mathbf{C})$.

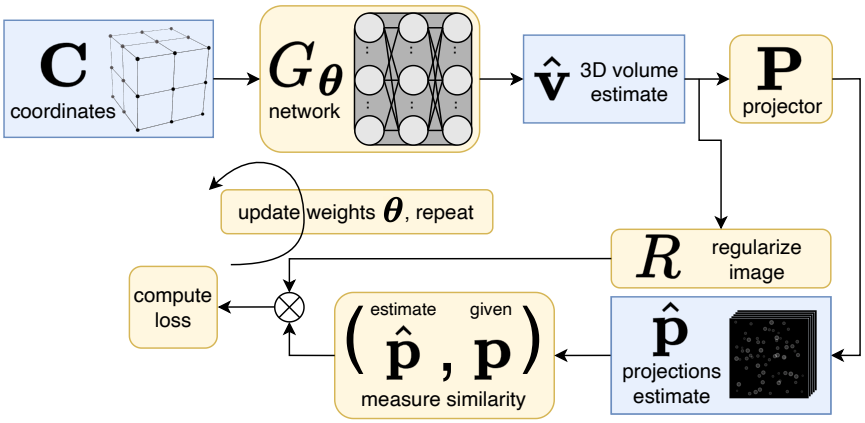

Figure 2 | Unsupervised reconstruction using a coordinate network (CN) to create a 3D volume estimate, $\hat{\mathbf{v}}$. This diagram depicts one iteration of training process, where network weights $\boldsymbol{\theta}$ are updated by constraining the estimated projections $\hat{\mathbf{p}}$ to match the given projections, $\mathbf{p}$. After the training process, $\hat{\mathbf{v}}$ closely matches the ground-truth tomogram.

Our methodology distinguishes itself by employing an unsupervised approach, thereby obviating the necessity for pretraining. Contrary to pretrained strategies that depend on supervising with data augmentation strategies—such as using subtomograms that replicate the effects of the missing wedge—our technique refines network parameters by directly using the experimental projection images. This direct optimization method effectively circumvents the common pitfalls of supervised learning, including the propensity for some types of artifact generation and structural inaccuracies [56–59]. Indeed, our approach leverages more dependable data—the experimental projection images themselves—avoiding the artifact-laden WBP reconstructions commonly utilized as a starting point for supervised methods. By ensuring a closer agreement between the reconstructions' reprojections and the original experimental projections, our method inherently reduces the likelihood of introducing hallucinated errors or artifacts.

2.3 Data

To compare our approach with other reconstruction methods, we carried out *in silico* experiments for which ground truth is known, enabling precise evaluation via quantitative, reference-based metrics. Tomograms were created with image processing tools available in EMAN2 [74], and their corresponding tilt series were generated by projecting through the volumes every 2° across the range of -60° to +60° using EMAN2's standard projector. See below for descriptions of each dataset ($x \times y \times z$):

Spheres ($1024 \times 1024 \times 256$): a collection of binarized hollow spheres of constant density and variable sizes (16 to 64 pixels in diameter). Compare to x and y, the smaller dimension in z renders a slab-shaped volume, geometrically mimicking a distribution of discrete objects in a thin layer of ice.

Mixed shapes ($1024 \times 1024 \times 256$): varied geometric shapes with heterogeneous structures. These binarized shapes include full spheres, ellipsoids, pyramids, cubes, rectangular prisms, circular discs, as well as 4- and 6-pointed 3D crosses. Similar to spheres, the slab shape of this tomogram mimics the geometry of a thin layer of ice.

P22 ($360 \times 360 \times 360$): single P22 phage particles. The P22 capsid displays icosahedral symmetry, while the virion tail exhibits pseudo-six fold symmetry. This map, accessed via the electron microscopy data bank (EMDB, accession number EMD-9008) [75], is sampled at 4.5 Å/pixel. We clipped the volume to a $360 \times 360 \times 360$ box size and threshold filtered it to eliminate negative densities.

Ubiquitin ($64 \times 64 \times 64$): the regulatory protein ubiquitin. We created this simulation using a map generated from an atomic model downloaded from the protein data bank (PDB)(PDB ID: 1UBQ) [76].

For final processing steps, each simulated tomogram (except for ubiquitin) was low-pass filtered to either render shape surfaces smooth (spheres, shapes) or dampen high-resolution features that would not be present in a raw cryoET tomogram without averaging (P22, ubiquitin). Finally, each tilt series of projection images

was normalized to be within the range of $[0, 1]$ to standardize input values provided to the network.

2.4 Experimental Setup

To find a set of weights $\boldsymbol{\theta}^*$ that minimize Equation 1, we construct a fully-connected coordinate network architecture in PyTorch [77]. This network has four hidden layers each with 256 features, positional encoding [66], and sinusoidal activation functions [65]. For simplicity, we employ this same architecture on all datasets and maintain a consistent ratio of network parameters to measurements (roughly $\frac{1}{8}$). This consistency is achieved by dividing the tilt series $\mathbf{p} \in \mathbb{R}^{l \times x \times y}$ into length-j subslices along the y-axis, *i.e.* $\mathbf{p}_{\text{sub}} \in \mathbb{R}^{l \times x \times j}$; in image space, this corresponds to a subvolume $\mathbf{v}^*_{\text{sub}} \in \mathbb{R}^{x \times j \times z}$. Subsequently we fit a separate set of network parameters to represent each subvolume. For example, given the aforementioned network with $4 * 256^2 = 262,144$ parameters and a tomogram of size $1024 \times 1024 \times 256$, obtaining a $\frac{1}{8}$ ratio would yield 128 networks, each fitting measurements for a subvolume of size $1024 \times 8 \times 256$. We then stitch these individual subvolumes together along the y-axis to obtain the final reconstructed volume.

Given sufficient memory, a sparser representation, or a smaller volume, one could reconstruct the entire volume with one single network as we demonstrate in Figure A4. However, we find our subvolume approach has several advantages: (1) it ensures the network has sufficient representational capacity for any tomogram, (2) it uses a small amount of memory——roughly 2-4 GB on our NVIDIA Quadro RTX 8000— making this method feasible on smaller GPUs, and (3) it enables us to leverage learned initializations [68], *i.e.* after fitting a network to one subvolume, those same parameters are used to initialize the network for the adjacent subvolume. This learned initialization strategy improves reconstruction quality by enhancing consistency along the y-direction and also reducing the number of gradient step iterations required for the network to fit the subvolume. As such, we use 2,000 iterations to fit the first subvolume and 400 iterations for all adjacent subvolumes which leverage learned initializations. For the first subvolume, we use an initial learning rate of 1×10^{-3} decayed logarithmically to 1×10^{-4}; for all adjacent subvolumes, we use an initial learning rate of 1×10^{-4} decayed logarithmically to 1×10^{-5}. By default, $\lambda = 0$ in Equation 1.

2.5 Baselines

We compare our algorithm to three well-established baselines previously introduced in Section 1: (1) WBP [41] reconstructions generated via IMOD [42], (2) Fourier inversion reconstructions generated using EMAN2 [78], and (3) reconstructions with missing wedge restoration by IsoNet [79], a supervised deep learning approach leveraging CNNs to fill in the missing wedge of the IMOD reconstruction provided as input. We choose IsoNet because it is widely regarded as the leading method for missing wedge compensation, outperforming approaches such as ICON [51] or MBIR [49]. We use default parameters for each reconstruction method.

2.6 Image Evaluation

To provide a comprehensive evaluation of signal preservation across frequencies in these reconstructions, we employ the Fourier shell correlation (FSC) metric, common in structural biology, as well as voxel-based metrics from the computational imaging literature. All metrics are reference-based, comparing the reconstructed image $\hat{\mathbf{v}}$ against a ground truth reference image $\mathbf{v}^*$. For each metric, higher values indicate superior quality.

2.6.1 Directional Fourier shell correlation (FSC)

FSC measures the similarity between the Fourier coefficients of a reconstructed image volume and those of its ground truth reference. This measurement is performed by selecting a specific radius in Fourier space and identifying points within a half-unit distance from the sphere's surface corresponding to that radius. These identified points contribute to the calculation of a normalized Pearson correlation coefficient, which is computed without subtracting the mean. The process involves incrementally adjusting the Fourier radius from one unit up to the Nyquist frequency, calculating the Pearson correlation at each step to assess the correlation across different spatial frequencies.

We first apply the conventional FSC metric across the entirety of the reconstructed volumes. Additionally, we introduce a directional FSC variant designed specifically for assessing the missing data regions. This innovative approach aims to directly highlight the effectiveness of a reconstruction algorithm in compensating

for the missing wedge. We characterize the "present data" region as the areas within a half-unit distance from any plane defined by the direction of the l acquired projections, effectively defining a slab for each tilt angle; the "missing data" regions correspond to the complement of the present data. While alternative methods for interpolating data within this geometric framework are possible, the FSC typically exhibits a smooth profile across these calculations. Consequently, we anticipate similar results with varying interpolation strategies.

2.6.2 Voxel-based metrics

This subsection outlines three metrics widely recognized in imaging research, each assessing distinct aspects of image quality.

Peak Signal-to-Noise Ratio (PSNR) quantifies the ratio between the maximum possible power of a signal and the power of corrupting noise that affects its representation. Measured in decibels, PSNR is derived from the mean-squared error (MSE) between a reconstructed image and its ground truth. The formula for PSNR is:

$$\text{PSNR} = 10\log_{10}\frac{I_m}{\|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2},$$

where I_m represents the maximum possible pixel value (*e.g.*, 255 in an 8-bit grayscale image) and the denominator corresponds to MSE. This metric is particularly suited to measuring similarity of low-frequency image components [80].

Structural Similarity Index (SSIM) [81] measures perceived image quality by evaluating aspects like structure (texture and pattern consistency), luminance (brightness levels), and contrast (voxel variance). SSIM values range from -1 to 1, with 1 signifying perfect similarity, *i.e.* $\hat{\mathbf{v}} = \mathbf{v}^*$. Compared to PSNR, SSIM offers a more nuanced evaluation of image quality, closely aligning with human visual perception [82].

Visual Information Fidelity (VIF) [83] quantifies how well the reconstructed image captures natural scene statistics corresponding to the human visual system (HVS). It measures mutual information (MI) between the input and outputs of both the reconstructed and reference volumes, $\mathbf{v}$ and $\mathbf{v}^*$. Hence the formula for VIF can be written as:

$$\text{VIF} = \frac{\text{MI}(\hat{\mathbf{v}}, \text{HVS}(\hat{\mathbf{v}}))}{\text{MI}(\mathbf{v}^*, \text{HVS}(\mathbf{v}^*))}.$$

Possible values of this metric can be between 0 and 1 (blurry $\hat{\mathbf{v}}$), 1 ($\hat{\mathbf{v}} = \mathbf{v}^*$), or greater than 1 ($\hat{\mathbf{v}}$ provides contrast enhancement of $\mathbf{v}^*$ without adding noise). This metric best captures similarity between higher frequency components of an image. In contexts such as magnetic resonance imaging, VIF has demonstrated alignment with radiologist preferences [84].

3 Results

Figure 3 shows xz-plane reprojections of small regions of the spheres and shapes datasets as well as the full span of a P22 volume. All orthogonal reprojections are shown in Figures A1, A2, and A3 with the xz-projection trimmed in size for display purposes. The missing wedge artifact due to anisotropic resolution manifests as elongation streaks in the xz-plane and blurriness along z in the yz-projections. Across methods, IMOD and EMAN2 reconstructions exhibit these artifacts most strongly. Such artifacts are substantially reduced by IsoNet and the least pronounced in reconstructions using our approach.

We now consider the accuracy of shape representation. For the spheres dataset, both IMOD and EMAN2 inaccurately render ellipsoidal contours elongated along the z-axis. For the P22 dataset, these methods fail to achieve the expected sharpness along the edges of the particle in z. Across all datasets, IsoNet reduces these distortions considerably. Meanwhile, our CN reconstruction preserves shape fidelity closest to ground truth, reducing distortions more than IsoNet.

Figure 3 exhibits high-frequency reconstruction artifacts for IsoNet in the form of tiling (spheres, shapes) and for our method in the form of high frequency streaks (spheres). In spite of this, both IsoNet and our method preserve overall shape fidelity at low and intermediate frequencies, which are most relevant in cryoET outside of high-resolution STA applications. We provide a thorough discussion of these artifacts in Section 4.

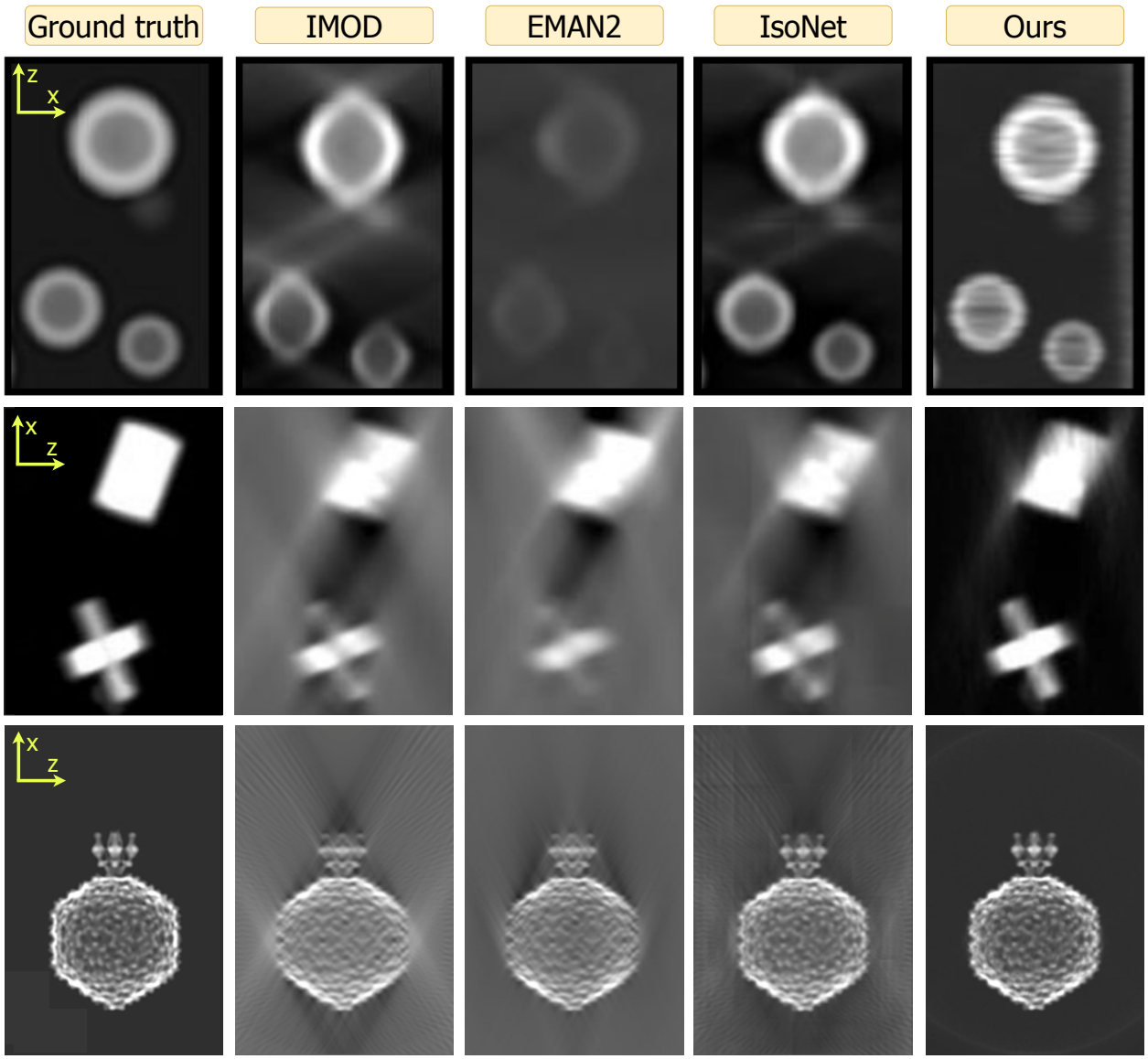

Figure 3 | Qualitative review of the spheres (top), shapes (middle), and P22 (bottom) datasets across all methods (columns). Images are zoom insets of projections in the xz-plane. Both IMOD and EMAN2 suffer from back-projection artifacts. Compared to IsoNet, our method better resolves these artifacts and produces higher shape fidelity. Both IsoNet (tile pattern) and ours (horizontal streaks) can produce artifacts due to computational constraints, as discussed in Section 4. Figures A1, A2, A3 contain all projection directions and a larger field of view.

To quantitatively evaluate reconstruction quality, we utilized several voxel-based metrics alongside the FSC (Figure 4). Our technique outperforms others in terms of PSNR, which assesses performance at lower frequencies, and SSIM, which evaluates structural integrity. However, IsoNet outperforms our method in VIF, indicative of higher frequency accuracy, in two out of the three datasets examined. This observation is consistent with FSC analysis, which demonstrates superior performance of our method at lower frequencies while trailing at higher frequencies. FSC curves for the missing data regions highlight the proficiency of both our method and IsoNet in compensating for the lack of information in the missing wedge compared to IMOD and EMAN2.

To further test robustness of different methods, we generated alternate tilt series from one of the P22 tomograms by varying two critical parameters: angular step (α) and angular range (β). CryoET data collection often

Dataset	Model	PSNR	SSIM	VIF	Runtime (min)
Spheres	EMAN2	26.3	0.94	0.64	2.68
	IMOD	27.3	0.93	0.78	0.15
	IsoNet	30.2	0.96	0.93	476
	Ours	31.4	0.97	0.86	168
Shapes	EMAN2	25.0	0.73	0.76	2.71
	IMOD	25.8	0.69	0.74	0.15
	IsoNet	27.4	0.69	0.84	473
	Ours	31.5	0.94	0.85	172
P22	EMAN2	30.9	0.98	0.76	3.18
	IMOD	28.7	0.92	0.70	0.17
	IsoNet	34.2	0.95	0.90	503
	Ours	36.1	0.99	0.88	19.6

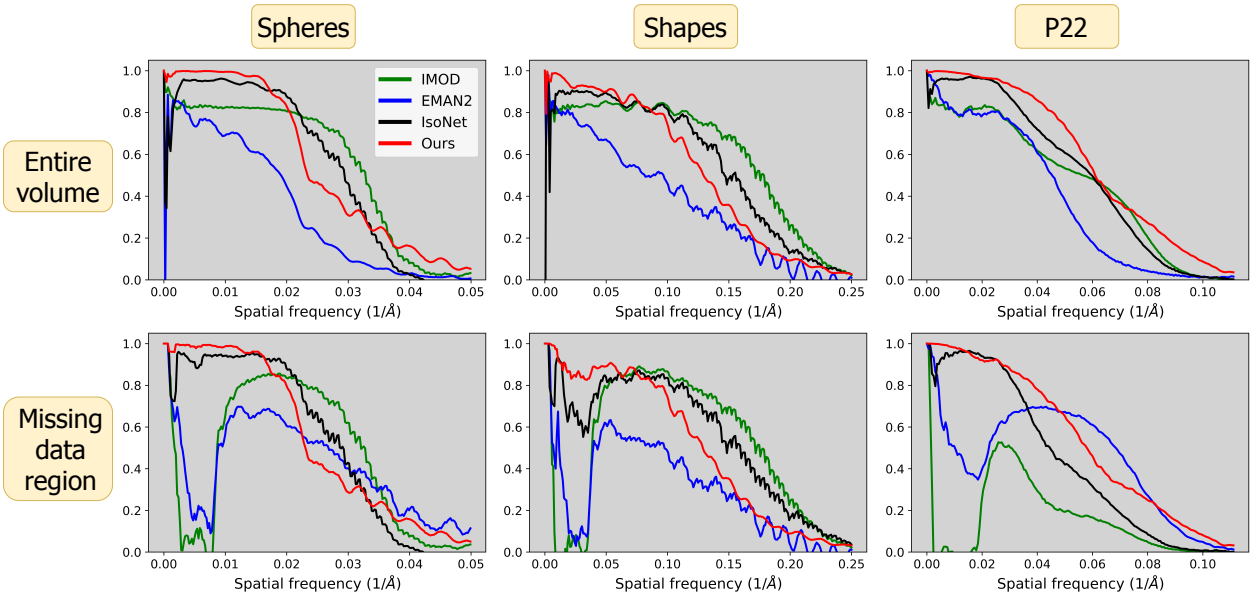

Figure 4 | Image quality metrics across all datasets and models. **Top:** Voxel-based metrics and algorithm runtime. Ours (CN) generally performs best except for the VIF metric, which measures high frequencies. Due to pretraining, IsoNet is the most computationally intensive. **Bottom:** Fourier shell correlation (FSC) curves for each dataset (columns) across the entire volume (top row) and missing data region (bottom row). Ours typically performs best at lower frequencies and is surpassed at higher frequencies. Only ours and IsoNet provide substantive information at lower frequencies.

uses smaller tilt steps of 1-2° for large, continuous specimens such as cells; however, tilt steps of 3-5° or larger can be used for sparse specimens such as macromolecules in solution destined to undergo STA [33]. Similarly, data collection ranges can be smaller than [-60°, +60°] for high-resolution STA, as high-tilt images may be too noisy and damaged by cumulative radiation dose.Figure 5 shows the impact of variations in α and β on the performance of each reconstruction method. Our findings affirm the adaptability of our method to different acquisition parameters, underscoring the potential utility of our approach in unique applications requiring varied or non-standard data collection parameters.

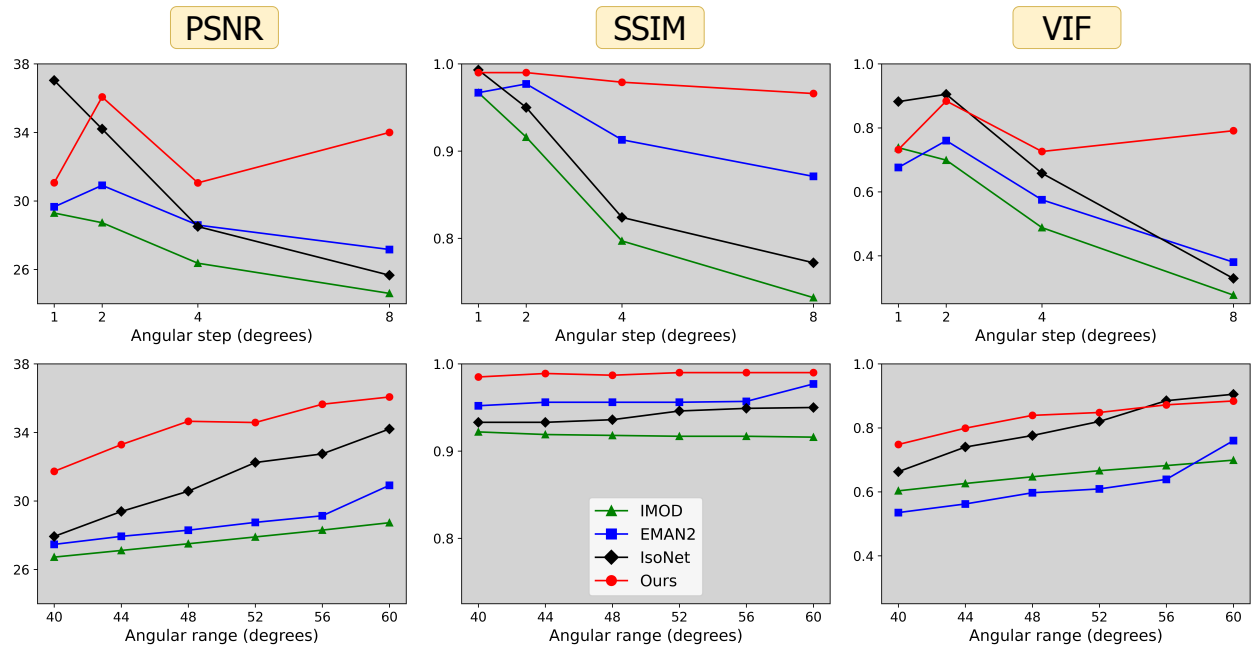

Figure 5 | Voxel-based metrics (columns) with varying acquisition parameters: angular step, *i.e.* number of degrees between tilts α (top row) and angular range β, corresponding to projections over the range of $[-\beta, \beta]$ degrees (bottom row). Default acquisition parameters are $\alpha = 2$ and $\beta = 60$ degrees.

4 Discussion

Our CN reconstruction method exhibits superior performance in low-frequency ranges as depicted in Figure 4, a trait consistent with observations in unsupervised learning methods noted for their low-frequency spectral bias [62–64, 85–87]. This may be advantageous for cryoET tasks that require shape integrity such as feature segmentation and particle picking—which can be challenging, slow, and inconsistent in typical cryoET tomograms due to missing-wedge-induced resolution anisotropy [88, 89]. At higher frequencies, our method performs worse than IMOD and IsoNet as deemed by the VIF metric and FSC curves in Figure 4. This suggests that our method may not benefit high-resolution applications, such as completing the missing wedge in single-particle cryoEM analyses of macromolecular complexes which exhibit preferred orientation; for this problem, various experimental [90] and algorithmic [91–93] approaches have been proposed. The fact that different methods perform better in different frequency ranges prompts considering an ensemble approach for future advancements, potentially integrating our unsupervised model's strengths in lower frequencies with IsoNet's proficiency in higher frequency details, to provide a more uniformly high-quality reconstruction.

The varied performance across frequency ranges also underscores the critical role of tailored evaluation metrics, such as our innovative use of directional FSC to evaluate information in the missing data region. This specific assessment underlines the advantage offered by neural network approaches like ours and IsoNet in compensating for the missing wedge. Given that no single reconstruction method performs best across all frequencies or samples, selecting varied test datasets and evaluation metrics is essential for a comprehensive assessment of reconstruction methods.

Our direct use of projection images bypasses the initial distortion introduced by WBP images that IsoNet uses for training. We hypothesize this promotes a higher fidelity to the original shapes within the tomograms, as exemplified by our spherical outcomes versus IsoNet's ellipsoidal tendencies in Figure 3 and by our superior performance at low frequencies and in structural similarity 4. This also reveals a trade-off in our reconstruction approach, as demonstrated in Figure 5. Our method's reliance on projections means that variations in the angular step (α) and range (β) directly alter the volume of information we process. In contrast, IsoNet's dependency on WBP images means that, while changes in α and β affect the quality of its WBP input

image, the amount of input data—essentially the image size —remain unchanged. Interestingly, when we decrease α to 1°, ostensibly increasing the available information, our method's performance unexpectedly dips. This likely stemmed from our network's capacity being held constant throughout these experiments, despite processing an increasing amount of projections with decreasing tilt step; i.e., the network capacity may have been exceeded going from 2° to 1°. This observation motivates our decision to dynamically adjust the network size in response to the dimensions of the tomogram and the number of projections available, as discussed in Section 2.4. This adaptability ensures that the network's representational capacity—essentially the ratio of network parameters to tomogram size—remains consistent. Such consistency is beneficial for achieving uniform quality across reconstructions of varying sizes and complexities. Additionally, this feature facilitates a scalable runtime, proportional to the tomogram's size. For instance, when transitioning from the larger spheres dataset to the smaller P22 dataset, we observed an 89% decrease in runtime (Figure 4), closely mirroring an 83% reduction in tomogram volume. In contrast, IsoNet exhibits a relatively uniform runtime across these datasets, highlighting a fundamental difference in our approaches.

Our method's capacity to adjust network size allows for customized reconstruction scale based on the hardware available: smaller subvolumes along the y-axis on more modest systems, and larger subvolumes on more powerful setups. In principle, dividing the tomogram into adjacent subvolumes to meet memory constrains of the GPU also allows for parallelization to derive future speed gains. However, this advantage does not come without challenges. Specifically, it can lead to the emergence of streaking artifacts along the y-axis, a phenomenon observable in the spheres reconstruction depicted in Figure 3. The occurrence of such artifacts, however, is not inherent in our CN methodology, as demonstrated by our experiment using the small regulatory protein ubiquitin in Figure A4, which shows these artifacts vanish when a single network is employed to reconstruct the entire volume. Yet, this single-network approach is not currently feasible for larger datasets like our spheres and geometric shapes simulations due to memory constraints on our hardware (NVIDIA Quadro RTX 8000). Future hardware developments or parallelization could alleviate this issue. Separately, future algorithmic improvements could more efficiently represent large volumes, using strategies such as Gaussian splatting [94] or Gaussian mixture models, as demonstrated for SPA cryoEM [95, 96].

IsoNet also grapples with computational limitations inherent to 3D CNNs, resulting in tiling artifacts in reconstructions (Figure 3). These artifacts stem from the necessity of dividing the training set into manageable 3D subtomograms, each measuring $64 \times 64 \times 64$ voxels by default. This methodological constraint underscores a shared challenge in cryoET reconstruction: balancing computational feasibility with the aim of computing artifact-free, high-fidelity reconstructions.

While our *in silico* results are promising, real-world application remains under development, presenting a frontier for future research. Real datasets introduce complexities such as noise and CTF, necessitating advanced unsupervised learning techniques for effective noise management and image enhancement [97, 98]. Indeed, the CTF is particularly challenging to correct for in cryoET datasets due to the defocus gradient present in cryoEM images of tilted specimens [99, 100], which worsens with increasing specimen thickness [101], tilt angle [102], and field of view such as at lower magnifications. CTF-induced artifacts can limit the resolution at which macromolecules and biological specimens in general can be visualized at, and artificially give an appearance of hollowness to solid objects depending on their shape, size, and the defocus amount. Moreover, IsoNet's comprehensive image processing pipeline, which includes preprocessing steps before reconstruction, such as CTF deconvolution [103] and masking, underscores the challenges to making objective, holistic comparisons between reconstruction methods. Here, we focus on characterizing the performance of unsupervised machine learning methods and evaluating trade-offs with supervised ones.

Given the increasing popularity of cryoET [104] owing to its demonstrated applications in cellular structural characterization [105] and histopathological clinical diagnoses [32], we expect for artificial intelligence developments that enhance tomographic reconstruction to become an increasingly active and impactful field of research. Since cryoET can provide 3D views of individual macromolecules and complexes across a wide range of sizes, as well as their distributions within organelles, cells, and tissues, improving tomographic reconstruction quality may enable novel structure-based diagnostics at the molecular level and accelerate drug development by assessing the effects of molecular therapeutics on phenotype.

5 Acknowledgments

This research was funded by a grant from the Chan-Zuckerberg Institute entitled "Resolving Biostructures In-situ via Cryogenic Light and Electron Microscopy" in addition to NIH grants U01 AI136680 and U54 AI170855.

References

1. Eisenstein, M. The field that came in from the cold. *Nature Methods* **13,** 19–23 (2016).
2. Crowther, R. A. The resolution revolution: recent advances in cryoEM (2016).
3. Shen, P. S. The 2017 Nobel Prize in Chemistry: cryo-EM comes of age. *Analytical and bioanalytical chemistry* **410,** 2053–2057 (2018).
4. De la Cruz, M. J. & Eng, E. T. Scaling up cryo-EM for biology and chemistry: The journey from niche technology to mainstream method. *Structure* (2023).
5. Saibil, H. R. Cryo-EM in molecular and cellular biology. *Molecular Cell* **82,** 274–284 (2022).
6. Lučić, V., Rigort, A. & Baumeister, W. Cryo-electron tomography: the challenge of doing structural biology in situ. *Journal of Cell Biology* **202,** 407–419 (2013).
7. Cope, J., Heumann, J. & Hoenger, A. Cryo-Electron Tomography for Structural Characterization of Macromolecular Complexes. *Current Protocols in Protein Science* **65,** 17–13 (2011).
8. Asano, S., Fukuda, Y., Beck, F., Aufderheide, A., Förster, F., Danev, R. & Baumeister, W. A molecular census of 26 S proteasomes in intact neurons. *Science* **347,** 439–442 (2015).
9. Shahmoradian, S. H., Galaz-Montoya, J. G., Schmid, M. F., Cong, Y., Ma, B., Spiess, C., Frydman, J., Ludtke, S. J. & Chiu, W. TRiC's tricks inhibit huntingtin aggregation. *Elife* **2,** e00710 (2013).
10. Darrow, M. C., Sergeeva, O. A., Isas, J. M., Galaz-Montoya, J. G., King, J. A., Langen, R., Schmid, M. F. & Chiu, W. Structural mechanisms of mutant huntingtin aggregation suppression by the synthetic chaperonin-like CCT5 complex explained by cryoelectron tomography. *Journal of Biological Chemistry* **290,** 17451–17461 (2015).
11. Bäuerlein, F. J., Saha, I., Mishra, A., Kalemanov, M., Martínez-Sánchez, A., Klein, R., Dudanova, I., Hipp, M. S., Hartl, F. U., Baumeister, W., *et al.* In situ architecture and cellular interactions of PolyQ inclusions. *Cell* **171,** 179–187 (2017).
12. Guo, Q., Lehmer, C., Martínez-Sánchez, A., Rudack, T., Beck, F., Hartmann, H., Pérez-Berlanga, M., Frottin, F., Hipp, M. S., Hartl, F. U., *et al.* In situ structure of neuronal C9orf72 poly-GA aggregates reveals proteasome recruitment. *Cell* **172,** 696–705 (2018).
13. Trinkaus, V. A., Riera-Tur, I., Martínez-Sánchez, A., Bäuerlein, F. J., Guo, Q., Arzberger, T., Baumeister, W., Dudanova, I., Hipp, M. S., Hartl, F. U., *et al.* In situ architecture of neuronal α-Synuclein inclusions. *Nature Communications* **12,** 2110 (2021).
14. Galaz-Montoya, J. G., Shahmoradian, S. H., Shen, K., Frydman, J. & Chiu, W. Cryo-electron tomography provides topological insights into mutant huntingtin exon 1 and polyQ aggregates. *Communications Biology* **4,** 849 (2021).
15. Liu, C., Mendonça, L., Yang, Y., Gao, Y., Shen, C., Liu, J., Ni, T., Ju, B., Liu, C., Tang, X., *et al.* The architecture of inactivated SARS-CoV-2 with postfusion spikes revealed by cryo-EM and cryo-ET. *Structure* **28,** 1218–1224 (2020).
16. Englmeier, R. & Förster, F. Cryo-electron tomography for the structural study of mitochondrial translation. *Tissue and Cell* **57,** 129–138 (2019).
17. Weiner, E., Pinskey, J. M., Nicastro, D. & Otegui, M. S. Electron microscopy for imaging organelles in plants and algae. *Plant Physiology* **188,** 713–725 (2022).
18. Sun, S. Y., Segev-Zarko, L.-a., Pintilie, G. D., Kim, C. Y., Staggers, S. R., Schmid, M. F., Egan, E. S., Chiu, W. & Boothroyd, J. C. Cryogenic electron tomography reveals novel structures in the apical complex of Plasmodium falciparum. *Mbio,* e02864–23 (2024).
19. Wu, G.-H., Smith-Geater, C., Galaz-Montoya, J. G., Gu, Y., Gupte, S. R., Aviner, R., Mitchell, P. G., Hsu, J., Miramontes, R., Wang, K. Q., *et al.* CryoET reveals organelle phenotypes in huntington disease patient iPSC-derived and mouse primary neurons. *Nature communications* **14,** 692 (2023).
20. Hsieh, C.-E., Marko, M., Frank, J. & Mannella, C. A. Electron tomographic analysis of frozen-hydrated tissue sections. *Journal of Structural Biology* **138,** 63–73 (2002).
21. Schaffer, M., Pfeffer, S., Mahamid, J., Kleindiek, S., Laugks, T., Albert, S., Engel, B. D., Rummel, A., Smith, A. J., Baumeister, W., *et al.* A cryo-FIB lift-out technique enables molecular-resolution cryo-ET within native Caenorhabditis elegans tissue. *Nature methods* **16,** 757–762 (2019).
22. Gaisin, V. A., Kooger, R., Grouzdev, D. S., Gorlenko, V. M. & Pilhofer, M. Cryo-electron tomography reveals the complex ultrastructural organization of multicellular filamentous Chloroflexota (Chloroflexi) bacteria. *Frontiers in microbiology* **11,** 548235 (2020).
23. Wang, S., Zhou, H., Chen, W., Jiang, Y., Yan, X., You, H. & Li, X. CryoFIB milling large tissue samples for cryo-electron tomography. *Scientific Reports* **13,** 5879 (2023).
24. Dudek, N. K., Galaz-Montoya, J. G., Shi, H., Mayer, M., Danita, C., Celis, A. I., Viehboeck, T., Wu, G.-H., Behr, B., Bulgheresi, S., *et al.* Previously uncharacterized rectangular bacterial structures in the dolphin mouth. *Nature Communications* **14,** 2098 (2023).
25. Grünewald, K. & Cyrklaff, M. Structure of complex viruses and virus-infected cells by electron cryo tomography. *Current opinion in microbiology* **9,** 437–442 (2006).
26. Bharat, T. A., Riches, J. D., Kolesnikova, L., Welsch, S., Krähling, V., Davey, N., Parsy, M.-L., Becker, S. & Briggs, J. A. Cryo-electron tomography of Marburg virus particles and their morphogenesis within infected cells. *PLoS biology* **9,** e1001196 (2011).
27. Murata, K., Zhang, Q., Gerardo Galaz-Montoya, J., Fu, C., Coleman, M. L., Osburne, M. S., Schmid, M. F., Sullivan, M. B., Chisholm, S. W. & Chiu, W. Visualizing adsorption of cyanophage P-SSP7 onto marine Prochlorococcus. *Scientific Reports* **7,** 44176 (2017).
28. Jin, J., Galaz-Montoya, J. G., Sherman, M. B., Sun, S. Y., Goldsmith, C. S., O'Toole, E. T., Ackerman, L., Carlson, L.-A., Weaver, S. C., Chiu, W., *et al.* Neutralizing antibodies inhibit chikungunya virus budding at the plasma membrane. *Cell host & microbe* **24,** 417–428 (2018).
29. Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., Stanifer, M. L., Boulant, S., Bartenschlager, R. & Chlanda, P. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nature communications* **11,** 5885 (2020).
30. Zimmermann, L. & Chlanda, P. Cryo-electron tomography of viral infection—from applications to biosafety. *Current Opinion in Virology* **61,** 101338 (2023).

31. Wang, Y., Huo, T., Tseng, Y.-J., Dang, L., Yu, Z., Yu, W., Foulks, Z., Murdaugh, R. L., Ludtke, S. J., Nakada, D., *et al.* Using Cryo-ET to distinguish platelets during pre-acute myeloid leukemia from steady state hematopoiesis. *Communications Biology* **5,** 72 (2022).
32. Galaz-Montoya, J. G. *The advent of preventive high-resolution structural histopathology by artificial intelligence-powered cryogenic electron tomography* Mar. 2024. osf.io/avr2z.
33. Briggs, J. A. Structural biology in situ—the potential of subtomogram averaging. *Current opinion in structural biology* **23,** 261–267 (2013).
34. Galaz-Montoya, J. G. & Ludtke, S. J. The advent of structural biology in situ by single particle cryo-electron tomography. *Biophysics reports* **3,** 17–35 (2017).
35. Mattei, S., Glass, B., Hagen, W. J., Kräusslich, H.-G. & Briggs, J. A. The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354,** 1434–1437 (2016).
36. Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature methods* **15,** 955–961 (2018).
37. Sutton, G., Sun, D., Fu, X., Kotecha, A., Hecksel, C. W., Clare, D. K., Zhang, P., Stuart, D. I. & Boyce, M. Assembly intermediates of orthoreovirus captured in the cell. *Nature communications* **11,** 4445 (2020).
38. Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nature methods* **18,** 186–193 (2021).
39. Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. & Ranson, N. A. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* **100,** 3–15 (2016).
40. Weis, F., Hagen, W. J., Schorb, M. & Mattei, S. Strategies for optimization of cryogenic electron tomography data acquisition. *JoVE (Journal of Visualized Experiments),* e62383 (2021).
41. Radermacher, M. Weighted back-projection methods. *Electron tomography: methods for three-dimensional visualization of structures in the cell,* 245–273 (2006).
42. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *Journal of structural biology* **116,** 71–76 (1996).
43. Frangakis, A. S. It's noisy out there! A review of denoising techniques in cryo-electron tomography. *Journal of Structural Biology* **213,** 107804 (2021).
44. Radermacher, M. Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *Journal of electron microscopy technique* **9,** 359–394 (1988).
45. Mastronarde, D. N. Dual-axis tomography: an approach with alignment methods that preserve resolution. *Journal of structural biology* **120,** 343–352 (1997).
46. Lanzavecchia, S., Cantele, F., Bellon, P., Zampighi, L., Kreman, M., Wright, E. & Zampighi, G. Conical tomography of freeze-fracture replicas: a method for the study of integral membrane proteins inserted in phospholipid bilayers. *Journal of structural biology* **149,** 87–98 (2005).
47. Moebel, E. & Kervrann, C. A Monte Carlo framework for missing wedge restoration and noise removal in cryo-electron tomography. *Journal of Structural Biology: X* **4,** 100013 (2020).
48. Paavolainen, L., Acar, E., Tuna, U., Peltonen, S., Moriya, T., Soonsawad, P., Marjomäki, V., Cheng, R. H. & Ruotsalainen, U. Compensation of missing wedge effects with sequential statistical reconstruction in electron tomography. *PloS one* **9,** e108978 (2014).
49. Yan, R., Venkatakrishnan, S. V., Liu, J., Bouman, C. A. & Jiang, W. MBIR: A cryo-ET 3D reconstruction method that effectively minimizes missing wedge artifacts and restores missing information. *Journal of structural biology* **206,** 183–192 (2019).
50. Goris, B., Van den Broek, W., Batenburg, K. J., Mezerji, H. H. & Bals, S. Electron tomography based on a total variation minimization reconstruction technique. *Ultramicroscopy* **113,** 120–130 (2012).
51. Deng, Y., Chen, Y., Zhang, Y., Wang, S., Zhang, F. & Sun, F. ICON: 3D reconstruction with 'missing-information'restoration in biological electron tomography. *Journal of structural biology* **195,** 100–112 (2016).
52. Ronneberger, O., Fischer, P. & Brox, T. *U-net: Convolutional networks for biomedical image segmentation* in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (2015), 234–241.
53. Buchholz, T.-O., Jordan, M., Pigino, G. & Jug, F. *Cryo-CARE: Content-aware image restoration for cryo-transmission electron microscopy data* in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (2019), 502–506.
54. Wiedemann, S. & Heckel, R. A Deep Learning Method for Simultaneous Denoising and Missing Wedge Reconstruction in Cryogenic Electron Tomography. *arXiv preprint arXiv:2311.05539* (2023).
55. He, J., Zhang, Y., Sun, W., Yang, G. & Sun, F. IsoVEM: Isotropic Reconstruction for Volume Electron Microscopy Based on Transformer. *bioRxiv,* 2023–11 (2023).
56. Antun, V., Renna, F., Poon, C., Adcock, B. & Hansen, A. C. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences* **117,** 30088–30095 (2020).
57. Laves, M.-H., Tölle, M. & Ortmaier, T. *Uncertainty estimation in medical image denoising with bayesian deep image prior* in *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis: Second International Workshop, UNSURE 2020, and Third International Workshop, GRAIL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2* (2020), 81–96.
58. Bhadra, S., Kelkar, V. A., Brooks, F. J. & Anastasio, M. A. On hallucinations in tomographic image reconstruction. *IEEE transactions on medical imaging* **40,** 3249–3260 (2021).
59. Huang, Y., Würfl, T., Breininger, K., Liu, L., Lauritsch, G. & Maier, A. *Some investigations on robustness of deep learning in limited angle tomography* in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I* (2018), 145–153.
60. Ader, N. R. & Kukulski, W. in *Methods in cell biology* 303–320 (Elsevier, 2017).
61. Kukulski, W., Schorb, M., Welsch, S., Picco, A., Kaksonen, M. & Briggs, J. A. Correlated fluorescence and 3D electron microscopy with high sensitivity and spatial precision. *Journal of Cell Biology* **192,** 111–119 (2011).
62. Ulyanov, D., Vedaldi, A. & Lempitsky, V. *Deep image prior* in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), 9446–9454.

63. Van Veen, D., Jalal, A., Soltanolkotabi, M., Price, E., Vishwanath, S. & Dimakis, A. G. Compressed sensing with deep image prior and learned regularization. *arXiv preprint arXiv:1806.06438* (2018).

64. Heckel, R. & Hand, P. Deep decoder: Concise image representations from untrained non-convolutional networks. *arXiv preprint arXiv:1810.03982* (2018).

65. Sitzmann, V., Martel, J., Bergman, A., Lindell, D. & Wetzstein, G. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems* **33,** 7462–7473 (2020).

66. Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. & Ng, R. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems* **33,** 7537–7547 (2020).

67. Lindell, D. B., Van Veen, D., Park, J. J. & Wetzstein, G. *Bacon: Band-limited coordinate networks for multiscale scene representation* in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022), 16252–16262.

68. Shen, L., Pauly, J. & Xing, L. NeRP: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems* (2022).

69. Liu, R., Sun, Y., Zhu, J., Tian, L. & Kamilov, U. Zero-shot learning of continuous 3D refractive index maps from discrete intensity-only measurements. *arXiv preprint arXiv:2112.00002* (2021).

70. Zang, G., Idoughi, R., Li, R., Wonka, P. & Heidrich, W. *IntraTomo: self-supervised learning-based tomography via sinogram synthesis and prediction* in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), 1960–1970.

71. Van Veen, D., Van der Sluijs, R., Ozturkler, B., Desai, A., Bluethgen, C., Boutin, R. D., Willis, M. H., Wetzstein, G., Lindell, D., Vasanawala, S., *et al.* Scale-agnostic super-resolution in mri using feature-based coordinate networks. *arXiv preprint arXiv:2210.08676* (2022).

72. Wolterink, J. M., Zwienenberg, J. C. & Brune, C. *Implicit neural representations for deformable image registration* in *International Conference on Medical Imaging with Deep Learning* (2022), 1349–1359.

73. Kunz, J. F., Ruschke, S. & Heckel, R. Implicit neural networks with fourier-feature inputs for free-breathing cardiac MRI reconstruction. *arXiv preprint arXiv:2305.06822* (2023).

74. Galaz-Montoya, J. G., Flanagan, J., Schmid, M. F. & Ludtke, S. J. Single particle tomography in EMAN2. *Journal of structural biology* **190,** 279–290 (2015).

75. Wang, C., Tu, J., Liu, J. & Molineux, I. *Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. Nat Microbiol 4: 1049–1056* 2019.

76. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8 Åresolution. *Journal of molecular biology* **194,** 531–544 (1987).

77. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L. & Lerer, A. Automatic differentiation in pytorch (2017).

78. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. & Ludtke, S. J. EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* **157,** 38–46 (2007).

79. Liu, Y.-T., Zhang, H., Wang, H., Tao, C.-L., Bi, G.-Q. & Zhou, Z. H. Isotropic reconstruction for electron tomography with deep learning. *Nature communications* **13,** 6482 (2022).

80. Shi, Z., Mettes, P., Maji, S. & Snoek, C. G. On measuring and controlling the spectral bias of the deep image prior. *International Journal of Computer Vision* **130,** 885–908 (2022).

81. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13,** 600–612 (2004).

82. Flynn, J. R., Ward, S., Abich, J. & Poole, D. *Image quality assessment using the ssim and the just noticeable difference paradigm* in *Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition: 10th International Conference, EPCE 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part I 10* (2013), 23–30.

83. Sheikh, H. R. & Bovik, A. C. *A visual information fidelity approach to video quality assessment* in *The first international workshop on video processing and quality metrics for consumer electronics* **7** (2005), 2117–2128.

84. Mason, A., Rioux, J., Clarke, S. E., Costa, A., Schmidt, M., Keough, V., Huynh, T. & Beyea, S. Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of MR images. *IEEE transactions on medical imaging* **39,** 1064–1072 (2019).

85. Akçakaya, M., Yaman, B., Chung, H. & Ye, J. C. Unsupervised deep learning methods for biological image reconstruction and enhancement: An overview from a signal processing perspective. *IEEE Signal Processing Magazine* **39,** 28–44 (2022).

86. Dimakis, A. G. in *Mathematical Aspects of Deep Learning* (eds Grohs, P. & Kutyniok, G.) 400–421 (Cambridge University Press, 2022).

87. Dimakis, A. G. Deep generative models and inverse problems.

88. Moussavi, F., Heitz, G., Amat, F., Comolli, L. R., Koller, D. & Horowitz, M. 3D segmentation of cell boundaries from whole cell cryogenic electron tomography volumes. *Journal of structural biology* **170,** 134–145 (2010).

89. Hecksel, C. W., Darrow, M. C., Dai, W., Galaz-Montoya, J. G., Chin, J. A., Mitchell, P. G., Chen, S., Jakana, J., Schmid, M. F. & Chiu, W. Quantifying variability of manual annotation in cryo-electron tomograms. *Microscopy and Microanalysis* **22,** 487–496 (2016).

90. Leschziner, A. E. & Nogales, E. The orthogonal tilt reconstruction method: an approach to generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles. *Journal of structural biology* **153,** 284–299 (2006).

91. Scheres, S. H., Melero, R., Valle, M. & Carazo, J.-M. Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. *Structure* **17,** 1563–1572 (2009).

92. Tan, Y. Z., Baldwin, P. R., Davis, J. H., Williamson, J. R., Potter, C. S., Carragher, B. & Lyumkis, D. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature methods* **14,** 793–796 (2017).

93. Sorzano, C. O. S., Semchonok, D., Lin, S.-C., Lo, Y.-C., Vilas, J. L., Jiménez-Moreno, A., Gragera, M., Vacca, S., Maluenda, D., Martínez, M., *et al.* Algorithmic robustness to preferred orientations in single particle analysis by CryoEM. *Journal of Structural Biology* **213,** 107695 (2021).

94. Kerbl, B., Kopanas, G., Leimkühler, T. & Drettakis, G. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* **42,** 1–14 (2023).

95. Chen, M. & Ludtke, S. J. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nature methods* **18,** 930–936 (2021).

96. Chen, M., Schmid, M. F. & Chiu, W. Improving resolution and resolvability of single-particle cryoEM structures using Gaussian mixture models. *Nature methods* **21,** 37–40 (2024).

97. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M. & Aila, T. Noise2Noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189* (2018).

98. Moran, N., Schmidt, D., Zhong, Y. & Coady, P. *Noisier2noise: Learning to denoise from unpaired noisy data* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), 12064–12072.

99. Winkler, H. & Taylor, K. A. Focus gradient correction applied to tilt series image data used in electron tomography. *Journal of structural biology* **143,** 24–32 (2003).

100. Fernandez, J., Li, S. & Crowther, R. CTF determination and correction in electron cryotomography. *Ultramicroscopy* **106,** 587–596 (2006).

101. Jensen, G. J. & Kornberg, R. D. Defocus-gradient corrected back-projection. *Ultramicroscopy* **84,** 57–64 (2000).

102. Galaz-Montoya, J. G., Hecksel, C. W., Baldwin, P. R., Wang, E., Weaver, S. C., Schmid, M. F., Ludtke, S. J. & Chiu, W. Alignment algorithms and per-particle CTF correction for single particle cryo-electron tomography. *Journal of structural biology* **194,** 383–394 (2016).

103. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature methods* **16,** 1146–1152 (2019).

104. Wagner, J., Schaffer, M. & Fernández-Busnadiego, R. Cryo-electron tomography—the cell biology that came in from the cold. *FEBS letters* **591,** 2520–2533 (2017).

105. Marx, V. Calling cell biologists to try cryo-ET. *Nature methods* **15,** 575–578 (2018).

A Appendix

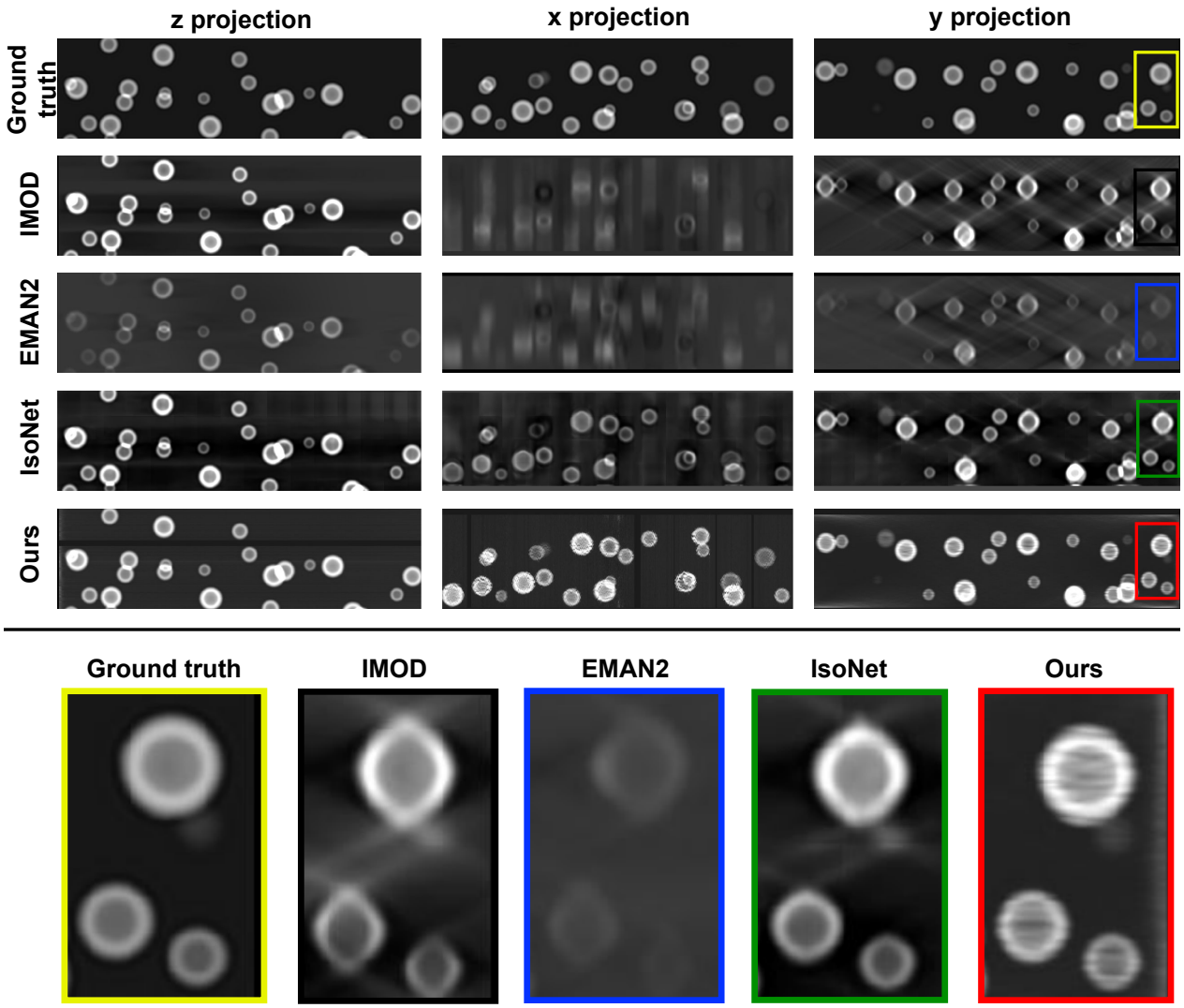

Figure A1 | Qualitative review of the spheres dataset. **Top:** Projections in each direction (columns) for each method (rows). **Bottom:** Zoom inset of the xz-plane. IMOD and EMAN2 suffer from back-projection artifacts. Compared to IsoNet, ours better resolves these artifacts and also produces a more spherical shape. Reconstruction artifacts are present in both IsoNet (tile pattern) and our (horizontal streaks) reconstructions due to computational constraints, as discussed in Section 4.

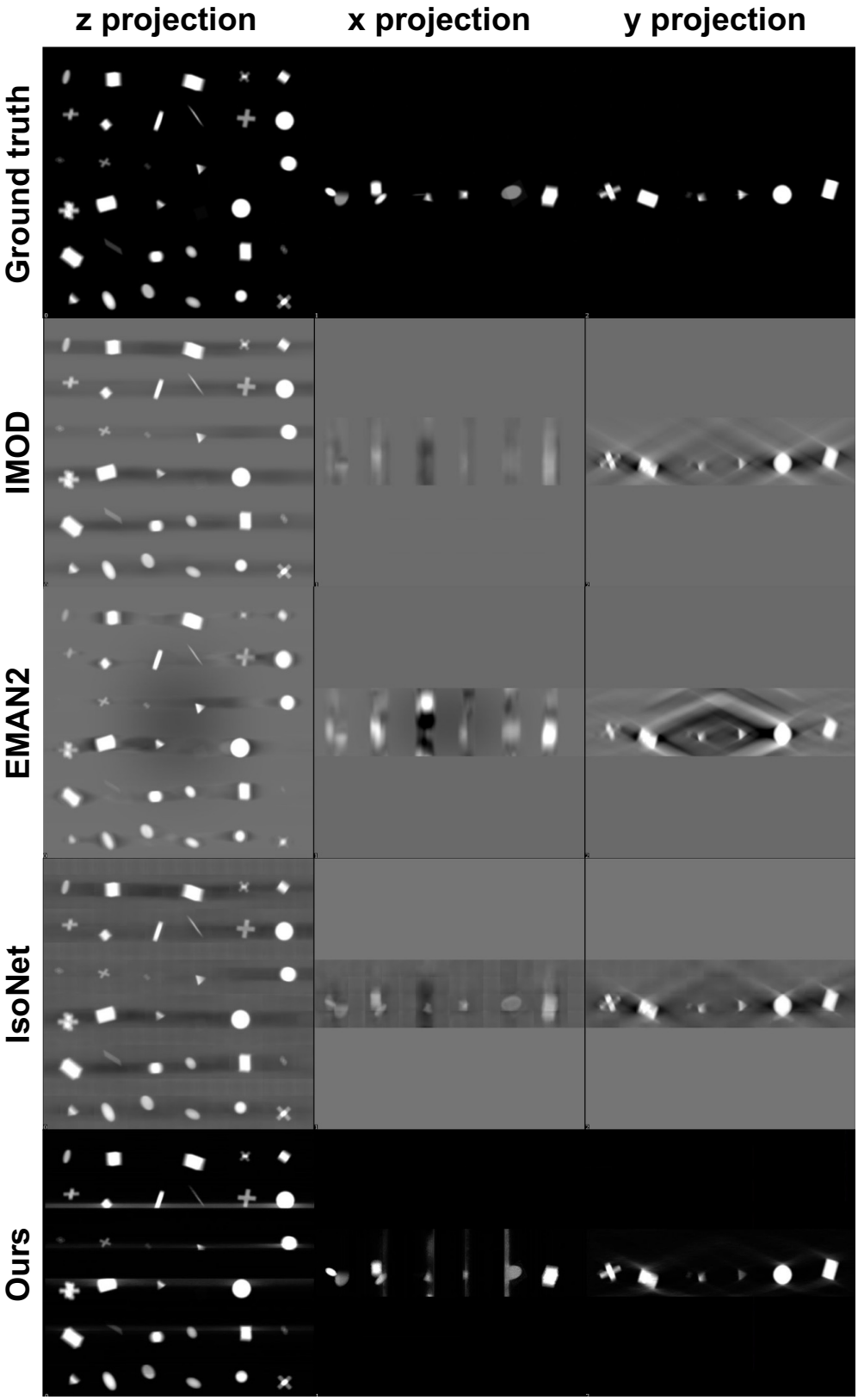

Figure A2 | Qualitative review of the geometric shapes dataset: projections in each direction (columns) for each method (rows). Ours provides the best shape completion and mitigation of back-projection artifacts.

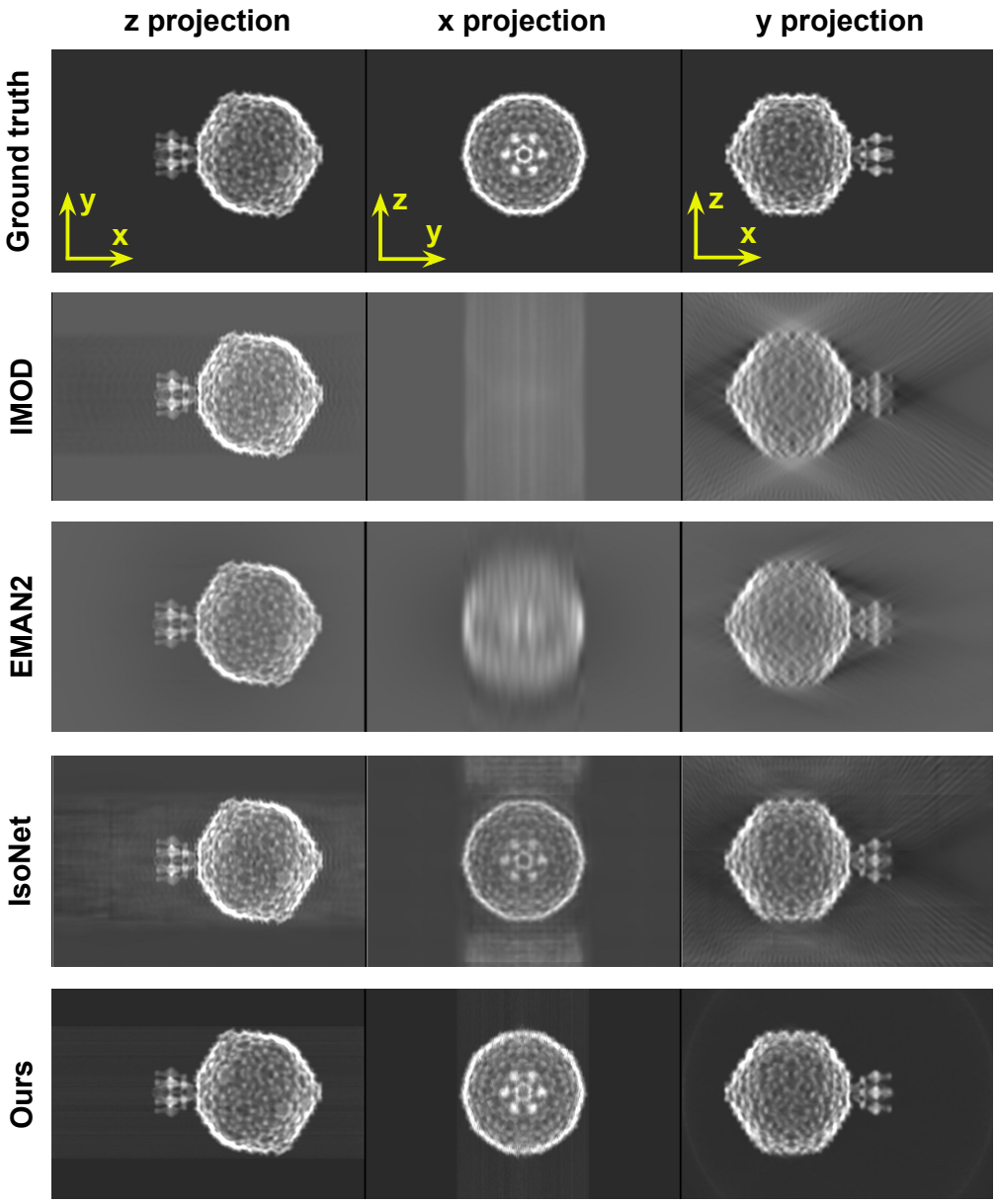

Figure A3 | Qualitative review of the P22 dataset: projections in each direction (columns) for each method (rows). Ours provides the best shape completion and mitigation of back-projection artifacts.

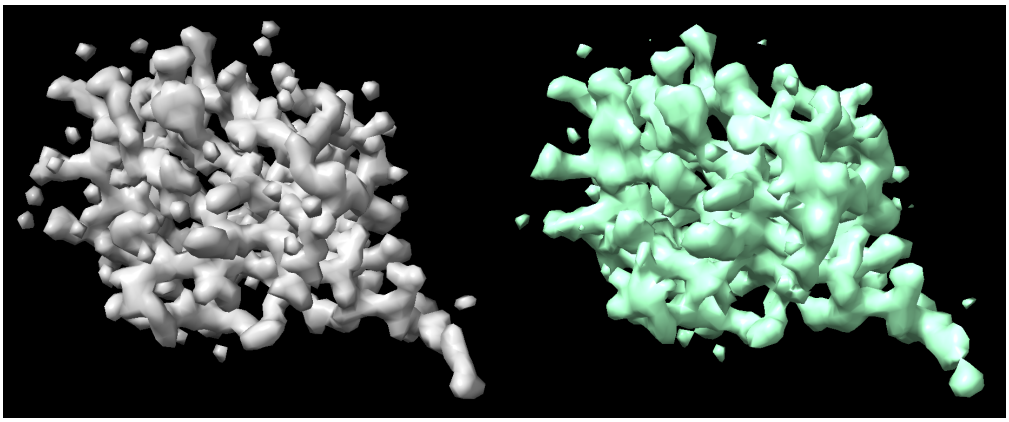

Figure A4 | Qualitative review of the ubiquitin dataset comparing 3D volumes of the ground truth (left) vs. ours (right) to emphasize that streaky artifacts are absent when a single network is used to fit the tomogram, as discussed in Section 2.4. IsoNet also exhibits reconstruction artifacts due to computational constraints, as displayed in Figure 3. This motivates potential improvement for both methods and the development of hybrid approaches to more efficiently and accurately represent large cryoET volumes.