Supplementary Information for
Scalable 3D Reconstruction for X-Ray Single Particle Imaging With Online Machine Learning

Jay Shenoy[1,2], Axel Levy[3,2], Kartik Ayyer[4,5,6], Frédéric Poitevin[2*], Gordon Wetzstein[3*]

This PDF file includes:

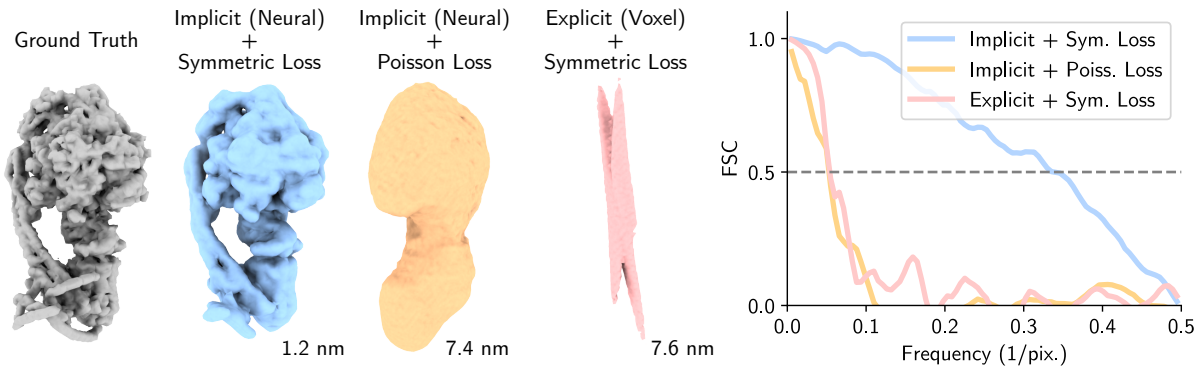
[1]Department of Computer Science, Stanford University, Stanford, CA, USA
[2]LCLS, SLAC National Accelerator Laboratory, Menlo Park, CA, USA
[3]Department of Electrical Engineering, Stanford University, Stanford, CA, USA
[4]Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany
[5]Center for Free-Electron Laser Science, Hamburg, Germany
[6]The Hamburg Centre for Ultrafast Imaging, Hamburg, Germany
*Corresponding authors. E-mail: frederic.poitevin@stanford.edu, gordon.wetzstein@stanford.edu

Supplementary Note 1: Ablation study

We conduct an ablation study and report the results in Fig. 1. Specifically, we analyze the importance of the implicit neural representation ($V_\theta : \mathbb{R}^3 \to \mathbb{R}_+$) by replacing it with an explicit representation ($V_\theta \in \mathbb{R}_+^{N\times N\times N}$) (i.e., a voxel grid) and that of the symmetric loss function by replacing it with a naive Poisson negative log-likelihood loss

$$\mathcal{L}(\psi, \theta; t) = \sum_{I\in\mathcal{B}_t} \text{PL}(\Gamma(V_\theta, f_\psi(Y)), I) \tag{1}$$

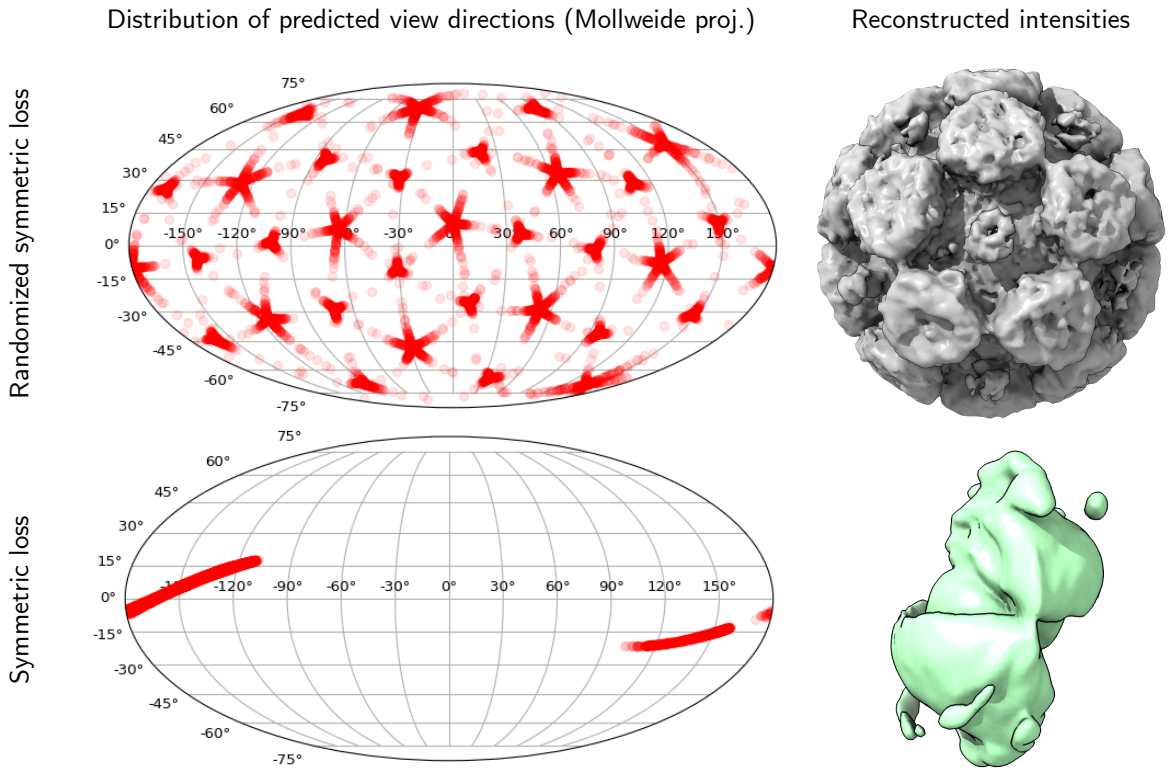
where $Y = \text{IFT}(\sqrt{I})$. Our results show that both components are necessary for converging to the correct density.



Supplementary Fig. 1 Ablation study on the symmetric loss function and on the implicit neural representation. (Left) Reconstructed densities with X-RAI (implicit representation + symmetric loss) and comparison with modified versions of the framework (Poisson negative log-likelihood loss instead of symmetric loss and voxel-based representation instead of neural representation). (Right) Associated Fourier Shell Correlations.

Supplementary Note 2: Replication scheme on PR772

We illustrate the importance of the randomized symmetric loss in Fig. 2. With the randomized symmetric loss, predicted poses are randomly rotated by a symmetry rotation of the icosahedron. This effectively enables the intensity volume to be supervised everywhere in $\mathbb{R}^3$, although the encoder only maps images to a subset of $\mathrm{SO}(3)$.



Supplementary Fig. 2 Importance of the randomized symmetric loss when performing reconstruction on the PR772 dataset [1]. Due to the icosahedral symmetry of the virus, diffraction patterns associated with different poses can be indistinguishable (except for the noise, which should not be a source of information). Therefore, the encoder maps images to a subset of $\mathrm{SO}(3)$ (bottom left). Without the replication scheme, the intensity volume, supervised on a subset of $\mathbb{R}^3$, is inaccurate (bottom right). With the replication scheme, the encoder spreads out predicted poses over $\mathrm{SO}(3)$ (upper left), and the intensity volume is supervised everywhere (upper right).

Supplementary Note 3: Datasets

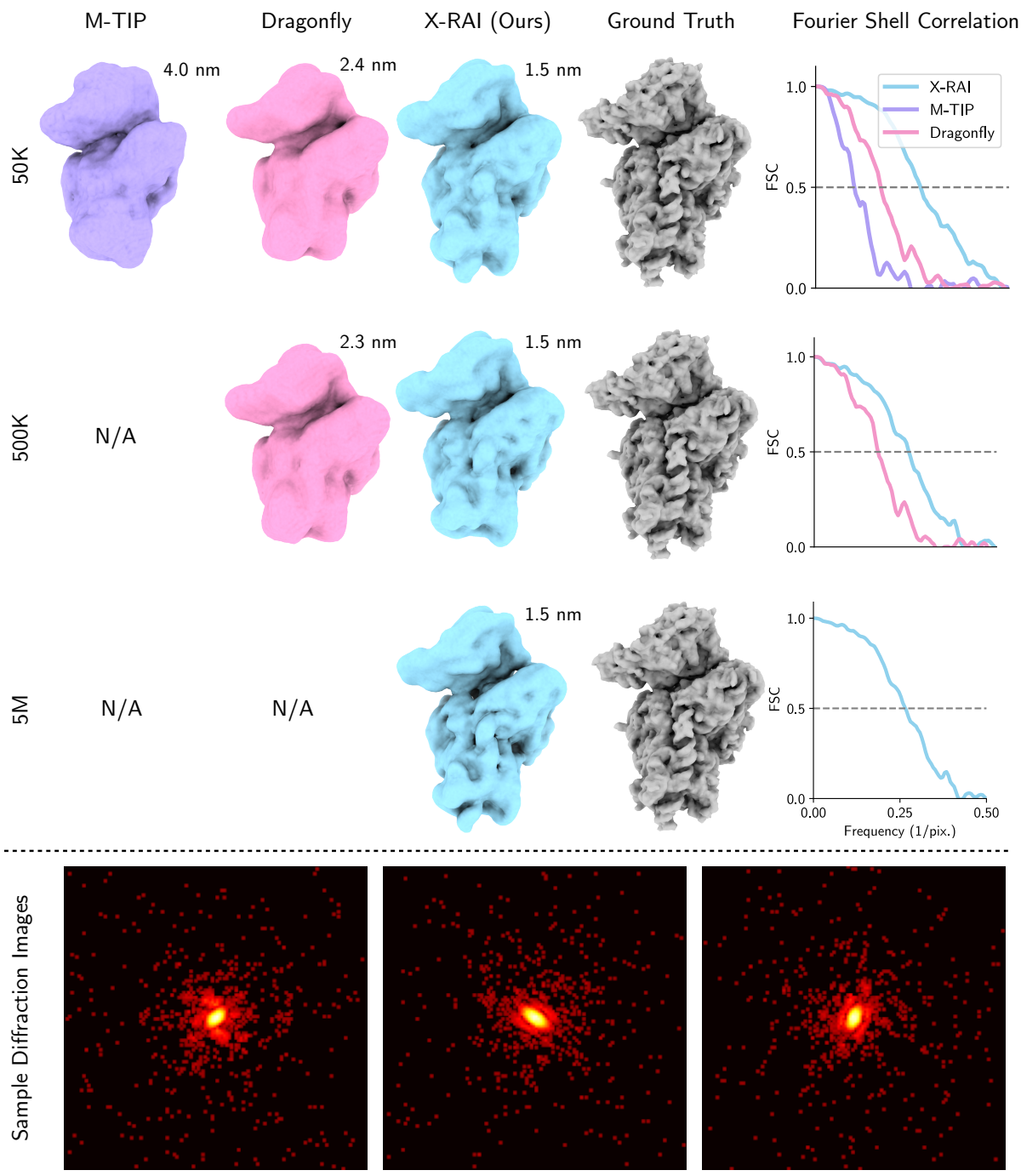
We simulate all synthetic datasets using the Skopi software package [2]. All the simulation parameters, as well as the statistics of the synthetic and experimental datasets, are detailed in Table 1.

ID	Type	# Images	Res. (pix.)	Photons per Pulse	Beam Fluence (mJ / μm^2)	Photons per Frame	Corner Res.
7R4X [3]	Synthetic	50K, 500K, 5M	128×128	10^{13}	9.38	1.1×10^4	8.0 Å
6J5I [4]	Synthetic	50K, 500K, 5M	128×128	10^{14}	93.8	3.3×10^4	8.0 Å
5O60 [5]	Synthetic	5M	128×128	6×10^{12}	5.63	8.7×10^3	8.0 Å
5O60	Synthetic	5M	128×128	10^{13}	9.38	1.5×10^4	8.0 Å
5O60	Synthetic	5M	128×128	10^{11}	0.02	3.3×10^3	1.4 Å
PR772 [1]	Experimental	14,772	260×257	10^{13}	—-	4.0×10^5	83 Å
Gold [6]	Experimental	566,390	331×319	2.6×10^{12}	—-	2.4×10^3	33 Å

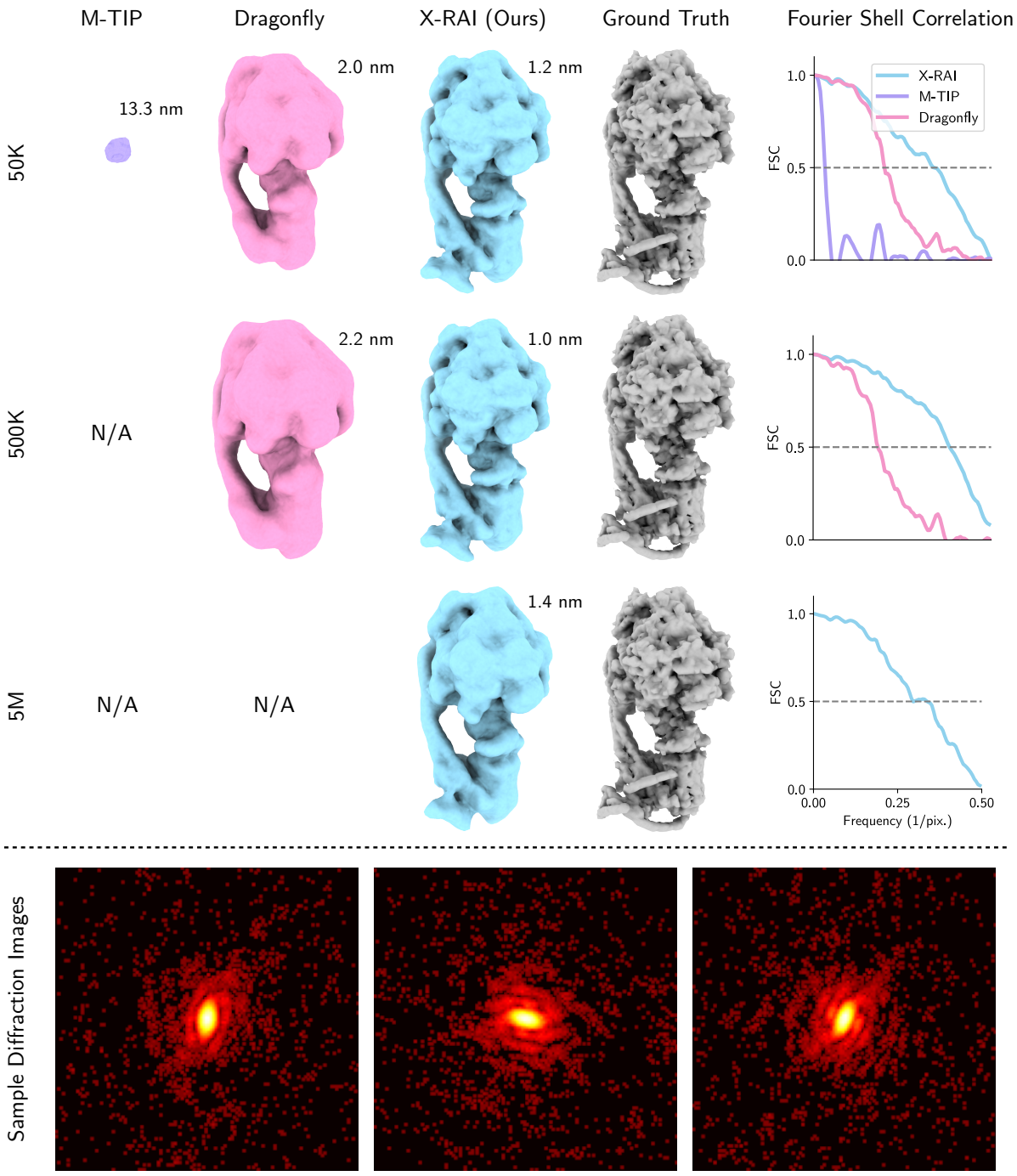
Supplementary Table 1 Overview of synthetic and experimental datasets used for reconstruction. For all simulations, we use a photon energy of 4.6 keV, a beam radius of 0.5 microns, and a detector with a side length of 0.1 meters that is placed 0.2 meters away from the sample. For the experimental PR772 dataset, we first crop the images to resolution 256×256 and then downsample them to 128×128 before running reconstruction. For the gold nanoparticle, we use the cub42 dataset from [6], filtering the original dataset of 2,234,232 images down to 566,390 images using 2D classification. We use the low-frequency portion of this dataset to determine the orientation estimates for all the images via end-to-end optimization with X-RAI and subsequently construct the intensity volume via trilinear interpolation on an explicit voxel grid using the full-resolution dataset.

Supplementary Note 4: Qualitative reconstruction results on all synthetic datasets

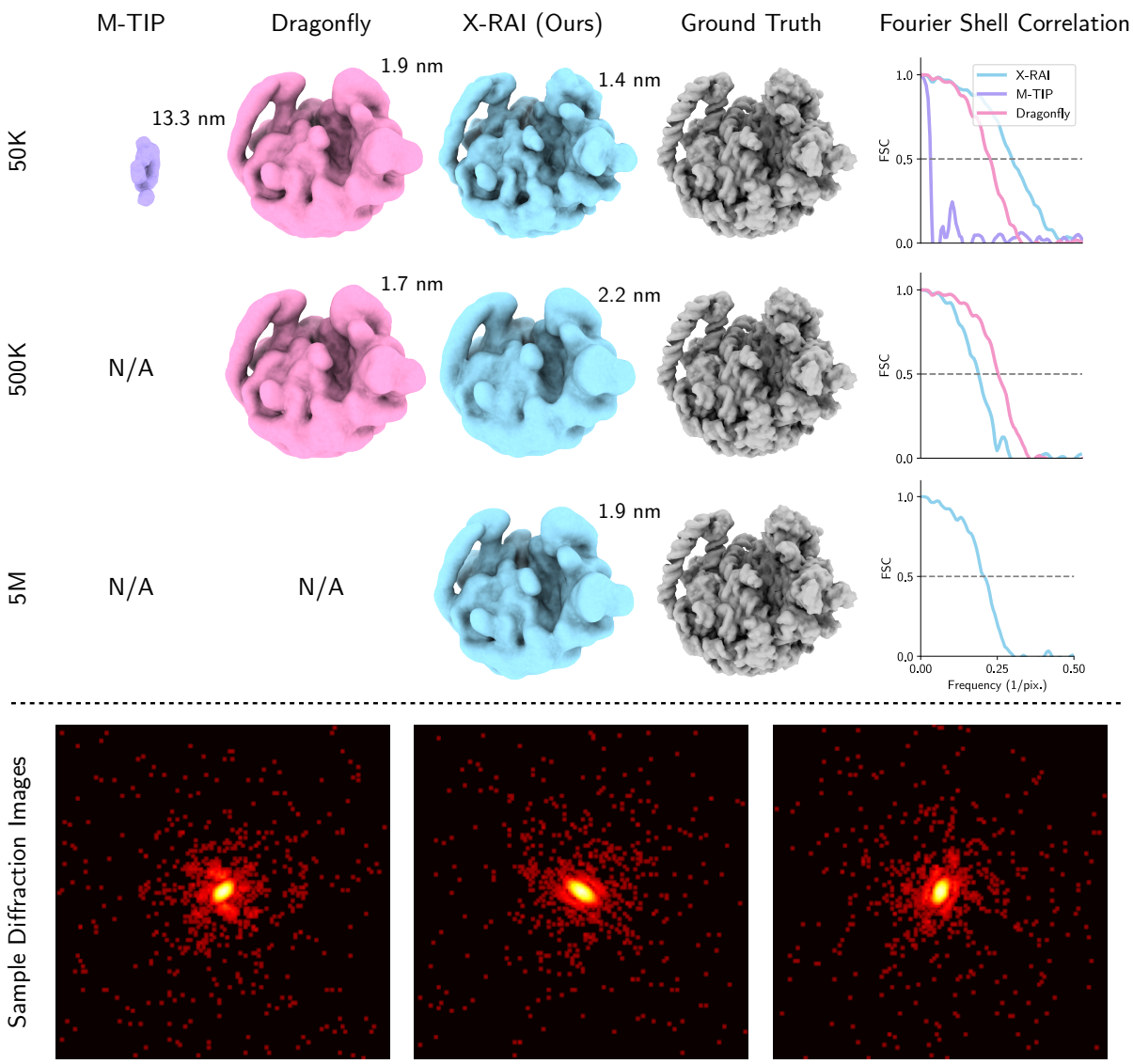
We present comprehensive reconstruction results on all of the synthetic datasets shown in Table 1 of the main paper. M-TIP fails to reconstruct datasets containing 500 thousand or 5 million images within the allotted time of 48 hours, while Dragonfly times out when processing 5 million images. The results are shown in figures 3, 4, and 5.



Supplementary Fig. 3 Comparison of X-RAI, M-TIP, and Dragonfly on reconstructing the 40S ribosomal subunit (PDB: 7R4X). The label "N/A" for M-TIP or Dragonfly indicates failure to process the respective datasets in the allotted time of 48 hours.



Supplementary Fig. 4 Comparison of X-RAI, M-TIP, and Dragonfly on reconstructing the ATP synthase molecule (PDB: 6J5I). The label "N/A" for M-TIP or Dragonfly indicates failure to process the respective datasets in the allotted time of 48 hours.

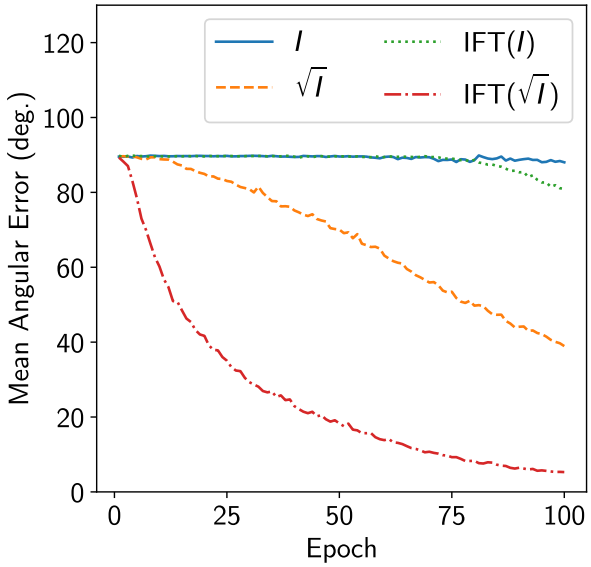


Supplementary Fig. 5 Comparison of X-RAI, M-TIP, and Dragonfly on reconstructing the 50S ribosomal subunit (PDB: 5O60). The label "N/A" for M-TIP or Dragonfly indicates failure to process the respective datasets in the allotted time of 48 hours.

Supplementary Note 5: Ablation study on input transformations to the encoder

If I denotes a single diffraction image, we find that inputting IFT($\sqrt{I}$) to the encoder yields the best performance for end-to-end reconstruction, where IFT is the inverse Fourier transform. To demonstrate the efficacy of this choice of input transformation, we perform an experiment in which we train the encoder to learn the mapping between images and poses by supervising directly on the true poses for each diffraction image. We compare the pose estimation accuracy for four different input transformations: I (no transform), $\sqrt{I}$, IFT(I), and IFT($\sqrt{I}$). As seen in figure 6, the encoder converges the fastest when using IFT($\sqrt{I}$) as input, when evaluated on the mean error of the out-of-plane viewing angle.

For training, we use a synthetic dataset of the ATP synthase molecule (PDB: 6J5I) containing 50,000 images simulated at a fluence of 93.8 mJ/μm^2, with an average of 3.3×10^4 photons per frame.



Supplementary Fig. 6 Comparison of four different input image transformations when training the encoder supervised directly on ground truth poses. The mean angular error of the out-of-plane viewing direction is plotted on the y-axis. The input IFT($\sqrt{I}$) yields the fastest convergence in this experiment.

References

[1] Reddy, H. K. *et al.* Coherent soft x-ray diffraction imaging of coliphage pr772 at the linac coherent light source. *Scientific data* **4**, 1–9 (2017).

[2] Peck, A. *et al.* Skopi: a simulation package for diffractive imaging of noncrystalline biomolecules. *Journal of Applied Crystallography* **55**, 1002–1010 (2022).

[3] Pellegrino, S., Dent, K. C., Spikes, T. & Warren, A. J. Cryo-em reconstruction of the human 40s ribosomal subunit at 2.15 å resolution. *Nucleic Acids Research* **51**, 4043–4054 (2023).

[4] Gu, J. *et al.* Cryo-em structure of the mammalian atp synthase tetramer bound with inhibitory protein if1. *Science* **364**, 1068–1075 (2019).

[5] Hentschel, J. *et al.* The complete structure of the mycobacterium smegmatis 70s ribosome. *Cell reports* **20**, 149–160 (2017).

[6] Ayyer, K. *et al.* 3d diffractive imaging of nanoparticle ensembles using an x-ray laser. *Optica* **8**, 15–23 (2020).