CRYOGEN: CRYOGENIC ELECTRON TOMOGRAPHY RECONSTRUCTION VIA GENERATIVE ENERGY-BASED MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Cryogenic electron tomography (Cryo-ET) is a powerful method for visualizing cellular structures in their native state (Lucic et al., 2005), but its effectiveness is limited by anisotropic resolution caused by the missing-wedge problem, complicating the interpretation of tomograms. IsoNet (Liu et al., 2022), a deep learning method, addresses these challenges by iteratively reconstructing missing-wedge information and improving the signal-to-noise ratio of tomograms. However, IsoNet relies on recursively updating its predictions, which can result in training instability and potential model collapse. In this study, we present CryoGEN, an enhanced energy-based method that effectively addresses resolution anisotropy without requiring recursive subtomogram generation. Our approach is about $10\times$ faster and offers a more stable and consistent methodology. Applying CryoGEN to various datasets, including immature HIV particles and ribosomes, demonstrates its capability to enhance structural interpretability. Moreover, CryoGEN holds significant potential for improving the functional interpretation of cellular tomograms in future high-resolution Cryo-ET studies, thereby providing substantial value and advancing progress in biological research.

1 INTRODUCTION

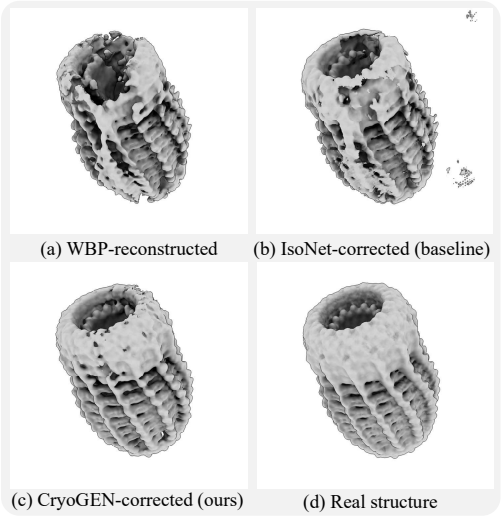


Figure 1: *Reconstructed 3D structure comparison among (a) WBP (b) IsoNet and (c) our method using simulated CryoET data of (d) C13 Vipp1 stacked rings (EMDB:18424).*

Cryo-ET is a cutting-edge technique that enables the visualization and analysis of the three-dimensional (3D) structure of biomolecules, cellular components, and even entire organisms in a near-native, hydrated state with near-atomic resolution. It offers unique insights into the molecular organization within cells, facilitating the precise identification and in-depth study of individual proteins and their interactions at subnanometer resolution. Recognizing the potential of Cryo-ET, the developers of AlphaFold3 (Abramson et al., 2024) anticipate that the increased availability of high-quality experimental data from this technique will significantly improve the model's performance in unraveling the complexity of molecular regulation within cells.

The rapid frozen, hydrated sample is imaged in a transmission electron microscope (TEM) as it is tilted through a series of angles, capturing a set of two-dimensional (2D) projections known as a "*tilt series*". These 2D images are then computationally reconstructed into a 3D model or tomogram of the sample. The most commonly used tomographic reconstruction technique is weighted back-projection (WBP) (Radermacher, 2006). However, due to the mechanical limitations of the TEM stage, the tilt range is typically restricted to around $\pm 60^\circ$. This limited tilt range results in a "*missing wedge*" in Fourier space, where data is not properly collected. The missing wedge leads to anisotropic resolution in the final WBP-reconstructed tomogram, as illustrated in Figure 1 (a), with the lowest resolution along the z-axis (the direction of the electron beam). This manifests as distortions and elongation of features

in the reconstructed 3D structure, making it challenging to accurately interpret the sample's native architecture. Figure 2 illustrates the core process of Cryo-ET.

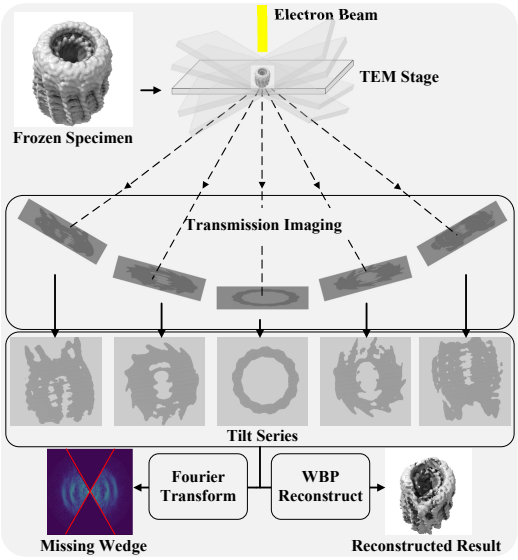


Figure 2: *Cryo-ET imaging and reconstruction.*

Indeed, the missing wedge effect is a fundamental limitation of Cryo-ET, as it is inherent to the data collection process. Efforts to mitigate this issue have included techniques such as dual-axis tomography, which collects two tilt series at perpendicular angles to reduce the missing wedge to a smaller missing pyramid. However, these methods add complexity to the data collection and processing workflows. Traditional computational approaches MBIR (Yan et al., 2019), ICON (Deng et al., 2016) utilize statistical and optimization methods to resolve the missing wedge problem. In addition to the missing wedge, Cryo-ET also suffers from a low signal-to-noise ratio (SNR) due to the low electron dose used to prevent specimen damage. Common solutions include applying various filters and deconvolution methods to partially denoise the tomograms. However, these computational strategies alone are insufficient to fully recover the lost information caused by the missing wedge, often leading to poor reconstructions with significant artefacts and distortions.

We introduce **CryoGEN**, a generative method that leverages energy models and deep learning neural networks. Our contributions can be summarized in three key aspects: a more precise problem formulation compared to previous work, the development of a novel framework for isotropic reconstruction in electron tomography, and improved generation quality by incorporating an energy-based model into our framework. This integration offers flexibility and compatibility with techniques such as generative adversarial networks (Goodfellow et al., 2014).

To be more specific, our approach consists of three main phases: first, we approach the problem from a probabilistic modeling perspective; second, we design an energy model E to capture the distribution of missing-wedge subtomograms; and third, we train a prediction model g_θ, a neural network parameterized by θ, to generate complete tomograms by combining the missing-wedge input with the energy model. Figure 1 provides an illustrative example where we compare our method with WBP and IsoNet (Liu et al., 2022), two state-of-the-art techniques.

The remainder of the paper is organized as follows: the Related Work section formulate Cryo-ET reconstruction from Bayesian perspective and further explains how it could be linked to energy-based models. The Motivation section outlines the rationale behind our method, accompanied by several illustrative examples. The Methodology section presents the objective and the complete algorithm. The Experiment section applies our method to real tomogram cases, and the final section provides the conclusion.

2 Related Work

2.1 Cryo-ET Reconstruction

Recently, deep learning-based methods have shown promise in recovering missing information and improving the signal-to-noise ratio, leading to higher contrast. IsoNet (Liu et al., 2022) has achieved notable success in filling in missing data, and many recent approaches build on it with minor modifications. Additionally, Noise2Noise-based denoising techniques, such as CryoCARE (Buchholz et al., 2019), Topaz (Bepler et al., 2020), and Warp (Tegunov & Cramer, 2019), have significantly enhanced volume clarity. More recently, several approaches (Wiedemann & Heckel, 2024; Zeng et al., 2024) have attempted to address denoising and missing wedge correction simultaneously, achieving performance comparable to two-step methods. However, current missing wedge correction techniques remain suboptimal, as they only partially restore the missing information, highlighting the need for more powerful tools.

2.2 BAYESIAN FRAMEWORK

In this context, we consider the original data x drawn from the domain $\mathcal{X}$, with the corresponding observation y from domain $\mathcal{Y}$. Specifically:

- $\mathcal{Y}$ represents the observation domain, with its data distribution denoted by p_y, where $y \sim p_y$ corresponds to a WBP-reconstructed tomogram with a missing wedge.
- $\mathcal{X}$ represents the source domain, with its data distribution denoted by p_x, where $x \sim p_x$ corresponds to an original sample whose missing wedge has been properly filled in.

We define a Cryo-ET transmission imaging operator $\mathcal{T}_M$, resulting in $\mathcal{Y} = \{\mathcal{T}_M(x) \mid x \in \mathcal{X}\}$. The imaging process is typically formulated as:

$$y = \mathcal{T}_M(x) + \epsilon_n, \quad \epsilon_n \sim \mathcal{N}(0, \sigma_n^2 I), \tag{1}$$

where ϵ_n represents additive Gaussian noise with zero mean and variance $\sigma_n^2 I$. Notably, $\mathcal{T}_M$ is generally a many-to-one operator. Our objective is to determine the x that generates the observed under-sampled y, which constitutes a classical inverse problem.

Notice that Equation (1) can be rewritten as:

$$y \sim \mathcal{N}(\mathcal{T}_M(x), \sigma_n^2 I) \tag{2}$$

By applying Bayes' theorem, we have $p(x|y) \propto p(y|x) \cdot p(x)$, which leads to:

$$\log p(x|y) = \log p(y|x) + \log p(x) + \text{constant}. \tag{3}$$

However, the distribution $p(x)$ is inaccessible in the Cryo-ET reconstruction scenario, which is also a limitation of current state-of-the-art methods. These issues will be addressed after we introduce the energy model in the following section.

2.3 ENERGY-BASED MODELS

Energy-based models (EBMs) (Lecun et al., 2006) are a type of probabilistic framework that formulates machine learning problems using the concept of *energy*. An energy function assigns lower energy values to configurations that are more likely or preferable, and higher energy values to less likely or undesirable configurations. The system's objective is to identify the states that minimize the energy and shape the energy landscape accordingly.

After defining the non-negative energy function E, the Boltzmann distribution (Boltzmann, 1974) can be expressed in terms of E as $p_x(x) = \frac{1}{Z}\exp(-E(x))$, where Z is the normalization constant, also referred to as the partition function. In this paper, we directly apply generative adversarial networks (GANs) (Goodfellow et al., 2014) to model this probabilistic framework, although other approaches, such as contrastive learning (Chen et al., 2020), could also be explored.

3 MOTIVATION

For each observation $y \in \mathcal{Y}$, we can obtain the optimal x^* through Equation 4:

$$x^*(y) = \arg\max_{x \in \mathcal{X}} \log p(x|y) = \arg\max_{x \in \mathcal{X}} \left[-\frac{1}{2\sigma_n^2} \|\mathcal{T}_M(x) - y\|_2^2 + \log p(x) \right] \tag{4}$$

However, since the sampling or optimization process is slow, a natural idea is to shift the objective towards finding a parameterized function g_θ with parameters θ that learns the mapping $g_\theta : y \mapsto x^*(y)$. This approach, which focuses on learning the mapping directly, is not immediately obvious but is fundamentally similar to the methods like (Johnson et al., 2016) (i.e., using a neural network to learn the high-probability targets rather than generating them iteratively on the fly). This concept also forms the foundation of IsoNet (Liu et al., 2022) and its variants (Buchholz et al., 2019), whose limitations we will discuss shortly, along with an explanation of why incorporating a generative model may be necessary. We consider a straightforward scenario with a single observation sample, that is, $\mathcal{Y} = \{y_0\}$. We assume that both x_1 and x_2 map exactly to y_0 under the operation $\mathcal{T}_M$, so:

$$\|\mathcal{T}_M(x_1) - y_0\|_2^2 = 0, \quad \|\mathcal{T}_M(x_2) - y_0\|_2^2 = 0.$$

Thus, the likelihood is maximized when x lies on the line connecting x_1 and x_2, as a result of the linear property of $\mathcal{T}_M$. Furthermore, we also assume $\sigma^2 < \frac{1}{4}\|x_1 - x_2\|_2^2$ (i.e., this may indicate that the features are relatively distinct in $\mathcal{X}$) and define the prior distribution of x as the following:

$$p(x) = \frac{1}{2\sqrt{2\pi\sigma^2}} \left[\exp\left(-\frac{\|x - x_1\|^2}{2\sigma^2} \right) + \exp\left(-\frac{\|x - x_2\|^2}{2\sigma^2} \right) \right].$$

Alternatively, this can be expressed using mixture of Gaussian distributions:

$$p(x) = \frac{1}{2} \left[\mathcal{N}(x \mid x_1, \sigma^2 I) + \mathcal{N}(x \mid x_2, \sigma^2 I) \right],$$

where $\mathcal{N}(x|x_i, \sigma^2 I)$ denotes a single Gaussian distribution centered at x_i with covariance matrix $\sigma^2 I$. We seek to maximize the posterior $p(g_\theta(y_0)|y_0)$, but according to IsoNet's steps (as described in Appendix A.1), the objective ultimately reduces to solving:

$$\tilde{\theta} = \arg\min_\theta \frac{1}{2} \left(\|g_\theta(y_0) - x_1\|_2^2 + \|g_\theta(y_0) - x_2\|_2^2 \right).$$

This formulation seeks the parameter $\tilde{\theta}$ that minimizes the mean squared error between the function's output $g_\theta(y_0)$ between both x_1 and x_2. However, this leads to $g_{\tilde{\theta}}(y_0) = \frac{1}{2}(x_1 + x_2)$, and thus

$$p\left(x = g_{\tilde{\theta}}(y_0)\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\|x_1 - x_2\|^2}{8\sigma^2} \right). \tag{5}$$

This result may be unfavorable because

$$p(x = x_1 \text{ or } x = x_2) = \frac{1}{2\sqrt{2\pi\sigma^2}} \left[1 + \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2} \right) \right]. \tag{6}$$

It can be shown that Equation (5) yields a lower probability compared to Equation (6) under our assumptions, meaning that the result learned by $g_{\tilde{\theta}}(y_0)$ is worse than directly choosing either x_1 or x_2. This typically happens when the probability density function of the prior distribution is *non-convex*, and thus simply averaging the minima lacks meaningful interpretation, highlighting a key limitation we have observed with IsoNet. By introducing an energy function as $E(x) = -\log p(x)$, the situation is depicted in Figure 3(a), where it is demonstrated that $E(\frac{1}{2}(x_1 + x_2))$ does not correspond to a low-energy state.

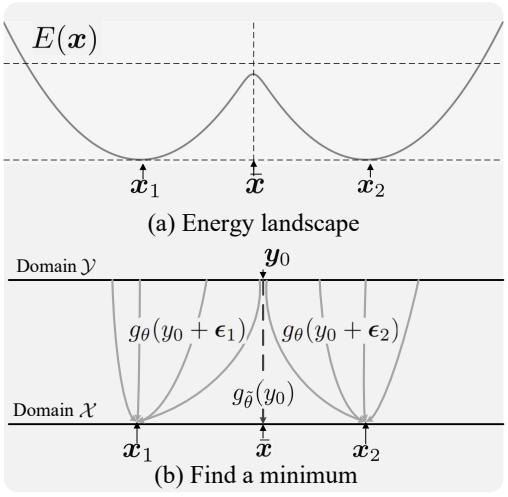


Figure 3: *Limitations of IsoNet formulation.*

The core issue stems from the one-to-many mapping, making it impossible to define a function that maps a single observation y to all possible minima x's. In other words, the problem arises because the cardinality of $g_\theta(\mathcal{Y})$ cannot exceed that of $\mathcal{Y}$, leading to the undesired outcome. A straightforward way to address this issue is by introducing a random noise through input augmentation. By adding noise to the input, represented as $y + \epsilon_g$, where $\epsilon_g \sim \mathcal{N}(0, \sigma_g^2 I)$ and the support of $\boldsymbol{\epsilon}_g$ lies in the whole real space (which has uncountably infinite cardinality).

By combining the energy model E along with $g_\theta(y + \boldsymbol{\epsilon}_g)$, we generate distributions that could match the desired target distribution, allowing us to identify an appropriate minimum rather than averaging across multiple minima. Additionally, this approach eliminates the need for real-time sampling and thus offers similar advantages to IsoNet while effectively addresses the previously noted issue. As shown in Figure 3(b), the added noise $\boldsymbol{\epsilon}_g$ helps guide the model toward different low-energy states, as we minimize E, directing the model to a local minimum instead of averaging multiple minima. More interestingly, we will demonstrate that this can be effectively approached as a generative task, as shown in Equation (9) in the next section.

4 METHODOLOGY

We define the desired source space as $\mathcal{X} = \{x \in \mathcal{X} \mid \mathcal{T}_M(x) \in \mathcal{Y}\}$, with the goal that g_θ generates samples within $\mathcal{X}$. However, we cannot directly train the energy model on the domain $\mathcal{X}$ due to the lack of real data; instead, we can only train it on observation space $\mathcal{Y}$. Consequently, if a generated sample x lies within $\mathcal{X}$, its energy $E(\mathcal{T}_M(x))$ should be low. Afterward, we define the distribution p_y using the trained energy model through the Boltzmann distribution, as described earlier.

Drawing inspiration from (Lecun et al., 2006), we do not explicitly define the energy function. Instead, we learn the energy function in a manner similar to contrastive learning, assigning low energy to samples and high energy to other regions. This method belongs to the class of *implicit probabilistic models* (Diggle & Gratton, 1984). In this section, we first define the objective to achieve this goal, followed by the presentation of the complete algorithm.

4.1 OBJECTIVE

We integrate two components for training our model: the *consistency loss* and the *posterior maximization*. The consistency loss, akin to the original loss used in IsoNet, may benefit the early stages of training but could negatively impact final model performance due to its drawbacks as mentioned in Section 3. In contrast, the energy penalty is a more sophisticated choice, but it converges more slowly.

4.1.1 CONSISTENCY LOSS

In general, we can assume that g_θ serves as an approximate inverse function of $\mathcal{T}_M$. Therefore, we first introduce a consistency loss to ensure that the inverse condition is met.:

$$\textit{Consistency Loss} = \mathbb{E}_{y\sim p_y, \epsilon\sim\mathcal{N}(0,\sigma_h^2 I)}\left[\frac{\lambda}{|\mathcal{R}|}\sum_{R\in\mathcal{R}}\|R^{-1}\circ g_\theta\circ\mathcal{T}_M\circ R\left(g_\theta(y)+\epsilon_h\right)-g_\theta(y)\|_2^2\right], \tag{7}$$

where $\circ$ denotes the composition of functions. However, if $\mathcal{T}_M$ is not a one-to-one mapping, which is likely the case, the issue outlined in Section 3 may arise. To address this during training, we can gradually reduce the weight λ of Equation (7).

Remark. Equation (7) closely resembles the objective function of IsoNet (Liu et al., 2022), except that there is no refinement step. Conversely, we decrease the penalty term λ for enforcing reconstruction, which fundamentally distinguishes our method from theirs.

4.1.2 MAXIMUM A POSTERIOR

Next, our goal is to generate results that maximize the log posterior $\log p(x|y)$ by incorporating the previously discussed energy penalty term. The objective is to minimize the error on $\mathcal{X}$ using a model trained on the $\mathcal{Y}$. Therefore, we must ensure that $E(\mathcal{T}_M(x))$ is low when $\mathcal{T}_M(x) \in \mathcal{Y}$ and high otherwise. Moreover, we assume that if $x \in \mathcal{X}$, then $R(x) \in \mathcal{X}$ as well, where R denotes a rotation operation selected from a predefined set of rotations $\mathcal{R}$ detailed in Appendix A.2. Consequently, for $x \in \mathcal{X}$, $\frac{1}{|\mathcal{R}|}\sum_{R\in\mathcal{R}} E(\mathcal{T}_M \circ R(x))$ should result in a low energy state as well. Therefore, we can define the posterior as:

$$\textit{Posterior} = \mathbb{E}_{y\sim p_y, \epsilon\sim\mathcal{N}(0,\sigma_g^2 I)}\left[-\frac{1}{|\mathcal{R}|}\sum_{R\in\mathcal{R}} E\left(\mathcal{T}_M\circ R\circ g_\theta(y+\epsilon_g)\right)+\frac{1}{2\sigma_n^2}\|\mathcal{T}_M\circ g_\theta(y+\epsilon_g)-y\|_2^2\right], \tag{8}$$

where σ_n^2 is a hyperparameter introduced in Equation (1).

4.2 ENERGY MODEL

The energy model E can be derived through several approaches. In this work, we opt to use GANs (Goodfellow et al., 2014). Specifically, we define our energy model E_ϕ as the discriminator proposed in (Goodfellow et al., 2014), which, like g_θ, is typically represented by parameterized neural

networks. This approach allows us to train both the energy model E_ϕ and g_θ simultaneously. Consequently, the energy model is learned through adversarial training, as described in Equation (9):

$$\max_{\phi} \min_{\theta} \left[\mathbb{E}_{y\sim p_y, \epsilon\sim p_\epsilon, \epsilon_g\sim\mathcal{N}(0,\sigma_g^2 I)} \frac{1}{|\mathcal{R}|} \sum_{R\in\mathcal{R}} E\left(\mathcal{T}_M \circ R \circ g_\theta(y+\epsilon_g)\right) - E(y+\epsilon) \right]. \tag{9}$$

Furthermore, we consider a similar formulation as (Arjovsky & Bottou, 2017). Let $\boldsymbol{y}$ follow the distribution p_y with support on $\mathcal{Y}$, and let $\boldsymbol{\epsilon}$ be an absolutely continuous distribution with density p_ϵ. Then, the distribution $p_{y+\epsilon}$ is also absolutely continuous with density:

$$p_{y+\epsilon}(z) = \mathbb{E}_{y\sim p_y}\left[p_\epsilon(z-y)\right] = \int_{\mathcal{Y}} p_\epsilon(z-y)\,\mathrm{d}p_y \tag{10}$$

Especially, If $\boldsymbol{\epsilon} \sim \mathcal{N}(0,\sigma^2 I)$, then $p_{y+\epsilon}(z) \propto \int_{\mathcal{Y}} e^{-\frac{\|z-y\|^2}{2\sigma^2}}\,\mathrm{d}p_y$, and Equation (9) reaches Nash equilibrium when $p_{g_\theta}(z) = p_{y+\epsilon}(z)$, as follows:

$$p_{g_\theta}(z) = \frac{1}{|\mathcal{R}|} \int_{y\in\mathcal{Y}} \int_{\epsilon_g\in\mathbb{R}^d} \sum_{R\in\mathcal{R}} \mathbb{1}_{z=\mathcal{T}_M\circ R\circ g_\theta(y+\epsilon_g)} \cdot p_{\epsilon_g}(\epsilon_g)\, p_y(y)\, d\epsilon_g\, dy. \tag{11}$$

This result is analogous to the conclusion presented in the original GAN paper by (Goodfellow et al., 2014), but in this case, the data distribution is obtained by convolving the original data distribution with a Gaussian.

4.3 Algorithm

The complete CryoGEN algorithm consists of two stages: training and inference. First, we train a generative model g_θ to minimize the pre-trained energy model. Then, we use this generative model to fill in the missing wedges of tomograms.

Building on the idea from Section 4.1, we combine both the *consistency loss* and *posterior* terms. The complete algorithm for training CryoGEN is presented in Algorithm 1.

Algorithm 1 Train prediction model.

Require: Tomogram dataset $\mathcal{Y}$, noise levels $\sigma^2, \sigma_g^2, \sigma_h^2 > 0$, estimated noise variance σ_n^2, penalty term $\lambda \geq 0$, energy model $E_\phi : \mathbb{R}^d \to \mathbb{R}^+$, prediction model $g_\theta : \mathbb{R}^d \to \mathbb{R}^d$, learning rate η.
repeat
  Randomly generate noise $\epsilon, \epsilon_g, \epsilon_h \sim \mathcal{N}(0,\sigma^2 I),\ \mathcal{N}(0,\sigma_g^2 I),\ \mathcal{N}(0,\sigma_h^2 I)$.
  Set $f(\phi,\theta) = -E_\phi\left(\mathcal{T}_M \circ R \circ g_\theta(y+\epsilon_g)\right) + 1/2\sigma_n^2 \cdot \|\mathcal{T}_M \circ g_\theta(y+\epsilon_g) - y\|_2^2$
  Update $\phi \leftarrow \phi - \eta \cdot \frac{\partial}{\partial_\phi}[E_\phi(y+\epsilon) - f(\phi,\theta)]$
  Update $\theta \leftarrow \theta - \eta \cdot \frac{\partial}{\partial_\theta}\left[f(\phi,\theta) - \lambda \cdot \|R^{-1} \circ g_\theta \circ \mathcal{T}_M \circ R\left(g_{\theta'}(y)+\epsilon_h\right) - [g_{\theta'}(y)]\,\|_2^2\right]$
  Reduce the penalty term λ and assign the value of θ to θ'
until convergence
return g_θ

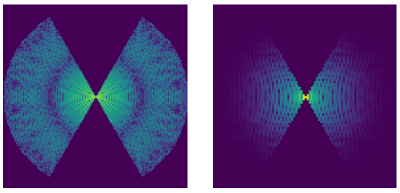
At the inference stage, we begin by cropping the complete tomogram into multiple overlapping subtomograms, which are then processed through g_θ to yield refined subtomograms. These refined subtomograms are subsequently reassembled into a complete tomogram, with the overlapping regions averaged using a weighted approach to mitigate edge effects.

5 Experiment

In this section, we compare our method to the state-of-the-art missing wedge correction technique, IsoNet, across various experiments. First, we validate our hypothesis using simple shapes, as detailed in Section 5.1. Next, we evaluate our algorithm on a simulated dataset and compare it with other approaches in Section 5.2. Finally, we apply our method to real-world examples to assess its robustness, as discussed in Section 5.3. Implementation and data processing details are provided in the Appendix, where we present results from the latest simultaneous missing wedge correction and denoising method, DeepDeWedge (Wiedemann & Heckel, 2024) as well.

5.1 SIMPLE SHAPES

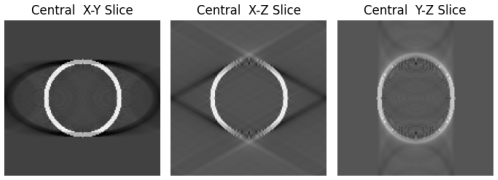
First, we apply the algorithm to simple shapes to demonstrate its effectiveness. The data is synthetically generated, with the ground truth available. Specifically, we create a 3D sphere and a triangular prism with an artificially introduced missing wedge. To further clarify, we transform the X-Z slice into the Fourier domain to highlight the presence of the missing wedge as shown in Figure 4.



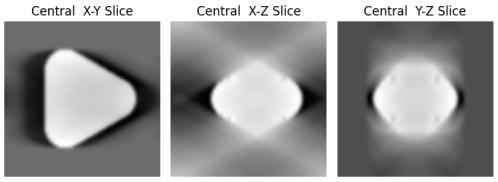
(a) *Sphere.* (b) *Prism.*

Figure 4: *Central Fourier slice of the original missing-wedged shapes.*

Our goal is to fill in these missing wedge regions, and we will demonstrate that our algorithm significantly outperforms the baseline in this task. It can be observed that the generated synthetic images exhibit lower resolution in the directions corresponding to the missing wedge (the X-Y slice closely matches the ground truth as designed) as illustrated in Figure 5.



(a) *Original missing-wedged sphere.*



(b) *Original missing-wedged prism.*

Figure 5: *Original missing-wedged shapes.*

We first apply both IsoNet and CryoGEN to the two shapes, with the results shown in Figure 6. Both methods aim to restore the corrupted regions. While IsoNet successfully reconstructs the sphere, some artefacts remain, and it fails to restore the prism, producing a distorted oval in the X-Z and Y-Z slices. In contrast, CryoGEN achieves an almost perfect restoration of both shapes, with the slices closely resembling the original clean images. Our experiments reveal that CryoGEN can effectively recover more of the missing wedge regions and capture high-frequency signals.

Next, to quantitatively assess the performance, we compute the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) between the ground truth and the generated results. The definitions of PSNR and SSIM are provided in Appendix A.3.1. We present the corrupted datasets, comparing the results corrected by IsoNet and DeepDeWedge to those corrected by CryoGEN. As demonstrated in Table 1, our method consistently outperforms the baseline.

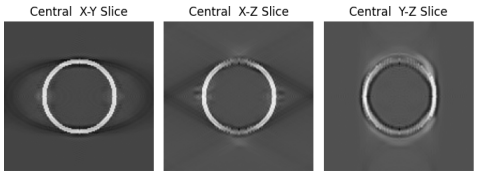
Table 1: *Quantitative evaluation of image quality for tomography reconstructions using different methods, comparing PSNR and SSIM metrics (**higher** values indicating better performance for both metrics) on sphere, prism and Vipp1 assembly datasets.*

Data State	Sphere		Prism		Vipp1 assembly	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Corrupted	21.12	0.8113	14.82	0.6931	26.68	0.8000
Iso-corrected	22.98	0.8770	19.11	0.8857	27.12	0.8191
Dewedge-corrected	23.17	0.8824	21.10	0.9278	28.75	0.8758
CryoGEN-corrected	**29.19**	**0.9706**	**32.69**	**0.9949**	**30.65**	**0.9199**

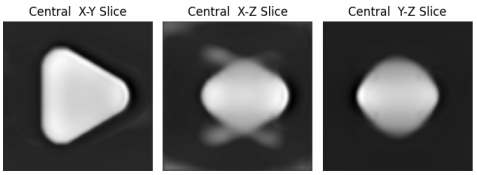
5.2 SIMULATED DATA

In this section, we applied our algorithm to more complex protein assemblies. Following the approach of IsoNet, we first evaluated our performance on the publicly available atomic model apoferritin (PDB:6Z6U) (Yip et al., 2020). Additionally, we selected the recently published electron microscopy dataset of C13 Vipp1 stacked rings (EMDB:18424) (Junglas et al., 2024). The results show that CryoGEN delivers more consistent outcomes in both the spatial and Fourier domains. Additionally, CryoGEN demands significantly less training time compared to IsoNet.

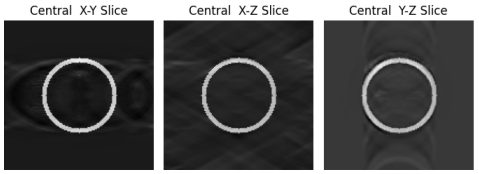
Apoferritin. We performed reconstructions using the atomic model of apoferritin (PDB:6Z6U), a widely-used benchmark in high-resolution CryoGEN. The simulated maps were then randomly rotated in ten different directions, and a missing wedge was applied in Fourier space, resulting in simulated subtomograms with missing wedge artefacts. In this experiment, CryoGEN delivered



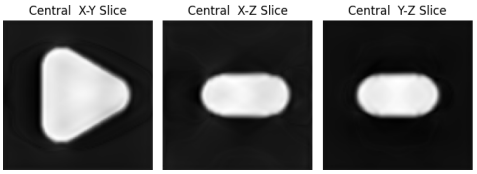
(a) *IsoNet-corrected sphere.*



(b) *IsoNet-corrected prism.*



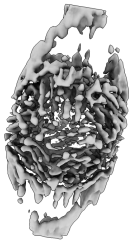
(c) *CryoGEN-corrected sphere.*



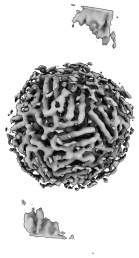
(d) *CryoGEN-corrected prism.*

Figure 6: *CryoGEN and IsoNet corrected shapes*.

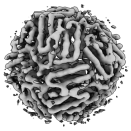
considerably better results than IsoNet, while also reducing the training time significantly. This improvement is clearly visible when visualizing the low-density volume using ChimeraX (Goddard et al., 2018), as shown in Figure 7.



(a) *WBP reconstructed missing-wedged apoferritin, displaying both corrupted and missing regions in the low-density volume.*



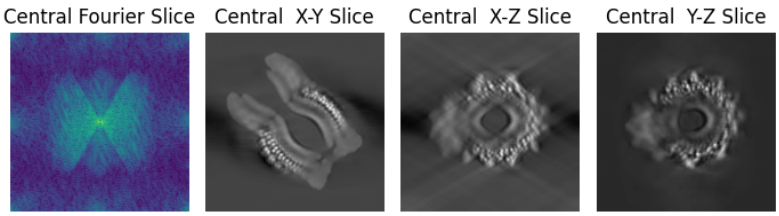
(b) *Structure generated by IsoNet, displaying a corrupted region with visible inconsistencies in the low-density volume.*



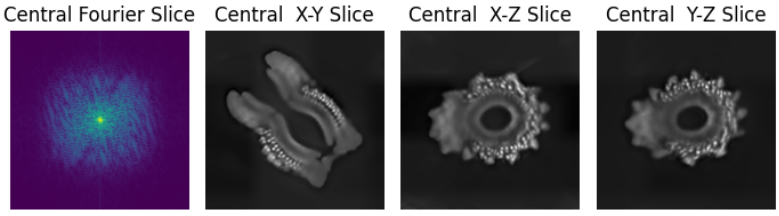
(c) *Structure generated by CryoGEN, showing a much smoother and coherent low-density volume representation.*

Figure 7: *Comparison of low-density volumes generated by WBP **(a)**, IsoNet **(b)** and CryoGEN **(c)**.*

C13 Vipp1 Stacked Rings. We evaluated our method on the recently published C13 Vipp1 stacked rings dataset (EMDB:18424), which represents complex assemblies. This dataset was sourced from the Electron Microscopy Data Bank (wwPDB Consortium).



(a) *IsoNet-corrected Vipp1 assembly.*



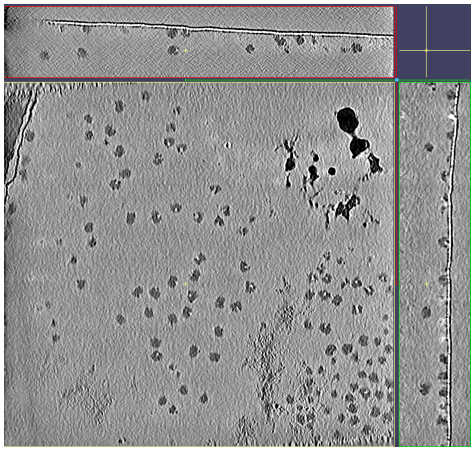
(b) *CryoGEN-corrected Vipp1 assembly.*

Figure 8: *CryoGEN and IsoNet corrected Vipp1 assembly.*

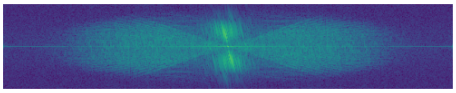
In line with IsoNet, we randomly rotate the tomography to create ten different samples before introducing missing wedge corruption. The results, as shown in Table 1, confirm that CryoGEN outperforms IsoNet, achieving superior PSNR and SSIM. We present the original results in Figure 1, along with both spatial and Fourier domain comparisons in Figure 8. In the Fourier domain, CryoGEN captures essential details more accurately and produces more consistent and symmetrical results. Additionally, the volume generated by IsoNet still contains corrupted regions, whereas CryoGEN produces a much smoother result in the spatial domain.

5.3 Real-World Examples

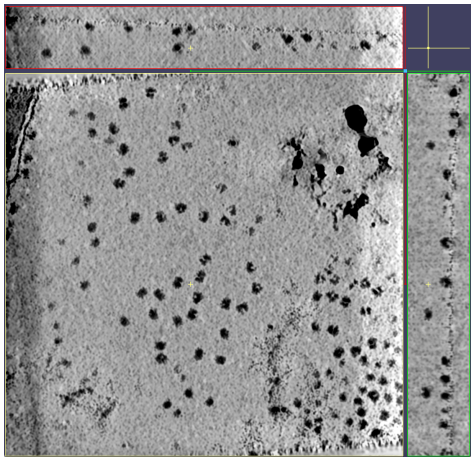
In this section, we apply the method to real-world examples to evaluate its effectiveness. We use a well-known Cryo-ET particle selection benchmark, specifically the dataset of purified ribosomes (Zhang et al., 2016), as well as the virus-like particle dataset of immature HIV-1 in both single-particle and tomography reconstruction (Schur et al., 2016). CryoGEN minimizes the ringing effect and achieves significantly higher contrast, while offering better compensation for the missing wedge compared to IsoNet's irregular distribution. *Moreover, CryoGEN completes the process in just two hours, compared to IsoNet's 20-hour runtime on an NVIDIA V100.*



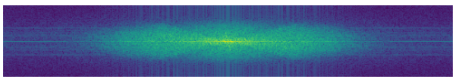
(a) *IsoNet-corrected purified ribosomes.*



(b) *Central Fourier slice of the IsoNet-corrected purified ribosomes.*



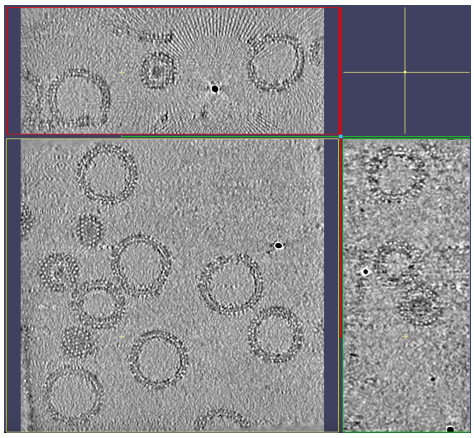
(c) *CryoGEN-corrected purified ribosomes.*



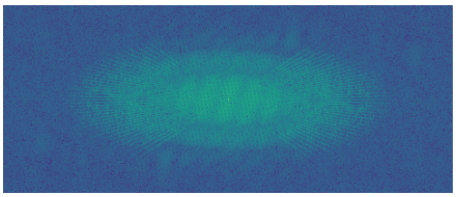
(d) *Central Fourier slice of the CryoGEN-corrected purified ribosomes.*

Figure 9: *Comparison of IsoNet-corrected and CryoGEN-corrected purified ribosomes, including their corresponding central fourier slices. The CryoGEN-corrected images exhibit higher contrast and reduced high-frequency features. While both methods effectively fill in the missing wedge, the IsoNet correction shows an irregular distribution of high-frequency components in the central region, whereas CryoGEN achieves a more consistent distribution.*

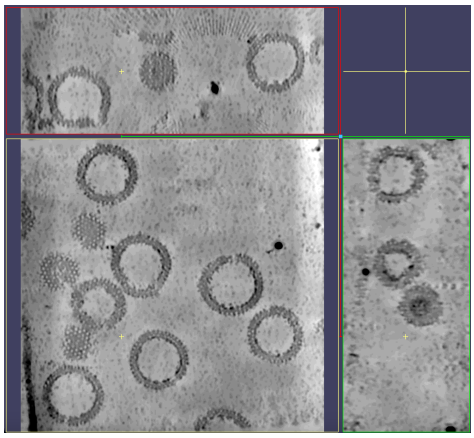
Purified Ribosomes. The ribosomes dataset is commonly used as a Cryo-ET benchmark. We collected all seven tilt series from the EMPIAR-10045 dataset and applied the same preprocessing steps as IsoNet, detailed in Appendix A.7.5. Figure 9 shows the correction results for both IsoNet and CryoGEN. Ribosomes in the CryoGEN-corrected volume appear clearer and exhibit higher contrast, which significantly aids in particle selection. Additionally, there is less noise and fewer sharp artifacts in the background. Notably, the IsoNet-corrected volume displays a frequency spectrum with an irregular concentration of high-frequency components in the central region, introducing noticeable noise and artifacts. In contrast, the CryoGEN-corrected volume shows a much smoother and more consistent frequency distribution, with better control over central frequencies. The more symmetrical pattern suggests reduced distortion and better alignment with the expected smooth behavior, indicating improved data integrity.



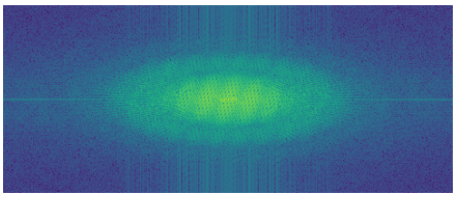
(a) *IsoNet-corrected immature HIV capsid.*



(b) *Central Fourier slice of the IsoNet-corrected immature HIV capsid.*



(c) *CryoGEN-corrected immature HIV capsid.*



(d) *Central Fourier slice of the CryoGEN-corrected immature HIV capsid.*

Figure 10: *Comparison of IsoNet-corrected and CryoGEN-corrected immature HIV capsids, along with their corresponding central fourier slices. The IsoNet-corrected images display noticeable ringing effects and bright white artefacts, while the CryoGEN images show minimal noise and a smoother background. Both methods fill in the missing wedge region, but the IsoNet correction results in more pronounced line artefacts.*

HIV Capsid. The results of HIV capsid dataset are presented in Figure 10. Following IsoNet's preprocessing procedure, we collected three tilt series from the EMPIAR-10164 dataset and processed the volume as detailed in Appendix A.7.6. The CryoGEN-corrected HIV capsid is noticeably clearer than the IsoNet-corrected version, with minimal noise and a much smoother background. In contrast, the IsoNet-corrected volume exhibits a pronounced ringing effect around the gold beads, with bright rings surrounding them and unwanted white dust scattered throughout the image. Our algorithm effectively eliminates all these artefacts. Additionally, CryoGEN compensates for the missing wedge region more effectively than IsoNet. As shown in the top windows of Figure 10 (a) and Figure 10 (c), the virus particle in the CryoGEN-corrected volume is more intact, with fewer defects compared to the IsoNet-corrected version. In the Fourier domain, while both methods attempt to fill the missing wedge region, IsoNet's correction introduces more noticeable line artefacts.

6 Conclusion

In this work, we introduce CryoGEN, a method for addressing the missing wedge problem in Cryo-ET using energy-based models. Our approach not only converges faster and more reliably than state-of-the-art techniques but also delivers significantly improved results. Moreover, even though developed for Cryo-ET, CryoGEN presents a more general framework for solving inverse problems by incorporating an energy model as a core component. To the best of our knowledge, this is the first energy-based framework proposed for tackling inverse problems in Cryo-ET reconstruction from a probabilistic perspective. Finally, our framework can integrate other advanced energy-based methods, such as Wasserstein GANs (Arjovsky et al., 2017) and energy-based diffusion models (Du et al., 2023), significantly broadening its potential applications. In the future, we plan to extend our method to efficiently handle larger and more complex datasets.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pp. 1–3, 2024.

Martin Arjovsky and Leon Bottou. Towards principled methods for training generative adversarial networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=Hk4_qw5xe.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 214–223. JMLR. org, 2017.

Tristan Bepler, Kotaro Kelley, Alex J Noble, and Bonnie Berger. Topaz-denoise: general deep denoising models for cryoem and cryoet. *Nature communications*, 11(1):5208, 2020.

Ludwig Boltzmann. *Theoretical Physics and Philosophical Problems: Selected Writings*, volume 5 of *Vienna Circle Collection*. Springer, 1974. ISBN 978-90-277-0287-8.

Tim-Oliver Buchholz, Mareike Jordan, Gaia Pigino, and Florian Jug. Cryo-care: Content-aware image restoration for cryo-transmission electron microscopy data. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 502–506. IEEE, 2019.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.

Yuchen Deng, Yu Chen, Yan Zhang, Shengliu Wang, Fa Zhang, and Fei Sun. Icon: 3d reconstruction with 'missing-information' restoration in biological electron tomography. *Journal of Structural Biology*, 2016. ISSN 1047-8477. doi: https://doi.org/10.1016/j.jsb.2016.04.004. URL https://www.sciencedirect.com/science/article/pii/S1047847716300636.

P. J. Diggle and R. J. Gratton. Monte carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(2):193–227, 1984.

Yilun Du, Conor Durkan, Robin Strudel, Joshua B. Tenenbaum, Sander Dieleman, Rob Fergus, Jascha Sohl-Dickstein, Arnaud Doucet, and Will Sussman Grathwohl. Reduce, reuse, recycle: Compositional generation with energy-based diffusion models and MCMC. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 8489–8510. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/du23a.html.

Thomas D Goddard, Conrad C Huang, Elaine C Meng, Eric F Pettersen, Gregory S Couch, John H Morris, and Thomas E Ferrin. Ucsf chimerax: Meeting modern challenges in visualization and analysis. *Protein science*, 27(1):14–25, 2018.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. volume 9906, pp. 694–711, 10 2016. ISBN 978-3-319-46474-9. doi: 10.1007/978-3-319-46475-6_43.

Benedikt Junglas, David Kartte, Mirka Kutzner, Nadja Hellmann, Ilona Ritter, Dirk Schneider, and Carsten Sachse. Structural basis for vipp1 membrane binding: From loose coats and carpets to ring and rod assemblies. *bioRxiv*, 2024. doi: 10.1101/2024.07.08.602470. URL https://www.biorxiv.org/content/early/2024/07/08/2024.07.08.602470.

Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, San Diega, CA, USA, 2015.

Yann Lecun, Sumit Chopra, and Raia Hadsell. *A tutorial on energy-based learning*. 01 2006.

Yun-Tao Liu, Heng Zhang, Hui Wang, Chang-Lu Tao, Guoqiang Bi, and Hong Zhou. Isotropic reconstruction for electron tomography with deep learning. *Nature Communications*, 13, 10 2022. doi: 10.1038/s41467-022-33957-8.

Vladan Lucic, Friedrich Förster, and Wolfgang Baumeister. Cryo-electron tomography of cells: connecting structure and function. *Nature Reviews Molecular Cell Biology*, 6(6):432–439, 2005. doi: 10.1038/nrm1613.

Michael Radermacher. Weighted back-projection methods. *Electron tomography: methods for three-dimensional visualization of structures in the cell*, pp. 245–273, 2006.

Florian KM Schur, Martin Obr, Wim JH Hagen, William Wan, Arjen J Jakobi, Joanna M Kirkpatrick, Carsten Sachse, Hans-Georg Kräusslich, and John AG Briggs. An atomic model of hiv-1 capsid-sp1 reveals structures regulating assembly and maturation. *Science*, 353(6298):506–508, 2016.

Dimitry Tegunov and Patrick Cramer. Real-time cryo-electron microscopy data preprocessing with warp. *Nature methods*, 16(11):1146–1152, 2019.

Simon Wiedemann and Reinhard Heckel. A deep learning method for simultaneous denoising and missing wedge reconstruction in cryogenic electron tomography. *Nature Communications*, 15(1): 8255, 2024.

The wwPDB Consortium. Emdb—the electron microscopy data bank.

Rui Yan, Singanallur V. Venkatakrishnan, Jun Liu, Charles A. Bouman, and Wen Jiang. Mbir: A cryo-et 3d reconstruction method that effectively minimizes missing wedge artifacts and restores missing information. *Journal of Structural Biology*, 2019. ISSN 1047-8477. doi: https://doi.org/10.1016/j.jsb.2019.03.002. URL https://www.sciencedirect.com/science/article/pii/S1047847719300474.

Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. Atomic-resolution protein structure determination by cryo-em. *Nature*, 587(7832):157–161, 2020.

Xiangrui Zeng, Yizhe Ding, Yueqian Zhang, Mostofa Rafid Uddin, Ali Dabouei, and Min Xu. Dual: deep unsupervised simultaneous simulation and denoising for cryo-electron tomography. *bioRxiv*, 2024.

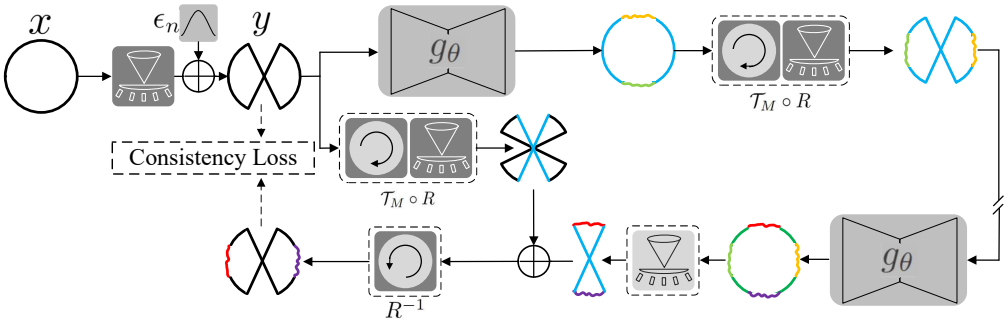
Chenwei Zhang, Anne Condon, and Khanh Dao Duc. Synthetic high-resolution cryo-em density maps with generative adversarial networks, 2024. URL https://arxiv.org/abs/2407.17674.

Xing Zhang, Mason Lai, Winston Chang, Iris Yu, Ke Ding, Jan Mrazek, Hwee L Ng, Otto O Yang, Dmitri A Maslov, and Z Hong Zhou. Structures and stabilization of kinetoplastid-specific split rrnas revealed by comparing leishmanial and human ribosomes. *Nature communications*, 7(1): 13223, 2016.

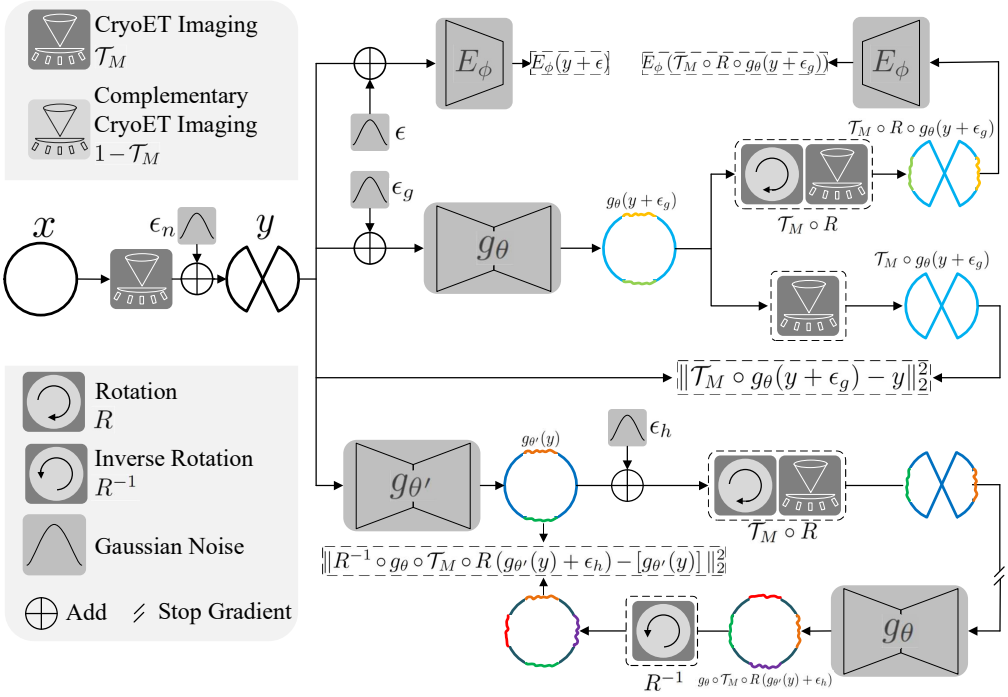
Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11. Springer, 2018.

A APPENDIX

A.1 ALGORITHM FLOWCHART



(a) *IsoNet*



(b) *CryoGEN*

Figure 11: *Algorithm flowchart of (a) IsoNet and (b) CryoGEN*

IsoNet formulates the task of filling y's missing wedge information to reconstruct x as an inpainting problem, solving it in a self-supervised manner as illustrated in Figure 11. It trains a U-Net, denoted as g_θ, where θ represents trainable parameters, following these steps:

1. y is processed by g_θ to obtain a missing-wedge-filled $\tilde{x} = g_\theta(y)$.
2. $\tilde{x}$ is rotated by a rotation operator R, randomly selected from a pre-defined rotation set $\mathcal{R}$, and then subjected to a missing wedge by a simulated $\mathcal{T}_M$ operation: $\tilde{y} = \mathcal{T}_M \circ R(\tilde{x})$.
3. $\tilde{y}$ is fed to g_θ, yielding $\hat{x} = g_\theta(\tilde{y})$.
4. The inpainted part is extracted by $(1 - \mathcal{T}_M)(\hat{x})$ and added with $\mathcal{T}_M \circ R(y)$, then rotated back to get $\hat{y} = R^{-1}((1 - \mathcal{T}_M)(\hat{x}) + \mathcal{T}_M \circ R(y))$
5. $\hat{y}$ and y form paired data to train g_θ, with y serving as ground truth. During training, no gradient is generated from $g_\theta(y)$.
6. These steps are iterated until convergence.

By constructing $(\hat{y}, y)$ pairs, IsoNet effectively makes a one-to-one mapping assumption of $\mathcal{T}_M$.

A.2 ROTATION LIST DEFINED IN THE ISONET

The cropped subtomograms are cube-shaped with six faces, resulting in 24 possible rotations for reorientation. However, we exclude the four rotations that maintain the same missing wedge in the X-Z direction as the original, unrotated subtomogram. Further details and a schematic diagram are provided in the supplementary information of IsoNet (Liu et al., 2022).

A.3 ADDITIONAL RESULTS

A.3.1 ADDITIONAL RESULTS FOR SECTION 5.1

Formally, $\hat{v}$ represents the predicted volume, and v^* denotes the ground truth. PSNR and SSIM are defined as follows:

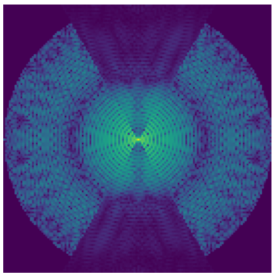
$$\text{PSNR} = 10 \log_{10} \frac{I_m}{\|\hat{v} - v^*\|^2}, \quad \text{SSIM}(\hat{v}, v^*) = \frac{(2\mu_{\hat{v}}\mu_{v^*} + C_1)(2\sigma_{\hat{v}v^*} + C_2)}{(\mu_{\hat{v}}^2 + \mu_{v^*}^2 + C_1)(\sigma_{\hat{v}}^2 + \sigma_{v^*}^2 + C_2)}.$$

In the PSNR formula, I_m represents the maximum possible pixel value, which we define as the maximum value of the ground truth image.

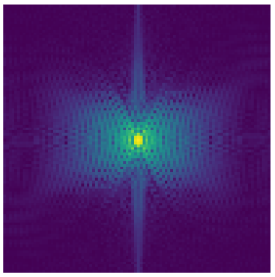
In the SSIM formula:

- $\mu_{\hat{v}}$ and μ_{v^*} are the mean intensities of images $\hat{v}$ and v^*.
- $\sigma_{\hat{v}}^2$ and $\sigma_{v^*}^2$ are the variances of images $\hat{v}$ and v^*.
- $\sigma_{\hat{v}v^*}$ is the covariance between images $\hat{v}$ and v^*.
- C_1 and C_2 are small constants used to stabilize the division when the denominator is close to zero.

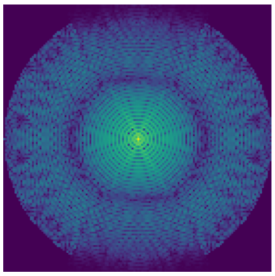
The central Fourier slices of the corrected sphere and prism are displayed in Figure 12, while the original clean sphere and prism are shown in Figure 13.



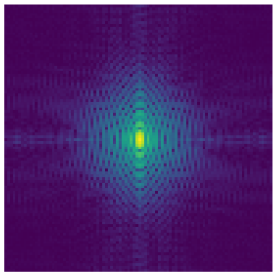
(a) Central fourier slice of the IsoNet-corrected sphere.



(b) Central fourier slice of the IsoNet-corrected prism.

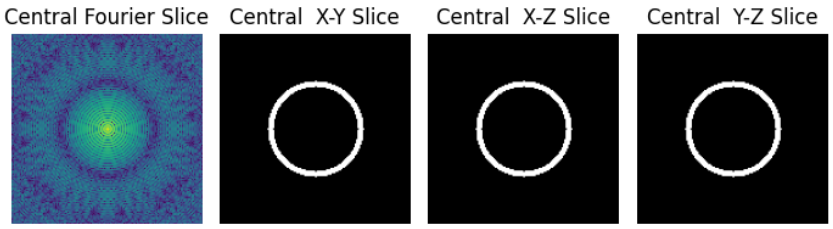


(c) Central fourier slice of the CryoGEN-corrected sphere.

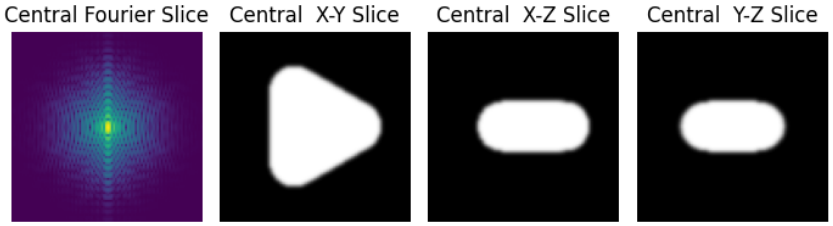


(d) Central fourier slice of the CryoGEN-corrected prism.

Figure 12: Central fourier slices of sphere and prism. The IsoNet's corrected has less information both at low and high-frequency signals with missing regions, while CryoGEN fills in most of the missing wedge. It is consistent with the spatial domain results.



(a) Orignal clean prism.

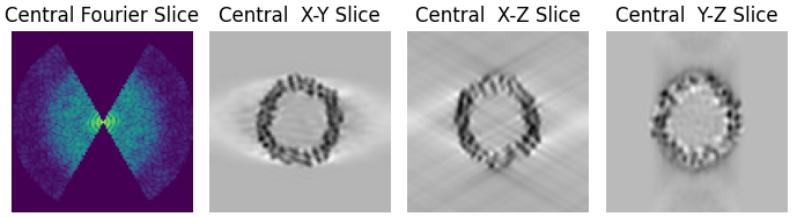


(b) Original clean prism.

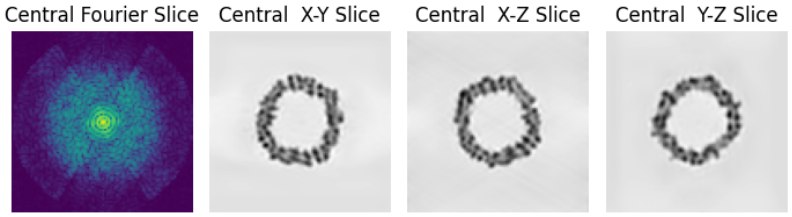
Figure 13: Original clean shapes.

Compared to the corrected tomograms, the CryoGEN-corrected versions more closely resemble the original clean shapes.

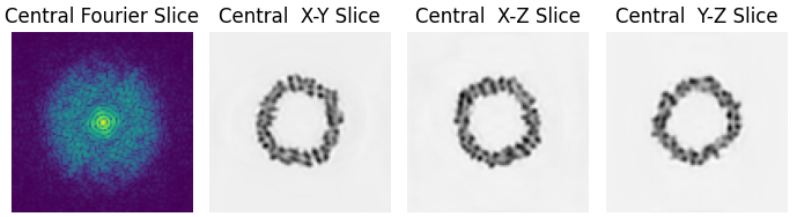
A.3.2 Additional Results for Section 5.2



(a) Original missing wedged apoferritin.



(b) IsoNet-corrected apoferritin.



(c) CryoGEN-corrected apoferritin.

Figure 14: Missing wedged, CryoGEN and Iso-Net corrected apoferritin.

We present the central X-Y, X-Z, Y-Z slices, as well as the central Fourier slices for the corrupted, IsoNet-corrected, and CryoGEN-corrected volumes in Figure 14. In the IsoNet-corrected X-Y slice, there are faint white artifacts in the background, consistent with Figure 7, and line artifacts in the X-Z slice, which are absent in the CryoGEN-corrected results. Additionally, the central Fourier slice of the IsoNet-corrected volume displays a distinct borderline, which is not present in the CryoGEN-corrected slice.

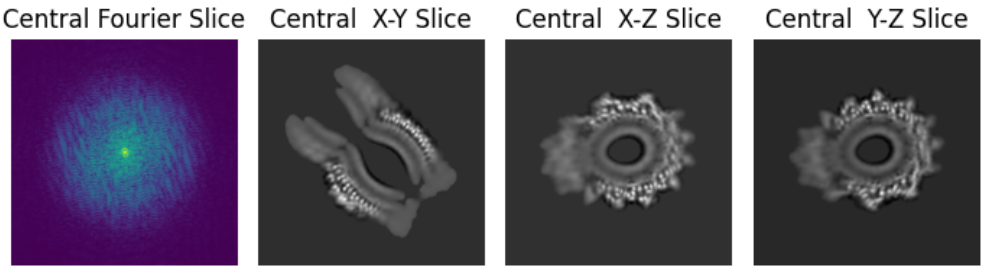
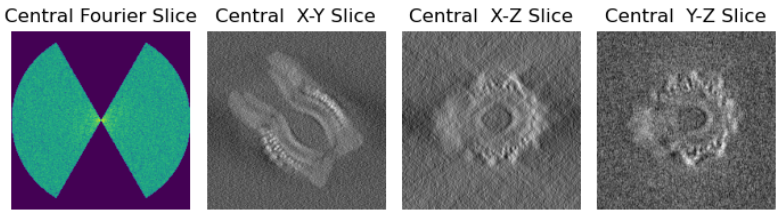


Figure 15: Original clean Vipp1 assembly.

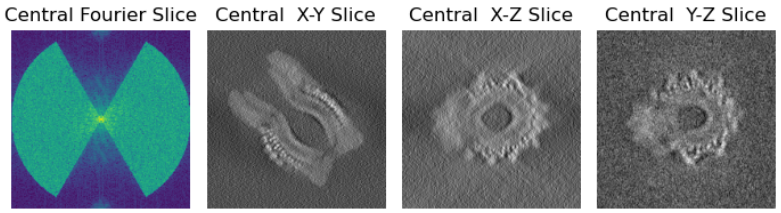
The original clean Vipp1 assembly are shown in Figure 15.

A.3.3 Reconstruction from Noisy Samples

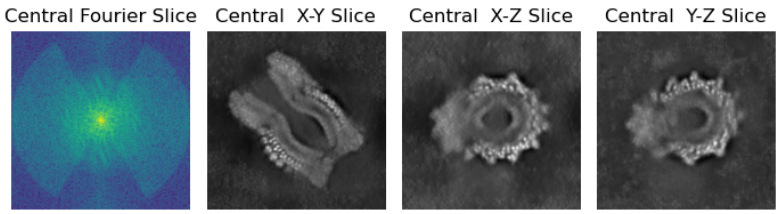
We also demonstrated the robustness of our algorithm under higher noise levels, such as SNR $= 0.2$. The results are presented in Figure 16. Even in this challenging scenario, CryoGEN outperforms IsoNet, demonstrating its strong denoising capabilities.



(a) *Original noisy Vipp1 assembly.*



(b) *IsoNet-corrected Vipp1 assembly.*

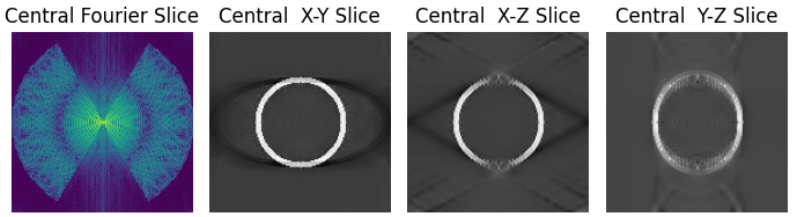


(c) *CryoGEN-corrected Vipp1 assembly.*

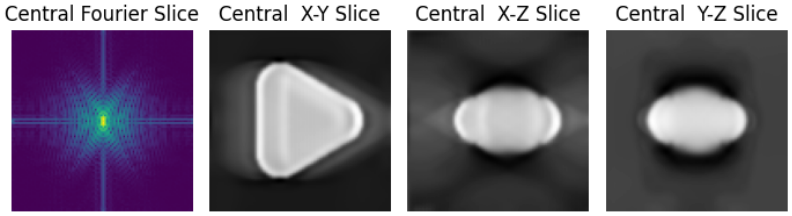
Figure 16: *CryoGEN and IsoNet corrected noisy Vipp1 assembly SNR* $= 0.2$.

A.4 ADDTIONAL RESULTS BY DEEPDEWEDGE

First, we present the DeepDeWedge-corrected simple shapes in Figure 17, where defects similar to those in the IsoNet-corrected versions are apparent. Both methods exhibit spatial domain artifacts, resulting in distorted volumes, and neither DeepDeWedge nor IsoNet effectively fills in the missing information in Fourier space.



(a) DeepDeWedge-corrected sphere.

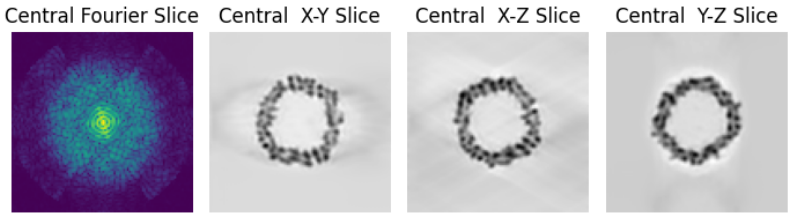


(b) DeepDeWedge-corrected prism.

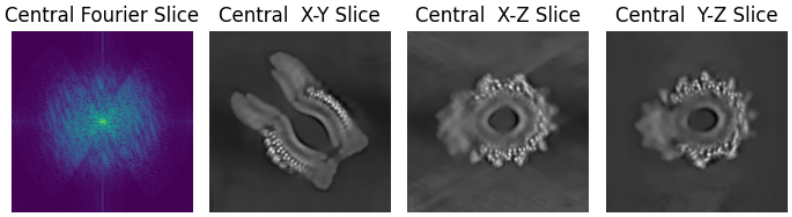
Figure 17: DeepDeWedge-corrected shapes.

Next, the DeepDeWedge-corrected simulated data is displayed in Figure 18. Similar to IsoNet, DeepDeWedge encounters the same issues, generating faint shadows and distinct line artifacts in the background, as well as a noticeable borderline in the central Fourier slice.

Finally, we test the DeepDeWedge on the real-world examples as shown in Figure 19. The DeepDeWedge-corrected ribosomes exhibit the same irregular distribution of high-frequency components in the central region and yield unsatisfactory results for the HIV capsid, with noticeable artifacts. A potential reason for the poor performance on the HIV capsid may be the loss of information caused by even-odd splits.

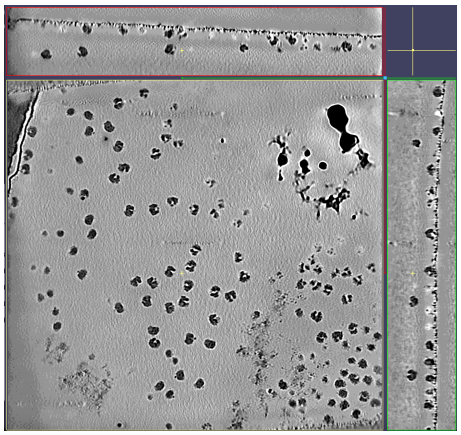


(a) DeepDeWedge-corrected apoferritin.

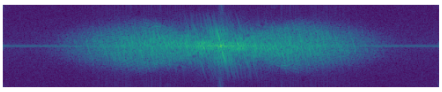


(b) DeepDeWedge-corrected Vipp1 assembly.

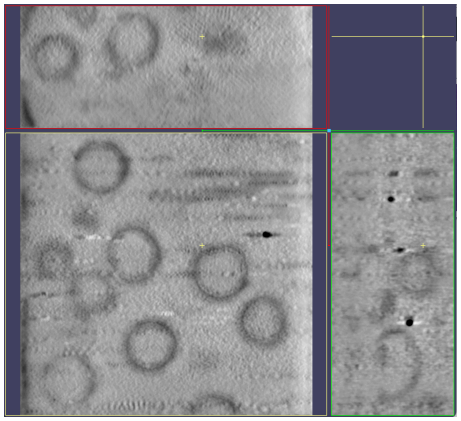
Figure 18: DeepDeWedge-corrected simulated data.



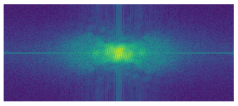
(a) DeepDeWedge-corrected purified ribosomes.



(b) Central fourier slice of the DeepDeWedge-corrected purified ribosomes.



(c) DeepDeWedge-corrected HIV capsid.



(d) Central fourier slice of the DeepDeWedge-corrected HIV capsid.

Figure 19: DeepDeWedge-corrected ribosomes and HIV capsid, along with their corresponding central Fourier slices.

A.5 VISUALIZATION OF A SINGLE RIBOSOME

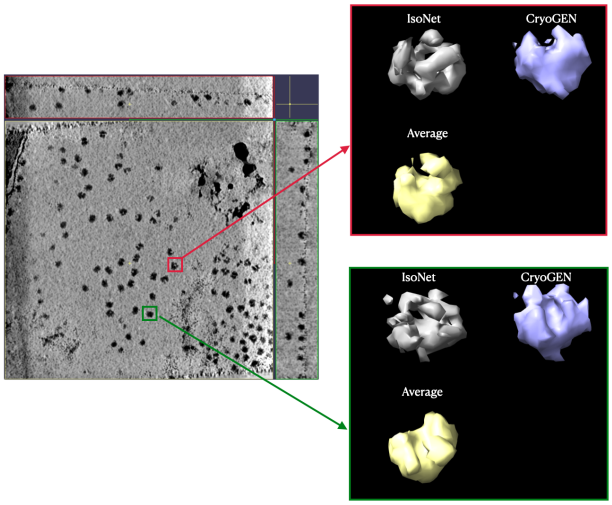


Figure 20: *Single Ribosome Visualization: We compared CryoGen with IsoNet and the averaged results of all ribosomes across two examples at different rotation angles.*

In Figure 20, we present the visualization of a single ribosome. While the IsoNet results may appear to show more details in 2D grayscale images, as illustrated in Figure 9, this could be due to noise-to-signal ratio issues or the enhanced denoising capabilities of CryoGEN (see Appendix A.3.3). Moreover, the comparison highlights that IsoNet results still suffer from residual effects of the missing wedge problem, which CryoGEN effectively addresses.

A.6 IMPLEMENTATION DETAILS

Following the struct2map GAN (Zhang et al., 2024), the architecture consists of a generator and discriminator with specific design choices. The generator is a modified U-Net architecture, known as U-Net++, which enhances the standard U-Net with dense skip connections for improved performance (Zhou et al., 2018). The discriminator is composed of four 3D convolutional layers, each using a $3 \times 3 \times 3$ kernel. Following the final convolutional layer, an adaptive average pooling layer reduces the dimensions of the feature map to $1 \times 1 \times 1$. This output is then flattened and fed into a series of three fully connected layers, with ReLU activations between each layer. The final layer produces a single output, which serves as the result of binary classification.

The CryoGEN are trained with the Adam optimizer (Kingma & Ba, 2015) with batch size one for simulated shapes and protein subtomograms and with batch size 4 for real-world examples. The learning rate is set to 10^{-4} with a linear warm-up phase in the initial one-tenth steps, which is

followed by a linear decay schedule thereafter. Different from IsoNet, which progressively increases the noise scale, we apply random noise levels across all training steps. Specifically, a random number is sampled from a uniform distribution within the range (0,1] and multiplied by the set noise scale for each step. Additionally, the penalty term λ is kept constant during the first epoch and then decays linearly throughout the subsequent epochs.

A.7 Experiment Details

A.7.1 Sphere

A hollow sphere with an outer diameter of 70 pixels and a thickness of 4 pixels is positioned at the center of a 140×140×140 volume. Corruption is applied by setting values to zero within the missing wedge angles in Fourier space. For training, the volume is split into ten 96×96×96 pixel subtomograms with randomly chosen origins. These subtomograms are then randomly cropped to 64×64×64 pixels before being input into the models.

A.7.2 Prism

A prism with a thickness of 20 pixels is placed inside a 96×96×96 volume. It is randomly rotated in ten directions. Corrupted prisms are generated by setting zero values within the missing wedge angles in Fourier space. The entire volume is directly fed into the model during training.

A.7.3 Simulated Apoferritin

Ten randomly rotated apoferritin datasets are downloaded from a link provided by IsoNet and generated using ChimeraX's molmap function. During training, the datasets are directly fed into the model without further modifications.

A.7.4 Simulated Stacked Rings

C13 Vipp1 stacked ring data are downloaded from the EMDB database and binned twice, resulting in 200×200×200 pixels. The data is randomly rotated in ten different directions. Corrupted stacked rings are generated by setting zero values within the missing wedge angles in Fourier space. For training, the data is split into ten 96×96×96 pixel subtomograms with random zero origins, then randomly cropped to 64×64×64 pixels before being fed into the models.

A.7.5 Ribosomes

Ribosome data is downloaded from the EMPIAR database and binned six times, yielding a pixel size of 13.02 Å. IsoNet's deconvolution is applied, following the same procedure as described by (Liu et al., 2022). To ensure that subtomograms contain sufficient data, IsoNet's mask generation tool is used to extract subtomograms with at least 40% non-zero pixels based on the density mask. For training, a tomogram is split into seventy 80×80×80 pixel subtomograms, resulting in a total of 490 subtomograms. These are randomly cropped to 64×64×64 pixels before being fed into the models.

A.7.6 HIV capsid

Raw tilt series for the HIV capsid is downloaded from the EMPIAR database. The movie stacks are drift-corrected and reconstructed using the WBP algorithm, aided by the latest tomogram processing tools such as Aretomo2. The processed tomograms, TS-01, TS-43, and TS-45, are then subjected to IsoNet's deconvolution, following the procedure outlined by (Liu et al., 2022). IsoNet's mask generation tool is applied to ensure that each subtomogram contains at least 50% non-zero pixels. During training, each tomogram is split into one hundred 96×96×96 pixel subtomograms, resulting in 300 subtomograms. These are randomly cropped to 64×64×64 pixels before being fed into the models.