A LATENT BACK-PROJECTION NETWORK FOR NOVEL PROJECTION SYNTHESIS FOR IMPROVED CRYO-ET

Robert Kiewisz *
Simons Machine Learning Center
New York Structural Biology Center
New York, NY, USA
rkiewisz@nysbc.org

Gabriel Meyer-Lee *
Department of Computer Science
University of Vermont
Burlington, VT, USA
gmeyerle@uvm.edu

Tristan Bepler
Simons Machine Learning Center
New York Structural Biology Center
New York, NY, USA
tbepler@nysbc.org

ABSTRACT

Cryo-electron tomography (Cryo-ET) is hindered by the missing wedge, a gap in Fourier-space information caused by limited tilt-series angular coverage, leading to anisotropic resolution loss and artifacts. Current methods, such as IsoNet, attempt to "inpaint" missing frequencies in reconstructed tomograms but are constrained by their reliance on pre-degraded data, often producing non-physical features. We present our proof-of-concept model, a generative latent back-projection autoencoder that bypasses traditional tomogram reconstruction and directly synthesizes novel projections from tilt-series data in the frequency domain. Our latent back-projection network encoder-decoder architecture maps raw projections to a 3D Fourier volume, leveraging the Fourier slice theorem to generate high-fidelity projections beyond the experimental tilt range. Evaluated on *E. coli* mini-cells Ishemgulova et al. (2023), our model achieves a lower MSE and higher correlation with ground-truth data. Crucially, our model robustly recovers withheld tilts (0° or ±15°) without retraining, outperforming IsoNet in MSE and correlation metrics. By mitigating the missing wedge through generating tilts at new angles, our proof-of-concept can potentially advances high-resolution in situ structural biology for radiation-sensitive specimens.

1 INTRODUCTION

Cryo-electron tomography (Cryo-ET) has emerged as a cornerstone technique for visualizing macromolecular complexes in their native cellular environments, offering unprecedented insights into mechanisms of viral infection, protein aggregation, and organelle dynamics Dutta & Priyamvada (2024); Nogales & Sjors (2015). The technique involves mechanically rotating the stage of the electron microscope to capture micrographs at differing angles (termed a tilt series), allowing 3D reconstruction of the imaged volume (**Fig. 1A**). During tilt series acquisition, the sample is incrementally tilted around a single axis, typically in 1–3° increments, and imaged at each angle Young & Villa (2023). These 2D projections (or tilts) are then computationally combined via tomographic reconstruction algorithms, such as weighted back-projection (WBP), to reconstruct a 3D volume Radermacher (2007); R et al. (2019).

However, the resolution of the technique is fundamentally constrained by the missing wedge, a gap in Fourier-space information caused by limited angular coverage in the tilt series. Mechanical stage restrictions, beam-induced sample warping, and cumulative radiation damage during imaging typically restrict tilt ranges to ±45°, resulting in anisotropic resolution loss and elongation artifacts

*equal contribution

that obscure critical structural features. Traditional reconstruction algorithms, such as WBP or simultaneous iterative reconstruction (SIRT), amplify these artifacts by back-projecting incomplete frequency data. While deep learning tools like IsoNet attempt post-reconstruction "inpainting" of missing frequencies, they operate on pre-reconstructed tomograms where the missing wedge has already irreversibly distorted the 3D frequency spectrum Liu et al. (2022). Consequently, IsoNet's adversarial training framework can only hallucinate plausible textures rather than recover genuine high-resolution information, often producing over-smoothed or non-physical results (**Fig. 1B-C**).

To address these limitations, we introduce our proof-of-concept model, a generative autoencoder that directly synthesizes novel tilt projections to fill the missing wedge by learning angle-related spatial relationship between input tilts series. Unlike existing generative approaches for Cryo-ET, such as GANrec Yang et al. (2023), which uses adversarial training to inpaint missing sinogram patches, our model operates as a latent reconstructor, learning a continuous mapping between 2D tilt-series projections and a generative 3D Fourier latent space. This latent space encodes the frequency spectrum of the imaged volume, enabling the synthesis of physically consistent projections at arbitrary angles. Our latent back-projection model comprises three key components:

1. ResNet-style Encoder-Decoder: A symmetric architecture processes input projections through a ResNet encoder, compressing spatial features by 16× while preserving high-frequency details via residual connections. This ensures that critical structural information (e.g., membrane pores) is retained for generative synthesis.

2. Differentiable Latent Back-Projection (DLBP): The encoder's latent representations are fused into a unified 3D Fourier volume via a differentiable back-projection layer. This volume is refined by 3D convolutions that resolve missing wedge regions while enforcing consistency with the Fourier slice theorem—a generative prior that prevents non-physical artifacts.

3. Angle-Conditioned Projection Generation: The decoder synthesizes new projections by slicing the 3D Fourier latent at user-specified angles, then upsampling to native resolution while minimizing a MSE loss. This ensures that generated projections are both novel (extending beyond observed angles) and consistent with ground-truth tilts.

By training directly on tilt-series data rather than reconstructed tomograms, our model avoids irreversible information loss, recovering missing frequencies through structured generation in Fourier space. Critically, our model generalizes to unseen specimens and angles without retraining, achieving lower MSE and higher correlation than IsoNet. This work establishes our proof-of-concept as a sample-agnostic generative framework for high-resolution Cryo-ET, bridging the gap between analytical reconstruction and deep generative modeling. Overall, our model achieved lower MSE and higher correlation than IsoNet, which relies on laborious per-sample adversarial optimization.

2 METHOD

Our proof-of-concept model is an autoencoder with differentiable latent back-projection (DLBP) transformation operation designed to synthesize new 2D projections from limited-angle tilt-series projections to mitigate missing-wedge artifacts. The encoder employs a ResNet-style architecture with strided convolutions, down-sampling input projections (e.g., 1440×1024 pixels) by 16× to generate a latent representation of shape N×C×90×64, where N is the number of projections and C denotes feature channels encoding spatial information. This latent space is then processed by a DLBP module (**Fig. 2A**), which aggregates 2D latent representations into a 3D latent representing the whole imaged volume (**Appendix A.1**). The DLBP refines this volume using 3D convolutions to model inter-slice dependencies, effectively addressing artifacts from WBP and recovering missing frequencies. Novel projections are synthesized by rotating the 3D latent to target angles and aggregating the latent content over the depth axis, simulating the physical projection operation that produced the input tilt series. A symmetric ResNet-style decoder up-samples these slices to the original resolution, preserving high-frequency details through residual connections and transposed convolutions.

While IsoNet is a reconstruction post-processing approach, our method is a pre-processing approach. The output synthesized projections can be incorporated into the input of any tomographic reconstruction algorithm as additional tilts. Although our method does incorporate latent-space reconstruction,

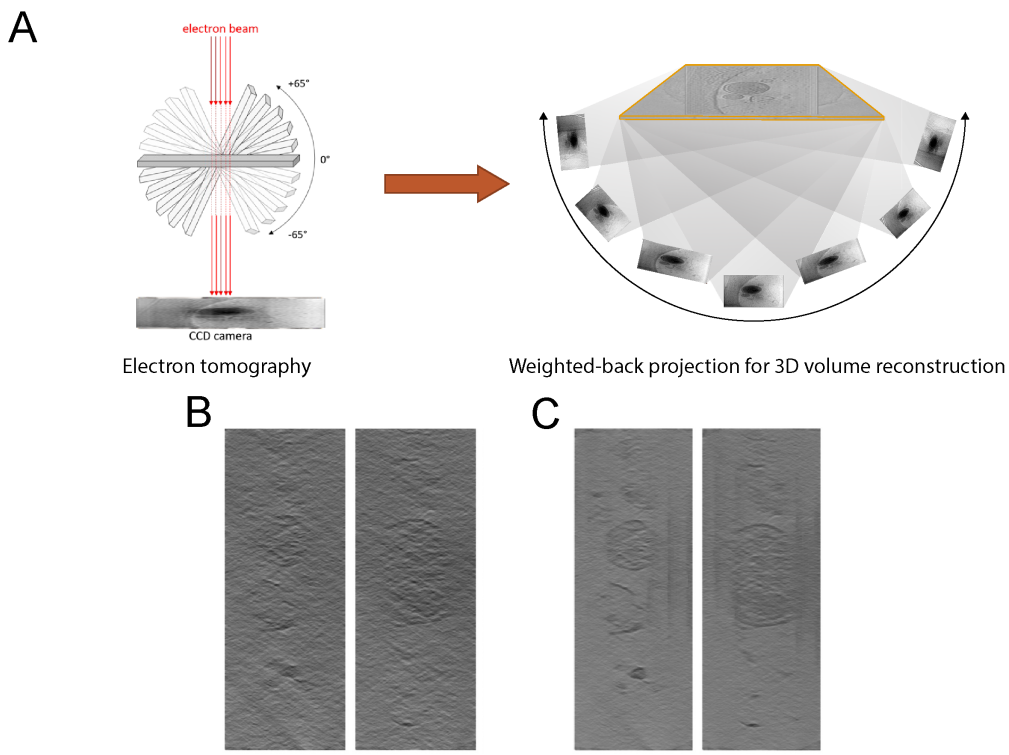

Figure 1: Illustration of technique used for a 3D volume reconstruction from Cryo-EM tilts angle. **A)** Cartoon showing process of image acquisition using Cryo-EM microscope. The sample is rotated inside the microscope to produce projections throw the volume at specific angles. This projections are than combine using algorithms like weighted back-projection to reconstruct a 3D volume of imaged sample. **B)** Shows XZ slice of original tomogram volume. **C)** XZ slice through tomogram volume corrected with IsoNet.

this is only used to ground the relationship between the input and generated tilts, not to directly produce a full reconstruction of the imaged volume.

For training, we optimized our model using the Adam optimizer with a learning rate of 0.00001, mean squared error (MSE) loss, and 1,000 epochs. To evaluate generalizability of our approach, we trained separate models on a single tomogram and five tomograms, mirroring IsoNet's limited tomogram training paradigm. We evaluated performance with mean MSE (MMSE) and correlation (MCorr) error over all projections. Additionally, we also withhold specific tilts (0° or ±15°) during training and quantified projection recovery accuracy. All experiments used a single NVIDIA A100 GPU.

3 Results

3.1 Generating original tilt range

We trained our model and IsoNet on both a single tomogram and five tomograms containing *E. coli* mini-cells collected via a TFS Titan Krios microscope Ishemgulova et al. (2023), comparing their ability to reconstruct projections within the original tilt range (±45°). We also generated new projections at ±75° to visually assess angular generalization (b and **Fig. 3**). When trained on a single tomogram, our model achieved a mean MSE of 0.69 and correlation of 0.65 for projections at ±45°, outperforming IsoNet (MSE: 1.22, correlation: 0.39; **Table 1**). Training on five tomograms further improved our model performance, reducing MSE to 0.66 and maintaining a correlation of 0.65, while IsoNet showed minimal improvement (MSE: 1.17, correlation: 0.41). These metrics are computed on projections (for which ground truth is available), not volumes. While the data is not

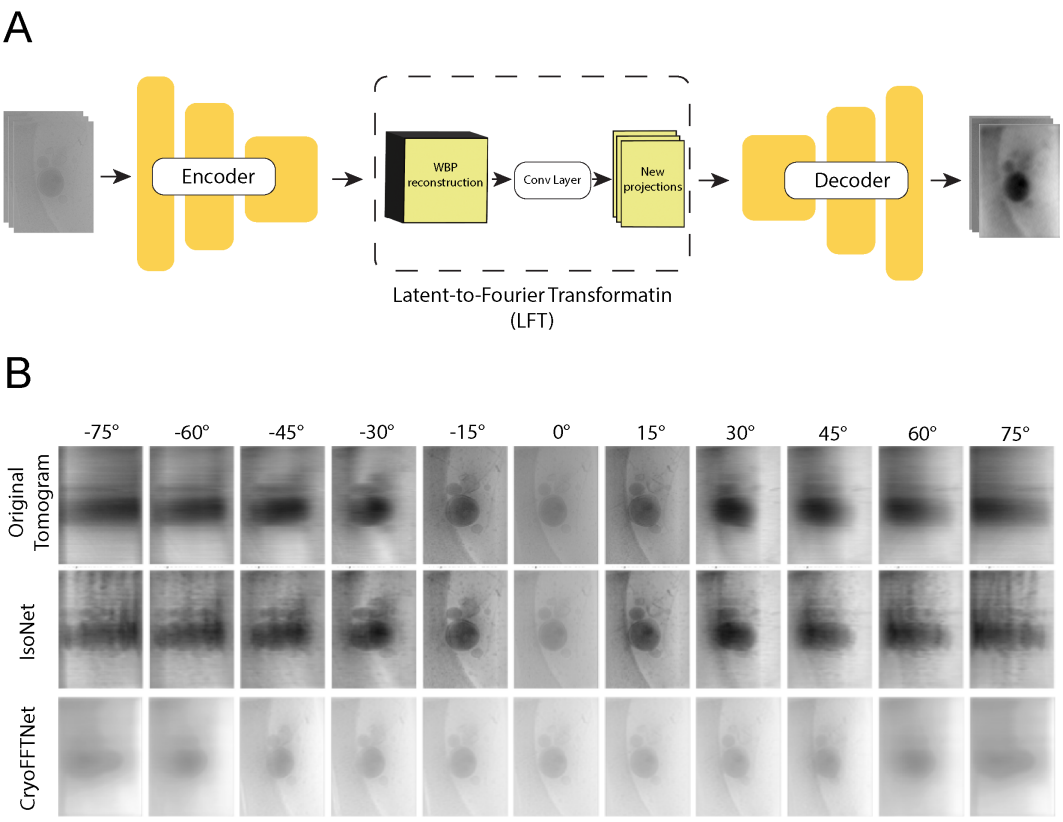

Figure 2: **A)** Our proof-of-concept model architecture. **B)** Comparison of projections digitally computed from original tomogram, IsoNet corrected tomogram, and our model.

simulated, we do simulate the projection operation on the corrected volume output by IsoNet. These results suggest our model benefits from multi-tomogram training, though future work could enhance performance with structural-aware loss functions.

Table 1: Comparison of MSE error for predicted Cryo-EM projections

Method	**Train on 1 tomogram**		**Train on 5 tomogram**	
	MMSE	**MCorr**	**MMSE**	**MCorr**
Our	0.69	0.65	0.66	0.65
IsoNet	1.22	0.39	1.17	0.41
Original tomogram	1.50	0.25	-	-

3.2 Generating missing tilts

Next, to evaluate robustness to incomplete data, we reconstructed tomograms with withheld tilts (0° or ±15°) and retrained both models. For missing 0° tilts, our model retained stable performance (MSE: 0.64, correlation: 0.76; **Table 2**), while IsoNet degraded significantly (MSE: 0.92, correlation: 0.54). For withheld ±15° tilts, our model maintained accuracy (MSE: 0.64, correlation: 0.66), whereas IsoNet failed catastrophically (MSE: 2.09, correlation: 0.05; **Table 3**), producing incoherent outputs (**Fig. 3D**). Notably, our model's MSE for missing 0° tilts was slightly higher than IsoNet's (0.76 vs. 0.67), highlighting room for refinement in angle extrapolation.

Table 2: Comparison of MSE and Correlation error for predicted Cryo-EM projections from tomogram with withhold tilt at 0 degree angle.

Method	MMSE	MSE at 0 angle°	MCorr	Corr at 0°
Our	0.64	0.76	0.76	0.59
IsoNet	0.92	0.67	0.54	0.66
Original tomogram	1.13	0.88	0.42	0.54

Table 3: Comparison of MSE and Correlation error for predicted Cryo-EM projections from tomogram with withhold tilt at -15 and 15 degree angle.

Method	MMSE	MSE at -15°	MSE at 15°	MCorr	Corr at -15°	Corr at 15°
Our	0.64	0.71	0.68	0.66	0.62	0.63
IsoNet	2.09	1.90	1.89	0.05	0.05	0.05
Original tomogram	2.53	2.33	2.32	0.05	0.05	0.05

3.3 Generalization to unseen tomograms

We also wanted to test how our proof-of-concept framework generalize to unseen tomograms, a task IsoNet was not designed to perform. When trained on five tomograms and tested on a held-out specimen, our model synthesized ±45° projections with MSE 0.33 and correlation 0.83, significantly outperforming IsoNet (MSE: 0.67, correlation: 0.66; **Table 4**).

Table 4: Comparison of MSE and correlation error for predicted Cryo-EM projections from new tomogram for recovering of masked tilts (M) and seen tilts (S).

	No masking		Randomly masked tilts			
Method	**MMSE**	**MCorr**	**MMSE (S)**	**MCorr (S)**	**MMSE (M)**	**MCorr (M)**
Our	0.33	0.83	0.33	0.83	0.34	0.82
IsoNet	0.67	0.66	-	-	-	-

The higher performance of our model compared to IsoNet stems from its explicit modeling of 3D Fourier geometry through a physics-aware inductive bias, encoded in the latent back-projection module. Unlike IsoNet, which treats tomograms as generic 3D image volumes, our latent back-projection module learns to aggregate 2D tilt-series projections into a unified 3D representation that inherently adheres to the Fourier slice theorem. This geometric prior ensures that high-frequency components are propagated orthogonally across slices, preserving spatial relationships critical for resolving anisotropic features (e.g., membrane curvature, **Fig. 3**).

4 Discussion

The experiments reveal the crucial conceptual difference between our approach and IsoNet. IsoNet trains on reconstructed tomograms which already contain missing wedge artifacts, synthetically introducing additional artifacts at different angles. By learning the pattern of artifacts induced by the synthetic missing wedge, IsoNet iteratively infers what might have been in the position of the actual missing wedge. By training on tomograms with synthetic missing wedges, IsoNet learns to mimic artifacts rather than recover genuine structural information. This creates a self-reinforcing cycle where artifacts from the initial reconstruction are amplified during inpainting. The discrepancy between IsoNet published results and our findings is most likely due to the limited tilt range of ±45° compared to the IsoNet benchmark dataset with a ±60° range. The use of tomograms as input meant that this relative lack of information was embedded in the input data.

In contrast, our approach operates on tilt series, where each tilt is a projection containing information on the entirety of the imaged volume. Because the angles of the tilt series do not span a complete rotation, retrieving this information precisely is ill-posed. Approximating solutions to ill-posed problems is, however, a classic use for deep learning models and exactly how our model is designed. Our encoder independently operates on the tilts and we combine them into a latent-space representation of the imaged volume via weighted back-projection. Of course, this reconstructed latent would have missing wedge artifacts and, like IsoNet, we utilize 3D convolution to correct for these artifacts. However, unlike IsoNet, we then compute projections from this 3D latent which are trained to match the original tilt series. By constraining the latent space in this way, our model avoids the ad-hoc texture hallucination of IsoNet, instead extrapolating missing wedge frequencies through structured interpolation in Fourier space. This inductive bias enables our model to generalize to unseen tomograms and angles, tasks where IsoNet fails due to its reliance on local pixel statistics rather than global geometric consistency.

This work is a proof of concept, outlining a scalable method for developing a deep-learning based approach for up-sampling Cryo-ET tilt series. Initial evidence showing that our method is able to benefit from training on a large number of tilt series to learn how to generally approximate rotation-limited back-projection. As the output is synthesized projections, these can be concatenated to the original tilt series and input into any Cryo-ET reconstruction algorithm to mitigate the missing wedge artifacts in the final tomogram.

5 ACKNOWLEDGMENTS

We thank Aygul Ishemgulova (NYSBC, New York) for providing E. coli mini-cell tomogram used in this paper. This work was performed in part at the Simons Electron Microscopy Center located at the New York Structural Biology Center and was supported by grants from the Simons Foundation (SF349247) and Chan Zuckerberg Initiative (2023-331615 (5022) GB-1597298). This material is based upon work supported in part by the National Science Foundation under the award DGE-1735316.

REFERENCES

Moumita Dutta and Acharya Priyamvada. Cryo-electron microscopy in the study of virus entry and infection. *Frontiers in molecular biosciences*, 1:677–89, 2024. doi: 10.3389/fmolb.2024.1429180.

Aygul Ishemgulova, Alex J. Noble, Tristan Bepler, and Alex De Marco. Preparation of labeled cryo-et datasets for training and evaluation of machine learning models. *NeurIPS 2023 Workshop: Machine Learning in Structural Biology Workshop*, 2023.

Yun-tao Liu, Heng Zhang, Hui Wang, Chang-Lu Tao, Guo-Qiang Bi, and Z. Hong Zhou. Isotropic reconstruction for electron tomography with deep learning. *Nature Communications*, 13(6482), 2022. doi: https://doi.org/10.1038/s41467-022-33957-8.

Eva Nogales and H W Scheres Sjors. Cryo-em: A unique tool for the visualization of macromolecular complexity. *Molecular cell*, 58:677–89, 2015. doi: 10.1016/j.molcel.2015.02.019.

Yan R, Venkatakrishnan SV, Liu J, Bouman CA, and Jiang W. Mbir: A cryo-et 3d reconstruction method that effectively minimizes missing wedge artifacts and restores missing information. *J Struct Biol*, 52(183-192), 2019. doi: 10.1016/j.jsb.2019.03.002.

M Radermacher. Weighted back-projection methods. In *Electron Tomography*. Springer, 2007. doi: doi.org/10.1007/978-0-387-69008-7_9.

Zhi Yang, Jiwei Qin, Chuan Lin, Yanping Chen, Ruizhang Huang, and Yongbin Qin. Ganrec: A negative sampling model with generative adversarial network for recommendation. *Expert Systems with Applications*, 52(119155), 2023. doi: doi.org/10.1016/j.eswa.2022.119155.

Lindsey N Young and Elizabeth Villa. Bringing structure to cell biology with cryo-electron tomography. *Annual review of biophysics*, 52(573-595), 2023. doi: 10.1146/annurev-biophys-111622-091327.

A APPENDIX

A.1 MATHEMATICAL JUSTIFICATION FOR WBP IN LATENT SPACE

Back-projection in latent space is mathematically well-suited for Cryo-ET due to its ability to address challenges such as noise, missing data, and the ill-posed nature of the reconstruction problem. The Fourier projection slice theorem forms the foundation of back-projection, stating that the Fourier transform of a 2D projection corresponds to a central slice of the 3D Fourier transform of the volume. Weighted back-projection (WBP) refines this process by compensating for uneven sampling in Fourier space caused by limited tilt angles. Mathematically, WBP reconstructs the 3D volume $V(x, y, z)$ as:

$$V(x, y, z) = \int_{\theta} W(k) P_{\theta}(x, y)\, d\theta,$$

where $W(k)$ is a weighting function to correct geometric distortions, and $P_{\theta}(x, y)$ represents the 2D projections at tilt angle θ. Additionally, Cryo-ET reconstruction is an ill-posed inverse problem due to incomplete data and noise. This can be formulated as:

$$\min_{V} \|AV - P\|^2 + \lambda R(V),$$

where A models the imaging process, P represents observed projections, $R(V)$ is a regularization term (e.g., enforcing smoothness or sparsity), and $\lambda > 0$ controls regularization strength. Latent space methods further enhance reconstruction by encoding noisy projections into a lower-dimensional representation that captures biologically relevant features while discarding noise. Noise reduction can be modeled as:

$$z_{\theta} = f(P_{\theta} + N), \quad f(P_{\theta} + N) = f(P_{\theta}),$$

where $N \sim \mathcal{N}(0, \sigma^2)$ represents additive Gaussian noise, and $f(\cdot)$ is an encoder function that isolates meaningful signal components from noise. Together, these allows us for an effective generalization of novel projections from reconstructed latent 3D volumes in Cryo-ET.

A.2 SUPPLEMENTARY FIGURES

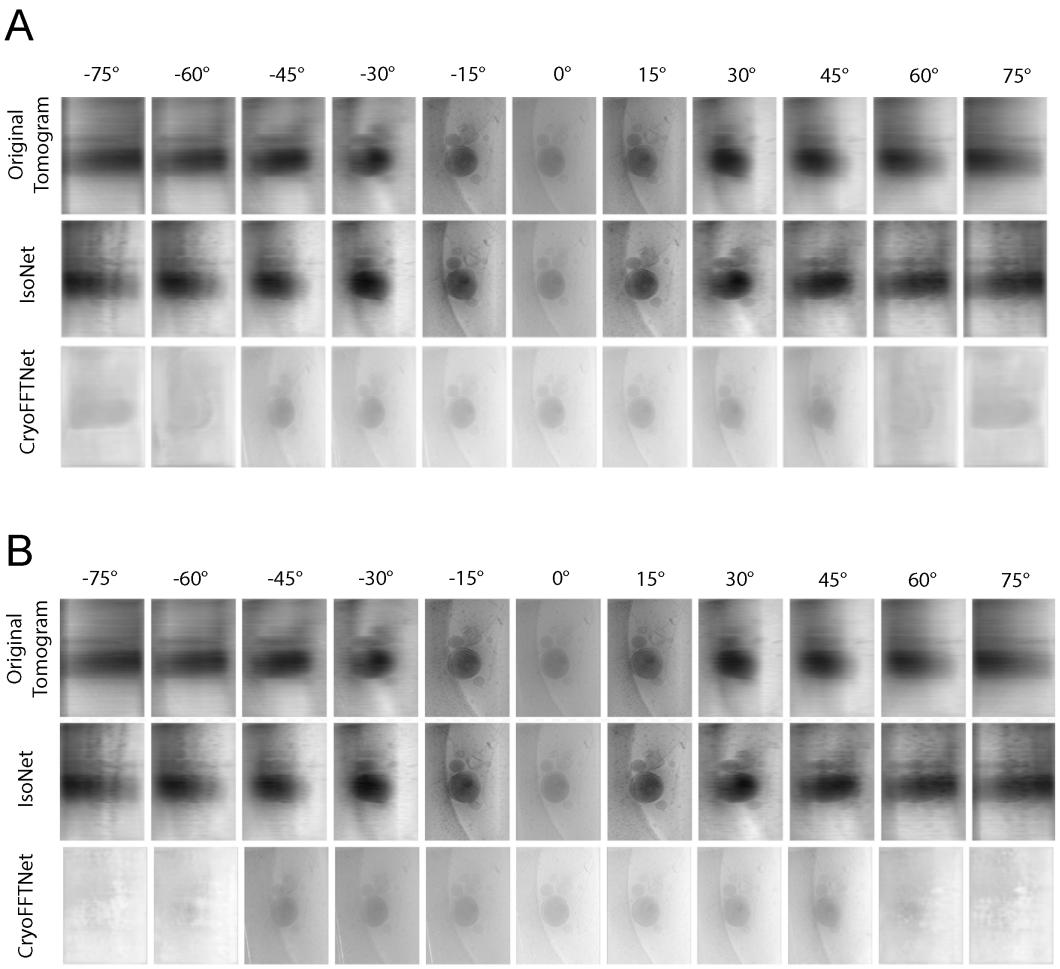

Figure 3: Synthesize tilts projection from different models. **A)** Tilts created by rotation of original tomogram and pre-trained IsoNet or CryFFTNet model trained on 5 tomograms. **B)** The same tilts angle as in A created by rotation of original tomogram and pre-trained IsoNet on 5 tomograms or CryFFTNet model trained on 129 tomograms.