bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MemBrain v2: an end-to-end tool for the analysis of membranes in cryo-electron tomography

Lorenz Lamm[1,2,3], Simon Zufferey[2], Ricardo D. Righetto[2], Wojciech Wietrzynski[2], Kevin A. Yamauchi[4,5], Alister Burt[6], Ye Liu[1,3], Hanyi Zhang[1,3], Antonio Martinez-Sanchez[7], Sebastian Ziegler[8,9], Fabian Isensee[8,9], Julia A. Schnabel[1,3,10], Benjamin D. Engel[2,*] Tingying Peng[1,3,*]

1 Helmholtz Munich – German Research Center for Environment and Health, Munich, Germany
2 Biozentrum, University of Basel, Basel, Switzerland
3 School of Computation, Information and Technology, Technical University of Munich, Munich, Germany
4 Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland
5 Swiss Institute of Bioinformatics, Basel, Switzerland
6 MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK.
7 Department of Information and Communications Engineering, Faculty of Computers Sciences, University of Murcia, Murcia, Spain
8 German Cancer Research Center (DKFZ), Division of Medical Image Computing, Heidelberg, Germany
9 Helmholtz Imaging, German Cancer Research Center (DKFZ), Heidelberg, Germany
10 School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

* Correspondence to: lorenz.lamm@helmholtz-munich.de, ben.engel@unibas.ch, tingying.peng@helmholtz-munich.de

ABSTRACT

MemBrain v2 is a deep learning-enabled program aimed at the efficient analysis of membranes in cryo-electron tomography (cryo-ET). The final v2 release of MemBrain will comprise three main modules: 1) MemBrain-seg, which provides automated membrane segmentation, 2) MemBrain-pick, which provides automated picking of particles along segmented membranes, and 3) MemBrain-stats, which provides quantitative statistics of particle distributions and membrane morphometrics.

This initial version of the manuscript is focused on the beta release of MemBrain-seg, which combines iterative training with diverse data and specialized Fourier-based data augmentations. These augmentations are specifically designed to enhance the tool's adaptability to a variety of tomographic data and address common challenges in cryo-ET analysis. A key feature of MemBrain-seg is the implementation of the Surface-Dice loss function, which improves the network's focus on membrane connectivity and allows for the effective incorporation of manual annotations from different sources. This function is beneficial in handling the variability inherent in membrane structures and annotations. Our ongoing collaboration with the cryo-ET community plays an important role in continually improving MemBrain v2 with a wide array of training data. This collaborative approach ensures that MemBrain v2 remains attuned to the field's needs, enhancing its robustness and generalizability across different types of tomographic data.

The current version of MemBrain-seg is available at https://github.com/teamtomo/membrain-seg, and the predecessor of MemBrain-pick (also called MemBrain v1) is deposited at https://github.com/CellArchLab/MemBrain. This preprint will be updated concomitantly with the code until the three integrated modules of MemBrain v2 are complete.

Keywords Cryo-electron tomography · Membranes · Segmentation · MemBrain · U-Net

1 Introduction

1.1 Background

Cryo-electron tomography (cryo-ET) is a powerful technique for imaging the molecular environment inside native cells in three dimensions (3D) at sub-nanometer resolution (Turk and Baumeister, 2020). Cryo-ET visualizes all densities inside a vitreously frozen cellular volume, creating vast potential to investigate the structures, interactions, and spatial arrangements of diverse organellar and macromolecular components. However, this biological complexity is also a central challenge for analyzing cryo-ET data (Young and Villa, 2023), and it remains difficult to accurately annotate the many features inside cellular tomograms. Membranes and their embedded protein complexes are fundamental to numerous biological processes that are visible by cryo-ET, including the ER-Golgi secretory pathway (Gemmer et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2023; Albert et al., 2020; Tran et al., 2021; Bykov et al., 2017), autophagy (M. Li et al., 2023; Bieber et al., 2022), vesicular transport (Foster et al., 2021), synaptic transmission (Radecke et al., 2023; Tao et al., 2018; Radhakrishnan et al., 2021; Liu, Tao, et al., 2020), energy generation in mitochondrial cristae and chloroplast thylakoids (Barad et al., 2023; Waltz et al., 2021; Huokko et al., 2021; Wietrzynski et al., 2020; Gupta et al., 2021), organelle contact sites (Wozny et al., 2023; Collado et al., 2019), cell membrane protrusions (Rodrigues-Oliveira et al., 2023; Sartori-Rupp et al., 2019), viral replication (Mendonça et al., 2021; Klein, Cortese, et al., 2020; Wolff et al., 2020; Klein, Golani, et al., 2023; Armbruster et al., 2023; Dahmane et al., 2022), and bacterial cell division (Navarro et al., 2022; Khanna et al., 2021). Accurate segmentation of membranes and localization of their associated proteins would greatly accelerate cryo-ET studies that link membrane architecture and molecular organization to cellular function.

The segmentation of membranes in cryo-electron tomograms poses significant challenges, notably due to the low signal-to-noise ratio inherent to cryo-ET data. Various denoising methods have been employed, including classical Wiener-like deconvolution (Tegunov and Cramer, 2019), as well as innovative deep learning-based approaches (Buchholz et al., 2019; Purnell et al., 2023; Bepler et al., 2020; Maldonado et al., 2023). For example, Cryo-CARE builds a noise model by comparing even and odd frames of a tilt-series (Buchholz et al., 2019), and hence can denoise images without ever seeing a noise-free example. An additional difficulty is anisotropic information from the "missing wedge" (Wan and Briggs, 2016; Schmid and Booth, 2008), a consequence of the incomplete angular sampling in cryo-ET data acquisition due to geometric limitations of tilting a slab of ice. This anisotropy leads to distortion or complete loss of certain features, particularly a problem for horizontally oriented membranes where densities can almost vanish from tomograms. Previous methods attempted to mitigate the missing wedge effects by iteratively imposing prior knowledge about the sample geometry on the 3D reconstruction (Biyani et al., 2018; Deng et al., 2016). More recently, IsoNet (Liu, H. Zhang, et al., 2022) and DeepDeWedge (Wiedemann and Heckel, 2023) tackle this problem by using deep learning models to fill in the absent information.

Additional challenges to membrane segmentation include the diverse nature of membrane architectures and the variation in tomogram appearances. Membranes exhibit a wide range of shapes and curvatures, often decorated with various proteins that modify their visual characteristics. Furthermore, different instrumentation setups (microscope, camera, energy filter) and acquisition parameters (tilt scheme, defocus, pixel size) influence the appearance and contrast of the tomograms.

This biological and technical diversity complicates the development of a universally effective segmentation method. Particularly for deep neural networks, which have become an intuitive choice due to the rise of deep learning in biomedical image analysis (Ronneberger et al., 2015; Isensee et al., 2021), such variability is demanding. For example, the training of a convolutional neural network (CNN) relies heavily on access to a comprehensive and varied dataset. This dataset must comprise the full spectrum of membrane types and tomographic appearances to ensure generalizability of the model. At present, however, such a dataset is not readily available.

1.2 Related Work

Various methodologies have been developed to address the challenge of membrane segmentation in cryo-ET. A prominent method is TomoSegMemTV (Martinez-Sanchez, Garcia, et al., 2014), a non-machine learning based computer vision algorithm. This framework utilizes the eigenvalues derived from the tomogram's second derivative and integrates tensor voting (Tong et al., 2004) to leverage contextual information from neighboring voxels. TomoSegMemTV operates under the premise that membranes are plane-like structure with smooth curvatures. This assumption becomes a limitation when dealing with membranes exhibiting increased curvature, often leading to inaccuracies in segmentation. Notably, the method's reliance on tensor voting for score propagation can result in erroneous outcomes, such as incorrectly extended segmentations (example in Figure 9).

ColabSeg (Siggel et al., 2023) is a novel membrane segmentation tool. It integrates TomoSegMemTV — and is anticipated to soon include MemBrain-seg — as a foundation for initial membrane segmentation. These preliminary segmentations are then transformed into point clouds, enabling the application of various parametric fitting techniques through interactive processes. This feature allows users to refine segmentations, such as enhancing membrane continuity or adjusting the shapes to match spherical membranes, thereby improving the accuracy and applicability of the segmentations.

Recent advancements in membrane segmentation have been significantly driven by U-Net-based approaches (Ronneberger et al., 2015), which excel in understanding the context and intricate shapes of membranes. DeePiCt (Teresa-Trueba et al., 2023) exemplifies this trend, employing various U-Nets for segmenting different cell compartments, including a 3D U-Net specifically for membrane segmentation. DeePiCt enhances its capability with a Fourier amplitude spectrum matching module designed to mimic diverse tomogram appearances. While DeePiCt offers pre-trained models and training data, such data are still limited to a single cell type.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

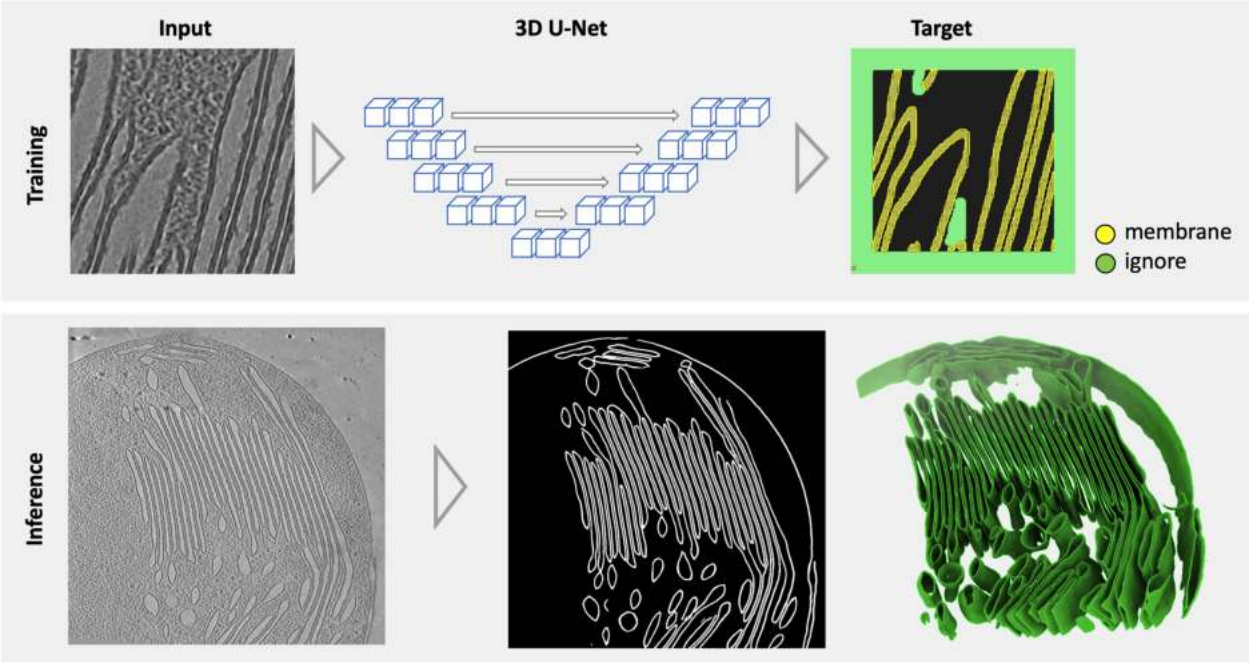

Figure 1: **Overview of MemBrain-seg**: MemBrain-seg applies a 3D U-Net, trained on precisely annotated data, to deliver robust membrane segmentations across diverse tomograms, eliminating the need for additional training at the inference stage.

DeepFinder (Moebel et al., 2021) and EMAN2 (M. Chen et al., 2017) utilize 3D and 2D U-Nets, respectively, to segment user-defined structures within tomograms. However, these tools do not come with pre-trained models or training datasets to their users. Another innovative approach (L. Zhou et al., 2023) uses synthetically generated tomograms to train CNNs to refine membrane segmentations, utilizing techniques like reinforcement learning and parametric fitting. However, this method is currently not publicly available.

Moreover, nnU-Net (Isensee et al., 2021), known for its self-configuring capabilities in biomedical image segmentation, and Dragonfly (Comet Technologies Canada Inc., 2022), with its emphasis on user-friendly interactive segmentation, are noteworthy contributions to this field. Both methods exemplify the progression towards more adaptable and user-centric segmentation tools in cryo-ET.

1.3 Our contribution

To enhance generalization across a broad spectrum of tomograms, we have meticulously curated an accurate and diverse dataset of membrane segmentations. This dataset serves as the training material for a 3D U-Net tailored for membrane segmentation tasks: MemBrain-seg is an intuitive, user-friendly tool designed for segmenting a wide array of membrane and tomogram appearances by the use of our generalizable pre-trained model.

A beneficial aspect of our approach is the incorporation of specialized data augmentation techniques. These include both missing wedge and Fourier amplitude augmentation, which can improve segmentation accuracy, particularly in low-signal areas and across diverse tomographic styles. When combined with our extensive and varied dataset, these augmentations result in robustly accurate membrane segmentations across a variety of membrane types and tomographic images. Furthermore, the Surface-Dice score, introduced in our study, serves as both an evaluation metric and a loss function, allowing for the analysis of membrane continuity and the ability to concentrate training efforts on enhancing this aspect.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2 Methods

2.1 MemBrain-seg Overview

MemBrain-seg, a core module of MemBrain v2, leverages a 3D-U-Net architecture optimized for cryo-ET membrane segmentation. In the following sections, we will describe MemBrain-seg's features:

- **3D-U-Net Based Architecture:** Incorporating a 3D-U-Net framework, MemBrain-seg is tailored for high-precision membrane segmentation in diverse cryo-ET data.
- **Specialized Data Augmentations:** The model employs missing wedge and Fourier amplitude augmentations, enhancing adaptability to a wide range of tomographic styles and improving segmentation in low-signal areas.
- **Surface-Dice Loss Function:** We integrate the SurfaceDice loss to improve membrane continuity.
- **Preprocessing Options:** MemBrain-seg offers various preprocessing options, including pixel-size matching, Fourier amplitude matching, and Wiener-like deconvolution, aiding in aligning diverse tomographic data with the model's training domain.
- **Iterative Dataset Generation:** We utilize an iterative approach to generate training annotations from various datasets, continuously refined for enhanced model performance.

2.2 Network Training

2.2.1 U-Net implementation and training

Our MemBrain-seg U-Net architecture is based on principles from nnU-Net (Isensee et al., 2021), a leading framework in the realm of biomedical image segmentation, and implemented using the MONAI framework (Cardoso et al., 2022). The U-Net comprises five downsampling and upsampling operations, respectively. Deep supervision (L. Wang et al., 2015) enhances gradient flow, ensuring robust network convergence. We employ a polynomial scheduler for the learning rate, decreasing from 0.01 to 0.0. Before training we normalize patch intensity by separately subtracting their means and dividing by their standard deviations.

For loss functions, we combine binary cross-entropy, Dice loss, and our Surface-Dice loss (see Section 2.2.2). An *ignore* feature in the loss calculation evaluates the loss only for voxels that do not carry the *ignore* label in ground truth annotations, detailed in Section 2.5.2.

During training, data augmentations are key to the success of our approach. We utilize a combination of standard and custom augmentations specifically tailored to cryo-ET data and membrane segmentation, described further in Section 2.3.

2.2.2 SurfaceDice

Motivation. Diversity in ground truth (GT) membrane annotations, arising from both the tomogram representation (e.g. pixel size, reconstruction algorithm, filtering) and the annotating expert, poses a challenge for accurate membrane prediction (illustrated in Supp. Figure 8). The thickness variations in GT segmentations can skew the evaluation of predicted membrane segmentations, and focus more on thickness differences than the topology of the membranes (see Figure 2C). During training, traditional Dice loss may prioritize thicker membranes due to their larger volume of positive samples. However, accurately capturing membrane continuity and overall shape is more critical than replicating thickness.

To address this, we introduce Surface-Dice, an advancement of Centerline-Dice (Shit et al., 2021) for 3D membrane segmentation. Surface-Dice serves as both an evaluative metric on binarized segmentation outputs and a loss function during network training by utilizing a differentiable skeletonization approach. This metric better reflects the continuity and integrity of membrane structures in 3D.

Loss Computation. Like Centerline-Dice, Surface-Dice employs skeletonizations of both predicted (M_{pred}) and ground truth (M_{GT}) segmentations, denoted as S_{pred} and S_{GT} respectively. In our approach, *skeletonization* reduces the membrane segmentation to a 1-voxel-thick surface or a 2D manifold in 3D space (illustrated in Figure 2A,B), as compared to reduction to a centerline (i.e., a 1D manifold).

To calculate Surface-Dice ($\text{Dice}_{\text{surf}}$), we define Surface-precision ($\text{Prec}_{\text{surf}}$) and Surface-recall ($\text{Rec}_{\text{surf}}$) as follows:

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

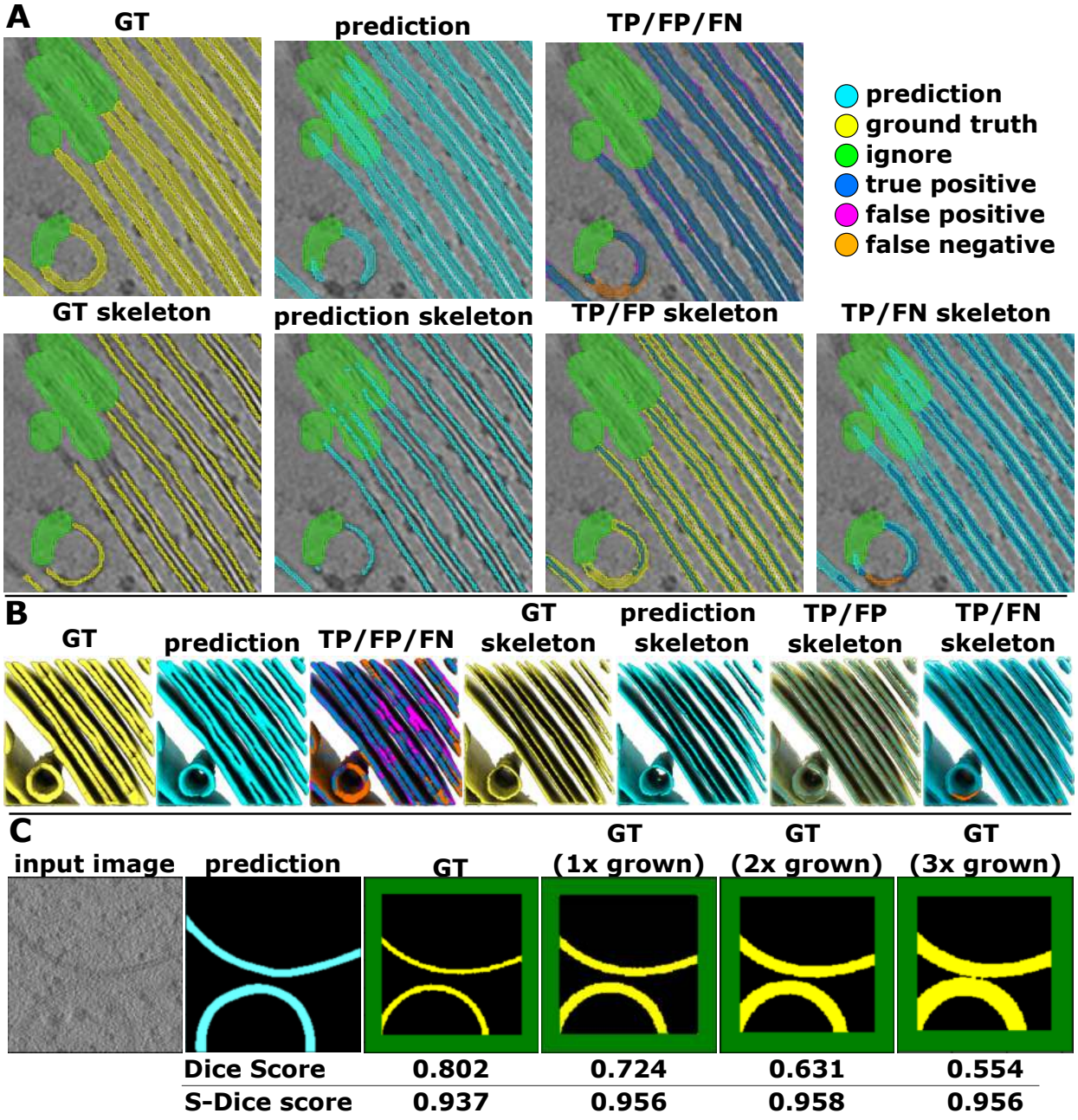

Figure 2: **Surface-Dice vs. Dice: A:** Top row shows ground truth (GT) and predicted membranes, with voxel classifications for Dice score calculation. Bottom row illustrates skeletons for GT and predictions, used to evaluate precision and recall in Surface-Dice. **B:** 3D renderings of segmentations and skeletons in A. **C:** Prediction compared against GTs of varying thickness to demonstrate Surface-Dice's consistency versus Dice score variability.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$\text{Prec}_{\text{surf}}(M_{\text{pred}}, M_{\text{GT}}) = \frac{|S_{\text{pred}} \cap M_{\text{GT}}|}{|S_{\text{pred}}|} \qquad \text{Rec}_{\text{surf}}(M_{\text{pred}}, M_{\text{GT}}) = \frac{|S_{\text{GT}} \cap M_{\text{pred}}|}{|S_{\text{GT}}|} \tag{1}$$

The $|\,.\,|$-operator sums up all positive voxels in the respective set. Subsequently, Surface Dice ($\text{Dice}_{\text{surf}}$) is computed as their harmonic mean:

$$\text{Dice}_{\text{surf}}(M_{\text{pred}}, M_{\text{GT}}) = 2 \times \frac{\text{Prec}_{\text{surf}}(M_{\text{pred}}, M_{\text{GT}}) \times \text{Rec}_{\text{surf}}(M_{\text{pred}}, M_{\text{GT}})}{\text{Prec}_{\text{surf}}(M_{\text{pred}}, M_{\text{GT}}) + \text{Rec}_{\text{surf}}(M_{\text{pred}}, M_{\text{GT}})} \tag{2}$$

Skeletonization. For using Surface Dice as a loss function in training, we avoid thresholding network scores, as this is a non-differentiable operation. Instead, we adopt a differentiable approach for skeletonization, similar to the method used in Centerline-Dice. This involves an iterative process where membrane erosion is performed to progressively thin the segmentations. Each iteration includes additional erosion and dilation of the segmentation, and we assess the differences between the segmentation before and after these operations. A difference implies that the erosion has removed parts of the membrane and thus the current iteration already depicts a thin surface in that specific region. Importantly, both erosion and dilation steps are implemented through min- and max-pooling, ensuring that these operations remain differentiable. For more details, refer to Shit et al., 2021.

2.3 Data Augmentation

Data augmentation plays an essential role in our network training by mitigating overfitting through artificially increasing the dataset size (Shorten and Khoshgoftaar, 2019). By exposing the network to significantly altered images, we enhance its robustness and ability to generalize across varying tomographic and membrane scenarios. Below, we detail the specific data augmentation techniques used for this network.

2.3.1 Conventional data augmentations

Our data augmentation strategy, largely in line with nnU-Net (Isensee et al., 2021), incorporates a mix of geometric and intensity transformations:

- **Geometric Transforms**: These include rotation at arbitrary angles, zooming within a range of 0.7 to 1.4, and both shuffling and flipping of axes.
- **Intensity Transforms**: Applied to alter image characteristics, these encompass median filtering, Gaussian blurring, Gaussian noise addition, brightness and contrast adjustments, low-resolution simulation, random erasing, additive brightness gradient, local Gamma transform, and sharpening.

These transformations, applied randomly and in combinations, significantly vary the appearance of tomographic images, contributing to the model's adaptability. For a visual representation of these transformations on a single patch, see Figure 3A.

2.3.2 Fourier Amplitude Spectrum augmentation

Fourier amplitude spectrum matching, as introduced in DeePiCt (Teresa-Trueba et al., 2023), emulates the style of various tomograms for normalization prior to segmentation. In MemBrain-seg, however, our objective is to enhance network generalizability without the extra step of normalization before inference. Style augmentations during training have been shown to be more effective for generalization compared to pre-prediction style normalization (Tellez et al., 2019; Wagner et al., 2021).

Consequently, we developed *Fourier amplitude spectrum augmentation* for training, simulating different tomogram styles. This is achieved by performing random Fourier amplitude matching using different normalization vectors. To prevent network bias towards specific tomogram styles, these vectors are randomly generated rather than derived from a limited set of tomograms.

The process involves creating a $1-D$ sequence x of length 80 (i.e., subvolume size / 2) through a random walk:

$$x_0 = 1.0 \tag{3}$$

$$x_{n+1} = |x_n + \Delta_n| \tag{4}$$

$$\Delta_n \sim \mathcal{N}(0, \sigma^2), \tag{5}$$

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

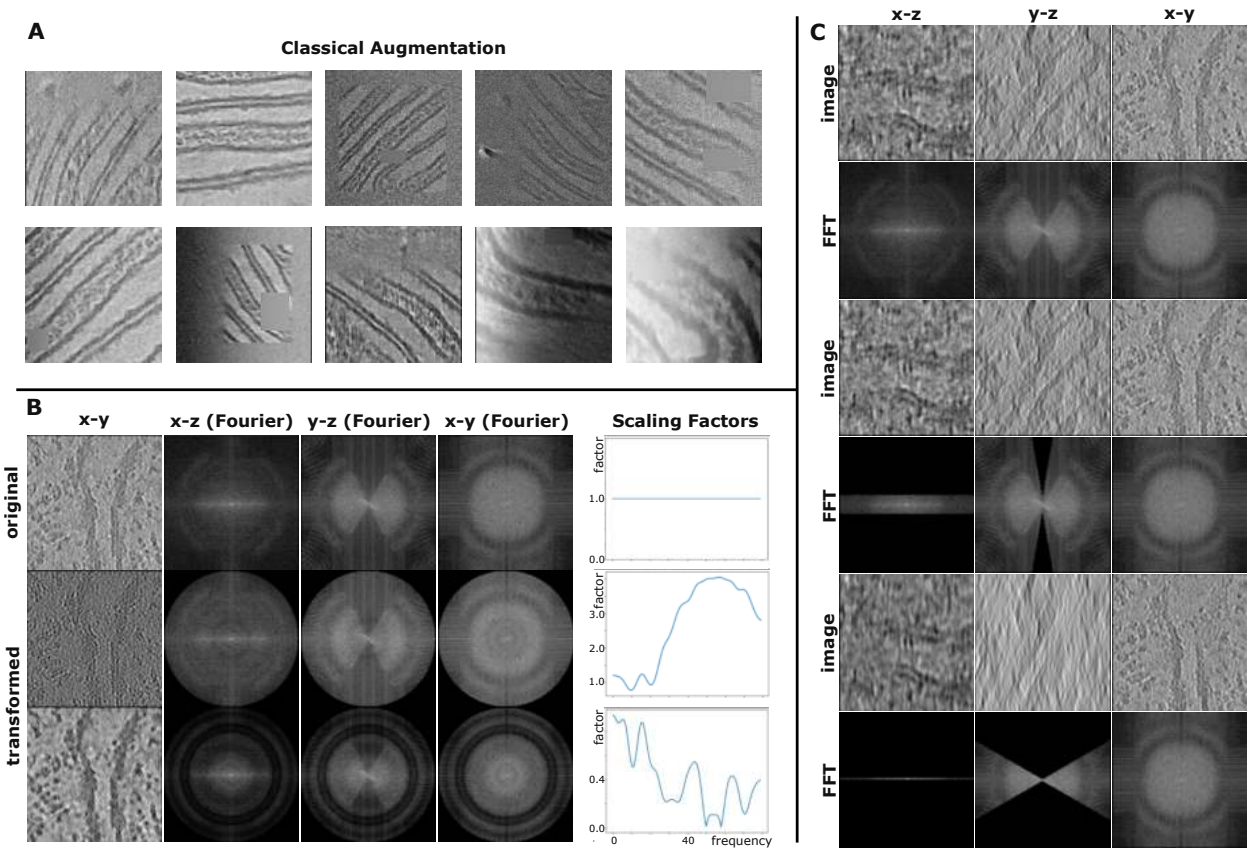

Figure 3: **Different Data Augmentations**: **A:** 2D slice of the same 3D training patch after random applications of our classical augmentations. **B:** Fourier Intensity Augmentation: The first column displays the x-y view of the original patch (top row) and its augmented states (following rows) in real space. Columns two to four depict the effects in Fourier space from different perspectives (x-z, y-z, x-y). The final column shows the variable scaling factors for each frequency component applied during augmentation. **C:** Missing Wedge augmentation: Rows 1 and 2 illustrate the x-z, y-z, and x-y views of the original patch in real and Fourier spaces, respectively. Rows 3 to 6 display these perspectives after augmentation, showcasing the introduction of the missing wedge effect.

where σ is empirically set to 0.2. We then convert this 1D plot into a 3D radial kernel, and multiply it with the Fourier space of the volume. The effects on image appearance are illustrated in Figure 3B: Up-scaling of higher frequencies leads to noisier images with less pronounced membranes (2nd row), whereas down-scaling of high frequencies leads to a stronger focus of the coarser features, including membranes (3rd row).

2.3.3 Missing Wedge Data augmentation

Accurately segmenting membranes in regions impacted by the missing wedge is a difficult challenge. To address this, we adopt a technique inspired by IsoNet (Liu, H. Zhang, et al., 2022), which involves randomly rotating input subvolumes and then introducing an *artificial missing wedge* by masking Fourier coefficients in a wedge-like shape. This process, as depicted in Figure 3C, simulates the distortion of membrane densities in the tomograms. In experimental data, this means that the missing wedge problem is made *worse*, but in a controllable direction. Utilizing the corresponding ground truth map for these artificially altered subvolumes, we generate training pairs that mimic the actual missing wedge's effect. This approach aims to enhance the model's performance in segmenting membranes in missing wedge-impacted regions.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2.4 Preprocessing Options

In case MemBrain-seg's out-of-the-box segmentation does not perform as desired, we offer several preprocessing tools to bring tomograms closer to our training domain, particularly when they deviate significantly. We describe these tools in the following.

2.4.1 Pixel-size Matching

Our training data spans pixel sizes from approximately 9Å to 14Å. We employ random zooming (factor range 0.7 to 1.4) during training for robustness against pixel size variations. For tomograms too far outside of this favorable range, we provide a pixel-size matching module, utilizing Fourier cropping and extension techniques (Smith et al., 1997) to rescale tomograms to a similar pixel size as the training domain.

2.4.2 Fourier Amplitude Matching

In addition to using Fourier Amplitude style transfer for data augmentation during training, we use an optional Fourier Amplitude Matching algorithm, adapted from DeePiCt (Teresa-Trueba et al., 2023), for normalizing tomograms to our training dataset's appearance.

2.4.3 Wiener-like Deconvolution

Although MemBrain-seg is trained also on non-denoised tomograms, enhancing contrast and reducing noise can improve segmentation. When applying Cryo-CARE (Buchholz et al., 2019) or similar noise-to-noise denoising approaches are not feasible, we suggest using Warp's Wiener-like deconvolution filter (Tegunov and Cramer, 2019) to improve tomogram clarity and ideally MemBrain-seg's segmentation performance. For the user's convenience, we make this filter available as a preprocessing option in the MemBrain-seg package.

2.5 Training Dataset

2.5.1 Challenges in generating training data

Currently, there's a lack of a comprehensive, accurately annotated, and publicly accessible datasets for membrane segmentation in cryo-ET. The DeePiCt dataset (Teresa-Trueba et al., 2023), while also providing membrane segmentations, primarily addresses general cell compartment segmentation, and contains inconsistencies in thickness and incompleteness in its membrane segmentations (see Figure 8).

The generation of a large cryo-ET membrane segmentation dataset faces significant challenges: Manual 3D annotation is labor-intensive, requiring constant navigation between multiple 2D slices to ensure accurate structural representation both in individual slices and in the overall 3D context. Additionally, the inherent noise and artifacts in cryo-ET images, particularly the *missing wedge* (Wan and Briggs, 2016; Schmid and Booth, 2008), often make membranes barely visible. This makes precise membrane delineation difficult, if not impossible, despite clear indications of their presence. Such image quality limitations also hinder the effectiveness of existing automated methods like TomoSegMemTV, which primarily relies on tomogram density values. This approach especially struggles to accurately identify membranes in regions with non-flat geometries (see also Supp. Figure 9).

2.5.2 Iterative approach for generating initial training data

To efficiently create a well-annotated membrane segmentation dataset, we adopted an iterative methodology based on active learning principles (Settles, 2009). Our approach was designed to minimize manual annotation efforts, concentrating specifically on areas where the network faced segmentation challenges. This iterative approach is schematically visualized in Figure 4A.

We initiated our project with TomoSegMemTV-annotated patches from our *Spinacia oleacera* dataset. Using MITK Workbench (MITK-Team, 2023; Wolf et al., 2004), we manually refined these patches, each sized $160 \times 160 \times 160$ voxels, assigning to every voxel either the *background*, *membrane*, or *ignore* classes - the latter not being evaluated by the loss function during training as it marks uncertain membrane regions, where exact delineation of membranes was challenging. The process of correcting predicted membrane segmentations is depicted in Figure 2.5.2B, where we remove false positive voxels from the segmentation, add false negatives to it, and coarsely assign the *ignore* class in uncertain regions.

After this primary round of annotating *Spinacia oleacera* patches, we trained a U-Net (Ronneberger et al., 2015; Isensee et al., 2021) model for membrane segmentation. This model, once applied to the same tomograms, outperformed

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

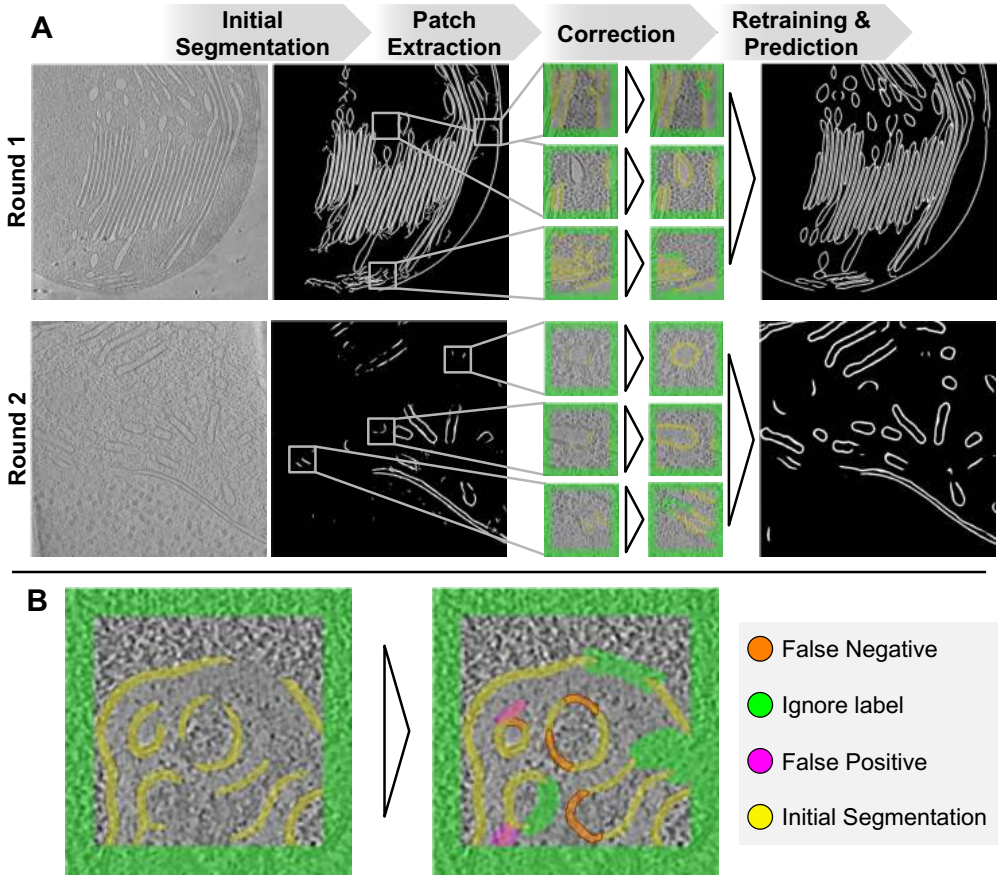

Figure 4: **Iterative Annotation and Training Process A:** The two leftmost columns show tomogram intensities together with TomoSegMemTV segmentation (row 1) or intermediate MemBrain-seg iteration (row 2). Patches of poor segmentations are extracted and manually corrected. After re-training MemBrain-seg, segmentations are improved (right column). **B:** Manual correction of extracted patches in MITK Workbench MITK-Team, 2023; Wolf et al., 2004: False negatives are added (orange) and false positives removed (magenta). In highly ambiguous regions, the *ignore* label is assigned.

TomoSegMemTV but still did not give satisfying segmentations. Consequently, we performed a second round of annotation, targeting patches with suboptimal segmentations, and integrating these into our training set. The process resulted in 61 accurately annotated patches of *Spinacia oleacera* chloroplasts (see Table 1 Round 1).

We then extended our approach to a dataset of *Chlamydomonas reinhardtii* tomograms, acquired with a different setup. Despite initial discrepancies in segmentation quality compared to our spinach results, iterative re-annotation of 16 patches (Table 1 Round 2; including thylakoid membranes, mitochondria, Golgi apparatus) enhanced the model's accuracy notably.

To further enhance MemBrain-seg's capabilities across various tomography settings and membrane structures, we integrated external datasets from publicly available sources. This expansion added 20 DeePiCt patches (Table 1 Round 4), where we extracted patches in regions where both MemBrain-seg's performance was limited and the available ground truth was well-annotated. Additionally, we generated 20 patches using synthetic tomogram generators CryoTomoSim (Purnell et al., 2023) and PolNet (Martinez-Sanchez, Jasnin, et al., 2023), respectively (Table 1 Round 5).

By collaborating with various research groups from the cryo-ET community, we are continuing to add more diverse samples to our training data. This includes samples with many different tomography setups, membrane architectures and cellular components like microtubules or cell walls.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

During the annotation process, we pay particular attention to the quality of the training dataset, i.e., the accuracy of the segmentation. The *ignore* label is particularly helpful here, as it prevents the model from learning uncertain regions simply as background. This annotation quality is a key factor for the generalization capability of our network.

2.5.3 Our integrated datasets

Our training datasets exhibit a large diversity, comprising examples from the following sources:

- *Spinacia oleacera:* The Spinach dataset, predominantly featuring sheet-like thylakoid membranes, served as our initial training base. We utilized 61 patches from this dataset, and progressively integrated additional datasets in subsequent iterations.
- *Chlamydomonas reinhardtii:* Developed in collaboration with the Max-Planck-Institute for Biochemistry, Martinsried, and Thermo-Fisher-Scientific, Eindhoven, this dataset captures various compartments of the single-cell green alga, including thylakoid membranes, the Golgi apparatus, and mitochondria. We extracted 16 patches with limited segmentation performance from this dataset, corrected them, and merged them into our training repertoire.
- *DeePiCt:* From these tomograms of *Schizosaccharomyces pombe*, we selected patches where our MemBrain-seg predictions were suboptimal, yet the DeePiCt ground truth (Teresa-Trueba et al., 2023) proved reliable. We thus extracted 20 patches from the tomograms and merged them into our training dataset. For some of the patches, we performed additional manual corrections using the *Ignore* label.
- *CryoTomoSim* (Purnell et al., 2023): Utilizing this simulation tool, we generated 20 new training patches and 10 for testing purposes. For the generation of these patches, we varied the simulated tomogram acquisition parameters (e.g. applied electron dose, defocus), as well as the number of inserted membranes and the types and concentrations of proteins.
- *PolNet* (Martinez-Sanchez, Jasnin, et al., 2023): To further increase the diversity of our training patches, additional 20 training and 10 test patches were created using the PolNet simulator. We again varied simulation factors like signal-to-noise ratios and different membrane types and protein presences.

3 Results

To evaluate the performance of MemBrain-seg's components, we conducted a series of experiments, with their outcomes depicted in Figure 5. Separate test sets were designated for each dataset—*Spinach*, *Chlamy*, *synthetic* (CryoTomoSim and PolNet combined), and *DeePiCt* — ensuring no overlap with training and validation data.

As illustrated in Figure 5, we analyzed the impact of various design elements on the model's performance across these test datasets, focusing on both Dice and Surface-Dice scores.

3.1 Incremental Training Improves Overall Model Performance

We conducted five training rounds, progressively incorporating new data to the training set (refer to Table 1), and assessed their outcomes on distinct test sets (Figure 5). Figure 6 additionally shows the effects of different training rounds on some test tomograms.

The *Spinach* test set consistently showed uniform performance across all phases, highlighting stable model behavior. In contrast, the *Chlamy* dataset demonstrated optimal Dice scores during the second phase, where the fraction of *Chlamy* patches in the training set was the highest, with subsequent rounds not achieving comparable results. This suggests that dataset-specific fine-tuning of networks might be optimal for achieving the best specialized outcomes.

Even though Round 3 integrated the same data as Round 2, supplemented with corresponding raw versions of the denoised (Buchholz et al., 2019) patches, the *Chlamy* dataset exhibited a decline in performance. This drop might be attributed to the fact that the *Chlamy* test set consists of only denoised patches. However, incorporating these mixed-quality patches enhanced the model's generalization capabilities, particularly evident in the improved results for the *DeePiCt* and *synthetic* datasets. The *DeePiCt* dataset, comprising both defocus and VPP tomograms without denoising, particularly benefited from this approach.

Overall, adding more training data steadily improved average performance, with only slight decreases on particular datasets (*Chlamy*, *DeePiCt*). This reflects our model's generalization capability, which can then be customized to perform best on specific datasets.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

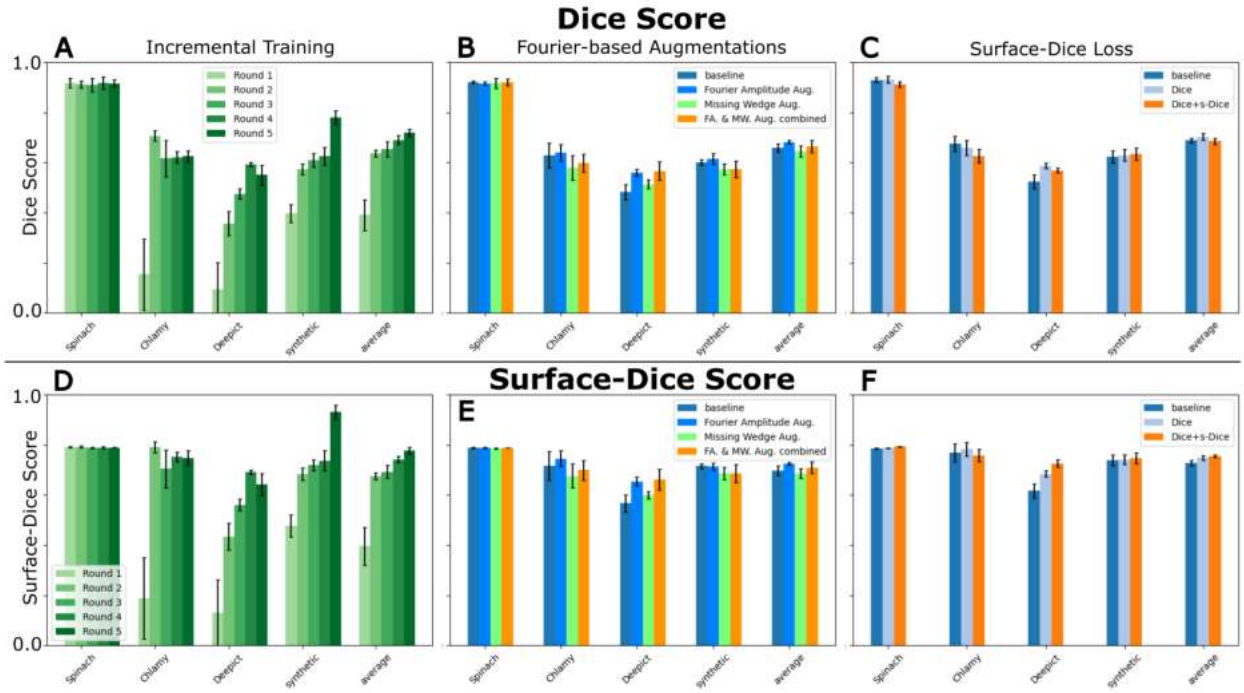

Figure 5: **Evaluation Scores**: Accumulated Dice scores (A-C) and accumulated Surface-Dice scores (D-F) for different experiments, evaluated on varying test datasets. **(A, D)** show incremental training results for Rounds 1-5. **(B, E)** depict results when training with our custom Fourier-based augmentations. Models in this experiment are trained with data from Round 3. **(C, F)** show the effects of using Surface-Dice as a loss function. Models are trained using data from Round 3 (baseline) or Round 4 (others), i.e., including data from DeePiCt, where the additional patches were included using either Dice loss or a combination of Dice and Surface-Dice losses.

Interestingly, incorporating *synthetic* data primarily benefited its corresponding test set, with minimal or even slightly decreased performance (*DeePiCt*). However, empirically, we observe an improvement using synthetic data in Figure 6: Round 5, trained with synthetic data, shows that training with missing-wedge-free synthetic patches appears to enhance segmentation completeness, particularly top and bottom segmentations of vesicles, as exemplified in tomogram EMD-18748 (Radecke et al., 2023).

3.2 Customized Data Augmentations Aid the Network Generalization

To explore the benefits of our custom Fourier-based data augmentations, we conducted an ablation study using the training data from Round 3 (refer to Table 1), i.e., excluding the *DeePiCt* dataset. The study compares a *baseline* model (without custom Fourier augmentations) against models trained with either *Fourier Amplitude augmentation*, *Missing Wedge augmentation*, or a combination of both. Results of these experiments are depicted in Figure 5B,E.

The *Spinach* test set maintained consistent results across models. The *Fourier Amplitude* augmentation notably improved performance, particularly for the *DeePiCt* test set, indicating better network generalization to unseen membrane appearances.

In contrast, the *Missing Wedge* augmentation did not consistently improve segmentation across datasets. Its impact was slightly negative for the *Chlamy* and *synthetic* datasets but showed some improvement for *DeePiCt*. We note that the evaluation of this augmentation is challenging due to the absence of missing-wedge-free ground truth in experimental datasets, and its effects on the data appearance might be overly harsh.

Combining both augmentations yielded mixed results, performing better than *Missing Wedge* alone but not as effective as *Fourier Amplitude*, which is in line with stacking the effects of each separate augmentation.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

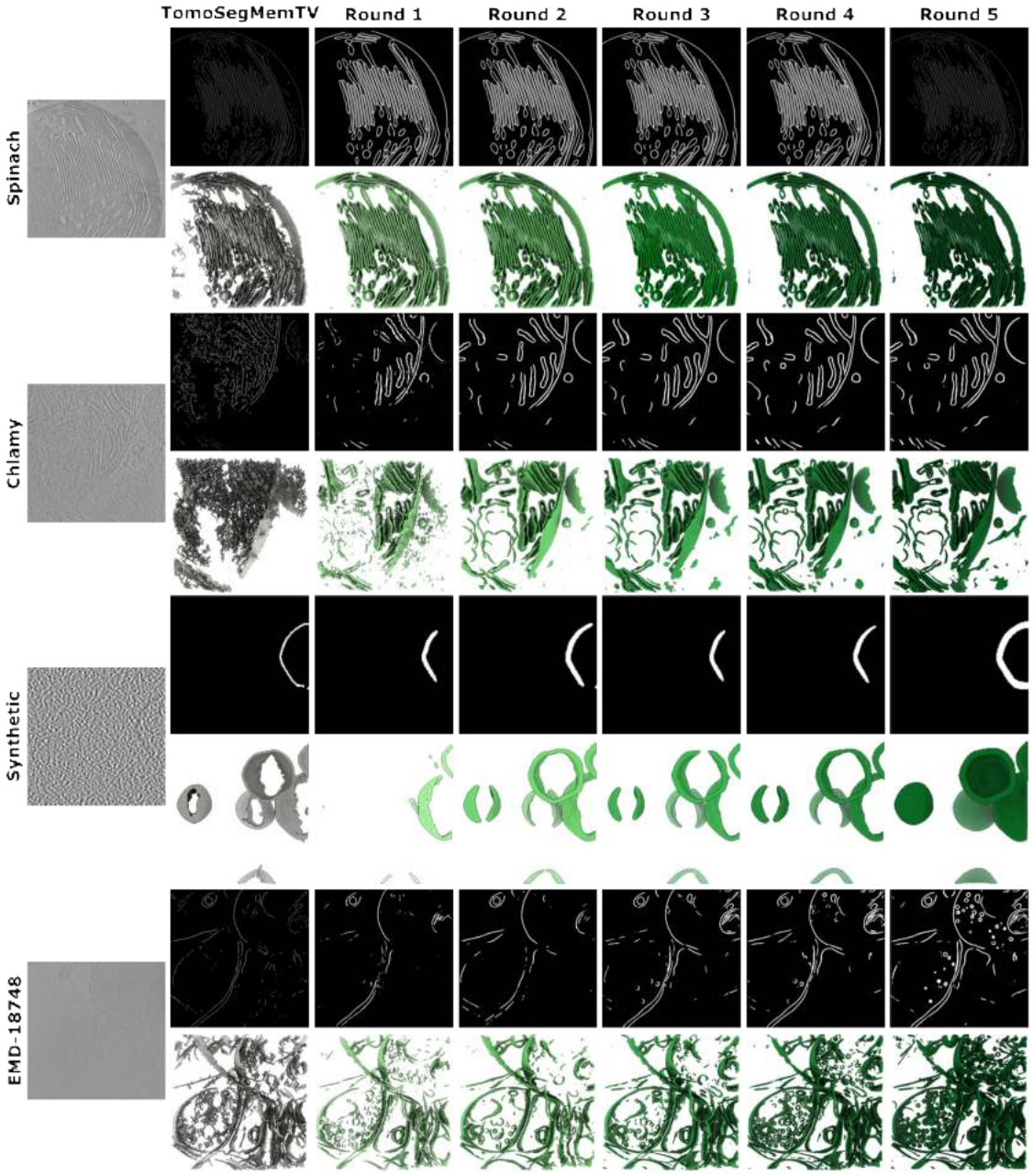

Figure 6: **Evolution of Segmentation Across Training Rounds.** Displayed are segmentation outcomes for four datasets over five incremental training rounds, alongside initial results from TomoSegMemTV for comparison. Starting with the original tomograms, each row progresses through successive rounds, showing the development of our segmentation method.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

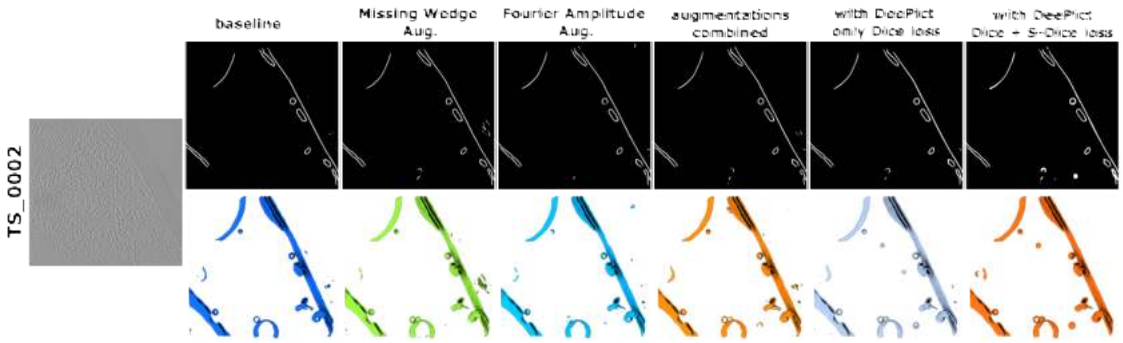

Figure 7: **Segmentation Performance with Varied Augmentations and Loss Functions.** The tomogram (TS_0002) is followed by segmentation results after training with data from Round 3 and using Missing Wedge Augmentation, Fourier Amplitude Augmentation, and a combination of both. The two rightmost columns showcase the results after training with data from Round 4, using only Dice loss, or a combination of Dice and Surface-Dice loss. Top: 2D segmentation masks. Bottom: 3D segmented membrane renderings.

3.3 SurfaceDice as a Loss Function Enables Integration of Non-uniform Annotations

In assessing the SurfaceDice as a loss function, we conducted an experiment (Figure 5) using a baseline model trained with Round 3 data and only Dice as loss function. Additionally, we implemented two training variations with Round 4 data: one using only *Dice* loss and the other combining *Dice* and *Surface-Dice* losses. Specifically, Round 3 patches were processed with Dice loss, while new Round 4 patches (*DeePiCt*) utilized SurfaceDice loss.

The Dice-only training maintained consistent Dice and SurfaceDice scores for *Spinach*, *Chlamy*, and *synthetic* datasets. For the *DeePiCt* dataset, both metrics are clearly improved compared to the baseline.

With the combined loss training, a slight decrease in Dice scores was observed for the *Chlamy* dataset, but SurfaceDice scores remained stable across *Spinach*, *Chlamy*, and *synthetic* datasets. Importantly, this approach yielded the best Surface-Dice scores for the *DeePiCt* dataset, highlighting the effectiveness of our Surface-Dice loss in optimizing connectivity for new datasets without compromising connectivity performance on existing ones.

Figure 7 underlines this outcome: The two rightmost columns show results of the two training settings on a test tomogram of the *DeePiCt* dataset. Training with a combination of Dice and Surface-Dice leads to smooth segmentations, and particularly helps with closing the segmentations of vesicles, which is not the case when training only with Dice loss.

4 Discussion

MemBrain v2 represents an important step forward in the segmentation of membranes in cryo-electron tomography. Our approach, combining iterative training with a large variety in data sources and specialized Fourier-based augmentations, has led to a notable improvement in the tool's ability to handle diverse tomographic data. The introduction of the Surface-Dice loss function provides a way to focus the network training on the membrane connectivity and allows to incorporate annotations from various sources into the training.

Importantly, we are continually working on improving MemBrain v2 and extend the training data in close collaboration with the cryo-ET community. This ongoing effort is crucial for enhancing the model's robustness and its capability to generalize across a wide range of tomographic data.

MemBrain v2 has the potential to significantly aid the cryo-ET community in unraveling and analyzing complex membrane structures more effectively. As we progress, we aim to not only refine MemBrain-seg's performance but also implement new functionalities that broaden its application, including the automated localization of membrane proteins (MemBrain-pick) and the statistical analysis of their organization (MemBrain-stats). The output segmentations from MemBrain-seg can also be fed into complementary software packages for the analysis of membrane curvature and morphometrics (Barad et al., 2023; Salfer et al., 2020).

We will continually update this preprint with the new functionalities added to MemBrain v2.

MemBrain-seg, together with our latest model, is available on GitHub: https://github.com/teamtomo/membrain-seg. It is also available as a plugin for Scipion (Jiménez de la Morena et al., 2022): https://github.com/scipion-em/scipion-

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

em-membrain. Until the release of MemBrain-pick, we recommend users pick protein complexes along segmented membranes with MemBrain v1 (Lamm et al., 2022), which is deposited at https://github.com/CellArchLab/MemBrain.

5 Acknowledgements

We are grateful to everyone who has tested and provided feedback on the beta version of MemBrain v2, including Matthias Pöge, Robert Brandt, Przemek Dutka, Wangbiao (Seven) Guo, Benoit Gallet, Thomas O'Sullivan, Xiyan Zhu, Zhen Hou, Peijun Zhang, Veijo Salo, Thomas Hoffmann, Giulia Zanetti, Jannis Anstatt, Jenny Sachweh, Abraham (Bram) Koster, Yunjie Chang, Yury Bykov, Pankaj Vilas Jadhav, Duolin Shepherd, and Christos Papantoniou.

We thank Abhay Kotecha and his team at Thermo Fisher Scientific for leading the Chlamydomonas visual proteomics project that generated much of the raw data used for initial training of MemBrain-seg.

Special thanks to users who also contributed training patches and annotations to improve MemBrain-seg's generalizability: Virly Ananda, Paula Perez Navarro, Benedikt Wimmer, Ohad Medalia, Ya-Ting (Atty) Chang, Michaela Medina, Ben Barad, Danielle Grotjahn, Malit Jessie James Limlingan, Martin Pilhofer, Lena Thärichen, Leanne de Jager, Marten Chaillet, Friedrich Förster, Ben Silva, Donghyun "Raphael" Park, Rory Hennell James, Thomas Marlovits, Sven Klumpe, Jasmina Redzovic, Fridolin Koch, Grigory Tagiltsev, and John Briggs. These contributions have significantly improved MemBrain-seg's capabilities.

Calculations were performed at sciCORE (https://scicore.unibas.ch/) scientific computing center at the University of Basel. L.L. acknowledges support from the Munich School for Data Science (MUDS) and a fellowship from the Boehringer Ingelheim Fonds. A.M-S is supported by the *Ramon y Cajal* program of the Spanish State Research Agency (AEI) [RYC2021-032626-I, 2021] and the University of Murcia [Attract-RYC, 2023]. B.D.E. acknowledges support from an HFSP Research Grant (RGP0005/2021) and ERC consolidator grant "cryOcean" (fulfilled by the Swiss State Secretariat for Education, Research and Innovation, M822.00045). A.B. is supported by the Medical Research Council [MC_UP_1201/6].

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Albert, Sahradha, Wojciech Wietrzynski, Chia-Wei Lee, Miroslava Schaffer, Florian Beck, Jan M Schuller, Patrice A Salomé, Jürgen M Plitzko, Wolfgang Baumeister, and Benjamin D Engel (2020). "Direct visualization of degradation microcompartments at the ER membrane". In: *Proceedings of the National Academy of Sciences* 117.2, pp. 1069–1080.

Armbruster, Emily, Jina Lee, Joshua Hutchings, Arica VanderWal, Eray Enustun, Benjamin Adler, Ann Aindow, Amar Deep, Zaida Rodriguez, Chase Morgan, et al. (2023). "Sequential membrane-and protein-bound organelles compartmentalize genomes during phage infection". In: *bioRxiv*, pp. 2023–09.

Barad, Benjamin A, Michaela Medina, Daniel Fuentes, R Luke Wiseman, and Danielle A Grotjahn (2023). "Quantifying organellar ultrastructure in cryo-electron tomography using a surface morphometrics pipeline". In: *Journal of Cell Biology* 222.4, e202204093.

Bepler, Tristan, Kotaro Kelley, Alex J Noble, and Bonnie Berger (2020). "Topaz-Denoise: general deep denoising models for cryoEM and cryoET". In: *Nature communications* 11.1, p. 5208.

Bieber, Anna, Cristina Capitanio, Philipp S Erdmann, Fabian Fiedler, Florian Beck, Chia-Wei Lee, Delong Li, Gerhard Hummer, Brenda A Schulman, Wolfgang Baumeister, et al. (2022). "In situ structural analysis reveals membrane shape transitions during autophagosome formation". In: *Proceedings of the National Academy of Sciences* 119.39, e2209823119.

Biyani, Nikhil, Sebastian Scherer, Ricardo D. Righetto, Julia Kowal, Mohamed Chami, and Henning Stahlberg (Aug. 2018). "Image processing techniques for high-resolution structure determination from badly ordered 2D crystals". In: *Journal of Structural Biology* 203.2. Publisher: Elsevier, pp. 120–134. ISSN: 10478477. DOI: 10.1016/j.jsb.2018.03.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S104784771830090X.

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

Bykov, Yury S, Miroslava Schaffer, Svetlana O Dodonova, Sahradha Albert, Jürgen M Plitzko, Wolfgang Baumeister, Benjamin D Engel, and John AG Briggs (2017). "The structure of the COPI coat determined within the cell". In: *Elife* 6, e32493.

Cardoso, M Jorge, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, et al. (2022). "Monai: An open-source framework for deep learning in healthcare". In: *arXiv preprint arXiv:2211.02701*.

Chen, Muyuan, Wei Dai, Stella Y Sun, Darius Jonasch, Cynthia Y He, Michael F Schmid, Wah Chiu, and Steven J Ludtke (2017). "Convolutional neural networks for automated annotation of cellular cryo-electron tomograms". In: *Nature methods* 14.10, pp. 983–985.

Collado, Javier, Maria Kalemanov, Felix Campelo, Clélia Bourgoint, Ffion Thomas, Robbie Loewith, Antonio Martínez-Sánchez, Wolfgang Baumeister, Christopher J Stefan, and Ruben Fernandez-Busnadiego (2019). "Tricalbin-mediated contact sites control ER curvature to maintain plasma membrane integrity". In: *Developmental cell* 51.4, pp. 476–487.

Comet Technologies Canada Inc. (2022). *Dragonfly*. Available at https://www.theobjects.com/dragonfly. Version 2022.2. [Computer software]. Comet Technologies Canada Inc., Montreal, Canada. URL: https://www.theobjects.com/dragonfly.

Dahmane, Selma, Adeline Kerviel, Dustin R Morado, Kasturika Shankar, Björn Ahlman, Michael Lazarou, Nihal Altan-Bonnet, and Lars-Anders Carlson (2022). "Membrane-assisted assembly and selective secretory autophagy of enteroviruses". In: *Nature communications* 13.1, p. 5986.

Deng, Yuchen, Yu Chen, Yan Zhang, Shengliu Wang, Fa Zhang, and Fei Sun (2016). "ICON: 3D reconstruction with 'missing-information'restoration in biological electron tomography". In: *Journal of structural biology* 195.1, pp. 100–112.

Foster, Helen E, Camilla Ventura Santos, and Andrew P Carter (2021). "A cryo-ET survey of microtubules and intracellular compartments in mammalian axons". In: *Journal of Cell Biology* 221.2, e202103154.

Gemmer, Max, Marten L Chaillet, Joyce van Loenhout, Rodrigo Cuevas Arenas, Dimitrios Vismpas, Mariska Gröllers-Mulderij, Fujiet A Koh, Pascal Albanese, Richard A Scheltema, Stuart C Howes, et al. (2023). "Visualization of translation and protein biogenesis at the ER membrane". In: *Nature* 614.7946, pp. 160–167.

Gupta, Tilak Kumar, Sven Klumpe, Karin Gries, Steffen Heinz, Wojciech Wietrzynski, Norikazu Ohnishi, Justus Niemeyer, Benjamin Spaniol, Miroslava Schaffer, Anna Rast, et al. (2021). "Structural basis for VIPP1 oligomerization and maintenance of thylakoid membrane integrity". In: *Cell* 184.14, pp. 3643–3659.

Huokko, Tuomas, Tao Ni, Gregory F Dykes, Deborah M Simpson, Philip Brownridge, Fabian D Conradi, Robert J Beynon, Peter J Nixon, Conrad W Mullineaux, Peijun Zhang, et al. (2021). "Probing the biogenesis pathway and dynamics of thylakoid membranes". In: *Nature communications* 12.1, p. 3475.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Isensee, Fabian, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein (2021). "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature methods* 18.2, pp. 203–211.

Jiménez de la Morena, J., P. Conesa, Y. C. Fonseca, F. P. de Isidro-Gómez, D. Herreros, E. Fernández-Giménez, D. Strelak, E. Moebel, T. O. Buchholz, F. Jug, A. Martinez-Sanchez, M. Harastani, S. Jonic, J. J. Conesa, A. Cuervo, P. Losana, I. Sánchez, M. Iceta, L. del Cano, M. Gragera, R. Melero, G. Sharov, D. Castaño-Díez, A. Koster, J. G. Piccirillo, J. L. Vilas, J. Otón, R. Marabini, C. O. S. Sorzano, and J. M. Carazo (June 2022). "ScipionTomo: Towards cryo-electron tomography software integration, reproducibility, and validation". en. In: *Journal of Structural Biology*, p. 107872. ISSN: 1047-8477. DOI: 10.1016/j.jsb.2022.107872. URL: https://www.sciencedirect.com/science/article/pii/S1047847722000429 (visited on 06/07/2022).

Khanna, Kanika, Javier Lopez-Garrido, Joseph Sugie, Kit Pogliano, and Elizabeth Villa (2021). "Asymmetric localization of the cell division machinery during Bacillus subtilis sporulation". In: *Elife* 10, e62204.

Klein, Steffen, Mirko Cortese, Sophie L Winter, Moritz Wachsmuth-Melm, Christopher J Neufeldt, Berati Cerikan, Megan L Stanifer, Steeve Boulant, Ralf Bartenschlager, and Petr Chlanda (2020). "SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography". In: *Nature communications* 11.1, p. 5885.

Klein, Steffen, Gonen Golani, Fabio Lolicato, Carmen Lahr, Daniel Beyer, Alexia Herrmann, Moritz Wachsmuth-Melm, Nina Reddmann, Romy Brecht, Mehdi Hosseinzadeh, et al. (2023). "IFITM3 blocks influenza virus entry by sorting lipids and stabilizing hemifusion". In: *Cell Host & Microbe* 31.4, pp. 616–633.

Lamm, Lorenz, Ricardo D Righetto, Wojciech Wietrzynski, Matthias Pöge, Antonio Martinez-Sanchez, Tingying Peng, and Benjamin D Engel (2022). "MemBrain: A deep learning-aided pipeline for detection of membrane proteins in Cryo-electron tomograms". In: *Computer methods and programs in biomedicine* 224, p. 106990.

Li, Meijing, Ishita Tripathi-Giesgen, Brenda A Schulman, Wolfgang Baumeister, and Florian Wilfling (2023). "In situ snapshots along a mammalian selective autophagy pathway". In: *Proceedings of the National Academy of Sciences* 120.12, e2221712120.

Liu, Yun-Tao, Chang-Lu Tao, Xiaokang Zhang, Wenjun Xia, Dong-Qing Shi, Lei Qi, Cheng Xu, Rong Sun, Xiao-Wei Li, Pak-Ming Lau, et al. (2020). "Mesophasic organization of GABAA receptors in hippocampal inhibitory synapses". In: *Nature neuroscience* 23.12, pp. 1589–1596.

Liu, Yun-Tao, Heng Zhang, Hui Wang, Chang-Lu Tao, Guo-Qiang Bi, and Z Hong Zhou (2022). "Isotropic reconstruction for electron tomography with deep learning". In: *Nature communications* 13.1, p. 6482.

Maldonado, Jeronimo Carvajal, Lorenz Lamm, Ye Liu, Yu Liu, Ricardo D Righetto, Julia A Schnabel, and Tingying Peng (2023). "F2FD: Fourier Perturbations for Denoising Cryo-Electron Tomograms and Comparison to Established Approaches". In: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, pp. 1–5.

Martinez-Sanchez, Antonio, Inmaculada Garcia, Shoh Asano, Vladan Lucic, and Jose-Jesus Fernandez (2014). "Robust membrane detection based on tensor voting for electron tomography". In: *Journal of structural biology* 186.1, pp. 49–61.

Martinez-Sanchez, Antonio, Marion Jasnin, Harold Phelippeau, and Lorenz Lamm (2023). "Simulating the cellular context in synthetic datasets for cryo-electron tomography". In: *bioRxiv*, pp. 2023–05.

Mendonça, Luiza, Andrew Howe, James B Gilchrist, Yuewen Sheng, Dapeng Sun, Michael L Knight, Laura C Zanetti-Domingues, Benji Bateman, Anna-Sophia Krebs, Long Chen, et al. (2021). "Correlative multi-scale cryo-imaging unveils SARS-CoV-2 assembly and egress". In: *Nature Communications* 12.1, p. 4629.

MITK-Team (Nov. 2023). *MITK*. Version v2023.12. URL: https://github.com/MITK/MITK.

Moebel, Emmanuel, Antonio Martinez-Sanchez, Lorenz Lamm, Ricardo D Righetto, Wojciech Wietrzynski, Sahradha Albert, Damien Larivière, Eric Fourmentin, Stefan Pfeffer, Julio Ortiz, et al. (2021). "Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms". In: *Nature methods* 18.11, pp. 1386–1394.

Navarro, Paula P, Andrea Vettiger, Virly Y Ananda, Paula Montero Llopis, Christoph Allolio, Thomas G Bernhardt, and Luke H Chao (2022). "Cell wall synthesis and remodelling dynamics determine division site architecture and cell shape in Escherichia coli". In: *Nature Microbiology* 7.10, pp. 1621–1634.

Purnell, Carson, Jessica Heebner, and Matt Swulius (2023). *Training Neural Networks With Simulated CryoET Data*.

Radecke, Julika, Raphaela Seeger, Anna Kádková, Ulrike Laugks, Amin Khosrozadeh, Kenneth N Goldie, Vladan Lučić, Jakob B Sørensen, and Benoıt Zuber (2023). "Morphofunctional changes at the active zone during synaptic vesicle exocytosis". In: *EMBO reports* 24.5, e55719.

Radhakrishnan, Abhijith, Xia Li, Kirill Grushin, Shyam S Krishnakumar, Jun Liu, and James E Rothman (2021). "Symmetrical arrangement of proteins under release-ready vesicles in presynaptic terminals". In: *Proceedings of the National Academy of Sciences* 118.5, e2024029118.

Rodrigues-Oliveira, Thiago, Florian Wollweber, Rafael I Ponce-Toledo, Jingwei Xu, Simon K-MR Rittmann, Andreas Klingl, Martin Pilhofer, and Christa Schleper (2023). "Actin cytoskeleton and complex cell architecture in an Asgard archaeon". In: *Nature* 613.7943, pp. 332–339.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, pp. 234–241.

Salfer, Maria, Javier F Collado, Wolfgang Baumeister, Rubén Fernández-Busnadiego, and Antonio Martínez-Sánchez (2020). "Reliable estimation of membrane curvature for cryo-electron tomography". In: *PLOS Computational Biology* 16.8, e1007962.

Sartori-Rupp, Anna, Diégo Cordero Cervantes, Anna Pepe, Karine Gousset, Elise Delage, Simon Corroyer-Dulmont, Christine Schmitt, Jacomina Krijnse-Locker, and Chiara Zurzolo (2019). "Correlative cryo-electron microscopy reveals the structure of TNTs in neuronal cells". In: *Nature communications* 10.1, p. 342.

Schmid, Michael F and Christopher R Booth (2008). "Methods for aligning and for averaging 3D volumes with missing data". In: *Journal of structural biology* 161.3, pp. 243–248.

Settles, Burr (2009). "Active learning literature survey". In.

Shit, Suprosanna, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze (2021). "clDice-a novel topology-preserving loss function for tubular structure segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16560–16569.

Shorten, Connor and Taghi M Khoshgoftaar (2019). "A survey on image data augmentation for deep learning". In: *Journal of big data* 6.1, pp. 1–48.

Siggel, Marc, Rasmus Kjeldsen Jensen, Julia Mahamid, and Jan Kosinski (2023). "ColabSeg: An interactive tool for editing, processing, and visualizing membrane segmentations from cryo-ET data". In: *bioRxiv*, pp. 2023–07.

Smith, Steven W et al. (1997). *The scientist and engineer's guide to digital signal processing*.

Tao, Chang-Lu, Yun-Tao Liu, Rong Sun, Bin Zhang, Lei Qi, Sakar Shivakoti, Chong-Li Tian, Peijun Zhang, Pak-Ming Lau, Z Hong Zhou, et al. (2018). "Differentiation and characterization of excitatory and inhibitory synapses by cryo-electron tomography and correlative microscopy". In: *Journal of Neuroscience* 38.6, pp. 1493–1510.

Tegunov, Dimitry and Patrick Cramer (2019). "Real-time cryo-electron microscopy data preprocessing with Warp". In: *Nature methods* 16.11, pp. 1146–1152.

Tellez, David, Geert Litjens, Péter Bándi, Wouter Bulten, John-Melle Bokhorst, Francesco Ciompi, and Jeroen Van Der Laak (2019). "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology". In: *Medical image analysis* 58, p. 101544.

Teresa-Trueba, Irene de, Sara K Goetz, Alexander Mattausch, Frosina Stojanovska, Christian E Zimmerli, Mauricio Toro-Nahuelpan, Dorothy WC Cheng, Fergus Tollervey, Constantin Pape, Martin Beck, et al. (2023). "Convolutional networks for supervised mining of molecular patterns within cellular context". In: *Nature Methods* 20.2, pp. 284–294.

Tong, Wai-Shun, Chi-Keung Tang, Philippos Mordohai, and Gerard Medioni (2004). "First order augmentation to tensor voting for boundary inference and multiscale analysis in 3D". In: *IEEE transactions on pattern analysis and machine intelligence* 26.5, pp. 594–611.

Tran, Ngoc-Han, Stephen D Carter, Ann De Mazière, Avi Ashkenazi, Judith Klumperman, Peter Walter, and Grant J Jensen (2021). "The stress-sensing domain of activated IRE1α forms helical filaments in narrow ER membrane tubes". In: *Science* 374.6563, pp. 52–57.

Turk, Martin and Wolfgang Baumeister (2020). "The promise and the challenges of cryo-electron tomography". In: *FEBS letters* 594.20, pp. 3243–3261.

Wagner, Sophia J, Nadieh Khalili, Raghav Sharma, Melanie Boxberg, Carsten Marr, Walter de Back, and Tingying Peng (2021). "Structure-preserving multi-domain stain color augmentation using style-transfer with disentangled representations". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*. Springer, pp. 257–266.

Waltz, Florent, Thalia Salinas-Giegé, Robert Englmeier, Herrade Meichel, Heddy Soufari, Lauriane Kuhn, Stefan Pfeffer, Friedrich Förster, Benjamin D Engel, Philippe Giegé, et al. (2021). "How to build a ribosome from RNA fragments in Chlamydomonas mitochondria". In: *Nature Communications* 12.1, p. 7176.

Wan, W and John AG Briggs (2016). "Cryo-electron tomography and subtomogram averaging". In: *Methods in enzymology* 579, pp. 329–367.

Wang, Liwei, Chen-Yu Lee, Zhuowen Tu, and Svetlana Lazebnik (2015). "Training deeper convolutional networks with deep supervision". In: *arXiv preprint arXiv:1505.02496*.

Wiedemann, Simon and Reinhard Heckel (Nov. 2023). *A Deep Learning Method for Simultaneous Denoising and Missing Wedge Reconstruction in Cryogenic Electron Tomography*. arXiv:2311.05539 [cs]. DOI: 10.48550/arXiv.2311.05539. URL: http://arxiv.org/abs/2311.05539 (visited on 11/14/2023).

Wietrzynski, Wojciech, Miroslava Schaffer, Dimitry Tegunov, Sahradha Albert, Atsuko Kanazawa, Jürgen M Plitzko, Wolfgang Baumeister, and Benjamin D Engel (2020). "Charting the native architecture of Chlamydomonas thylakoid membranes with single-molecule precision". In: *Elife* 9, e53740.

Wolf, Ivo, Marcus Vetter, Ingmar Wegner, Marco Nolden, Thomas Bottger, Mark Hastenteufel, Max Schobinger, Tobias Kunert, and Hans-Peter Meinzer (2004). "The medical imaging interaction toolkit (MITK): a toolkit facilitating the

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.05.574336; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

creation of interactive software by extending VTK and ITK". In: *Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display*. Vol. 5367. SPIE, pp. 16–27.

Wolff, Georg, Ronald WAL Limpens, Jessika C Zevenhoven-Dobbe, Ulrike Laugks, Shawn Zheng, Anja WM de Jong, Roman I Koning, David A Agard, Kay Grünewald, Abraham J Koster, et al. (2020). "A molecular pore spans the double membrane of the coronavirus replication organelle". In: *Science* 369.6509, pp. 1395–1398.

Wozny, Michael R, Andrea Di Luca, Dustin R Morado, Andrea Picco, Rasha Khaddaj, Pablo Campomanes, Lazar Ivanović, Patrick C Hoffmann, Elizabeth A Miller, Stefano Vanni, et al. (2023). "In situ architecture of the ER–mitochondria encounter structure". In: *Nature*, pp. 1–5.

Young, Lindsey N and Elizabeth Villa (2023). "Bringing Structure to Cell Biology with Cryo-Electron Tomography". In: *Annual Review of Biophysics* 52, pp. 573–595.

Zhou, Li, Chao Yang, Weiguo Gao, Talita Perciano, Karen M Davies, and Nicholas K Sauter (2023). "A machine learning pipeline for membrane segmentation of cryo-electron tomograms". In: *Journal of Computational Science* 66, p. 101904.