Supplementary Material for "MiShape: 3D Mitochondrial Shape Modelling for Fluorescence Microscopy"

Abhinanda R. Punnakkal[1], Suyog S. Jadhav[1], Biswajoy Ghosh[2], Alexander Horsch[1], Krishna Agarwal[2], Dilip K. Prasad[1]

[1]Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway
[2]Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway
*Corresponding author: *abhinanda.r.punnakkal@uit.no*

Abstract

In this supplementary document, we first provide details of the microscope parameters used in our various experiments. We then give details of the architecture, baselines, and metrics for the unconditional generative models and reconstruction models. Finally, we describe and provide the results of the experiment, Microscope-to-microscope transformation.

Microscope Parameters - simulated and experimental

Experiments	**Conf2**	**Epi1**	**Conf1**
Image to shape reconstruction (Figure 2, main paper)	X		
Stack to shape (Figure 3, main paper)	×		
Generalization (Figure 4, main paper)			
Topology Conformation (Figure 5, main paper)	x		
Reconstruction on real data (Figure 6, main paper)		x	
Microscope-to-microscope (sec. *1*, supplementary)		x	x
Type of microscope	Confocal	Epifluorescence	Confocal
Emission wavelength	488 nm	688 nm	600 nm
Pixel Size	48 nm	109 nm	70 nm
Numerical Aperture	1.4	1.42	1.4
Magnification	63	60	63
Optical Resolution (nm)	124	245	152
Optical Resolution (pixels)	2.71	2.22	2.16

Table 1: Microscope Settings used for various experiments.

We have simulated one epifluorescence microscope and one confocal microscope image stack in this paper, in addition to what is available in 3DMSL[1]. Table 1 lists the optical parameters of these microscopes. It also includes the sections in the main manuscript in which these microscopes were used. The publicly available data creation protocol provided in 3DMSL was used for simulating the fluorescence microscopy images and stacks. Noise was added to the simulated microscopy images to emulate a signal-to-background ratio between 2 and 4. For this, the value of background b was set around 100.

Architecture, implementation details and additional results for- Unconditional Generation, Image-to-shape, and Stack-to-shape Reconstruction

We outline the general architecture of the MiShape decoder in Figure 1 (b) (main paper), with the base architecture similar to occupancy networks[2]. The encoder-decoder modules form a VAE-like construction. All the models take the output $c \in \mathbb{R}^C$ of the encoder network that differs for different experiments, and a batch of T input 3D points.

Dataset and Splits

The dataset used for training the different models in the experiments in the manuscript is obtained from 3DMSL[1] (3D Mitochondria Shape Library). 3DMSL contains >27k instances of diverse mitochondria shapes in different 3D shape representation formats of meshes, point clouds and implicit shape. These are derived from the individual components extracted from the segmentation masks available from EMPIAR-10791[3].The 3D instances are at a resolution of 24 nm. 27272 individual shape instances are split as 70-10-20 percent of the whole between the training, validation, and test subsets.

Unconditional generative model

Architecture

The architecture of the encoder for this model is outlined in Figure 1. The occupancy encoder receives the batches of query points p and their respective occupancies as input. The encoder consists of a positional encoding layer followed by 2 fully connected layers encoding the input points and their respective occupancies. This is followed by a stack of two fully-connected (FC) layers and max pooling layers, concatenating their outputs to the next layer. The output of this stack is then fed to an FC and max pool layer. Finally, the output is then fed to two FC layers to produce the mean and log of the standard deviation of the 128-dimensional latent code z. The decoder with the architecture described in Figure 1 (b) (main paper), predicts the occupancies of the query points.

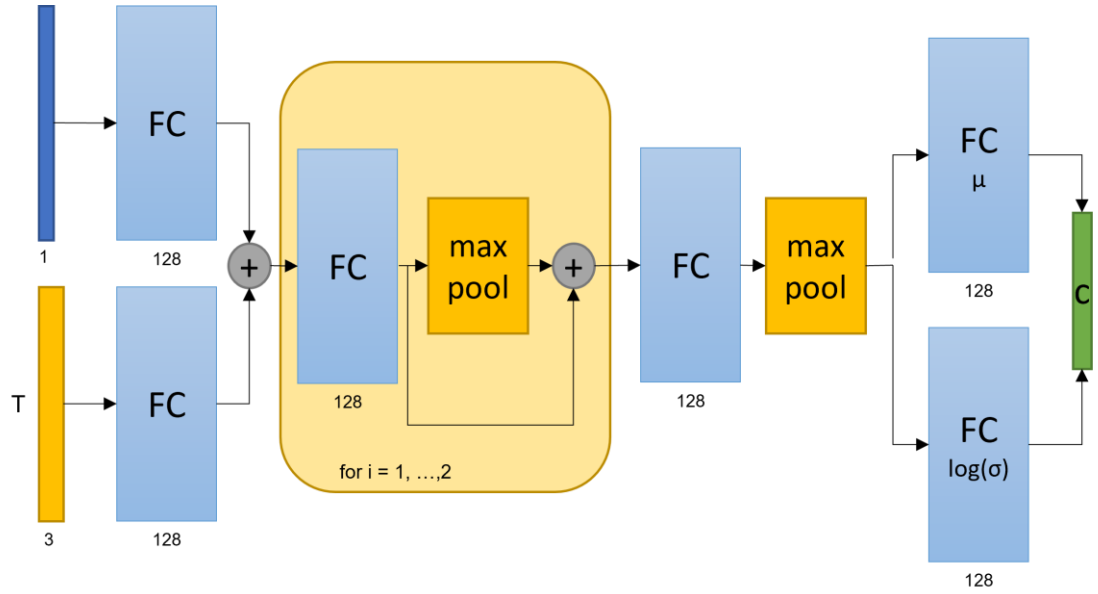

Figure 1: Encoder for the unconditional generative model used to learn the implicit shapes of Mitochondria, adapted from occupancy networks.

Expert Qualitative Evaluation

The confusion matrix in Figure 1(e) in the main paper summarizes the results of the experiment where 10 domain experts were shown pictures of real and synthetic mitochondria shapes and asked to predict if they are 'real' or 'synthetically generated'. These experts are working with biological studies of mitochondria and are familiar with the different shapes that mitochondria exhibit. We see that irrespective of whether they are presented with real or generated 3D shapes, they predict the shape to be real more than 50% times, but only barely more than that. This indicates that they cannot tell the real and generated shapes apart from each other. The results of the experiment indicate that generated structures are similar to the real ones, at least in the sample size of 100 real and 100 generated images, and 10 experts. Potentially a crowdsourced, larger study may reveal other insights, but this is relegated to the future.

Image-to-shape Reconstruction model

The overall architecture for the Image-to-shape reconstruction model follows the general decoder architecture, with the basic ResNet block in the decoder is comprised of Con1D layers. In this mode, we use the simulated microscope image of the input shape to condition the decoder. The encoder architecture of image-to-shape reconstruction models, which is a based on ResNet architecture, is outlined in Figure 2.

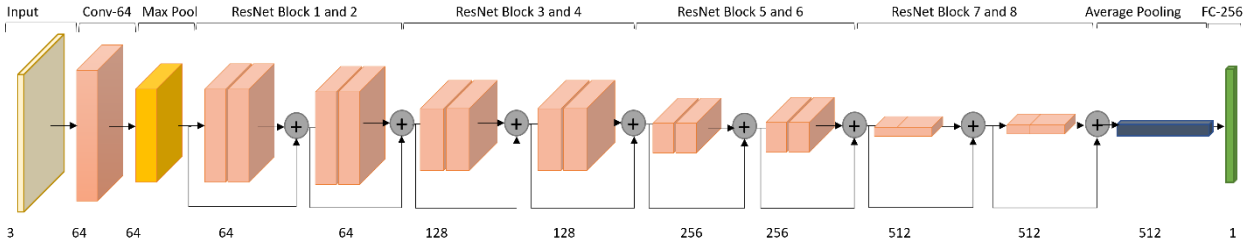

Figure 2: Encoder architecture for the image-conditioned model used for image-to-shape experiments.

3D Reconstruction Metrics

The metric used for the evaluation of the image-to-shape and the stack-to-shape models are volumetric IoU, Chamfer-L_1 distance, and normal consistency as defined in occupancy networks[2]. To obtain resolution-independent estimates of volumetric IoU, the intersection over union is per on randomly sampled 100k points from the bounding volume and determining if the points lie inside or outside the ground truth / predicted mesh. If M_{pred} and M_{gt} are the set of all points that are inside or on the mesh of the predicted and the ground truth mesh, the volumetric IoU is defined as,

$$IoU(M_{pred}, M_{gt}) \equiv \frac{|M_{pred} \cap M_{gt}|}{|M_{pred} \cup M_{gt}|}$$

The Chamfer-L_1 distance is defined as the mean of accuracy and completeness measures, where accuracy is the mean distance of the predicted mesh to their nearest neighbors in the ground truth mesh and completeness is the mean distance of the ground truth mesh to their nearest neighbors in the predicted mesh. This is written as,

$$\text{Chamfer-}L_1\ (M_{pred}, M_{gt}) \equiv \frac{1}{2|\partial M_{gt}|}\int_{\partial M_{gt}} \min_{q \in M_{pred}} ||p-q||\ dq + \frac{1}{2|\partial M_{pred}|}\int_{\partial M_{pred}} \min_{q \in M_{gt}} ||p-q||\ dp$$

By this definition, lower values for Chamfer-L_1 are better. Finally, the normal consistency score is used to measure how well the higher-order information is captured in the predicted mesh. It is defined as the absolute dot product of the normals in one mesh and the corresponding nearest neighbors in the other. If $n(p)$ and $n(q)$ denote the unit normal vectors on the mesh surface M_{pred} and M_{gt} respectively, normal consistency is written as,

$$\text{Normal-Consistency}(M_{pred}, M_{gt}) \equiv \frac{1}{2|\partial M_{gt}|}\int_{\partial M_{gt}} |\langle n(proj_1(q)), n(q)\rangle| dq + \frac{1}{2|\partial M_{pred}|}\int_{\partial M_{pred}} |\langle n(p), n(proj_2(p))\rangle| dp$$

where $<.,.>$ denotes the inner product, and $n(proj_2(p))$ and $n(proj_1(q))$ denote projections of p and q onto M_{pred} and M_{gt} respectively. Consequently, higher volumetric IoU and normal-consistency scores are better and lower values of Chamfer-L_1 distance is better.

Topological Metric

We refer to the topological loss defined as Scalar Function Topology Divergence (SFTD)[4] for comparing the topology of two 3D objects. We compare two shapes by first converting them into scalar fields within the same domain. For our meshes, we use voxelization on a shared, axis-aligned grid to create two 3D voxelized occupancy grids representing the reconstructed and ground-truth meshes. The two binary occupancy grids (F and G, each shaped $n_x \times n_y \times n_z$) are embedded into a 3-channel volume D, where; D[0] = min(F, G): the intersection region, D[1] = F: the reconstructed occupancy field and D[2] is a global minimum value replicated over the grid that serves as a baseline reference level. A cubical complex is constructed from this 3D volumetric structure using GUDHI[5], and persistent homology is computed over this grid-based cell complex. For each homology dimension of interest (H0 and H1: H0 for connected components, H1 for loops/holes), the algorithm extracts persistence pairs (birth, death) from the cubical complex. The loss is then computed as the sum of the top-k persistence values, raised to the power p (we use p=1). Thus, this produces a single scalar that grows with the magnitude and number of unmatched topological features. This captures the strongest topological features and their stability, effectively penalizing topological discrepancies while remaining computationally efficient and differentiable.

Baseline-DISPR

DISPR is a diffusion-based generative model that tries to solve the problem of estimating 3D shape from 2D microscopy images. The authors utilize a diffusion approach inspired by an improved denoising diffusion model[6], with an autoencoder architecture adapted from SHAPR[7]. During training, noise is successively added to the 3D ground truth, and the 2D input image is concatenated separately at each step. The decoder takes this noisy 3D image and tries to reconstruct the ground truth by successively denoising it. During inference, the encoder is unused, and a random 3D noisy image is concatenated with the 2D input image and 2D mask. This is used as the input to the decoder. After successively denoising 1000 steps, the 3D prediction is obtained.

For comparison with this method, we utilized the public code published by the authors to train DISPR on a total of 9443 training images. The model was trained with the parameters outlined in the paper, and the results were evaluated using a test dataset containing 2698 images. The surface is then extracted using the same isosurface extraction algorithm used earlier. The dataset used for this baseline is a subset of the dataset used for training our models since larger mitochondria had to be excluded for confining to the data size. The voxel resolution used for the data is 24nm.

Baseline-ImageJ

We employed the standard reconstruction workflow commonly used by biologists in Fiji (ImageJ) [8] to derive shapes from image stacks. The procedure applies basic threshold-based 3D segmentation[9], followed by the marching-cubes algorithm to generate surface reconstructions

Baseline Nelli

Nellie [10] is a GUI-based pipeline that combines adaptive multiscale image enhancement with hierarchical segmentation and marker-based tracking to extract multilevel spatial and temporal features from 2D/3D live-cell microscopy for different organelles. First, images are preprocessed with a voxel-aware, multiscale structural filter (a modified Frangi) and thresholder to produce semantic masks; instances are then skeletonized and decomposed into branches and nodes to form a hierarchical representation (voxels to nodes to branches to organelles). Nellie computes and exports a large suite of hierarchical

morphology and motility features, supports downstream graph construction and representation learning, and is distributed as a Napari plugin.

Nellie receives the input of the pixel size of the input image to adapt the settings of the filters. We use the segmentation module of Nellie on a 3-stack input followed by marching cubes for mesh reconstruction from 3D voxel segmentation.

Stack-to-shape Reconstruction model

The stack-to-image model follows the same basic architecture as the image-to-shape reconstruction model but the encoder is a simple 3D voxel encoder adapted from occupancy network in place of a 2D-ResNet used for image-to-shape reconstruction.

Microscope-to-microscope transformation

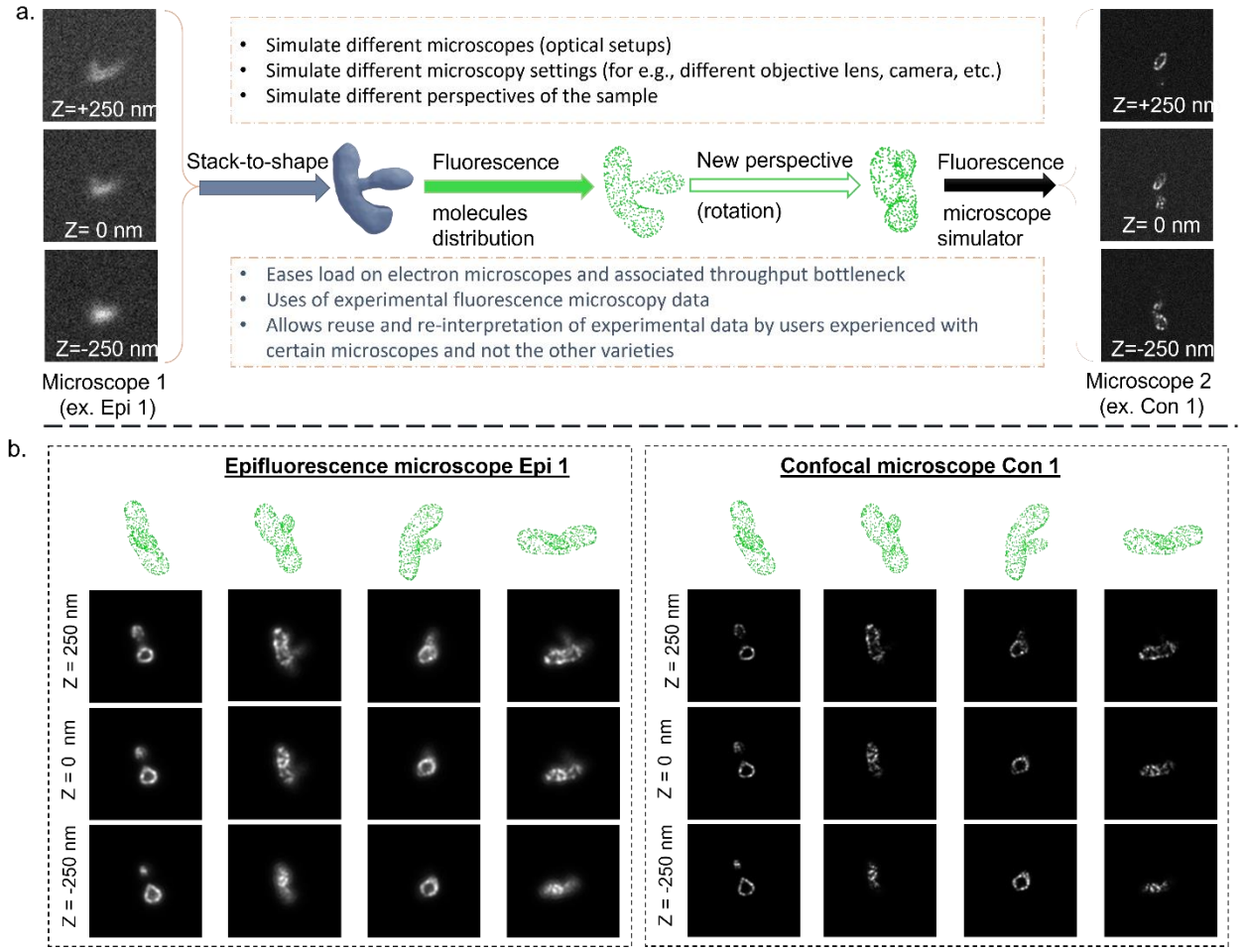

Figure 3: Microscope-to-microscope transformation (a) The concept of microscope-to-microscope transformation is illustrated using an example. (b) Microscope-to-microscope transformation results for the shape used in (a) Simulated (noise-free) images created for two different microscopes using different perspectives of the shapes learnt by the stack-to-shape model are presented.

Here, we consider the problem of using 2D fluorescence images or z-stack generated by one fluorescence microscope to simulate the microscopy images or z-stack corresponding to another fluorescence microscope.

We presented the methodology to derive 3D shapes from 2D fluorescence images or fluorescence z-stack in the main paper. For the purpose of fluorescence microscope-to-microscope translation, we propose to first, use the fluorescence image or z-stack from the source microscope to construct the corresponding 3D shape. Then, we scale the 3D shape to the actual physical units. The scale factor can be computed through a minimization approach if it is unknown. Here, we assume that the scale factor is known since we intend to illustrate the concept and feasibility. Having brought the implicit shape into the relevant physical dimensions, it may be used as it is or subject to rotation, translation, or creation of collage of many such structures. Then the process of generating the fluorescence molecule distribution and the fluorescence images or z-stacks presented in 3DMSL are applied to generate the images. Figure 3(a) outlines the steps and Figure 3(b) shows the results of this experiment.

Such a transformation is useful to avoid using simulation-only datasets for AI and incorporating experimental evidence without resorting to low throughput and expertise-demanding EM imaging. Moreover, experiments in biology are conducted over timelines of months to years. In this span, it often happens that imaging has to be done on different microscopes considering the simultaneous availability of microscopes and biological samples. This often prohibits one-to-one comparison of biological images related to the same experiment. This problem can be solved through the proposed approach. Similarly, it can help in situations where the imaging was done in one microscopy setting (for example, one objective lens) but it is desirable to study the image in another setting.

Further, it can be used to support the reinterpretation of the data by a biologist more used to one kind of microscope, which is different from the microscope from which the data was acquired. This is useful because the biologists learning to interpret data from one microscope also learn the limitations of the microscope and how these should be incorporated in interpretation and deriving inferences about biological questions. However, this abstract heuristic knowledge is not easily transferable across microscopes. Also, it can reduce the dependence on relatively lower throughput, costly, and heavily used confocal microscopes by forming confocal image equivalents using cheaper and high-throughput epifluorescence microscopes.

References

1. Punnakkal, A. R. *et al.* 3DMSL - 3D Mitochondria Shape Library for Optical Microscopy. doi:doi:10.18710/JX6JXF.

2. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S. & Geiger, A. Occupancy Networks: Learning 3D Reconstruction in Function Space. in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (2019).

3. Parlakgül, G. *et al.* Regulation of liver subcellular architecture controls metabolic homeostasis. *Nat. 2022 6037902* **603**, 736–742 (2022).

4. Trofimov, I., Voronkova, D., Tulchinskii, E., Burnaev, E. & Barannikov, S. Scalar Function Topology Divergence: Comparing Topology of 3D Objects. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **15146 LNCS**, 274–290 (2024).

5. The GUDHI Project. *GUDHI User and Reference Manual*. (GUDHI Editorial Board, 2026).

6. Nichol, A. & Dhariwal, P. Improved Denoising Diffusion Probabilistic Models. *Proc. Mach. Learn. Res.* **139**, 8162–8171 (2021).

7. Waibel, D. J. E. *et al.* SHAPR predicts 3D cell shapes from 2D microscopic images. *iScience* **25**, 105298 (2022).

8. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods 2012 97* **9**, 676–682 (2012).

9. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).

10. Lefebvre, A. E. Y. T. *et al.* Nellie: automated organelle segmentation, tracking and hierarchical feature extraction in 2D/3D live-cell microscopy. *Nat. Methods 2025 224* **22**, 751–763 (2025).