Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

MiShape: 3D Mitochondrial Shape Modelling for Fluorescence Microscopy

Abhinanda Punnakkal

abhinanda.r.punnakkal@uit.no

UiT The Arctic University of Norway https://orcid.org/0009-0006-8135-9606

Suyog Jadhav

UiT The Arctic University of Norway

Biswajoy Ghosh

UiT The Arctic University of Norway https://orcid.org/0000-0003-1309-1661

Alexander Horsch

UiT The Arctic University of Norway

Krishna Agarwal

UiT The Arctic University of Norway https://orcid.org/0000-0001-6968-578X

Dilip K. Prasad

Bio-AI Lab, Department of Computer Science UiT The Arctic University of Norway, Tromso

Article

Keywords:

Posted Date: July 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10325775/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

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

The problem of accurate reconstruction of the mitochondrial shape is important for understanding the life-sustaining mechanisms at the cellular level that are indicative of many illnesses and poor health conditions. Current approaches are restricted to the problem of 2D segmentation of fluorescent images and do not address the inherent 3D structure. We present MiShape, a 3D shape prior for modeling mitochondrial shapes in fluorescence microscopy images. MiShape is representative of the real distribution of mitochondrial shapes as it is trained using shapes sourced from high-resolution Electron Microscopy images. MiShape uses an implicit representation to model shapes, offering a versatile solution for accurately representing the diverse shapes of sub-cellular structures. To our knowledge, MiShape is the first framework to use a learnt 3D mitochondrial shape prior to reconstruct 3D mitochondrial morphology from a single 2D fluorescence microscopy image. Across simulated, external, and real microscopy datasets, MiShape outperforms conventional and state-of-the-art baselines in terms of geometry-based and topological reconstruction metrics.

Introduction

Mitochondria are cell organelles responsible for producing most of the chemical energy required for the cell. Their morphology, number, and size vary greatly between different cell types and organisms, and within a single cell[1,2]. Mitochondrial morphological differences are correlated to several diseases such as Alzheimer's, Parkinson's, Huntington's diseases[3] and cardiovascular diseases[4] among many others[5]. The fluorescent optical microscope is the most commonly used instrument for observing mitochondrial morphology. In addition to capturing 2D slices of 3D structures, these instruments are limited by their spatial resolution, which is comparable to the sizes of mitochondria. Consequently, most of the work so far has reduced the problem of modeling the shape of mitochondria to that of 2D segmentation from fluorescent microscopes[6,7].

A comprehensive and current list of mitochondria analysis tools for fluorescence microscopy images is provided in Chue et.al[8], which indicates the significance of mitochondrial analysis. Current techniques for analyzing mitochondrial morphology are based on image processing[6,9–11] for segmentation, or mitochondria are often simplified as tubular structures[6,7,12] for network analysis. Deep learning-based segmentation[7,13–15] suffers from a lack of labeled data[7,16]. Simulation supervision[7,18] and semi-supervision[19] have emerged as a workaround to this problem; however, no holistic or universal model for modeling the 3D shape of the subcellular structure exists. Current methods are useful for studying networks of mitochondria, but not the nanoscale 3D morphology, which is vital to identify dynamics and gain insight into behavioral changes. Mitochondria undergo processes like fission, fusion, etc.,

which are markers used to infer mitochondrial health[20–22]. However, detecting potential sites of mitochondrial dynamics demands accurate 3D shape reconstruction. We are motivated by recent advances in generative modeling that aim to provide universal models for a particular problem domain.

Driven by the need for a shape model for mitochondria that extends to 3D shapes and does not constrain quantitative analysis to 2D, we propose MiShape: 3D Mitochondrial Shape modelling for fluorescence microscopy. MiShape is a generative shape prior trained on a large dataset of high-resolution data that is representative of the real diversity of mitochondrial shapes. MiShape uses an implicit representation[23–26] that provides a computationally efficient model as well as scaling of the shapes without the loss of details. Implicit representation models surfaces as a boundary decision of a volume, allowing a lightweight parameterization for large volumes of data. Central to MiShape is the Variational Auto Encoder (VAE) based architecture designed to work on implicit shape representation. MiShape learns a distribution of mitochondrial shapes from 3DMSL[27], a multi-modal dataset containing 27K instances of 3D shapes of mitochondria sourced from high-resolution Electron Microscopy (EM) data[28], that acts as a high-quality data source for ground truth shapes of organelles with paired fluorescent microscope images. By training a generative model using implicit representations with high-quality ground truth, we create a shape prior that retains the realistic diversity and complex details of mitochondrial shapes.

We observe that MiShape can capture the diversity and complexity exhibited by mitochondrial shapes because it learns from an extensive high-resolution EM source. MiShape generates different classes of mitochondria identified in the literature, like dots, rods, networks, donuts, and vases, etc. The shapes generated by MiShape are empirically closer to ground-truth shapes than those created by previous methods[7] . Furthermore, MiShape shows promising 3D shape predictions for the problem of 3D shape reconstruction from a single fluorescence image, with performance increasing with the increasing number of input image stacks. We evaluated the 3D shape predictions for single-image input with conventionally used solutions and observed superior performance in MiShape, both quantitatively and qualitatively. Further experiments demonstrate that MiShape can be generalized to other datasets and cell types and can be used to infer shapes in real datasets, too. Overall, we believe that MiShape is a significant advancement in the extraction of accurate 3D mitochondrial shapes from fluorescence microscopy images, providing a route towards the 3D shape analysis of subcellular organelles.

Results

MiShape

We present the general workflow of training MiShape in Figure 1 (a). MiShape is trained on 3DMSL[27,29], which provides ground truth data for generating simulated fluorescence images using a computational microscope simulator for 27K mitochondria instances. MiShape can be trained in both conditional and unconditional ways, the conditioning being a simulated microscope image/stack, depending on the application. Figure 1 (a) depicts the workflow of a conditioned MiShape model used for 3D reconstruction experiments. The basic architecture of MiShape is a VAE composed of a set of residual blocks adapted to handling implicit shape representation. MiShape takes sampled 3D query points, together with a latent shape code or image/stack-derived conditioning, as input and predicts the occupancy probability of each point as output, thereby defining the mitochondrial surface as a continuous decision boundary. Figure 1 (b) shows the architecture of the decoder used in MiShape. All versions of MiShape models are trained using the training subset of the 3DMSL[27] dataset and evaluated using the test set of the 3DMSL dataset.

Mitochondria exhibit different phenotypes, such as compact dots, elongated rods, donuts, branched networks, etc., and the different phenotypes have been observed to have functional relevance[30–33].

We train MiShape without any conditioning to determine its modelling capacity and examine the diversity of shapes generated. Figure 1 (c) shows qualitative results in comparison to Shekh et.al[7] for mitochondrial shape generation . 3D rendering of MiShape-generated shapes are provided in supplementary GIF 1 . Shekh et al.[7] use a parametric representation of tubular structures created using cylindrical curves around B-spline curves. It is observed that the shapes generated by MiShape are more complex and varied than the state-of-the-art, Shekh et al.[7]. We also verified that our structures visually match the different phenotypes reported in Miyazo et al.[34]. In Figure 1 (d), we compare common shape metrics used to analyze and compare geometric properties of shapes. We compare these metrics for 3000 shapes generated using MiShape and 3000 shapes using Shekh et al. Although not statistically significant, we see that MiShape metrics are closer to the ground truth shape than Shekh et.al in all four metrics of aspect ratio, sphericity, elongation, and ellipticity.

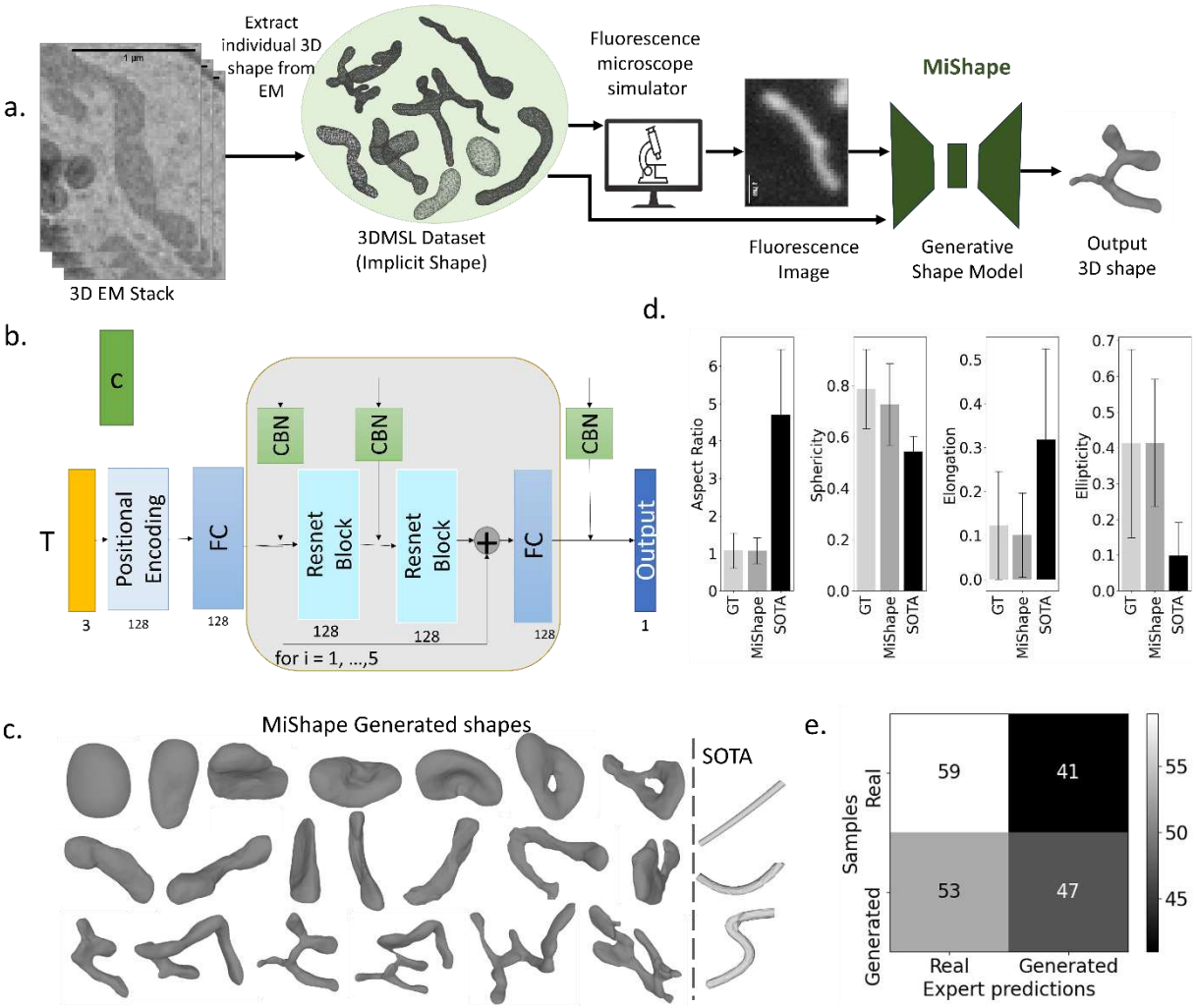


Figure 1: MiShape:3D Mitochondrial Shape Modelling for Fluorescence Microscopy. (a) Overview of MiShape. 3D shapes extracted from high-resolution EM images are used as ground truth shapes, which are then used to create simulated fluorescence images. MiShape is trained with pairs of 3D ground truth shapes and simulated microscope images to learn a generative shape prior of 3D mitochondrial shapes. (b) Model architecture: MiShape is trained on implicit shape representations. The decoder architecture consists of a positional encoding module followed by five fully connected ResNet blocks, where the basic block is a FC. The final layer of the decoder is a fully connected layer of output size, one. The decoder predicts the occupancies of the query points. (c) MiShape-generated shapes present greater diversity and complexity than SOTA[7], they include different topologies of mitochondria identified in the literature: dots, rods, donuts, and networks. (d) Shapes generated by MiShape have shape parameters of aspect ratio, sphericity, elongation, and ellipticity closer to the ground truth than SOTA[7]. (e) Results of qualitative expert evaluation of realism of 3D shapes by comparing 100 instances of real and generated mitochondria by experts.

We also validate that our generator is effective in modeling and generating realistic mitochondrial shapes via aggregating qualitative expert evaluations. Expert evaluators who are already familiar with mitochondrial shapes are shown gifs of revolving 3D mitochondrial shapes and are asked to answer if they think the shapes are real or generated. The result of the study is presented in Figure 1 (e). 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% of the time, but only barely more than that. This suggests that they cannot tell the real and generated shapes apart from each other.

MiShape can generate 3D Shapes from a single fluorescence image

Generating 3D shapes from single images is not new in computer visions[23,35,36], but is less explored in the microscopy domain[37,38], and to the best of our knowledge, non-existent for nanoscopic structures like mitochondria. Here, we address this highly challenging problem of generating 3D shapes from a single 2D fluorescence microscopy image. Conditioning individual mitochondria shapes with their counterpart fluorescence microscopic image or z-stack permits learning the shape correspondence between microscope images and the corresponding 3D shapes. As shown in Figure 2(a), the Point Spread Function (PSF) of a microscope has a 3D nature that is generally rotationally symmetric in the lateral x-y plane, but presents z-dependent blurring to the sample. Therefore, even when using a single 2D fluorescence image, the 3D context information of the sample within the depth of field (DOF) of the microscope is gathered, along with a form of z-position encoding through the PSF. Consequently, the 2D image to 3D shape reconstruction problem is not as ill-conditioned as it appears at first sight. Nevertheless, the 3D shape beyond the DOF does not get imaged and therefore the reconstructed shapes are only accurate within the DOF. We therefore leverage the 3D context embedded in the PSF and the prior knowledge learnt by MiShape as ways to constrain the predicted shapes to plausible shapes matching a given input image.

We train a conditioned MiShape model with pairs of 3D shape and their corresponding fluorescent image available in 3DMSL[27]. We use the images with the Conf2 microscope configuration from 3DMSL (configurations provided in the supplementary text). We compare our model with the typical 3D analysis and reconstruction pipeline followed by biologists, i.e., Fiji (ImageJ)[39], the state-of-the-art computational method Nellie[40], and also a deep learning-based state-of-the-art in the task of single image 3D reconstruction in microscopy, DISPR[38]. The ImageJ pipeline is a combination of simple thresholding-based 3D segmentation followed by marching cubes for reconstruction. Nellie is an automated pipeline for segmentation, tracking, and feature extraction of diverse intracellular structures. For segmentation, Nellie first applies a multiscale, structure-enhancing preprocessing pipeline to boost organelle contrast across sizes and morphologies and then performs semantic segmentation on the preprocessed image using a thresholding scheme to obtain a foreground organelle mask. DISPR is a denoising diffusion method-based generative model trained to predict 3D shapes. It is trained with pairs of 2d microscope images of cell nuclei and their 3D ground truth with a voxel grid size of 64x64x64. We train and test on a subset of our dataset (9443 training samples and 1207 testing samples) to adapt the architecture to the microscopic resolution of mitochondria, maintaining the ratio of resolutions of the original method. It is to be noted that both the non-deep-learning baseline of ImageJ and Nellie cannot generate 3D shape from a single image. Therefore, in our comparisons, they use a 3-image stack as input, as opposed to MiShape, which uses a single image input. DISPR requires an image and its 2D segmentation mask as input to predict 3D shape at inference.

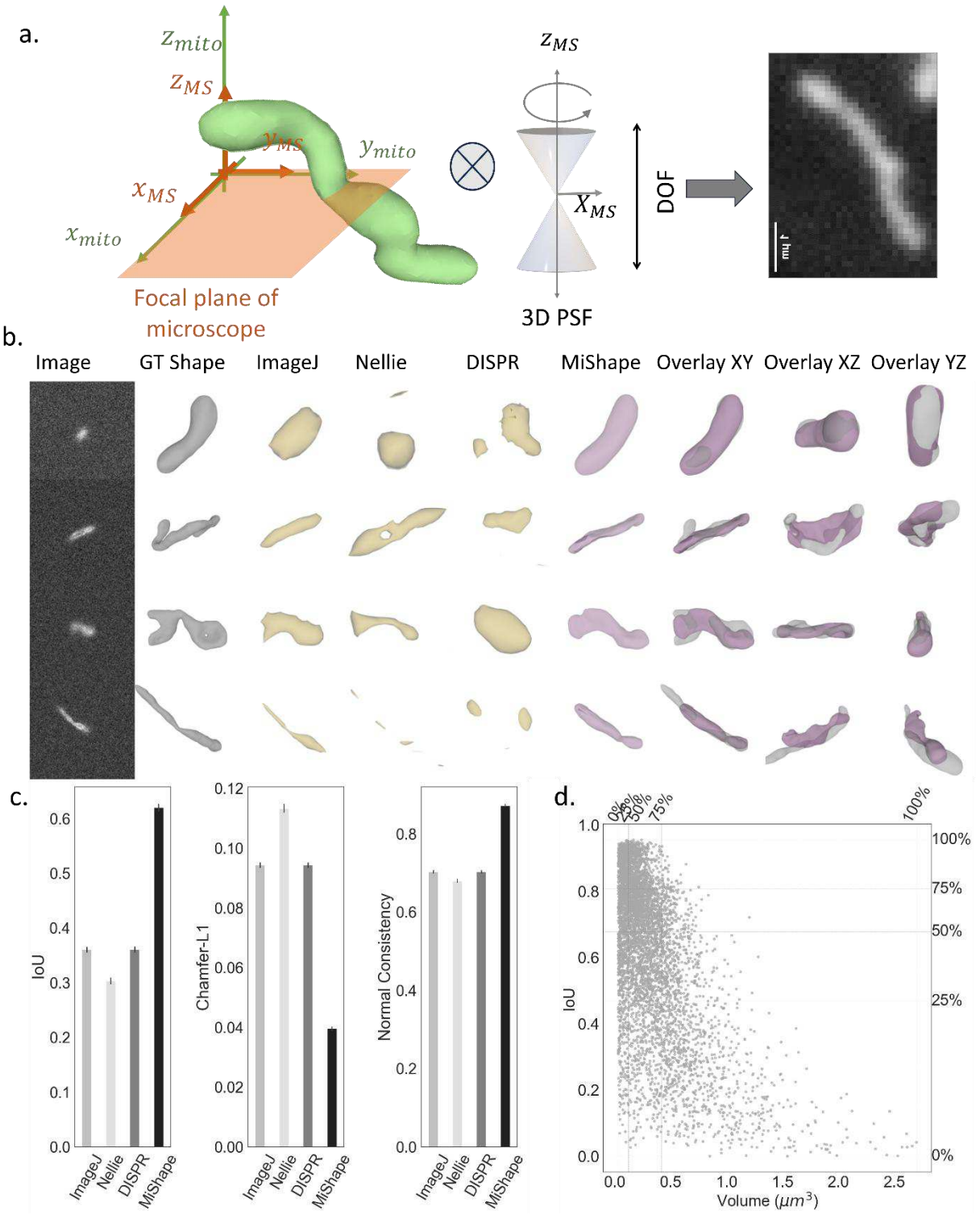


Figure 2: Image to shape reconstruction. (a) 3D context encoded in 2D image as PSF is a z-dependent function – it encodes blurred details of out-of-focus portions of the 3D object into the 2D image (b) Examples of 3D shapes generated by MiShape and other baselines, given a single 2D confocal microscope image as input. (c) Quantitative comparison of the performance of 3D shape reconstruction, measured by metrics of volumetric IoU, Chamfer L1 distance, and normal consistency. All metrics indicate superior performance of MiShape. (d) Variation of the quantitative metric with the size of the ground truth shape indicates that larger structures produce more errors.

Qualitative comparisons of MiShape with the chosen baselines, along with the ground-truth shape of the image, are presented in Figure 2 (b), and 3D renderings are provided in supplementary GIFs 2-7. Ground truth shapes are shown in gray, MiShape results in plum, and the baselines in yellow. The

results demonstrate that MiShape is able to generate shapes closer to the ground-truth shapes than other baselines. Nellie produces thinner structures, often leading to holes (row 2) or broken structures (row 4). ImageJ results look visually closer to the ground truth than other baselines. DISPR predictions often contain multiple components (sometimes generate pure noise too) and are comparatively dissimilar to GT. MiShape predicts a good estimate of the ground-truth shape for simpler structures (row 1), but the error increases as the size of the mitochondria increases and is out of the DOF (row 2, 4) of the image. We see that the error also increases with increasing complexity of the shape (row 2). To compare the different methods quantitatively, given in Figure 2 (c), we use the metrics volumetric IoU, Chamfer L1 distance, and normal consistency (metrics are defined in the supplementary text), commonly used to evaluate 3D predictions[23]. Higher volumetric IoU and normal-consistency scores are better, and lower values of Chamfer-L1 distance is better. These results also indicate that MiShape has superior performance to all the baselines in all three metrics, despite the two methods of ImageJ and Nellie having more informative inputs of a 3-stack input image compared to a single image input for MiShape. In Figure 2 (d), we show how the reconstruction metric of volumetric IoU varies with the increasing size of the ground truth shape. This shows an inverse relationship between the size of the mitochondria and volumetric IoU. The plot also shows the unbalanced nature of the dataset, with many mitochondria of a smaller size. For very large shapes of sizes from 1.3 μm^3 – 2.7 μm^3, the IoU is not greater than 0.5, which is not surprising, as the (DOF) does not span more than 500nm in a fluorescence microscope, which leads to only a portion of the mitochondrion being captured by the input image. Qualitative examples of larger structures are provided in rows 2 and 4 which have volumes of 1.6 μm^3 and 1.3 μm^3, respectively.

MiShape reconstruction improves with increasing input stack size

The use of additional slices of the z-stack presents better conditioning than the use of a single 2D image as it allows spanning a larger axial extent of the sample, providing more information than a single image. Therefore, we train three separate models with an increasing number of input z-stack images as conditioning. We refer to these models as 3-Stack, 5-Stack, and 7-Stack models, named based on the number of input images in the stack used for training. We follow the protocol described in 3DMSL to create these stacks. The z-slices are 0.14 µm apart, with each image having a DOF of 0.754 µm[41]. With this spacing and DOF, any point near the center can be captured in up to five slices (DOF/spacing ≈ 0.754/0.14 ≈ 5.4; with symmetric DOF, at most 5 slices contain the point). Thus, moving from 1 → 3 → 5 slices increases both axial coverage and the number of in-focus observations per point; moving from 5 → 7 slices mainly extends the axial span with limited additional in-focus redundancy at the center. By allowing the central region of the shape to be sampled multiple times from the overlapping DOFs in the input image stack, we expect improvements in reconstruction quality in this region. We restrict the experiment to 7 stack models due to limitations of the computational model used for simulation[42].

In Figure 3, we show the results of our experiments on increasing z-stack image input. The qualitative result shown in Figure 3(a) shows the ground truth in gray and the MiShape results for single image, 3-Stack, 5-Stack, and 7-Stack models (3D renderings are provided in supplementary GIFs 8-11). We observe that with an increasing number of stack-image inputs, both the overall conformity of the shape to the ground truth shape and also more detailed structures are present in the predictions (top row: protrusion present in the result for 7-stack model, middle row: new branch reconstruction in the 7-stack model result). The performance improvement is also reflected by the positive trends in all three quantitative metrics of volumetric IoU, Chamfer-L1 distance, and normal consistency, as shown in Figure 3(b).

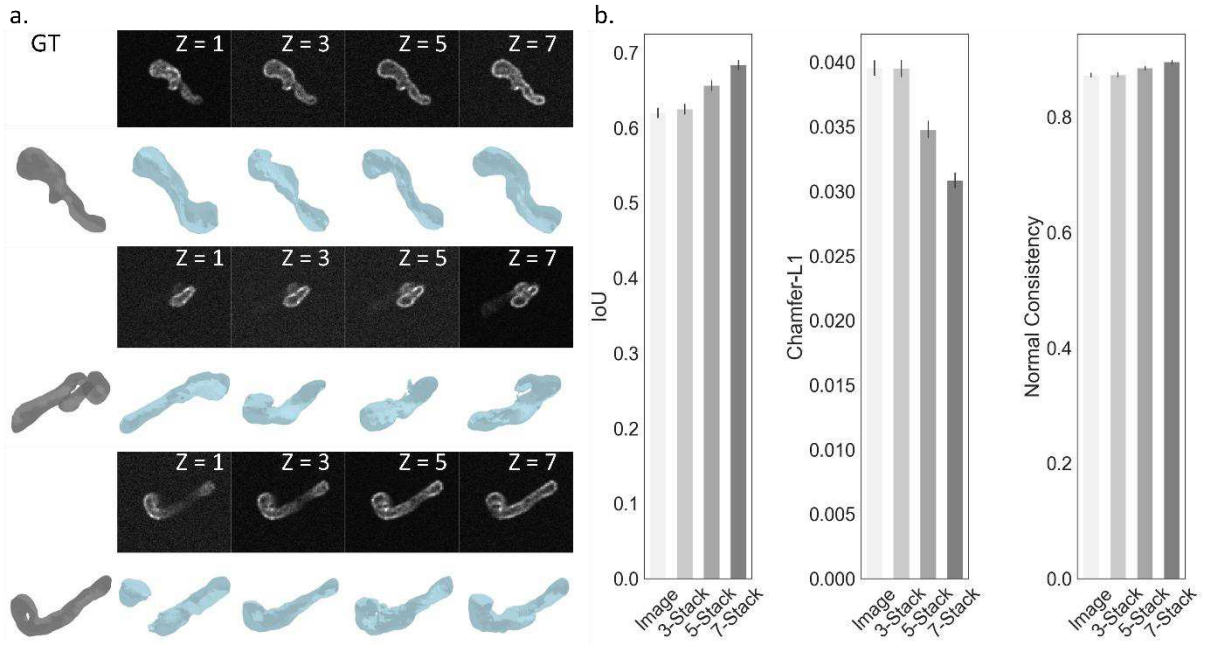


Figure 3: Reconstruction quality increases with the number of conditioning z-slices. (a) Three examples of 3D shapes generated by MiShape given stack of images. The leftmost column shows the 3D ground truth shape. The top rows show images in the stack, the bottom rows in blue, from left to right, show the predictions of MiShape models trained on 1, 3, 5, and 7-image stack input. (b) The quantitative comparison of MiShape models of varying input stack size.

MiShape generalizes to other datasets

After analyzing the performance of MiShape on the test set of 3DMSL data, we explore its generalization capability. For this, we compare MiShape's inferences with those of other baselines on a different dataset from the one it was trained on. Specifically, we create synthetic images following the 3DMSL protocol for the C. elegans dataset of the EMPIAR 10982 dataset[43]. The 123 test images are generated with the same microscope configuration, Epi1, as that of the model used for training. Therefore, this test dataset differs from the training data in terms of cell type and sample preparation protocol and thus is expected to have a different distribution of mitochondrial shapes than our training dataset. It is to be noted that the majority of the shapes in the chosen test dataset are highly spherical (dot-like shapes) and are not complex (network-shaped, with multiple branches) or elongated (rods). Despite this, we use this dataset to test the generalizability of MiShape due to the scarcity of other large EM datasets with available segmentation labels.

We show the generalization capabilities of the MiShape reconstruction model on the test set created using the EMPIAR 10982 dataset in Figure 4. We show qualitative results of the baseline models and MiShape in Figure 4(a) and 3D renderings are provided in supplementary GIFs 12-16. We can observe that while ImageJ and Nellie capture the basic shape of the test samples, we can observe that MiShape results better match the curvatures of the ground truths and provide more detailed coverage (reflected in the IoU metric in Figure 4(b)). Similar to the reconstruction results on the 3DMSL dataset, both the non-deep-learning baseline of ImageJ and Nellie generate 3D shape from a 3-image stack as input, whereas DISPR and MiShape use a single image. We also compare the different methods quantitatively and against performance on the training dataset 3DMSL in Figure 4(b) (2nd, 4th and 6th plots titled 10982), using volumetric IoU, Chamfer L1 distance, and normal consistency. When comparing metrics between 3DMSL and EMPIAR 10982 datasets, we observe an improvement in the performance of Nellie (but slightly worse metrics observed in ImageJ), which suggests that Nellie is better at recovering simpler shapes (like the ones in EMPIAR 10982). When looking into the deep-learning-based baseline,

DISPR and MiShape on the training dataset (3DMSL) and the new dataset (EMPIAR 10982), we see worse performance for both DISPR and MiShape. For MiShape, the volumetric IoU lowers from 0.62 to 0.55, Chamfer-L1 distance increases from 0.04 to 0.06, and normal consistency lowers from 0.89 to 0.85. This anticipated drop in performance is attributed to morphological shifts and differences in the sample preparation protocol between datasets. However, in the case of DISPR, this drop in performance is greater. DISPR was performing better than the non-deep-learning-based baselines of Nellie and ImageJ on 3DMSL, but on the EMPIAR-10982 dataset, it performed worse. However, despite the drop in performance on a new dataset, MiShape still performs better than all other methods (similar to results on 3DMSL). From these observations, we conclude that MiShape better generalizes than DISPR is a better candidate for 3D reconstruction than conventional methods on new datasets. We also note that since the EMPIAR-10982 set is dominated by near-spherical mitochondria, the performance on highly branched networks remains to be validated on an external dataset with such shapes.

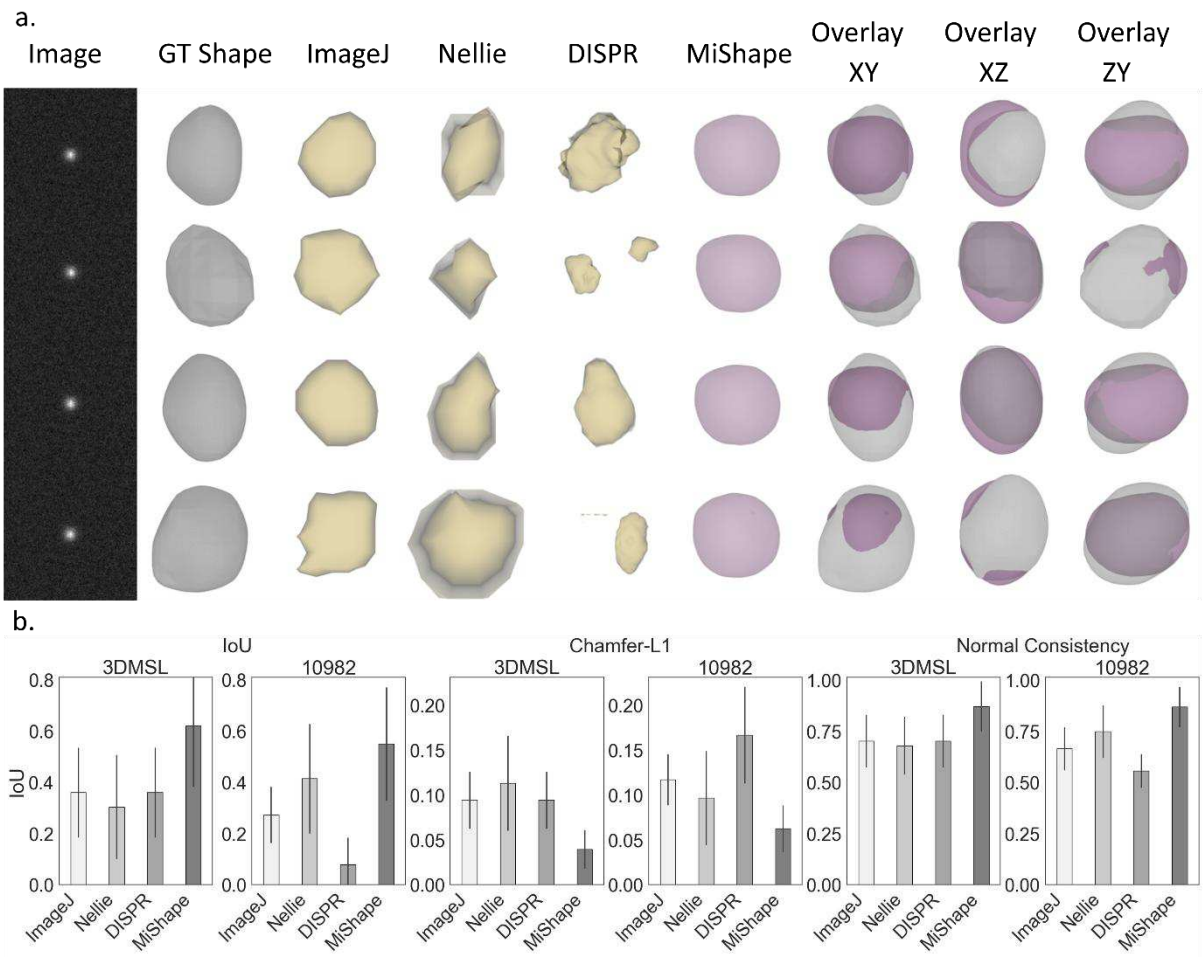


Figure 4: MiShape generalizes to other datasets. MiShape generalizes to other datasets. (a) Qualitative results comparing predictions of MiShape and other baselines for 3D reconstruction. (b) Quantitative metrics of IoU, Chamfer -L1 distance, and Normal Consistency, comparing the performance of MiShape with the baselines for the test set of the training data (3DMSL) and the new dataset used for checking zero-shot generalization (10982).

MiShape reconstruction reveals significant topological features

We consider MiShape's ability to preserve topology in reconstructed mitochondria and why this matters for interpretation. Mitochondrial morphology is highly diverse and dynamically regulated by motility, fusion, and fission; the physiological roles of specific arrangements remain incompletely understood. Beyond overall abundance, distinct phenotypes carry different functional and disease

associations. For example, altered branching has been reported in cancers, and toroidal ("donut") mitochondria have been observed in the brains of individuals with cognitive decline [31]. Consequently, accurate recovery of topology is essential for phenotype-level analysis.

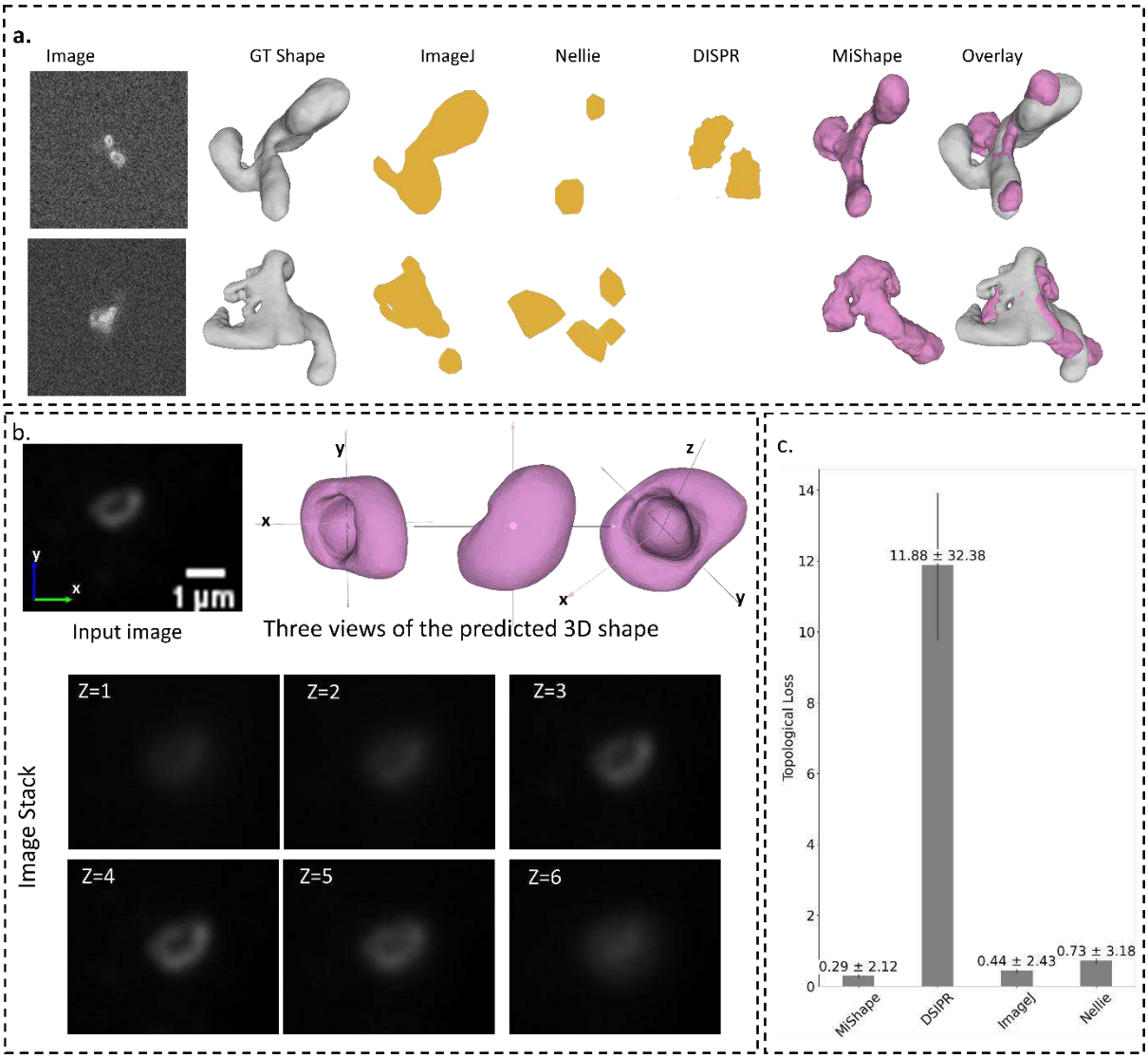


Figure 5: Topology correctness in MiShape predictions. (a) Qualitative results showing MiShape's ability to predict complex topology with multiple branches (top row) and holes in complex topology (bottom row) accurately, even though these are not evident from the input image. (b) MiShape predicts the bowl shape from a single input image, which has often been mistaken for a donut structure in microscope image analysis[44]. The vase shape of the mitochondrion is revealed upon inspecting the image stack (bottom row).

Figure 5 (a) illustrates MiShape reconstructions from a single input image for specimens with complex topology (3D renderings are provided as supplementary GIFs 17-18). In both the top and bottom examples, the mitochondria contain multiple branches that are not clearly visible in the input but are recovered by MiShape. In the bottom row of Figure 5 (a), MiShape also recovers an internal hole (toroidal structure) that is not captured by any baseline. These results indicate that MiShape better preserves global connectivity and local features from limited input, supporting more reliable phenotype-specific interpretations. We also include a quantitative comparison of the topology in Figure 5 (c). The topological loss used here is the Scalar Function Topology Divergence (SFTD)[45] loss (defined in the supplementary text), which is a measure of dissimilarity for topological features (number of components, loops, and holes) between two shapes; therefore, lower values are desirable. MiShape performs better than baseline methods. Nellie and DISPR tend to fragment the object or miss

branches, resulting in poorer topological fidelity. DISPR predicts very noisy (many small components) output for some inputs, which results in its high average topological loss.

We present an interesting result in Figure 5 (b) of MiShape on real microscope image (3D rendering included as supplementary GIF 19). It has been pointed out that mitochondria undergo mitochondrial fragmentation, resulting in structural changes from a tubular to a more rounded and swollen or ring/doughnut mitochondria)[1]. While it was believed that most mitochondria changed from a tubular to a ring shape, it was later discovered that most ring mitochondria did not have a through hole[1]. This was proven by acquiring correlated images of the same regions from fluorescence and electron microscopes. The electron microscope provided higher resolution in both lateral and axial dimensions, thus enabling high-quality reconstruction. However, electron microscopes are costly, require complex sample preparation, and cannot be applied to live cells. Therefore, when presented with the input image as shown in Figure 5 (b) (top image), it is likely that even humans predict a ring/doughnut-shaped mitochondrion. But on examining the stack of images containing that instance (image stack in Figure 5 (b)), it is observed that the bowl shape (without a through hole) is accurate.

Zero-shot MiShape generalizes to real fluorescence microscopy and dynamic mitochondrial events

Quantitative validation of MiShape on real data is challenging due to the scarcity of data with 3D ground-truth. The gold standard data for this would be Correlated Light-Electron Microscopy (CLEM) sdata, which combines the modality of electron microscopy, which can have a resolution of a few nanometers. However, the use of CLEM requires specialized knowledge and costly equipment that is not widely available, and the process is time-consuming, requiring a complex sample preparation protocol. We therefore resort to openly available datasets. Another hindering factor is a limitation of MiShape; MiShape requires ROI's that are not densely populated with mitochondria. Since the model has been trained on single mitochondria instances per image in the training data, providing MiShape with densely populated fluorescence images at test time will inflict a large data distribution shift for the model. Therefore, we narrow our search further to openly available CLEM datasets of fluorescent microscopy to ones that have suitably sparse ROI to test MiShape. (CLEM datasets with Super-resolution microscopy data are more abundant than for fluorescent microscopy.)

Figure 6 (a) shows the steps used to extract ground truth shape from a publicly available 3D CLEM dataset[46] of hiPSC-derived neurons for a quantitative comparison of MiShape on real microscope images. The EM stacks have an isotropic resolution of 10nm, and the confocal image stack has a pixel size of 0.048 µm and an axial resolution of 0.140 µm. We select a region of interest (ROI) of size 128x128 pixels in the confocal image stack containing one mitochondrion instance and locate the corresponding region in the EM stack. We obtain the mitochondrial surface boundary from the EM stack using EMPANDA[2], a tool for deep learning-based panoptic segmentation of 2D and 3D electron microscopy images of cells. The resulting 3D EM segmented stacks are then converted to a 3D mesh using the marching cubes[47] algorithm. Since the confocal and EM stacks in the dataset were not registered, the GT shape was aligned using the Iterative Closest Point (ICP) algorithm[48]. The ground truth shape is then further processed using the data processing steps of 3DMSL to create the desired input format for comparison. Figure 6 (b) shows the quantitative and qualitative results of MiShape (input is a single fluorescence confocal image) along with the baselines (3-image stack input). The reconstruction metrics of IoU, Chamfer-L1, and normal consistency indicate superior zero-shot reconstruction quality of MiShape on real microscope images without the need for fine-tuning. However, we acknowledge that this result does not provide concrete experimental validation for comparison on real datasets, as

it is based on a single, relatively simple mitochondrial shape. Nevertheless, given the scarcity of the particular ground-truth datasets for comparison, we present this as proof of usability on real datasets.

Mitochondria are highly dynamic organelles and undergo shape shifts over time, depending upon their current functions. We present results of MiShape reconstructions on video datasets of commonly observed mitochondrial motions. We use publicly available fluorescence microscope videos of H9C2 cardiomyoblasts[15]. In Figure 6, we show MiShape predictions on real fluorescence data of two sequences (c) Mitochondrial fission: one mitochondrion splitting up into smaller parts, and (d) Curl – where a long mitochondrion curls up to form a loop-like structure. MiShape is not trained explicitly using video or sequential data, and therefore, the predictions are made per image. In Figure 6 (c), we observe that the predictions follow the pattern observed in the image over time, i.e, initially a small pinch is made to the center portion of the structure, which later on thins out this part further and leads to forming two separate structures. Although some of the individual predictions (for example, T= 9, 11) have some artefacts and abnormalities, the overall evolution of the shape is reflected in the predictions. In the predictions of the curl example presented in Figure 6 (d), we observe the shortening of the elongated structure and subsequent curling and eventually forming a loop structure.

Discussion

Why Implicit Shapes

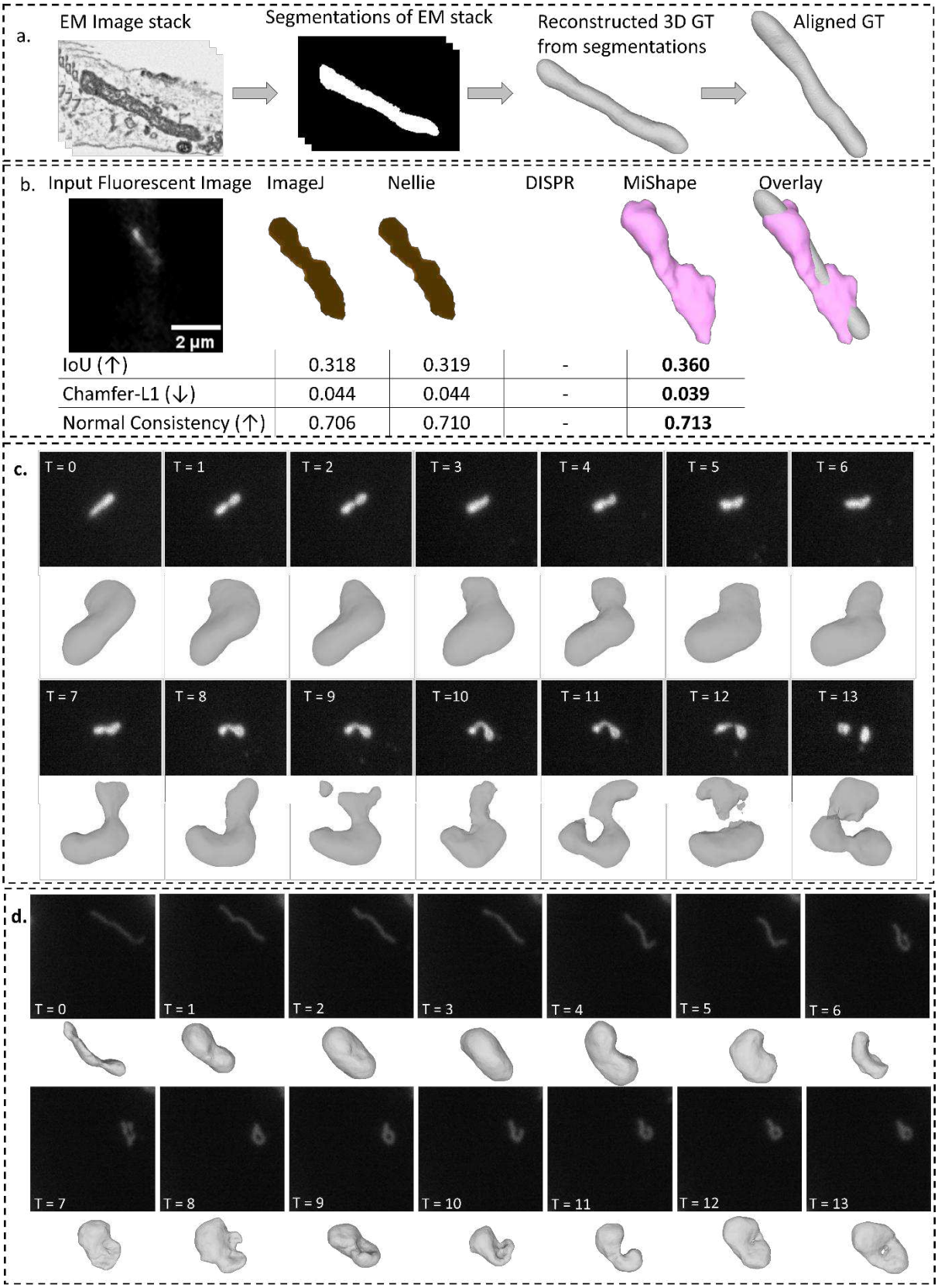
Traditionally used shape representations such as voxels are computationally expensive to train deep learning models and require resizing (as mitochondrial sizes vary greatly), leading to loss of detail in the shape. Learning on meshes involves deforming a template, which does not work for varying topologies, and point clouds lack connectivity information and require post-processing to extract the shapes. Implicit representation has better representation power for complex, irregular, and varying shapes exhibited by cell organelles.

Limitations of MiShape

We identify the following limitations of MiShape. First, we do not have accurate estimates of the absolute scale of the reconstructed shapes, as the data is scaled to a unit cube for training. This restricts the use of MiShape for quantitative analysis with absolute unit measurements. The scale factor can be computed through a minimization approach or regressed by modifying the model. Another limitation of MiShape is that the 3D prediction does not work for highly dense/crowded input images. This arises from the fact that the training data for MiShape is per-instance images. MiShape can handle 2-3 mitochondria in a single input image, but does not work for over-crowded network images.

Considerations for using an EM Data source

The biggest advantage of 3D electron microscopy is its resolution. However, the versatile use of 3D EM datasets for learning sub-cellular shape priors is difficult due to several limitations. The sample preparation for FIB-SEM images is quite challenging, especially if the morphology of the cells and sub-cellular organelles is to be preserved to match their living state. This can be achieved by using protocols that allow damage-free preservation of samples that match live cell morphology. Second, the lack of specificity in EM implies that everything in the cell gets imaged, and we require tools to segment objects of interest. Correlative light electron microscopy supplemented with modern processing helps but requires sophisticated experimental protocols. Imaging single cells requires several hours, sometimes days of imaging. Public datasets are emerging, yet the availability of 3D EM images with good segmentations is still scarce.



IoU (↑)	0.318	0.319	-	**0.360**
Chamfer-L1 (↓)	0.044	0.044	-	**0.039**
Normal Consistency (↑)	0.706	0.710	-	**0.713**

Figure 6: MiShape reconstruction on real fluorescent mitochondria images. (a) Steps to obtain a 3D shape from the uncorrelated EM and fluorescence image stack include obtaining segmentations of EM images, reconstructing the 3D shape from the segmentations using the marching cubes algorithm, and aligning the GT mesh to match the orientation of the fluorescent image results. (b) Qualitative and quantitative results comparing MiShape with baselines for results on real image. (c) Qualitative results of MiShape following a fission event (d) Qualitative result of MiShape on a curl event.

Possibility of using 3D Fluorescence Microscopy Data instead of EM data for constructing Shape Priors

3D fluorescence microscopes, such as confocal, airy-scan light-sheet, and lattice light sheets, are good candidates for large structures such as nuclei, where resolution is not a competing factor. However, super-resolution optical microscopy may be the only alternative to EM data for nanoscale structures. Stimulated Emission Depletion (STED) microscopy provides resolution as small as 20-30 nm in 3D, albeit with a compromise in throughput. Alternatively, higher throughput and almost isotropic resolution can be achieved using Structured Illumination Microscopy (SIM), although with poor resolution (~ 120 nm). Notably, nanoscopy of mitochondria images is a recent development. Nonetheless, we look forward to the emergence of large public datasets of nanoscopy images and their use in learning shape priors of sub-cellular organelles.

Future Directions

Microscope-to-microscope transformation: In the supplementary text, we present an additional exciting application of MiShape, namely, generating fluorescence images of mitochondria corresponding to a target microscope from fluorescence images or stacks of the same mitochondria acquired using another microscope.

We surmise that MiShape will open the path to developing an implicit shape atlas of cell organelles. By integrating information from the shape prior and sparse views, MiShape assembles structural detail that individual images alone cannot reveal, resolving the inherent ambiguity in many microscopy reconstructions. This ability reduces experimental and computational costs for biologists — train or calibrate once and reuse the model to recover shapes that were previously inaccessible. This will facilitate the simulation of realistic cell and sub-cellular morphologies, and their dynamic deformations and interactions. We believe that MiShape will enable computer vision and artificial intelligence to play a larger role in knowledge discovery in life sciences.

Methods

Data

Fluorescence microscopy images used for training are created from the mitochondrial shape of the 3DMSL dataset as per the protocol stated in 3DMSL[27]. Specifically, we use the confocal microscope with Conf2 configuration for the shape reconstruction and stack reconstruction experiments on 3DMSL. We also create epifluorescence images with the configuration of Epi1 to create the test set for the one-shot generalization experiment. More details of 3DMSL and exact parameters used for different microscope configurations used are provided in the supplementary text.

The real data used for the topology experiment (Figure 5b) were acquired in-house using a DeltaVision OMX V4 Blaze widefield fluorescence imaging system equipped with a 60×/1.42 NA oil-immersion objective (Olympus; immersion oil RI = 1.516). Images were acquired using Alexa Fluor 647/Cy5 filter settings with a z-step size of 300 nm. Human primary kidney mesangial cells were cultured on poly-L-lysine-coated No. 1.5 coverslips, fixed with 4% paraformaldehyde, permeabilized, and labeled with TOMM20 rabbit monoclonal antibody (EPR15581-54, Abcam). Goat anti-rabbit IgG (H+L) highly cross-adsorbed Alexa Fluor 647 secondary antibody (A-21245, Thermo Fisher Scientific) was used for detection. Coverslips were mounted onto glass slides using ProLong™ Diamond Antifade Mountant (Thermo Fisher Scientific).

Real data used in the experiment of quantitative comparison of zero-shot generalization on real data (Figure 6 (a) and (b)) are taken from a published CLEM dataset[46] of hiPSC-derived neurons imaged with the intention to reveal structural organizations and location of alpha-synuclein (α-Syn) aggregation hotspots inside the neuron. We summarize the information from the published description. The human iPSC-derived neuronal cells were cultured on gridded glass-bottom dishes and imaged using a Zeiss LSM 880 confocal microscope with Airyscan processing. The region of interest was acquired using a 63×/1.4 NA oil-immersion DIC UV-VIS-IR M27 objective, while lower-magnification tiling scans for relocation were acquired using a 25×/0.8 NA LD LCI Plan-Apochromat Imm Korr DIC M27 objective. The sample included Hoechst-labeled nuclei, mitoGFP-labeled mitochondria, and AF594-labeled human recombinant α-synuclein. Following fluorescence imaging, the same cells were processed for correlative electron microscopy by fixation, osmium-based heavy-metal staining, dehydration, Durcupan embedding, and relocation of the region of interest using the gridded substrate. The EM counterpart provides high-resolution (10 nm isotropic resolution) ultrastructural information for the selected mitochondrion and was used to obtain the reference 3D mitochondrial shape for evaluating MiShape reconstruction from the fluorescence image.

Real data from the experiment for zero-shot generalization on video data (Figure 6 (c) and (d)) is taken from publicly available microscopy videos of mitochondria in cardiomyoblasts[49]. We summarize the information from the published description. H9c2 rat cardiomyoblasts grown under standard glucose conditions, expressing a dual-tag outer mitochondrial membrane marker (mCherry–EGFP–OMP25). Cells were treated with 10 µM CCCP to depolarize mitochondria. Imaging was performed as a 3D dual-color fluorescence time-lapse (z-stacks over time) to monitor mitochondrial dynamics and EGFP loss in acidic compartments. The time-lapse microscopy data were acquired using a DeltaVision OMX V4 Blaze imaging equipped with a 60X 1.42NA oil-immersion objective (Olympus) and three sCMOS cameras for rapid multi-channel imaging.

Implicit Shape Representation

We base our work on occupancy networks[23] for shape representation of mitochondria. Occupancy-based implicit representation is especially suited for our purpose since it allows us to shift to isomorphic representations adaptable to different target resolutions from the voxelized segmentations of EM images. Occupancy is an implicit 3D shape representation that represents 3D surfaces as the continuous decision boundary of a deep neural network classifier. For a shape S, occupancy function o maps every possible point in the 3D space $\mathbb{R}^3$ as the occupancy probability between 0 and 1 which can be denoted by,

$$o: \mathbb{R}^3 \rightarrow \{0,1\}$$

A deep neural network can be used to approximate this function denoted by f_θ, which takes an observation $x \in \chi$ as input and outputs the occupancy for every point in $\mathbb{R}^3$.

$$f_\theta: \chi \times \mathbb{R}^3 \rightarrow [0,1]$$

This network is termed as occupancy network. A hierarchical isosurface extraction algorithm is used to extract the high-resolution meshes from the occupancy network.

Unconditional MiShape

We train an unconditional generative model to sample infinite instances from the shape distribution for fluorescence dataset creation and fluorescence microscope-to-microscope translation experiments. We use the same methodology as occupancy networks for our unconditional generative model, denoted by

$$f_\theta(h, p) = o: \mathbb{R}^c \times \mathbb{R}^3 \rightarrow \mathbb{R}$$

where h is a c- dimensional latent vector encoding of the shape S and $p \in \mathbb{R}^3$ is the query point and $o \in [0,1]$, is the probability of occupancy.

The sampled points are used as input to the model and the model is trained to predict the occupancy values of the query point. To learn a latent distribution or shape prior, the model is constructed as a variational autoencoder. The loss is a combination of classification cross-entropy loss between the true occupancies and the predicted occupancies and the KL divergence between the latent space and a Gaussian distribution. Once the training is complete, the model can be used in the inference stage to randomly generate new shapes by passing Gaussian noise as input. The architecture and implementation details are provided in the supplementary text.

The architecture of MiShape decoder is given in Figure 1 (b) as is derived from occupancy networks. The encoder-decoder modules form a VAE-like construction. All the models take the output $c \in \mathbb{R}^C$ of the encoder network, which differs for different experiments and a batch of T input 3D points. The points are projected to a higher-dimensional space with a Fully Connected (FC) layer. This high-dimensional vector is fed to 5 ResNet blocks. Each ResNet block is conditioned through a Conditional Batch Normalization (CBN) layer. The output is then passed to an FC layer and a CBN layer. The final FC layer projects the output to a 1-dimensional output vector. All models are trained on a sum of binary cross-entropy loss between the predicted occupancies and true occupancies and the KL-divergence loss between the latent distribution and a normal distribution. The distinct differences from the original occupancy network and MiShape are: (i) the introduction of positional encoding from the work NeRF[50], which is meant to help with high-frequency variations in shape. These are essentially sine and cosine multiples of the model input, with frequencies varying from 1 to 5. (ii) The use of weighted sampling during training to counteract the unbalanced nature of the dataset 3DMSL. The weighting factor is the inverse of the class frequency of the instance for which the classification is done, as stated in 3DMSL. So, the most abundant class, like the dot mitochondria, has a lower probability of being sampled during training than a scarce shape like a nano-tunnel. The details of encoder is given in the supplementary text.

MiShape with Image/Z-stack conditioning

Conditioning individual mitochondria shapes with their counterpart fluorescence microscopic image or z-stack permits learning the shape correspondence between microscope images and the corresponding 3D shapes. This can be written as,

$$f_\theta(h, p, m) = o: \mathbb{R}^M \times \mathbb{R}^c \times \mathbb{R}^3 \rightarrow \mathbb{R}$$

where m is the encoding of the fluorescent microscope image or z-stack corresponding to the shape S. For this, we train an occupancy model conditioned on the encoding of simulated microscope images. We train multiple models, for mapping a single 2D fluorescence image to 3D shape (image-to-shape) and the others for mapping 3-slice, 5-slice, and 7-slice fluorescence z-stack to 3D shape (stack-to-shape).

For the image conditioned model, the simulated image is input to a 2D ResNet- 18[51] to produce the conditioning c. The conditioning with the image input is done using conditional batch normalization on each of the ResNet blocks of the decoder. The architecture of the stack encoder is adapted from the voxel encoder architecture of the occupancy network. It consists of 4 consecutive layers of 3D convolutions with 32, 64, 128, 256, and 512 kernels of size 3, followed by a linear layer of size 128 (dimension of the condition input).

Data Availability

The training data MiShape is available in Dataverse at (https://doi.org/10.18710/JX6JXF). Additional data formats and used to perform quantitative evaluation are available in Zenodo at (https://zenodo.org/records/21159675)

Code Availability

The code development for MiShape was done using Python 3.6.7 on Linux ecosystem. The source code and scripts are available through GitHub at (https://github.com/bioailab/MiShape). The pretrained models are available for download at (https://zenodo.org/records/21159675)

References

1. Miyazono, Y. *et al.* Uncoupled mitochondria quickly shorten along their long axis to form indented spheroids, instead of rings, in a fission-independent manner. *Sci. Rep.* **8**, (2018).
2. Conrad, R. & Narayan, K. Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model trained on a diverse dataset. *Cell Syst.* **14**, 58-71.e5 (2023).
3. Harmuth, T. *et al.* Mitochondrial Morphology, Function and Homeostasis Are Impaired by Expression of an N-terminal Calpain Cleavage Fragment of Ataxin-3. *Front. Mol. Neurosci.* **11**, (2018).
4. Ong, S. B. & Hausenloy, D. J. Mitochondrial morphology and cardiovascular disease. *Cardiovasc. Res.* **88**, 16–29 (2010).
5. Chinnery, P. F. Primary Mitochondrial Disorders Overview. in (eds. Adam, M. P. et al.) (1993).
6. Viana, M. P., Lim, S. & Rafelski, S. M. Quantifying mitochondrial content in living cells. *Methods Cell Biol.* **125**, 77–93 (2015).
7. Sekh, A. A. *et al.* Physics-based machine learning for subcellular segmentation in living cells. *Nat. Mach. Intell. 2021 312* **3**, 1071–1080 (2021).
8. Chu, C. H., Tseng, W. W., Hsu, C. M. & Wei, A. C. Image Analysis of the Mitochondrial Network Morphology With Applications in Cancer Research. *Front. Phys.* **10**, 289 (2022).
9. Lefebvre, A. E. Y. T., Ma, D., Kessenbrock, K., Lawson, D. A. & Digman, M. A. Automated segmentation and tracking of mitochondria in live-cell time-lapse images. *Nat. Methods 2021 189* **18**, 1091–1102 (2021).
10. Lihavainen, E., Mäkelä, J., Spelbrink, J. N. & Ribeiro, A. S. Mytoe: automatic analysis of mitochondrial dynamics. *Bioinformatics* **28**, 1050–1051 (2012).
11. 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).
12. Zamponi, N. *et al.* Mitochondrial network complexity emerges from fission/fusion dynamics. *Sci. Reports 2017 81* **8**, 363- (2018).
13. Fischer, C. A. *et al.* MitoSegNet: Easy-to-use Deep Learning Segmentation for Analyzing Mitochondrial Morphology. *iScience* **23**, (2020).

14. Ding, Y. *et al.* Mitochondrial segmentation and function prediction in live-cell images with deep learning. *Nat. Commun. 2025 161* **16**, 1–15 (2025).

15. Punnakkal, A. R. *et al.* Analyzing Mitochondrial Morphology Through Simulation Supervised Learning. *J. Vis. Exp.* **2023**, e64880 (2023).

16. Ma, J. & Wang, B. Towards foundation models of biological image segmentation. *Nat. Methods 2023 207* **20**, 953–955 (2023).

17. Punnakkal, A. R. *et al.* Analyzing Mitochondrial Morphology Through Simulation Supervised Learning. *J. Vis. Exp.* (2023) doi:10.3791/64880-V.

18. Godtliebsen, G. *et al.* High-resolution visualization and assessment of basal and OXPHOS-induced mitophagy in H9c2 cardiomyoblasts. *Autophagy* **19**, 2769–2788 (2023).

19. Peiris, H., Hayat, M., Chen, Z., Egan, G. & Harandi, M. Uncertainty-guided dual-views for semi-supervised volumetric medical image segmentation. *Nat. Mach. Intell. 2023 57* **5**, 724–738 (2023).

20. Detmer, S. A. & Chan, D. C. Functions and dysfunctions of mitochondrial dynamics. *Nat. Rev. Mol. Cell Biol. 2007 811* **8**, 870–879 (2007).

21. Yapa, N. M. B., Lisnyak, V., Reljic, B. & Ryan, M. T. Mitochondrial dynamics in health and disease. *FEBS Lett.* **595**, 1184–1204 (2021).

22. Tilokani, L., Nagashima, S., Paupe, V. & Prudent, J. Mitochondrial dynamics: overview of molecular mechanisms. *Essays Biochem.* **62**, 341 (2018).

23. 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).

24. Park, J. J., Florence, P., Straub, J., Newcombe, R. & Lovegrove, S. Deepsdf: Learning continuous signed distance functions for shape representation. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vols 2019-June (2019).

25. Chen, Z. & Zhang, H. Learning implicit fields for generative shape modeling. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vols 2019-June 5932–5941 (2019).

26. Michalkiewicz, M., Pontes, J. K., Jack, D., Baktashmotlagh, M. & Eriksson, A. P. Deep level sets: Implicit surface representations for 3d shape inference. CoRR abs/1901.06802 (2019). (1901).

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

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

29. Punnakkal, A. R. *et al.* 3D Mitochondria Shape Library for Optical Microscopy (3DMSL): A multimodal dataset for deep learning based mitochondrial analysis. *Data Br.* **65**, 112507 (2026).

30. Jenkins, B. C. *et al.* Mitochondria in disease: changes in shapes and dynamics. *Trends Biochem. Sci.* **49**, 346–360 (2024).

31. Knott, A. B., Perkins, G., Schwarzenbacher, R. & Bossy-Wetzel, E. Mitochondrial fragmentation in neurodegeneration. *Nat. Rev. Neurosci. 2008 97* **9**, 505–518 (2008).

32. Wakabayashi, T. Megamitochondria formation - physiology and pathology. *J. Cell. Mol. Med.* **6**, 497 (2007).

33. Fenton, A. R., Jongens, T. A. & Holzbaur, E. L. F. Mitochondrial dynamics: Shaping and remodeling an organelle network. *Curr. Opin. Cell Biol.* **68**, 28–36 (2021).

34. Miyazono, Y. *et al.* Uncoupled mitochondria quickly shorten along their long axis to form indented spheroids, instead of rings, in a fission-independent manner. *Sci. Reports 2017 81* **8**, 1–14 (2018).

35. Yang, J., Wickramasinghe, U., Ni, B. & Fua, P. ImplicitAtlas: Learning Deformable Shape Templates in Medical Imaging.

36. Punnakkal, A. R., Bhansal, A. & Prasad, D. K. On Generalization of 3D Generative Models. in *8th Workshop on Generalization in Planning, AAAI 202* (2024).

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

38. Waibel, D. J. E., Röell, E., Rieck, B., Giryes, R. & Marr, C. A Diffusion Model Predicts 3D Shapes from 2D Microscopy Images. (2022) doi:10.48550/arxiv.2208.14125.

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

40. 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).

41. Novotny, L. & Hecht, B. *Principles of nano-optics*. (Cambridge university press, 2012).

42. Xue, F., Li, J. & Blu, T. Fast and accurate three-dimensional point spread function computation for fluorescence microscopy. *JOSA A, Vol. 34, Issue 6, pp. 1029-1034* **34**, 1029–1034 (2017).

43. Conrad, R. & Narayan, K. Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model. *bioRxiv* 2022.03.17.484806 (2022) doi:10.1101/2022.03.17.484806.

44. Miyazono, Y. *et al.* Uncoupled mitochondria quickly shorten along their long axis to form indented spheroids, instead of rings, in a fission-independent manner. *Sci. Reports 2017 81* **8**, 1–14 (2018).

45. 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).

46. Choi, M. L. *et al.* Pathological structural conversion of α-synuclein at the mitochondria induces neuronal toxicity. *Nat. Neurosci.* **25**, 1134–1148 (2022).

47. Lempitsky, V. Surface extraction from binary volumes with higher-order smoothness. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 1197–1204 (2010) doi:10.1109/CVPR.2010.5539832.

48. Liu, J. *et al.* Automatic Reconstruction of Mitochondria and Endoplasmic Reticulum in Electron Microscopy Volumes by Deep Learning. *Front. Neurosci.* **14**, 599 (2020).

49. Opstad, I. Fluorescence microscopy videos of mitochondria in H9c2 cardiomyoblasts. doi:doi:10.18710/11LLTW.

50. Mildenhall, B. *et al.* NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis.

Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) **12346 LNCS**, 405–421 (2020).

51. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2016-December**, 770–778 (2015).

Acknowledgements

This work was supported H2020 FET Open RIA project OrganVision (id. 964800), and UiT Tematisk Satsinger project Virtual Stain (Cristin id. 2061348). Open access funding provided by UiT The Arctic University of Norway (incl University Hospital of North Norway).

We would like to thank all the participants of the human evaluation experiment for their voluntary help and support. We wish to acknowledge the fruitful discussions with Dr. Nikhil Jayakumar (UiT), Dr. Mitchell J. Fulton (UiT) and Dr. Rohit Agarwal (UiT).

Author Information

Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway
Abhinanda R. Punnakkal, Suyog S. Jadhav, Alexander Horsch, Dilip K. Prasad

Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway
Biswajoy Ghosh, Krishna Agarwal

A.R.P, K. A, and D.K.P. conceptualized the study. A.R.P developed the code and performed experiments. S. J and A.R.P conducted the benchmarking comparison. B. G performed imaging. A.R. P and K. A drafted the manuscript. K. A, D.K.P and A.H provided supervision. A.R.P and K.A drafted the manuscript. K.A, A. H and D.K.P acquired funding for the project. All authors reviewed the manuscript and contributed to discussions.

Corresponding authors: Abhinanda Ranjit Punnakkal

Ethics Declarations

Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Disclosure of Use of AI-assisted Tools, including Generative AI

We utilize an AI-assisted tool, Grammarly, for academic proofreading and spell check. We also used coding assistance in the form of GitHub Copilot predictions. All research ideas were developed solely by the authors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.pdf
- Supplementaryvideos.zip