This microtubule does not exist: Super-resolution microscopy image generation by a diffusion model

Alon Saguy[1], Tav Nahimov[1], Maia Lehrman[1], Estibaliz Gómez-de-Mariscal[2], Iván Hidalgo-Cenalmor[2], Onit Alalouf[1], Ricardo Henriques[2,3], Yoav Shechtman[1,4,†]

[1] Department of Biomedical Engineering, Technion – Israel Institute of Technology, Haifa, Israel

[2] Optical cell biology group, Instituto Gulbenkian de Ciência, Oeiras, Portugal

[3] UCL Laboratory for Molecular Cell Biology, University College London, London, UK

[4] Department of Mechanical Engineering, University of Texas at Austin, Austin, TX

[†] Corresponding Author

Abstract

Generative models, such as diffusion models, have made significant advancements in recent years, enabling the synthesis of high-quality realistic data across various domains. Here, we explore the adaptation and training of a diffusion model on super-resolution microscopy images from publicly available databases. We show that the generated images resemble experimental images, and that the generation process does not memorize existing images from the training set. Additionally, we compare the performance of a deep learning-based deconvolution method trained using our generated high-resolution data versus training using high-resolution data acquired by mathematical modeling of the sample. We obtain superior reconstruction quality in terms of spatial resolution using a small real training dataset, showing the potential of accurate virtual image generation to overcome the limitations of collecting and annotating image data for training. Finally, we make our pipeline publicly available, runnable online, and user-friendly to enable researchers to generate their own synthetic microscopy data. This work demonstrates the potential contribution of generative diffusion models for microscopy tasks and paves the way for their future application in this field.

Introduction

Deep learning algorithms have been extensively used in the past decade to solve various microscopy challenges[1–7]. These algorithms outperform traditional computer vision methods in terms of reconstruction quality, analysis time, and classification, among many others. However, deep learning solutions are hungry for data. To train a model, one should typically acquire and annotate hundreds or even thousands of images manually, a highly time and resource consuming process. An alternative approach is to produce synthetic data based on mathematical models that describe the structure of the biological specimen[1,3,7–9]. Yet, tuning the data generation parameters is a cumbersome process that leads to non-realistic features in the synthetic images due to parameter estimation errors and model inaccuracies, which is critical to train highly generalizable and accurate models.

Recently, the field of generative models has seen a significant surge in terms of both development and application[10–13]. Generative models have moved far beyond their initial application in producing artificial images and are now being used to create synthetic datasets that can effectively mimic real-world data in diverse domains[14]. Two major contributors to this advancement have been Denoising Diffusion Probabilistic Models (DDPM)[10] and Denoising Diffusion Implicit Models (DDIM)[13]. DDPM and DDIM offer

a dynamic approach for the generation of synthetic data, relying on stochastic processes to create totally new images that still capture the inherent image statistics present in the training dataset.

The capacity of diffusion models to accurately create realistic visual data is profoundly impacting many computer vision applications[15], including microscopic imaging, where overcoming the existing challenges to gather high-quality large training datasets is invaluable. Indeed, several studies already incorporate diffusion models to microscopy to reconstruct 3D biomolecule structures in Cryo-EM images[16], predict 3D cellular structures out of 2D images[17], or drug molecule design[18], among others.

Here, we propose the application of generative diffusion models in the field of super-resolution microscopy. First, we show the ability of diffusion models to generate realistic, high-quality, super-resolution microscopy images of microtubules and mitochondria. Then, we assess the capacity of the models to learn the intricate nature of the data domain by validating that the network does not memorize images from the training data. Next, we utilize the generated dataset to train a single-image super-resolution deep learning model and show superior reconstruction quality compared to the same model trained on model-based simulated data. The diffusion model approach proposed here is publicly available on the ZeroCostDL4Mic platform[19], enabling non-expert researchers to benefit from it.

Results

We base our work on a previously reported[20] diffusion model which we adapt to super-resolution microscopy. We trained two diffusion models on different biological samples, microtubules and mitochondria, sourced from a publicly available database (ShareLoc.xyz[21–23]). We split our data into 60% and 40% for training and validation of the performance, obtaining a 7:5 training:validation image ratio for the microtubule data and a 3:2 training:validation image ratio for the mitochondria data. Furthermore, we split each image into patches of 256x256 pixels and transformed them using random horizontal flips and rotations of 90, 180, and 270 degrees to augment the training data. The augmentation step yielded a total of 2000 training patches for the microtubule data and 800 training patches for the mitochondria data. Training details are further specified in the Methods section.

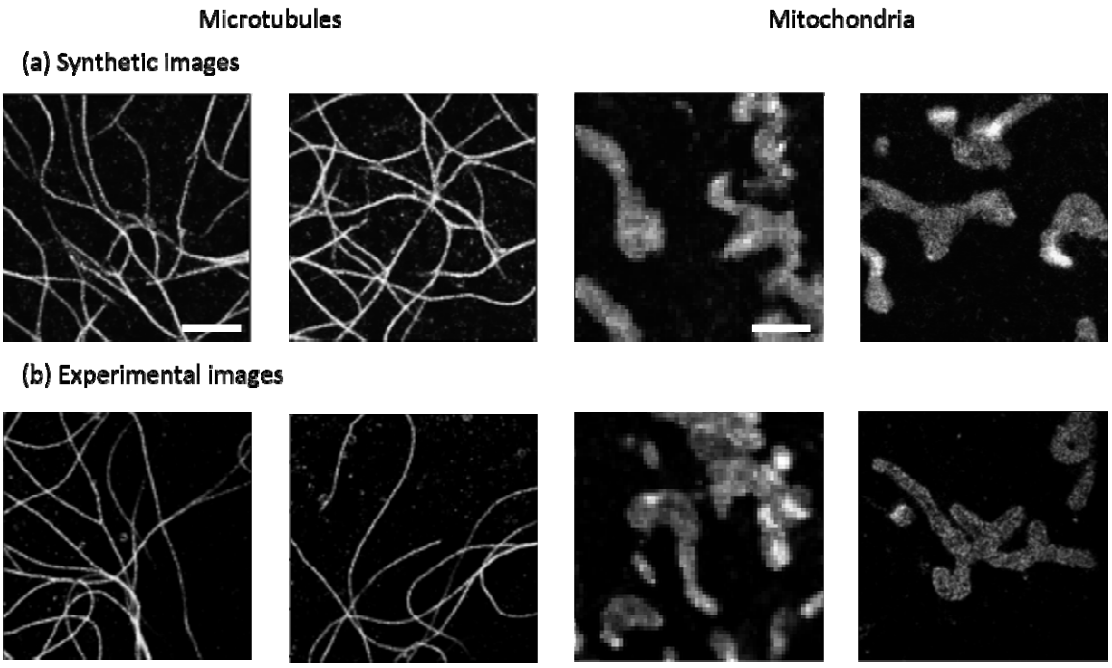

Figure 1: Qualitative comparison of experimental microscopy data versus data generated using our generative diffusion model. (a) Example synthetic images of microtubules (alpha-tubulin – Alexa647) and mitochondria (TOM 22 – Alexa647) generated by our diffusion model. (b) Example experimental super-resolution images, used as training data. Scale bars = 2.5 .

The images generated by our DDPM qualitatively resemble the training data used for the training, as can be clearly seen in the examples in Figure 1. To validate that our model does not memorize images, namely, copy existing images from the training set and generate them as network outputs, we calculated the normalized cross-correlation between every generated image (a total of 50 images), including rotated and flipped versions of the images, and the augmented patches used for training. The maximal normalized cross-correlation, calculated between all generated images and the training data, was 0.345 (0.682) for the synthetic microtubule (mitochondria) images. The mean normalized cross-correlation value was 0.336 (0.631) for the microtubules (mitochondria) images. For comparison, repeating this process of normalized cross-correlation calculation between experimental images, taken from different datasets (a total of 10 images), and all other training data, yielded a mean value of 0.372 (0.414) and max value of 0.483 (0.510) for the microtubule (mitochondria) data. Then, we overlaid the training images that obtained the highest cross-correlation score with the generated images to verify that our synthetic images are sufficiently new and different from the training images (Figure 2).

Notably, the cross-correlation values are similar for experimental microtubule images from another dataset (imaged in similar conditions) and the microtubule images that were generated by our diffusion model, showing the expected variability between different and independent datasets. In the case of mitochondria images, the cross-correlation values were slightly higher than those obtained when comparing with images from a different experimental dataset (see in-depth analysis in the discussion section).

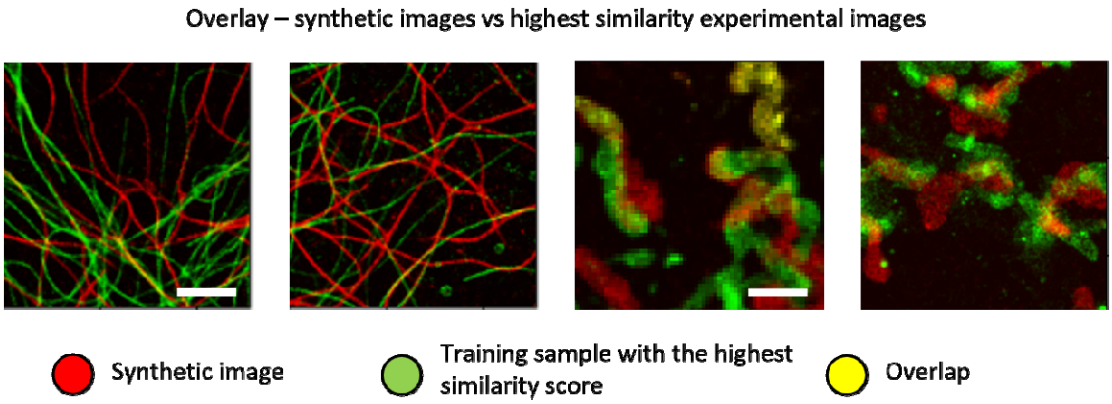

Figure 2: Diffusion models do not memorize training images. Overlay between each reconstructed image and the training image with highest resemblance (maximal cross-correlation score). Red marks generated data, green marks the closest training sample, and yellow marks overlap between the two images. Scale bars = 2.5 .

Next, we tested the applicability of our generated data to improve deep learning-based methods. We used our generated data to train an application of Content-Aware Restoration (CARE)[1], specifically, a deconvolution method aiming to transform a low-resolution image to a high-resolution image based on prior knowledge of image-statistics. Notably, obtaining single-image-based super-resolution algorithmically is yet an unsolved problem in microscopy, with results strongly dependent on the prior information provided, and is no match to physics based super-resolution microscopy methods (SMLM, STED, SIM, etc.[24–27]). Nevertheless, we use this task to demonstrate the potential of diffusion model-based data generation in virtual super-resolution microscopy imaging. We trained two CARE models for each biological sample using 1) synthetic images generated by a mathematical model and 2) images generated by our diffusion model. During the training stage, we simulated synthetic high-resolution images either by our model or a mathematical model; next, we obtained low-resolution images by forward passing the high-resolution images through a model of our optical system (see methods section for more details). Ultimately, we used these low-resolution – high-resolution pairs to train CARE.

Visually, the CARE model trained on data generated by the diffusion model yielded a better reconstruction in comparison to the traditionally trained network (Figures 3, 4). Moreover, we have analyzed the spatial resolution we obtained in both reconstructions using the Fourier Ring Correlation (FRC) plug-in for ImageJ[28]. In brief, FRC is a similarity measure that seeks the maximal spatial frequency in which the reconstructed and ground truth images are similar up to a predefined threshold. The similarity is quantified by the normalized cross-correlation between the Fourier transforms of both images inside a torus with increasing radius. A high cross-correlation value within the torus indicates high similarity between the images, in the corresponding spatial frequency band.

The mean spatial resolution of the reconstructed images, as quantified by FRC, using a 1/7 threshold[28] when training on microtubule images generated by our diffusion model was 100 nm, while the mean spatial resolution obtained when training on synthetic microtubules generated via a mathematical model was 140 nm.

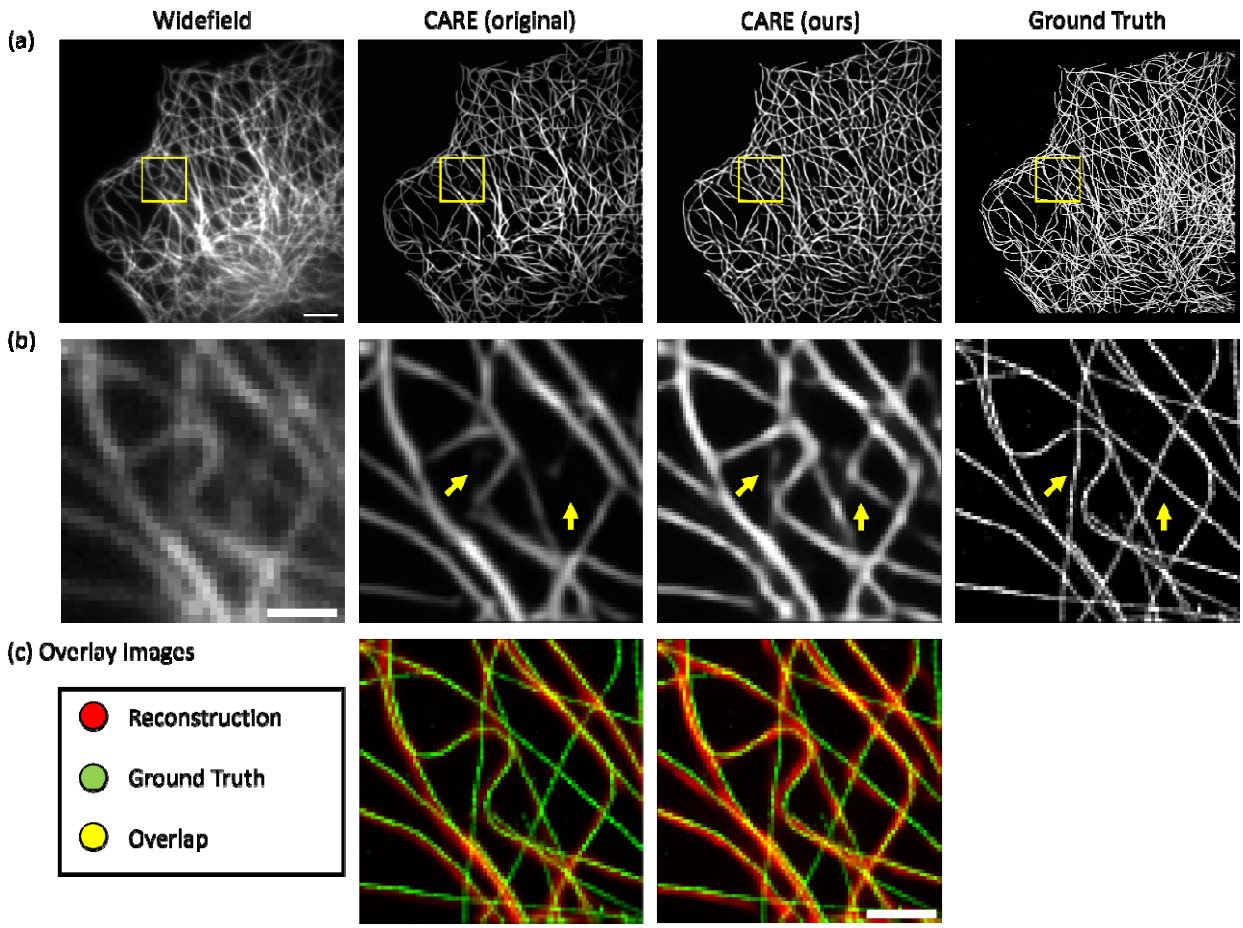

Figure 3: Performance of CARE trained on synthetic microtubule images generated by a mathematical model vs. training on microtubules generated by our diffusion model. (a) Left to right: widefield image, CARE reconstruction when trained on mathematical simulations, CARE reconstruction when trained on our synthetic data, and ground truth. Scale bar = 5 . (b) Regions of interest (marked by yellow squares in (a)), yellow arrows mark areas in which CARE trained on our data outperformed the previous method. (c) Left: overlay between CARE trained on mathematical simulations (red) and the ground truth (green). Right: overlay between CARE trained on our diffusion model-based synthetic data (red) and the ground truth (green). Scale bar = 1 .

Notably, microtubule images can be simulated with relatively high fidelity by a variety of well-established mathematical models[29]. However, for an arbitrary type of biological specimen, it is not easy to obtain a simple mathematical model describing its shape and characteristics. Therefore, the most remarkable feature of diffusion model-based data generation is the ability to generate synthetic data from non-mathematically defined biological specimen. Additionally, diffusion models might also contribute to the understanding of biological structures in a data driven manner by interpretation of the generated image statistics.

We validate this claim by training our diffusion model on publicly available super-resolution images of mitochondria[22] (Figure 4). The spatial resolution obtained by CARE trained on our generated mitochondria was 110 nm. Unlike for microtubules, there is no available mathematical model to generate mitochondria images. Therefore, training CARE traditionally would require obtaining enough extended super-resolution images.

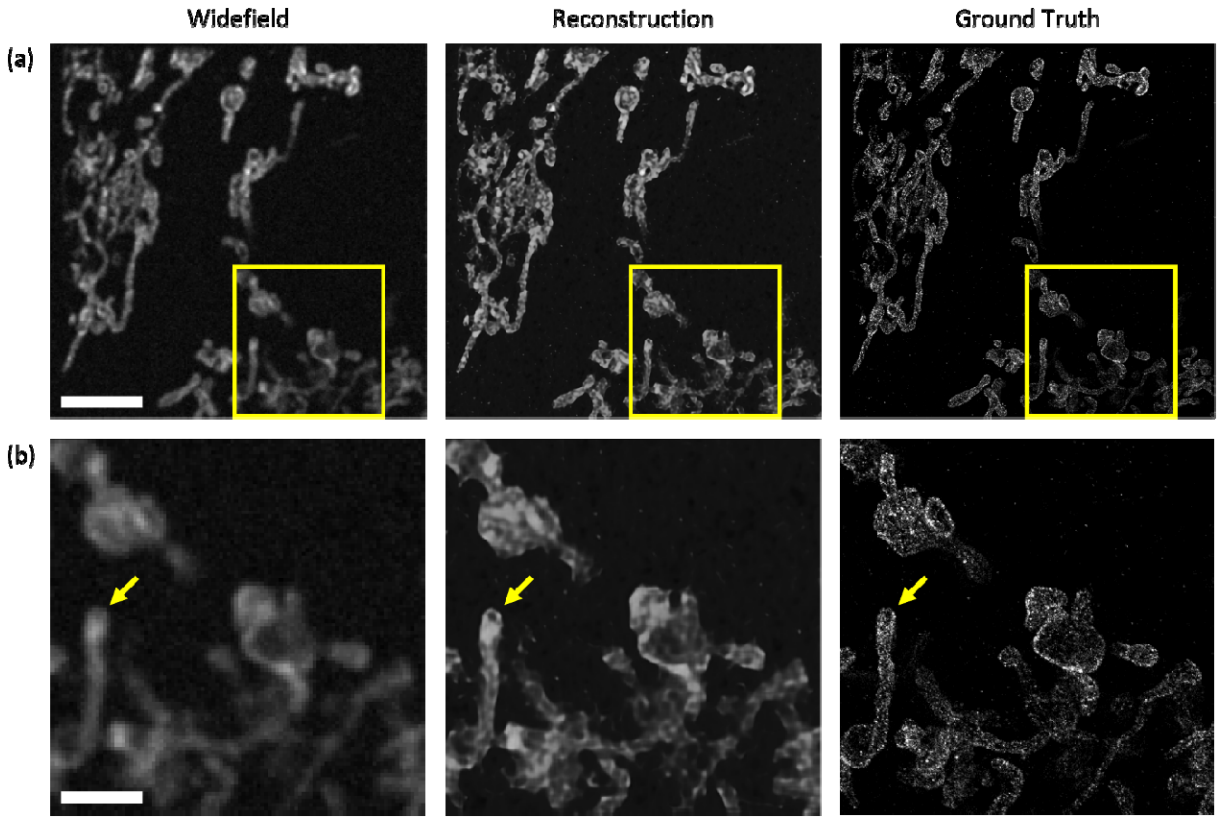

Figure 4: Performance of CARE trained on mitochondria generated by our diffusion model. (a) Left to right: widefield image, CARE reconstruction when trained on our synthetic data, and ground truth. Scale bar = 5 . (b) Region of interest; yellow arrow marks a subtle feature not visible in widefield imaging, which is made visible in our reconstruction. Scale bar = 2 .

Discussion

In this work, we demonstrate the potential of diffusion models to generate large super-resolution microscopy datasets by relying on a relatively small number of super-resolution images. Given only 7 microtubule images we managed to generate realistic microtubule images that looked different from the original training data, while resembling similar experimental data distribution. Notably, when training on yet a smaller dataset, i.e., 3 mitochondria images, some parts of the image were memorized, yielding slightly higher cross-correlation values when using the generated images than the values obtained when using mitochondria images from a different dataset. This observation is in par with existing work in this field[30,31], implying that larger training sets prevent memorization and increase the uniqueness of the generated data. Therefore, we suggest using quantitative sanity checks (such as the cross-correlation metric) on the generated data as a tool to evaluate whether enough images were used to train the diffusion model. Additionally, when choosing the number of images for training, one should also take into consideration the data complexity and the size of the field-of-view of each image.

Creating synthetic images of biological data that are highly realistic and representative of the original data has important implications. For example, diffusion models enable efficient generation of super-resolution datasets that could be transformed to low-resolution observations by forward passing through an optical model of the imaging system; then, one may perform supervised model training

without the need for extensive experimental data acquisition, often a limiting factor due to the impractical duration of the acquisition process. The contribution of our method is particularly relevant for the general case where no simple mathematical model is available for synthetic image generation.

Moreover, the ability to learn the complexity of biological structures and reproduce them to create realistic simulated images is key to identifying and interpreting biological phenomena and reinforcing data-driven discoverability. Our experiments have shown that diffusion models can be of great value in this area; these models have the ability to produce synthetic images that are of high quality and closely resemble real microscopy data, even when dealing with complex structures such as microtubule networks. To obtain accurate and highly generalizable models, it is essential to train deep learning models with the most realistic and extensive data possible, covering most of the natural experimental domain. These highly effective models, also known as foundation models, require massive amounts of image data to be trained effectively, which could be partially alleviated by smart and accurate data generation.

While the task chosen in this work to demonstrate the potential of the approach is single-image super-resolution, the applicability of diffusion model-based image generation for microscopy is naturally much broader. Numerous potential applications exist, including denoising, multi-image super-resolution, cross-modality imaging, live-cell dynamic imaging, and more. On the other side, quantitative evaluation of biological image data generation in the lack of annotated images is still an open question in the field that requires further work and consensus.

We share an easy-to-use notebook via the ZeroCostDL4Mic[18] platform to enable researchers to replicate our pipeline and harness diffusion model capabilities. We also distribute the pretrained models that allow the generation of data similar to the data presented in this work. Of note, training diffusion models is time consuming due to the large number of stochastic operations involved in the learning process.

In light of the encouraging results obtained from this study, future research should continue to focus on further optimizing and evaluating diffusion models for generating more types of synthetic microscopy data and on finding the applications where these capabilities are most impactful. Furthermore, due to the capacity of diffusion models to create virtual representations of nanoscale cellular structure, they can potentially predict prospective multi-structural spatial relationships that will guide observations and discovery in the field of microscopy. The emergence of generative models for microscopy represents an exciting phase for bio-medical research and holds promising potential for advancements in the near future.

Funding

This research was supported in part by funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 802567-ERC-Five-Dimensional Localization Microscopy for Sub-Cellular Dynamics. Y.S. is supported by the Zuckerman Foundation and by the Donald D. Harrington Fellowship. E.G.M., I.H.C., R.H. is supported by the support of the Gulbenkian Foundation (Fundação Calouste Gulbenkian), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 101001332 to R.H.) and the European Union through the Horizon Europe program (AI4LIFE project with grant agreement 101057970-AI4LIFE, and RT-SuperES project with grant agreement 101099654-RT-SuperES to R.H.). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. Our

work was also supported by the European Molecular Biology Organization (EMBO) Installation Grant (EMBO-2020-IG-4734 to R.H.), an EMBO Postdoctoral Fellowship (EMBO ALTF 174-2022 to E.G.M.), the Chan Zuckerberg Initiative Visual Proteomics Grant (vpi-0000000044 with doi:10.37921/743590vtudfp to R.H.). R.H. also acknowledges the support of LS4FUTURE Associated Laboratory (LA/P/0087/2020).

Methods

Optical model for low-resolution image generation

To train CARE on low-resolution – high-resolution image pairs, we used high-resolution data and passed it through a model of our optical system to obtain low-resolution images. In this work, we use a simple model to simulate a 2D low-resolution image based on a 2D high-resolution image. Let the imaged structure be depicted by $S(x, y)$ and let $H(x, y)$, the point spread function (PSF) of the optical system, be modeled as a 2D Gaussian:

$$H(x, y) = A \cdot e^{-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}}$$

Where A is the amplitude of the PSF, x_0, y_0 are the position of the emitter, and $\sigma_x = \sigma_y = \sigma$ represents the PSF width.

The low-resolution image formed at the camera is described by the convolution of the imaged structure with the system's PSF equation:

$$I(x, y) = P\big(S(x, y) * H(x, y)\big) + G(x, y)$$

Where * indicates a convolution operator, $P(x, y)$ indicates a Poisson distribution of the emitted number of photons, and $G(x, y)$ indicates a Gaussian noise simulating the camera read noise.

Diffusion model architecture and training details

We have adopted the network architecture presented by Nichol, et al[20] We used a single residual network (ResNet) block and we changed the input and output layers of the model to fit monochromatic data. To decrease the network size, we also changed the channel multiplication between different layers of the ResNet, namely, instead of (1, 1, 2, 2, 4, 4) multiplication we used (1, 1, 2, 2, 2, 2) multiplication, where the initial channel number is 64. Additionally, we changed the number of diffusion steps to 2000, set the batch size to 10, the learning to $1e^{-5}$, and employed a cosine noise schedule. To train the network, we used 7 super-resolution localization lists of microtubule experiments and 3 of mitochondria experiments, all publicly available (ShareLoc[19]); then, we generated from each localization list a super-resolved image scaled by a factor of 4 in comparison to the diffraction limited data, yielding pixel sizes of 27 nm and 32 nm for the microtubule and mitochondria images, respectively.

Next, we split the input images to multiple overlapping patches of size $256x256\ pixels^2$ and augmented the patches by flipping and rotating the images horizontally and vertically. The total number of training patches we used is 2000 and 800 for the microtubule and mitochondria networks respectively. Finally, we trained the generative diffusion model over $80{,}000$ steps for 8 hours on a single NVIDIA 32GB Titan RTX GPU. Ultimately, generation of a single super-resolution image depends on the image size, e.g. 30 seconds for images of size $256x256$ pixels^2.

CARE training details

We obtained super-resolution training data based on: 1) the mathematically simulated data presented in CARE paper; 2) the data generated by our trained diffusion model. To generate the low-resolution data needed for training CARE network, we followed a similar scheme as described in the CARE paper by convolving the super-resolution data with a gaussian microscope PSF model and adding Perlin noise, shot noise and gaussian noise. Importantly, we made sure that images generated by the two methods described above shared properties such as signal-to-noise ratio, sample size, etc. Finally, we trained the CARE network on 5000 synthetic low-resolution-high-resolution image pairs for the microtubule reconstruction and 2000 for the mitochondria reconstruction. To maintain a fair comparison between CARE trained on our data vs CARE trained on the mathematically generated microtubules, we used the same training set size in both cases.

References

1. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods* **15**, 1090–1097 (2018).
2. Nehme, E., Weiss, L. E., Michaeli, T. & Shechtman, Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* **5**, 458 (2018).
3. Saguy, A. *et al.* DBlink: Dynamic localization microscopy in super spatiotemporal resolution via deep learning. *Nature Methods* 1–10 (2023).
4. Nogin, Y. *et al.* DeepOM: single-molecule optical genome mapping via deep learning. *Bioinformatics* **39**, btad137 (2023).
5. Nehme, E. *et al.* DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning. *Nature Methods* **17**, 734–740 (2020).
6. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnology* **36**, 460–468 (2018).
7. Saguy, A. *et al.* Deep-ROCS: from speckle patterns to superior-resolved images by deep learning in rotating coherent scattering microscopy. *Optics Express* **29**, 23877–23887 (2021).
8. Dunn, K. W. *et al.* DeepSynth: Three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data. *Scientific reports* **9**, 1–15 (2019).

9. Xie, W., Noble, J. A. & Zisserman, A. Microscopy cell counting and detection with fully convolutional regression networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization* 1–10 (2016) doi:10.1080/21681163.2016.1149104.

10. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. in *Advances in Neural Information Processing Systems* vols 2020-Decem (2020).

11. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C. & Chen, M. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* (2022).

12. Gu, S. *et al.* Vector quantized diffusion model for text-to-image synthesis. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 10696–10706 (2022).

13. Song, J., Meng, C. & Ermon, S. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).

14. Guo, Z. *et al.* Diffusion Models in Bioinformatics: A New Wave of Deep Learning Revolution in Action. *arXiv preprint arXiv:2302.10907* (2023).

15. Yang, L. *et al.* Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys* **56**, 1–39 (2023).

16. Kreis, K., Dockhorn, T., Li, Z. & Zhong, E. Latent Space Diffusion Models of Cryo-EM Structures. *arXiv preprint arXiv:2211.14169* (2022).

17. Waibel, D. J. E., Röoell, E., Rieck, B., Giryes, R. & Marr, C. A diffusion model predicts 3d shapes from 2d microscopy images. *arXiv preprint arXiv:2208.14125* (2022).

18. Igashov, I. *et al.* Equivariant 3d-conditional diffusion models for molecular linker design. *arXiv preprint arXiv:2210.05274* (2022).

19. Chamier, L. von *et al.* ZeroCostDL4Mic: an open platform to use Deep-Learning in Microscopy. *bioRxiv* 2020.03.20.000133 (2020) doi:10.1101/2020.03.20.000133.

20. Nichol, A. Q. & Dhariwal, P. Improved denoising diffusion probabilistic models. in *International Conference on Machine Learning* 8162–8171 (PMLR, 2021).

21. Ouyang, W. *et al.* ShareLoc — an open platform for sharing localization microscopy data. *Nature Methods* **19**, 1331–1333 (2022).

22. Wei OUYANG. Mitochondrial protein TOM22 in COS7 cells. https://doi.org/10.5281/zenodo.5512636.

23. Manish SINGH. Microtubule stained in alexa 647 treated with 1 um of Nocodazole. https://doi.org/10.5281/zenodo.6948554 (2022).

24. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).

25. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793–795 (2006).

26. Hell, S. W. & Wichmann, J. Breaking the Diffraction Resolution Limit By Stimulated-Emission - Stimulated-Emission-Depletion Fluorescence Microscopy. *Optics Letters* **19**, 780–782 (1994).

27. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of Microscopy* **198**, 82–87 (2000).

28. Nieuwenhuizen, R. P. J. *et al.* Measuring image resolution in optical nanoscopy. *Nature Methods* **10**, 557–562 (2013).

29. Shariff, A., Murphy, R. F. & Rohde, G. K. A generative model of microtubule distributions, and indirect estimation of its parameters from fluorescence microscopy images. *Cytometry Part A* **77**, 457–466 (2010).

30. Somepalli, G., Singla, V., Goldblum, M., Geiping, J. & Goldstein, T. Understanding and mitigating copying in diffusion models. *Advances in Neural Information Processing Systems* **36**, 47783–47803 (2023).

31. Somepalli, G., Singla, V., Goldblum, M., Geiping, J. & Goldstein, T. Diffusion art or digital forgery? investigating data replication in diffusion models. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 6048–6058 (2023).