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

Alon Saguy[1], Tav Nahimov[1], Maia Lehrman[1], Onit Alalouf[1], Yoav Shechtman[1,†]

[1] Department of Biomedical Engineering, Technion – Israel Institute of Technology, Haifa, 3200003, Israel

[†] 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 of a diffusion model to super-resolution microscopy images and train it on images from a publicly available database. We show that the generated images resemble experimental images, and that the generated images do not copy existing images from the training set. Additionally, we compare the performance of a deep-learning-based deconvolution method when trained on our generated data versus training on mathematical model based data and show superior reconstruction quality in means of spatial resolution. These findings demonstrate the potential contribution of generative diffusion models for microscopy tasks and pave the way for their future application in this field.

Introduction

In the past decade, deep learning algorithms have been extensively used to solve various challenges in microscopy[1–6]. These algorithms outperformed traditional computer vision methods in terms of reconstruction quality, analysis time and more. However, one major drawback of deep learning solutions is their hunger for data. To train a model, one should manually acquire and annotate hundreds or even thousands of images, a time-consuming process that requires significant resources. Therefore, researchers often train such networks on synthetic data, based on mathematical models that describe the structure of the biological specimen[1,3,7–9]; naturally, for optimal results, the training data should resemble experimental data as much as possible.

Recently, the field of generative models has seen a significant surge in terms of both development and application[10–12]. 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. 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 to the generation of synthetic data, relying on stochastic processes to create data that captures the inherent image statistics present in the training dataset.

Although the impact of diffusion models has been transformative across many sectors, in the field of microscopy imaging, their contribution is invaluable since large high-quality datasets can be extremely challenging to obtain. Not surprisingly, numerous studies have been done to incorporate diffusion models to microscopy images[14], e.g., in the attempt to reconstruct 3D biomolecule structure in Cryo-EM images[15], in prediction of 3D cellular structure out of 2D images[16], in drug molecule design[17], and more. However, to the best of our knowledge, diffusion models have yet to be utilized in a variety of microscopy data analysis domains and specifically in super-resolution microscopy objectives.

Here, we apply a generative diffusion model 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 test our generated data for overfitting, namely, we validate that the network does not memorize images from the training data. Finally, we utilize the generated dataset to train an existing single image super-resolution algorithm and show superior reconstruction quality, compared to training on model-based simulated data.

Results

To explore the potential of diffusion models in super-resolution microscopy, we adapted a previously reported[18] diffusion model to the context of microscopy images (see Methods section for more details). All training data was obtained by single molecule localization microscopy (SMLM) imaging, sourced from a publicly available database (ShareLoc.xyz[19]); we used different images for training/ testing data.

The images generated by our diffusion model qualitatively resemble the training data used for our model training, as can be clearly seen in the examples in figure 1 (see many more examples online[20]).

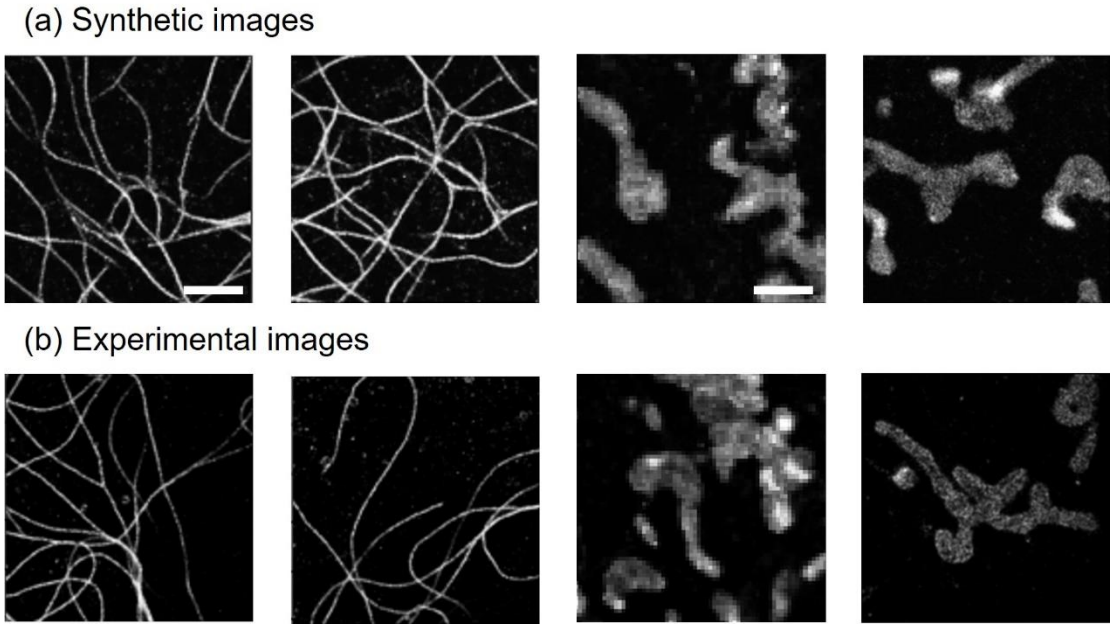

Figure 1: Qualitative comparison of experimental microscopy data versus data generated using our generative diffusion model. (a) Synthetic images generated by our diffusion model. (b) Experimental images used as training data. Scale bars = 2.5 μm.

Importantly, it is crucial to validate that our model does not memorize images, namely, copies existing images from the training set and generates them as network outputs. To do so, we have calculated the normalized cross-correlation between every generated image (including rotated and flipped versions of the images) and the augmented patches used for training. Then, we overlaid the training images with the highest cross-correlation score with the generated ones to validate that our synthetic images are original (figure 2).

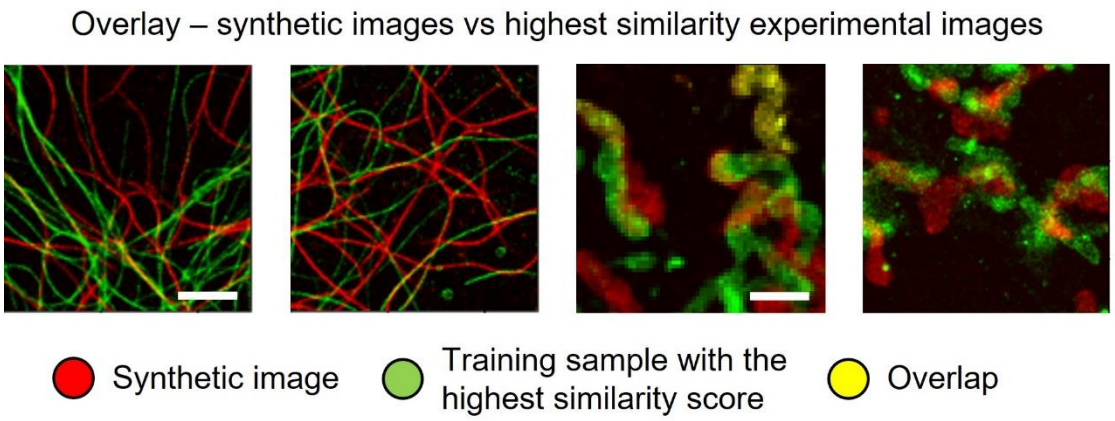

Figure 2: Validating that the diffusion model does not memorize training images as outputs. Overlay between each reconstructed image and the training image with highest resemblance (maximal cross-correlation score). Red marks generated data, green marks the closest training sample, and yellow marks overlap between the two images. Scale bars = 2.5 μm.

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, single-image-based algorithmic super-resolution is an extremely difficult task, and it is not expected to yield a reconstruction at SMLM quality. Here, we use this task as a demonstration of the potential of diffusion model-based data generation. We compare two CARE models 1) trained on synthetic images generated by a mathematical model; 2) trained on images generated by our diffusion model.

Visually, CARE trained on our generated data yielded 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 Image-J[21]. In brief, FRC is a similarity measure that seeks the maximal spatial frequency in which the reconstructed image and the ground truth image agree. The agreement 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 spatial resolution of the reconstructed images, as quantified by FRC, when training on microtubule images generated by the diffusion model was 100 nm, while the spatial resolution obtained when training on synthetic microtubules generated via a mathematical model was 140 nm.

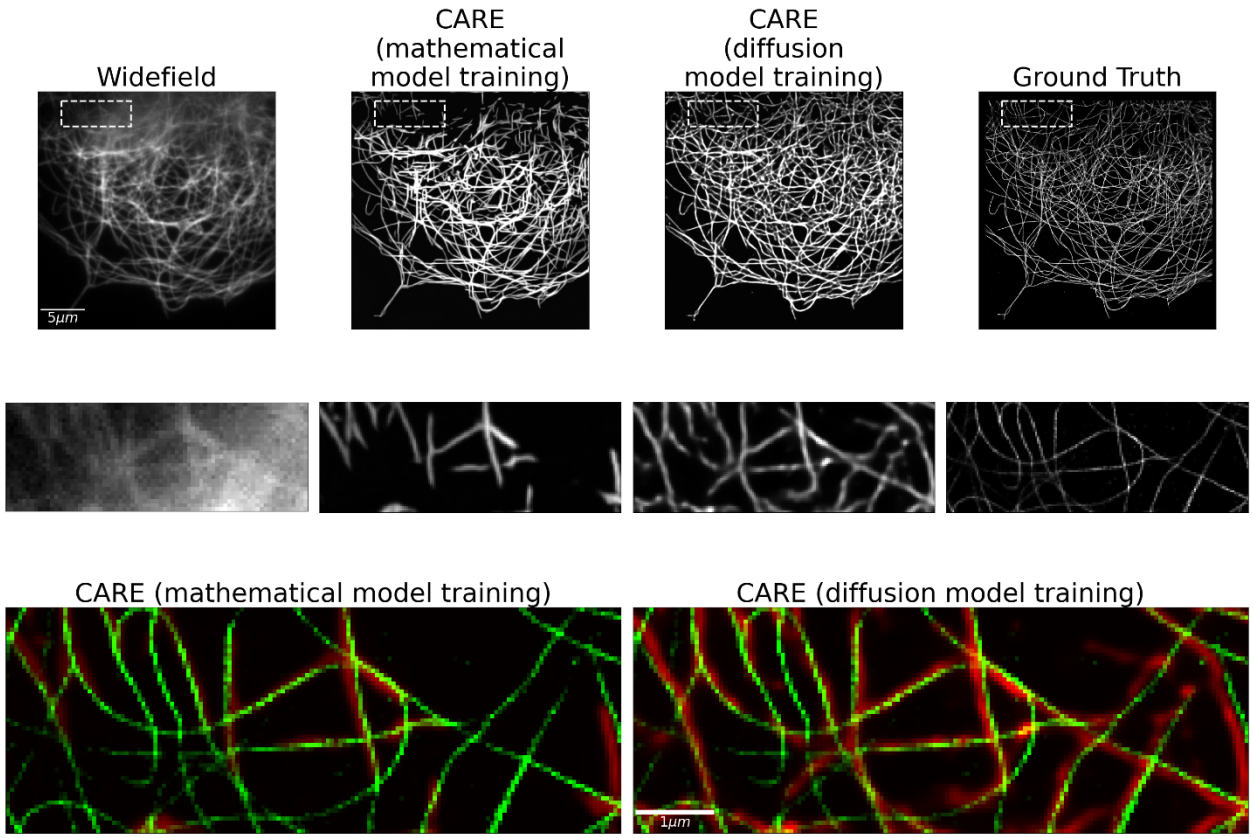

Figure 3: Performance of CARE trained on synthetic microtubule images generated by a mathematical model, vs. training on microtubules generated by our diffusion model. **Upper row**. Left to right: widefield image, CARE reconstruction when trained on mathematical simulations, CARE reconstruction when trained on our synthetic data, and ground truth. Scale bar = 5 μm. **Middle row**. Regions of interest, yellow arrows mark areas in which CARE trained on our data outperformed the previous method. **Bottom row**. 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 μm.

Notably, microtubule data can be simulated with relatively high fidelity by a variety of well-established mathematical models. However, in general, it is not easy to obtain a simple mathematical model describing the shape and characteristics for an arbitrary type of biological specimen. Therefore, the most significant contribution of diffusion model-based data generation would be the ability to generate synthetic data of any biological specimen, even when we cannot describe it mathematically in a convenient manner.

We validate this claim by training our diffusion model on super-resolution mitochondria data available online[19] and testing the network reconstruction on different mitochondria data acquired separately (figure 4). The average spatial resolution we obtained over 3 test experiments was 190 nm when training CARE on a dataset generated by our diffusion model. Notably, unlike microtubule data, for this data type (mitochondria) there is no available mathematical model to use for image generation, therefore training CARE traditionally would require many super-resolution and diffraction limited image pairs.

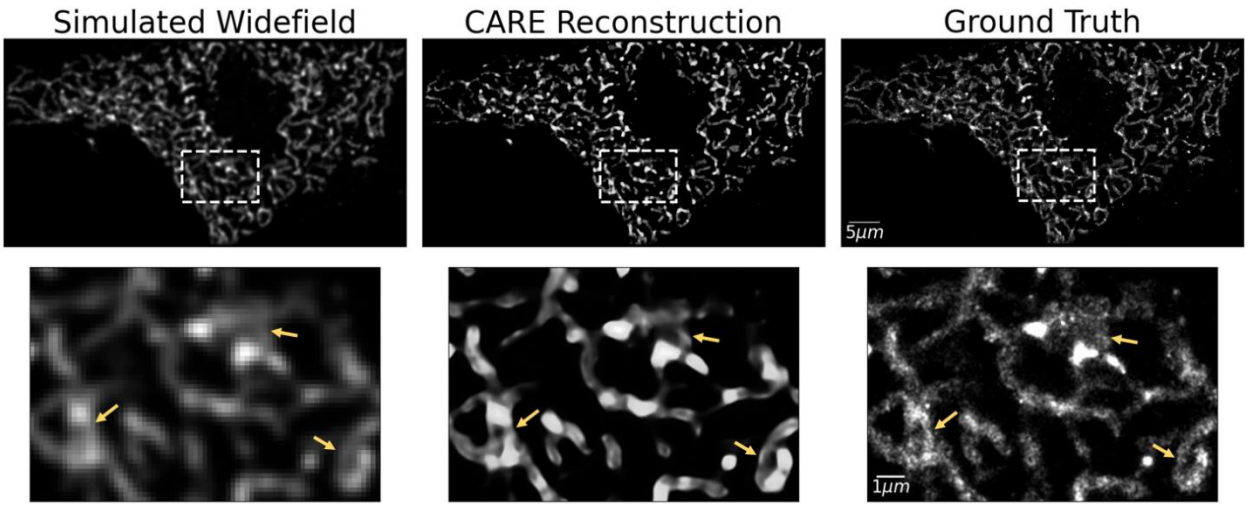

Figure 4: Performance of CARE trained on synthetic mitochondria images generated by a mathematical model, vs. training on mitochondria generated by our diffusion model. **Upper row**. Left to right: widefield image, CARE reconstruction when trained on our synthetic data, and ground truth. Scale bar = 5 μm. **Bottom row**. Region of interest, yellow arrows mark regions in which the resolution improvement in the reconstruction is visible. Scale bar = 1 μm.

Discussion

Creating synthetic images of biological data that are highly realistic and representative of the original data has important implications. For example, high-quality and annotated synthetic datasets can enable efficient algorithm training without the need for extensive real-world data, often a limiting factor due to the cumbersome process of data acquisition and annotation. In this work, we have demonstrated a method to obtain big super-resolution microscopy datasets by relying on relatively low number of annotated images (7 and 8 for the microtubules and mitochondria networks respectively). The contribution of our method is particularly relevant for the general case where there is no simple mathematical model available for synthetic image generation.

Moreover, as shown in this work, training models on more realistic simulated images may lead to enhanced algorithm performance and assist in identification and interpretation of complex biological structures and events. Diffusion models, as demonstrated by our experiment, are likely to be key contributors to this paradigm. Their ability to generate high-quality synthetic images that closely resemble actual microscopy data, even complex structures like microtubule networks, is a testament to their potential.

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 and numerous potential applications exist, such as denoising, multi-image super-resolution, cross-modality imaging, live-cell dynamic imaging and more.

In light of the encouraging results obtained from this study, future research should continue to focus on the further optimization and evaluation of diffusion models in generating more types of synthetic microscopy data, and on finding the applications where these capabilities are most impactful. The

emergence of generative models for microscopy imaging signifies an exciting phase for bio-medical research and holds promising potential for advancements in the near future.

Methods

Mitochondria experiments – sample preparation and imaging

Cover glasses of size 22X22 mm, 170 μm (Deckgläser, No.1.5H) were cleaned in an ultrasonic bath with 5% Contrad 70 (Decon) at 60°C for 30 min, then washed twice with double distilled water (incubated shaking for 10 min each time), incubated shaking in ethanol absolute for 30 min, sterilized with filtered 70% ethanol for 30 min and dried in a biological cabinet. COS7 cells at a concentration of ~20,000 $\frac{cells}{ml}$ in Dulbecco's Modified Eagle Medium (DMEM) with 1 $\frac{g}{l}$ D-glucose (Sartorius, 01-050-1A), supplemented with fetal bovine serum (Biological Industries, 04-007-1A), penicillin-streptomycin (Biological Industries, 03-031-1B) and glutamine (Biological Industries, 03- 020-1B), were grown for 24 hr in a 6-well plate (Thermo Fisher, Nunclon Delta Surface) containing 6 ml of the cell suspension and the cleaned cover glasses, at 37°C, and 5% CO_2. The cells were fixed with 4% paraformaldehyde and 0.2% glutaraldehyde in PBS, pH 6.2, for 60 min, washed and incubated in 0.3M glycine/PBS solution for 10 minutes. The cover glasses were transferred into a clean 6-well plate and incubated in a blocking solution for 2 hr (10% goat serum, 3% BSA, 2.2% glycine, and 0.1% Triton-X in PBS, filtered with 0.45 μm Millex PVDF filter unit). The cells were then immune-stained with 1:500 diluted anti-TOMM20-AF647 antibody (Abcam, ab209606) in the blocking buffer overnight at 4°C and washed five times with PBS. For super-resolution imaging, a PDMS chamber (22x22x3 mm, with a 13x13 mm hole cut in the middle) was attached to the cover glass containing the fixed and stained COS7 cells to create a pool for the blinking buffer. Blinking buffer (50 mM Cysteamine hydrochloride (Sigma, M6500), 20% sodium lactate solution (Sigma, L1375), and 3% OxyFluor (Sigma, SAE0059) in PBS, pH 8-8.5) was added and a cover glass was placed on top while ensuring minimal air bubbles.

We used the Nikon eclipse Ti inverted microscope equipped with a TIRF module (Nikon), oil objective (Nikon, SR Apo TIRF 100x/1.49 Oil WD 0.12), and a multi-bandpass dichroic (Chroma TRF 89902-NK) and an EMCCD camera (Andor, iXON Life) for image acquisition. The sample was illuminated by a 100 mW 640 nm laser and increasing illumination of 405 nm (~2-30 mW) with 50 ms exposure.

Diffusion model architecture and training details

We have adopted the network architecture presented by Nichol, et al[18]. 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, and employed a cosine noise schedule. To train the network, we used 7 super-resolution localization lists of microtubule experiments and 8 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 mictorubule 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 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 Titan RTX GPU. Ultimately, generation of a single super-resolution image takes few seconds depending on the image size, e.g. 8 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 1500 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. *Nat 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. *bioRxiv* 2022.07.01.498428 (2022).

4. Nogin, Y. *et al.* DeepOM: Single-molecule optical genome mapping via deep learning. *bioRxiv preprint* (2022).

5. Nehme, E. *et al.* DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning. *Nat Methods* **17**, 734–740 (2020).

6. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat Biotechnol* **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. *Opt 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. *Sci Rep* **9**, 1–15 (2019).

9. Xie, W., Noble, J. A. & Zisserman, A. Microscopy cell counting and detection with fully convolutional regression networks. *Comput Methods Biomech Biomed Eng Imaging Vis* **6**, 283–292 (2018).

10. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Adv Neural Inf Process Syst* **33**, 6840–6851 (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. Kreis, K., Dockhorn, T., Li, Z. & Zhong, E. Latent Space Diffusion Models of Cryo-EM Structures. *arXiv preprint arXiv:2211.14169* (2022).

16. 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).

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

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

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

20. https://www.clever-microscopy.com/this-microtubule-is-not-real. (2023).

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