# Task-Assisted GAN for Resolution Enhancement and Modality Translation in Fluorescence Microscopy

Catherine Bouchard[1,2,3], Theresa Wiesner[2,3,4], Andréanne Deschênes[2,3,4], Flavie Lavoie-Cardinal[2,3,5*], and Christian Gagné[1,2*]

[1]Département de génie électrique et de génie informatique, Université Laval, Québec (QC), Canada

[2]Institut Intelligence et Données (IID), Université Laval, Québec (QC), Canada

[3]CERVO Brain research centre, Québec (QC), Canada

[4]Département de biochimie, microbiologie et bio-informatique, Université Laval, Québec (QC), Canada

[5]Département de psychiatrie et de neurosciences, Université Laval, Québec (QC), Canada

[*]corresponding authors: christian.gagne@gel.ulaval.ca

flavie.lavoie-cardinal@cervo.ulaval.ca

July 19, 2021

## Abstract

We introduce a deep learning model for resolution enhancement and prediction of super-resolved biological structures, which is based on a Generative Adversarial Network (GAN) assisted by a complementary segmentation task. It is applied to predict biological nanostructures from diffraction-limited images and to guide microscopists for quantitative fixed- and live-cell STimulated Emission Depletion (STED) microscopy. More specifically, we show that the use of a complementary segmentation task improves the accuracy of the predicted nanostructures over state-of-the art resolution enhancement generative approaches, allowing quantitative analysis of the sub-diffraction structures in the resulting generated images.

## Main

Some of the forefront progresses induced by deep learning lie in generating synthetic images indistinguishable from real ones, mainly through the introduction of Generative Adversarial Networks (GAN) [1]. Beyond the prevalent and visually-appealing application of such generative models for creating entirely new human faces, animals or objects [2], their uses also extend to image-to-image translation [3]. For natural images, these methods can improve the spatial resolution by increasing the number of pixels and sharpening finer details [4]. For biomedical imaging, they can be applied to improve the signal-to-noise ratio of a given imaging modality [5]. But in optical microscopy, the spatial resolution is limited by the diffraction barrier that cannot be surpassed by adding pixels or improving the contrast [6]. In this context, it is challenging to generate *new* sub-diffraction structures that are not optically resolved in the input image [7]. For quantitative microscopy image analysis of nanoscopic

structures in biological samples, the super-resolution method needs to be reliable at generating sub-diffraction structures of interest. New methods for deep learning-based super-resolution in microscopy have been proposed [8–12], but concerns and skepticism arise regarding their applicability to characterize biological structures at the nanoscale [7,13].

Generative super-resolution methods for microscopy are trained by making a direct comparison between the generated and the ground truth images using pixel-wise metrics (e.g. mean squared error [9], absolute error [8,10], structural similarity index [11,12]) (Suppl. fig. S2a). These approaches perform well to remove blurring artifacts or noise from images of simple structures such as microtubules [8] and granules [10]. Yet, the generic models fall short at generating details specifically relevant to the biological interpretation (e.g., Suppl. fig. S1). Guiding the generative model with a more specific task helps steer it to generate images with the sought after relevant information. Recent works using GANs for natural images use one or multiple tasks to provide spatial guidance to the generator, ensuring that the generated images are consistent with the target annotations [14–16].

We propose a resolution-enhancing approach relying on a task-assisted GAN (TA-GAN) to ensure accurate generation of biological nanostructures of interest. The TA-GAN is optimized to perform well over a complementary segmentation task of the nanoscopic structures that could not be performed on diffraction-limited images. For this, it relies on the output of a complementary network used to compare the real and the generated images according to the task-specific loss (Fig. 1a and Suppl. fig. S2b,c). The TA-GAN not only generates realistic images, but is guided by a criterion directly aiming at accurately generating the biological structures of interest. We apply the TA-GAN method to STimulated Emission Depletion (STED) microscopy of fixed and living neurons. Our results demonstrate how it improves characterization of protein organization at the nanoscale in comparison to other GAN-based super-resolution approaches. Specifically, our proposal is useful to 1) provide supplementary information from diffraction-limited images (e.g., for guiding quantitative analysis of nanostructures), 2) ease switching between imaging modalities or biological contexts by generating new datasets while eliminating the annotation burden, and 3) improve the efficiency of live-cell STED imaging.

To validate the proposed method we first use pairs of confocal and STED images of the F-actin cytoskeleton, more specifically the F-actin rings in axons of fixed hippocampal neurons [17,18] (Suppl. fig. S3). We measure a significant increase in segmentation performance when using the complementary task compared to the standard conditional GAN architecture (Fig. 1b and c). We then evaluate the TA-GAN performance on a more complex F-actin cytoskeleton in dendrites, where the complementary task is the semantic segmentation of dendritic F-actin longitudinal fibers and periodical rings. The proportion of these nanostructures in dendrites varies depending on neuronal activity levels and cannot be determined from confocal images alone [17]. Compared to real STED images, quantitative analysis of synthetic STED images generated with the TA-GAN shows similar segmentation masks and proportions for F-actin rings and fibers (Fig. 1d and e, Suppl. Fig. S4). In both the real STED images and the ones generated by the TA-GAN, the area of the periodical lattice significantly decreases as the neuronal activity increases, while the opposite is observed for fibers (Fig. 1e). A similar conclusion cannot be drawn from images generated by a standard conditional GAN architecture since the rings and fibers generated are not realistic nor precise enough to be segmented by either an expert or a segmentation network (Suppl. Fig. S1). This experiment highlights the reliability of the generator trained with a loss function built specifically to retain the information of interest; here, the distribution of dendritic F-actin rings and fibers. We next evaluate the TA-GAN approach on a dataset of immunostained synaptic protein pairs (PSD95 - Homer 1c and Bassoon - PSD95) in fixed hippocampal neurons [19] (Suppl. fig. S5). For this dataset, we rely on automatically generated wavelet segmentation masks [19,20] for the complementary task, thereby eliminating the need for manual annotation (Suppl. fig. S6 and S7).

For many applications of STED microscopy in fixed cells, imaging speed and photobleaching are not necessarily major concerns that would motivate the generation of synthetic images from confocal acquisitions over real STED acquisitions. For live-cell imaging however, super-resolution microscopy can induce phototoxicity and photobleaching effects due to repeated light

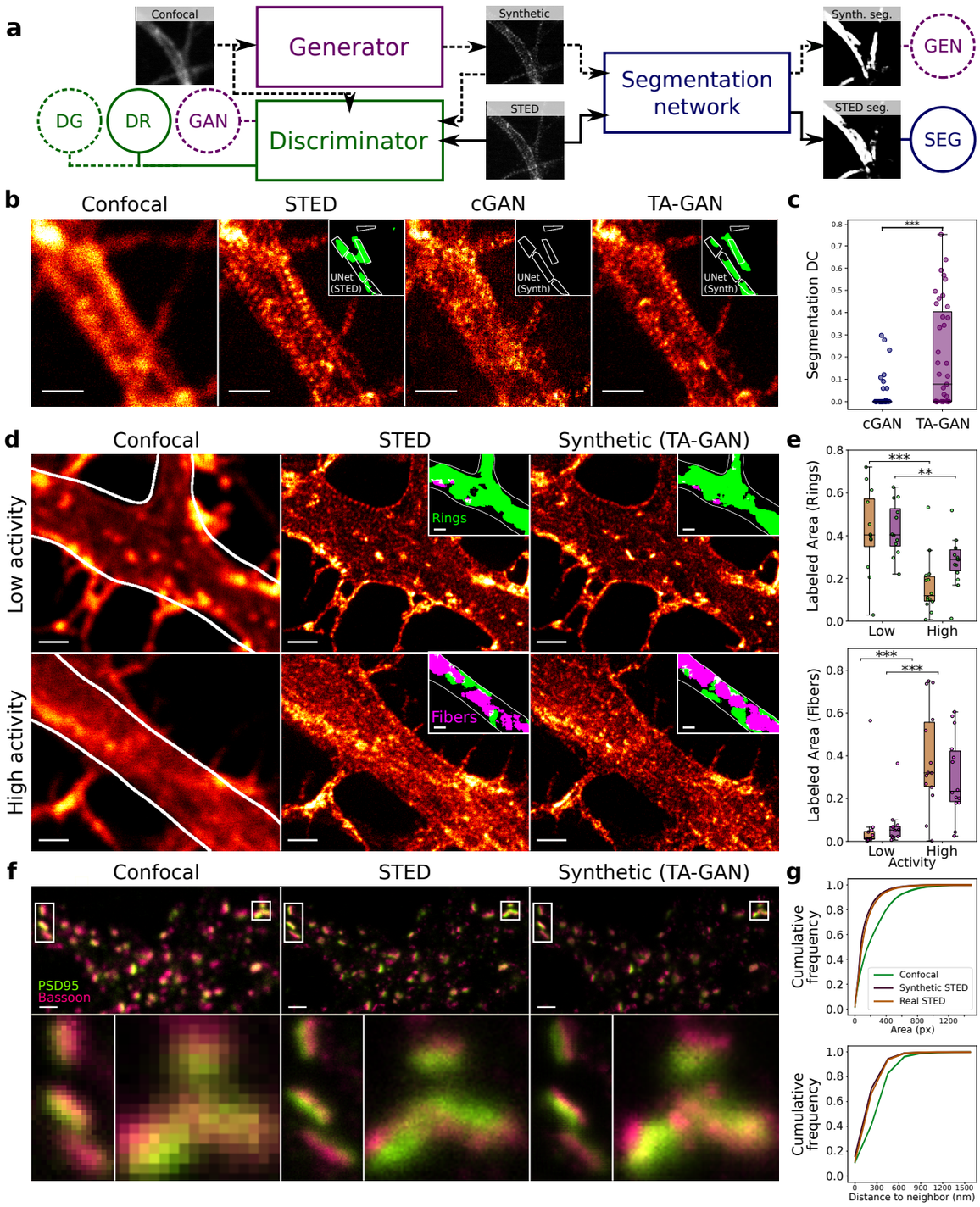


Figure 1: **TA-GAN applied to quantitative analysis of nanostructures in synthetic STED images. a,** Architecture of the TA-GAN. Backpropagation of the losses (circles) to the corresponding networks (rectangles) is shown for the generator (violet, 9 blocks ResNet[21]) (Suppl. fig. S8), the discriminator (green, PatchGAN[3]), and the segmentation network (blue, 6 blocks ResNet). Dashed line: flow of the input confocal. Solid line: flow of the gound truth STED image. **b,** Synthetic and real STED images of the axonal F-Actin periodical lattice in neurons. Insets show the segmentation masks (green) of the F-actin rings obtained from a segmentation CNN trained on real STED images[17] together with the outline of the manual expert annotations (white line). From left to right: input confocal, corresponding ground truth real STED, synthetic STED generated using a standard cGAN architecture[3], and using our TA-GAN architecture. **c,** The Dice coefficient metric comparing the segmentation maps of the generated and real STED images shows a significant ($p \sim 10^{-5}$) improvement when using TA-GAN compared to cGAN (see Suppl. fig. S4). Statistical significance assessed with an independent samples t-test. **d,** Confocal, ground truth STED, and synthetic STED images showing dendrites at low (top) and high (bottom) neuronal activity level (see Methods Sec. 4). Insets show the regions identified as rings (green) and fibers (magenta) by the segmentation CNN trained on real STED images[17]. **e,** The measured proportion of F-actin rings in dendrites is significantly larger at low neuronal activity in both real (yellow) and synthetic (purple) STED images (real STED: $p = 0.0007$, synthetic STED: $p = 0.004$). The opposite is observed for F-Actin fibers, with an increase observed with increasing neuronal activity for both the real STED images ($p = 0.0006$) and the synthetic STED images ($p = 0.0009$). Statistical significance is computed with an independent samples t-test (*** $P < 0.001$, ** $P < 0.01$). **f,** Generated upsampled two-color STED images showing the synaptic protein pair PSD95 (green) and Bassoon (pink). Confocal, synthetic and STED sub-regions showing the similarity between the resolution, synaptic protein organization, and cluster shape of the real and synthetic STED images. The crops in the second row are $1 \times 2\,\mu m$ (left) and $1 \times 1\,\mu m$ (right). **g,** Cumulative frequency plots of the area of the clusters and the distance between clusters for Bassoon, computed with pySODA[22] using the confocal, synthetic STED and real STED images (n = 12 images) (Suppl. fig. S6 and S7). (All scale bars: 1 µm).

exposure. Consequently, it can hinder quantitative image analysis and long term measurements of dynamic nanostructures in biological samples (Suppl. fig. S9). To overcome this difficulty, the TA-GAN method is adapted for live-cell imaging of the F-Actin cytoskeleton, benefiting from the reduced photobleaching from confocal acquisitions while still obtaining super-resolved information from the synthetic STED images.

Although the structures in living and fixed cells are very similar, the images still differ too much to be directly segmented using a network trained on images of fixed cells (Fig. 2b). We therefore address how a suitable annotated dataset can be generated with no additional expert annotations. For this, we adapt the TA-GAN method to unpaired images: STED images of fixed and live cells. We developed a *fixed to live* ($F \rightarrow L$) modality translation framework to convert already annotated real fixed-cell STED images of F-Actin into synthetic live-cell images of the same protein (Methods sec. 3.1, Fig. 2a and Suppl. fig. S10). The generated live-cell images can be associated with the annotations from the corresponding real fixed cell image. The translated annotated $F \rightarrow L$ images are used to train a segmentation network for F-actin rings in live-cell images (Fig. 2b). The segmentation network for live cells is then used to train the TA-GAN model for live-cell imaging (Methods sec. 3 and Suppl. fig. S11).

The trained generator network from the TA-GAN model for live-cell imaging is included in the acquisition process of the STED microscope (Fig. 2c). The live F-actin nanostructures are first imaged in confocal and STED mode for low neuronal activity (high $Mg^{2+}$/low $CaMg^{2+}$). The neurons are next stimulated using a chemical long term potentiation (cLTP) stimulation (adapted from[23]), and imaged every subsequent minute in confocal mode (Methods sec. 4). For each confocal image, the region of lowest confidence inferred by the generator is acquired with the STED microscope. It is transferred (along with the confocal image of the full field of view) to the generator to obtain a corresponding synthetic image. The generated full field of view image includes the sub-region that was acquired using the STED modality. This approach guides the acquisition process to only perform super-resolution imaging on the sub-regions with the lowest confidence. Super-resolution imaging of other regions with high confidence is spared from the imaging process to minimize light exposition on the sample and photobleaching (Suppl. fig. S12). To further limit photobleaching, a central region of interest is defined where no STED crops are acquired over the whole imaging loop. A final confocal and STED image pair is acquired at the end of the sequence.

This method is demonstrated over a biological process previously imaged and analyzed on fixed cells that could not be reproduced in live cells due to photobleaching effects: the activity-dependent remodelling of the F-actin based sub-membrane lattice (F-actin rings) into longitudinal fibers in dendrites[17]. The TA-GAN assisted acquisition enables the generation of synthetic STED images of the F-Actin nanostructures for the full field of view using the confocal images and STED sub-regions acquired at each frame (Suppl. fig. S13). Using a segmentation CNN, the proportion of F-actin rings and fibers is monitored using the generated synthetic STED images. It strongly minimizes photobleaching and consequently maintains sufficient contrast and image quality over the whole imaging process (Fig. 2 b, d & e, and Suppl. fig. S14). The same dynamic transformation cannot be observed by imaging the whole field of view repeatedly with STED microscopy, since the fluorescence of the SiR-Actin dye rapidly decays, impairing the detection and quantification of the F-actin rings and fibers (Suppl. fig. S9). The dynamic remodelling of the F-actin lattice can be quantified over time and our approach allows the visualization of the dynamic processes occurring between the initial and final STED acquisitions (Fig. 2d).

We show how the TA-GAN method can be used for quantitative analysis of sub-diffraction elements from diffraction-limited acquisitions, to minimize photobleaching effects for live-cell imaging and to recycle datasets and annotations to train networks on new imaging modalities. Quantitative analysis of elements smaller than the diffraction limit such as the spacing between synaptic protein clusters and the F-actin periodical lattice can be performed from diffraction-limited images of fixed and living cells. TA-GAN assisted live-cell imaging also allows the monitoring of a dynamic process at the nanoscale, while strongly limiting negative effects related to prolonged high-intensity light exposure. Such an approach based on the prediction

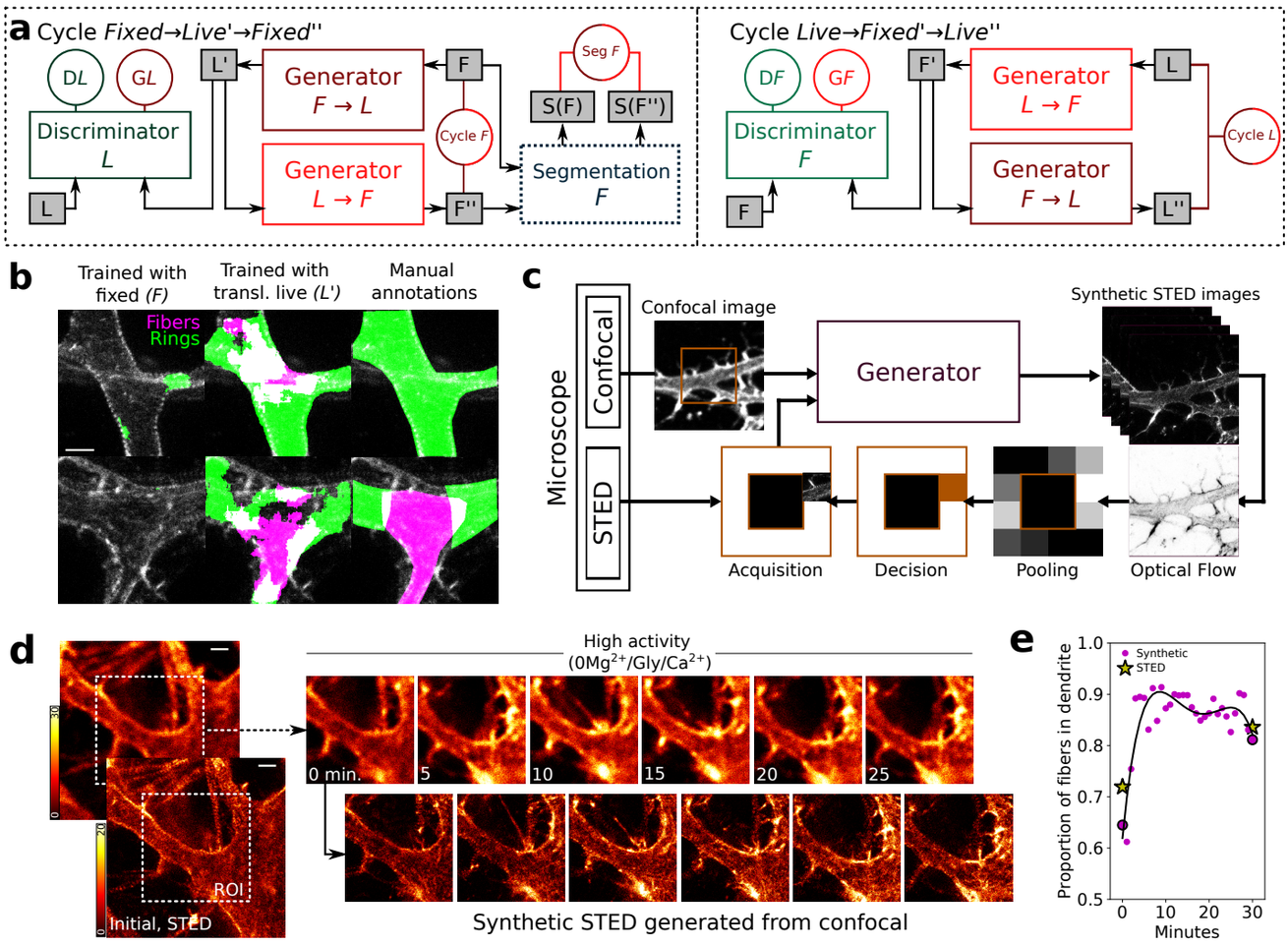


Figure 2: **Cycle-GAN applied to live-cell imaging and modality translation. a,** The architecture used is a Cycle-GAN[24] with a segmentation task used to optimize the translation networks as in TA-GAN. Adding the segmentation loss (*Seg F*) to the optimization process ensures that the structures of interest - here F-actin rings and fibers - are preserved throughout the whole cycle. Since no expert annotations were available to train a segmentation network on live-cell images prior to this step, the segmentation loss is only applied on fixed-cell images, but backpropagated through both translators to balance the number of optimization steps between both networks. Once trained, the *fixed-to-live* translator is used in inference to create a complete dataset of live-cell synthetic images. This new dataset, along with the segmentation labels from the initial fixed-cell dataset, is used to train a segmentation network from scratch. **b,** Real STED acquisitions of live-cell images and the segmentation of fibers (magenta) and rings (green) as computed by a segmentation network trained only on images of fixed cells and on synthetic images of live cells. The U-Net trained on fixed-cell images does not recognize the rings nor the fibers, whereas the U-Net trained on the synthetically translated live-cell images does. **c,** Image acquisition workflow integrating the generator to super-resolve the acquired confocal and to decide the region of lowest confidence that needs to be imaged with the STED modality. **d,** Timelapse imaging of F-actin nanostructures in living neurons using the TA-GAN assisted acquisition workflow. The first row shows confocal acquisitions in the central ROI, the second row shows corresponding generated STED images. Scalebar: 1 µm. All images from a given row are normalized to the same intensity value to show the effect of the acquisitions on the loss of fluorescence. **e,** The segmentation network trained on synthetic live-cell images is used to compute the proportion of fibers in the dendrite for each synthetic image. In this specific case, there is an abrupt activity-dependent increase over the first few minutes that then reaches a plateau (See Suppl. fig. S14 for additional examples).

of biological nanostructures, that cannot be directly inferred from diffraction-limited images, could become an important guiding tool towards the design of intelligent super-resolution microscopes that provide the capability to specifically decide when and where to image specific regions depending on the predictions of the TA-GAN. The results suggest the possibility for a neural network to extract information when it is trained in a fashion that encourages the decoding of specific biologically relevant information to improve the efficiency of super-resolution microscopy image acquisition, annotation and analysis.

# Acknowledgments

Francine Nault and Sarah Pensivy for neuronal cell culture. Anthony Bilodeau for the design of the website. Annette Schwerdtfeger for proofreading the manuscript. Funding was provided by grants from the Natural Sciences and Engineering Research Council of Canada (F.L.C. and C.G.), the CERVO Foundation (F.L.C.), and the Neuronex Initiative (National Science Foundation, Fond de recherche du Québec - Santé) (F.L.C.). C.G. is a CIFAR Canada AI Chair and F.L.C. is a Canada Research Chair Tier II. C.B. is supported by scholarships from the Fonds de Recherche Nature et Technologie (FRQNT) Quebec, from the FRQNT strategic cluster UNIQUE, and from the Natural Sciences and Engineering Research Council (NSERC), and a Leadership and Scientific Engagement Award from Université Laval. T.W. is supported by postdoctoral research funding from the FRQNT strategic cluster UNIQUE.

# Author contributions

C.B., F.L.C. and C.G. designed the method. C.B. implemented the live imaging automatic acquisitions, performed all deep learning experiments, and analysed the results. A.D. and T.W. performed the live imaging experiments. C.B., A.D., T.W., F.L.C. and C.G. wrote the manuscript.

# Competing interests

The authors declare no competing interests.

# Data and code availability

All of the datasets used to train and test the TA-GAN model and the baselines are available to download at `https://s3.valeria.science/flclab-tagan/index.html`. All of the programs, trained models and sample images needed to test the TA-GAN architecture are available at `https://github.com/FLClab/TA-GAN`. Instructions on how to adapt the dataloaders and networks to new images are also provided.

# References

[1] Goodfellow, I. *et al.* Generative adversarial nets. *Advances in neural information processing systems* (2014).

[2] Choi, Y., Uh, Y., Yoo, J. & Ha, J.-W. StarGAN v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8188–8197 (2020).

[3] Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1125–1134 (2017).

[4] Yang, W. *et al.* Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia* **21**, 3106–3121 (2019).

[5] Kaji, S. & Kida, S. Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiological physics and technology* **12**, 235–248 (2019).

[6] Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).

[7] Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature Methods* **16**, 1215–1225 (2019).

[8] Chen, J. *et al.* Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nature Methods* **18**, 678–687 (2021).

[9] Fang, L. *et al.* Deep learning-based point-scanning super-resolution imaging. *Nature Methods* **18**, 406–416 (2021).

[10] Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods* **15**, 1090–1097 (2018).

[11] Qiao, C. *et al.* Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature Methods* **18**, 194–202 (2021).

[12] Wang, H. *et al.* Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods* **16**, 103–110 (2019).

[13] Hoffman, D. P., Slavitt, I. & Fitzpatrick, C. A. The promise and peril of deep learning in microscopy. *Nature Methods* **18**, 131–132 (2021).

[14] Zhang, Y. *et al.* DatasetGAN: Efficient labeled data factory with minimal human effort. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10145–10155 (2021).

[15] Jiang, S., Tao, Z. & Fu, Y. Segmentation guided image-to-image translation with adversarial networks. In *IEEE International Conference on Automatic Face & Gesture Recognition*, 1–7 (2019).

[16] Zhang, C. *et al.* Multitask GANs for semantic segmentation and depth completion with cycle consistency. *IEEE Transactions on Neural Networks and Learning Systems* (2021).

[17] Lavoie-Cardinal, F. *et al.* Neuronal activity remodels the F-actin based submembrane lattice in dendrites but not axons of hippocampal neurons. *Scientific reports* **10**, 1–17 (2020).

[18] Xu, K., Zhong, G. & Zhuang, X. Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. *Science* **339**, 452–456 (2013).

[19] Wiesner, T. *et al.* Activity-dependent remodeling of synaptic protein organization revealed by high throughput analysis of STED nanoscopy images. *Frontiers in neural circuits* **14** (2020).

[20] Olivo-Marin, J.-C. Extraction of spots in biological images using multiscale products. *Pattern recognition* **35**, 1989–1996 (2002).

[21] He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (2016).

[22] Lagache, T. *et al.* Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. *Nature Communications* **9**, 1–15 (2018).

[23] Lu, W.-Y. *et al.* Activation of synaptic NMDA receptors induces membrane insertion of new AMPA receptors and LTP in cultured hippocampal neurons. *Neuron* **29**, 243–254 (2001).

[24] Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, 2223–2232 (2017).

[25] Cherian, A. & Sullivan, A. Sem-GAN: semantically-consistent image-to-image translation. In *IEEE Winter Conference on Applications of Computer Vision*, 1797–1806 (2019).

[26] Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

[27] Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE transactions on image processing* **26**, 3142–3155 (2017).

[28] Zhang, Y. *et al.* A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11710–11718 (2019).

[29] Huo, Y. *et al.* Synseg-net: Synthetic segmentation without target modality ground truth. *IEEE transactions on medical imaging* **38**, 1016–1025 (2018).

[30] Li, C. & Wand, M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *European conference on computer vision*, 702–716 (Springer, 2016).

[31] Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241 (2015).

[32] Nault, F. & De Koninck, P. Dissociated hippocampal cultures. In *Protocols for Neural Cell Culture*, 137–159 (2009).

[33] Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference*, 11 – 15 (2008).

[34] Dijkstra, E. W. A note on two problems in connexion with graphs. *Numerische mathematik* **1**, 269–271 (1959).

# Methods

## 1 TA-GAN model

The Task-Assisted Generative Adversarial Network (TA-GAN) is a conditional GAN model with an added segmentation network (Fig. 1a)[25]. The segmentation network is used to compute a generation loss that is defined by an analysis task of biological interest. Three networks are trained simultaneously: the generator, the discriminator and the segmentation network. A random initialization is used for all networks. The model is trained from pairs of confocal and STED images

acquired simultaneously. The generator translates the confocal image into its synthetic STED version. The synthetic and the real STED are processed independently by the segmentation network, which outputs the respective segmentation maps of the biological structure of interest. The generation loss (*GEN*), which backpropagated to the generator, is defined as the mean squared difference between the segmentation maps of the synthetic images and the manual weak labels (Fig. 1a and Suppl. fig. S2). The segmentation loss (*SEG*), which optimizes the segmentation network, is defined as the mean squared difference between the segmentation maps of the real STED images and the labels. Both the real and synthetic images are also independently passed through the discriminator along with the initial confocal image to be classified as either real or synthetic. Three losses result from this classification: *DG* (correct classification of the generated images as synthetic), *DR* (correct classification of the real images as real) and *GAN* (misclassification of the generated images as real) (Fig. 1a). *DG* and *DR* are used to optimize the discriminator, while *GAN* optimizes the generator.

TA-GAN can be trained using one of two modes: 1) training the segmentation network from scratch (Suppl. fig. S2b), or 2) using a pre-trained and frozen segmentation task (Suppl. fig. S2c). Training from scratch is simpler as the whole model is optimized through a single training phase (Fig. 1a). Starting from a pre-trained segmentation network leads to faster training of the generator since the gradients do not need to be computed through the frozen segmentation network. For all training experiments using mode 1), the generator and segmentation networks were residual networks[21]. The generator had 9 residual blocks (Suppl. fig. S8) and the segmentation network had 6 residual blocks.

## 1.1 Single nanostructure generation of axonal F-actin rings

The axonal F-actin rings dataset from Lavoie-Cardinal *et al.*[17] was split into 377 images for training, 56 for validation and 52 for testing, all 224×224 pixels. Each image had been acquired in confocal and STED modalities that were spatially well aligned. The axonal F-actin periodical lattice (F-actin rings) was manually segmented from the STED images using bounding boxes weak labels. These annotations were used to train the segmentation network. Automated segmentation of this nanostructure served as the segmentation task to compute the generation loss.

The performance of the generator for the F-actin rings dataset was assessed using the task it was trained with: the segmentation of F-actin axonal rings. The 52 confocal images kept for testing were passed through the generator to output as many synthetic STED images. The mean squared error with the ground truth STED image was computed for each generated image (Suppl. fig. S4b). All generated images were also passed through a segmentation network for axonal F-actin rings that was trained on real STED images[17]. This network is available at `https://github.com/FLClab/STEDActinFCN`. The segmentation maps were compared to the bounding box labels using the Dice coefficient metric (Suppl. fig. S4d). The statistical significance results reported in Fig. 1c and S4b,d were computed from a two-sided t-test over independent samples.

## 1.2 Dual nanostructure generation of dendritic F-actin rings and fibers

The dataset of dendritic F-actin rings and fibers images from Lavoie-Cardinal *et al.*[17] was split into 304 images for training and 54 for validation. Fixation, immunostaining and STED imaging are described in Lavoie-Cardinal *et al.*[17]. Prior to training, these images were cropped using a sliding window of size 224x224. If less than 1% of the pixels of the crop were identified by the bounding boxes as containing a structure of interest (rings and/or fibers), the crop was discarded. This operation resulted in 4,331 crops for training and 659 crops for validation. Each image data had a confocal image, a spatially-aligned STED image, and segmentation annotations for rings and for fibers. The output of the segmentation network was a two-channel segmentation map, as were the bounding box annotations. The second column of Table 1 presents all hyperparameters used

to train this model.

The TA-GAN model for the dendritic F-actin rings and fibers dataset was evaluated using the segmentation of F-actin rings and fibers in the synthetic images generated from the confocal images. We used the segmentation network published in Lavoie-Cardinal *et al.*[17] and available at `https://github.com/FLClab/STEDActinFCN`, trained and tested using the same images and segmentation labels as TA-GAN. We reproduced their findings regarding the remodelling of F-actin rings into longitudinal fibers when neuronal activity increases using the same 26 testing images. Two conditions were compared, one where the neuronal activity is low ($n_{lowactivity} = 12$), and one where it is high ($n_{highactivity} = 14$). The statistical significance results reported in Fig. 1e were computed from a two-sided t-test over independent samples.

## 1.3 Upsampling and segmentation of synaptic protein clusters

A fraction of the dataset of synaptic proteins from Wiesner *et al.*[19] was split into 63 images for training (28 PSD95/Homer, 35 PSD95/Basssoon) and 19 for validation (7 PSD95/Homer and 12 PSD95/Bassoon). The confocal images were acquired using a pixel size of 60 nm, while STED images were acquired with a pixel size of 15 nm. To facilitate the loss computation between the confocal and STED images, the confocal images were rescaled by a factor of 4 using nearest-neighbor interpolation. The field of view of these images measured up to 100 µm which resulted in long time lapses between the acquisition of the confocal and STED images, creating a perceptible shift between the two. To ensure the network was not trained on shifted pairs, the confocal crop was selected by matching it to the STED crop. Squared crops of size 512 pixels were selected from the STED image with a sliding window with a stride of 256 pixels. If the mean value of the photon count was below 0.5, the crop was discarded. At each iteration, a 256 pixels square crop centered on the same coordinates was taken from the confocal image and matched to the larger STED region (Fig. S15) using the template matching library from cv2[26]. The translations that maximized the match between the STED and confocal crops were applied to the STED region and the segmentation masks, resulting in 6-channel (2 confocal, 2 STED and 2 masks) images where all channels were spatially aligned (Suppl. fig. S15). This process of rescaling, cropping and registering resulted in 6,046 crops for training and 1,830 for validation, all 256×256 pixels.

The clusters were automatically segmented from the non-cropped images using wavelet transform decomposition[20]. Contrary to Wiesner *et al.*[19], no segmented clusters were discarded based on size or position, following the intuition that even the smallest outliers should be generated. The resulting segmentation masks (examples of which are shown in Suppl. fig. S6 and S7) were used to train the complementary network for this dataset. The third column of Table 1 presents all the hyperparameters used to train this model.

The TA-GAN for synaptic proteins was trained and tested on the same pairs of pre- and post-synaptic proteins: PSD95/Bassoon and PSD95/Homer. The testing dataset contains 19 images (12 PSD95/Bassoon, 7 PSD95/Homer) of different sizes. The pixel size is the same as for the training images: confocal images have been acquired with 60 nm pixels and STED images with 15 nm pixels. The area, perimeter, eccentricity and distance between neighbors were computed using SODA-analysis[22] for the protein clusters from the confocal images, STED images and generated STED images (Suppl. fig. S6 and S7).

| **Hyperparameters** | | **F-actin rings** [17] | **Dendritic F-actin** [17] | **Synaptic Proteins** [19] | **Live F-actin** |
|---|---|---|---|---|---|
| **Training crops** | **#** | 377 | 4,331 | 6,046 | 753 |
| | **Size (px)** | 224 x 224 | 224 x 224 | 256 x 256 | variable |
| **Validation crops** | **#** | 56 | 659 | 1,830 | 47 |
| | **Size (px)** | 224 x 224 | 224 x 224 | 256 x 256 | variable |
| **Assisting Task** | **Task** | Segmentation of actin rings | Segmentation of actin rings and fibers | Segmentation of synaptic clusters | Segmentation of actin rings and fibers |
| | **Labels** | Bounding boxes | Bounding boxes | Wavelet seg. | N/A |
| | **Pretrain.** | No | No | No | Yes |
| **Networks** | **G** | ResNet 9-blocks | ResNet 9-blocks | ResNet 9-blocks | ResNet 9-blocks |
| | **S** | ResNet 6-blocks | ResNet 6-blocks | ResNet 6-blocks | U-Net 128 |
| | **D** | PatchGAN | PatchGAN | PatchGAN | PatchGAN |
| **Batch size** | | 8 | 32 | 32 | 16 |
| **Learning rate** | | 0.0002 | 0.0002 | 0.0002 | 0.0002 |
| **Lambda** | **GAN** | 1 | 1 | 1 | 1 |
| | **SEG** | 10 | 1 | 1 | 10 |
| **Data augmentation** | **Methods** | flip, rotation, crop | flip, rotation, crop | flip, rotation, crop | flip, rotation, crop |
| | **Crop size** | 128 | 128 | 128 | 256 |
| **Epochs** | **#** | 1000 | 500 | 1000 | 5000 |

Table 1: Hyperparameters used to train the TA-GAN model on all four datasets presented.

# 2 Training the baselines

## 2.1 DnCNN

The pretrained version of DnCNN[27] available at `https://github.com/yinhaoz/denoising-fluorescence` was directly applied to our confocal images. A version of the network trained on the fluorescence microscopy denoising dataset[28] was used. The network was not trained on our specific images. It was included as a baseline to show how the confocal to STED transformation of F-actin nanostructures was not a denoising task, but a structure generation task.

## 2.2 CARE

Content-Aware Image REstoration (CARE)[10] was implemented from the public GitHub repository (`https://github.com/CSBDeep/CSBDeep`). The residual U-Net generator was optimized from scratch with the same training and validation images as the TA-GAN for dual nanostructure generation of dendritic F-actin rings and fibers. All default hyperparameters were used and the model was trained for 100 epochs using a mean absolute error loss. The epoch that reached the lowest validation loss was kept for testing.

## 2.3 3D-RCAN

Three-dimensional residual channel attention networks (3D-RCAN)[8] was implemented with Tensorflow and Keras. The code was taken from the publicly available GitHub repository (`https://github.com/AiviaCommunity/3D-RCAN`). The model was trained on the same dataset of dendritic F-actin rings and fibers as the TA-GAN. All default hyperparameters were used and the model was trained over 300 epochs to insure convergence of the validation loss. The model reaching the lowest validation loss (epoch 221) was used for testing.

### 2.4 Pix2Pix

Pix2pix[3] was implemented with Pytorch from the publicly available GitHub repository (`https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix`). For each experiment, the same hyperparameters and datasets as for the TA-GAN were used for training (Table 1), replacing only the generation loss with a pixel-wise mean squared error loss between the ground truth and generated STED images (Suppl. fig. S2a). The results from this baseline (cGAN) are compared to the TA-GAN for the generation of axonal F-actin rings in Fig. 1c and for the generation of dendritic F-actin rings and fibers in Suppl. fig. S4.

## 3 TA-GAN model for live-cells

### 3.1 Translation from fixed- to live-cell images

Task-related annotations were required to apply the TA-GAN approach on a new modality or structure of interest. The annotations could be generated either manually (e.g. bounding boxes for the F-Actin dataset) or automatically (e.g. wavelet segmentation for the synaptic protein dataset). To apply TA-GAN on live-cell images without having an expert perform the time-consuming annotation task, an adaptation of TA-GAN was developed. Huo *et al.*[29] applied a method based on a Cycle-GAN to segment structures in computerized tomography scans by using only segmentation maps labeled from non-corresponding magnetic resonance imaging scans. We used this method to segment live-cell images using the segmentation labels from non-corresponding fixed-cell images, by translating fixed-cell images into live-cell images. The translation architecture was a cycle-GAN[24] with a segmentation task that optimizes the translation networks, similarly to the TA-GAN (Fig. 2a). The input fixed-cell image is translated to a live-cell version, and translated back to a fixed-cell version. The mean squared difference between the input and output fixed-cell versions is the *Cycle F* loss that optimizes both generators (F $\rightarrow$ L and L $\rightarrow$ F). The same cyclic translation is done on the input live-cell image (*Cycle L*). Two discriminators, one for images of fixed cells and one for images of live cells, were optimized using the classification of real images as real and generated images as generated (combined into *DF* for fixed cell and *DL* for live-cell images). The generators (F $\rightarrow$ L and L $\rightarrow$ F) were also optimized by GAN losses (*GL* and *GF*). Finally, the live-cell image translated from a real fixed cell image was passed through a pre-trained segmentation network for F-actin rings and fibers. The resulting segmentation map was compared with the segmentation of the input fixed cell image to compute the *Seg F* loss, which was backpropagated through both generators (F $\rightarrow$ L and L $\rightarrow$ F) to optimize their weights. The segmentation network was not optimized. It was pretrained on the same set of images of fixed cells as the whole translation architecture (refer to Lavoie-Cardinal *at al.*[17]). The segmentation of the F-Actin rings and fibers in living neurons reached a higher performance when using the translation TA-GAN approach compared to a segmentation network trained only on fixed cells (Fig. 2b).

Both translation networks are ResNets 9-blocks[21] (Suppl. fig. S8) and both discriminators are 70x70 PatchGANs[30]. The CycleGAN was trained with the axonal F-Actin dataset (Methods sec. 1.1), and the live F-actin dataset (Methods sec. **??**). The network was trained for 1000 epochs with a fixed learning rate of 0.0002. The fixed-cell training images were translated to live-cell images using the translation network of epoch 500, which was identified qualitatively as the best iteration from the generated validation images.

### 3.2 Training the segmentation network for live imaging

The live-cell segmentation network is built around a U-Net-128[31] architecture with batch normalization and two output channels (F-actin rings and fibers). A random subset (2,069 training crops and 277 validation crops) of the dendritic F-actin rings and fibers images (Methods sec. 1.2) was translated into live-cell images using the $F \rightarrow L$ generator (Methods sec. 3.1, Suppl. fig. S10). The translated live-cell images had corresponding annotations from the fixed cell images they were generated from (Suppl. fig. S10a). These translated images and their corresponding annotations are used to train the segmentation network for live cells. Random crops of 128×128 pixels, horizontal and vertical flips were used for data augmentation. Due to class imbalance in the training set, the segmentation loss for fibers is weighted by a factor of 2.5, which corresponded to the ratio between the total number of pixels labeled as the two classes. The segmentation network was trained for 1000 epochs and the iteration with the lowest segmentation loss over the validation set was kept for further use and testing.

### 3.3 Dual nanostructure generation of dendritic F-actin rings and fibers in live-cell images

For the generation of live-cell STED images, the TA-GAN model was trained with the pre-trained and frozen complementary segmentation network (Methods Sec. 3.2) to compute the generation loss. The segmentation network was not trained together with the generator for this dataset because the live-cell images dataset was not labeled. The live-cell TA-GAN was trained with 753 pairs of confocal and STED images (see Table 1 for the hyperparameters).

## 4 STED imaging of F-actin in living neurons

### 4.1 STED live-cell imaging and labeling

Dissociated rat hippocampal neurons were prepared as described previously[17,32] in accordance with and approved by the animal care committee of Université Laval. The dissociated cells were plated on PDL-Laminin coated glass coverslips (18 mm) at a density of 322 cells/mm$^2$ and used for imaging at DIV 12-16.

Super-resolution imaging was performed on a 4-color Abberior STED microscope (Abberior Instruments, Germany) using a 40 MHz pulsed 640 nm excitation laser, a ET685/70 (Chroma, USA) fluorescence filter, and a 775 nm pulsed (40 MHz) depletion laser. Scanning was conducted using a pixel dwell time of 5 µs, a pixel size of 20 nm, and 8 line repetition sequence. The STED microscope is equipped with a motorized stage and auto-focus unit. The neurons were preincubated in HEPES buffered artificial cerebrospinal fluid (aCSF) at 33°C with SiR-Actin(0.5 µM, SpiroChrome) for 8 minutes and washed once gently in SiR-Actin -free media. Imaging was performed in HEPES buffered aCSF of high $Mg^{2+}$/low $Ca^{2+}$ (in mM: NaCl 98, KCl 5, HEPES 10, $CaCl_2$ 0.6, Glucose 10, $MgCl_2$ 5). After identification of the region of interest, the perfusion solution was switched to HEPES buffered aCSF containing high $Ca^{2+}$ , Glycine and without $Mg^{2+}$ (high $Ca^{2+}$: in mM: NaCl 98, KCl 5, HEPES 10, Glycine 0.2, $CaCl_2$ 2.4, Glucose 10). Solutions were adjusted to Osmolality: 240 mOsm per kg and pH: 7.3.

### 4.2 Integration of the generator with the microscope

As shown in Fig. 2c, the trained generator was directly integrated in the acquisition process of the STED microscope to 1) generate the synthetically super-resolved version of the acquired confocal image and 2) select the most informative region from

the confocal field of view to be acquired with the STED modality in the subsequent iteration. For the live-cell experiment, a large field of view was first imaged at low-resolution from which three 500 pixels×500 pixels (10×10 µm) regions of interest were selected by the expert user. Each region was first acquired with both modalities (confocal and STED). For subsequent iterations, 1) a confocal image of the region was acquired, 2) the confocal image was passed through the network to determine the sub-region (2×2 µm) of lowest confidence (see Methods section 4.2.1), 3) a STED image of this sub-region was acquired, and 4) the confocal image of the full region and the STED sub-region were passed through the generator to produce a synthetically super-resolved image of the full region. Note that the last step could be done post-acquisition if the user did not want real-time visualization of the synthetically enhanced image. This whole process was repeated for 15 or 30 iterations at 1 iteration/minute. For the last iteration, a STED and confocal image of the full region were acquired to produce paired images for which the STED image served as ground truth for the synthetic STED image.

Steps 2) and 4) needed to be computed with a graphical processing unit (GPU) to avoid computation induced delays. To do so, the commands from steps 2) and 4) were sent from the microscope's control computer to a GPU-equipped computer using the Flask web framework Python module. All automated acquisitions were programmed using the specpy Python library to interface with the Imspector software (Abberior Instruments, Germany).

#### 4.2.1 STED sub-region selection

The selection of the sub-region that should be acquired with STED was based on the hypothesis that regions of higher uncertainty for the network are the most informative. The network's uncertainty was computed by generating 20 synthetic STED images from the one input confocal image and computing the optical flow between these generations. The optical flow was computed using a Python implementation of the Horn–Schunck method with the Python multiprocessing library, parallelizing the computations on 8 CPUs to increase the speed and avoid delays. The optical flow was considered a better method to compute the variation between the generations than the pixel-wise standard deviation, the latter being mostly proportional to the value of the pixel and not the normalized spread of its values (Suppl. fig. S13). The resulting optical flow was pooled using mean pooling from a 500×500 pixels image to a 5×5 map and the sub-region with a maximum mean optical flow was selected as the most uncertain. This sub-region was then acquired with the STED modality to feed more information into the generator as per step 4) of the workflow presented in Methods sec. 4.2.

### 4.3 Synthetic frame selection for live-cell imaging

The stochasticity induced by the drop-out layers in the generator network gave the possibility of generating multiple synthetic STED counterparts to a single confocal image. A single inference pass generated one image at random from the distribution of possibilities. To generate image sequences that follow the best transition between individual frames, multiple images were generated for each frame and the best was selected. To do so, a graph was constructed using the NetworkX Python library [33], with each synthetic frame represented as a node in the graph. Each frame was connected to all generated images from the immediate previous and next frames by an edge. The weight of each edge was defined by the mean squared difference between the segmentation maps of the images it connects. The final synthetic STED image series was generated from the path minimizing the weight between an initial and a final frame, computed using Dijkstra's algorithm [34]. The supplementary videos (available at `https://github.com/FLClab/TA-GAN`) were constructed using 50 generations per frame.

# 5 Supplementary Figures

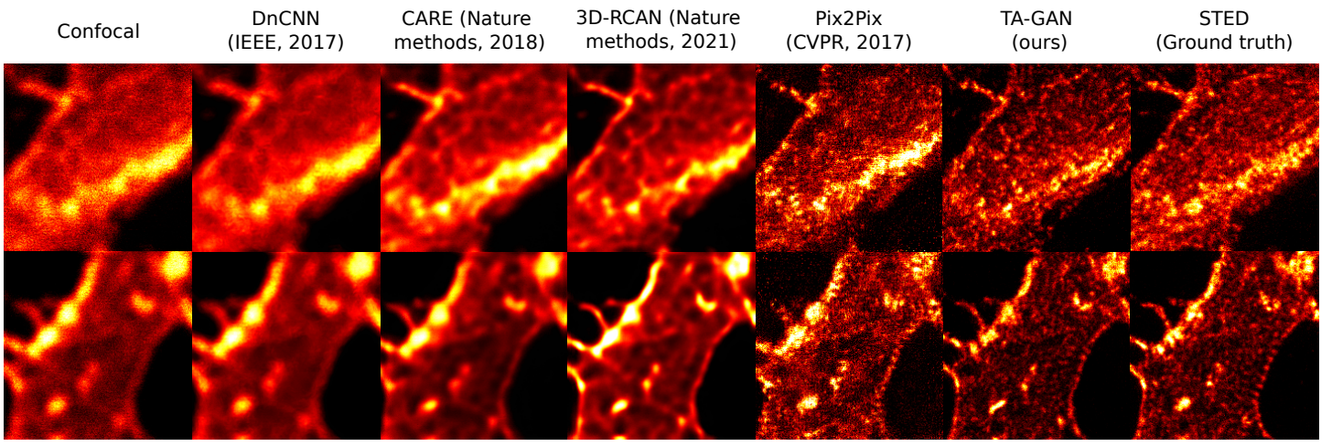


Figure S1: Comparison of different deep learning methods for resolution enhancement and denoising on dendritic F-actin nanostructures. The confocal image is the low-resolution input and the STED image is the aimed ground truth. DnCNN (denoising convolutional neural networks)[27] is a state-of-the-art method for natural images; CARE (content-aware image restoration)[10] uses a U-Net to deblur and denoise but fails at reconstructing the nanostructures that are not resolved in the confocal image; 3D-RCAN[8] uses residual channel attention networks to denoise and sharpen fluorescence microscopy image volumes; Pix2Pix[3] is a state-of-the-art method for image-to-image translation in natural images but is not compelled to reconstruct the nanostructures authentically beyond realism. Our proposed TA-GAN is better at reproducing the nanostructures.

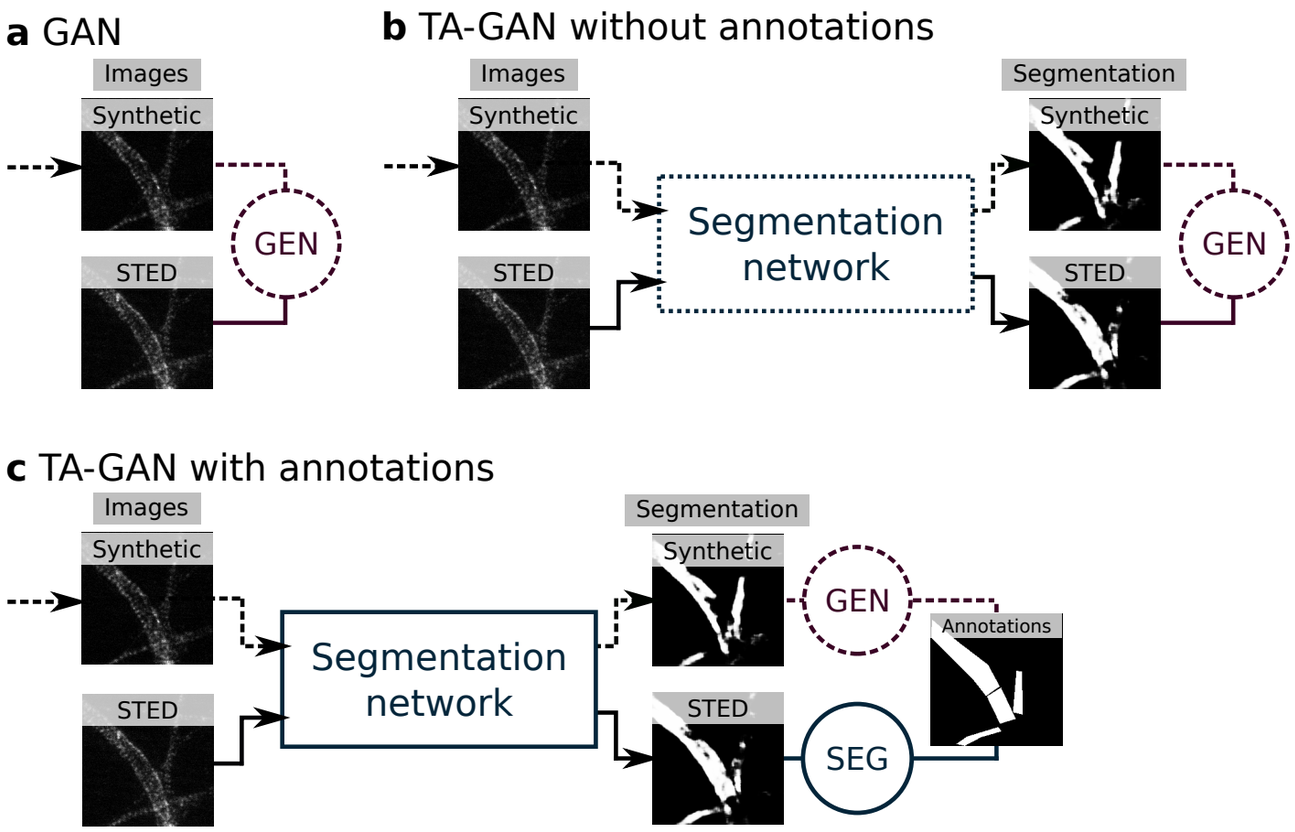


Figure S2: Computation methods of the generation loss. **a,** In the standard GAN architecture, the loss is computed by comparing the synthetic image with its ground truth using a pixel-wise metric such as MSE, L1 or SSIM. **b,** When manual annotations are not available (such as for the live F-actin dataset), the segmentation network has to be pre-trained and its weights are frozen. The generation loss is computed by comparing the output of the segmentation network for the synthetic and real STED images. **c,** When manual annotations are available for the complementary task – pictured here as the bounding boxes for the segmentation of F-actin rings in axons – the segmentation network is optimized by comparing the output of the network for the real STED image with the annotations. The generator is optimized by comparing the segmentation of the synthetic image with the same annotations.

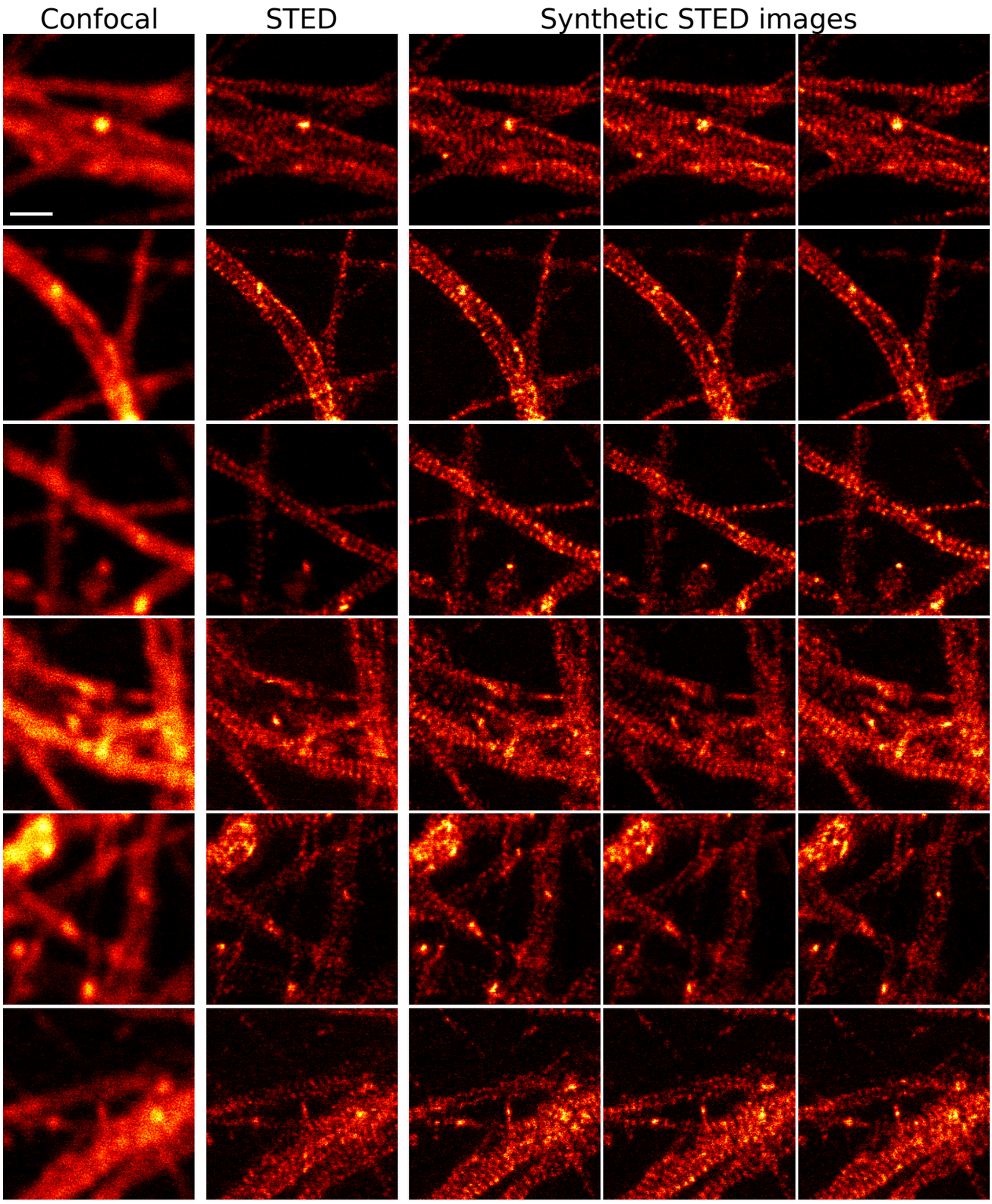


Figure S3: Example results obtained with test images from the axonal F-actin rings dataset. From left to right: confocal image, STED image and three randomly sampled synthetic STED images made with TA-GAN. All images are normalized to their own maximum. The scalebar is illustrated in the first image only but applies to all images and is 1 µm.

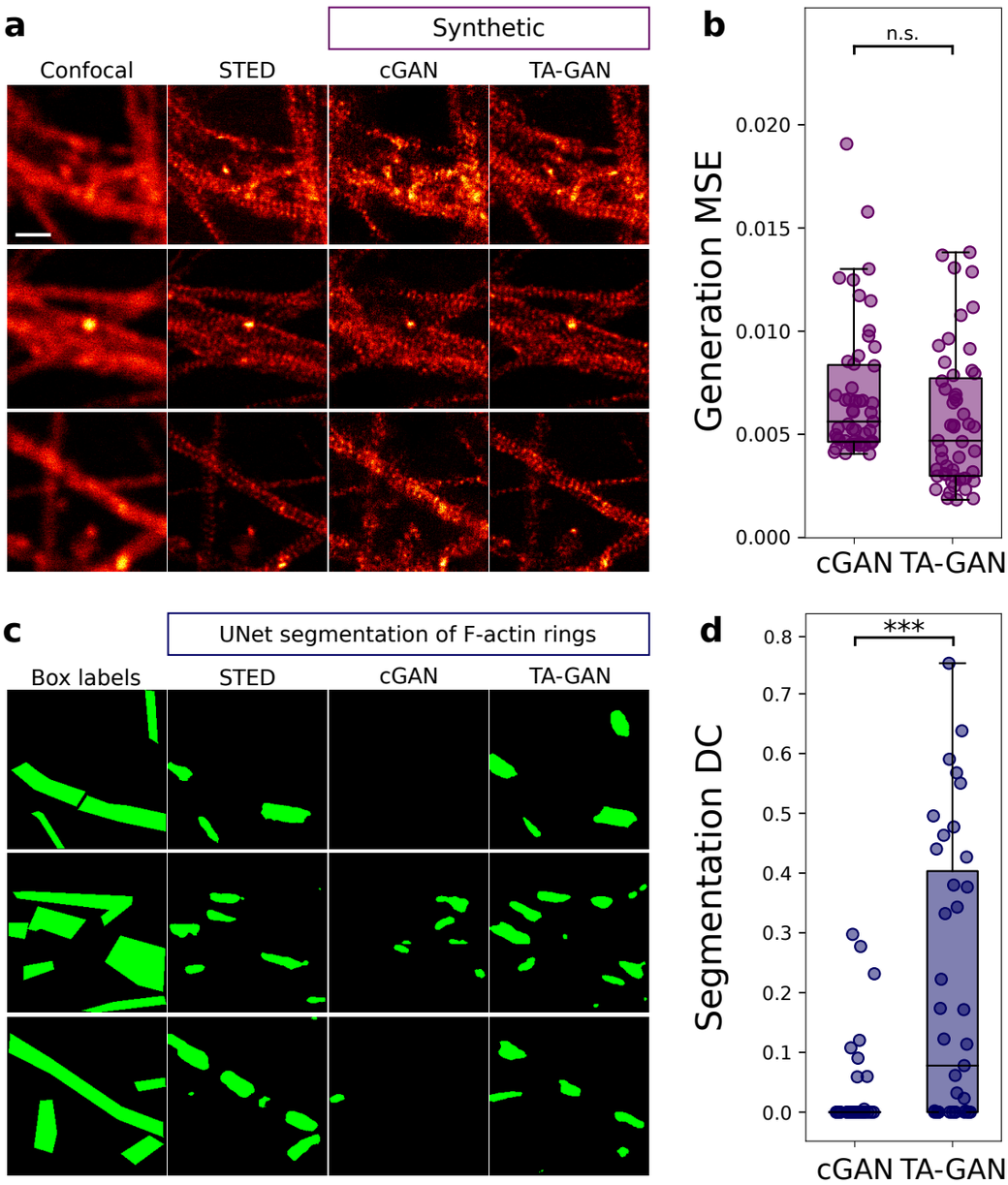


Figure S4: Comparison of the mean square error (MSE) and Dice coefficient (DC), evaluation metrics. **a,** Example results of generated images with a standard conditional GAN (Pix2Pix)[3] and with our TA-GAN method (Scalebar: 1 µm). **b,** Distributions of the mean squared error (MSE) between the synthetic and real STED images over the test set (n = 52), for both methods ($p = 0.06$, not significant). **c,** The segmentation maps output by a segmentation network trained on real STED images[17] show that the F-actin rings generated by the conditional GAN are not recognized, whereas they are recognized for the images generated by the TA-GAN method. **d,** Distributions of the Dice coefficient between the segmentation maps of the real and synthetic STED images over the test set, ignoring empty segmentation maps for which the DC is not defined (n = 39), for both methods ($p = 6 \times 10^{-5}$). Even though the comparison of the MSE distributions shows no significant performance improvement, the DC distributions show that the generation of F-actin rings by the TA-GAN method is much more accurate than the standard cGAN. This result highlights the importance of choosing the right training losses and evaluation metrics. (Statistical significance computed with the independent samples t-test, *** $p < 0.001$, n.s. $p > 0.05$)

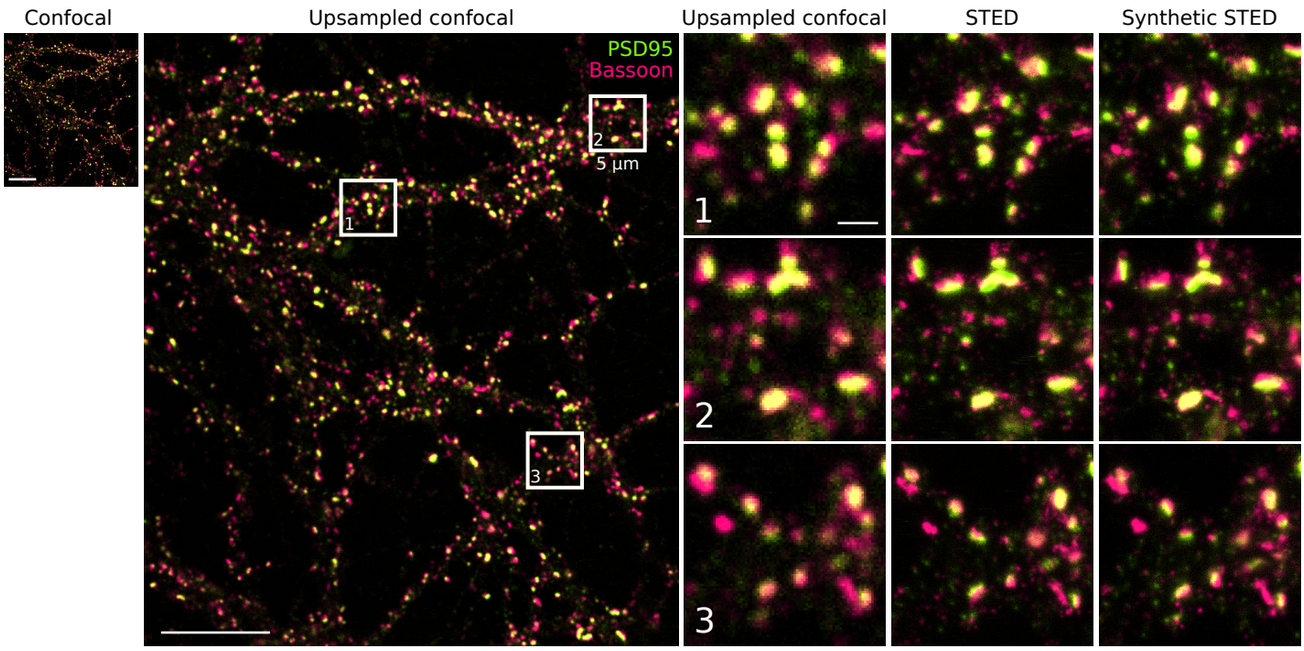


Figure S5: **a,** Example results obtained with test images from the synaptic proteins dataset with 4x upsampling (confocal images have 60 nm pixels and STED images have 15 nm pixels). From left to right: confocal image as it was acquired, upsampled confocal image with nearest-neighbor interpolation, confocal crops, STED crops and corresponding synthetic crops of the regions identified in the upsampled confocal image. The contrast is adjusted so that 1% of the pixels from each image or crop is saturated for better visualization. Scalebars: 10 µm (full image) and 1 µm (ROIs).

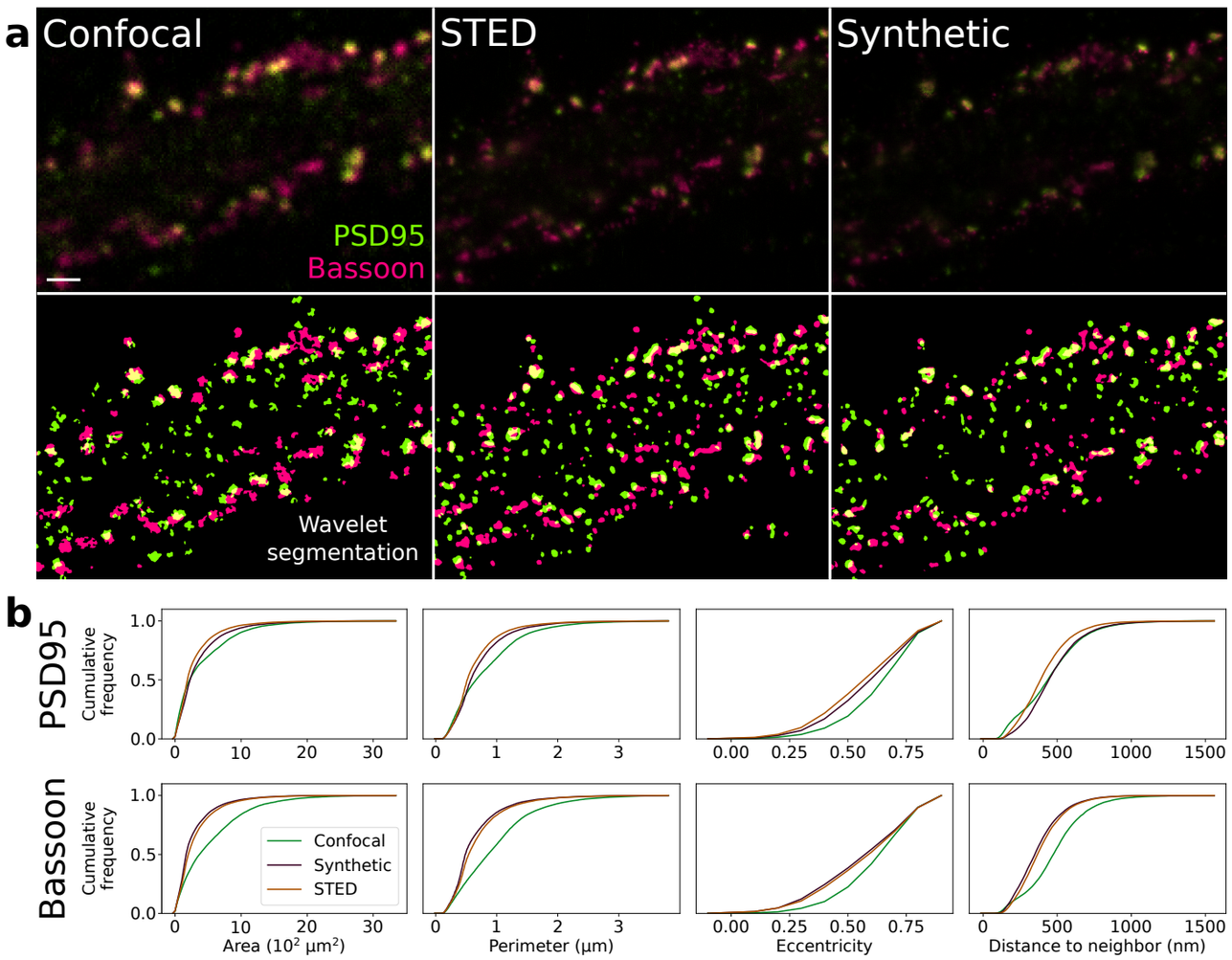


Figure S6: Cluster analysis of PSD95-Bassoon **a,** Example image of PSD95-Bassoon protein pairs in confocal (left), STED (middle) and synthetic STED generated with the TA-GAN (right). **b,** Evaluation of morphological features (area, perimeter, eccentricity and distance to the neighbor from the same channel) of clusters in confocal (green), STED (yellow) and synthetic STED (purple) for PSD95 (top) and Bassoon (bottom). Scalebar: 1 µm.

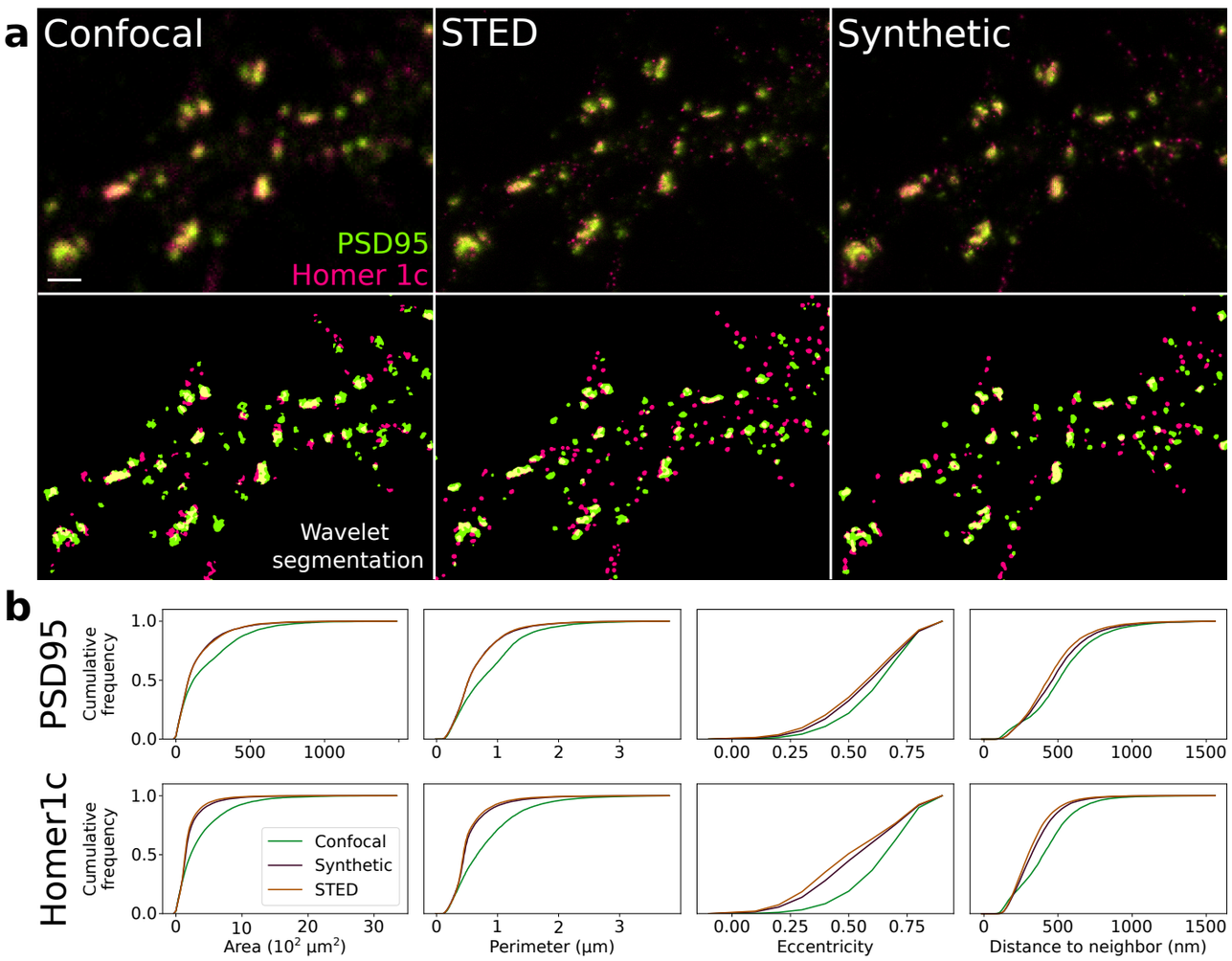


Figure S7: Cluster analysis of PSD95-Homer **a,** Example image of PSD95-Homer1c protein pairs in confocal (left), STED (middle) and synthetic STED generated with the TA-GAN (right). **b,** Evaluation of morphological features (area, perimeter, eccentricity and distance to the neighbor from the same channel) of clusters in confocal (green), STED (yellow) and synthetic STED (purple) for PSD95 (top) and Homer1c (bottom). Scalebar: 1 µm.

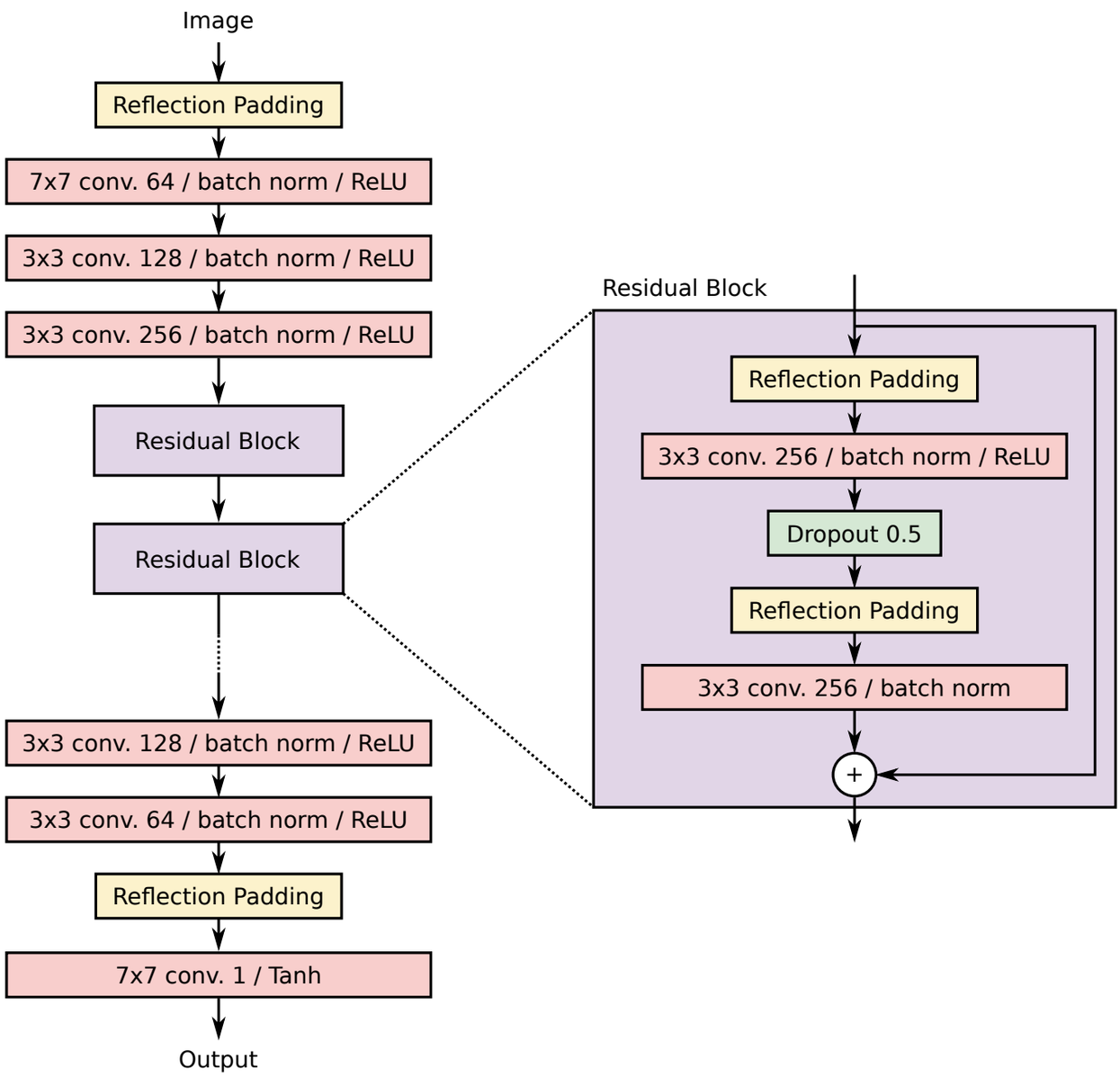


Figure S8: Architecture of the Residual Networks (ResNet). This architecture is referred to as ResNet $X$-blocks, where $X$ refers to the number of residual blocks used.

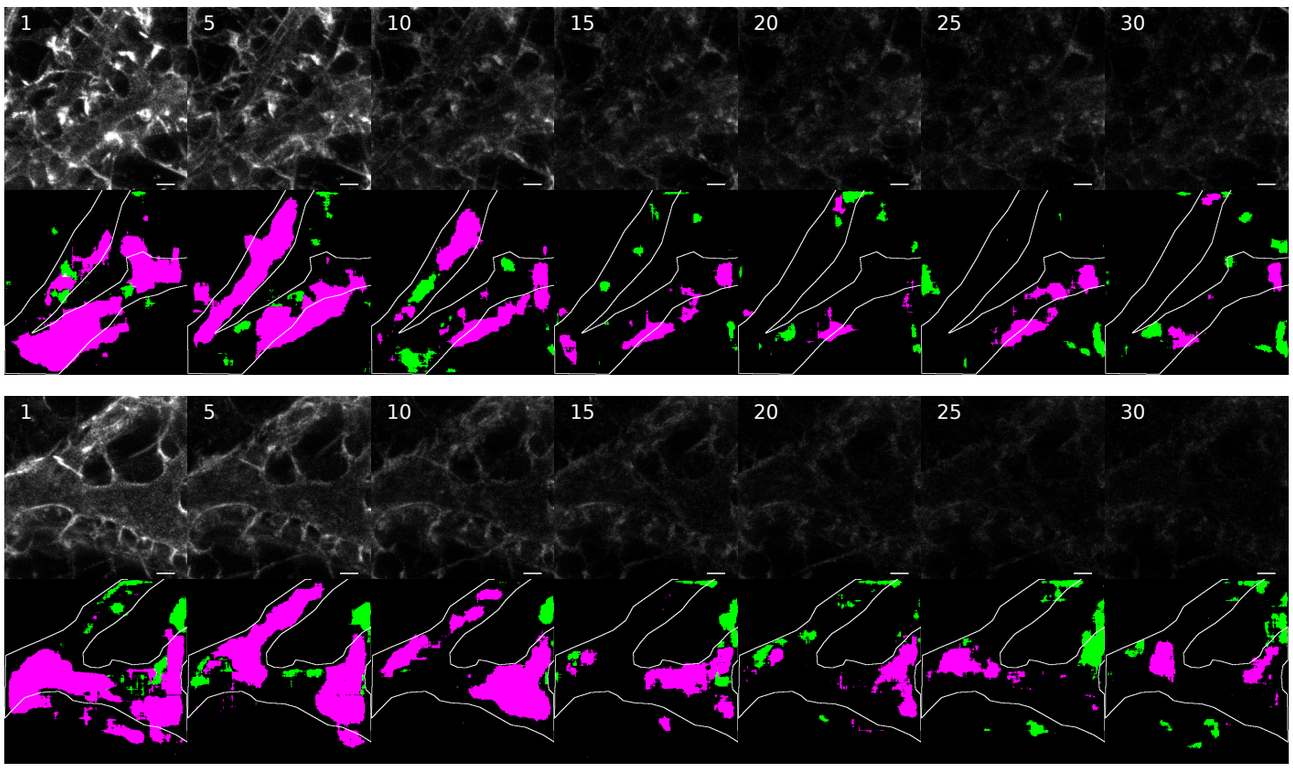


Figure S9: Photobleaching effects are observed when imaging the complete field of view in STED, which strongly affects the detection and segmentation of the nanostructures. The second row shows the output of the segmentation network for live cells (F-actin rings in green and fibers in magenta). As the contrast decreases, the network (as do human experts) has more difficulty in identifying the structures. The frame number is identified in the top left corner, with one image being acquired every minute. (Scalebar: 1 µm)

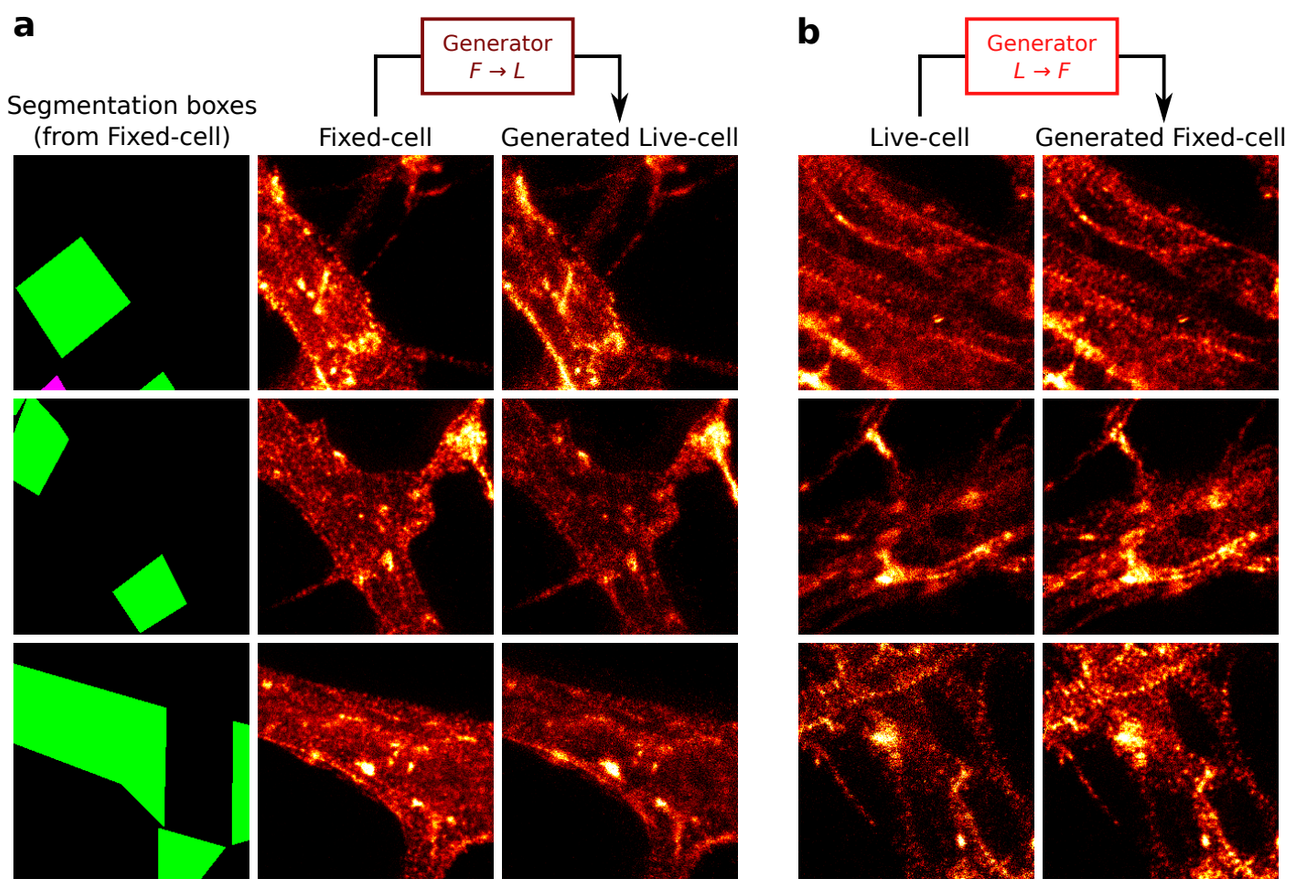


Figure S10: Fixed-cell and live cell images translation. **a,** Real fixed-cell images are translated to synthetic live cell images. The segmentation labels then correspond to both the fixed-cell image and the generated live-cell image. The generated live-cell images are later used to train a segmentation network for live-cell images without requiring additional labeling effort. Live-cell images translated to fixed-cell images (right).

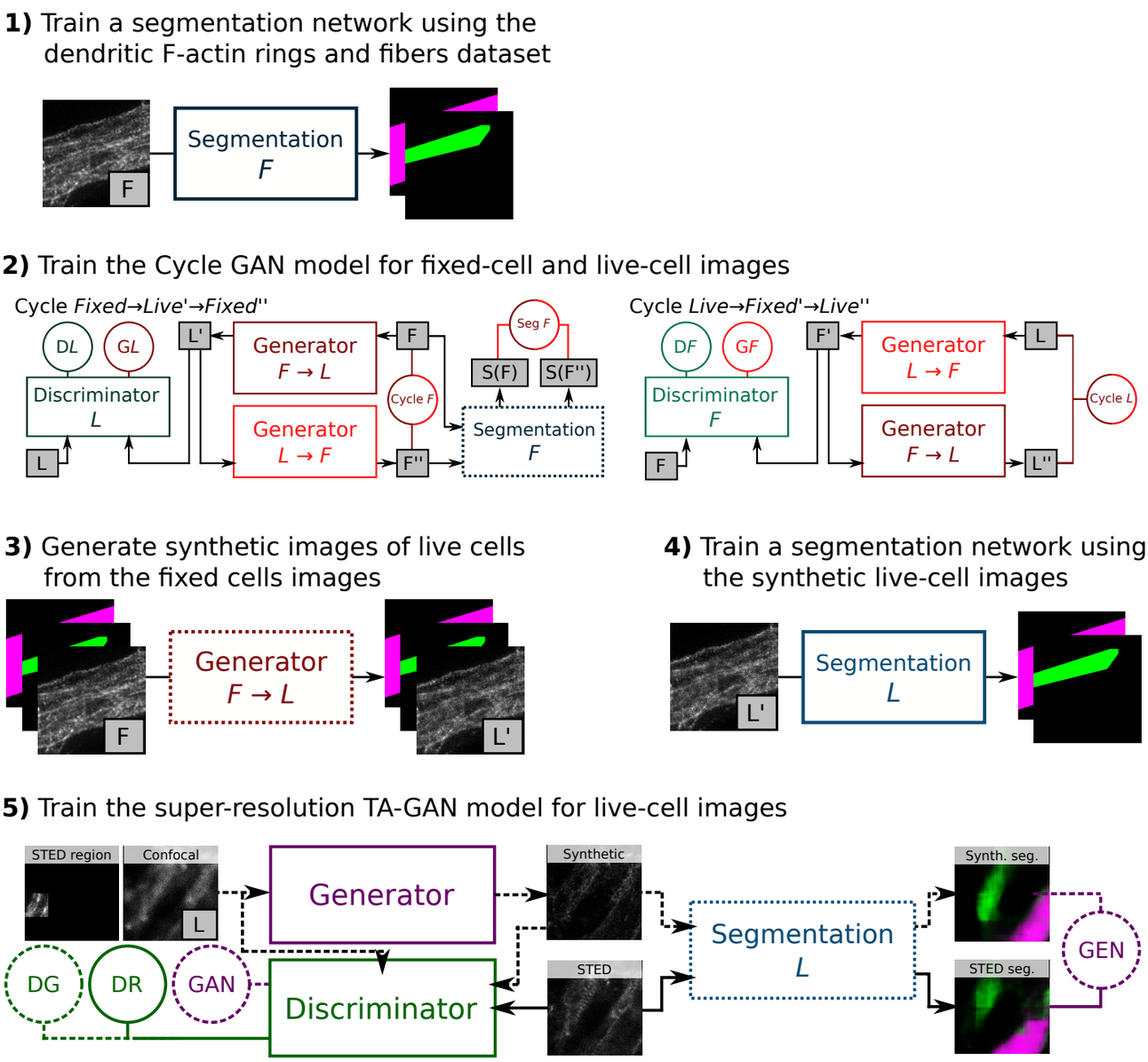


Figure S11: Because the live F-actin dataset has no segmentation labels, additional steps are required to train the TA-GAN model for super-resolution. **Step 1,** the dendritic F-actin rings and fibers dataset is used to train a segmentation network for fixed-cell images. **Step 2,** the trained segmentation network for fixed-cell images is used for the cycle GAN model to learn a mapping from fixed-cell images to live-cell images, and vice versa. **Step 3,** the $F \rightarrow L$ generator translates the training dataset of dendritic F-actin rings and fibers from fixed-cell images to live-cell images. **Step 4,** the synthetic live-cell images are used, along with the segmentation labels from the fixed-cell images they are generated from, to train a segmentation network for live-cell images. **Step 5,** the trained segmentation network for live-cell images is used to compute the generation loss in the TA-GAN model for confocal to STED generation.

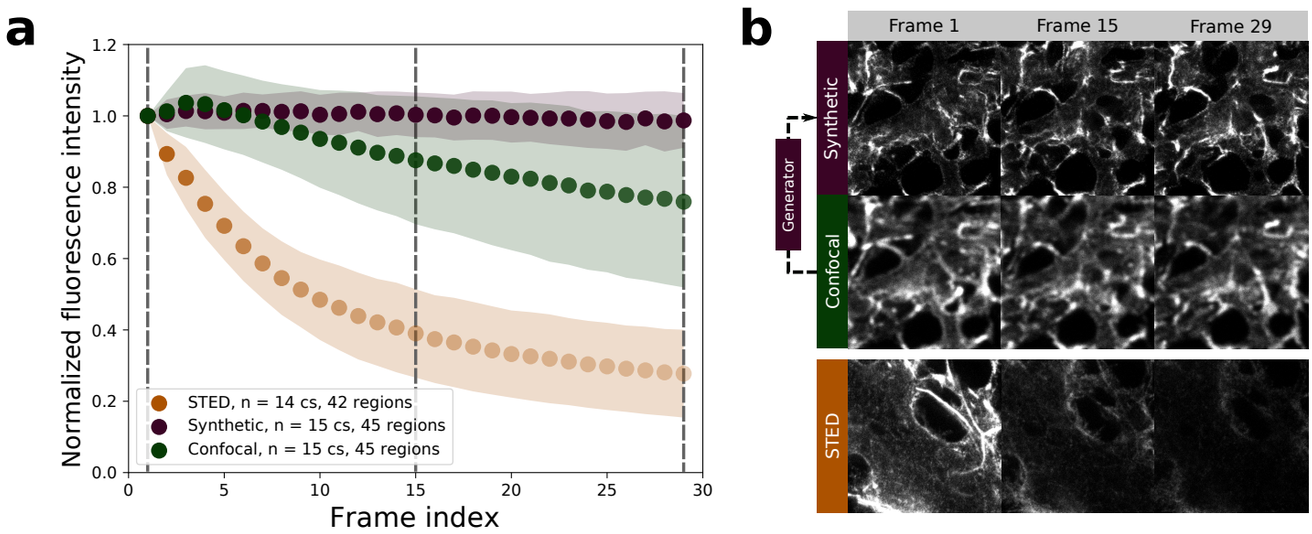


Figure S12: Photobleaching effects compared between STED, confocal and TA-GAN assisted live-cell STED imaging using synthetic STED images. The generate synthetic STED images do not suffer from photobleaching effects as the TA-GAN was trained to generate non-photobleached STED images regardless of the fluorescence level of the confocal image (see Methods sec. 3.3).

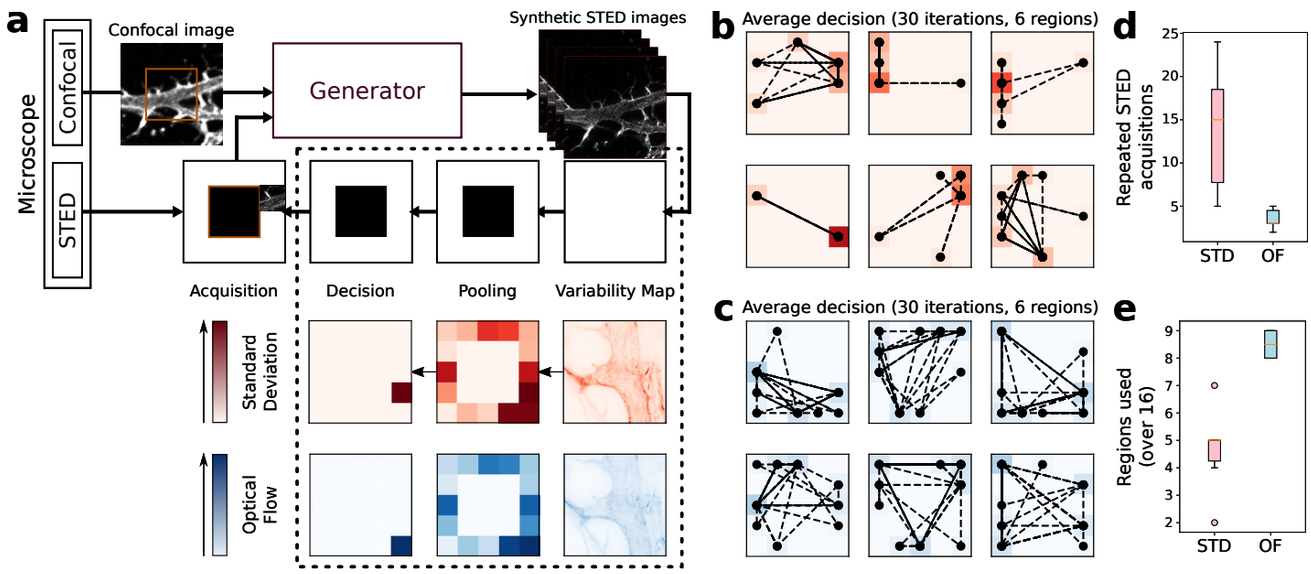


Figure S13: Comparison between standard deviation (SD, red) and optical flow (OF, blue) metrics for the decision process in live imaging. Both STD and OF can be used to measure the variation in the synthetic generations. **a,** For each confocal, the generator produces 50 STED versions. From these 50 generations, a single variability map is computed using either STD or OF, which is then pooled in 25 candidate regions (shade of color is proportional to STD or OF, both proportional to the variability). The central 9 regions are discarded to avoid damaging STED acquisitions over the region of most interest. From the 16 remaining regions, the one with the maximum mean variability dictates the next STED region to acquire. **b-c,** Regions acquired over 30 iterations for 6 different regions from two different coverslips. The shade of each region is proportional to the number of times it was acquired (darker regions were acquired more often). The black dotted lines follow the order of acquisition. In **b** the variability map is computed using STD, and in **c** using OF. **d,** Number of times the same STED region is acquired for two successive frames, over the same 6 regions, for STD and OF. **e,** Number of regions used out of 16 over the 30 iterations. Using OF instead of STD leads to more variability in the choice of region, and therefore less photobleaching from repetitive acquisitions over the same region. It also avoids choosing the same region over successive frames, leading to more diverse and informative content input to the generator.

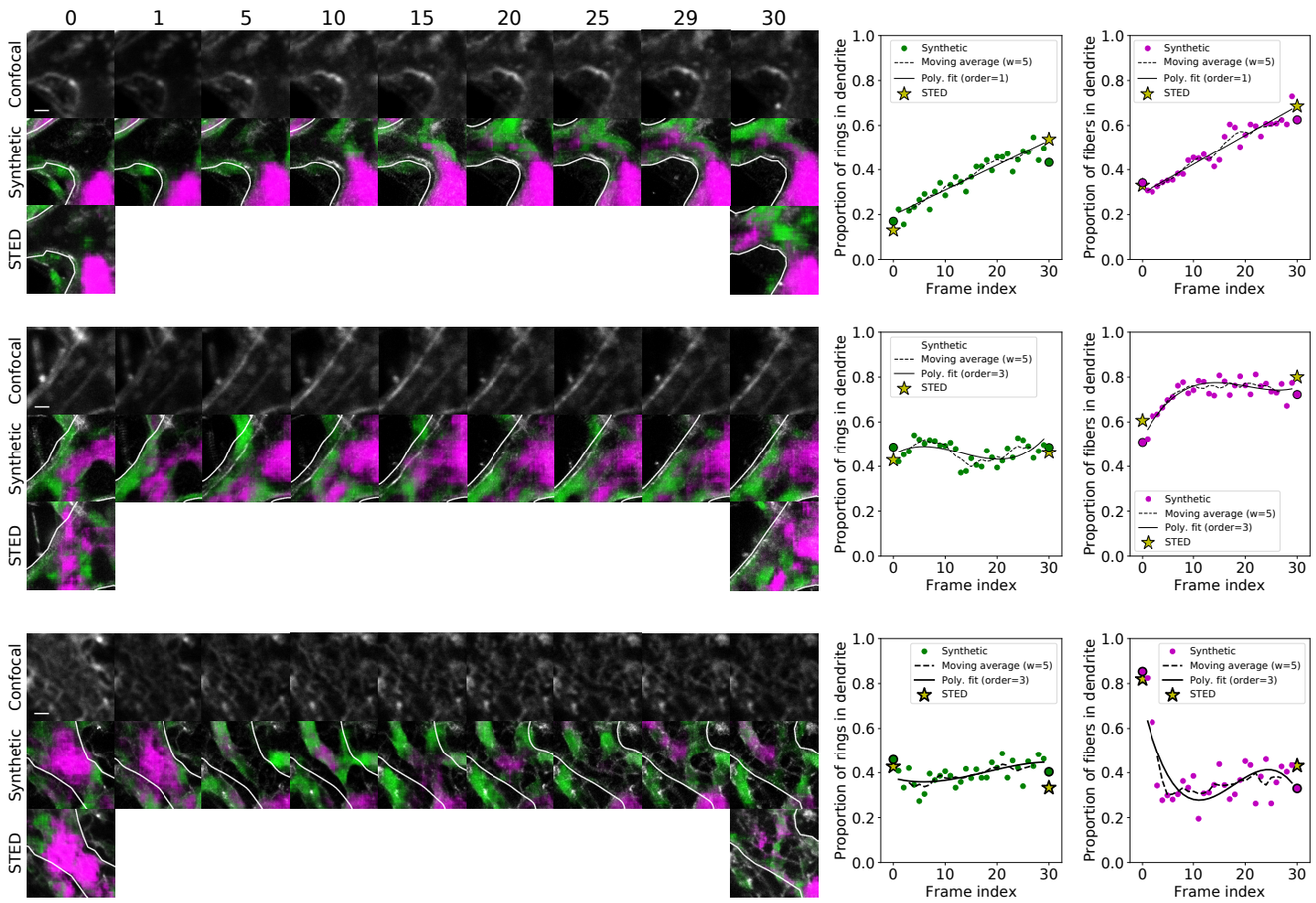


Figure S14: Example of different F-Actin dynamics observed in live-cell imaging following $0Mg^{2+}$/Gly stimulation (Methods). First row: the proportion of rings (green) and fibers (magenta) both increase linearly. Second row: As the proportion of rings stay constant, the proportion of fibers increases rapidly in the first 10 frames and plateaus for the following 20. Third row: after a few frames, the proportion of rings and fibers both decrease. In that case due to the overall decrease of the fluorescence signal, both fibers and rings are difficult to detect after a few frames. The TA-GA does not predict fibers or rings as confirmed by the high correspondance between the last real and synthetic STED images. Observing the transition between the initial and final states is only made possible by generating synthetic STED frames from the confocal images, as complete STED acquisitions would rapidly photobleach the fluorophores. Scalebar: 1 µm.

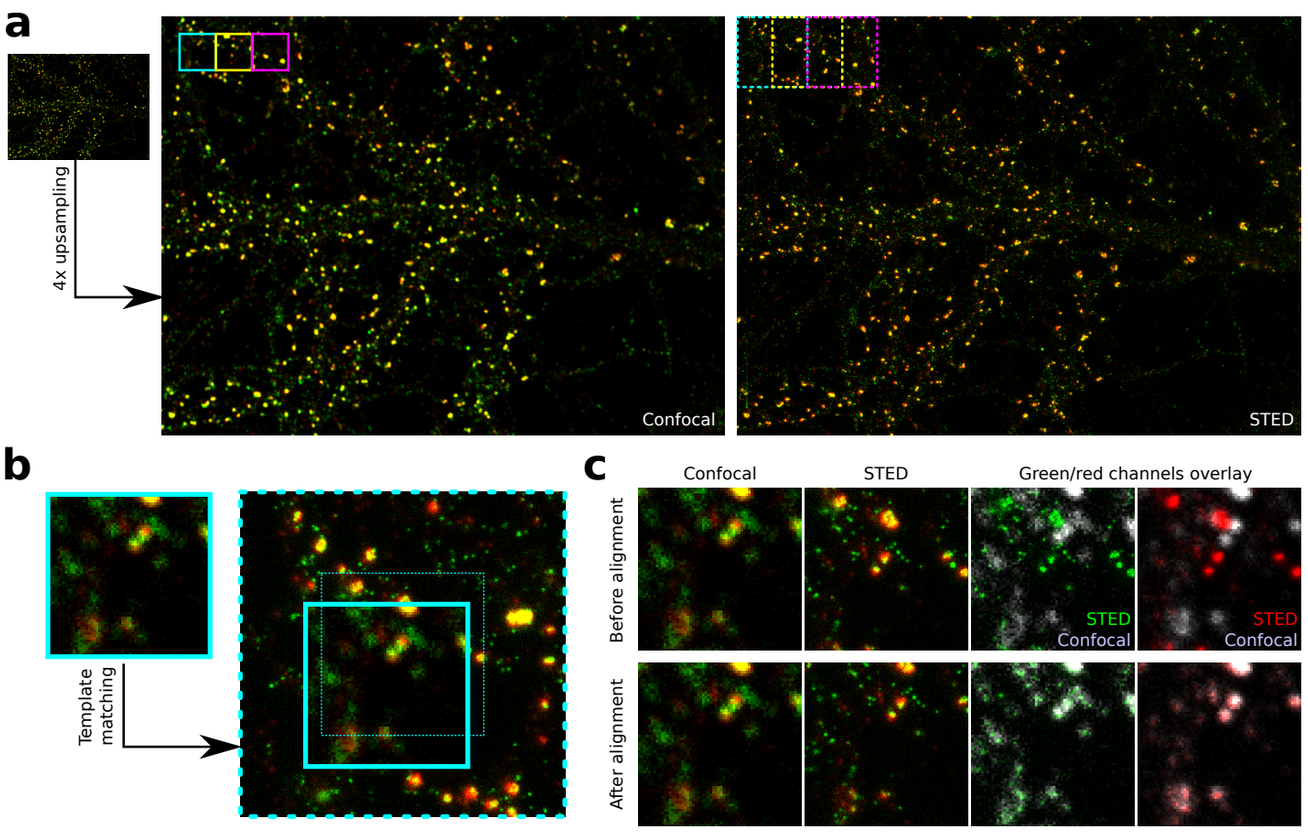


Figure S15: Registration method to create aligned pairs of confocal and STED images for the synaptic protein dataset. **a,** The confocal image is upsampled without interpolation (nearest-neighbor interpolation) to match the size of the STED image. The three regions identified show the sliding window method to iteratively choose sections to match. **b,** For each 512 px square region of the STED image, the centered 256 px square region from the confocal image is matched using the *OpenCV* [26] template matching library. The dashed square identifies the center of the region, and the full line identifies the region which matches the confocal crop; notice the offset between the two. **c**) The confocal crop, the STED corresponding crop and a channel-wise overlay of the two before and after alignment.