Resolution Enhancement with a Task-Assisted GAN to Guide Optical Nanoscopy Image Analysis and Acquisition

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

[1]Institute Intelligence and Data (IID), Université Laval, Québec, QC, Canada.
[2]CERVO Brain Research Center, Québec, QC, Canada.
[3]Département de génie électrique et de génie informatique, Université Laval, Québec, QC, Canada.
[4]Département de psychiatrie et de neurosciences, Université Laval, Québec, QC, Canada.

*Corresponding author(s). E-mail(s): christian.gagne@gel.ulaval.ca; flavie.lavoie-cardinal@cervo.ulaval.ca;
Contributing authors: catherine.bouchard.22@ulaval.ca; theresa.wiesner@gmail.com; andreanne.deschenes.1@ulaval.ca; anthony.bilodeau.1@ulaval.ca;

Abstract

We developed a deep learning model that predicts super-resolved versions of diffraction-limited microscopy images. Our model, named Task-Assisted Generative Adversarial Network (TA-GAN), incorporates an auxiliary task (e.g. segmentation, localization) that is closely related to the characterization of the observed biological nanostructures. We evaluate how TA-GAN improves generative accuracy over unassisted methods using images acquired with two modalities: confocal (diffraction-limited) and STimulated Emission Depletion (STED, super-resolved) microscopy. The generated synthetic STED images show an accurate distribution of the F-actin nanostructures, and they replicate the nanoscale morphology

of synaptic protein clusters. We expand the applicability of the TA-GAN to online live-cell imaging assistance. Incorporated directly into the acquisition pipeline of the microscope, the TA-GAN informs the user on the nanometric content of the field of view without requiring the acquisition of a super-resolved image. This information is used to automatically select time points of interest for the visualization of biological change, optimizing the acquisition sequence and reducing light exposure. TA-GAN partially replaces super-resolution imaging acquisitions, enables the creation of domain-adapted labeled datasets requiring minimal manual annotation, and assists microscopy users by taking online decisions regarding the choice of imaging modality and region of interest.

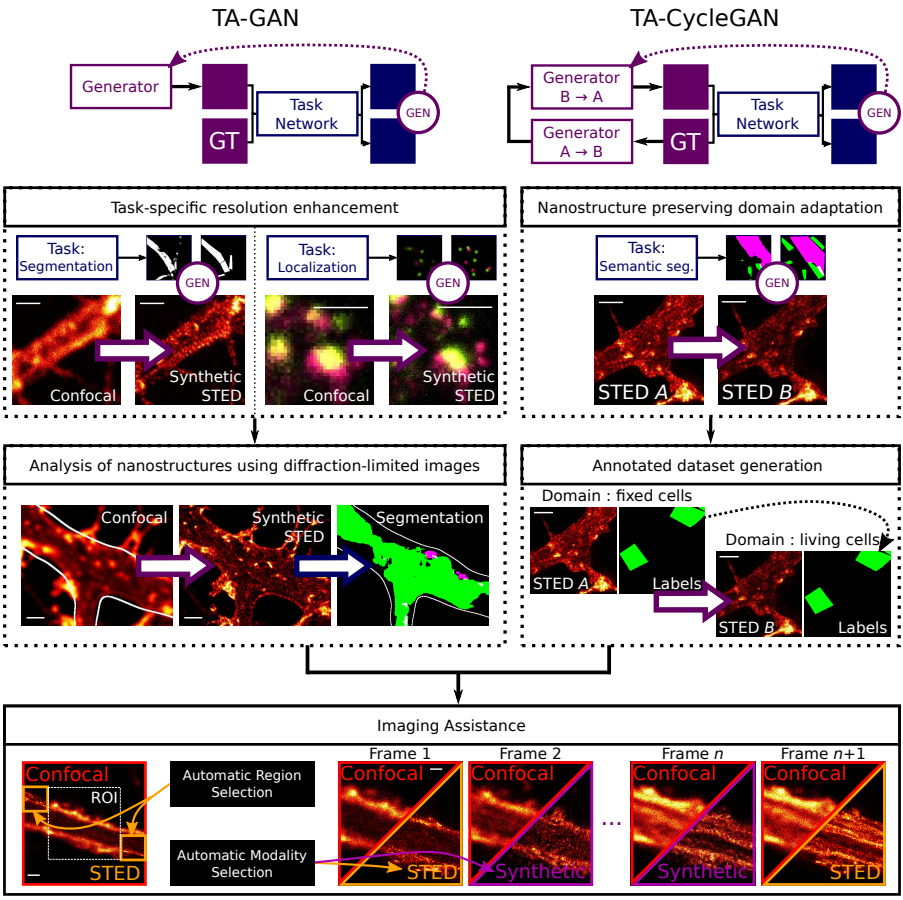

Fig. 0 Graphical Abstract. The proposed model has two general use cases: TA-GAN, for paired datasets, and TA-CycleGAN, for unpaired datasets. *Top-left:* The TA-GAN uses a task adapted to each dataset for accurate resolution enhancement. *Middle-left:* The generated synthetic STED images are used to analyze the distribution of nanostructures that were not resolved in the original confocal image. *Top-right:* Domain adaptation using the TA-CycleGAN enables the generation of large annotated synthetic image datasets from a new domain. *Middle-right:* The task network generates the corresponding annotations on the synthetic dataset. *Bottom:* Both models can be used for microscopy acquisition guidance. The TA-GAN model, trained using a TA-CycleGAN generated dataset, can automatically identify regions and frames of interest from the low-resolution images. Automatic switching between low- and high-resolution imaging modalities is guided by the TA-GAN predictions. Scale bars: 1 µm.

1 Introduction

The development of super-resolution optical microscopy techniques to study the nanoscale organisation of biological structures has transformed our understanding of cellular and molecular processes [1]. Such techniques, including STimulated Emission Depletion (STED) microscopy [2], are compatible with live-cell imaging, enabling the monitoring of sub-cellular dynamics with unprecedented spatio-temporal precision. In the design of super-resolution microscopy experiments, multiple and often conflicting objectives (e.g. achievable spatial resolution, acquisition speed, light exposure, and signal-to-noise ratio) must be considered [3]. For example, high spatio-temporal resolution is often achieved at the expense of an increased light dose on the sample [4]. Machine learning-assisted optical microscopy approaches have been proposed to improve the acquisition processes, mostly by limiting light exposure [3, 5, 6]. In parallel, several supervised [7, 8] and weakly supervised [3, 9] deep learning approaches have been developed for quantitative and high-throughput analysis of super-resolution optical microscopy images.

Deep learning approaches were also recently applied for microscopy image generation [11–15]. Conditional Generative Adversarial Networks (cGAN) [16] generate data instances based on a different input value, capturing some of the features of this input value to guide the creation of a new instance that fits the target domain. This conditioning makes cGANs well suited for image-to-image translation tasks [2], including super-resolution. For natural images, cGANs have been used to enhance the spatial resolution by increasing the number of pixels and sharpening the finer details of the upsampled image [18]. For medical imaging, cGANs have been applied to improve the signal-to-noise ratio [19], to augment limited data [20–22], and to sharpen raw noisy data [23, 24]. cGAN-based super-resolution approaches have been gaining similar interest for microscopy images, as they achieve high performance in reducing blurring artifacts and noise [13, 14]. Their applicability has been demonstrated on images of biological structures such as microtubules [14] and mitochondria [13]. Yet, these cGAN models are trained to generate images that *appear* to belong to the target higher-resolution domain, without specifically requiring the biological features of interest to be accurately generated [25].

Super-resolution microscopy techniques exploit the ability to modulate (or switch) the emission properties of fluorescent molecules to overcome the diffraction barrier of light microscopy, which cannot be surpassed by adding pixels or improving the contrast [26]. In this context, it is challenging to rely on algorithmic methods to generate images of sub-diffraction structures that are not optically resolved in the original image [27]. Deep learning-based super-resolution [5, 11–14] and domain adaptation [28] approaches have been proposed recently for optical microscopy, but concerns and skepticism arise regarding their applicability to characterize biological structures at the nanoscale [25, 27, 29]. Yet, the possibility to super-resolve microscopy images post-acquisition would unlock opportunities that were otherwise out of reach,

favorably alleviating some of the compromises between the spatial resolution, light exposure, and acquisition speed [11, 30].

In its regular form, the generator network of a cGAN is optimized to produce images that are realistic enough to be indiscernible from real images. This is well adapted for the generation of natural images, where fooling a discriminant network (discriminator) with a realistic output is the sole objective used to optimize the generator. However, for microscopy images, realism does not ensure that the images are usable for further field-specific analysis. The primary goal for generating super-resolved microscopy images is to produce reliable nanoscale information on the biological structures of interest. Optimizing a network using complementary tasks, or multi-task learning, can guide the generator to produce content that matters for the current context [31]. Various applications of cGANs for image-to-image translation use complementary tasks such as semantic segmentation [32, 33], attributes segmentation [34], or foreground segmentation [35], to provide spatial guidance to the generator, ensuring that the generated images are consistent with the target annotations. We adapt this idea in the context of microscopy, where structure-specific annotations can direct the attention to subtle features that are only recognizable by trained experts.

We propose to guide the image generation process using a complementary task that is closely related to the biological question at hand. This approach improves the applicability of synthetic data generation using deep learning in microscopy and ensures that the generated features in synthetic images are consistent with the observed biological structures in real images. Microscopy image analysis tasks that are already routinely solved with deep learning [36] (e.g. segmentation, detection, and classification) can guide a cGAN to preserve the biological features of interest in the generated synthetic images, gearing towards undifferentiated analysis between the synthetic and the ground truth super-resolved images. We designed a Task-Assisted GAN (TA-GAN) for confocal-to-STED super-resolution image generation, which is optimized to perform well over a complementary task associated with the nanostructures of interest that are unresolved in confocal images. We expand the applicability of the method with a variation called TA-CycleGAN, based on the CycleGAN model [37], applicable to unpaired datasets. Here, the TA-CycleGAN is applied to domain adaptation for STED microscopy of fixed and living neurons. Our results demonstrate that the TA-GAN and TA-CycleGAN models improve the synthetic representation of biological nanostructures in comparison to other deep learning-based super-resolution approaches. Specifically, our method is useful to 1) guide the quantitative analysis of nanostructures from diffraction limited ground truth images, 2) generate synthetic datasets of different modalities for data augmentation or to reduce the annotation burden, and 3) predict regions of interest for machine learning-assisted live-cell STED imaging.

2 Results

TA-GAN: Task-assisted super-resolution image generation

Generative methods designed for synthetic microscopy image generation, where direct comparisons between the generated and the ground truth images are made, have been shown to be very effective for deblurring and denoising of confocal images [11, 12, 14]. To increase the accuracy of deep learning-assisted super-resolved image generation of complex nanoassemblies (e.g. the F-actin periodical lattice in neurons [38]), we considered the combination of a cGAN with an additional convolutional neural network, the task network, targeting an image analysis task relevant to the biological structures of interest (Figure 1a). Three individual networks form the TA-GAN model: 1) the generator, 2) the discriminator, and 3) the task network. They are trained in parallel and updated sequentially in each iteration (Methods). Pairs of confocal (low-resolution) and STED (super-resolution) images are used to train the generator and the discriminator. The TA-GAN is tested on different nanostructures in fixed hippocampal neurons for the translation of confocal to STED images. The absence of movement in fixed cells allows for the acquisition of paired confocal and STED images. The task network relies on manual expert annotations highlighting the nanostructures of interest on the ground truth STED images for training (Figure 1a). The error between the task network predictions and the expert annotations is back-propagated to the generator to optimize its parameters (Methods, Figure S1).

The first TA-GAN model is trained to generate STED images of the axonal F-actin lattice from confocal images (Figure 1b). The membrane-associated periodic skeleton in axons is formed by F-actin rings and spectrin tetramers that are periodically spaced by around 180 nm and cannot be resolved with confocal microscopy [38]. The complementary task we identified to train the TA-GAN on the *axonal F-actin dataset* (Methods) was the segmentation of periodical F-actin rings in axons (Figure 1a). This complementary task is used to compute the generation loss and to evaluate the generation performance at test time (Methods). Different denoising and super-resolution baselines were applied to the translation of the input confocal image into a STED image (Figure 1b, first row). For all baselines, the generated F-actin nanostructures in the synthetic denoised and super-resolved images could not be recognized by a segmentation U-Net trained to recognize the F-actin rings on real STED images (Figure 1b, second row). For the TA-GAN image generation, using the segmentation loss to optimize the generator's weights forces the generated F-actin nanostructures to be realistic enough to be recognized by the segmentation U-Net (Figure 1b). We compared the results of the TA-GAN with the best performing baseline, pix2pix, a cGAN developed for image modality translation [2]. The pixel-wise mean square error (MSE) between generated and ground truth (GT) STED images shows no significant difference when comparing with (TA-GAN) and without (pix2pix) task-assistance during training

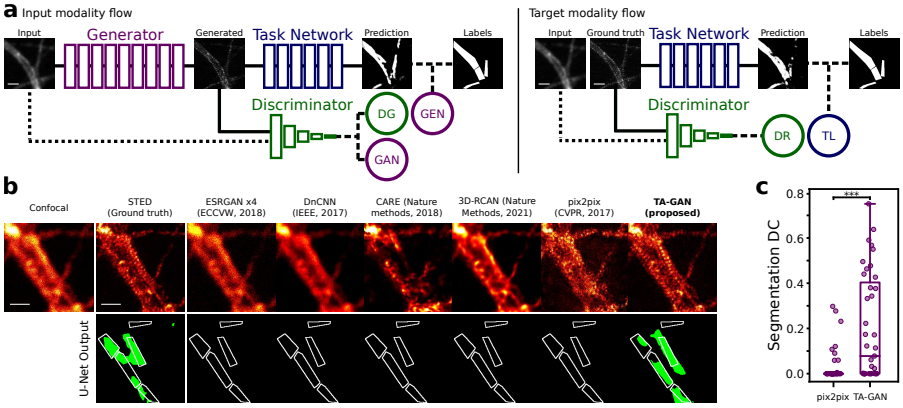

Fig. 1 The TA-GAN method. a, Architecture of the TA-GAN model. The losses (circles) are backpropagated to the networks of the same color: the generator (violet, ResNet with 9 blocks [39]), the discriminator (green, PatchGAN [40]), and the task network (blue, ResNet with 6 blocks). The model is split for easier visualization: the left section shows the flow of the input image and its related losses; the right section shows the flow of the ground truth (GT) target image and its related losses (Methods). **b,** Comparison of different deep learning methods for resolution enhancement and denoising on axonal F-actin nanostructures. (*Top*) The confocal image is the low-resolution input and the real STED image is the aimed ground truth (STED GT). ESRGAN x4 (Enhanced Super-Resolution Generative Adversarial Networks) [41] is a state-of-the-art method for upsampling natural images; DnCNN (denoising convolutional neural networks) [42] is a state-of-the-art denoising method for natural images; CARE (content-aware image restoration) [11] uses a U-Net for debluring and denoising; 3D-RCAN [12] uses residual channel attention networks to denoise and sharpen fluorescence microscopy image volumes; pix2pix [2] is a state-of-the-art method for image-to-image translation in natural images. (Methods) (*Bottom*) Segmentation of the axonal F-actin rings (green) predicted by a U-Net trained on real STED images [3]. Bounding boxes (white line) correspond to the manual expert annotations for axonal F-actin rings [3]. See Figure S3 for more TA-GAN results. **c,** The DC evaluates the overlap of the predicted segmentation masks for the generated synthetic and real STED images (STED GT). It shows a significant improvement ($p \sim 10^{-5}$, Mann-Whitney U test [1]) when using TA-GAN compared to pix2pix for F-actin rings segmentation on synthetic super-resolved images (Figure S2). Scale bars: 1 µm

(Figure S2a). The Dice coefficient (DC) however, computed between the output of a U-Net trained on real STED images[3] for the generated and GT STED images, shows a significant improvement for TA-GAN compared to pix2pix (Figure 1c, Figure S2b). These results confirm that the pixel-wise MSE loss is not sufficient to guide the generator to produce synthetic super-resolved images that are reliable for the analysis of the axonal F-actin nanostructure distribution.

The performance of the TA-GAN is next evaluated on a more complex task, which is the semantic segmentation of two nanostructures that can be differentiated only with super-resolution optical microscopy: dendritic F-actin rings and fibers [3, 44]. It was recently shown that, in dendrites, the prevalence of these nanostructures varies in an activity-dependent manner [3]. The TA-GAN

is trained on the *dendritic F-actin dataset* (Methods) using the semantic segmentation of F-actin rings and fibers as the complementary task (Figure 2a). The TA-GAN generates synthetic nanostructures in dendrites that are successfully segmented by a U-Net trained on real STED images[3] (Figure 2b). Similarly to results previously obtained from real STED images [3], segmentation of the synthetic images shows that the area of the F-actin rings significantly decreases as the neuronal activity increases, while the opposite is observed for F-actin fibers (Figure 2c). A similar conclusion cannot be drawn from images generated by a standard conditional GAN architecture since the rings and fibers generated are not accurate enough to be segmented by the U-Net trained on real STED images (Figure S4). This experiment highlights the reliability of the TA-GAN to guide quantitative image analysis of different sub-diffraction structures for varying biological conditions.

The complementary task in the TA-GAN model requires expert annotations, which results in a trade-off between the cost of producing these annotations and the gain in generation accuracy they provide. Model-based and machine learning-based tools have been developed to automatically create simple labels for an extensive number of specific analysis applications, like segmentation labels [45–47] for cell-wise morphology features extraction, and detection labels [36] for cell counting. When applicable, these resources can eliminate the task-assistance trade-off by eliminating the manual annotation requirement, replacing it with automatic annotations.

We use automatically generated localization annotations to train the TA-GAN for confocal-to-STED super-resolution of the *Synaptic protein dataset* consisting of two-channels images of synaptic proteins clusters [4] (Figure 3a,b, Figure S5, Methods). We localize the centers of synaptic proteins clusters (Pre : Bassoon; Post : PSD95, Homer1c) [4] in STED images relying on an automatic segmentation pipeline [5] (Methods). We evaluate the performance of the TA-GAN on the comparison of synaptic protein cluster morphology features in real and synthetic STED images (Figure 3c, Methods). We measure similar population-level distributions for the cluster area, perimeter, and eccentricity in super-resolved real and synthetic images (Figure 3c, Figure S7 and Figure S8).

TA-CycleGAN: Domain adaptation on unpaired datasets

For many microscopy modalities, paired *and labeled* training datasets are not directly available or would be associated with a high annotation burden from highly qualified experts. Based on the results obtained using confocal and STED image pairs on fixed neurons, we wanted to expand the applicability of the TA-GAN to unpaired datasets – here, images of fixed and living cells. We generated a dataset of F-actin nanostructures in living neurons using the far-red fluorogenic dye SiR-Actin (Methods) that reveals the F-actin rings and fibers in living neurons [50].

Using our task-assisted strategy, we trained a CycleGAN [2] model, as it was precisely developed for image domain translation for unpaired datasets.

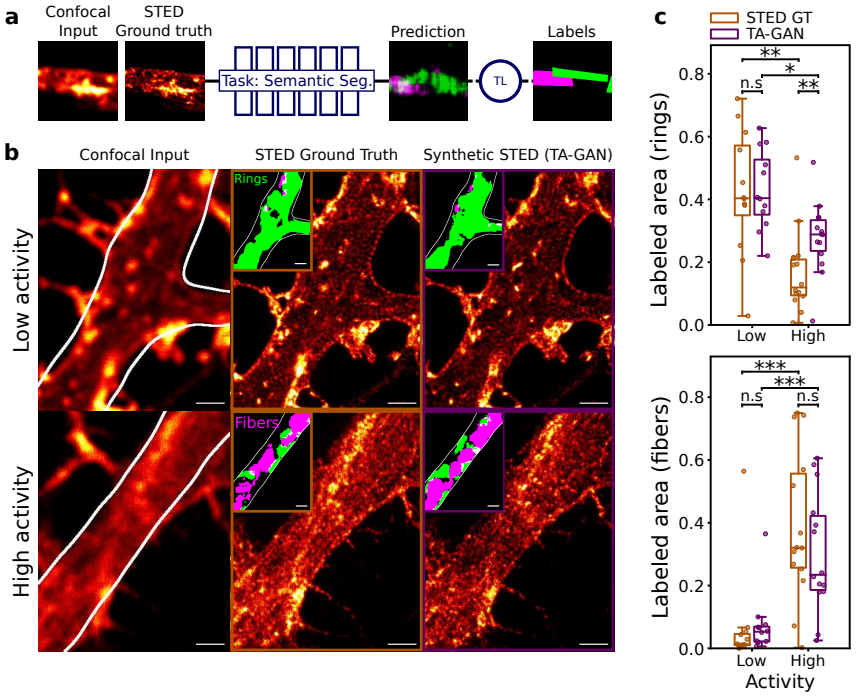

Fig. 2 Resolution improvement by TA-GAN with an auxiliary semantic segmentation task. a, The semantic segmentation of F-actin rings (green) and fibers (magenta) is used as the complementary task to train the TA-GAN (Methods). Annotations identifying F-actin rings and fibers are polygonal bounding boxes generated by an expert (Methods). **b,** Confocal, real STED (ground truth (GT)), and synthetic STED images showing dendrites for low (top) and high (bottom) neuronal activity level (Methods). Insets show the regions identified as rings (green) and fibers (magenta) by the segmentation U-Net trained on real STED images [3]. **c,** The measured proportion of F-actin rings in dendrites is significantly larger at low neuronal activity compared to high activity in both real (orange, STED GT) and TA-GAN generated synthetic (purple, TA-GAN) STED images (STED GT: $p = 0.0007$, TA-GAN: $p = 0.004$). The opposite is observed for F-Actin fibers for both the real ($p = 0.0006$) and the synthetic STED images ($p = 0.0009$). Statistical significance computed with the The Mann-Whitney U test [1] ($*** \; p < 0.001$, $** \; p < 0.01$, $* p < 0.05$, n.s. $p > 0.05$). Scale bars: 1 µm.

The TA-CycleGAN includes two generators that are trained to first perform a complete cycle between the two domains (fixed- and live-cell STED imaging), and then to compare the ground truth input image with the generated end-of-cycle image. In the generic CycleGAN model, the losses are minimized when the generated images appear to belong to the target domain and the MSE loss between input and output is minimized. No weight is given to the preservation of the substructures of interest through the translation. With the

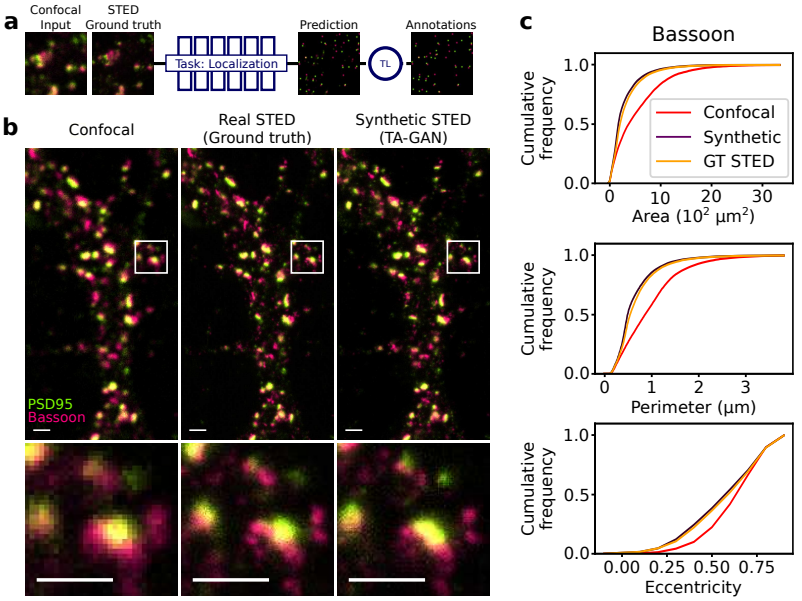

Fig. 3 TA-GAN predicts synaptic protein cluster nanoscale morphology through an automatically labeled localization task. a, Prediction of the position of the weighted centroids of PSD95 (green) and Bassoon (magenta) clusters is used as the complementary task (localization). **b,** Two-channel confocal (left), real STED (center), and synthetic STED (right) images of the synaptic protein pair PSD95 and Bassoon. **c,** Evaluation of morphological features (area, perimeter, and eccentricity) of clusters in confocal (red), GT STED (orange) and synthetic STED (purple) for Bassoon clusters (n = 12 images, Figure S7). The difference between the distributions is statistically significant using a one-sided ANOVA test (Methods) between GT STED and confocal (area : $p \sim 10^{-11}$, perimeter : $p \sim 10^{-27}$, eccentricity : $p \sim 10^{-52}$) and between synthetic STED and confocal (area : $p \sim 10^{-9}$, perimeter : $p \sim 10^{-26}$, eccentricity : $p \sim 10^{-63}$). Difference between GT and synthetic STED is not significant (area : $p = 0.3$, perimeter : $p = 0.7$, eccentricity : $p = 0.07$). (Methods, Figure S6, S7 and S8). Scale bars: 1 µm.

TA-CycleGAN, a task-related loss is added to ensure that the biological structures (here, F-actin nanostructures) of the input image are similarly located and distributed in the translated and end-of-cycle images (Figure 4a).

A STED image of F-actin acquired in *fixed (F)* neurons is given as input to the $F \rightarrow L$ generator of the TA-CycleGAN which translates it into an image of the same structure but with image features (e.g. spatial resolution, signal-to-noise ratio, background level) corresponding to the domain of *live-cell (L)* images (Figure S9). The generated image is translated back into an image of fixed cells by the $L \rightarrow F$ generator (Figure 4a) and compared to the original image with a cycle-consistency loss and a generation loss (*CYC* and *GEN* in Figure 4a).

To reduce the annotation burden, we train the TA-CycleGAN using only the annotation of the fixed cell dataset (Methods). We then used the trained

TA-CycleGAN to infer a new synthetic dataset of F-actin in living neurons with its corresponding annotations for F-actin rings and fibers (Figure 4b, Methods). This dendritic *F-actin dataset adapted to live F-actin domain*, along with the expert annotations from the initial *dendritic F-actin dataset*, is used to train a segmentation network from scratch. This network, even if trained with synthetic images only, can perform well on the segmentation task of F-actin rings and fibers on real STED images acquired on living neurons (Figure 4c, Methods). To confirm that training on synthetic domain-adapted images generalizes to real live-cell STED images, a user study was conducted where an expert was asked to annotate F-actin rings and fibers on a set of 28 images using bounding box annotations (Methods). The segmentation network is tested on this set of unseen live-cell STED images, achieving an area under the Receiver Operating Characteristic (ROC) curve (AUROC) of 0.76 for rings and 0.83 for fibers using the user-study annotations as ground truth (Figure 4c, d, Figure S11). In comparison, a U-Net trained on images from fixed neurons achieves only an AUROC of 0.60 for rings and 0.59 for fibers (Figure 4c) on the same live-cell test images. Thus, domain adaptation with TA-CycleGAN enables the use of synthetic images to train a modality-specific segmentation network when no real annotated dataset is available for training. This strongly facilitates the cumbersome step in the training of any supervised machine learning method: creating data specific annotations.

Imaging Assistance: Automated modality selection with TA-GAN

Optimizing light exposure is of particular concern for live-cell imaging, where multiple acquisitions over an extended period of time might be required to observe a dynamic process. We integrate the generator of the TA-GAN for resolution enhancement of live F-actin in the acquisition loop of a STED microscope for time-lapse imaging of F-actin in living neurons (Methods). The TA-GAN-assisted microscope uses the generated synthetic images to guide the acquisition schemes of STED experiments by predicting the remodelling of F-actin nanostructures during live-cell STED acquisition sequences. The light exposure on the sample is reduced by relying on the synthetic STED images to decide at which time points the acquisition of a real super-resolved image of the region of interest (ROI) is necessary. At the beginning of an imaging sequence, a reference STED and confocal image pair is acquired. Thereafter, the TA-GAN uses the confocal image of the full field of view (FOV, 10 x 10 µm) to generate synthetic STED images of the same FOV (Figure 5a, Methods). The TA-GAN generates 10 possible synthetic STED images of the FOV for each confocal acquisition. The optical flow is measured from these synthetic images to determine the most variable sub-region (2 x 2 µm) (Methods). It serves as an estimation of the TA-GAN's uncertainty over the generated nanostructures. This most variable sub-region is imaged using the STED modality and given as input along with the confocal FOV to the generator of the TA-GAN. It primarily serves to minimize the signal variations encountered in live-cell

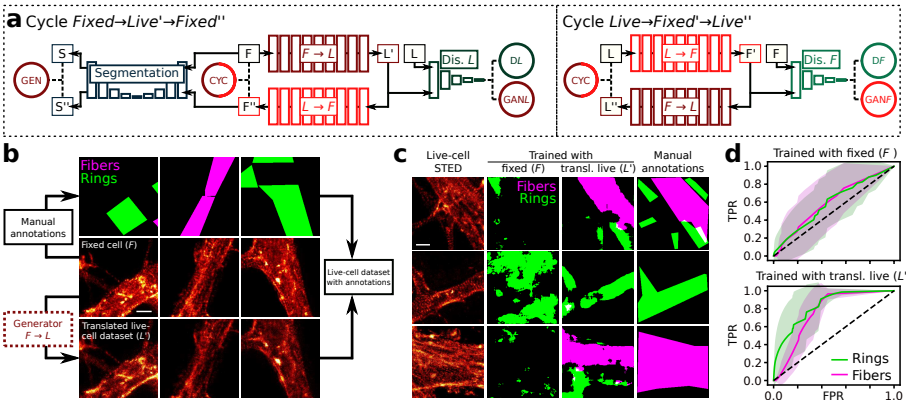

Fig. 4 TA-CycleGAN for unpaired domain adaptation. a, The TA-CycleGAN consists of a CycleGAN combined with a segmentation network as in TA-GAN. Losses: generation (GEN), cycle consistency (CYC), discrimination for live (DL) and fixed (DF) cells, GAN for live (GANL) and fixed (GANF) cells. Networks: fixed-to-live cell generator ($F \rightarrow L$, dark red), live-to-fixed cell generator ($L \rightarrow F$, light red), discriminator for live (Dis. L, dark green) and for fixed cells (Dis. F, light green), and segmentation for fixed cells (Segmentation, blue). **b,** The *dendritic F-actin dataset* (middle row), previously annotated by an expert with polygonal bounding boxes [3] (F-Actin rings: green, F-Actin fibers: magenta, top row), is translated to the live-cell domain using the fixed-to-live cell generator (bottom row, Methods). The annotations (top row) and the generated *live-cell F-actin dataset* (bottom row) are combined to train a network for the generation and semantic segmentation of F-actin nanostructures in living neurons (Methods). **c,** Segmentation of F-actin nanostructures on live-cell STED acquisitions. F-actin fibers and rings in living neurons are not properly segmented by a U-Net trained only with images of F-actin in fixed neurons (second column). The segmentation network trained with synthetic images generated by the TA-CycleGAN produces segmentation masks (third column) that are similar to the manual expert annotations (fourth column, Methods). See Figure S10 for more examples and DC distributions. **d,** ROC curves (TPR: true positive rate, FPR: false positive rate) for the segmentation of F-actin rings (green) and fibers (magenta) averaged over the manually annotated test set of real live-cell STED images (N=28). Top: the U-Net is trained on fixed-cell images. Bottom: The U-Net is trained on the same images translated to the live-cell domain using the TA-CycleGAN (Methods). Shaded region is the standard deviation. Dashed line represents the performance of a random classifier (AUROC=0.5). Scale bars: 1 µm.

imaging (Figure 5b, Methods). At each time point, the TA-GAN provides a prediction of the corresponding STED image for the full FOV. The synthetic STED image is segmented by the complementary task network to measure the dynamic remodelling of F-actin nanostructures (Figure 5c).

We also rely on the generated synthetic STED images for automated modality (confocal or STED) selection. The acquisition of a complete frame using the STED modality is triggered when either 1) the segmentation of the synthetic STED images changes when compared to the segmentation of the previous STED image (Figure 6, Methods), or 2) the TA-GAN confidence is low on the nanostructure predictions (Figure 7, Methods).

We apply TA-GAN-assisted image acquisition to detect the activity-dependent remodeling of dendritic F-actin from periodical rings into fibers in

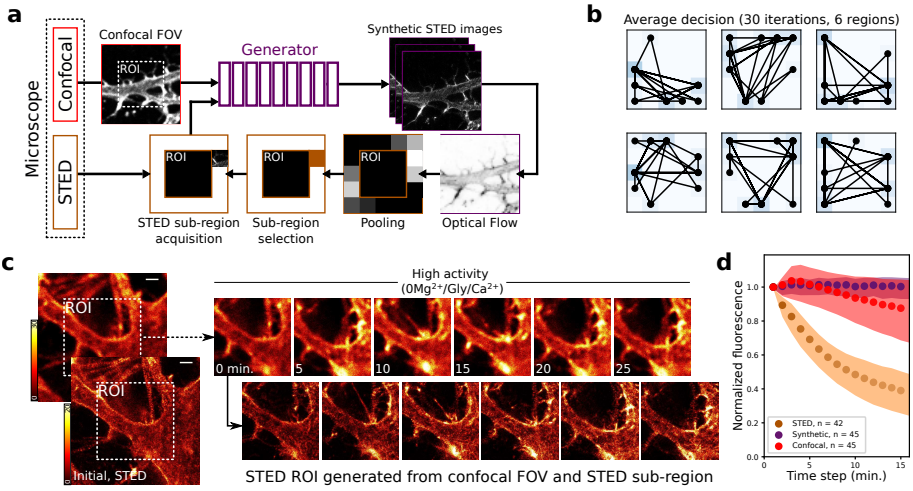

Fig. 5 Real-time assistance for live-cell STED microscopy. a, Image acquisition workflow integrating the TA-GAN for resolution enhancement in living neurons. Using optical flow on the full field of view (FOV), the sub-region with the highest uncertainty outside the region of interest (ROI) is identified and imaged with the STED modality (Methods). The acquired sub-region is used to improve the prediction of the nanostructures in the synthetic image of the ROI. **b,** STED sub-regions acquired over 30 iterations for 6 different FOV from two coverslips. The shade of each region is proportional to the number of times it was chosen using the optical flow measurement (darker regions were acquired more often). Subsequently acquired regions are connected with a black line. **c,** Time-lapse 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. All images from a given row are displayed using the same intensity scale (min:0, max:30 for confocal, min:0, max:20 for STED) to show the minimal effect of the acquisitions on the fluorescence intensity. **d,** Normalized fluorescence intensity after 15 confocal acquisitions over the full FOV (red, N=45 regions) and associated synthetic STED signal (purple, N=45 regions) in comparison to acquisitions using the STED modality (orange, N=42 regions). The TA-GAN predictions compensate for the fluorescence intensity decrease in the synthetic STED images. Scale bars: 1 µm.

living neurons. This remodeling was previously shown in fixed neurons but could not be monitored in living neurons due to technical limitations [3]. Monitoring F-actin nanoscale dynamics is particularly challenging as a strongly reduced concentration of the jasplaskinoid-based fluorogenic dye SiR-Actin is required to limit the F-actin stabilization effects associated with jasplakinoide [51]. This results in a low signal-to-noise ratio for time-lapse STED imaging of F-actin in living neurons. Additionally, photobleaching effects associated with repeated STED acquisitions can impair our ability to detect and quantify the presence of dim nanostructures such as F-actin rings (Figure 5d).

For the first acquisition scheme, we use the measurement of the *expected* activity-dependent remodelling of the F-actin lattice [3] to guide the imaging sequence. The acquisition of a complete STED image is triggered when changes

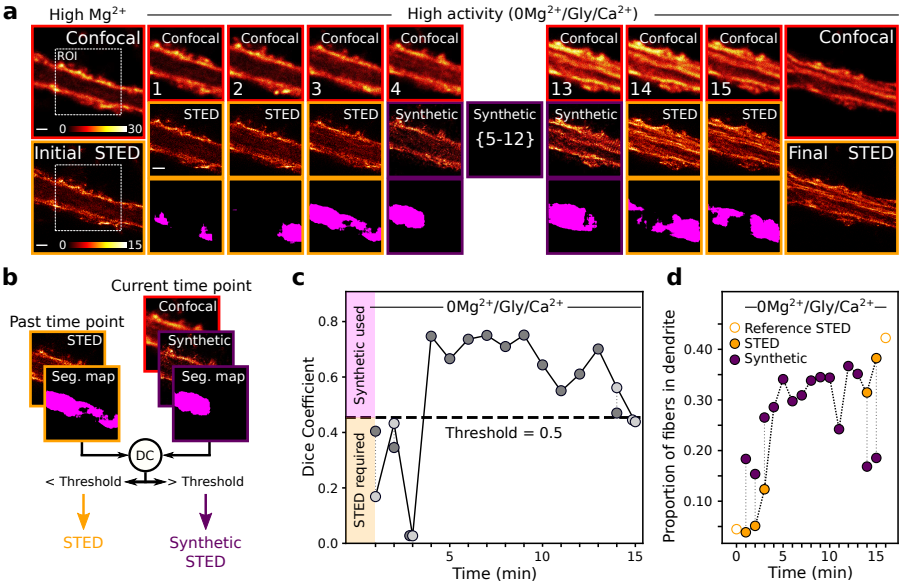

Fig. 6 Prediction of F-actin remodelling with TA-GAN a, Live-cell imaging of dendritic F-actin before (initial), during (frames 0-15) and after (final) application of a $0Mg^{2+}/Gly/Ca^{2+}$ solution, which promotes synaptic NMDA receptor activity. Shown are the confocal images (red, top row), synthetic STED images (purple, middle row), and real STED images when acquired (orange, middle row), and corresponding segmentation masks for F-actin fibers (bottom row) (See Figure S12 for all 15 frames. At each time point, a confocal 5 x 5 µm image is acquired and a synthetic STED image of the same FOV is generated. The synthetic STED image is segmented by the segmentation network, and the DC is computed using the segmentation of the last acquired real STED image as reference. Bottom row shows the segmentation output for the central ROI for the real STED if acquired, or for the synthetic STED otherwise. **b,** Modality selection method : a STED is acquired if the DC between the segmentation map of the last acquired real STED image and the segmentation map of the synthetic STED image of the current time point falls below a predefined threshold. **c,** DC measured at each time point. When the DC falls below a manually preset threshold (here 0.5), a real STED image is acquired (for this example, real STED images are acquired for frames 1, 2, 3, 14 and 15). Dark gray points indicate that the last acquired real STED (used to compute the DC) is from a previous time step; light gray points indicate that a new STED is acquired at this time step allowing to update the reference. **d,** Proportion of dendritic F-actin fibers at each time point segmented using the segmentation network trained on synthetic live-cell images (Methods) measured on either the real STED (orange) or the synthetic STED (purple). When a real STED acquisition is triggered, the proportion of fibers in both images is measured (dotted line). Initial and final reference STED images (empty orange bullets) are acquired at each round. In this specific case, there is an activity-dependent increase between the second and third minutes. See Figure S14 for an additional example of imaging sequence. Scale bars: 1 µm.

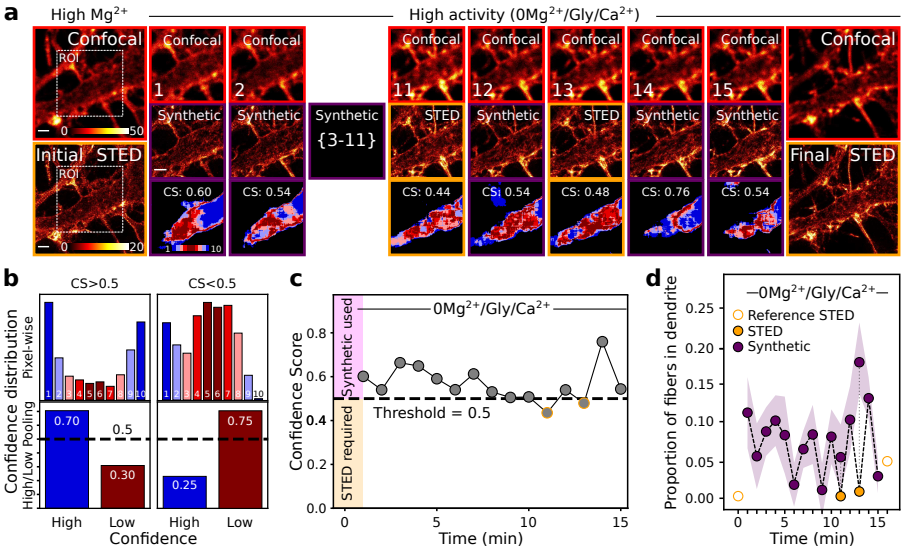

Fig. 7 Real-time assistance for live-imaging: monitoring the generator's confidence. **a,** Live-cell imaging of dendritic F-actin before and during a $0Mg^{2+}/Gly/Ca^{2+}$ stimulation. Shown are the confocal images (red, top row), synthetic STED images (purple, middle row), real STED images when acquired (orange, middle row), and the TA-GAN confidence maps (bottom row) (Methods, see Figure S13 for all 15 frames). Pixels predicted to belong to the same class (fibers or not fibers) in $\geq$80 % of the TA-GAN predictions are defined as high-confidence pixels (blue) and pixels predicted to belong to the same class in $<$80 % of the TA-GAN predictions are defined as low-confidence pixels (red). **b,** The high confidence pixels (blue) are binned together against the low-confidence pixels (red). **c,** The TA-GAN confidence score corresponds to the proportion of high-confidence pixels at each frame. This score falls below 0.5 when the number of low-confidence pixels exceeds the number of high-confidence pixels (b, right), which triggers the acquisition of an image with the STED modality, as seen at time steps 11 and 13. **d,** Proportion of fibers on the dendrite on either the real STED (orange) or the synthetic STED (purple). Initial and final reference STED images (empty orange bullets) are acquired at each round. The mean (dashed line) and standard deviation (shaded area) in the proportion of F-actin fibers is computed for the 10 predicted synthetic images. In this specific case, there is no activity-dependent increase. Scale bars: 1 µm.

in the segmentation masks of the generated synthetic image are detected. Full confocal and STED FOVs are first acquired (Figure 6a). The segmentation output of the first STED image is used as the initial reference. For each subsequently acquired confocal image, a synthetic STED is generated and segmented for F-actin fibers (Methods). The DC between the segmentation masks of the synthetic STED image and the reference STED image is computed (Figure 6b). A real STED image is acquired if the DC between the segmentation mask of the synthetic and the reference real STED image is below a predefined threshold (Figure 6c, threshold fixed at 0.5 for this example). The proportion of fibers in the dendrite is measured in the real and synthetic STED

images using the segmentation network trained on synthetic live-cell images (Figure 6d, Methods). At the end of the image acquisition sequence, confocal and STED images of the full FOV are acquired to validate the formation of fibers and compare with the synthetic generations (Figure 6a, last frame). The value of the DC threshold is chosen based on preliminary imaging trials and previous knowledge about the remodelling extent and dynamics, which is not always available prior to the imaging experiment. Using the segmentation masks to trigger a STED acquisition requires the user to know what structure is expected to either appear, disappear or change in the course of the imaging sequence. For the remodelling of F-actin rings, we use the segmentation of fibers because, based on experimental results on fixed cells, the distribution of pixels segmented as fibers is expected to increase following the change of activity condition [3]. This requirement of prior knowledge could possibly prove to be cumbersome to image an unexpected transformation.

We developed a second method to trigger STED image acquisitions, which is based on the confidence of the TA-GAN with respect to the image generation task. This approach is particularly useful when no previous knowledge on the expected structural change is available prior to the experiment. For each frame, a confocal image of the full FOV is acquired from which 10 synthetic super-resolved images are predicted by the TA-GAN (Figure 7a). A STED image of the full FOV should be acquired when the predicted nanostructures strongly vary between the 10 synthetic images, which could be an indicator for a structural remodelling. We compare the segmentation maps of 10 generated synthetic STED images and calculate the variability in the predicted class for each pixel in the 10 images. We define pixels that are assigned in more than 80 % of the synthetic images to the same class (fibers or background) as having a low variability (high confidence of the TA-GAN, Figure 7b). Otherwise, when the segmentation network assigns a pixel in less than 80 % of the generated images to the same class, the variability is high (low confidence of the TA-GAN). The proportion of high confidence pixels in the image corresponds to the confidence score. When the confidence score is below 0.5, a full STED image is acquired (Figure 7c,d, frames 11 and 13). This approach can be beneficial to detect unexpected patterns and rare events, but can lead to the acquisition of multiple consecutive frames if the measured confocal pattern is associated with a high variability in the generated images. Modulation of the STED modality acquisition frequency can be achieved by adapting the DC or confidence thresholds to the experimental settings.

3 Discussion

We have introduced TA-GAN for resolution enhancement and domain adaptation and have demonstrated its applicability to STED nanoscopy. We have shown that a complementary task assisting the training of a generative network for resolution enhancement and domain adaptation improves the reconstruction accuracy of nanoscopic structures in optical microscopy images. Using a

segmentation loss to optimize the image generation process results in more reliably generated F-actin nanostructures in terms of general appearance and, most importantly, distribution in axons (Figure 1) and dendrites (Figure 2). The applicability of our method was also demonstrated using a paired confocal and STED microscopy dataset of synaptic protein pairs in fixed neurons (Figure 3). We showed that the TA-GAN method is flexible and can be trained with different complementary tasks such as binary segmentation, semantic segmentation, and localization. For unpaired datasets, we have introduced the TA-CycleGAN model and demonstrated how the structure preserving domain adaptation opens up the possibility to create paired datasets of annotated images that cannot be acquired simultaneously (Figure 4). We have used the TA-CycleGAN generated *synthetic* STED images of F-actin to train a segmentation network that performs well for the segmentation of F-actin nanostructures in *real* STED images, without the need for manual reannotations of the new live-cell imaging dataset. This segmentation network was also used to train a TA-GAN for resolution enhancement in living neurons. The resulting confocal to STED generator for live-cell imaging was incorporated into the acquisition loop of a STED microscope (Figure 5). We have validated how the trained TA-GAN model can be helpful in assisting a microscopist by automating decisions in live-cell optical nanoscopy acquisition sequences. The resolution enhancement TA-GAN provides increases the informative value of each confocal acquisition and automatically triggers the acquisition of a STED image only in the regions and time-steps where this acquisition is informative enough due to variations or uncertainties in the predicted nanostructures (Figures 6 and 7).

We have used the variability in the output of the generator to quantify its confidence and to trigger the acquisition of a STED image. Future work in calibrating the network's probabilistic output could lead to an improved quantification of its confidence. Multiple successive frames could also be given as input to the generator to introduce temporal information instead of using static frames individually. This could enable the generator to decode the rate of biological change and introduce this knowledge to the next frame prediction, leading to smoother transitions between synthetic images. The TA-GAN model, as presented here, enables the visualization of biological dynamics over longer sequences with reduced photobleaching effects. Thus, TA-GAN-assisted STED nanoscopy can guide microscopists for optimized acquisition schemes and reduced light exposure.

Acknowledgments. Francine Nault and Sarah Pensivy for neuronal cell culture. Gabriel Leclerc for the FIJI macro for segmentation. Annette Schwerdtfeger for proofreading the manuscript. Funding was provided by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) (F.L.C. and C.G.), the Canada First Research Excellence Fund (F.L.C. and C.G.), the Canadian Institute for Health Research (CIHR) (F.L.C.), and the Neuronex Initiative (National Science Foundation 2014862, 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 NSERC, from the Fonds de Recherche Nature et Technologie (FRQNT) Quebec, from the FRQNT strategic cluster UNIQUE, and by a Leadership and Scientific Engagement Award from Université Laval. T.W. is supported by postdoctoral research funding from the FRQNT strategic cluster UNIQUE. A.B. is supported by scholarships from NSERC, FRQNT and from the FRQNT strategic cluster UNIQUE.

Declarations

Availability of data and materials

The datasets used to train and test the TA-GAN model and the baselines are available for download at https://s3.valeria.science/flclab-tagan/index.html.

Code availability

The codes, 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 train the networks for new

Competing interests

The authors declare no competing interests.

Author contributions

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

Methods

4 Sample preparation and STED microscopy

Cell culture

Dissociated rat hippocampal neurons were prepared as described previously [3, 52] in accordance with and approved by the animal care committee of Université Laval. For live-cell STED imaging, the dissociated cells were plated on PDL-Laminin coated glass coverslips (18 mm) at a density of 322 cells/mm^2 and used at DIV 12-16.

STED microscopy

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 was equipped with a motorized stage and auto-focus unit.

Live-cell imaging

The cultured neurons were pre-incubated 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) using a gravity driven perfusion system. After identification of the regions of interest, the perfusion solution was switched to HEPES buffered aCSF containing high Ca^{2+}, Glycine and without Mg^{2+} (in mM: NaCl 98, KCl 5, HEPES 10, Glycine 0.2, $CaCl_2$ 2.4, Glucose 10). Solutions were adjusted to an osmolality of 240 mOsm per kg and a pH of 7.3.

5 Image Datasets

Axonal F-actin dataset

The *Axonal F-actin dataset* was used to train the TA-GAN for confocal-to-STED super-resolution of axonal F-actin rings using a binary segmentation task. The original dataset consisted of 516 paired confocal and STED images (224 x 224 pixels) of axonal F-actin from Lavoie-Cardinal *et al.* [3]. 31 images from the original dataset were discarded for not containing annotated F-actin rings or fiber structures. The remaining images were split into a training set (377 images), a validation set (56 images), and a testing set (52 images). The manual polygonal bounding box annotations of the axonal F-actin periodical lattice (F-actin rings) from the original dataset were retained (Figure 1b).

Dendritic F-actin dataset

The *Dendritic F-actin dataset* was used to train the TA-GAN for confocal-to-STED super-resolution of dendritic F-actin nanostructures using a semantic segmentation task as well as to train the TA-CycleGAN for live and fixed domain adaptation. The original dataset from Lavoie-Cardinal *et al.* [3] was split into a training set (304 images), a validation set (54 images), and a testing set (26 images, 12 for low activity and 14 for high activity). This dataset consisted in paired confocal and STED images (20 nm pixel size for both) of the dendritic F-actin cytoskeleton in fixed hippocampal dissociated neurons, which had been manually annotated using polygonal bounding boxes (Figure 2a). The training and validation crops were taken from large STED images (between 500 × 500 and 3000 × 3000 pixels) using a sliding window of size 224 x 224 pixels with no overlap. If less than 1 % of the pixels of the crop were annotated as containing a structure of interest (F-actin rings and/or fibers), the crop was discarded from the set. This operation resulted in 4,331 crops for training and 659 crops for validation.

Synaptic protein dataset

Images from the original dataset from Wiesner *et al.* [4] were split into a training set (81 images), a validation set (22 images) and a testing set (12 PSD95-Bassoon and 7 PSD95-Homer1c images). The confocal and STED images from the training and validation sets were first registered using the pipeline presented in Figure S15, resulting in 1,841 crops for training and 509 crops for validation. The localization maps were generated by automatically segmenting the STED images using wavelet transform decomposition [5] with the same parameters as Wiesner *et al.* [4] (scales 3 and 4). No segmented clusters were discarded based on size or position, following the intuition that even the smallest structures should be generated. The localization map was created from a black image by placing a white pixel at the position of the intensity-weighted centroid of each segmented cluster, and then applying a gaussian filter with a variance of 2 (Figure S5).

Live F-actin dataset

The *Live F-actin dataset* was used to train the TA-CycleGAN for live and fixed domain adaptation and the segmentation network for live-cell images. A large FOV was first imaged at low-resolution from which regions were manually selected by an expert. For each region selected, a confocal and a STED image were acquired sequentially. The *Live F-actin dataset* consists in 904 paired STED and confocal images of F-actin stained with the fluorogenic dye SiR-Actin (Spirochrome, US) in living hippocampal cultured neurons. The dataset was split into a training set (833 images) and a validation set (71 images). The images were of variable size (from a minimum width of 2.76 to a maximum of 49.1 µm). We chose to exploit the variable size of the input images for data

augmentation by performing a cropping step during training. Refer to the Cell Culture and STED microscopy sections for cell cultures and imaging protocols.

Dendritic F-actin dataset adapted to the live-cell STED imaging domain

The *Domain adapted dendritic F-actin dataset* was used to train the resolution enhancement TA-GAN for live-cell images. This dataset corresponds to the *Dendritic F-actin dataset* adapted to the live-cell STED imaging domain using the TA-CycleGAN for fixed-to-live domain adaptation (TA-CycleGAN for domain adaptation section). It contains the same number of images, the same train/valid/test splits, and the same image characteristics (crop size, pixel size, annotations) as the *Dendritic F-actin dataset*.

Training procedures

TA-GAN for resolution enhancement

The TA-GAN was developed from the conditional GAN model for image-to-image translation [2], available at https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix. It differs by its computation of the generation loss. Comparable methods using cGANs for enhancing the resolution of microscopy images are trained using pixel-wise metrics such as mean squared error (MSE) [5], absolute error [11, 12] or structural similarity index [13, 14] (Figure S1a). For the TA-GAN, the generation loss is computed by comparing the output of a complementary task network applied on the real and synthetic images (Figure S1b,c). The training of TA-GAN has two dataset-specific aspects: the task and the data format. Table 1 summarizes the settings for the four resolution enhancement experiments presented in this paper, and table 2 presents the hyperparameters used for training the TA-GAN for each of these experiments.

	Axonal F-actin [3]	**Dendritic F-actin [3]**	**Synaptic Proteins [4]**	**Live F-actin**
Input	Confocal (1 channel)	Confocal (1 channel)	Confocal (2 channels)	Confocal + decision matrix + STED sub-region
Output	STED (1 channel)	STED (1 channel)	STED (2 channels)	STED (1 channel)
Task	Seg. of axonal F-actin rings	Semantic seg. of dendritic F-actin rings and fibers	Localization of synaptic protein cluster centroids	Semantic seg. of dendritic F-actin rings and fibers
Annotations	Bounding boxes (1 class)	Bounding boxes (2 classes)	Wavelet detected centroids (2 channels)	None
Pretrained Task net.	No	No	No	Yes

Table 1 Summary of experimental settings using the TA-GAN for confocal-to-STED resolution enhancement.

Hyperparameters	Axonal F-actin [3]	Dendritic F-actin [3]	Synaptic Proteins [4]	Live F-actin
Number of training images	377	4,331	1,841	753
Number of validation images	56	659	509	47
Image size (px)	224 x 224	224 x 224	512 x 512 x 2	variable
Generator architecture	9-blocks ResNet [39]			
Task network architecture	6-blocks ResNet			U-Net 128 [53]
Discriminator architecture	PatchGAN [40]			
Batch size	8	32	32	16
Learning rate	0.0002			
Weight of GAN loss	1			
Weight of GEN loss	10	1	1	10
Data augmentation methods	Flip, 90 rotations, random crops			
Crop size (px)	128 x 128	128 x 128	128 x 128 x 2	256 x 256 x 3
Number of epochs	1000	500	2000	5000

Table 2 Hyperparameters used to train the TA-GAN model on the four datasets presented.

Resolution enhancement using a segmentation task

The TA-GAN models for confocal-to-STED resolution enhancement were first trained using a segmentation network to compute the generation loss. The output of the segmentation network was compared with manual bounding box annotations of the same structure produced by an expert [3] using a MSE loss. The loss computed from the real STED image (Task Loss, TL in Figure 1) was back-propagated to the segmentation network to optimize its weights, and the loss from the synthetic STED image (GEN) optimizes the generator. The other losses computed were standard cGAN losses : the GAN loss (GAN: misclassification of synthetic images as real images), the discriminator losses (DR: classification of real images as real, and DG: classification of generated images as synthetic). The validation losses were not used for early-stopping because of the adversarial nature of GANs. The validation images were instead used as a qualitative assessment of the training progress to select the best iteration for testing the model.

For the *Axonal F-actin dataset* (Axonal F-actin dataset section), the complementary task was the segmentation of axonal F-actin rings. The output of the segmentation network was the predicted segmentation maps of F-actin rings. For the *Dendritic F-actin dataset* (Dendritic F-actin dataset section), the complementary task was the semantic segmentation of dendritic F-actin rings and fibers. The output of the segmentation network was a two-channel image, with the predicted segmentation of rings in the first channel and of fibers in the second channel. The two first columns of Table 2 present the hyperparameters used to train these two models.

Resolution enhancement using a localization task

A TA-GAN model for confocal-to-STED resolution enhancement was trained using a localization network to compute the generation loss. The localization network took a STED image as input to output a map of dots indicating the

intensity-weighted centroids of all detected clusters in the STED image. This model was trained on the *Synaptic protein dataset* (Synaptic protein dataset section) using the hyperparameters presented in Table 2, column 3. The input of the generator was a two-channel confocal image rescaled and registered to the STED image. The generation loss (GEN in Figure 1) was the MSE between the wavelet-generated localization maps on the real STED image and the localization predictions from the task network on the synthetic image.

Resolution enhancement for live-cell imaging

The TA-GAN for resolution enhancement of live-cell STED imaging was trained on the *Live F-actin dataset* (Live F-actin dataset section). The complementary task was the semantic segmentation of dendritic F-actin rings and fibers. Since the *Live F-actin dataset* did not include any labels, the segmentation network was pretrained (U-Net for the segmentation of F-actin nanostructures in living neurons section) on the *Domain adapted dendritic F-actin dataset* (Dendritic F-actin dataset adapted to the live-cell STED imaging domain and TA-CycleGAN for domain adaptation section). This pretrained network was not updated during the TA-GAN training, but it was used to compute the MSE generation loss between the segmentation prediction of the real and the synthetic STED images. The fourth column of Table 2 shows the hyperparameters used for this training. The input of the generator was the confocal image concatenated with two channels of the same size. The first additional channel contained a sub-region acquired with the STED modality, with the rest of the image filled with 0s (Figure 5a, *STED sub-region acquisition*). The second additional channel was filled with 0s, with the position of the STED sub-region indicated by 1s (Figure 5a, *STED sub-region selection*). The selection of the STED sub-region served as data augmentation; at each iteration, one of the 16 possible sub-regions was selected from the full STED image. Training using this three-channel input enabled the generator to learn features from the STED sub-region and turned the resolution enhancement task into an image completion task.

TA-CycleGAN for domain adaptation

The TA-CycleGAN model was developed from the CycleGAN model [2]. As for the standard CycleGAN, the TA-CycleGAN consists of 4 networks : two generators (one that translates the domain of fixed-cell STED imaging (F) into domain of live-cell STED imaging (L), the other domain F into domain L), and two discriminators (one for domain F, the other for domain L), which are combined with a fifth network : the complementary task network (Figure 4a). The TA-CycleGAN was applied to non-paired images, where the prediction of the generator for a given input cannot be compared to a corresponding ground truth. Instead, the generated synthetic image was passed through a second generator and converted back to the input domain where it was compared to the initial image (ground truth) for the computation of losses.

The TA-CycleGAN for fixed-to-live domain adaptation was trained using two datasets: STED images of the *Dendritic F-actin dataset* (F, Dendritic F-actin dataset section) and the *Live F-actin dataset* (L, Live F-actin dataset section). The complementary task was the semantic segmentation or F-actin rings and fibers on the fixed-cell STED images, for which manual bounding box annotations were available [3]. The U-Net segmentation network was already optimized for the semantic segmentation of F-actin rings and fibers in fixed-cell STED images [3]. The generation loss was the MSE between the U-Net segmentation prediction on the real fixed cell image (F), and the end-of-cycle fixed cell image (F") (Figure 5a).

U-Net for the segmentation of F-actin nanostructures in living neurons

The live-cell domain segmentation network was built around a U-Net-128 [53] 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 dataset* (Dendritic F-actin dataset section) was translated into live-cell domain using the $F \rightarrow L$ generator (TA-CycleGAN for domain adaptation section, Figure S9). The manual annotation from the fixed cell images were associated with the corresponding synthetic images from the live-cell domain (Figure S9a).

Random crops of 128 x 128 pixels of the translated synthetic images and their corresponding annotations were used to train a segmentation network on images of the live-cell domain. Horizontal and vertical flips were used for data augmentation. Due to class imbalance in the training set, the segmentation loss for fibers was weighted by a factor of 2.5, which reflected the ratio of total annotated pixels for each class. 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. The segmentation maps were binarized to compute the DC (TA-GAN assisted monitoring of expected structural change) and the summed prediction (Monitoring the pixel-wise generator's confidence for previously unknown structural remodelling section). The optimal threshold to binarize the segmentation prediction was determined as the value that reached the optimal DC over the validation set (-0.53 for the raw output predictions).

Baselines

The baselines were, unless specified, trained and tested on the *Axonal F-actin dataset* (Axonal F-actin dataset section) to produce the results in Figure 1 and on the *Dendritic F-actin dataset* (Dendritic F-actin dataset section) for the results in Figure S4.

Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN x4)

The trained version of Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) [41], available at https://github.com/xinntao/ESRGAN, was directly applied to the test images from the *Axonal F-actin dataset* (Figure 1b). The network was not trained on our datasets and was not expected to achieve great performance, since it was trained on natural images. This baseline was used to demonstrate that the confocal-to-STED transformation is not an upsampling task, justifying the need for methods that differ from the super-resolution methods developed for natural images.

Denoising convolutional neural networks (DnCNN)

The trained version of DnCNN [42] available at https://github.com/yinhaoz/denoising-fluorescence was directly applied to our test images from the *Axonal F-actin dataset* (Figure 1b). A version of the network trained on the fluorescence microscopy denoising dataset [54] 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 is not a denoising task, but a generation task.

Content-Aware image REstoration (CARE)

Content-Aware image REstoration (CARE) [11] was implemented from the public GitHub repository (https://github.com/CSBDeep/CSBDeep). We used the standard CARE network for image restoration and enhancement. The residual U-Net generator was optimized from scratch with the same training and validation images as the TA-GAN. All default hyperparameters were used and the model was trained for 300 epochs using a mean absolute error loss. The epoch that reached the lowest validation loss was used for testing.

Three-dimensional residual channel attention networks (3D-RCAN)

Three-dimensional residual channel attention networks (3D-RCAN) [12] was implemented with Tensorflow and Keras. The code was taken from the publicly available GitHub repository (https://github.com/AiviaCommunity/3D-RCAN). All default hyperparameters were used and the model was trained over 300 epochs to ensure convergence of the validation loss. The model reaching the lowest validation loss was used for testing.

Conditional GAN for image-to-image translation (pix2pix)

Pix2pix [2] 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 2), replacing only the generation loss with

a pixel-wise MSE loss between the ground truth and generated STED images (Figure S1a). The results from this baseline are compared to the TA-GAN for the generation of axonal F-actin rings in Figure 1c and Figure S2, and for the generation of dendritic F-actin in Figure S4.

Evaluation of networks performance

Segmentation of axonal F-actin nanostructures in synthetic STED images

Synthetic images were generated by the TA-GAN from the confocal images from the test set of the *Axonal F-actin dataset* (52 images) and of the *Dendritic F-actin dataset* (26 images) (Axonal F-actin dataset and Dendritic F-actin dataset sections). The MSE was computed between the ground truth and synthetic STED images of the test set of the *Axonal F-actin dataset* (Figure S2b). Additionally, a U-Net that was trained on real STED images only [3] (available at https://github.com/FLClab/STEDActinFCN) was used to produce segmentation masks of F-Actin nanostructures on the ground-truth and synthetic STED image pairs (Figure S2c and Figure 2). The segmentation masks were compared with manual bounding box annotations using the DC [55] (Figure S2d). The statistical significance results reported in Figure 1c and Figure S2b,d were computed with the Mann-Whitney U test[1].

Assessment of synaptic protein cluster morphology

We characterized the morphological features of the generated synaptic protein clusters on the *Synaptic protein dataset*. The area, perimeter, and eccentricity of the protein clusters were measured in the confocal images, STED images and synthetic STED images (Figure 3, S7 and S8). The distribution of each morphological feature over all associated clusters was computed using a Python library for Statistical Object Distance Analysis (pySODA) [4]. The same parameters as in Wiesner *et al.* [4] were used for the analysis: wavelet segmentation scales of 3 and 4, a minimum cluster area of 5 pixels, and minimum cluster width/height of 3 pixels. The weighted centroids of the detected clusters were calculated on the raw STED images. A foreground mask was generated following Wiesner *et al.* [4]: applying a gaussian blur (standard deviation of 10) on the sum of both STED channels, and thresholding the image using 50% of the mean intensity value. Only clusters from the foreground mask were considered for the analysis. The statistical significance results reported in Figure 3, Figure S7, and Figure S8 were computed from a one-sided ANOVA test over all clusters identified as coupled by pySODA analysis.

User-study for the segmentation of live F-actin images

A set of 28 STED images (224 x 224 pixels) from the *Live F-actin dataset* test set was labeled by an expert using a FIJI [56] macro to test the performance of

the segmentation network trained on synthetic live-cell images. The 28 images were chosen among a set of 150 test images. To avoid testing on images that do not include at least one of the two nanostructures of interest (rings or fibers), the output of the segmentation network was used to make this selection: the 14 images that had the most pixels identified as rings, and the 14 images that had the most pixels identified as fibers, were kept for the user-study. Each STED image was presented to the expert who was prompted to draw bounding boxes that enclose all regions identified as F-actin rings, then fibers, using the polygon tool. The expert had access to all FIJI functions to facilitate the process, such as contrast enhancing and zooming. In addition to the set of live-cell images, a second set of 28 synthetic live-cell STED images, selected from the *Domain adapted dendritic F-actin dataset* validation set using the same selection criteria, was also included in the user-study. The expert was presented with an image from one of the two sets, without being informed whether the image was real or synthetic.

TA-GAN-assisted live-cell STED microscopy

TA-GAN integration in the acquisition loop

The generator trained for resolution enhancement for live-cell imaging (Resolution enhancement for live-cell imaging section) was directly integrated in the imaging acquisition process of the STED microscope (Figure 5a). A region of interest (ROI) of 10 x 10 µm was selected and reference STED and confocal images were acquired. Next, at each time point (to monitor F-actin remodelling, a new image was acquired every minute), the steps described in Table 3 were performed.

1	A confocal image of the ROI is acquired
2	10 synthetic images of the ROI are generated (Resolution enhancement for live-cell imaging section)
3	The sub-region (2x2 µm) of lowest confidence within the ROI is identified (Optical flow computation for sub-region selection section)
4	A real STED image of this sub-region is acquired
5	The confocal image of the full ROI and of the STED sub-region are used by the TA-GAN to produce 10 synthetic STED image of the full ROI (Resolution enhancement for live-cell imaging section)

Table 3 Steps performed at each time point for automated TA-GAN assistance.

Note that step 5 can be done post-acquisition if real-time visualization of the synthetically enhanced image is not required. This whole process was repeated for 15 iterations at 1 iteration/minute. For the last iteration, a STED and confocal images of the full region were acquired.
Steps 2, 3 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, 3 and 4 were sent from the microscope's control computer to a GPU-equipped computer using the Flask [57] web framework Python module. All

automated acquisitions were programmed using the SpecPy Python library to interface with the Imspector software (Abberior Instruments, Germany).

Optical flow computation for sub-region selection

The optical flow was used to choose the most informative sub-region to guide TA-GAN synthetic image generation. It was computed using a Python implementation of the Horn–Schunck method [58] with the Python multiprocessing library, parallelizing the computations on 8 CPUs to increase the computation 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 intensity values of the pixels (Figure S16). To translate the pixel-wise optical flow to a region-wise map of image variability, the 500 x 500 pixel optical flow image is downsampled to a 5 x 5 map using the mean of the region. The sub-region with a maximum mean optical flow was selected as the most variable (Figure 5b, Figure S16).

Live-cell imaging decision guidance using the TA-GAN

The TA-GAN predictions were used for decision guidance on the optimal STED and confocal acquisition sequence. Two approaches could be used to determine when to acquire a STED image of the full ROI depending on the type of structural remodelling studied.

TA-GAN assisted monitoring of expected structural change

The proof-of-concept experiment targeted the *expected* activity-dependent remodelling of dendritic F-actin rings into fibers [3]. Since the area of F-actin fibers was expected to increase following a neuronal stimulation, structural remodelling was monitored by comparing the area of segmented F-actin fibers on the synthetic and the reference real STED images. F-actin fibers were segmented on the synthetic STED images by the U-Net described in the U-Net for the segmentation of F-actin nanostructures in living neurons section. At each time point, steps 1-5 are performed as described in Table 3. After step 5, 10 synthetic images of the ROI were generated and segmented by the U-Net. The mean of the 10 segmentation maps predicted for the synthetic images was compared to the segmentation map predicted for the reference real STED image using the DC metric. A low DC was indicative of a change in respect to the reference image in the spatial distribution of the dendritic fibers. A full real STED image was acquired if the DC fell below 0.5. The value of 0.5 was chosen by performing several trials on live-cell F-actin imaging. The value of the DC threshold should be adapted to the type of structural remodelling observed. Each time the acquisition of a STED on the full ROI was triggered, the STED reference image was updated for subsequent comparison of the segmentation maps.

Monitoring the pixel-wise generator's confidence for previously unknown structural remodelling

The pixel-wise generator's confidence can be used when the structural remodelling cannot be precisely defined prior to the experiments. At each time point, steps 1-5 were performed as described in Table 3. Following step 5, 10 synthetic STED images of the ROI were generated. The 10 synthetic images were segmented by the U-Net, resulting in 10 segmentation maps for F-actin rings and fibers. The 10 segmentation maps of F-actin fibers were binarized (U-Net for the segmentation of F-actin nanostructures in living neurons section) and summed. Pixels in the summed segmentation prediction had a value between 0 and 10. The confidence of the generator on the segmentation prediction was evaluated from the summed segmentation prediction. High confidence pixels were associated with those having the same value for at least 80% of the predicted segmentation maps (1-2 : no fibers; 9-10 : fibers). Pixels with values in between (3-7, inclusive) were considered low confidence since the segmentation of the F-actin fibers was not consistent between the synthetic STED images. The distribution of high confidence and low confidence pixels (Figure 7b) was compared for each image (pixels with a value of zero for the summed segmentation maps were not considered in the distribution as they mostly corresponded to background pixels). The proportion of high confidence pixels was defined as the confidence score (CS). If the CS was above 0.5, the generator was consistent in its predictions and therefore confident. If the CS was below 0.5, the synthetic STED images were not consistent and therefore not reliable; a STED acquisition was therefore required. The threshold of 0.5 was chosen because it corresponds to the tipping point where the number of low confidence pixels exceeds the number of high confidence pixels.

References

[1] Sahl, S. J., Hell, S. W. Jakobs, S. Fluorescence nanoscopy in cell biology. *Nature reviews Molecular cell biology* 18, 685–701 (2017).

[2] Hell, S. W. Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters* 19, 780–782 (1994).

[3] Durand, A. et al. A machine learning approach for online automated optimization of super-resolution optical microscopy. *Nature communications* 9, 1–16 (2018).

[4] Laissue, P. P., Alghamdi, R. A., Tomancak, P., Reynaud, E. G. Shroff, H. Assessing phototoxicity in live fluorescence imaging. *Nature methods* 14, 657–661 (2017).

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

[6] Wu, Y. et al. Multiview confocal super-resolution microscopy. *Nature* 1–6 (2021).

[7] Nehme, E., Weiss, L. E., Michaeli, T. Shechtman, Y. Deep-storm: super-resolution single-molecule microscopy by deep learning. *Optica* 5, 458–464 (2018).

[8] von Chamier, L. et al. Democratising deep learning for microscopy with zerocostdl4mic. *Nature communications* 12, 1–18 (2021).

[9] Bilodeau, A. et al. Microscopy analysis neural network to solve detection, enumeration and segmentation from image-level annotations. Nature Machine Intelligence 1–12 (2022).

[10] 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).

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

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

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

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

[15] Baniukiewicz, P., Lutton, E. J., Collier, S. Bretschneider, T. Generative adversarial networks for augmenting training data of microscopic cell images. *Frontiers in Computer Science* 10 (2019).

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

[17] 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).

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

[19] Li, Y., Zhang, K., Shi, W., Miao, Y. Jiang, Z. A novel medical image denoising method based on conditional generative adversarial network. *Computational and Mathematical Methods in Medicine* 2021 (2021).

[20] Sandfort, V., Yan, K., Pickhardt, P. J. Summers, R. M. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific reports* 9, 1–9 (2019).

[21] Kaiser, B. Albarqouni, S. MRI to CT translation with GANs. *arXiv preprint arXiv:1901.05259* (2019).

[22] Wu, Z., Wei, J., Yuan, W., Wang, J. Tasdizen, T. Inter-slice image augmentation based on frame interpolation for boosting medical image segmentation accuracy. *arXiv preprint arXiv:2001.11698* (2020).

[23] Calivá, F. et al. Breaking speed limits with simultaneous ultra-fast MRI reconstruction and tissue segmentation. In *Medical Imaging with Deep Learning*, 94–110 (2020).

[24] 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).

[25] Cohen, J. P., Luck, M. Honari, S. Distribution matching losses can hallucinate features in medical image translation. In *International conference on medical image computing and computer-assisted intervention*, 529–536 (Springer, 2018).

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

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

[28] Li, X. et al. Unsupervised content-preserving transformation for optical microscopy. *Light: Science Applications* 10, 1–11 (2021).

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

[30] Pawley, J. B. Fundamental limits in confocal microscopy. In *Handbook of biological confocal microscopy*, 20–42 (2006).

[31] Ruder, S. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098* (2017).

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

[33] Ren, M., Dey, N., Fishbaugh, J. Gerig, G. Segmentation-renormalized deep feature modulation for unpaired image harmonization. *IEEE Transactions on Medical Imaging* 40, 1519–1530 (2021).

[34] 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).

[35] Jaiswal, A. et al. Controlling BigGAN image generation with a segmentation network. In *International Conference on Discovery Science*, 268–281 (2021).

[36] Falk, T. et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nature methods* 16, 67–70 (2019).

[37] 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).

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

[39] 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).

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

[41] ESRGAN: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, 0–0 (2018).

[42] 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).

[43] Mann, H. B. Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics* 50–60 (1947).

[44] D'Este, E., Kamin, D., Göttfert, F., El-Hady, A. Hell, S. W. STED nanoscopy reveals the ubiquity of subcortical cytoskeleton periodicity in living neurons. *Cell reports* 10, 1246–1251 (2015).

[45] Caicedo, J. C. et al. Data-analysis strategies for image-based cell profiling. *Nature methods* 14, 849–863 (2017).

[46] Vu, Q. D. et al. Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in bioengineering and biotechnology* 53 (2019).

[47] Perez, A. J. et al. A workflow for the automatic segmentation of organelles in electron microscopy image stacks. *Frontiers in neuroanatomy* 8, 126 (2014).

[48] 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).

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

[50] Lukinavičius, G. et al. Fluorogenic probes for live-cell imaging of the cytoskeleton. *Nature methods* 11, 731–733 (2014).

[51] Bubb, M. R., Senderowicz, A., Sausville, E. A., Duncan, K. Korn, E. D. Jasplakinolide, a cytotoxic natural product, induces actin polymerization and competitively inhibits the binding of phalloidin to f-actin. *Journal of Biological Chemistry* 269, 14869–14871 (1994).

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

[53] 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).

[54] 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).

[55] Dice, L. R. Measures of the amount of ecologic association between species. *Ecology* 26, 297–302 (1945).

[56] Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nature methods* 9, 676–682 (2012).

[57] Grinberg, M. Flask web development: developing web applications with python (2018).

[58] Horn, B. K. Schunck, B. G. Determining optical flow. *Artificial intelligence* 17, 185–203 (1981).