

# Resolution enhancement with a task-assisted GAN to guide optical nanoscopy image analysis and acquisition

Received: 10 August 2022

Accepted: 12 June 2023

Published online: 27 July 2023

 Check for updates

Catherine Bouchard<sup>1,2</sup>, Theresa Wiesner<sup>1,2</sup>, Andréanne Deschênes<sup>1,2</sup>, Anthony Bilodeau<sup>1,2</sup>, Benoît Turcotte<sup>1,2</sup>, Christian Gagné<sup>1,3</sup> ✉ & Flavie Lavoie-Cardinal<sup>1,2,4</sup> ✉

Super-resolution fluorescence microscopy methods enable the characterization of nanostructures in living and fixed biological tissues. However, they require the adjustment of multiple imaging parameters while attempting to satisfy conflicting objectives, such as maximizing spatial and temporal resolution while minimizing light exposure. To overcome the limitations imposed by these trade-offs, post-acquisition algorithmic approaches have been proposed for resolution enhancement and image-quality improvement. Here we introduce the task-assisted generative adversarial network (TA-GAN), which incorporates an auxiliary task (for example, segmentation, localization) closely related to the observed biological nanostructure characterization. We evaluate how the TA-GAN improves generative accuracy over unassisted methods, using images acquired with different modalities such as confocal, bright-field, stimulated emission depletion and structured illumination microscopy. The TA-GAN is incorporated directly into the acquisition pipeline of the microscope to predict the nanometric content of the field of view without requiring the acquisition of a super-resolved image. This information is used to automatically select the imaging modality and regions of interest, optimizing the acquisition sequence by reducing light exposure. Data-driven microscopy methods like the TA-GAN will enable the observation of dynamic molecular processes with spatial and temporal resolutions that surpass the limits currently imposed by the trade-offs constraining super-resolution microscopy.

The development of super-resolution optical microscopy (optical nanoscopy) techniques to study the nanoscale organization of biological structures has transformed our understanding of cellular and molecular processes<sup>1</sup>. Such techniques, including stimulated emission depletion (STED) microscopy<sup>2</sup>, are compatible with

live-cell imaging, enabling the monitoring of subcellular dynamics with unprecedented spatio-temporal precision. In the design of optical nanoscopy experiments, multiple and often conflicting objectives (for example, spatial resolution, acquisition speed, light exposure and signal-to-noise ratio) must be considered<sup>3,4</sup>.

<sup>1</sup>Institute Intelligence and Data (IID), Université Laval, Quebec City, Quebec, Canada. <sup>2</sup>CERVO Brain Research Center, Quebec City, Quebec, Canada.

<sup>3</sup>Département de génie électrique et de génie informatique, Université Laval, Quebec City, Quebec, Canada. <sup>4</sup>Département de psychiatrie et de neurosciences, Université Laval, Quebec City, Quebec, Canada. ✉ e-mail: [christian.gagne@gel.ulaval.ca](mailto:christian.gagne@gel.ulaval.ca); [flavie.lavoie-cardinal@cervo.ulaval.ca](mailto:flavie.lavoie-cardinal@cervo.ulaval.ca)

Machine learning-assisted microscopy approaches have been proposed to improve the acquisition processes, mostly by limiting light exposure<sup>3,5,6</sup>. In parallel, several supervised<sup>7–10</sup> and weakly supervised<sup>11–13</sup> deep learning approaches have been developed for high-throughput analysis of microscopy images. Deep learning-based super-resolution<sup>5,14–18</sup> and domain adaptation<sup>19</sup> approaches have also been proposed recently for optical microscopy, but concerns and scepticism arise regarding their applicability to characterize biological structures at the nanoscale<sup>20–22</sup>.

Optical nanoscopy techniques exploit the ability to modulate the emission properties of fluorescent molecules to overcome the diffraction limit of light microscopy<sup>23</sup>. In this context, it is challenging to rely on algorithmic methods to generate images of subdiffraction structures that are not optically resolved in the original image<sup>20</sup>. Methods that are optimized for generating images that appear to belong to the target higher-resolution domain do not specifically guarantee that the biological features of interest are accurately generated<sup>22</sup>. Yet, the possibility to super-resolve microscopy images post-acquisition would favourably alleviate some of the compromises between the acquisition parameters in optical nanoscopy<sup>16,24</sup>.

Among the methods developed for algorithmic super-resolution, conditional generative adversarial networks (cGAN)<sup>25</sup> generate data instances based on a different input value, capturing some of its features to guide the creation of a new instance that fits the target domain. However, the realism of the synthetic images does not ensure that the images are usable for further field-specific analysis, which is limiting their use in optical microscopy. 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 auxiliary tasks, or multi-task learning, can guide the generator to resolve content that matters for the current context<sup>26</sup>. Various applications of cGANs for image-to-image translation use auxiliary tasks such as semantic segmentation<sup>27,28</sup>, attributes segmentation<sup>29</sup> or foreground segmentation<sup>30</sup> to provide spatial guidance to the generator. 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.

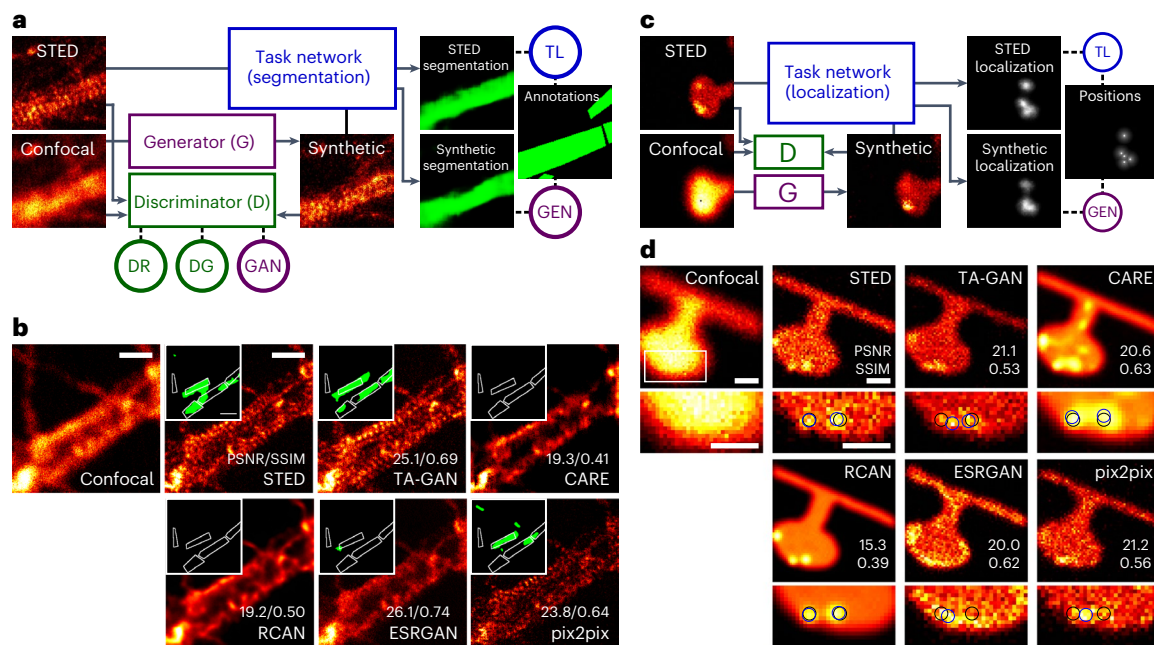
We propose to guide the image-generation process using an auxiliary task that is closely related to the biological question at hand. This approach improves the applicability of algorithmic super-resolution and ensures that the generated features in synthetic images are consistent with the observed biological structures in real nanoscopy images. Microscopy image analysis tasks that are already routinely solved with deep learning<sup>9</sup> (for example, segmentation, detection and classification) can guide a cGAN to preserve the biological features of interest in the generated synthetic images. Here we introduce a task-assisted GAN (TA-GAN) for resolution-enhanced microscopy image generation. The TA-GAN relies on an auxiliary task associated with structures that are unresolved by the input low-resolution modalities (for example, confocal or bright-field microscopy) but are easily distinguishable in the targeted super-resolution modalities (for example STED or structured illumination microscopy (SIM)). We expand the applicability of the method with a variation called TA-CycleGAN, based on the CycleGAN model<sup>31</sup>, applicable to unpaired datasets. Here the TA-CycleGAN is applied to domain adaptation for STED microscopy of fixed and living neurons. Our results show that the TA-GAN and TA-CycleGAN models improve the synthetic representation of biological nanostructures compared with other algorithmic super-resolution approaches. Specifically, our method is useful to (1) guide the quantitative analysis of nanostructures, (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.

## Results

**Task-assisted super-resolution image generation.** Deep learning methods designed for synthetic microscopy image generation have been shown to be effective for deblurring and denoising confocal images<sup>15,16,18</sup>. To increase the accuracy of resolution enhancement approaches applied to the generation of complex nanoassemblies, we consider the combination of a cGAN with an additional convolutional neural network, the task network (Fig. 1a), targeting an image analysis task relevant to the biological structures of interest. Three individual networks form the TA-GAN model: (1) the generator, (2) the discriminator and (3) the task network (Fig. 1a). The chosen auxiliary task should be achievable using the high-resolution modality only, ensuring that it is informative about content that is not resolved in the low-resolution input modality. The error between the task network predictions and the ground-truth annotations is backpropagated to the generator to optimize its parameters (Methods). The TA-GAN is trained using pairs of low-resolution (confocal or bright field) and super-resolution (STED or SIM) images.

The first TA-GAN model, TA-GAN<sub>ax</sub>, is trained on the axonal F-actin dataset<sup>13</sup> to generate STED images of the axonal F-actin lattice from confocal images (Fig. 1b). The auxiliary task identified to train the TA-GAN<sub>ax</sub> is the segmentation of the axonal F-actin rings, which cannot be resolved with confocal microscopy<sup>32</sup> (Fig. 1a). The segmentation network output is used to compute the generation loss and to evaluate the generation performance at the test time. The image super-resolution baselines content-aware image restoration (CARE)<sup>16</sup>, residual channel attention network (RCAN)<sup>15</sup>, enhanced super-resolution generative adversarial networks (ESRGAN)<sup>33,34</sup> and pix2pix<sup>35</sup> are trained on the axonal F-actin dataset and applied to the generation of a synthetic resolution-enhanced image from an input confocal image (Fig. 1b, first row). We additionally evaluate the performance of the image denoising baselines denoising convolutional neural network (DnCNN)<sup>36,37</sup> and Noise2Noise<sup>37,38</sup> on the confocal-to-STED image translation task (Supplementary Fig. 1). Comparison between the results of the TA-GAN<sub>ax</sub> with the baselines reveals that the pixel-wise mean square error (MSE), structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) between generated and ground-truth STED images are either improved or similar using the TA-GAN<sub>ax</sub> (Extended Data Fig. 1 and Supplementary Figs. 2 and 3). To evaluate the accuracy of each baseline in the generation of the nanostructure of interest, we evaluate the ability of an independent deep learning model trained on real STED images only<sup>13</sup>, which we refer to as U-Net<sub>fixed-ax</sub>, to segment the F-actin rings in the synthetic images over a held-out subset of the dataset that was not used for training the TA-GAN<sub>ax</sub>. The TA-GAN<sub>ax</sub> model uses the segmentation loss to optimize the generator's weights, which forces the generated F-actin nanostructures to be realistic enough to be recognized by the task network during training, and by U-Net<sub>fixed-ax</sub> during testing. The U-Net<sub>fixed-ax</sub> is applied to the synthetic and real STED images, and the similarity between the resulting pairs of segmentation maps is computed using the Dice coefficient (DC) and intersection over union (IOU) metrics. The improvement in similarity is significant for TA-GAN<sub>ax</sub> compared with all baselines (Extended Data Fig. 1).

We created a dataset of nanodomains in simulated shapes of dendritic spines using the pySTED simulation platform<sup>39</sup> to characterize the conditions where the TA-GAN outperforms the baselines in a controlled environment. The task used to train the TA-GAN for synaptic nanodomain generation (TA-GAN<sub>nano</sub>) is the localization of the centres of the simulated nanodomains (Fig. 1c). We compare the generated images with the ground-truth datamaps for two analysis tasks: (1) the localization of two nanodomains that are spaced by less than 100 nm, which is too close to be resolved with a standard deconvolution approach (Richardson Lucy<sup>40</sup>), and (2) the counting of nanodomains (2 to 6) separated by variable distances. The localization of the nanodomains can be performed using the TA-GAN synthetic images with similar accuracy to the one obtained using the simulated STED images from the



**Fig. 1 | The TA-GAN method.** **a**, Architecture of the TA-GAN<sub>ax</sub>. The losses (circles) are backpropagated to the networks of the same colour: the generator (violet), the discriminator (green) and the task network (blue). DG, discriminator loss for generated images; GEN, generation loss; GAN, GAN loss; DR, discriminator loss for real images; TL, task loss. The TA-GAN<sub>ax</sub> is applied to the axonal F-actin dataset using the segmentation of F-actin rings as an auxiliary task to optimize the generator. **b**, Representative example chosen out of 52 test images for the comparison of the TA-GAN<sub>ax</sub> and algorithmic super-resolution baselines on the axonal F-actin dataset. The confocal image is the low-resolution input and the STED image is the aimed ground truth. Insets: segmentation of the axonal F-actin

rings (green) predicted by the U-Net<sub>fixed-ax</sub> with the bounding boxes (white line) corresponding to the manual expert annotations<sup>13</sup>. PSNR and SSIM metrics are written on the generated images. Scale bars, 1  $\mu$ m. **c**, The TA-GAN<sub>nano</sub> is trained on the simulated nanodomain dataset using the localization of nanodomains as the auxiliary task. **d**, Representative example chosen out of 75 test images for the comparison of the TA-GAN<sub>nano</sub> with the baselines for nanodomain localization. The black circles represent the position of the nanodomains on the ground-truth datamap and the blue circles represent the nanodomains identified by an expert on images from the test set (Methods). The intensity scale is normalized for each image by its respective minimum and maximum values. Scale bars, 250 nm.

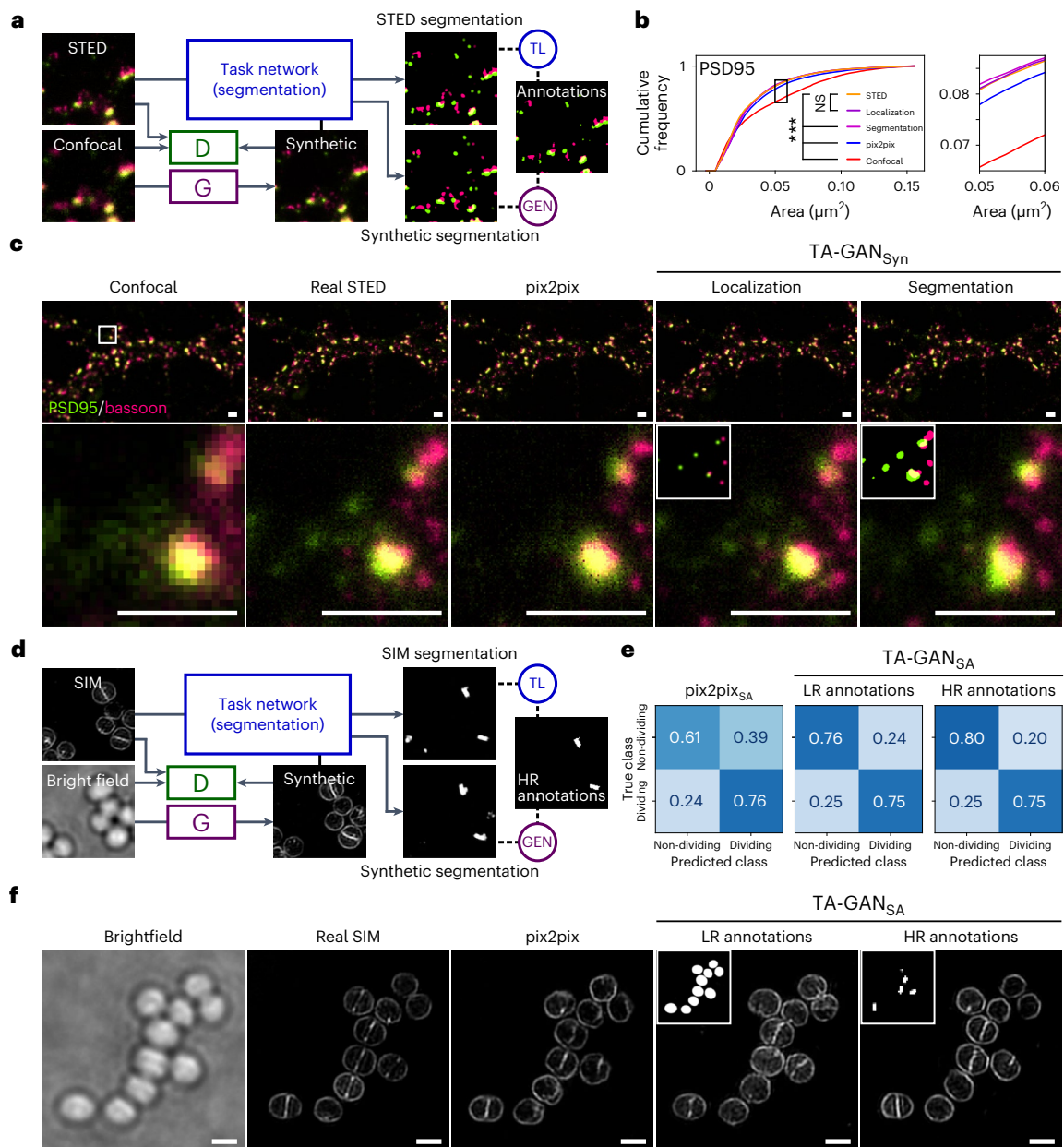
pySTED platform (Supplementary Fig. 4a). For the counting task, the images generated by the TA-GAN, RCAN and pix2pix allow to count up to six nanodomains that cannot be resolved in the simulated confocal images within a simulated spine (Supplementary Fig. 4b). Similarly to the results obtained on the axonal F-actin dataset, TA-GAN and pix2pix are the two algorithmic super-resolution approaches that generate synthetic images with the highest similarity to the target domain for the simulated nanodomain dataset, preserving image features such as the signal-to-noise ratio, background level and spatial resolution (Fig. 1d and Supplementary Fig. 5).

The TA-GAN model requires the definition of a task that steers the training of the generator towards the accurate extraction of sub-resolution information. The addition of this task is what differentiates TA-GAN from baselines such as pix2pix. We therefore evaluate how the choice of task impacts the performance using two different datasets. For the synaptic proteins dataset<sup>41</sup>, we evaluate the approach using a localization (Fig. 2a) and a segmentation task (Supplementary Fig. 6). The annotations are automatically generated using the pySODA analysis strategy<sup>41</sup>. For the localization task, we use the weighted centroids of the clusters, whereas for the segmentation task the masks are generated with wavelet segmentation<sup>42</sup>. We show that both tasks can be used to guide the synthetic image generation (Fig. 2b,c), but that the localization task allows to generate synaptic protein clusters with morphological features that are more similar to the one observed in the real images (Supplementary Figs. 7 and 8).

We evaluate how the precision of the labels used for the task impacts the generation accuracy using the publicly available dataset of *Staphylococcus aureus* cells from DeepBacs<sup>43,44</sup>. *S. aureus* bacteria are very small (around 1  $\mu$ m diameter), and monitoring their morphology changes and cell division processes requires subdiffraction

resolution<sup>45</sup>. The TA-GAN<sub>SA</sub> is trained for bright-field-to-SIM resolution enhancement using a classification task based either on: (1) low-resolution (LR) annotations generated from the bright-field modality (Supplementary Fig. 6) or (2) high-resolution (HR) annotations of dividing cell boundaries obtained from the SIM modality (Fig. 2d). We evaluate how the images generated with the algorithmic super-resolution approaches can be used for the classification of dividing and non-dividing bacterial cells, a task that is not achievable using only bright-field microscopy images (Supplementary Fig. 9). Training the TA-GAN<sub>SA</sub> model using the HR annotations leads to an improved classification performance combined with improved realism of the synthetic images (Fig. 2e,f).

**Domain adaptation on unpaired datasets.** For many microscopy modalities, paired and labelled training datasets are not directly available, or would require a high annotation burden from highly qualified experts. On the basis of 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 first validate that the TA-GAN can be applied to the dendritic F-actin dataset<sup>13</sup> using the semantic segmentation of F-actin rings and fibres in dendrites of fixed neurons (Fig. 3a). The trained TA-GAN<sub>Dend</sub> generates synthetic nanostructures that are successfully segmented by the U-Net<sub>fixed-dend</sub>, which recognizes dendritic F-actin rings and fibres in real STED images<sup>13</sup> (Fig. 3b). Similar to results previously obtained from real STED images<sup>13</sup>, the segmentation of the synthetic images with U-Net<sub>fixed-dend</sub> shows that the area of the F-actin rings significantly decreases as the neuronal activity increases, whereas the opposite is observed for F-actin fibres (Supplementary Fig. 10). Using our task-assisted strategy, we next trained a CycleGAN<sup>35</sup> model, as it was



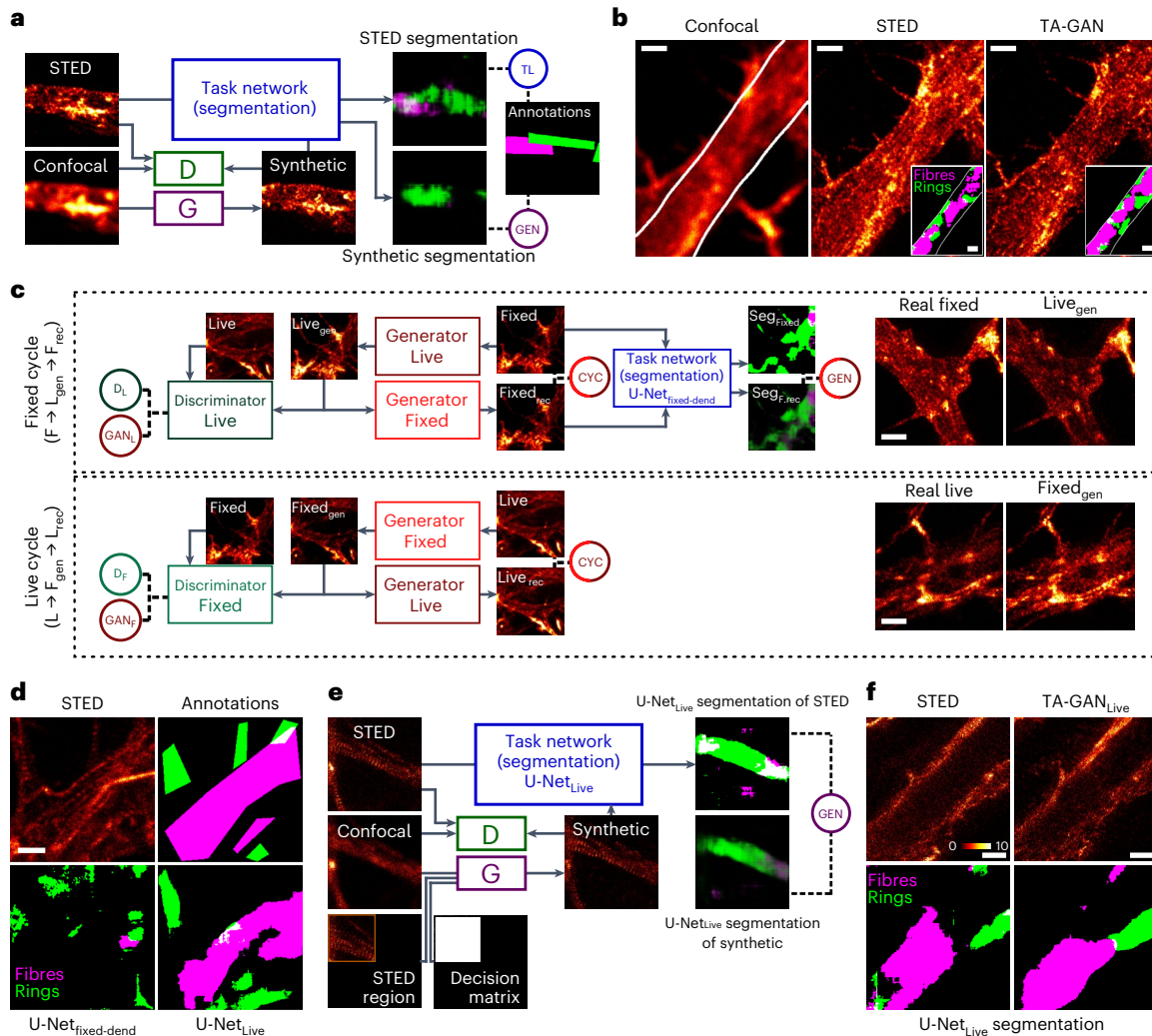
**Fig. 2 | Dataset-specific tasks drive reliable resolution enhancement with the TA-GAN approach.** **a**, Two TA-GAN models designed for the synaptic protein dataset are trained using one of two auxiliary tasks: the segmentation of the protein clusters (shown) or the localization of the weighted centroids (Supplementary Fig. 6). **b**, Comparison between the different approaches for the characterization of synaptic cluster morphological features. Shown is the cumulative distribution of the cluster area for PSD95 (see Supplementary Fig. 7 for other features). Statistical analysis: two-sided two-sample Kolmogorov–Smirnov test for the null hypothesis that the continuous distribution underlying the results for each baseline is the same as the one underlying the STED results (\*\* $P < 0.001$ , not significant (NS)  $P > 0.05$ ). **c**, Representative crop chosen from one of the nine test images for the generation of synthetic two-colour images of PSD95 and bassoon using the non-task-assisted baseline (pix2pix), the TA-GAN<sub>Syn</sub> with the localization task and the TA-GAN<sub>Syn</sub> with the segmentation task. Insets:

localization and segmentation annotations used to train the two TA-GAN<sub>Syn</sub> models. Scale bars, 1  $\mu\text{m}$ . Each crop is normalized to the 98th percentile of its pixel values for better visualization of dim clusters. **d**, The TA-GAN<sub>SA</sub> models designed for the *S. aureus* dataset are trained using a segmentation task with annotations requiring only the LR bright-field image or annotations requiring the HR SIM image. **e**, Confusion matrices for the classification of dividing and non-dividing cells on the test set of the *S. aureus* dataset ( $n = 410$  cells in five images). The TA-GAN<sub>SA</sub> trained with HR annotations achieves better performance in generating the boundaries between dividing bacterial cell, a morphological feature visible only with SIM microscopy, compared with pix2pix and the TA-GAN<sub>SA</sub> trained with LR annotations. **f**, Representative crop chosen from one of the five test images of the *S. aureus* dataset generated with pix2pix and the TA-GAN<sub>SA</sub> trained with LR and HR annotations. Insets: LR and HR annotations used to train the two TA-GAN<sub>SA</sub> models. Scale bars, 1  $\mu\text{m}$ .

precisely developed for image domain translation on unpaired datasets. The TA-CycleGAN can be applied to the translation between two microscopy modalities or experimental conditions in which the same biological structure can be observed (here the F-actin cytoskeleton in cultured live and fixed neurons) without the need for paired images.

To this aim we generated the live F-actin dataset consisting of confocal and STED images of F-actin nanostructures in living neurons using the far-red fluorogenic dye SiR-actin<sup>46</sup>.

The TA-CycleGAN includes two generators that are trained to first perform a complete cycle between the two domains (fixed- and



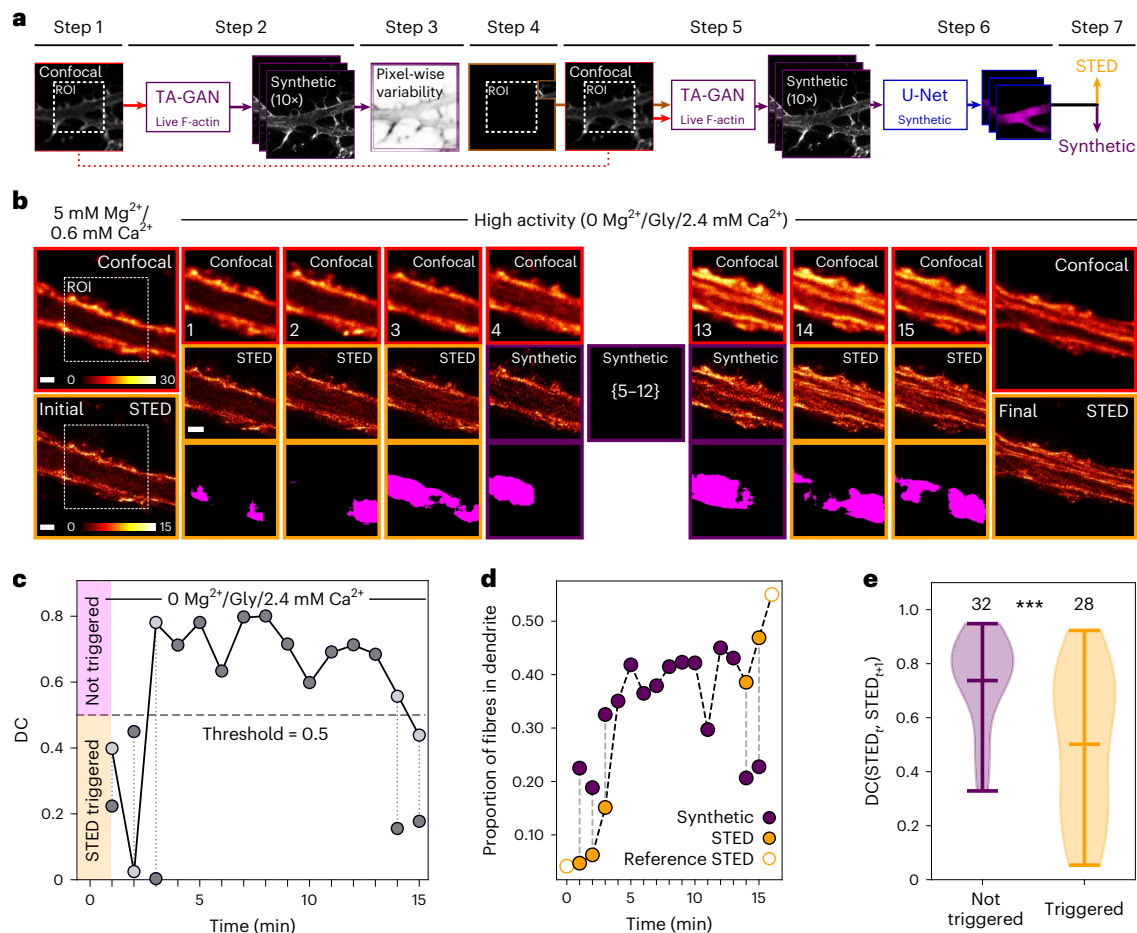
**Fig. 3 | Domain adaptation.** **a**, The semantic segmentation of F-actin rings (green) and fibres (magenta) is used as the auxiliary task to train the TA-GAN<sub>Dend</sub>. **b**, Example of confocal, real STED and TA-GAN<sub>Dend</sub> synthetic images chosen among 26 test images. Insets: the regions identified as rings and fibres by the U-Net<sub>fixed-dend</sub> trained on real STED images<sup>13</sup>. White solid line shows the border of the dendritic mask generated from the MAP2 channel, following the methods presented in ref. 13. **c**, The same semantic segmentation task is used to train the TA-CycleGAN. The reference to compute the TL is the segmentation of real fixed-cell STED images by U-Net<sub>fixed-dend</sub>. The fixed cycle (top) uses U-Net<sub>fixed-dend</sub> to encourage semantic consistency between the input fixed-cell image and the end-of-cycle reconstructed image. The live cycle (bottom) does not use a task network, enabling the use of non-annotated images from the live F-actin dataset. Once trained, the TA-CycleGAN can generate domain-adapted datasets (right). D<sub>L</sub>, discriminator loss for live-cell images; D<sub>F</sub>, discriminator loss for fixed cell images; GAN<sub>L</sub>, GAN loss for live-cell images; GAN<sub>F</sub>, GAN loss for fixed cell images; CYC, cycle loss; GEN, generation loss; L<sub>rec</sub>, live reconstructed; L<sub>gen</sub>, live

generated; F<sub>rec</sub>, fixed reconstructed; F<sub>gen</sub>, fixed generated; Live<sub>gen</sub>, generated live-cell image; Fixed<sub>gen</sub>, generated fixed cell image. **d**, Representative example chosen among 28 annotated live-cell STED test images for the segmentation of F-actin nanostructures. The nanostructures on the live-cell STED images (top left) are not properly segmented by the U-Net<sub>fixed-dend</sub> (bottom left). The U-Net<sub>Live</sub> is trained with synthetic images generated by the TA-CycleGAN to segment the F-actin nanostructures on real live-cell STED images. The segmentation predictions generated by the U-Net<sub>Live</sub> (bottom right) are similar to the manual expert annotations (top right). **e**, The semantic segmentation task is used to train the TA-GAN<sub>Live</sub>. The generator of the TA-GAN<sub>Live</sub> takes as input the confocal image as well as a STED subregion and a decision matrix indicating the position of the STED subregion in the FOV (Methods). **f**, Representative example of real and synthetic live-cell STED images of F-actin generated with TA-GAN<sub>Live</sub>, chosen among the initial images from 159 imaging sequences. The annotations of both real and synthetic images are obtained with the U-Net<sub>Live</sub>. Colour bar: raw photon counts. Scale bars, 1 μm.

live-cell STED imaging), and then to compare the ground-truth input image with the generated end-of-cycle image (Fig. 3c). In the generic CycleGAN model, the losses are minimized when the generated images appear to belong to the target domain and the MSE between the input and output is minimized. In TA-CycleGAN we add a task network, here the U-Net<sub>fixed-dend</sub>, which performs the semantic segmentation of dendritic rings and fibres. The U-Net<sub>fixed-dend</sub> is applied to the real fixed STED images and the end-of-cycle reconstructed fixed STED images (Fig. 3c). The generation loss is computed as the MSE between these segmentation masks. At inference, the trained TA-CycleGAN translates images of a given structure (here F-actin) but with image features (for example,

spatial resolution, signal-to-noise ratio, background level) corresponding to the target domain (here live-cell imaging). The translated F-actin dataset was generated by applying the TA-CycleGAN to the dendritic F-actin dataset (Supplementary Fig. 11).

The translated F-actin dataset, along with the expert annotations from the initial dendritic F-actin dataset, is used to train the U-Net<sub>Live</sub> segmentation network to segment F-actin structures in images from the live-cell domain without requiring annotation of the live F-actin dataset (Fig. 3d and Supplementary Fig. 11). To confirm that training on synthetic domain-adapted images generalizes to real live-cell STED images, the area under the receiver operating characteristic (AUROC)



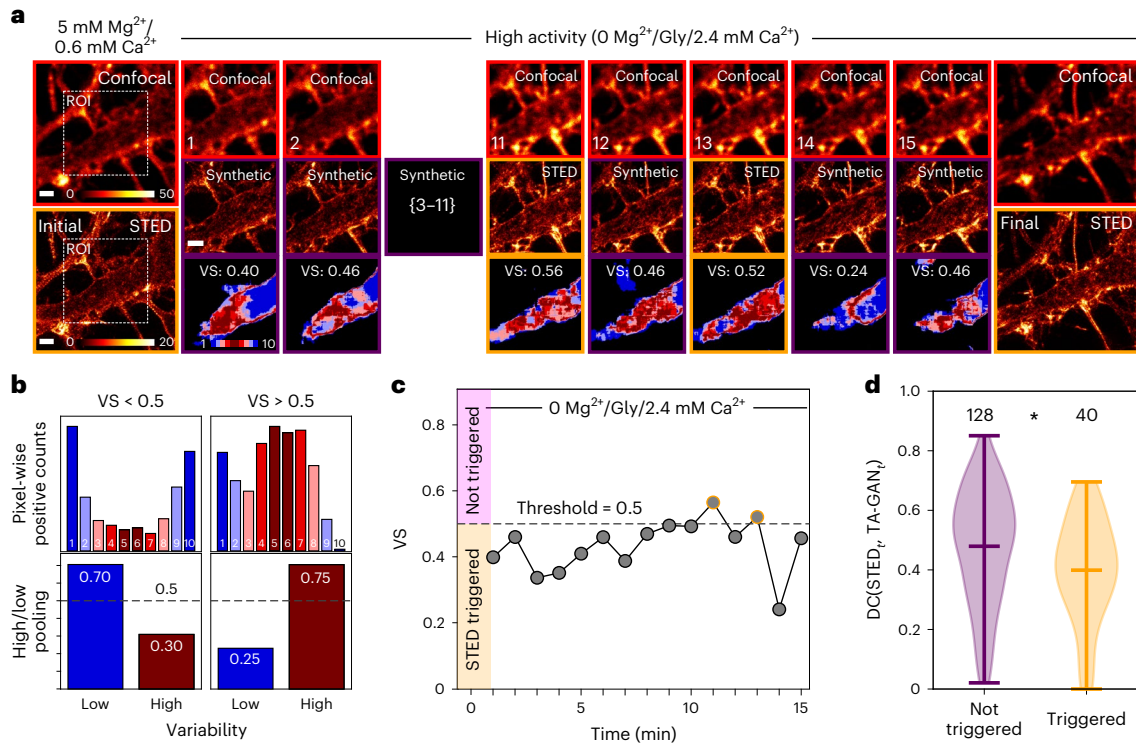
**Fig. 4 | Monitoring change with the TA-GAN<sub>Live</sub>.** **a**, Step-by-step imaging-assistance pipeline using the TA-GAN<sub>Live</sub> in the live-cell acquisition loop. **b**, Live-cell imaging of dendritic F-actin before (initial), during (frames 1–15) and after (final) application of a stimulation solution (0 Mg<sup>2+</sup>/Gly/2.4 mM Ca<sup>2+</sup>). Shown are the confocal (red, top row), synthetic (purple, middle row) and real (orange, middle row) STED images when acquired, and corresponding segmentation masks for F-actin fibres (magenta, bottom row). The series was chosen as a representative example from a total of 72 series. Colour bars: raw photon counts. **c**, The DC at each time point measured between the current synthetic image and the last acquired reference STED image for the sequence shown in **b**. Dark grey points indicate that the last acquired real STED (used as reference) is from a previous time step and light grey points connected with a vertical dashed line indicate that a new STED is acquired at this time step, and the DC is recomputed with this new reference. **d**, Proportion of dendritic F-actin fibres at each time

point segmented by the U-Net<sub>Live</sub> on either the real STED (orange) or the synthetic STED (purple) images. When a real STED acquisition is triggered, the proportion of fibres in both images is compared (dotted line). Initial and final reference STED images (empty orange circles) are acquired at each round. **e**, The DC is computed for the F-actin fibre segmentation on control sequences of two consecutive real STED images (time points  $t$  and  $t + 1$ ). The segmentation of the STED <sub>$t$</sub>  image is used as reference and the DC is computed with the segmentation mask on the STED <sub>$t+1$</sub>  image. When a real STED image acquisition would not have been triggered by the threshold-based approach, the DC between the segmentation masks of the two real STED is higher.  $n = 60$  control sequences of two consecutive confocal–STED pairs. Violin plots show the minimum, maximum and mean. Statistical analysis: two-sided Mann–Whitney  $U$  test<sup>62</sup> for the null hypothesis that the two distributions are the same (\*\*\* $P = 0.0004$ ). Scale bars, 1  $\mu$ m.

was computed between the U-Net<sub>Live</sub> segmentation masks and manual ground-truth annotations generated on 28 images by an expert in a user study (0.76 for rings and 0.83 for fibres; Extended Data Fig. 2 and Supplementary Figs. 12 and 13). In comparison, when applied to live-cell STED, the U-Net<sub>fixed</sub> trained only on real images of fixed neurons achieves an AUROC of only 0.60 and 0.59 for the segmentation of rings and fibres, respectively. Thus, domain adaptation with TA-CycleGAN enables the use of synthetic images to train a modality-specific segmentation network (here U-Net<sub>Live</sub>) when no real annotated dataset is available for training. This facilitates the cumbersome step in the training of any supervised machine learning method: creating data-specific annotations. We next train TA-GAN<sub>Live</sub> for resolution enhancement of live F-actin confocal images using the live F-actin dataset and the pre-trained U-Net<sub>Live</sub> as the auxiliary task network (Fig. 3e and Methods). The annotations generated by the U-Net<sub>Live</sub> are used to compute the generation loss. Thus our image translation approach allows to train

a TA-GAN to generate synthetic images from live-cell confocal images of F-actin in neurons (TA-GAN<sub>Live</sub>) as well as a segmentation network adapted to the live-cell imaging domain (U-Net<sub>Live</sub>), without the need to annotate the live F-actin dataset (Fig. 3f and Supplementary Fig. 14).

**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. In super-resolution microscopy, repeated imaging with high-intensity illumination can cause photobleaching, which quickly diminishes the signal quality (Supplementary Fig. 15 and Extended Data Fig. 3). We evaluate how the integration of the TA-GAN<sub>Live</sub> in the acquisition loop of an STED microscope can guide imaging sequences for time-lapse live-cell microscopy. We apply our approach to detect the activity-dependent remodelling of dendritic F-actin from periodical rings into fibres in living neurons, which was



**Fig. 5 | Monitoring prediction variability with the TA-GAN<sub>Live</sub>.** **a**, Live-cell imaging of dendritic F-actin using the same stimulation as in Fig. 4. The TA-GAN<sub>Live</sub> variability maps are shown on the bottom row. The series was chosen as a representative example from a total of 87 series. Colour bars: raw photon counts. **b**, Example histograms of the pixel-level positive counts over the segmentation of ten synthetic images (top) and high- and low-variability pooling. On the left, the VS is below 0.5 (dashed line, no trigger); on the right, the VS is above 0.5 (STED triggered). **c**, The VS at each time point for the sequence shown in **a**. When the VS is above 0.5, the number of high-variability pixels exceeds the number of

low-variability pixels (**b**, VS > 0.5, right), which triggers the acquisition of a real STED image (orange circles). **d**, The DC is computed between the segmentation masks of synthetic and real STED image from the same time point ( $n = 168$  pairs of real and synthetic images). When an STED acquisition would have been triggered using the VS criterion, the DC between the two corresponding images is lower. Violin plots show the minimum, maximum and mean. Statistical analysis: two-sided Mann–Whitney  $U$  test for the null hypothesis that the two distributions are the same ( $*P = 0.014$ ). Scale bars, 1  $\mu\text{m}$ .

previously observed in fixed neurons but could not be monitored in living neurons due to technical limitations<sup>13</sup>. For a given image acquisition sequence, we first acquire a confocal image (Fig. 4a, step 1). We next use a Monte Carlo dropout approach<sup>47</sup> to generate ten possible synthetic STED images with the TA-GAN<sub>Live</sub>. We apply a different random dropout mask for each image generated (Fig. 4a, step 2). This use of MC dropout with GANs has been previously demonstrated on natural images<sup>48,49</sup> and serves as an estimation of the variability of TA-GAN<sub>Live</sub> over the generated nanostructures. We next measure the optical flow between the ten synthetic images (Fig. 4a, step 3, and Methods). The subregion with the highest mean optical flow is acquired with the STED modality (Fig. 4a, step 4) and given as an input to the TA-GAN<sub>Live</sub> together with the corresponding confocal image of the full field of view (FOV; Fig. 4a, step 5). This step helps to minimize the effect of signal variations encountered in live-cell imaging. The TA-GAN<sub>Live</sub> generates, with different dropout masks, ten new synthetic images of the region of interest (ROI), which are segmented by the U-Net<sub>Live</sub> to detect the presence of F-actin fibres (Fig. 4a, step 6). The segmentation predictions of the U-Net<sub>Live</sub> for the synthetic images are used to decide whether or not a real STED image should be acquired at a given time point (Fig. 4a, step 7). The acquisition of a complete frame using the STED modality is triggered when either (1) the segmentation prediction on the synthetic STED image is different from the one obtained on the last acquired real STED image (Fig. 4b–e) or (2) there is high variability in the segmentation predictions on the ten synthetic STED images (Fig. 5 and Methods).

For the first acquisition scheme we calculate at each time point the mean DC between the segmentation masks from the ten generated

synthetic images and the last real STED image (Fig. 4b and Supplementary Figs. 16 and 17). A new STED image is acquired if the mean DC between the synthetic and the reference real STED images is below a predefined threshold of 0.5 (Fig. 4c). Using paired confocal and real STED images acquired at the end of the imaging sequence (15 min), we measure an increase in the proportion of F-actin fibres in living neurons (Fig. 4b, last frame, and Extended Data Fig. 4). On the basis of control sequences of two consecutive STED and confocal images pairs, we measure that the segmentation masks of those real STED images are more similar for sequences that would not have triggered a new real STED image acquisition, indicating that STED acquisitions are triggered at time points of higher biological change (Fig. 4e and Supplementary Fig. 18a,b). The value of the DC threshold is chosen based on preliminary imaging trials and previous knowledge about the remodelling extent and dynamics, which, depending on the experimental context, is not always available before the imaging experiment. With this acquisition scheme an average of 1.6 STED images are acquired per sequence (15 confocal images per sequence, 72 sequences). It reduces the light dose in average by 89% in the central ROI compared with acquiring 15 consecutive STED images.

We developed a second method to trigger STED image acquisitions, which is based on the variability in the predictions of the TA-GAN<sub>Live</sub>. This approach is particularly useful when not enough previous knowledge on the expected structural change is available to define a threshold for the DC before the experiment. With this acquisition scheme, for each confocal acquisition, we measure the pixel-wise variability of the segmentation predictions on the

ten generated synthetic STED images (Fig. 5a). Pixels predicted to belong to the same class (fibres or not fibres) in  $\geq 80\%$  of the synthetic images are defined as low-variability pixels, and pixels predicted to belong to the same class in  $< 80\%$  by the TA-GAN<sub>Live</sub> are defined as high-variability pixels (Fig. 5b). The proportion of high-variability pixels corresponds to the variability score (VS; Supplementary Fig. 19). When the VS is higher than 0.5 for the ROI, a full STED image is acquired (Fig. 5 and Supplementary Fig. 20). We validate the VS criterion on a set of real STED reference images and their corresponding synthetic counterparts. On these images, we measure a higher DC between the segmentation masks when a real STED image acquisition would not have been triggered by the VS threshold (Fig. 5d and Supplementary Fig. 18c,d). This indicates that the VS is a good indicator of the similarity between the real and synthetic STED image at a given time point. This approach can be beneficial to detect unexpected patterns and rare events. An average of 3.8 STED images were acquired for each sequence (87 sequences) using the variability-based triggers, which reduces the light dose in average by 74% in the central ROI compared with acquiring an STED image of the ROI at every frame. For both approaches, modulation of the STED modality acquisition frequency can be achieved by adapting the DC or VS thresholds. The resulting frame rate with TA-GAN assistance is comparable to acquiring sequences of paired confocal and STED images (Extended Data Table 1).

## Discussion

We introduce TA-GAN for resolution enhancement and domain adaptation. We demonstrate its applicability to optical nanoscopy (Extended Data Fig. 5) and show that an auxiliary task assisting the training of a generative network improves the reconstruction accuracy of nanoscopic structures. The applicability of our method is demonstrated for paired confocal and STED microscopy datasets of F-actin in axons and dendrites, synaptic protein clusters, simulated nanodomains as well as for paired bright-field and SIM images of dividing *S. aureus* bacterial cells. We show that the TA-GAN method is flexible and can be trained with different auxiliary tasks such as binary segmentation, semantic segmentation and localization. For unpaired datasets, we introduce the TA-CycleGAN model and demonstrate how the structure-preserving domain adaptation opens up the possibility to create paired datasets of annotated images that cannot be acquired simultaneously. The synthetic STED images from the live-cell domain can be used to train a neural network that performs well for the segmentation of F-actin nanostructures in real STED images, without the need for manual re-annotations of the new live-cell imaging dataset. The TA-GAN for resolution enhancement in living neurons can be integrated into the acquisition loop of an STED microscope (Figs. 4 and 5). We validate how this TA-GAN model can be helpful in assisting microscopists by automatically taking decisions that optimize the photon budget and reduce photobleaching (Extended Data Fig. 3) in live-cell optical nanoscopy acquisition sequences. The TA-GAN increases the informative value of each confocal acquisition and automatically triggers the acquisition of an STED image only in the regions and time steps where this acquisition is informative due to variations (Fig. 4) or uncertainties in the predicted nanostructures (Fig. 5).

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.

## Methods

### Sample preparation and STED microscopy

**Cell culture.** Dissociated Sprague Dawley rat hippocampal neurons were prepared as described previously<sup>13,50</sup> in accordance with and approved by the animal care committee of Université Laval. For live-cell STED imaging, the dissociated cells were plated on poly-D-lysine-laminin-coated glass coverslips (18 mm) at a density of 322 cells per mm<sup>2</sup> and used at 12–16 days in vitro.

**STED microscopy.** Live-cell super-resolution imaging was performed on a four-colour STED microscope (Abberior Instruments) using a 40 MHz pulsed 640 nm excitation laser, an ET685/70 (Chroma) fluorescence filter and a 775 nm pulsed (40 MHz) depletion laser. Scanning was conducted using a pixel dwell time of 5  $\mu$ s, a pixel size of 20 nm and an 8 line repetition sequence. The STED microscope was equipped with a motorized stage and auto-focus unit. The imaging parameters used are described in Supplementary Table 1.

The cultured neurons were pre-incubated in HEPES buffered artificial cerebrospinal fluid (aCSF) at 33 °C with SiR-actin (0.5  $\mu$ M, Spiro-Chrome) for 8 min and washed once gently in SiR-actin-free media. Imaging was performed in HEPES buffered aCSF of 5 mM Mg<sup>2+</sup>/0.6 mM Ca<sup>2+</sup> (NaCl 98 mM, KCl 5 mM, HEPES 10 mM, CaCl<sub>2</sub> 0.6 mM, glucose 10 mM, MgCl<sub>2</sub> 5 mM) using a gravity-driven perfusion system. Neuronal stimulation was performed with an HEPES buffered aCSF containing 2.4 mM Ca<sup>2+</sup>, glycine and without Mg<sup>2+</sup> (NaCl 98 mM, KCl 5 mM, HEPES 10 mM, glycine 0.2 mM, CaCl<sub>2</sub> 2.4 mM, glucose 10 mM). Solutions were adjusted to an osmolality of 240 mOsm per kg and a pH of 7.3.

### Datasets

**Axonal F-actin dataset.** The publicly available axonal F-actin dataset<sup>13</sup> was used to train the TA-GAN<sub>Ax</sub> for confocal-to-STED resolution enhancement of axonal F-actin nanostructures using the binary segmentation of F-actin rings as the auxiliary task. The original dataset consisted of 516 paired confocal and STED images (224  $\times$  224 pixels, 20 nm pixel size) of axonal F-actin in fixed cultured hippocampal neurons from ref. 13. Thirty-one images from the original dataset were discarded for not containing annotated axonal F-actin rings. The remaining images were randomly split into a training set (377 images), a validation set (56 images) and a testing set (52 images), which was not used for training. The manual polygonal bounding box annotations of the axonal F-actin periodical lattice (F-actin rings) from the original dataset were retained (Fig. 1b).

**Dendritic F-actin dataset.** The publicly available dendritic F-actin dataset was used to train the TA-GAN<sub>Dend</sub> for confocal-to-STED resolution enhancement of dendritic F-actin nanostructures using the semantic segmentation of F-actin rings and fibres as the auxiliary task. The dendritic F-actin dataset was also used to train the TA-CycleGAN for domain adaptation. The original dataset from ref. 13 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). We used the same testing split as the original publication to compare the segmentation results over the same images (Supplementary Fig. 10). The dataset consists of paired confocal and STED images of the dendritic F-actin cytoskeleton in fixed cultured hippocampal neurons, which had been manually annotated using polygonal bounding boxes. The training and validation crops were taken from large STED images (between 500  $\times$  500 pixels and 3,000  $\times$  3,000 pixels, 20 nm pixel size) using a sliding window of size 224  $\times$  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 fibres), the crop was discarded from the set. This operation resulted in 4,331 crops for training and 659 crops for validation.

**Simulated nanodomains dataset.** We used the pySTED image simulation platform<sup>39</sup> to create a simulated dataset of nanodomains within a dendritic spine. The pySTED simulator requires as input a matrix



providing the position and number of fluorescent molecules for each pixel in the FOV, referred to as a datamap. Each datamap ( $64 \times 64$  pixels or  $1.28 \times 1.28 \mu\text{m}$ ) consisted of a mushroom spine-like shape (between  $0.12 \mu\text{m}^2$  and  $0.48 \mu\text{m}^2$ ) containing  $N(1-6)$  regions ( $20 \times 20 \text{ nm}$ ) with a higher fluorophore concentration, which we refer to as nanodomains. In the majority of the images, the simulated STED modality was required to resolve all nanodomains. The position of the nanodomains was randomly distributed on the edge of the synapse ( $<140 \text{ nm}$  away from the edge) with a minimal distance of  $40 \text{ nm}$  between nanodomains. We allowed random rotation and translation of the spine making sure that the nanodomains were kept within the FOV. For training, we generated a total of 1,200 simulated datamaps (200 for each number of nanodomains). The training and validation datasets were split using a 90/10 ratio. The localization maps are matrices of size  $64 \times 64$  pixels, where the value of each pixel is the cubic root of the distance to the closest nanodomain. Two testing datasets were created. The first consisted of 75 simulated datamaps with different numbers of nanodomains (2–6, 15 images per number of nanodomains). The second consisted of 80 images with two nanodomains, where the distance between the pair of nanodomains varies from  $40 \text{ nm}$  to  $450 \text{ nm}$ .

**Synaptic protein dataset.** The publicly available synaptic protein dataset consists of paired two-colour STED and confocal images of the synaptic protein pair PSD95 (postsynapse) and bassoon (presynapse) in fixed hippocampal neurons obtained from ref. 41. The dataset was split into a training set (32 images), a validation set (2 images) and a testing set (9 images). The confocal and STED images from the training and validation sets were first registered using the pipeline presented in Supplementary Fig. 21, resulting in 690 crops for training and 35 crops for validation. The segmentation maps were generated by automatically segmenting the STED images using wavelet transform decomposition<sup>42</sup> with the same parameters (scales 3 and 4) as in ref. 41. No segmented clusters were discarded based on size or position, following the intuition that even the smallest structures should be generated. The localization maps were 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 standard deviation of 2 (Supplementary Fig. 22).

***S. aureus* dataset.** We used the bright-field images and the corresponding SIM images from the publicly available *S. aureus* dataset for segmentation from ref. 44. This dataset includes 12 images (6 for training, 1 for validation and 5 for testing) with manual whole-cell annotations. The bright-field images ( $80 \text{ nm}$  per pixel) were rescaled to the size of the SIM images ( $40 \text{ nm}$  per pixel) using bilinear interpolation, and the cell annotations were rescaled using nearest-neighbour interpolation. The whole-cell annotations were converted to binary segmentation maps with pixel values of 0 for background and 1 for cells. These whole-cell segmentation maps were used as LR annotations. HR annotations highlighting the cell division boundary were generated from the SIM images. To generate the HR annotations, we first applied a Sobel filter to the SIM images to find the outer and inner edges of the cells, followed by a Gaussian filter with a standard deviation of 1. We next applied a threshold corresponding to 20% of the maximum value of the filtered result. This resulted in a binary mask of the boundary between dividing cells as well as of the cell outer membrane. We similarly applied a Sobel filter to the LR annotations, followed by a Gaussian filter with a standard deviation of 1 and applied a threshold of 0 to generate a binarized cell border mask. The binarized cell border mask was subtracted from the mask of the outer and inner cell borders to generate the final HR annotations.

The training crops were generated using a sliding window of size  $256 \times 256$  pixels with an overlap of 128 pixels. Crops were discarded if they contained less than 3% annotated pixels. The validation crops were generated using the same sliding window method, but for a size of  $128 \times 128$  pixels without overlap. All validation crops were considered,

regardless of the percentage of annotated pixels. The resulting dataset comprised 202 training crops and 64 validation crops.

**Live F-actin dataset.** The live F-actin dataset was acquired for this study and was used to train: (1) the TA-CycleGAN for live and fixed domain adaptation and (2) the TA-GAN<sub>Live</sub>. The live F-actin dataset consists of 800 paired STED and confocal images of F-actin stained with the fluorogenic dye SiR-actin (Spirochrome) in living hippocampal cultured neurons (Supplementary Table 1). The dataset was split into a training set (753 images) and a validation set (47 images). The images were of variable size (from a minimum width of  $2.76 \mu\text{m}$  to a maximum of  $49.1 \mu\text{m}$ , pixel size is always  $20 \text{ nm}$ ).

**Translated F-actin dataset.** The translated F-actin dataset was used to train the TA-GAN<sub>Live</sub>. 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. It contains the same number of images, the same training, validation and testing splits, and the same image characteristics (crop size, pixel size, annotations) as the dendritic F-actin dataset.

### TA-GAN training procedure

The TA-GAN was developed from the cGAN model for image-to-image translation `pix2pix`<sup>35</sup>, available at <https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>. All the functions for training and testing TA-GAN use `pytorch`<sup>51</sup> (1.0.0), `torchvision` (0.2.1), `numpy` (1.19.2), `Pillow` (8.3.1), `tiffio` (2020.0.3), `scipy` (1.5.4) and `scikit-image` (0.17.2). Comparable methods using cGANs for enhancing the resolution of microscopy images are trained using pixel-wise generation losses to compare the generated image with the ground truth, such as MSE<sup>5</sup>, absolute error<sup>15,16</sup> or structural similarity index<sup>17,18</sup>. For the TA-GAN, the generation loss is computed by comparing the output of an auxiliary task network applied on the real (ground truth) and generated (synthetic) images (Fig. 1a). The other standard losses for conditional GANs<sup>35</sup> are also used for TA-GAN: the discrimination losses for the classification of the real and generated images, and the GAN loss for the misclassification of generated images. The networks (generator, discriminator and task network) are optimized using the Adam optimizer with momentum parameters  $\beta_1 = 0.5$  and  $\beta_2 = 0.999$  for all TA-GAN models. We follow the same approach as the `pix2pix` paper<sup>35</sup>: at each epoch we alternate between one gradient descent step on the discriminator, then one step on the generator, then one step on the task network. Supplementary Table 2 summarizes the settings for the resolution-enhancement experiments presented in this paper, and Supplementary Table 3 presents the hyperparameters used for training the TA-GAN for each of these experiments.

**TA-GAN training with segmentation auxiliary tasks.** The TA-GAN<sub>Ax</sub>, TA-GAN<sub>Dendr</sub>, TA-GAN<sub>Syn</sub> and TA-GAN<sub>SA</sub> were trained for resolution enhancement using the segmentation of subdiffraction biological structures as the auxiliary task. The output of the segmentation network was compared with the ground-truth annotations using an MSE loss. The loss computed from the real STED image (task loss, TL in Fig. 1a) 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 TA-GAN<sub>Ax</sub>, the auxiliary task was the segmentation of axonal F-actin rings. The output of the auxiliary task network was the predicted segmentation maps of F-actin rings. The spatial resolution of the real and synthetic images were not significantly different (Supplementary Table 4 and Supplementary Fig. 23).

For the TA-GAN<sub>Dend</sub>, the auxiliary task was the semantic segmentation of dendritic F-actin rings and fibres. The output of the auxiliary task network was a two-channel image, with the predicted segmentation maps of F-actin rings in the first channel and of F-actin fibres in the second channel. The spatial resolution of the real and synthetic images were not significantly different (Supplementary Table 4 and Supplementary Fig. 23).

For the TA-GAN<sub>SA</sub>, the auxiliary task was either whole-cell segmentation (LR annotations) or the segmentation of the boundary between dividing cells (HR annotations). The output of the auxiliary task network is a one-channel image with the predicted segmentation maps of either whole cells or the dividing cell boundaries, respectively.

For the TA-GAN<sub>Syn</sub> trained using a segmentation task, the output of the segmentation network is a two-channel image with the predicted segmentation maps of PSD95 clusters in the first channel and bassoon in the second channel. The spatial resolution of the real and synthetic images were not significantly different (Supplementary Table 4 and Supplementary Fig. 23).

**TA-GAN training with localization auxiliary tasks.** The TA-GAN<sub>Syn</sub> and TA-GAN<sub>Nano</sub> for confocal-to-STED resolution enhancement were trained using a localization network to compute the generation loss. The localization network took an STED image as input to output a map of dots indicating the intensity-weighted centroids of all detected clusters in the STED image.

The TA-GAN<sub>Syn</sub> was trained on the synaptic protein dataset using the two-channel confocal image rescaled and registered to the STED image. The generation loss (GEN in Fig. 1) was the MSE between the weighted centroids of the real STED image and the localization predictions from the task network on the synthetic image. The spatial resolution of the real and synthetic images were not significantly different (Supplementary Table 4 and Supplementary Fig. 23).

TA-GAN<sub>Nano</sub> was trained on the simulated nanodomain dataset using the simulated confocal image as input. The generation loss was the MSE between the localization maps from the ground-truth datamaps and the localization predictions from the task network on the synthetic image.

**TA-CycleGAN training for domain adaptation.** The TA-CycleGAN model was developed from the CycleGAN model<sup>35</sup>. As for the standard CycleGAN, the TA-CycleGAN consists of four networks: two generators (one that translates the domain of fixed-cell STED imaging (F) into the domain of live-cell STED imaging (L), and one that translates domain L into domain F), and two discriminators (one for domain F, the other for domain L), which are combined with a fifth network, the task network (Fig. 3c). The TA-CycleGAN was applied to non-paired images, where the prediction of the generator for a given input cannot be compared with 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 with the initial image (ground truth) for the computation of losses.

The TA-CycleGAN for fixed-to-live domain adaptation was trained using two datasets: the dendritic F-actin dataset (F) and the live F-actin dataset (L). The auxiliary task was the semantic segmentation or F-actin rings and fibres on the dendritic F-actin dataset, for which manual bounding box annotations were available<sup>13</sup>. The U-Net<sub>fixed-dend</sub> was already optimized for the semantic segmentation of F-actin rings and fibres in fixed-cell STED images<sup>13</sup>. The generation loss was the MSE between the U-Net<sub>fixed-dend</sub> segmentation prediction on the real fixed-cell image (fixed) and the end-of-cycle fixed-cell image (fixed<sub>rec</sub>) (Fig. 3c).

### Training procedures of resolution enhancement and denoising baselines

**Enhanced super-resolution generative adversarial network.** ESRGAN x4 (ref. 33) is a state-of-the-art method for upsampling natural images. ESRGAN was implemented from the public GitHub repository

(<https://github.com/xinntao/Real-ESRGAN>). We fine-tuned ESRGAN on two of our datasets, the axonal F-actin dataset and the simulated nanodomains dataset, using the code and pretrained weights released with the most recent iteration of the model, Real-ESRGAN<sup>34</sup>. For both datasets, the input of the model is the confocal image, and the target output is the corresponding STED image upsampled four times using nearest-neighbour interpolation. Even though the confocal and STED images are the same size, the upsampling had to be kept in the model to use the pretrained weights. ESRGAN was fine-tuned for 50,000 iterations. The model was applied to the validation images to ensure training had converged after 50,000 iterations. All default parameters proposed by the authors were used, except for the input crop size and batch size (128 pixels and 4 for the axonal F-actin dataset, 64 pixels and 16 for the simulated nanodomains dataset).

**Content-aware image restoration.** CARE<sup>16</sup> uses a U-Net for deblurring, denoising and enhancing fluorescence microscopy images. CARE 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 on our datasets. The original CARE model does not use data augmentation, as it is trained on unlimited simulated images. We augmented our datasets before training the CARE models so that the number of training images is similar to the one used for the original model trained on simulated images (8,000 synthetic pairs of 128 × 128 pixels). For the axonal F-actin dataset, each image from the training set is augmented 32 times by cropping the four corners into 128 × 128 crops and applying the 8 possible flips and rotations to each corner crops. The 377 224 × 224 images were augmented into 12,064 different crops. For the simulated nanodomains dataset, the 64 × 64 images were too small to be further cropped, but were instead augmented 8 times using flips and rotations. The 1,080 training images were augmented into 8,640 different images. The patience parameter for the learning-rate decay function was adjusted from 10 to 20 epochs after noticing that the learning rate was reduced too abruptly to allow the training loss to properly converge. Except for the patience of the learning-rate decay function, 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 used for testing.

**Residual channel attention network.** RCAN<sup>52</sup> uses residual channel attention networks to increase the resolution of natural images. 3D-RCAN<sup>15</sup> adapts the original model to denoise and sharpen fluorescence microscopy image volumes. We used the code implemented with TensorFlow and Keras from the publicly available GitHub repository (<https://github.com/AiviaCommunity/3D-RCAN>). We used the same patch size as for training the TA-GAN<sub>ax</sub> (128 × 128 pixels) and the TA-GAN<sub>Nano</sub> (64 × 64 pixels). We trained different RCAN models using configurations of hyperparameters that were inspired by both the two-dimensional (2D)<sup>52</sup> and the three-dimensional (3D)<sup>15</sup> versions. We first trained a model on the axonal F-actin dataset with the hyperparameters from the 2D RCAN version. Even though both training and validation losses had converged, the output obtained with the weights from the epoch of lowest validation loss (epoch 205 out of 1,000) is an unrecognizable and smoothed version of the input. We hypothesize that this version of the model is too deep (15 million trainable parameters) for the number of training images. We trained a second version of RCAN using the hyperparameters from ref. 15. The loss when training this model quickly converges to a minimum (epoch 34 out of 300) and the resulting images are smoothed versions of the input confocal image. This simplified version of RCAN might be too lightened for the 2D context. The architecture that ended up performing the best with our datasets mixes hyperparameters from both implementations. (1) We used 2D convolutions because our images are 2D, as in RCAN. (2) We set the number of residual groups to 10 in the residual in residual

structure, as in 3D-RCAN. (3) The residual channel attention blocks were set to 20, as in RCAN. (4) We set the number of convolution layers in the shallow feature extraction and residual in residual structure to 32, as in 3D-RCAN. (5) We set the reduction ratio to eight as in 3D-RCAN. (6) The upscaling module was removed because the confocal and STED images are the same size, as is the case for 3D-RCAN. This RCAN model was trained for 1,000 epochs for both datasets to ensure convergence of the validation loss. The model reaching the lowest validation loss (epoch 838 for the simulated nanodomains dataset, epoch 398 for the axonal F-actin dataset) was used for testing.

**cGAN for image-to-image translation.** pix2pix<sup>35</sup> is a state-of-the-art method for image-to-image translation in natural images. It was implemented with Pytorch from the publicly available GitHub repository (<https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>). The TA-GAN and pix2pix share the same architecture with or without the task assistance. For each experiment, the same hyperparameters and datasets as for the TA-GAN were used for training (Supplementary Table 3), replacing only the generation loss with a pixel-wise MSE loss between the ground-truth and generated STED images. The results from this baseline are compared to the TA-GAN for all fixed-cell datasets.

**Denoising convolutional neural network.** The denoising convolutional neural network (DnCNN)<sup>36</sup> is a state-of-the-art denoising method for natural images. The trained version of DnCNN<sup>36</sup> available at <https://github.com/yinhaoz/denoising-fluorescence> was directly applied to our test images for all datasets (Supplementary Fig. 1). The datasets used in this study do not provide the required characteristics to retrain DnCNN (that is, lack of images with different noise levels); therefore, a published version of the DnCNN trained on the fluorescence microscopy denoising dataset<sup>37</sup> was used as is. It was included as a baseline to show how the confocal-to-STED and bright-field-to-SIM transformations are not denoising tasks.

**Noise2Noise.** Noise2Noise<sup>38</sup> is a state-of-the-art deep learning denoising method that does not require clean (denoised) data for training. Like DnCNN, we used the training version available at <https://github.com/yinhaoz/denoising-fluorescence> and directly applied it to the test images from our datasets, without retraining or fine-tuning (Supplementary Fig. 1).

## Evaluation of networks performance

**Segmentation of F-actin nanostructures in synthetic STED images.** The performance of the TA-GAN<sub>Ax</sub> was measured on the images from the test set of the axonal F-actin dataset, which were held-out and not used for training the TA-GAN<sub>Ax</sub> or the baselines. The MSE, PSNR and SSIM were computed between the ground-truth and synthetic STED images of the test set (Extended Data Fig. 1). In addition, U-Net<sub>fixed-ax</sub>, a U-Net that was trained to segment axonal F-actin rings on real STED images only<sup>13</sup> (available at <https://github.com/FLClab/STEDActinFCN>), was used to produce segmentation masks of axonal F-actin rings on the real and synthetic STED image pairs (Extended Data Figs. 1 and 3), which were compared using the DC and IOU metrics. We used the trained weights provided and did not retrain U-Net<sub>fixed-ax</sub> specifically for this work.

The performance of the TA-GAN<sub>Dend</sub> was evaluated on the test set of the dendritic F-actin dataset, which was held-out and not used to train the TA-GAN<sub>Dend</sub> or the U-Net<sub>fixed-dend</sub>. The U-Net<sub>fixed-dend</sub>, a U-Net that was trained for the semantic segmentation of dendritic F-actin rings and fibres on real STED images only<sup>13</sup> (available at <https://github.com/FLClab/STEDActinFCN>), was used to segment the real and synthetic STED images. The segmentation masks of both F-actin rings and fibres were compared on the real and synthetic STED image pairs (Supplementary Fig. 10). We used the trained weights provided and did not retrain U-Net<sub>fixed-dend</sub> specifically for this work.

**Assessment of synaptic protein cluster morphology.** The perimeter, eccentricity, area, distance to nearest neighbour from the same channel and distance to nearest neighbour from the other channel of the protein clusters from the synaptic protein dataset were measured in the confocal, STED images and synthetic images (Fig. 2 and Supplementary Fig. 7). The distribution of each morphological feature over all associated clusters from the test set images was computed using a Python library for Statistical Object Distance Analysis (pySODA)<sup>41</sup> (Supplementary Fig. 7). A foreground mask was generated following ref. 41: 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 same parameters as in ref. 41, which were optimized for real STED images of synaptic protein clusters, were used for the analysis: wavelet segmentation scales of 3 and 4, a minimum cluster area of 5 pixels, and minimum cluster width and height of 3 pixels. The weighted centroids of the detected clusters were calculated on the raw STED images.

**Classification of *S. aureus* cells.** The TA-GAN<sub>SA</sub> performance was evaluated using the classification of dividing bacterial cells, which is a task that cannot be achieved using only the bright-field images. A simple threshold optimization applied on bright-field images was not sufficient to classify the cells as dividing or not (Supplementary Fig. 9). A dividing bacterial cell is defined as having a clear boundary between the two dividing cells that can be identified in the SIM image. We trained the ResNet<sub>SA</sub> using the SIM images (training set) and the HR annotations, to segment the dividing cell boundaries. The ResNet<sub>SA</sub> is a ResNet-9 architecture trained for 200 epochs using an MSE loss, a learning rate of 0.0002 and the Adam optimizer. All real SIM images and synthetic SIM images generated from pix2pix<sub>SA</sub>, TA-GAN<sub>SA</sub> trained with LR annotations and TA-GAN<sub>SA</sub> trained with HR annotations are segmented by ResNet<sub>SA</sub>.

The dividing/non-dividing cells classification was based on the segmentation of the ResNet<sub>SA</sub>: (1) dividing if the segmentation mask contained at least 20 positive pixels and (2) non-dividing if the segmentation mask was empty for a given cell. For segmentation masks containing 1–19 pixels, the cells were identified as ambiguous and discarded. On the real SIM images test set, 251 cells were identified as non-dividing (single cells) and 159 as dividing (showing a clear boundary between the dividing cells).

**User study for the segmentation of live F-actin images.** A set of 28 STED images (224 × 224 pixels) from the live F-actin dataset test set was labelled by an expert using a Fiji<sup>53</sup> macro to test the performance of the U-Net<sub>live</sub> trained on the domain-adapted dendritic F-actin dataset for the segmentation of real live-cell STED images. In addition, a second set of 28 synthetic images, selected from the domain-adapted dendritic F-actin dataset was 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. For each image, the expert draws polygonal bounding boxes that enclosed all regions identified as F-actin rings and fibres.

**User study for the localization of nanodomains.** The positions of the nanodomains in the real and synthetic test images of the simulated nanodomains dataset were identified by an expert to compare the localization performance of the TA-GAN<sub>Nano</sub> with the baseline methods. The expert was presented with an image without being informed whether the image was real or synthetic, or by which model it was generated. For each image, the expert selects the pixel identified as the centre of each nanodomain detected. To compute the F1 score, a detection is defined as a true positive if it is within 3 pixels of the ground-truth position of a nanodomain centre.

### TA-GAN-assisted live-cell STED microscopy

**Training of the TA-GAN<sub>Live</sub>.** The TA-GAN<sub>Live</sub> for resolution enhancement of live-cell STED imaging was trained on the new and not previously annotated live F-actin dataset. The auxiliary task was the semantic segmentation of dendritic F-actin rings and fibres. The original live F-actin dataset did not include any manual annotations. To circumvent this limitation, the U-Net<sub>Live</sub> segmentation network was pretrained on the domain-adapted dendritic F-actin dataset. The pretrained U-Net<sub>Live</sub> was frozen during the TA-GAN<sub>Live</sub> training and was used to compute the MSE generation loss between the segmentation prediction of the real and the synthetic STED images.

To better adapt to cell-to-cell signal variations and experimental variability in live-cell STED images, the input of the generator has three channels: (1) the confocal image, (2) a real STED subregion acquired in the vicinity of the ROI and (3) an image indicating the position of the STED subregion (Fig. 3e). Training using this three-channel input enables the generator to learn features from the STED subregion and turns the resolution-enhancement task into an image-completion task.

**Training of the U-Net<sub>Live</sub>.** The U-Net<sub>Live</sub> was built around a U-Net-128 (ref. 54) architecture with batch normalization and two output channels (F-actin rings and fibres) for the segmentation of F-actin nanostructures in living neurons.

The training of the U-Net<sub>Live</sub> required an annotated dataset of images of the live-cell domain. A random subset (2,069 training crops and 277 validation crops) of the dendritic F-actin dataset was translated into the live-cell domain using the generator<sub>Live</sub> (Supplementary Fig. 11). This resulted in the domain-adapted F-actin dataset. The manual annotation from the fixed-cell images were associated with the corresponding synthetic images from the live-cell domain (Supplementary Fig. 11a).

Random crops of 128 × 128 pixels of the domain-adapted F-actin dataset and their corresponding annotations were used to train U-Net<sub>Live</sub> 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 fibres was weighted by a factor of 2.5, which reflected the ratio of total annotated pixels for each class. The U-Net<sub>Live</sub> was trained for 1,000 epochs and the iteration with the lowest segmentation loss over the validation set was kept for further use and testing. 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).

**TA-GAN integration in the acquisition loop.** The TA-GAN<sub>Live</sub> trained for resolution enhancement for live-cell imaging was directly integrated in the imaging acquisition process of the STED microscope (Fig. 4a). At the beginning and at the end of each experiments, an FOV of 10 × 10 μm was selected and reference STED and confocal images were acquired. The reference images were used to monitor the dendritic F-actin activity-dependent remodelling in living neurons (Extended Data Fig. 4). Similarity measurements between the synthetic and real STED images do not show time-dependent changes in the generation accuracy over all imaging sequences (Supplementary Fig. 24).

For each time point, as a first step (Table 1), a confocal image of the ROI is acquired to serve as the input to the TA-GAN<sub>Live</sub> for the generation of ten synthetic resolution-enhanced images of the ROI (step 2, Table 1). The ten synthetic images of steps 2 and 5 are generated using different random dropout masks created with the default dropout rate of 0.5 from ref. 55 and confirmed to be appropriate when applied on GANs by ref. 49.

The third step is the selection of an STED subregion outside the ROI (step 3, Table 1), which is given as input, along with the confocal FOV, to the TA-GAN<sub>Live</sub> to account for signal variation in live-cell imaging. In the fourth step, the STED subregion is acquired on the microscope. Finally, this subregion (step 5, Table 1) is given as input to the TA-GAN together with the confocal image as described in the previous section. The STED

**Table 1 | Steps performed at each time point for automated TA-GAN assistance**

1	A confocal image of the FOV (10×10 μm) is acquired.
2	Using dropout, ten synthetic images of the FOV are generated.
3	A subregion (2×2 μm) of highest variability outside the ROI is identified with optical flow.
4	A real STED image of this subregion is acquired.
5	The confocal images of the FOV and of the STED subregion are used by the TA-GAN to produce ten new synthetic STED images of the FOV.
6	The fibres in the ROI (central region of 6×6 μm) are segmented by U-Net <sub>Live</sub> .
7	The segmentation predictions are used to decide if an STED image should be acquired based on either (1) the mean DC with the segmentation of the last acquired STED or (2) the variability between the ten segmentation predictions.

images generated by TA-GAN<sub>Live</sub> more closely match the ground-truth STED ROI when an STED subregion is given as input along with the confocal FOV (Supplementary Fig. 25).

In our imaging-assistance framework, we choose for step 3 (Table 1) to compute the pair-wise optical flow (OF) between the ten synthetic images generated with the TA-GAN<sub>Live</sub> using dropout. The OF is computed using a Python implementation of the Horn–Schunck method<sup>56</sup> with the Python multiprocessing library, parallelizing the computations on eight central processing units to increase the computation speed and avoid delays. The OF is computed between each pair of the ten synthetic images (1–2, 2–3, ...). To translate the pixel-wise OF to a region-wise maps, the 500 × 500 pixels OF image was downsampled to a 5 × 5 map using the mean of each 100 × 100 pixels region. The subregion with the highest mean displacement is imaged with the STED modality. We decided to use OF as a measure of disparity between the synthetic generations, but other measures (for example, standard deviation, SSIM, mean intensity) could be used for experiments where computation time needs to be minimized (Supplementary Table 5). The sequence of acquiring the full confocal (2.6 s), generating ten synthetic STED images (2.5 s), computing the OF (6.1 s), acquiring the STED subregion (1.3 s), generating ten synthetic STED images again (2.5 s) and taking the decision requires a total of around 15.0 s per 500 × 500 pixels regions (10 × 10 μm). In comparison, acquiring an STED image requires 13.6 s using the same parameters (pixel size, pixel dwell time and size of the FOV).

Steps 2, 3, 5 and 6 are computed with a graphics processing unit to avoid computation induced delays. To do so, the commands from steps 2, 3, 5 and 6 are sent from the microscope's control computer to a graphics-processing-unit-equipped computer using the Flask<sup>57</sup> web framework Python module, version 2.0.3. All automated acquisitions use the SpecPy Python library version 1.2.1 to interface with the Imspector software (Abberior Instruments).

### Live-cell imaging decision guidance using the TA-GAN

The TA-GAN<sub>Live</sub> predictions are used for decision guidance on the optimal STED and confocal acquisition sequence and applied to the imaging of F-actin remodelling dynamics in cultured hippocampal neurons. For a region size of 6 × 6 μm (300 × 300 pixels), as used for the live-cell experiments, a confocal acquisition applies to the sample a photon dose of 1.168 × 10<sup>13</sup> photons per second, compared with 1.543 × 10<sup>18</sup> photons per second with STED. The TA-GAN assistance aims at reducing the light dose by limiting STED acquisitions to only the time points where a structural change is predicted.

### TA-GAN assisted monitoring of expected structural change.

The proof-of-concept experiment targets the expected activity-dependent remodelling of dendritic F-actin rings into fibres<sup>13</sup>. On the

basis of previous findings, the area of F-actin fibres was expected to increase following a neuronal stimulation<sup>13</sup>. The structural remodeling is monitored by comparing the area of segmented F-actin fibres on the synthetic and the reference real STED images. F-actin fibres are segmented on the synthetic STED images by U-Net<sub>Live</sub>. At each time point, steps 1–5 are performed as described in Table 1. To decide, following step 5, whether or not an STED image of the full ROI should be acquired, ten synthetic images of the ROI (acquired with the confocal modality) are generated and segmented by the U-Net<sub>Live</sub>. The mean of the ten segmentation maps is compared with the segmentation map predicted for the last acquired real STED image (reference STED) using the DC metric. A low DC is indicative of changes in the F-actin nanostructures in respect to the reference STED. A full real STED image is acquired if the DC falls below a pre-established threshold of 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 an STED on the full ROI is triggered, the STED reference image is updated for subsequent comparison of the segmentation maps.

Monitoring the TA-GAN<sub>Live</sub> generator's variability. The pixel-wise generator's variability can also be monitored to trigger the imaging of a full ROI with the STED modality. At each time point, steps 1–5 are performed as described in Table 1. The ten synthetic images generated at step 5 are segmented by the U-Net<sub>Live</sub>, resulting in ten segmentation maps for F-actin rings and fibres. The ten segmentation maps of F-actin fibres are binarized and summed. Pixels in the summed segmentation prediction has a value between zero and ten (zero when the presence of fibres was predicted in none of the synthetic images and ten when it is predicted in all). The variability of the generator on the segmentation prediction is evaluated from the summed segmentation prediction. Low-variability pixels are the pixels having the same value for at least 80% of the predicted segmentation maps (values of 1–2 (no fibres), 9–10 (fibres), positive counts). High-variability pixels are those having positive counts in between three and eight, inclusive. The distribution of high- and low-variability pixels from the foreground (Fig. 4b) is compared for each image. Pixels with zero positive counts (mostly background) are not considered. The proportion of low-variability pixels in the foreground is defined as the variability score (VS). A VS below 0.5 corresponds to images for which the predictions of the U-Net<sub>Live</sub> on the ten synthetic images are consistent for the majority of foreground pixels. If the VS is above 0.5, the ten synthetic STED images are not consistent and an STED acquisition is triggered. The threshold of 0.5 was chosen because it corresponds to the tipping point where the number of high-variability pixels exceeds the number of low-variability pixels.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The *S. aureus* dataset from refs. 43,44 is available at <https://zenodo.org/record/5550933#.Y6lhFNLMJH4> (ref. 43) and <https://zenodo.org/record/5551141#.Y6ljBdLMJH5> (ref. 58). The live F-actin dataset introduced here is available to download at <https://zenodo.org/record/7908914> (ref. 59) and <https://s3.valeria.science/flclab-tagan/index.html>. Other datasets can be requested from their respective publications: the axonal F-actin dataset<sup>13</sup>, the dendritic F-actin dataset<sup>13</sup>, the synaptic protein dataset<sup>41</sup> and the *S. aureus* dataset<sup>43,44</sup>. The processed versions of those datasets, as used to train the TA-GAN models, can be downloaded from <https://s3.valeria.science/flclab-tagan/index.html>. Sample test images are available at <https://github.com/FLClab/TA-GAN> in the 'test' subfolders of each dataset. Source data are provided with this paper.

### Code availability

The codes, trained weights and instructions on how to train, test and adapt for new experiments the TA-GAN model are available at <https://github.com/FLClab/TA-GAN> and <https://doi.org/10.5281/zenodo.7908818> (ref. 60). The trained U-Net<sub>Live</sub> model is available to download at <https://doi.org/10.5281/zenodo.7909304> (ref. 61). The code, seeds and parameters used to generate the simulated nanodomains dataset are available at <https://github.com/FLClab/TA-GAN> and <https://doi.org/10.5281/zenodo.7908818> (ref. 60).

### References

- Sahl, S., Hell, S. & Jakobs, S. Fluorescence nanoscopy in cell biology. *Nat. Rev. Mol. Cell Biol.* **18**, 685–701 (2017).
- Hell, S. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Lett.* **19**, 780–782 (1994).
- Durand, A. et al. A machine learning approach for online automated optimization of super-resolution optical microscopy. *Nat. Commun.* **9**, 5247 (2018).
- Laissue, P., Alghamdi, R., Tomancak, P., Reynaud, E. & Shroff, H. Assessing phototoxicity in live fluorescence imaging. *Nat. Methods* **14**, 657–661 (2017).
- Fang, L. et al. Deep learning-based point-scanning super-resolution imaging. *Nat. Methods* **18**, 406–416 (2021).
- Wu, Y. et al. Multiview confocal super-resolution microscopy. *Nature* **600**, 279–284 (2021).
- Chamier, L. et al. Democratizing deep learning for microscopy with ZeroCostDL4Mic. *Nat. Commun.* **12**, 2276 (2021).
- Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
- Falk, T. et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).
- Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
- Lu, M. et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**, 555–570 (2021).
- Bilodeau, A. et al. Microscopy analysis neural network to solve detection, enumeration and segmentation from image-level annotations. *Nat. Mach. Intell.* **4**, 455–466 (2022).
- Lavoie-Cardinal, F. et al. Neuronal activity remodels the F-actin based submembrane lattice in dendrites but not axons of hippocampal neurons. *Sci. Rep.* **10**, 11960 (2020).
- Nehme, E., Weiss, L., Michaeli, T. & Shechtman, Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* **5**, 458–464 (2018).
- Chen, J. et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nat. Methods* **18**, 678–687 (2021).
- Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090–1097 (2018).
- Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat. Methods* **18**, 194–202 (2021).
- Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).
- Li, X. et al. Unsupervised content-preserving transformation for optical microscopy. *Light Sci. Appl.* **10**, 44 (2021).
- Belthangady, C. & Royer, L. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods* **16**, 1215–1225 (2019).

21. Hoffman, D. P., Slavitt, I. & Fitzpatrick, C. A. The promise and peril of deep learning in microscopy. *Nat. Methods* **18**, 131–132 (2021).
22. Cohen, J. P., Luck, M. & Honari, S. Distribution matching losses can hallucinate features in medical image translation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference* 529–536 (Springer International Publishing, 2018).
23. Hell, S. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
24. Pawley, J. B. In *Handbook Of Biological Confocal Microscopy* (ed. Pawley, J.) 20–42 (Springer, 2006); [https://doi.org/10.1007/978-0-387-45524-2\\_2](https://doi.org/10.1007/978-0-387-45524-2_2)
25. Mirza, M. & Osindero, S. Conditional generative adversarial nets. Preprint at <http://arxiv.org/abs/1411.1784> (2014).
26. Ruder, S. An overview of multi-task learning in deep neural networks. Preprint at <http://arxiv.org/abs/1706.05098> (2017).
27. Zhang, C., Tang, Y., Zhao, C., Sun, Q., Ye, Z. & Kurths, J. Multitask GANs for semantic segmentation and depth completion with cycle consistency. *IEEE Trans. Neural Netw. Learn.* **32**, 5404–5415 (2021).
28. Ren, M., Dey, N., Fishbaugh, J. & Gerig, G. Segmentation-renormalized deep feature modulation for unpaired image harmonization. *IEEE Trans. Med. Imaging* **40**, 1519–1530 (2021).
29. 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 (IEEE, 2019).
30. Jaiswal, A. et al. Controlling BigGAN image generation with a segmentation network. In *International Conference On Discovery Science* (eds Soares, C. & Torgo, L.) 268–281 (Springer, 2021).
31. Zhu, J. Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. IEEE International Conference on Computer Vision* 2223–2232 (IEEE, 2017).
32. Xu, K., Zhong, G. & Zhuang, X. Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. *Science* **339**, 452–456 (2013).
33. Wang, X. et al. Esrgan: enhanced super-resolution generative adversarial networks. In *Proc. European Conference on Computer Vision (ECCV) Workshops* (Springer International Publishing, 2018).
34. Wang, X., Xie, L., Dong, C. & Shan, Y. Real-ESRGAN: training real-world blind super-resolution with pure synthetic data. In *Proc. IEEE/CVF International Conference on Computer Vision* 1905–1914 (IEEE, 2021).
35. 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 (IEEE, 2017).
36. Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **26**, 3142–3155 (2017).
37. Zhang, Y. et al. A Poisson–Gaussian denoising dataset with real fluorescence microscopy images. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 11710–11718 (IEEE, 2019).
38. Lehtinen, J. et al. Noise2Noise: learning image restoration without clean data. In *Proc. 35th International Conference on Machine Learning* (Eds Dy, J. & Krause, A.) Vol. 80, 2965–2974 (PMLR, 2018).
39. Turcotte, B., Bilodeau, A., Lavoie-Cardinal, F. & Durand, A. pySTED: a STED microscopy simulation tool for machine learning training. In *Annual AAAI Workshop on AI to Accelerate Science and Engineering* (AI2ASE, 2022).
40. Richardson, W. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.* **62**, 55–59 (1972).
41. Wiesner, T. et al. Activity-dependent remodeling of synaptic protein organization revealed by high throughput analysis of STED nanoscopy images. *Front. Neural Circuits* **14**, 57 (2020).
42. Olivo-Marin, J. Extraction of spots in biological images using multiscale products. *Pattern Recognit.* **35**, 1989–1996 (2002).
43. Pereira, P. & Pinho, M. DeepBacs—*Staphylococcus aureus* widefield segmentation dataset. Zenodo <https://zenodo.org/record/5550933> (2021).
44. Spahn, C. et al. DeepBacs for multi-task bacterial image analysis using open-source deep learning approaches. *Commun. Biol.* **5**, 688 (2022).
45. Saraiva, B. et al. Reassessment of the distinctive geometry of *Staphylococcus aureus* cell division. *Nat. Commun.* **11**, 4097 (2020).
46. Lukinavičius, G. et al. Others fluorogenic probes for live-cell imaging of the cytoskeleton. *Nat. Methods* **11**, 731–733 (2014).
47. Gal, Y. & Ghahramani, Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *International Conference on Machine Learning* (Eds Balcan, M. F. & Weinberger, K. Q.) Vol. 48, 1050–1059 (PMLR, 2016).
48. Palakkadavath, R. & Srijith, P. Bayesian generative adversarial nets with dropout inference. In *Proc. 3rd ACM India Joint International Conference on Data Science and Management of Data* 92–100 (ACM, 2021).
49. Wieluch, S. & Schwenker, F. Dropout induced noise for co-creative GAN systems. In *Proc. IEEE/CVF International Conference on Computer Vision Workshops* (IEEE, 2019).
50. Nault, F., De Koninck, P. & De Koninck, P. In *Protocols for Neural Cell Culture* 4th edn, 137–159 (Springer, 2010).
51. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32**, 8024–8035 (2019).
52. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. & Fu, Y. Image super-resolution using very deep residual channel attention networks. In *Proc. European Conference on Computer Vision* 286–301 (2018).
53. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
54. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science* (eds Navab, N. et al.) Vol. 9351, 234–241 (Springer, 2015).
55. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
56. Horn, B. & Schunck, B. Determining optical flow. *Artif. Intell.* **17**, 185–203 (1981).
57. Grinberg, M. *Flask Web Development: Developing Web Applications with Python* (O’Reilly Media, 2018).
58. Pereira, P. M. & Pinho, M. DeepBacs—*S. aureus* SIM prediction dataset and CARE model. Zenodo <https://doi.org/10.5281/zenodo.5551141> (2021).
59. Bouchard, C., Gagné, C. & Lavoie-Cardinal, F. Confocal and STED live F-actin dataset (version 1). Zenodo <https://doi.org/10.5281/zenodo.7908914> (2023).
60. Bouchard, C., Bilodeau, A., Deschênes, A. & Lavoie-Cardinal, F. FLClab/TA-GAN: TA-GAN (version v2023). Zenodo <https://doi.org/10.5281/zenodo.7908818> (2023).
61. Bouchard, C., Gagné, C. & Lavoie-Cardinal, F. U-Net live: aegmentation network for F-actin nanostructures in STED images of living neurons (version 1). Zenodo <https://doi.org/10.5281/zenodo.7909304> (2023).
62. Mann, H. & Whitney, D. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60 (1947).

## Acknowledgements

We thank F. Nault and S. Pensivy for the neuronal cell culture, G. Leclerc for the Fiji macro for segmentation and A. Schwerdtfeger for proofreading the paper. Funding was provided by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-06704-2019 to F.L.-C. and RGPIN-2019-06706 to C.G.), Fonds de Recherche Nature et Technologie (FRQNT) Team Grant (2021-PR-284335 to F.L.-C and C.G.), Sentinel North Initiative funded by Canada first Research Excellence Fund (2020-2024 Major Call for Proposal to 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 295824 to 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. was supported by a postdoctoral scholarship from the FRQNT strategic cluster UNIQUE. A.B. is supported by scholarships from NSERC and FRQNT, and from the FRQNT strategic cluster UNIQUE.

## 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.B. and B.T. generated the simulated nanodomain dataset. A.D., T.W. and C.B. performed the live imaging experiments. A.D. contributed to the integration of the TA-GAN on the microscope. A.B. provided trained networks, helped manage the data, analysed results on the simulated nanodomain dataset and designed the website. C.B., F.L.-C. and C.G. wrote the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s42256-023-00689-3>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42256-023-00689-3>.

**Correspondence and requests for materials** should be addressed to Christian Gagné or Flavie Lavoie-Cardinal.

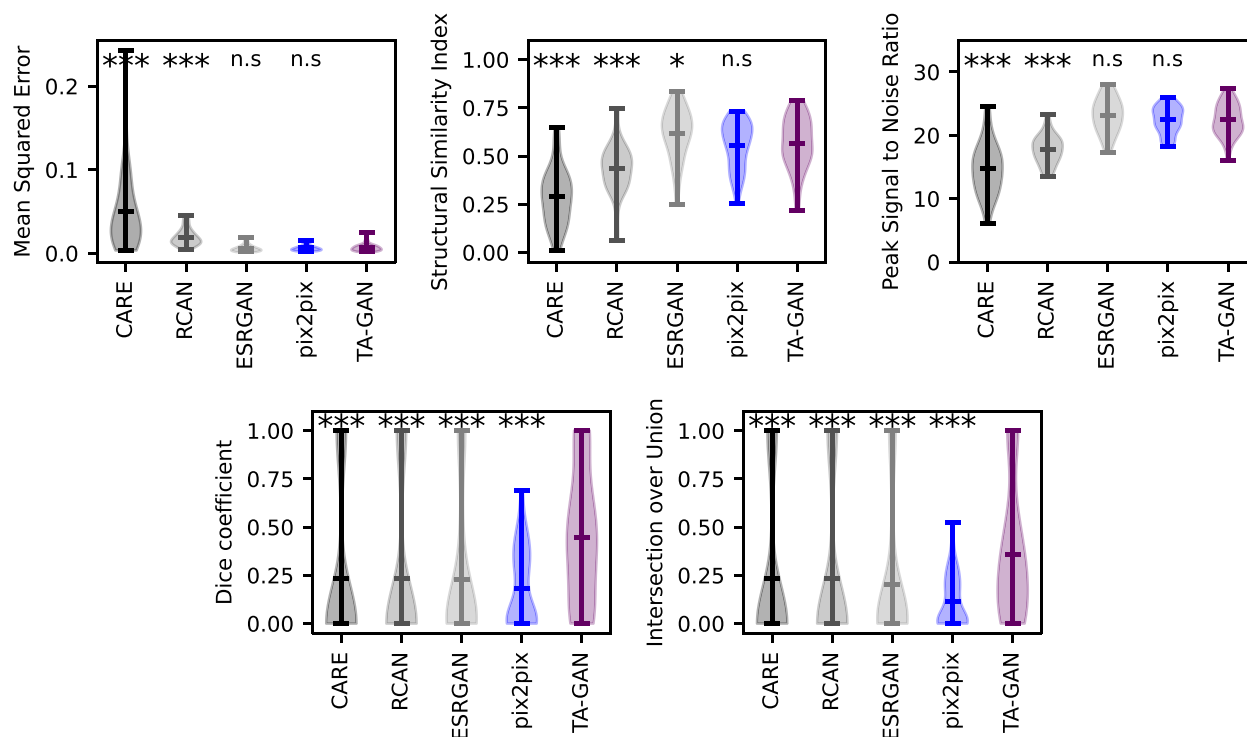
**Peer review information** *Nature Machine Intelligence* thanks Ruiming Cao, Jiji Chen, Ricardo Henriques and Estibaliz Gómez de Mariscal for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

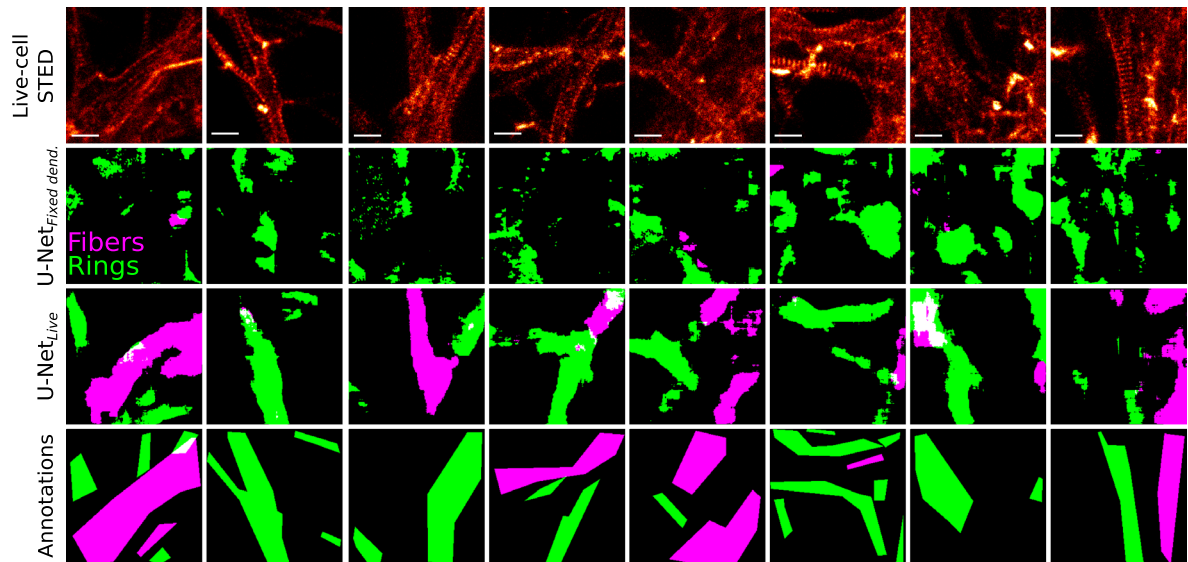
© The Author(s) 2023



**Extended Data Fig. 1 | Generation accuracy of TA-GANx compared with resolution enhancement baselines.** Comparison of TA-GAN<sub>ax</sub> with the resolution enhancement baselines using three image evaluation metrics: 1) Mean squared error (MSE), 2) Structural Similarity Index Measure (SSIM), 3) Peak Signal to Noise Ratio (PSNR), and two segmentation evaluation metrics: 1) Dice Coefficient (DC), 2) Intersection over Union (IOU). For the image metrics, images are normalized to 0-1 using min-max normalization. The segmentation predictions are computed with the U-Net<sub>fixed-ax</sub> on the synthetic images generated with each approach. Metrics are computed using the real STED image

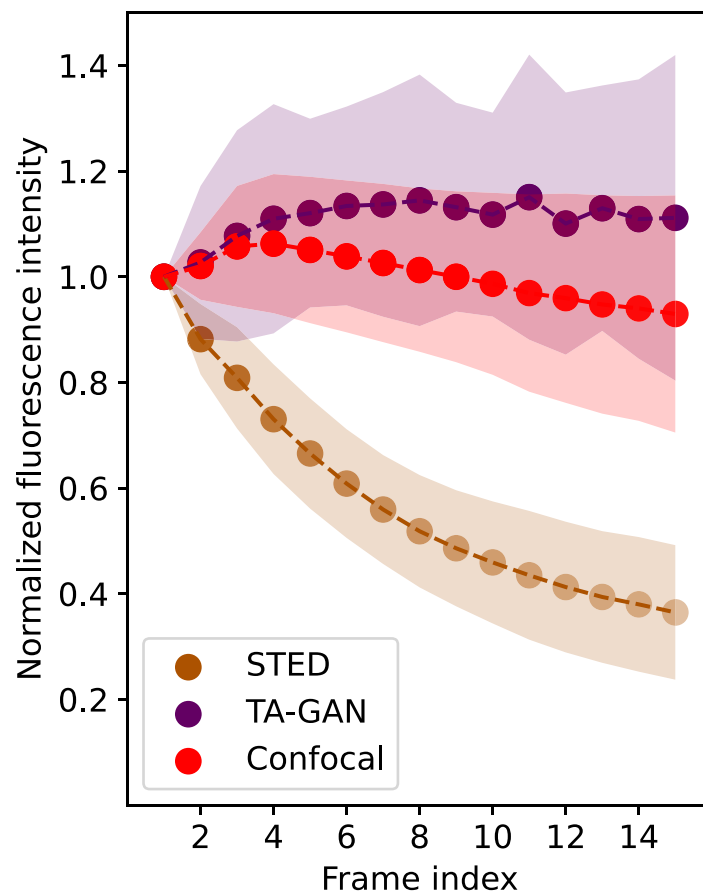
and its segmentation by U-Net<sub>fixed-ax</sub> as the reference. The score for DC and IOU is 1 if both the reference and prediction are empty. The performance of the TA-GAN is significantly better than all baseline for both segmentation metrics. For the image similarity metrics, TA-GAN performs significantly better than CARE and RCAN, and is similar to ESRGAN and pix2pix. Statistical analysis: Mann-Whitney U test<sup>62</sup> for the two-sided hypothesis that the distribution underlying the results for each baseline is the same as the distribution underlying the TA-GAN results. Violin plots show the minimum, maximum and mean of each distribution. (\*\*\*)  $p < 0.001$ , n.s.  $p > 0.05$ .  $n=52$  independent images.





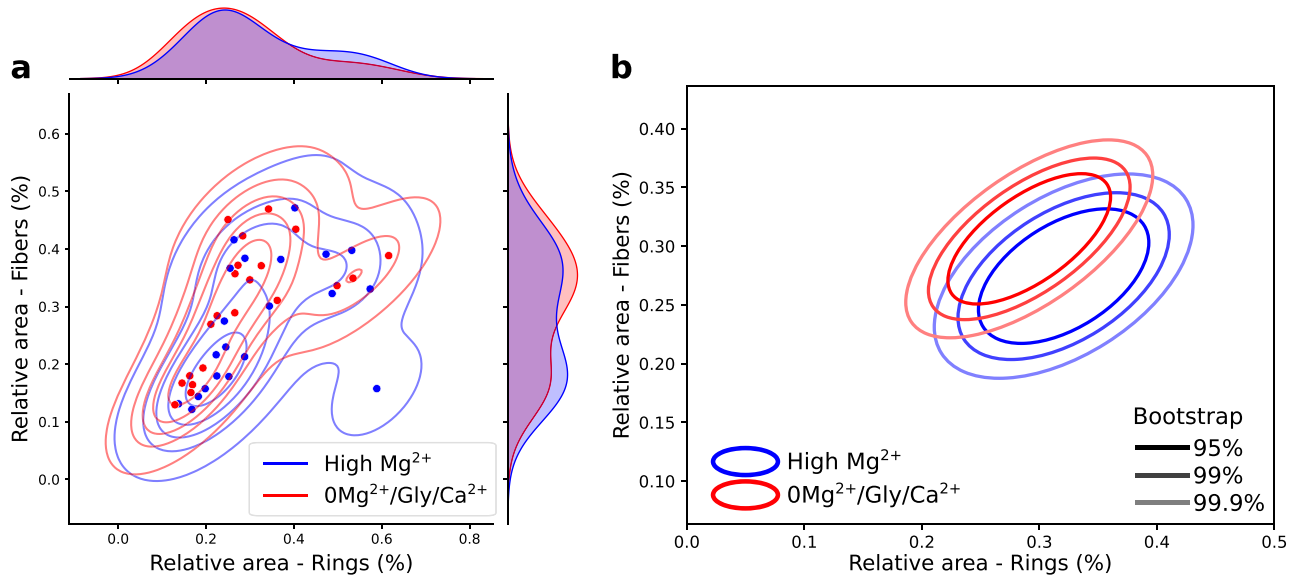
**Extended Data Fig. 2 | U-NetLive example results for the segmentation of F-actin nanostructures in live-cell STED images.** Segmentation predictions by U-Net<sub>Fixed dend.</sub><sup>13</sup> and U-Net<sub>Live</sub> on 8 representative images chosen from 28 annotated live-cell STED test images. Annotations were created for testing purposes and

were not used for training U-Net<sub>Live</sub>. The U-Net<sub>Live</sub> trained only on synthetic images from the *Translated F-actin dataset* succeeds in segmenting F-actin nanostructures on real STED images. Scale bars: 1  $\mu\text{m}$ .



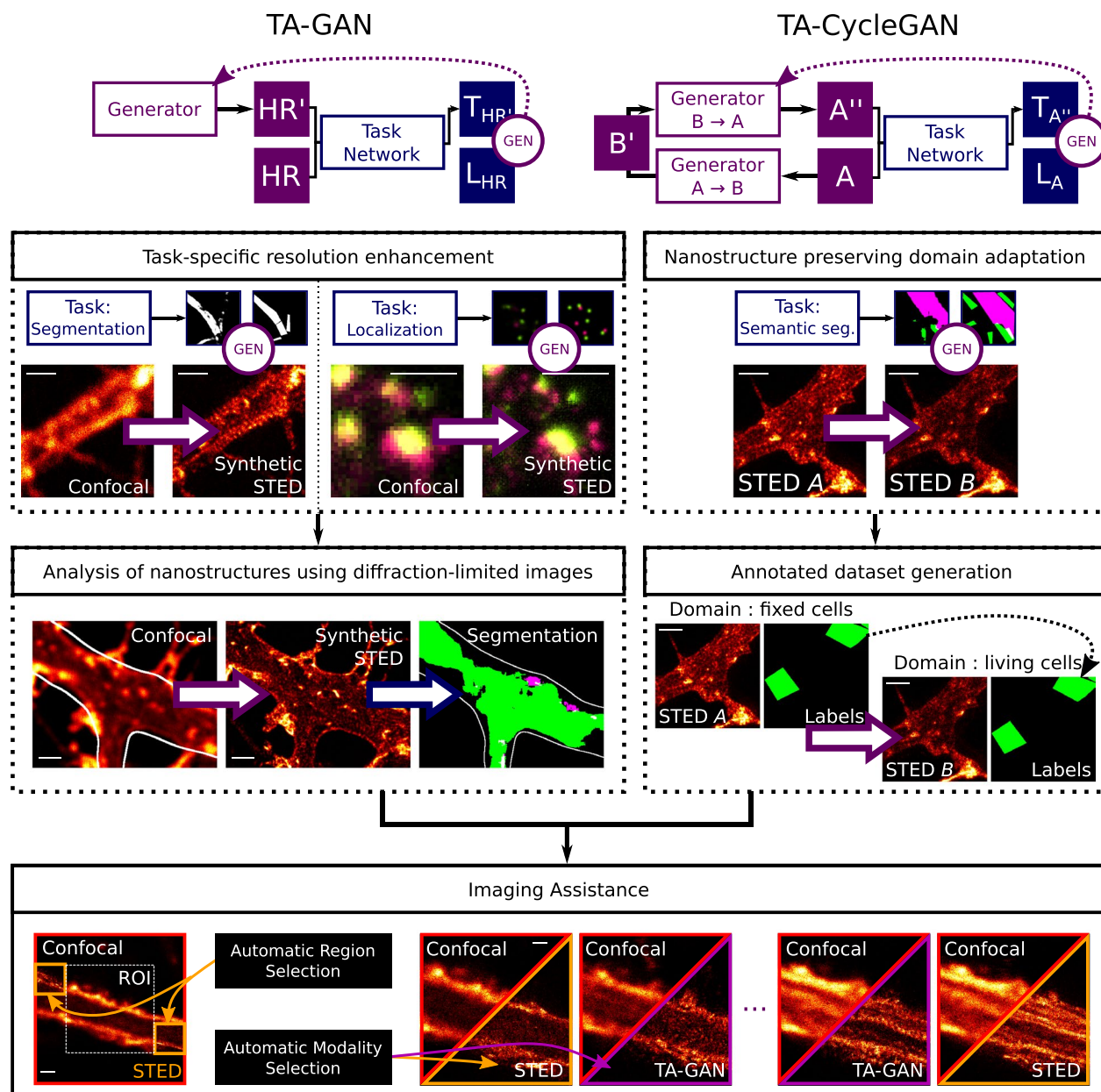
**Extended Data Fig. 3 | Comparison of photobleaching effects for consecutive confocal and STED acquisitions.** Normalized fluorescence intensity after 15 confocal acquisitions (red, N=45 regions) and, associated synthetic STED signal (purple, N=45 regions) over the central ROI ( $300 \times 300$  pixels) in comparison to acquisitions using the STED modality at each frame (orange, N=45 regions).

Dots show the average and shaded regions cover the standard deviation. The TA-GAN<sub>live</sub> predictions compensate for the fluorescence intensity decrease in the synthetic STED images. The 15th consecutive STED image has  $36 \pm 12\%$  of the initial STED image intensity and  $92 \pm 16\%$  for the sequence of confocal images for the corresponding TA-GAN generated images.



**Extended Data Fig. 4 | Observation of F-actin remodeling in living cells. a,** Kernel density estimate of the F-actin fibres and rings dendritic area distribution for after 30 minutes in a solution reducing neuronal activity (high  $Mg^{2+}$ /low  $Ca^{2+}$ , blue) or following a stimulation ( $0Mg^{2+}/Glu/Ca^{2+}$ , from  $t = 1-15min$ , red). **b,** Bootstrapped distributions of the results shown in **a**. Shown are the regions

comprising 95%, 99% and 99.9% of the data point distribution. Following the  $0Mg^{2+}/Glu/Ca^{2+}$  stimulation, we observe a small increase in the proportion of F-actin fibres and a decrease in the proportion of rings. High  $Mg^{2+}$   $N=21$ ,  $0Mg^{2+}/Glu/Ca^{2+}$   $N=21$ .



**Extended Data Fig. 5 | 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. The generation loss (GEN circle) is computed from the comparison between the output of the task network for the synthetic high-resolution image ( $T_{HR'}$ ) and the labels obtained from the ground truth image ( $L_{HR}$ ). The loss is backpropagated to the generator (dashed arrow). *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, even if labels are only available in one domain. The generation loss (GEN circle) is computed from the

comparison between the output of the task network for the image and the synthetic version ( $T_{A''}$ ) and the labels obtained from the input domain A image ( $L_A$ ). The loss is backpropagated to the generator (dashed arrow). *Middle-right:* Labeled datasets from domain A (e.g. fixed cells) are adapted to the unlabeled domain B (e.g. live cells) to obtain a labeled dataset from domain B, which can be used to train a super-resolution TA-GAN. *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<sub>Live</sub> predictions. Scale bars: 1  $\mu\text{m}$ .

**Extended Data Table 1 | Processing time with and without the TA-GAN-assisted implementation**

	Confocal acquisition	Generation	Optical flow computation	Sub-region acquisition	Generation & segmentation	STED acquisition	Total
<b>With</b> TA-GAN assistance	2.6 seconds	2.5 seconds	6.1 seconds	1.3 seconds	2.5 seconds	13.6 sec. (for 11-25% of frames)	15.0 sec. (for 75-89% of frames)
<b>Without</b> TA-GAN assistance	2.6 seconds					13.6 sec. (Every frame)	16.2 seconds

The number of STED images acquired with the TA-GAN assistance varies between sequences and averages 11% for the DC-based triggers and 25% for the variability-based triggers.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Confocal and STED images were acquired on a 4 color Abberior Expert-Line STED microscope (Abberior Instruments GmbH, Germany), equipped with a 100x 1.4 NA oil objective and using pulsed (40 MHz) excitation (640 nm) and depletion (775 nm) lasers. Pixel size was set to 20 nm for the Axonal F-actin dataset, the Dendritic F-actin dataset and the Live F-actin dataset for both confocal and STED images. For the Synaptic protein dataset, pixel size was set to 15 nm for the STED images and 60 nm for the confocal images.

Data analysis

OpenCV 3.1.0 was used for image registration. Flask 2.0.3 was used for communicating between the microscope computer and the GPU-equipped computer. Specpy 1.2.1 was used to automate the acquisitions on the Abberior STED microscope. For training and testing TAGAN, we used Python 3.6 with the package versions mentioned in the project Dockerfile (<https://github.com/FLClab/TA-GAN/blob/main/TAGAN-Docker/Dockerfile>): pytorch (1.0.0), torchvision (0.2.1), numpy (1.19.2), Pillow (8.3.1), tifffile (2020.0.3), scipy (1.5.4), and scikit-image (0.17.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

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>. The publicly available *S. aureus* dataset from Spahn et al. 2022 is available at <https://zenodo.org/record/5550933#.Y6lhFNLMJH4> and <https://zenodo.org/record/5551141#.Y6ljBdLMJH5>.

Sample test images are available at <https://github.com/FLClab/TA-GAN> in the "test" subfolders of each dataset. Results from the paper can be reproduced with the complete test sets, all available to download at <https://s3.valeria.science/flclab-tagan/index.html> (10.5281/zenodo.7908914).

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	<input type="text" value="N/A"/>
Population characteristics	<input type="text" value="N/A"/>
Recruitment	<input type="text" value="N/A"/>
Ethics oversight	<input type="text" value="N/A"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences     Behavioural & social sciences     Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Training of the TA-GAN on already existing and published datasets was performed using all available images. The size of these public datasets determined the sample size for each experiment. This includes the Axonal F-actin dataset, Dendritic F-actin dataset, Synaptic protein dataset, and the DeepBacs dataset. For the live F-actin dataset, no sample size calculation was performed for the acquisition of the dataset. To estimate the number of images required to train the TA-CycleGAN, the network was trained on partial datasets and its performance was evaluated on a validation dataset. New images were acquired until the synthetic STED images generated from the validation confocal images could not be told apart from real STED images by experts. The live F-actin training dataset consists of 753 paired confocal and STED images of F-actin structures of 168 neurons from 6 independent neuronal cell culture preparation. The performance of the TA-GAN on images from each batch was compared and no difference between the batches was measured.
Data exclusions	Axonal F-actin dataset : we excluded 31 images containing mostly background, where no structures had been annotated. Dendritic F-actin dataset : All images from the original dataset were included in this study. Synaptic protein dataset : All images for three pairs of proteins (PSD95-Bassoon, PSD95-Homer1c and Bassoon-Homer1c) from the original dataset were included in this study. Live F-actin dataset : Some images were discarded based on visual inspection (loss of focus, sample drift, imaging artifacts). The imaging sequences included in the manuscript were selected to best illustrate specific use cases of the approach (e.g. unexpected biological change, expected biological change). Other imaging sequences were included in the supplementary materials.
Replication	Time-lapse STED imaging using the TA-GAN assistance were replicated on 53 neurons from 5 independent primary neuronal cultures. On each coverslip, the experiment was performed on 3 regions from the same neuron.  Once optimal hyperparameters had been selected from multiple trials, all trainings of the neural networks were replicated twice with different initial randomization of the initial weights, using the same train/valid splits and the same hyperparameters, to verify convergence and ensure reproducibility of the results. All results from the paper are produced from the first trained models, with the second trained model only used for validation of the reproducibility. Time-lapse STED imaging using the TA-GAN assistance were replicated on 53 neurons from 5 independent primary neuronal cultures. On each coverslip, the experiment was performed on 3 regions from the same neuron.

Once optimal hyperparameters had been selected from multiple trials, all trainings of the neural networks were replicated twice with different initial randomization of the initial weights, using the same train/valid splits and the same hyperparameters, to verify convergence and ensure reproducibility of the results. All results from the paper are produced from the first trained models, with the second trained model only used for validation of the reproducibility.

**Randomization** The train/valid splits of each dataset are random. For the datasets taken from previous publications, the same test sets were used so that results could be compared.

**Blinding** All user-studies were blind; the generation approach as well as the name and acquisition parameters of the images were not displayed to the participant, and the participant was not aware of whether the image was real or generated.

For all experiments performed on already acquired datasets, blinding was not relevant as all images were associated to the same experimental group and training/validation set split was done randomly, while keeping the same test set as in the original publication.

Blinding was not required for the acquisition of the live F-actin dataset as images were acquired from samples belonging to a single experimental group. Split of the live F-actin dataset in training/validation/test sets was done randomly.

For the acquisition of the TA-GAN assisted imaging sequences, investigators were blinded during the region selection process concerning which of 3 approaches would be used on a specific region: 1) TA-GAN assisted imaging, 2) control sequences of paired STED and confocal images, 3) only STED acquisitions.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- n/a  Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology and archaeology
- Animals and other organisms
- Clinical data
- Dual use research of concern

### Methods

- n/a  Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

## Antibodies

### Antibodies used

#### Axonal F-actin dataset :

Mouse-anti-SMI31, Biolegend, cat. 8016011:250, <https://www.biolegend.com/it-it/products/purified-anti-neurofilament-h-nf-h-phosphorylated-antibody-11476>

Goat anti mouse STAR580, Abberior 2-0002-005-1, Dilution 1:250, <https://abberior.shop/abberior-STAR-580-goat-anti-mouse-igg-500-ll-1-mg-ml>

Phalloidin-STAR635, Abberior, cat. 2-0205-002-5, Dilution 1:50, <https://abberior.shop/abberior-STAR-635-phalloidin-20-ig>

#### Dendritic F-actin dataset :

Rabbit-anti-MAP2, Milipore Sigma, cat. AB5622, Dilution 1:1000, <https://www.sigmaaldrich.com/CA/en/product/mm/ab5622>

Goat anti Rabbit-STAR488, Abberior 2-0012-006-5, Dilution 1:250, <https://abberior.shop/abberior-STAR-488-goat-anti-rabbit-igg-500-ll-1-mg-ml>

Phalloidin-STAR635, Abberior, cat. 2-0205-002-5, Dilution 1:50, <https://abberior.shop/abberior-STAR-635-phalloidin-20-ig>

#### Synaptic protein dataset :

##### Primary antibodies :

Mouse anti-PSD95 (6G6-1C9), Abcam MA1-045, Dilution 1:500, <https://www.thermofisher.com/antibody/product/PSD-95-Antibody-clone-6G6-1C9-Monoclonal/MA1-045>

Rabbit anti-Bassoon, Synaptic Systems 141003, Dilution 1:500, <https://sysy.com/product/141003>

Rabbit anti-Homer1c, Synaptic Systems 160023, Dilution 1:500, <https://sysy.com/product/160023>

Oregon Green 488 Phalloidin, ThermoFisher O7466, Dilution 1 : 50

<https://www.thermofisher.com/order/catalog/product/O7466?SID=srch-hj-O7466>

Goat anti Mouse STAR 635P, Abberior 2-0002-007-5, Dilution 1 : 250, <https://abberior.shop/abberior-STAR-635P-goat-anti-mouse-igg-500-ll-1-mg-ml>

Goat anti Mouse Alexa Fluor 594, ThermoFisher A11005, Dilution 1:100, <https://www.thermofisher.com/antibody/product/Goat-anti-Mouse-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11005>

Goat anti Rabbit STAR 635P, Abberior 2-0012-007-5, Dilution 1 : 250, <https://abberior.shop/abberior-STAR-635P-goat-anti-rabbit-igg-500-ll-1-mg-ml>

Goat anti Rabbit Alexa Fluor 594, ThermoFisher A11037, Dilution 1:100, <https://www.thermofisher.com/antibody/product/Goat-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11037>



Live F-actin dataset :  
SiR-Actin, Supplier: Cytoskeleton inc., Cat: CY-SC001, Manufacturer's website : <https://www.cytoskeleton.com/sir-actin>.

## Validation

Axonal F-actin dataset and Dendritic F-actin dataset : 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).  
Synaptic protein dataset : 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).  
Live F-actin dataset : Manufacturer's website : <https://www.cytoskeleton.com/sir-actin>

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

Wild animals

Reporting on sex

Field-collected samples

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.