nature methods

ARTICLES

https://doi.org/10.1038/s41592-021-01080-z

Deep learning-based point-scanning super-resolution imaging

Linjing Fang[1], Fred Monroe[2], Sammy Weiser Novak[1], Lyndsey Kirk[3], Cara R. Schiavon[1], Seungyoon B. Yu[4], Tong Zhang[1], Melissa Wu[1], Kyle Kastner[5], Alaa Abdel Latif[6], Zijun Lin[6], Andrew Shaw[6], Yoshiyuki Kubota[7], John Mendenhall[3], Zhao Zhang[8], Gulcin Pekkurnaz[4], Kristen Harris[3], Jeremy Howard[6] and Uri Manor[1]✉

Point-scanning imaging systems are among the most widely used tools for high-resolution cellular and tissue imaging, benefiting from arbitrarily defined pixel sizes. The resolution, speed, sample preservation and signal-to-noise ratio (SNR) of point-scanning systems are difficult to optimize simultaneously. We show these limitations can be mitigated via the use of deep learning-based supersampling of undersampled images acquired on a point-scanning system, which we term point-scanning super-resolution (PSSR) imaging. We designed a 'crappifier' that computationally degrades high SNR, high-pixel resolution ground truth images to simulate low SNR, low-resolution counterparts for training PSSR models that can restore real-world undersampled images. For high spatiotemporal resolution fluorescence time-lapse data, we developed a 'multi-frame' PSSR approach that uses information in adjacent frames to improve model predictions. PSSR facilitates point-scanning image acquisition with otherwise unattainable resolution, speed and sensitivity. All the training data, models and code for PSSR are publicly available at 3DEM.org.

An essential tool for understanding the spatiotemporal organization of biological systems, microscopy is nearly synonymous with biology itself. Microscopes suffer from the so-called 'eternal triangle of compromise', which dictates that image resolution, illumination intensity (and consequent sample damage) and imaging speed are all in tension with one another. In other words, it is impossible to optimize one parameter without compromising another. This is particularly problematic for point-scanning systems, for example scanning electron and laser scanning confocal microscopes, for which higher-resolution images require higher numbers of sequentially acquired pixels to ensure proper sampling, thus increasing the imaging time and sample damage in direct proportion to the image resolution. Nonetheless, point-scanning systems remain perhaps the most common imaging modality in biological research due to their versatility and ease of use. Thus, the ability to effectively supersample undersampled point-scanning microscope images could be transformative.

Deep learning (DL) has been extensively used to 'supersample' the pixels in computationally downsampled digital photographs[1–4]. For microscopy, DL has long been established as an invaluable method for image analysis and segmentation[5]. More recently, DL has been used with spectacular results in restoring relatively low signal-to-noise ratio (SNR) or low resolution fluorescence microscopy acquisitions[6–10]. DL has also been used for accelerating single-molecule localization microscopy[10–12]. Similarly, low-resolution (LR) or low SNR electron microscopy (EM) data have been restored using DL[8,13–16]. Finally, DL has been used to restore blurry, low optical resolution images to sharp, high 'optical' resolution images via either supervised training of fluorescence samples[10,17,18], or via more generalizable deconvolution algorithms[19,20].

Increasing the *xy* pixel resolution of real-world undersampled point-scanning microscope images presents a unique set of both challenges and opportunities. Undersampling point-scanning microscope images in the *xy* plane while maintaining a constant pixel dwell time increases the imaging speed and decreases the sample damage, in addition it results in not only a lower pixel resolution image, but also a lower SNR, since the total photons or electrons detected are in this case proportional to the number of pixels collected. Thus, restoring undersampled point-scanning microscope images with DL requires simultaneous supersampling and denoising, a challenging task that requires large amounts of high-quality training data.

To accomplish this, we developed a DL-based training workflow that uses a new 'crappifier' that simultaneously injects noise while downsampling the pixel resolution of high-resolution (HR) training data. This crappifier circumvents the need to acquire real-world image pairs for training, which is difficult and expensive for large datasets, and is practically impossible for live samples with quickly moving structures (for example, subcellular organelles). These crappified images are then paired with their HR counterparts to train models that can successfully supersample and denoise real-world undersampled, noisy images. We found crappifiers using additive-Gaussian noise performed best for training models that effectively restore real-world images. We also found that DL restored fluorescence time-lapse images of fast-moving subcellular organelles suffer from flickering artifacts when restoring very LR, low

[1]Waitt Advanced Biophotonics Center, Salk Institute for Biological Studies, La Jolla, CA, USA. [2]Wicklow AI Medical Research Initiative, San Francisco, CA, USA. [3]Department of Neuroscience, Center for Learning and Memory, Institute for Neuroscience, University of Texas at Austin, Austin, TX, USA. [4]Neurobiology Section, Division of Biological Sciences, University of California San Diego, La Jolla, CA, USA. [5]Montreal Institute for Learning Algorithms, Université de Montréal, Montréal, Canada. [6]Fast.AI, University of San Francisco Data Institute, San Francisco, CA, USA. [7]Division of Cerebral Circuitry, National Institute for Physiological Sciences, Okazaki, Japan. [8]Texas Advanced Computing Center, University of Texas at Austin, Austin, TX, USA. ✉e-mail: umanor@salk.edu

SNR images. To address this, we developed a 'multi-frame' approach for high-spatiotemporal resolution time-lapse data, which reduces flickering artifacts and generates more accurate super-resolution images by using information from neighboring video frames in time-lapse acquisitions. Thus, our point-scanning super-resolution (PSSR) software provides a practical and powerful framework for simultaneously increasing the sensitivity, pixel resolution, 'optical' resolution and acquisition speed of any point-scanning imaging system.

Results

Three-dimensional electron microscopy (3DEM) is a powerful technique for determining the volumetric ultrastructure of tissues. In addition to serial section EM (ssEM)[21] and focused ion beam–scanning EM (FIB–SEM)[22], one of the most common tools for high-throughput 3DEM imaging is serial blockface scanning EM (SBFSEM)[23], wherein a built-in ultramicrotome iteratively cuts ultrathin sections off the surface of a blockface after it is imaged with a scanning electron probe. This technology facilitates relatively automated, high-throughput 3DEM imaging with minimal post-acquisition image alignment. Unfortunately, higher electron doses cause sample charging, which renders the sample too soft to section reliably (Supplementary Video 1). Furthermore, the extremely long imaging times and large file sizes inherent to HR volumetric imaging present a substantial bottleneck. For these reasons, most 3DEM datasets are acquired with sub-Nyquist pixel sampling (for example, pixel sizes ≥4 nm), which precludes the reliable detection or analysis of smaller subcellular structures, such as roughly 35-nm presynaptic vesicles. While low-pixel resolution 3DEM datasets can be suitable for many analyses (for example, cellular segmentation), the ability to mine targeted regions of pre-existing large datasets for higher-resolution ultrastructural information would be extremely valuable. Unfortunately, many 3DEM imaging approaches are destructive, and HR ssEM can be slow and laborious. Thus, the ability to increase the pixel resolution of these 3DEM datasets post hoc is highly desirable.

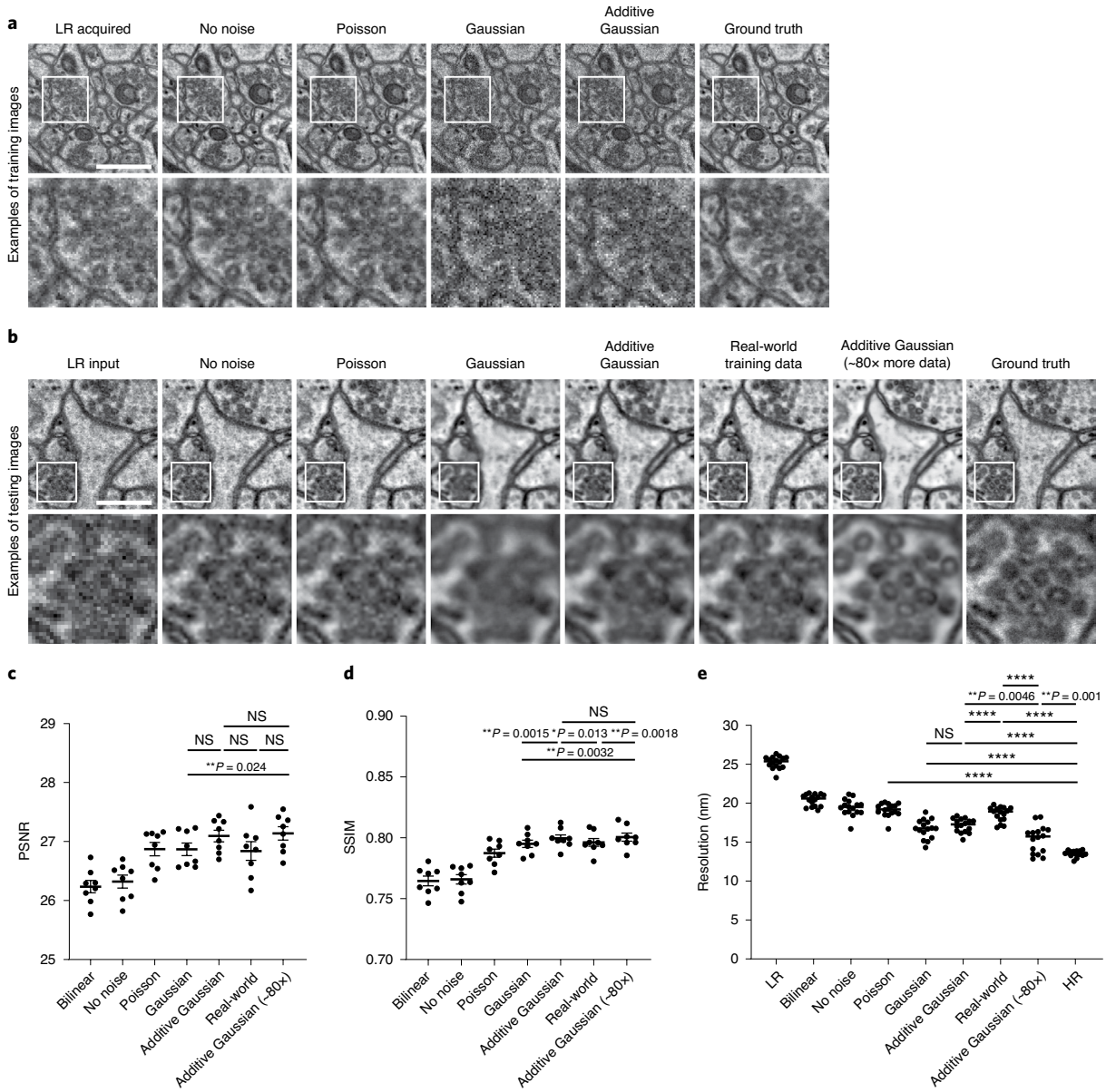
Frustrated by our inability to perform SBFSEM imaging with the desired 2-nm pixel resolution and SNR necessary to reliably detect presynaptic vesicles, we decided to test whether a deep convolutional neural net model (PSSR) trained on 2-nm-pixel HR images could 'super-resolve' 8-nm-pixel LR images. We also wished to enable imaging with equal pixel dwell times as the HR images, facilitating a 16× increase in imaging speed, but at the cost of a 16× decrease in detected electrons, greatly lowering the SNR. Thus, our PSSR model would need to simultaneously increase the SNR and pixel resolution of the LR images. To train a DL model for this purpose, a very large number of perfectly aligned HR and LR image pairs would be required. However, EM imaging of ultrathin sections results in nonlinear distortions that are impossible to predict or control. Therefore, it is extremely difficult and sometimes impossible to perfectly align two sequential acquisitions of the same field of view. Furthermore, sequentially acquiring a large enough number of images for training data is time intensive and cost prohibitive. Thus, instead of manually acquiring HR and LR image pairs for training, we opted to generate training data by computationally 'crappifying' HR images to be paired with their HR counterparts.

We hypothesized that to simulate real-world, low-pixel resolution acquisitions with lower electron doses, our 'crappifier' must add noise while also decreasing the pixel resolution. To create a suitable crappifier, we performed an ablation study where we compared the performance of models trained with image pairs generated with different crappifiers, while all other factors were kept constant (Fig. 1a–e and Methods). Specifically, we compared models crappified with no noise (that is, downsampling only), Poisson noise, Gaussian noise (independently and identically distributed) and additive-Gaussian-distributed noise, respectively. Among all models trained with crappified pairs, additive Gaussian yielded the best results. The additive-Gaussian model also outperformed the model trained with manually acquired training pairs (real world) across all metrics. We further compared additive Gaussian with 'additive Gaussian (roughly 80×)', where we used approximately 80× more training data, which did not substantially increase the peak-signal-to-noise ratio (PSNR) or structural similarity (SSIM) measurements, but did further increase the resolution as measured by Fourier ring correlation (FRC) analysis. Notably, it would have taken >480 h of imaging time and cost >US$16,000 to acquire the same amount of real-world training data pairs as was used in our 80× additive-Gaussian model. But by using pre-existing gold standard HR data that had already been acquired for separate experimental purposes, we were able to generate training data in only 2 h, at a cost of only US$16 (Fig. 1f). This highlights the use of the crappifier method, which facilitates the generation of much larger amounts of training data at a fraction of the cost. Therefore, for our training data we 'crappified' approximately 130 GB of training data from 2-nm-pixel transmission-mode scanning electron microscope (tSEM[24]) images of 40-nm ultrathin sections of rat CA1 hippocampal tissue. We then trained our image pairs with a ResNet-based U-Net architecture (Fig. 2a and Extended Data Fig. 1: see Methods and tables for full details). Using a mean squared error (MSE) loss function yielded excellent results as determined by visual inspection as well as PSNR, SSIM and FRC analyses. Overall, the PSSR-restored images from our semisynthetic pairs contained more high-frequency information than our LR images, and yet displayed less noise than both our LR and HR images, making it easier to identify smaller objects such as 35-nm presynaptic vesicles (Fig. 2b,c).

We next tested whether our PSSR model was effective on real-world LR acquisitions. DL-based models are notoriously sensitive to variations in training versus testing data, usually precluding the use of models generated from training images acquired in one condition on images acquired in another (that is, data generated using a different sample preparation technique, type or on a different microscope). Similarly, there is a risk that models trained on artificially injected noise will learn how to remove artificial noise yet fail to remove real-world noise.

Our training images were generated from 40-nm sections of rat CA1 tissue acquired with a tSEM detector. But for our testing data, we acquired HR and LR images of 80-nm sections of dentate gyrus tissue from a mouse brain, imaged with a backscatter detector. Based on several metrics including PSNR, SSIM, FRC (Fig. 2b,c), NanoJ-SQUIRREL error mapping analysis (Extended Data Fig. 2)[25], visual inspection and comparison to the block-matching and 3D filtering (BM3D) denoising method (Extended Data Fig. 3)[26], we found PSSR successfully restored real-world LR images (Fig. 2c). Thus, our PSSR model is effective for real-world data, and is not restricted to data acquired in the exact same fashion as the training set.

We next asked whether we could sufficiently restore 8-nm SBFSEM datasets to 2 nm using PSSR, since high-quality 2-nm SBFSEM imaging is currently difficult or impossible for us to achieve. Using the same PSSR model described above we were able to restore an 8-nm-pixel SBFSEM 3D dataset to 2 nm (Fig. 3a and Supplementary Video 2). Our PSSR model also worked very well on mouse, rat and fly samples imaged on four different microscopes in four different laboratories (Fig. 3a–d). In addition to our SBFSEM and SEM imaging systems, PSSR processing appeared to restore images acquired on a Zeiss FIB–SEM (from the Hess laboratory at Janelia Farms, Fig. 3c and Supplementary Video 3) and a Hitachi Regulus field emission–SEM (FE–SEM) (from the Kubota laboratory at National Institute for Physiological Sciences, Fig. 3d). PSSR processing also performed well on a $10 \times 10 \times 10\,nm^3$ resolution



f

Training data type	Training image aquisition		Total time (for 80 GB of EM data)	Difficulty	Total cost[a]
	HR	LR			
Real-world data pairs	Acquired	Acquired	≥480 imaging hours	Data must be perfectly aligned	~US$16,800
Semisynthetic data pairs	Acquired, if not available	Crappified	~240 imaging hours + ~2 h of computational crappification	No alignment necessary	~US$8,416
	Not acquired, if available	Crappified	0 imaging hours + ~2 h of computational crappification	No alignment necessary	~US$16

[a]Cost estimate is based on US$35 per h for EM imaging, and US8 per h for computation.

Fig. 1 | Evaluation of crappifiers with different noise injection on EM data. a, Different crappifiers applied to HR, high SNR images, including 'no noise' (no added noise, downsampled pixel size only), Poisson, Gaussian and additive-Gaussian noise. The real-world acquired LR (LR acquired) and HR (ground truth) images are also shown for comparison. Each training set contains 40 image pairs, achieving similar results. **b**, Visualized restoration performance of PSSR models that were trained on each different crappifier (no noise, Poisson, Gaussian and additive Gaussian), as well as a model trained with manually acquired LR versions of the same samples used for the HR semisynthetic training data (real-world training data). Results from a model that used the same crappifier as additive Gaussian, but with roughly 80× more training data (additive Gaussian (roughly 80×)) are also displayed. LR input and ground truth of the example testing regions of interest (ROI) are also shown. Scale bars, 0.4 µm. **c**–**e**, Experiments were repeated with 8–16 images, achieving similar results. PSNR (**c**), SSIM (**d**) and resolution as measured by FRC analysis (**e**) (PSNR and SSIM, $n = 8$ independent images; FRC resolution, $n = 16$ independent images). **f**, A table that compares the devoted time, cost and difficulty level between experiments with manually acquired training pairs and experiments using our crappification method. All values are shown as mean ± s.e.m. P values are specified in the figure for $0.0001 < P < 0.05$. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, NS, not significant; two-sided paired t-test.

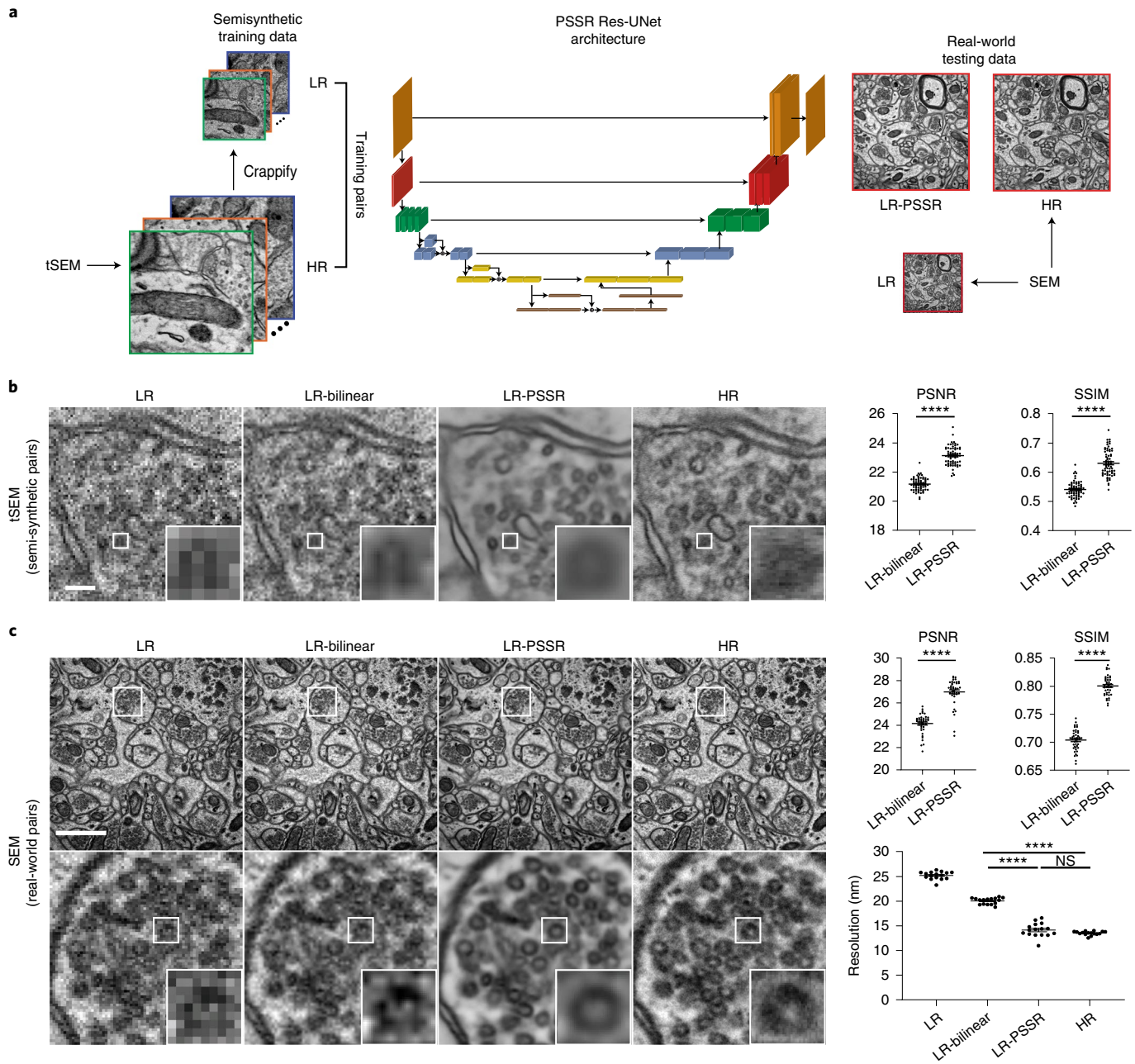


Fig. 2 | Restoration of semisynthetic and real-world EM testing data using PSSR model trained on semisynthetically generated training pairs. a, Overview of the general workflow. Training pairs were semisynthetically created by applying a degrading function to the HR images taken from a tSEM to generate LR counterparts (left column). Semisynthetic pairs were used as training data through a dynamic ResNet-based U-Net architecture. Layers of the same *xy* size are in the same color (middle column). Real-world LR and HR image pairs were both manually acquired under a SEM (right column). The output from PSSR (LR-PSSR) when LR is served as input is then compared to HR to evaluate the performance of our trained model. **b**, Restoration performance on semisynthetic testing pairs from tSEM. Shown is the same field of view of a representative bouton region from the synthetically created LR input with the pixel size of 8 nm (left column), a 16× bilinear-upsampled image with 2 nm pixel size (second column), 16× PSSR upsampled result with 2 nm pixel size (third column) and the HR ground truth acquired at the microscope with the pixel size of 2 nm (fourth column). A close view of the same vesicle in each image is highlighted. The PSNR and the SSIM quantification of the semisynthetic testing sets are shown (right) ($n = 66$ independent images). Scale bar, 0.1 μm **c**, Restoration results of manually acquired SEM testing pairs. Shown is the comparison of the LR input acquired at the microscope with a pixel size of 8 nm (left column), 16× bilinear-upsampled image (second column), 16× PSSR upsampled output (third column) and the HR ground truth acquired at the microscope with a pixel size of 2 nm (fourth column). The bottom row compares the enlarged region of a presynaptic bouton with one vesicle highlighted in the inset. Graphs comparing PSNR, SSIM and image resolution are also displayed (right). The PSNR and SSIM values were calculated between an upsampled result and its corresponding HR ground truth ($n = 42$ independent images). Resolution was calculated with the FRC plugin in NanoJ-SQUIRREL by acquiring two independent images at low and HR ($n = 16$ independent images). Scale bar, 1 μm. All values are shown as mean ± s.e.m. *P* values are specified in the figure for $0.0001 < P < 0.05$. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, NS, not significant; two-sided paired *t*-test.

FIB–SEM fly brain dataset, resulting in a $2\times2\times10\,nm^3$ resolution dataset with higher SNR and resolution (Fig. 3b). Thus, PSSR can provide 25× super-resolution with useful results, effectively increasing the lateral resolution and speed of FIB–SEM imaging by a factor of at least 25×.

The main concern with DL-based image processing is the possibility of false positives (also known as 'hallucinations')[5,8,27,28]. To further test the accuracy and use of the PSSR output within a more concrete, biological context, we next randomized bilinear interpolated (LR-bilinear), LR-PSSR and HR images, then distributed them in a blinded manner to two human experts for manual segmentation of presynaptic vesicles, which are difficult to detect with 8-nm pixel resolution images (see Supplementary Notes and Methods for full details). We found the LR-PSSR segmentation was notably more accurate than the LR-bilinear (Fig. 3e). While the LR-PSSR output substantially reduced false negatives, the LR-PSSR output had a slightly higher number of 'false positives' than the LR-bilinear. However, the variance between the LR-PSSR and HR results was similar to the variance between the two expert human results on HR data (Fig. 3e), which is the current gold standard. Notably, we found PSSR images were much easier to segment: a major bottleneck for analyzing 3DEM datasets (Supplementary Video 4).

Similar to SBFSEM, laser scanning confocal microscopy also suffers from a direct relationship between pixel resolution and sample damage (that is, phototoxicity/photobleaching)[29]. This can be a large barrier for cell biologists who wish to study the dynamics of smaller structures such as mitochondria, which regularly undergo fission and fusion but also show increased fission and swelling in response to phototoxicity (Supplementary Video 5 and Extended Data Fig. 4).

Laser scanning microscopy also suffers from an inverse relationship between pixel resolution and imaging speed, making live-cell imaging of faster processes (for example, organelle motility in neurons) challenging if not impossible. Nonetheless, Nyquist sampling criteria often necessitates the use of smaller pixels to resolve smaller structures—this is particularly true for higher-resolution imaging methods that depend on postprocessing pixel reassignment and/or deconvolution. Thus, we sought to determine whether PSSR might enable acquisitions with decreased pixel resolution to optimize the imaging speed and SNR of live laser scanning confocal microscope imaging. To train a DL model using image pairs of mitochondria in live cells is virtually impossible because they are constantly moving and changing shape. Thus, our 'crappification' approach is particularly useful generating training data for live-cell imaging datasets.

To generate our ground truth training dataset, we used a Zeiss laser scanning confocal microscope (LSM) 880 in Airyscan mode (Methods). Similar to our EM model, an ablation study that compared crappifiers with different noise distributions was conducted (Extended Data Fig. 5 and Methods). We found the crappifier injected with Salt&Pepper and additive-Gaussian noise yielded overall best performance. For the real-world LR test data, we acquired images in confocal mode at 16× lower pixel resolution with a 2.5 AU pinhole on a photomultiplier (PMT) confocal detector, without any additional image processing. To ensure minimal phototoxicity, we also decreased the laser power for our LR acquisitions by a factor of 4 or 5 (see Tables 1 and 2 for more details), resulting in a net laser dose decrease of roughly 64–80×. Thus, our PSSR model was trained to restore LR, low SNR, low-pixel resolution confocal images to high SNR, high-pixel resolution, high 'optical' (that is, deblurred) resolution Airyscan-equivalent image quality. To start, we trained on live-cell time-lapse acquisitions of mitochondria in U2OS cells. As expected, imaging at full resolution resulted in serious bleaching and phototoxicity-induced mitochondrial swelling and fission (Supplementary Video 5). However, the LR acquisitions were extremely noisy and pixelated due to undersampling. However, the LR scans showed far less photobleaching (Extended Data Fig. 4). Similar to our EM data, LR-PSSR images had higher SNR and resolution compared to LR acquisitions, as determined by testing on both semisynthetic and real-world LR versus HR image pairs. Notably, LR-PSSR images also had higher SNR than HR images (Fig. 4d).

We observed severe 'flickering' in LR-PSSR time-lapses (Supplementary Video 6 and Fig. 4b) due to noise-induced variations in signal detection and image reconstruction, causing both false breaks and merges in mitochondrial networks (Fig. 4c, white boxes and Fig. 4d, red and yellow arrows), making it impossible to accurately detect bona fide mitochondrial fission or fusion events. This temporal inconsistency was reflected in neighboring-frame cross-correlation analysis (Fig. 4b, see Methods for full details). One strategy for increasing the SNR of images is to average multiple scans, for example 'frame averaging'; this method can also be used to reduce 'flickering' effects in videos (Fig. 4b). However, this approach is problematic for live imaging of quickly moving objects: If objects move greater than half the distance of the desired spatial resolution between individual frames, temporal Nyquist criteria are no longer satisfied, resulting in blurring artifacts and loss of both spatial and temporal resolution[30,31]. This loss of information is compounded if spatial Nyquist criteria are also unmet, that is when subsampling pixels, as is the case in LR acquisitions. However, although simple frame averaging approaches may lose resolution in exchange for higher SNR, more sophisticated computational approaches can take advantage of multi-frame acquisitions to increase the resolution of individual frame reconstructions[32–35].

We hypothesized that the PSSR network could learn the additional information contained in sequential video frames, even when grossly undersampled in both space and time, and could thus be used to reduce flickering artifacts, while also improving the restoration accuracy and resolution of PSSR-processed time-lapse videos. To test this hypothesis, we exploited the multi-dimensional capabilities of our PSSR Res-U-Net architecture by training on five sequential timepoint inputs for each single timepoint output (multi-frame PSSR, or PSSR-MF, Fig. 4a). As measured by PSNR, SSIM, FRC and NanoJ-SQUIRREL error mapping, as well as

Fig. 3 | PSSR model is effective for multiple EM modalities and sample types. a–d, Shown are representative LR, LR-bilinear and PSSR-restored (LR-PSSR) images from mouse brain sections ($n=75$ sections in one image stack, *xy* dimension 240×240 pixels) imaged with a Zeiss Sigma VP Gatan SBSEM system (**a**), fly sections ($n=50$ sections in one image stack, *xy* dimension $1{,}250\times1{,}250$ pixels) acquired with Zeiss/FEI FIB–SEM (**b**), mouse sections ($n=563$ sections in one image stack, *xy* dimension 240×240 pixels) from Zeiss/FEI FIB–SEM (**c**) and rat sections (one montage, *xy* dimension $2{,}048\times1{,}024$ pixels) imaged with a Hitachi Regulus ssEM (**d**). **e**, Validation of presynaptic vesicle detection. LR, LR-bilinear, LR-PSSR and ground truth HR images of a representative bouton region as well as their color-labeled vesicle counts are shown. Vesicles colored magenta represent false negatives, green are false positives and white are true positives. Docked vesicles are outlined in yellow. The percentage of each error type is shown in the pie chart. Vesicle counts from two humans were plotted (dashed line, Human-1; solid line, Human-2), with the average total error ± s.e.m. displayed above. Experiments were conducted with $n=10$ independent bouton regions in all conditions, achieving similar results. The linear regression between LR-bilinear and HR, LR-PSSR and HR, and two human counters of HR are shown in the third row. The equation for the linear regression, the goodness-of-fit (R^2) and the *P* value of each graph are displayed. Scale bars, 1.5 µm.

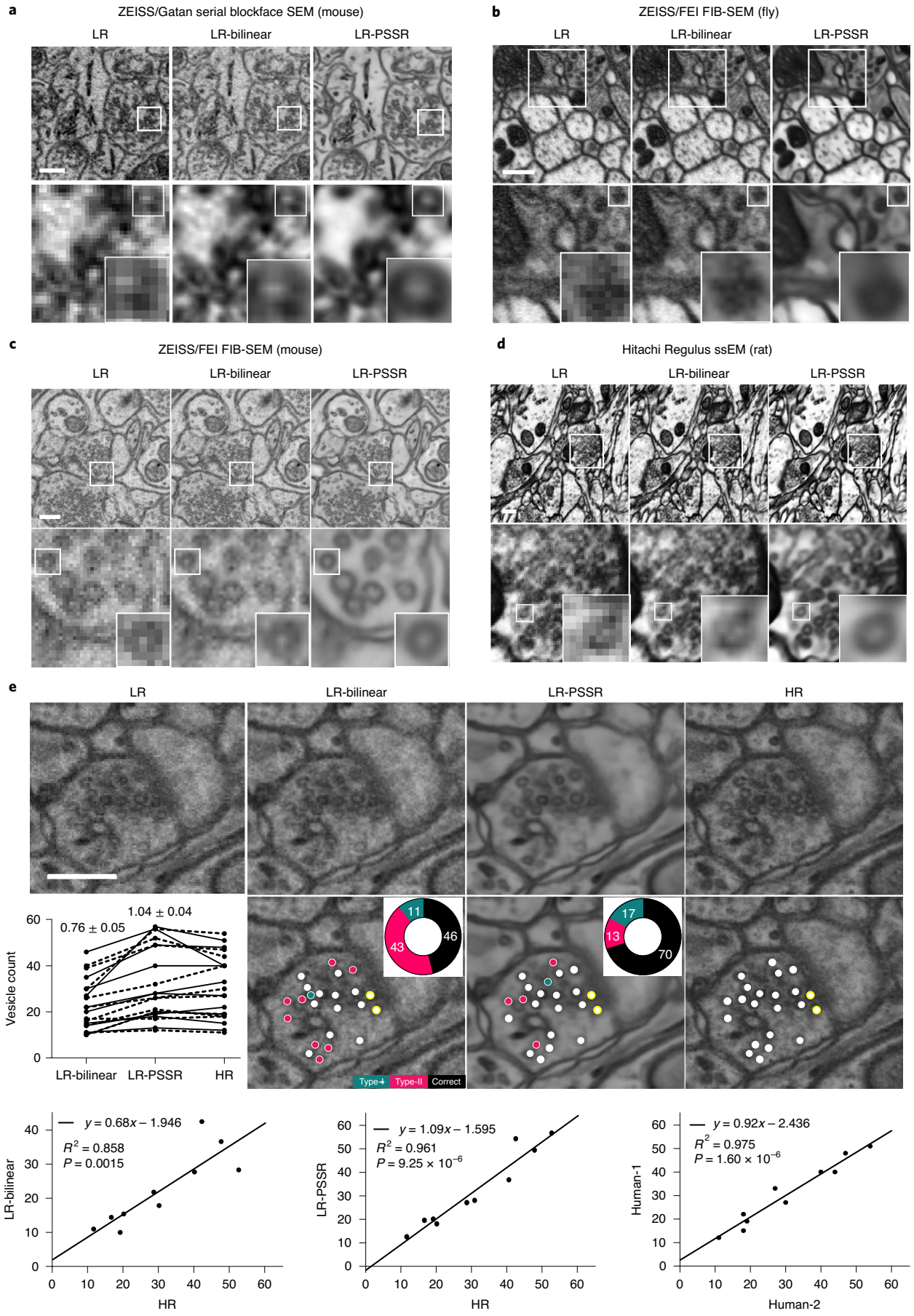

a
ZEISS/Gatan serial blockface SEM (mouse)
LR
LR-bilinear
LR-PSSR
b
ZEISS/FEI FIB-SEM (fly)
LR
LR-bilinear
LR-PSSR
c
ZEISS/FEI FIB-SEM (mouse)
LR
LR-bilinear
LR-PSSR
d
Hitachi Regulus ssEM (rat)
LR
LR-bilinear
LR-PSSR
e
LR
LR-bilinear
LR-PSSR
HR
1.04 ± 0.04
0.76 ± 0.05
Vesicle count
LR-bilinear
LR-PSSR
HR
11
46
43
17
13
70
Type-I
Type-II
Correct
y = 0.68x – 1.946
R2 = 0.858
P = 0.0015
LR-bilinear
HR
y = 1.09x – 1.595
R2 = 0.961
P = 9.25 × 10−6
LR-PSSR
HR
y = 0.92x – 2.436
R2 = 0.975
P = 1.60 × 10−6
Human-1
Human-2

Table 1 | Details of fluorescence PSSR training experiments

Experiment	U2OS MitoTracker PSSR-SF	U2OS MitoTracker PSSR-MF	Neuron Mito-dsRed PSSR-MF
Input size (*x*, *y*, *z*)	(128, 128, 1)	(128, 128, 5)	(128, 128, 5)
Output size (*x*, *y*, *z*)	(512, 512, 1)	(512, 512, 1)	(512, 512, 1)
No. of image pairs for training	5,000	5,000	3,000
No. of image pairs for validation	200	200	300
No. of GPUs	2	2	2
Training datasource size (GB)	9.4	9.4	9.5
Validation datasource size (GB)	0.44	0.44	1.8
Training dataset size (GB)	1.38	1.7	1.03
Validation dataset size (GB)	0.06	0.07	0.1
Batch size per GPU	8	8	6
No. of epochs	100	100	50
Training time (h)	3.76	3.77	1.25
Learning rate	4×10^{-4}	4×10^{-4}	4×10^{-4}
Best model found at epoch	33	86	39
Normalized to ImageNet statistics?	No	No	No
ResNet size	ResNet34	ResNet34	ResNet34
Loss function	MSE	MSE	MSE

compared with the BM3D denoising algorithm, PSSR-MF processing of LR acquisitions (LR-PSSR-MF) showed significantly increased resolution and SNR compared to the raw input (LR), 16× bilinear interpolated input (LR-bilinear) and single-frame PSSR (LR-PSSR-SF) (Fig. 4d,f and Extended Data Figs. 7 and 8). Since it is conceivable that simply averaging five frames would yield similar improvements due to reduced flickering and increased SNR, we tested whether using a five-frame rolling average over the LR-PSSR-SF output (LR-PSSR-SF-RA) could yield similar results to LR-PSSR-MF. Although LR-PSSR-SF-RA output displayed reduced flickering (Fig. 4b,c), we found that LR-PSSR-SF-RA performed significantly worse than LR-PSSR-MF in terms of both resolution and accuracy (Fig. 4c, Extended Data Fig. 6 and Supplementary Video 6). We also compared PSSR models with content-aware image restoration (CARE), a gold standard DL-based image restoration algorithm (Fig. 4b,c, Extended Data Fig. 6 and Methods). Specifically, CARE trained results (LR-CARE) and its rolling average postprocessed version (LR-CARE-RA) were compared with LR-PSSR-MF, LR-PSSR-SF and its rolling average postprocessed results (LR-PSSR-SF-RA). As expected, the LR-CARE and LR-PSSR-SF results were similar. However, we found the multi-frame PSSR approach yielded better results than CARE, both before and after the rolling average processing (Extended Data Fig. 6). We therefore concluded multi-frame PSSR significantly enhances the fidelity and resolution of LR images beyond any standard frame-by-frame image restoration approach.

For all time-lapse PSSR we used PSSR-MF and refer to it as PSSR for the remainder of this article. The improved speed, resolution and SNR enabled us to detect mitochondrial fission events that were not detectable in the LR or LR-bilinear images (yellow arrows, Fig. 4e and Supplementary Video 7). Additionally, the relatively high laser dose during HR acquisitions raises questions as to whether observed fission events are artifacts of phototoxicity. We validated the accuracy of our fission event detection with semisynthetic data quantified by two expert humans (Fig. 4g,h). We found a notable improvement in detecting fission events with relatively minor increases in false positives (Fig. 4h–k). We again found the variance between the PSSR and HR results was similar to the variance between the two expert human results on HR data. Thus, our PSSR model provides an opportunity to detect very fast mitochondrial fission events with fewer phototoxicity-induced artifacts than standard HR Airyscan imaging using normal confocal optics and detectors.

As mentioned above, in addition to phototoxicity issues, the slow speed of HR scanning confocal imaging often results in temporal undersampling of fast-moving structures such as motile mitochondria in neurons (Supplementary Fig. 1 and Supplementary Videos 8 and 9). However, relatively fast LR scans do not provide sufficient pixel resolution or SNR to resolve fission or fusion events, or individual mitochondria when they pass one another along a neuronal process, which can result in faulty analysis or data interpretation (Supplementary Video 8). Thus, we next tested whether PSSR provided sufficient restoration of undersampled time-lapse imaging of mitochondrial trafficking in neurons.

As measured by PSNR, SSIM, FRC and NanoJ-SQUIRREL error mapping, the overall resolution and SNR improvement provided by PSSR enabled us to resolve adjacent mitochondria as well as fission and fusion events (Fig. 5a–c, Extended Data Figs. 9 and 10 and Supplementary Video 9). Since our LR acquisition rates are 16× faster than HR, instantaneous motility details were preserved in LR-PSSR whereas in HR images they were lost (Fig. 5d, Supplementary Fig. 1 and Supplementary Video 9). The overall total distance mitochondria traveled in neuronal processes was the same for both LR and HR (Fig. 5f). However, we were able to obtain unique information about how they translocate when

Table 2 | Details of fluorescence PSSR testing data for PSNR, SSIM and error mapping

	U2OS MitoTracker			Neuron Mito-dsRed		
	Semisynthetic	Real world		Semisynthetic	Real world	
	HR	LR	HR	HR	LR	HR
Microscopy	Airyscan	Confocal	Airyscan	Airyscan	Confocal	Airyscan
Laser power (μW)	35	7	28	82	11	55
Datasource size (MB)	1,250	7.15	200	5,920	10.7	305
Dataset size (MB)	325	6.39	191	1,970	10.4	305
Total number of different cells	6	10	10	7	10	10

imaging at a 16× higher frame rate (Fig. 5g). A larger range of velocities was identified in LR-PSSR than both LR and HR images. Overall, LR-PSSR and HR provided similar values for the percentage of time mitochondria spent in motion (Fig. 5h). Smaller distances traveled were easier to detect in our LR-PSSR images, and therefore there was an overall reduction in the percentage of time mitochondria spent in the paused position in our LR-PSSR data (Fig. 5i). Taken together, these data show PSSR provides a means to detect critical biological events that would not be possible with conventional HR or LR imaging.

Discussion

We have demonstrated DL-based pixel super-resolution can be a viable strategy of particular value for both optical and electron point-scanning microscopes. Acquiring suitably aligned pairs of high- and low-quality images for training is incredibly expensive and difficult (Fig. 1f). Thus, we introduced a new 'crappifier' for generating noisy, LR training data from HR, high SNR ground truth data. This enables the use of large, pre-existing gold standard datasets for training new models without acquiring any new data. We hope the open-source availability of our crappifier will be reciprocated by open-source sharing of high-quality imaging data, which can then be used to train new DL models. We did not fully explore all possibilities for a crappifier and believe this is an open and fruitful area for future studies.

We discovered that DL-based restoration of noisy time-lapses suffers from temporal inconsistency (flickering) artifacts due to noise-induced randomness in pixel values between frames. To address this, we introduced a multi-frame super-resolution approach that leverages the information in previous and future timepoints to better infer the frame of interest. We found this multi-frame approach not only reduces flickering artifacts, but also provides better overall image restoration for each independent frame. Notably, this approach would be impossible without the crappifier, which provides the ability to generate training data from videos of rapidly moving structures in live cells for which it is impossible to acquire perfectly aligned image pairs.

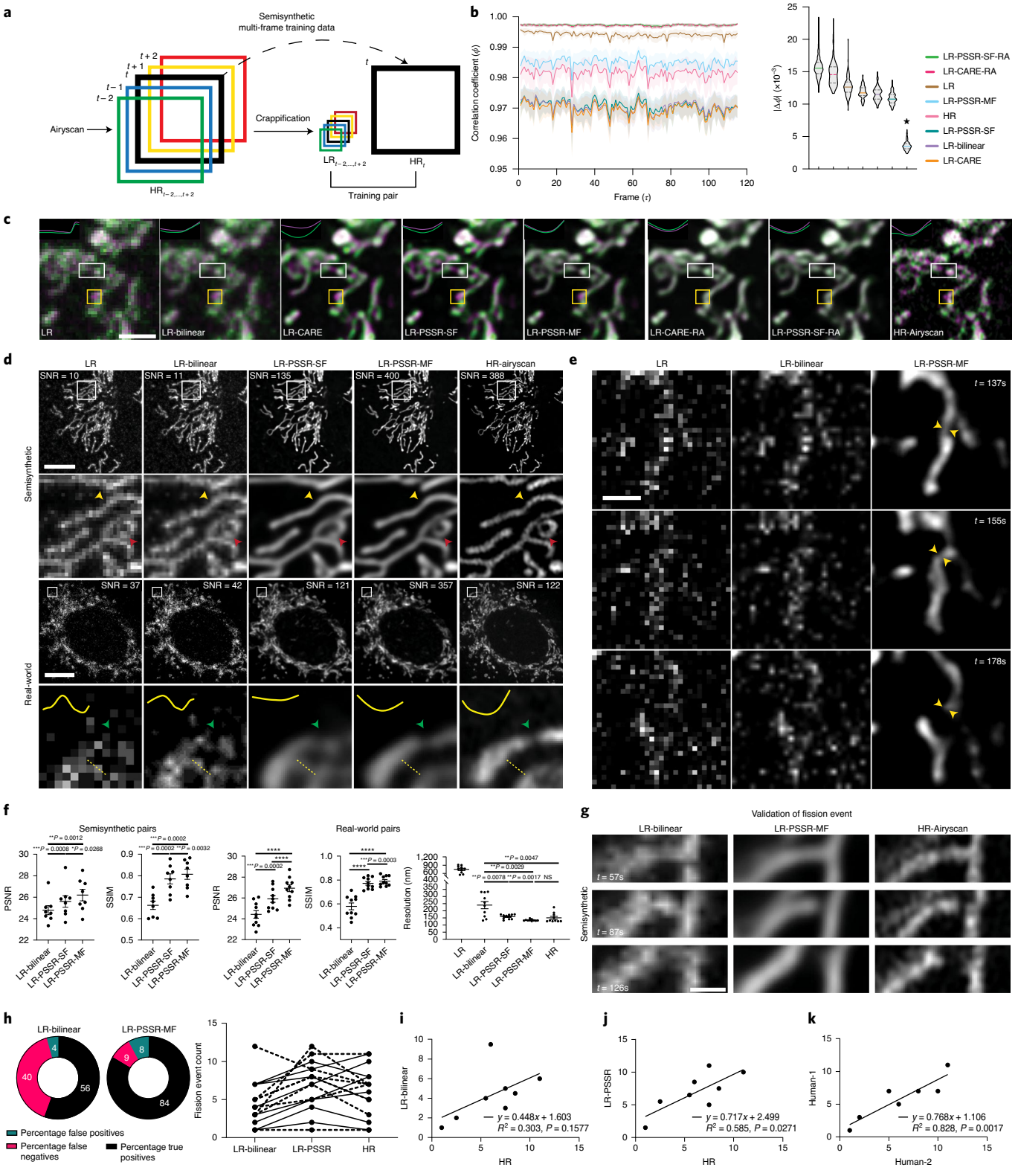
Any output from a DL super-resolution model is a prediction, is never 100% accurate and is always highly dependent on sufficient correspondence between the training versus experimental data[5,8,28,36]. Whether the level of accuracy of a given model for a given dataset is satisfactory is ultimately dependent on the tolerance for error in the measurement being made. For example, our EM PSSR model was validated by segmenting vesicles in presynaptic boutons. But we did not rule out the possibility that other structures or regions in the same sample may not be restored by our model with the necessary accuracy for any arbitrary measurement. Thus, it is essential to validate the accuracy of the model for the specific task at hand before investing further. Similarly, we observed that no single performance metric reliably captures the 'best' model. Thus, model performance must be evaluated by a combination of metrics, segmentation and, of course, visual inspection by human experts. An important future direction may be to develop better metrics for evaluating models.

Although the accuracy of DL approaches such as PSSR is technically imperfect, real-world limitations on acquiring ground truth data may render PSSR the best option. Our results show the PSSR approach can in principle enable higher speed and resolution imaging with the fidelity necessary for biological research. The ability to use DL to supersample undersampled images provides an opportunity that extends to any point-scanning system, including ion-based imaging systems[37,38] or HR cryoSTEM[39].

Fig. 4 | Multi-frame PSSR time-lapses of mitochondrial dynamics. a, Overview of multi-frame PSSR training data generation method. Five consecutive frames ($HR_i, i \in [t-2, t+2]$) from a HR Airyscan time-lapse video were synthetically crappified to five LR images ($LR_i, i \in [t-2, t+2]$), which together with the HR middle frame at time t (HR_t), form a five-to-one training 'pair'. **b**, Temporal consistency analysis. Neighboring-frame cross-correlation coefficient ($\emptyset(X_\tau, X_{\tau+1})$) that corresponds to frame τ on the x axis denotes the correlation coefficient of frame τ (X_τ) and frame $\tau+1$ ($X_{\tau+1}$) (left). Absolute error against HR ($|\Delta\emptyset|$) for each condition was compared ($|\Delta\emptyset| = \left|\emptyset_\tau - \emptyset_\tau^{HR}\right|$, right). $n=6$ independent time-lapses with $n=80$–120 timepoints each. Colored shades show standard error. The asterisk above LR-PSSR-MF denotes that LR-PSSR-MF is significantly more consistent with HR than all other conditions ($P<0.0001$). All violin plots show lines at the median and quartiles. **c**, Examples of false mitochondrial network merges (white boxes) due to the severe flickering artifacts in single-frame models (LR-bilinear, LR-CARE and LR-PSSR-SF), and loss of temporal consistency and resolution (yellow boxes) in models postprocessed with a 'rolling frame averaging' method (LR-CARE-RA and LR-PSSR-SF-RA). Two consecutive frames of an example region from semisyntheticly acquired LR, LR-bilinear, CARE (LR-CARE), five-frame rolling average postprocessed CARE output (LR-CARE-RA), LR-PSSR-SF, single-frame PSSR postprocessed with a five-frame rolling average (LR-PSSR-RA), five-frame multi-frame PSSR (LR-PSSR-MF) and ground truth HR (Airyscan) time-lapses are color coded in magenta ($t=0$ s) and green ($t=5$ s). Insets show the intensity line plot of the two frames drawn in the center of the white box in each condition. The yellow box shows an example of temporal resolution loss in rolling average conditions (LR-CARE-RA and LR-PSSR-SF-RA) only. Magenta pixels represent signal that only exists in the $t=0$ s frame, but not in the $t=5$ s, while green pixels represent signal present only in the $t=5$ s frame. Scale bar, 2 µm. **d**, Restoration performance on semisynthetic and real-world testing pairs. For the semisynthetic pair, LR was synthetically generated from Airyscan HR videos. Enlarged ROI show an example of well resolved mitochondrial structures by PSSR, agreeing with Airyscan ground truth images. Red and yellow arrowheads show two false connecting points in LR-bilinear and LR-PSSR-SF, which were well separated in LR-PSSR-MF. In the real-world example, green arrowheads in the enlarged ROI highlight a well restored gap between two mitochondria segments in the LR-PSSR-MF output. Normalized line-plot cross-section profile (yellow) highlights false bridging between two neighboring structures in LR-bilinear and LR-PSSR-SF, which was well separated with our PSSR-MF model. SNR measured using the images in both semisynthetic and real-world examples are indicated. Scale bars, 10 µm. **e**, PSSR output captured a transient mitochondrial fission event. Shown is a PSSR-restored dynamic mitochondrial fission event, with three key time frames displayed. Arrows highlight the mitochondrial fission site. Scale bar, 150 nm. **f**, PSNR and SSIM quantification of the semisynthetic ($n=8$ independent time-lapses with $n=80$–120 timepoints each) as well as the real world ($n=10$ independent time-lapses of fixed samples with $n=10$ timepoints each) testing sets discussed in **d**. FRC values measured using two independent LR versus HR acquisitions from multiple cells are indicated ($n=10$). **g**, Validation of fission event captures using semisynthetic data. An example of a fission event that was detectable in LR-PSSR but not LR-bilinear. Experiments were repeated with eight time-lapses, achieving similar results. Scale bar, 1 µm. **h**, For fission event detection, the number of false positives, false negatives and true positives detected by expert humans was quantified for eight different time-lapses. Distribution was shown in the pie charts. Fission event counts from two humans were plotted (dashed line, Human-1; solid line, Human-2). **i**–**k**, Linear regression between LR-bilinear and HR, LR-PSSR and HR, and two human counters of HR are shown ($n=8$ independent time-lapses with $n=80$–120 timepoints each). The linear regression equation, the goodness-of-fit (R^2) and the P value of each graph are displayed. All values are shown as mean ± s.e.m. P values are specified in the figure for $0.0001<P<0.05$. $^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, $^{****}P<0.0001$; NS, not significant; two-sided paired t-test.

For future uses of PSSR, we propose an acquisition scheme wherein a relatively limited number of 'ground truth' HR images are acquired for fine-tuning pretrained models. More importantly, the performance of generalized, unsupervised or 'self-supervised' denoising approaches[7,9] as well as DL-enabled deconvolution approaches[19,20] suggests we may one day be able to generate a more generalized model for a specific imaging system, instead of a specific sample type.

Structured illumination microscopy, single-molecule localization microscopy and pixel reassignment microscopy demonstrate the power of configuring optical imaging schemes with a specific postprocessing computational strategy in mind. The power of deep convolutional neural networks for postprocessing image data presents a new opportunity for redesigning imaging systems to exploit these capabilities to minimize costs traditionally considered necessary for extracting meaningful imaging data. Similarly,



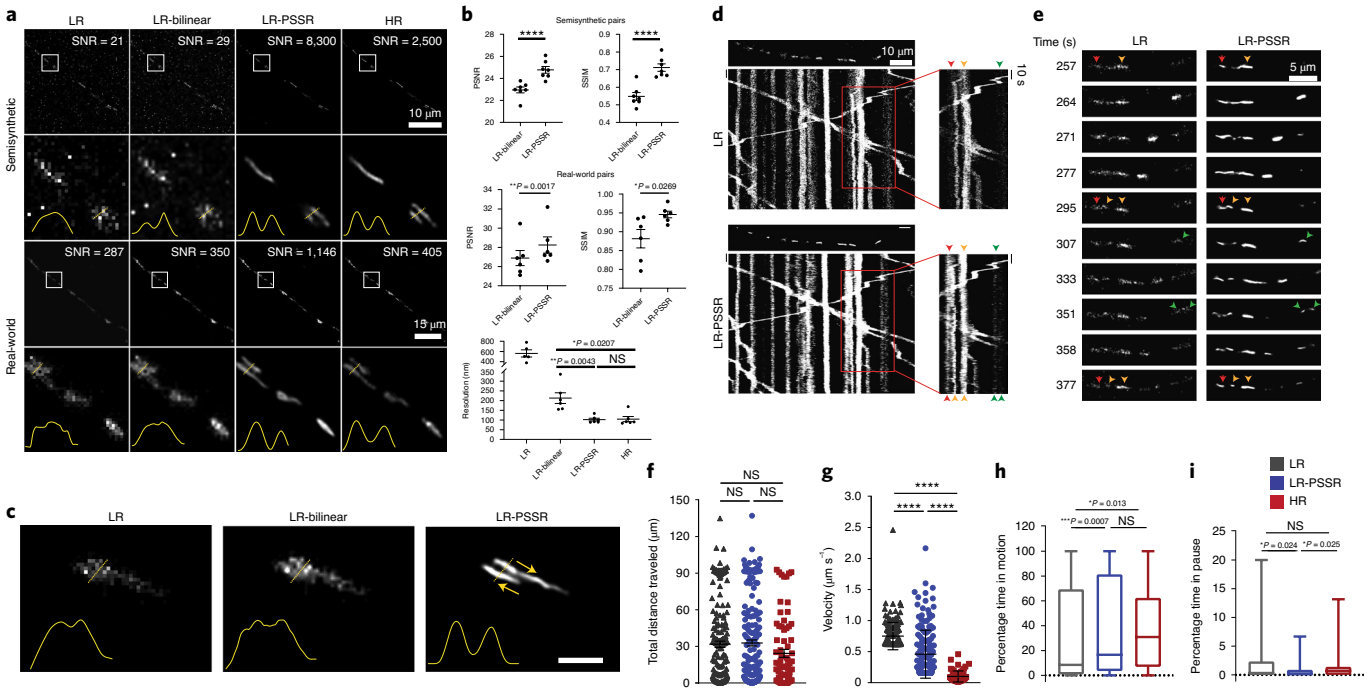


Fig. 5 | Spatiotemporal analysis of mitochondrial motility in neurons. PSSR facilitates high-spatiotemporal resolution imaging of mitochondrial motility in neurons. **a**, Comparison of LR-PSSR results versus LR-bilinear interpolation on semisynthetic ($n = 7$ independent time-lapse videos with $n = 100$ independent timepoints each) and real-world testing pairs ($n = 6$ independent time-lapse videos with $n = 12$ independent timepoints each). Enlarged ROI from representative images show PSSR resolved two mitochondria in both semisynthetic and real-world testing sets, quantified by normalized line-plot cross-section profiles. Scale bars, 10 µm (upper) and 15 µm (lower). **b**, PSNR (top) and SSIM (middle) quantification of the datasets in **a**. FRC resolution measured from two independent acquisitions of the real-world overview dataset discussed in **a** is indicated (bottom). **c**, PSSR restoration of LR time-lapses resolves mitochondria moving past one another in a neuronal process (arrows indicate direction of movement). Scale bar, 2 µm. **d**, Representative kymographs of mitochondrial motility in hippocampal neurons transfected with Mito-DsRed ($n = 7$ independent LR time-lapse videos processed to LR-PSSR). First frame of each time-lapse video is shown above each kymograph. Different color arrowheads indicate mitochondria going through fission and fusion events. Each color represents a different mitochondrion. Scale bar, 10 µm. **e**, Enlarged areas of **d**, capturing mitochondrial fission and fusion events in real-time. Scale bar, 5 µm. **f**–**i**, Mitochondrial motility was quantified from time-lapse videos as demonstrated in Supplementary Video 8. For each mitochondrial trajectory the total distance mitochondria traveled (**f**), mitochondrial velocity (**g**), percentage time mitochondria spent in motion (**h**) and in pause (**i**) was quantified ($n = 76 - 216$ mitochondria from four neurons and three independent experiments). All values are shown as mean ± s.e.m. *P* values are specified in the figure for $0.0001 < P < 0.05$. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; NS, not significant; two-sided paired *t*-test (**b**) and Kruskal–Wallis test followed by Dunn's multiple comparison test (**f**–**i**).

automated real-time corrections to the images and real-time feedback to the imaging hardware are now within reach. This is an exciting area of active investigation in our laboratory and others[40].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-021-01080-z.

Received: 22 October 2019; Accepted: 28 January 2021;
Published online: 08 March 2021

References

1. Wang, Z., Chen, J. & Hoi, S. C. H. Deep learning for image super-resolution: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* https://doi.org/10.1109/TPAMI.2020.2982166 (2019).
2. Jain, V. et al. Supervised learning of image restoration with convolutional networks. In *2007 IEEE 11th International Conference on Computer Vision* 1–8 (IEEE, 2007).
3. Romano, Y., Isidoro, J. & Milanfar, P. RAISR: Rapid and Accurate Image Super Resolution. *IEEE Trans. Comput. Imaging* **3**, 110–125 (2016).
4. Shrivastava, A. et al. Learning from simulated and unsupervised images through adversarial training. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2107–2116 (IEEE, 2017).
5. Moen, E. et al. Deep learning for cellular image analysis. *Nat. Methods* https://doi.org/10.1038/s41592-019-0403-1 (2019).
6. Buchholz, T.-O., Jordan, M., Pigino, G. & Jug, F. Cryo-CARE: Content-Aware Image Restoration for Cryo-Transmission Electron Microscopy Data. In *Proc. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 502–506 (IEEE, 2019).
7. Krull, A., Buchholz, T.-O. & Jug, F. Noise2void-learning denoising from single noisy images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2129–2137 (IEEE, 2019).
8. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090–1097 (2018).
9. Batson, J. & Royer, L. Noise2self: blind denoising by self-supervision. *Proc. 36th Int. Conf. Machine Learning, PMLR 97* 524–533 (2019).
10. Li, Y. et al. DLBI: deep learning guided Bayesian inference for structure reconstruction of super-resolution fluorescence microscopy. *Bioinformatics* **34**, i284–i294 (2018).
11. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol.* **36**, 460–468 (2018).
12. Nelson, A. J. & Hess, S. T. Molecular imaging with neural training of identification algorithm (neural network localization identification). *Microsc. Res. Tech.* **81**, 966–972 (2018).
13. Buchholz, T. O. et al. Content-aware image restoration for electron microscopy. *Methods Cell. Biol.* **152**, 277–289 (2019).

14. Heinrich, L., et al. Deep learning for isotropic super-resolution from non-isotropic 3D electron microscopy. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017* 135–143 (Springer, 2017).
15. de Haan, K., Ballard, Z. S., Rivenson, Y., Wu, Y. & Ozcan, A. Resolution enhancement in scanning electron microscopy using deep learning. *Sci. Rep.* **9**, 12050 (2019).
16. Sreehari, S. et al. Multi-resolution data fusion for super-resolution electron microscopy. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR) Workshops* 88–96 (IEEE, 2017).
17. Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).
18. Chen, J. et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.27.270439 (2020).
19. Guo, M. et al. Rapid image deconvolution and multiview fusion for optical microscopy. *Nat. Biotechnol.* **38**, https://doi.org/10.1038/s41587-020-0560-x (2020).
20. Kobayashi, H., Solak, A. C., Batson, J. & Royer, L. A. Image deconvolution via noise-tolerant self-supervised inversion. Preprint at https://arxiv.org/abs/2006.06156 (2020).
21. Horstmann, H., Korber, C., Satzler, K., Aydin, D. & Kuner, T. Serial section scanning electron microscopy (S3EM) on silicon wafers for ultra-structural volume imaging of cells and tissues. *PLoS ONE* **7**, e35172 (2012).
22. Xu, C. S. et al. Enhanced FIB–SEM systems for large-volume 3D imaging. *eLife* https://doi.org/10.7554/eLife.25916 (2017).
23. Denk, W. & Horstmann, H. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* **2**, e329 (2004).
24. Kuwajima, M., Mendenhall, J. M., Lindsey, L. F. & Harris, K. M. Automated transmission-mode scanning electron microscopy (tSEM) for large volume analysis at nanoscale resolution. *PLoS ONE* **8**, e59573 (2013).
25. Culley, S. et al. Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nat. Methods* **15**, 263–266 (2018).
26. Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. Image denoising with block-matching and 3D filtering. In *Proc. SPIE 6064, Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning* 606414 (2006).
27. von Chamier, L., Laine, R. F. & Henriques, R. Artificial intelligence for microscopy: what you should know. *Biochem Soc. Trans.* https://doi.org/10.1042/BST20180391 (2019).
28. Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods* https://doi.org/10.1038/s41592-019-0458-z (2019).
29. Phototoxicity revisited. *Nat. Methods* **15**, 751 (2018).
30. Jonkman, J. & Brown, C. M. Any way you slice it—a comparison of confocal microscopy techniques. *J. Biomol. Tech.* **26**, 54–65 (2015).
31. Kner, P., Chhun, B. B., Griffis, E. R., Winoto, L. & Gustafsson, M. G. Super-resolution video microscopy of live cells by structured illumination. *Nat. Methods* **6**, 339–342 (2009).
32. Wronski, B. et al. Handheld multi-frame super-resolution. *ACM Trans. Graph.* **38**, 28 (2019).
33. Huang, X. et al. Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy. *Nat. Biotechnol.* **36**, 451–459 (2018).
34. Carlton, P. M. et al. Fast live simultaneous multiwavelength four-dimensional optical microscopy. *Proc. Natl Acad. Sci. USA* **107**, 16016–16022 (2010).
35. Arigovindan, M. et al. High-resolution restoration of 3D structures from widefield images with extreme low signal-to-noise-ratio. *Proc. Natl Acad. Sci. USA* **110**, 17344–17349 (2013).
36. Barbastathis, G., Ozcan, A. & Situ, G. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).
37. Keren, L. et al. MIBI-TOF: a multiplexed imaging platform relates cellular phenotypes and tissue structure. *Sci. Adv.* **5**, eaax5851 (2019).
38. Arrojo, E. D. R. et al. Age mosaicism across multiple scales in adult tissues. *Cell Metab.* **30**, 343–351 e343 (2019).
39. Wolf, S. G. & Elbaum, M. CryoSTEM tomography in biology. *Methods Cell. Biol.* **152**, 197–215 (2019).
40. Mi, L. et al. Learning guided electron microscopy with active acquisition. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020* (eds Martel, A. L. et al.) (Springer, 2020); https://doi.org/10.1007/978-3-030-59722-1_8

Methods

Semisynthetic training image generation. HR images were acquired using scanning electron or Airyscan confocal microscopes. Due to the variance of image properties (for example, format, size, dynamic range and depth) in the acquired HR images, data cleaning is indispensable for generating training sets that can be easily accessed during training. In this article, we differentiate the concept of 'data sources' and 'datasets', where data sources refer to uncleaned acquired HR images and datasets refer to images that are generated and preprocessed from data sources. HR datasets were obtained after preprocessing HR images from data sources, LR datasets were generated from HR datasets using a 'crappifier' function.

Preprocessing. Tiles of predefined sizes (for example, 256×256 and 512×512 pixels) were randomly cropped from each frame in image stacks from HR data sources. 'Refection padding' was used if the image size in the data sources is smaller than the predefined tile size. All tiles were saved as separate images in .tif format, which together formed a HR dataset.

Image crappification. A 'crappifier' was then used to synthetically degrade the HR datasets to LR images, with the goal of approximating the undesired and unavoidable pixel intensity variation in real-world LR and low SNR images of the same field of view directly taken under an imaging system. These HR images together with their corrupted counterparts served as training pairs to facilitate 'deCrappification'. The crappification function can be simple, but it materially improves both the quality and characteristics of PSSR outputs.

Image sets were normalized from 0 to 1 before being 16× downsampled in pixel resolution (for example, a 1,000×1,000 pixel image would be downsampled to 250×250 pixels). To mimic the image quality degradation caused by 16× undersampling on a real-world point-scanning imaging system, Salt&Pepper noise and Gaussian additive noise with specified local variance were randomly injected into the HR images. The degraded images were then rescaled to 8-bit for viewing with normal image analysis software.

EM crappifier. Random Gaussian-distributed additive noise ($\mu_{EM}=0$, $\sigma_{EM}=3$) was injected. The degraded images were then downsampled using spline interpolation of order 1.

MitoTracker and neuronal mitochondria crappifier. The crappification of MitoTracker and neuronal mitochondria data followed a similar procedure. Salt&Pepper noise was randomly injected in 0.5% of each image's pixels replacing them with noise, which was followed by the injection of random Gaussian-distributed additive noise ($\mu_{LSM}=0$, $\sigma_{LSM}=5$). The crappified images were then downsampled using spline interpolation of order 1.

Data augmentation. After crappified LR images were generated, we used data augmentation techniques such as random cropping, dihedral affine function, rotation, random zoom to increase the variety and size of our training data[41].

Multi-frame training pairs. Unlike imaging data of fixed samples, where we use traditional one-to-one HR and LR images as training pairs, for time-lapse videos, five consecutive frames ($HR_i, i \in [t-2, t+2]$) from a HR Airyscan time-lapse video were synthetically crappified to five LR images ($LR_i, i \in [t-2, t+2]$), which together with the HR middle frame at time t (HR_t), form a five-to-one training 'pair'. The design of five-to-one training 'pairs' leverages the spatiotemporal continuity of dynamic biological behaviors (Fig. 4a).

Crappifier comparison. *EM crappifier comparison.* Four crappifiers including 'no noise', 'Poisson', 'Gaussian' and 'additive Gaussian' were used to generate semisynthetic training pairs from the same set of HR SEM images. The 'no noise' crappifier simply downsampled HR image pixel sizes by a factor of 16× (4×4) without adding any noise, while the Poisson, Gaussian and additive-Gaussian crappifiers added random Poisson noise, random Gaussian noise ($\mu_{EM}=0$, $\sigma_{EM}=0.15$) and random Gaussian-distributed additive noise ($\mu_{EM}=0$, $\sigma_{EM}=3$), respectively, before applying pixel downsampling. The 'additive Gaussian (roughly 80×)' used the same crappifier as additive Gaussian, but with roughly 80× more training data. We also compared the models described above with 'real world', a model trained with real-world pairs, whose HR images are the same as the HR images of the semisynthetically generated training pairs, but whose LR images were manually acquired at the microscope.

MitoTracker crappifier comparison. Five crappifiers including no noise, Salt&Pepper, Gaussian, additive Gaussian and 'Salt&Pepper + additive Gaussian' were used to generate semisynthetic training pairs from the same set of HR Airyscan MitoTracker time-lapse videos. The no noise crappifier downsampled HR image pixel sizes by a factor of 16× (4×4) without adding any noise, while the Salt&Pepper, Gaussian, additive Gaussian and Salt&Pepper + additive-Gaussian crappifiers added random Salt&Pepper noise (0.5%), random Gaussian noise ($\mu_{LSM}=0$, $\sigma_{LSM}=0.025$), random Gaussian-distributed additive noise ($\mu_{LSM}=0$, $\sigma_{LSM}=5$) and the combination of Salt&Pepper and additive Gaussian, respectively, before the bilinear downsampling.

Neural networks. *Single-frame neural network (PSSR-SF).* A ResNet-based U-Net was used as our convolutional neural network for training[42]. Our U-Net is in the form of encoder-decoder with skip-connections, where the encoder gradually downsizes an input image, followed by the decoder upsampling the image back to its original size. For the EM data, we used ResNet pretrained on ImageNet as the encoder. For the design of the decoder, the traditional handcrafted bicubic upscaling filters are replaced with learnable subpixel convolutional layers[43], which can be trained specifically for upsampling each feature map optimized in LR parameter space. This upsampling layer design enables better performance and largely reduces computational complexity, but at the same time causes unignorable checkerboard artifacts due to the periodic time-variant property of multirate upsampling filters[44]. A blurring technique[45] and a weight initialization method, known as subpixel convolution initialized to convolution neural network resize[46] that was designed for the subpixel convolution upsampling layers, were implemented to remove checkerboard artifacts. In detail, the blurring approach introduces an interpolation kernel of the zero-order hold with the scaling factor after each upsampling layer, the output of which gives out a nonperiodic steady-state value, which satisfies a critical condition ensuring a checkerboard artifact-free upsampling scheme[45]. Compared to random initialization, in addition to the benefit of removing checkerboard artifacts, subpixel convolution initialized to convolution neural network resize also empowers the model with higher modeling power and higher accuracy[46]. A self-attention layer inspired by Zhang et al.[47] was added after each convolutional layer.

Multi-frame neural network (PSSR-MF). A similar yet slightly modified U-Net was used for time-lapse video training. The input layer was redesigned to take five frames simultaneously while the last layer still produced one frame as output.

Training details. *Loss function.* MSE loss was used as our loss function.

Optimization methods. Stochastic gradient descent with restarts (SGDR)[48] was implemented. Aside from the benefits we are able to get through classic stochastic gradient descent, this method resets the learning rate to its initial value at the beginning of each training epoch and allows it to decrease again following the shape of a cosine function, yielding lower loss with higher computational efficiency.

Cyclic learning rate and momentum. Instead of having a gradually decreasing learning rate as the training converges, we adopted cyclic learning rates[49], cycling between the upper and lower bounds, which helps oscillate toward a higher learning rate, thus avoiding saddle points in the hyper-dimensional training loss space. In addition, we followed The One Cycle Policy[50], which restricts the learning rate to only oscillate once between the upper and lower bounds. Specifically, the learning rate linearly increases from the lower bound to the upper bound as the momentum decreases from its upper bound to the lower bound linearly. In the second half of the cycle, the learning rate fits a cosine annealing decreasing from the upper bound to zero while the momentum increases from its lower bound to the upper bound following the same annealing. This training technique achieves superior regularization by preventing the network from overfitting during the middle of the learning process, as well as enabling super-convergence[51] by allowing large learning rates and adaptive momentum.

Progressive resizing (used for EM data only). Progressive resizing was applied during the training of the EM model. Training was executed in two rounds with HR images scaled to *xy* pixel sizes of 256×256 and 512×512 and LR images scaled to 64×64 and 128×128 progressively. The first round was initiated with an ImageNet pretrained ResU-Net, and the model trained from the first round served as the pretrained model for the second round. The intuition behind this is it quickly reduces the training loss by allowing the model to see lots of images at a small scale during the early stages of training. As the training progresses, the model focuses more on picking up high-frequency features reflected through fine details that are only stored within larger scale images. Therefore, features that are scale-variant can be recognized through the progressively resized learning at each scale.

Discriminative learning rates (used for EM data only). To better preserve the previously learned information, discriminative learning was applied during each round of training for the purpose of fine-tuning. At the first stage of training, only the parameters from the last layer were trainable after loading a pretrained model, which either came from a large-scaled trained publicly available model (that is, pretrained ImageNet), or from the previous round of training. The learning rate for this stage, lr_1, was fixed. Parameters from all layers were set as learnable in the second stage. A linearly spaced learning rate range lr_2 was applied. The learning rate gradually increased across the layers of the entire network architecture. The number of training epochs at each round is denoted as (N_1, N_2), where N_1 and N_2 denote the epoch number used at stage one and stage two separately:

EM training data			
Round of progressive resizing	1	2	3
Input size (*x*, *y*, *z*)	(32, 32, 3)	(64, 64, 3)	(128, 128, 3)
Output size (*x*, *y*, *z*)	(128, 128, 3)	(256, 256, 3)	(512, 512, 3)
Batch size per GPU	64	16	8
No. of epochs (N_1, N_2)	(1, 1)	(3, 3)	(3, 3)
Learning rate (lr_1, lr_2)	(1×10^{-3} (1×10^{-5}, 1×10^{-3}))	(1×10^{-3} (1×10^{-5}, 1×10^{-3}))	(1×10^{-3} (1×10^{-5}, 1×10^{-4}))
Training dataset size (GB)	3.9	15.56	62.24
Validation dataset size (GB)	0.96	3.89	15.56
Training datasource size (GB)	105		
Validation datasource size (GB)	26		
No. of image pairs for training	80,000		
No. of image pairs for validation	20,000		
Total training time (h)	16		
Normalized to ImageNet statistics?	Yes		
ResNet size	ResNet34		
Loss function	MSE		
No. of GPUs	2		

Best model preservation (used for fluorescence data only). Instead of saving the last model after training a fixed number of epochs, at the end of each training epoch, PSSR checks if the validation loss goes down compared to the loss from the previous epoch and will only update the best model when a lower loss is found. This technique ensures the best model will not be missed due to local loss fluctuation during the training.

Elimination of tiling artifacts. Testing images often need to be cropped into smaller tiles before being fed into our model due to the memory limit of graphic cards. This creates tiling edge artifacts around the edges of tiles when stitching them back to the original images. A Gaussian blur kernel ($\mu_{tile}=0$, $\sigma_{tile}=1$) was applied to a ten-pixel wide rectangle region centered in each tiling edge to eliminate the artifacts.

Technical specifications. Final models were generated using fast.ai v.1.0.55 library (https://github.com/fastai/fastai), PyTorch on two NVIDIA TITAN RTX graphical processing units (GPUs). Initial experiments were conducted using NVIDIA Tesla V100s, NVIDIA Quadro P6000s, NVIDIA Quadro M5000s, NVIDIA Titan Vs, NVIDIA GeForce GTX 1080s or NVIDIA GeForce RTX 2080Ti GPUs.

Evaluation metrics. *PSNR and SSIM quantification.* Two classic image quality metrics, PSNR and SSIM, known for their properties of pixel-level data fidelity and perceptual quality fidelity, correspondingly, were used for the quantification of our paired testing image sets.

PSNR is inversely correlated with MSE, numerically reflecting the pixel intensity difference between the reconstruction image and the ground truth image, but it is also famous for poor performance when it comes to estimating human perceptual quality. Instead of traditional error summation methods, SSIM is designed to consider distortion factors such as luminance distortion, contrast distortion and loss of correlation when interpreting image quality[52].

SNR quantification. SNR was quantified for LSM testing images (Figs. 4b and 5a) by

$$\mathrm{SNR} = \frac{I_{\mathrm{MAX}} - \mu_{\mathrm{bg}}}{\sigma_{\mathrm{bg}}}$$

where I_{MAX} represents the maximum intensity value in the image, μ_{bg} and σ_{bg} represent the mean and the standard deviation of the background, respectively[17].

FRC analysis. NanoJ-SQUIRREL[25] was used to calculate image resolution using FRC method on real-world testing examples with two independent acquisitions of fixed samples (Figs. 2b,c, 4c and 5b).

Resolution-scaled error and resolution-scaled Pearson's coefficient. NanoJ-SQUIRREL[25] was used to calculate the resolution-scaled error and resolution-scaled Pearson's coefficient for both semisynthetic and real-world acquired LR, LR-bilinear and PSSR (LR-PSSR) images versus ground truth HR images. Difference error maps were also calculated (Extended Data Figs. 2, 7 and 10).

EM imaging and analysis. *tSEM HR training data acquisition.* Tissue from a perfused 7-month-old Long–Evans male rat was cut from the left hemisphere stratum radiatum of CA1 of the hippocampus. The tissue was stained, embedded and sectioned at 45 nm using previously described protocols[53]. Sections were imaged using a scanning transmission EM detector on a Zeiss Supra 40 scanning electron microscope with a 28 kV accelerating voltage and an extractor current of 102 µA (gun aperture 30 µm). Images were acquired with a 2 nm pixel size and a field size of 24,576×24,576 pixels with Zeiss Atlas. The working distance from the specimen to the final lens was 3.7 mm, and the dwell time was 1.2 µs.

EM testing sample preparation and image acquisition. EM datasets were acquired from multiple systems at multiple institutions for this study.

For our testing ground truth data, paired LR and HR images of the adult mouse hippocampal dentate gyrus middle molecular layer neuropil were acquired from ultrathin sections (80 nm) collected on silicon chips and imaged in a Zeiss Sigma VP FE–SEM[21]. All animal work was approved by the Institutional Animal Care and Use Committee of the Salk Institute for Biological Studies. Samples were prepared for EM according the NCMIR protocol[54]. Pairs of 4×4 µm² images were collected from the same region at pixels sizes of both 8 and 2 nm using Fibics Atlas software (InLens detector, 3 kV; dwell time, 5.3 µs; line averaging, 2; aperture, 30 µm and working distance, 2 mm).

SBFSEM images were acquired with a Gatan 3View system installed on the Zeiss Sigma VP FE–SEM. Images were acquired using a pixel size of 8 nm on a Gatan backscatter detector at 1 kV and a current of 221 pA. The pixel dwell time was 2 µs with an aperture of 30 µm and a working distance of 6.81 mm. The section thickness was 100 nm and the field of view was 24.5×24.5 µm².

Mouse FIB–SEM data sample preparation and image acquisition settings were previously described in the original manuscript the datasets were published[22]. Briefly, the images were acquired with 4 nm voxel resolution. We downsampled the lateral resolution to 8 nm, then applied our PSSR model to the downsampled data to ensure the proper 8-to-2 nm transformation for which the PSSR was trained.

Fly FIB–SEM data sample preparation and image acquisition settings were previously described in the original manuscript the datasets were published[55]. Briefly, images were acquired with 10 nm voxel resolution. We first upsampled the *xy* resolution to 8 nm using bilinear interpolation, then applied our PSSR model to the upsampled data to ensure the proper 8-to-2-nm transformation for which the PSSR model was originally trained.

The rat SEM data sample was acquired from an 8-week old male Wistar rat that was anesthetized with an overdose of pentobarbital (75 mg kg^{-1}) and perfused through the heart with 5–10 ml of a solution of 250 mM sucrose 5 mM $MgCl_2$ in 0.02 M phosphate buffer (pH 7.4) followed by 200 ml of 4% paraformaldehyde containing 0.2% picric acid and 1% glutaraldehyde in 0.1 M phosphate buffer. Brains were then removed and oblique horizontal sections (50 µm thick) of frontal cortex/striatum were cut on a vibrating microtome (Leica VT1200S) along the line of the rhinal fissure. The tissue was stained and cut to 50-nm sections using ATUMtome (RMC Boeckeler) for SEM imaging using the protocol described in the original publication for which the data were acquired[56]. The Hitachi Regulus rat SEM data were acquired using a Regulus 8240 FE–SEM with an acceleration voltage of 1.5 kV, a dwell time of 3 µs, using the backscatter detector with a pixel resolution of 10×10 nm². We first upsampled the *xy* resolution to 8 nm using bilinear interpolation, then applied our PSSR model to the upsampled data to ensure the proper 8-to-2-nm transformation for which the PSSR model was originally trained.

EM segmentation and analysis. Image sets generated from the same region of the neuropil (LR-bilinear, LR-PSSR and HR) were aligned rigidly using the ImageJ plugin Linear stack alignment with SIFT[57]. Presynaptic axonal boutons ($n=10$) were identified and cropped from the image set. The bouton image sets from the three conditions were then assigned randomly generated filenames and distributed in a blinded manner to two human experts for manual segmentation of presynaptic vesicles. Vesicles were identified by having a clear and complete membrane, being round in shape and of approximately 35 nm in diameter. For consistency between human experts, vesicles that were embedded in or attached to obliquely sectioned axonal membranes were excluded. Docked and nondocked synaptic vesicles were counted as separate pools. Vesicle counts were recorded and unblinded and grouped by condition and by expert counter. Linear regression analyses were conducted between the counts of the HR images and the corresponding images of the two different LR conditions (LR-bilinear and LR-PSSR) to determine how closely the counts corresponded between the HR and LR conditions. Linear regression analysis was also used to determine the variability between counters.

Fluorescence imaging and analysis. *U2OS cell culture.* U2OS cells were purchased from ATCC. Cells were grown in DMEM supplemented with 10% fetal bovine serum at 37 °C with 5% CO_2. Cells were plated onto either eight-well no. 1.5 imaging chambers or no. 1.5 35-mm dishes (Cellvis) coated with 10 µg ml^{-1} fibronectin in PBS at 37 °C for 30 min before plating. Next, 50 nm MitoTracker Deep Red or CMXRos Red (ThermoFisher) was added for 30 min then washed for at least 30 min to allow for recovery time before imaging in FluoroBrite (ThermoFisher) media.

Airyscan confocal imaging of U2OS cells. To generate our ground truth training and testing dataset we used a Zeiss Airyscan LSM 880, an advanced confocal microscope that uses a 32-detector array and postprocessing pixel reassignment to generate images roughly 1.7× higher in resolution (roughly 120 nm) and roughly 8× higher SNR than a standard confocal system. All HR ground truth training data were acquired with a ×63 objective with at least 2× Nyquist sampling pixel sizes (roughly 50 ± 10 nm pixels), then Airyscan processed (deconvolved) using Zeiss Zen software. For the real-world LR test data, we acquired images at 16× lower pixel resolution (roughly 200 nm pixel sizes) with a 2.5 AU pinhole on a PMT confocal detector, without any additional image processing. We maintained equal pixel dwell times for the HR versus LR testing acquisitions, resulting in overall 16× shorter exposure times for the LR images. To ensure minimal phototoxicity, we also decreased the laser power for our LR acquisitions by a factor of 4 or 5 (see Tables 1 and 2 for more details), resulting in a net laser dose decrease of roughly 64–80× (for example, 5× lower laser power and 16× shorter exposure time yields a 80× lower laser dose). Furthermore, our LR testing data were not deconvolved and used a much larger effective pinhole size than the HR Airyscan ground truth data, resulting in a blurrier image with lower optical resolution. Cells were imaged with a ×63 1.4 numerical aperture (NA) oil objective on a Zeiss 880 LSM Airyscan confocal system with an inverted stage and heated incubation system with 5% CO_2 control. For both HR and LR images, equal or lower (when indicated) laser power and equal pixel dwell time of roughly 1 µs per pixel was used. For testing PSSR-MF, at least ten sequential frames of fixed samples were acquired with HR and LR settings to facilitate PSSR-MF processing.

Neuron preparation. Primary hippocampal neurons were prepared from E18 rat (Envigo) embryos as previously described. Briefly, hippocampal tissue was dissected from embryonic brain and further dissociated to single hippocampal neuron by trypsinization with Papain (Roche). The prepared neurons were plated on coverslips (Bellco Glass) coated with 3.33 µg ml^{-1} laminin (Life Technologies) and 20 µg ml^{-1} poly-L-lysine (Sigma) at the density of 7.5×10^4 cells per cm^2. The cells were maintained in Neurobasal medium supplemented with B27 (Life Technologies), penicillin/streptomycin and L-glutamine for 7–21 d in vitro. Two days before imaging, the hippocampal neurons were transfected with Lipofectamine 2000 (Life Technologies).

Temporal consistency analysis. Given a preprocessed time-lapse video with N total number of frames, the cross-correlation coefficient ($\emptyset(X_i, X_{i+1}), i \in [1, N)$) was calculated for two neighboring frames (X_i and X_{i+1}) repeatedly across each video with a step size of $i = 1$ (Fig. 4b). To ensure comparisons between PSSR output and ground truth data were not biased by high-frequency artifacts and noise, each time-lapse was first preprocessed with a Gaussian blur filter ($\sigma = 5$).

Fission event detection and analysis. Given fission events cannot be precisely defined across different evaluators, a HR time-lapse of MitoTracker stained cells was first given to two human experts as a pilot experiment to examine and correct the inspection performance of all experts. Three conditions (LR-bilinear, LR-PSSR and HR) of the same Airyscan time-lapses ($n = 6$) were then sequentially assigned in a blinded manner to two human experts for mitochondrial fission event detection. Fission event counts were recorded and unblinded and grouped by condition and by expert counter. Linear regression analyses were conducted between the counts of the HR images and the corresponding images of the two different LR conditions (LR-bilinear and LR-PSSR) to determine how closely the counts corresponded between the HR and LR conditions. Linear regression analysis was used to determine the variability between counters.

Neuronal mitochondria imaging and kymograph analysis. Live-cell imaging of primary neurons was performed using a Zeiss LSM 880 confocal microscope, enclosed in a temperature control chamber at 37 °C and 5% CO_2, using a ×63 (NA 1.4) oil objective in SR-Airyscan mode (that is, 0.2 AU virtual pinhole). For LR conditions, images were acquired with a confocal PMT detector with a pinhole size of 2.5 AU at 440 × 440 pixels at 0.5× Nyquist (170 nm per pixel) every 270.49 ms using a pixel dwell time of 1.2 µs and a laser power ranging between 1 and 20 µW. For HR conditions, images were acquired at 1,764 × 1,764 pixels at 2× Nyquist (roughly 42 nm per pixel) every 4.33 s using a pixel dwell time of 1.2 µs and a laser power of 20 µW. Imaging data were collected using Zen Black software. HR images were further processed using Zen Blue's 2D-SR-Airyscan processing. Kymograph analysis of the time-lapse videos were conducted using ImageJ plugin Kymolyzer as described previously[58].

Fluorescence photobleaching quantification. Normalized mean intensity over time was measured using Fiji software. Given a time-lapse video with N frames, a background region was randomly selected and remained unchanged across frames. The normalized mean intensity ($\bar{I}^{(i)}_{\text{norm}}$) can be expressed as:

$$\bar{I}^{(i)}_{\text{norm}} = \frac{\bar{I}^{(i)}_{\text{img}} - \bar{I}^{(i)}_{\text{bg}}}{\text{MAX}\left(\bar{I}^{(i)}_{\text{img}} - \bar{I}^{(i)}_{\text{bg}}\right)}, i \in [1, N]$$

where i represents the frame index, $\bar{I}^{(i)}_{\text{bg}}$ represents the mean intensity of the selected background region at frame i and $\bar{I}^{(i)}_{\text{img}}$ represents the intensity mean of the entire frame i.

Comparing PSSR with other methods. *Block-matching and 3D filtering (BM3D) denoising algorithm.* We compared PSSR with BM3D, on both EM and fluorescence MitoTracker data. Application of BM3D before (LR-BM3D-bilinear) and after (LR-bilinear-BM3D) bilinear upsampling of pixel sizes were both tested. A wide range of Sigma ($\sigma \in (0, 95]$), the key parameter that defines the assumed zero-mean white Gaussian noise in the images, was thoroughly explored. The exact same test set was used for quantification of PSSR versus BM3D results.

CARE and rolling average methods. A semisynthetically crappified dataset of MitoTracker data was used to train both CARE and PSSR networks (PSSR-SF and PSSR-MF) in a consistent manner. We applied the trained models to semisynthetically generated LR testing time-lapses. Testing data were similarly crappified as training data. Five-frame rolling average processing was further applied to CARE and single-frame testing output.

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

Data availability

Example training data and pretrained models are included in the GitHub release (https://github.com/BPHO-Salk/PSSR). The entirety of our training and testing datasets and data sources are available at Texas Data Repository (https://doi.org/10.18738/T8/YLCK5A). Source data are provided with this paper.

Code availability

PSSR source code and documentation are available for download on GitHub (https://github.com/BPHO-Salk/PSSR) and are free for use under the BSD 3-Clause License.

References

41. Perez, L. & Wang, J. The effectiveness of data augmentation in image classification using deep learning. Preprint at https://arxiv.org/abs/1712.04621 (2017).
42. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)* **9351**, 234–241 (2015).
43. Shi, W. et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1874–1883 (IEEE, 2016).
44. Harada, Y., Muramatsu, S. & Kiya, H. Multidimensional multirate filter without checkerboard effects. In *Proc. 9th European Signal Processing Conf. (EUSIPCO 1998)* 1–4 (IEEE, 1998).
45. Sugawara, Y., Shiota, S. & Kiya, H. Super-resolution using convolutional neural networks without any checkerboard artifacts. In *Proc. 25th IEEE Int. Conf. Image Processing (ICIP)* 66–70 (IEEE, 2018).
46. Aitken, A. et al. Checkerboard artifact free sub-pixel convolution: a note on sub-pixel convolution, resize convolution and convolution resize. Preprint at https://arxiv.org/abs/1707.02937 (2017).
47. Zhang, H., Goodfellow, I., Metaxas, D. & Odena, A. Self-attention generative adversarial networks. *Proc. 36th Int. Conf. Machine Learning, PMLR 97* 7354–7363 (2019).
48. Loshchilov, I. & Hutter, F. SGDR: Stochastic Gradient Descent with Warm Restarts. In *Proc. Int. Conf. Learning Representations (ICLR)* (2017).
49. Smith, L. N. Cyclical learning rates for training neural networks. In *Proc. IEEE Winter Conf. Applications of Computer Vision (WACV)* 464–472 (IEEE, 2017).
50. Smith, L. N. A disciplined approach to neural network hyper-parameters: part 1—learning rate, batch size, momentum, and weight decay. Preprint at https://arxiv.org/abs/1803.09820 (2018).
51. Smith, L. N. & Topin, N. Super-convergence: very fast training of residual networks using large learning rates. In *Proc. Artif. Intell. Mach. Learn. Multi-Domain Oper. Appl. 6* (2019).
52. Hore, A. & Ziou, D. Image quality metrics: PSNR vs. SSIM. In *Proc. 20th Int. Conf. Pattern Recognition* 2366–2369 (IEEE, 2010).
53. Kuwajima, M., Mendenhall, J. M. & Harris, K. M. Large-volume reconstruction of brain tissue from high-resolution serial section images acquired by SEM-based scanning transmission electron microscopy. *Methods Mol. Biol.* **950**, 253–273 (2013).
54. Deerinck, T. J., Bushong, E. A., Thor, A. & Ellisman, M. H. *NCMIR Methods for 3D EM: A New Protocol for Preparation of Biological Specimens for Serial Block Face Scanning Electron Microscopy* (National Center for Microscopy and Imaging Research, 2010).

55. Takemura, S. Y. et al. Synaptic circuits and their variations within different columns in the visual system of *Drosophila*. *Proc. Natl Acad. Sci. USA* **112**, 13711–13716 (2015).
56. Kubota, Y. et al. A carbon nanotube tape for serial-section electron microscopy of brain ultrastructure. *Nat. Commun.* **9**, 437 (2018).
57. Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**, 91–110 (2004).
58. Pekkurnaz, G., Trinidad, J. C., Wang, X., Kong, D. & Schwarz, T. L. Glucose regulates mitochondrial motility via Milton modification by O-GlcNAc transferase. *Cell* **158**, 54–68 (2014).

Acknowledgements

We thank J. Sedat, T. Sejnowski, A. Pinto-Duarte, F. Jug, M. Weigert, K. Prakash, S. Saalfeld and the entire NSF NeuroNex consortium for invaluable advice and critical feedback on our data and the manuscript. We also thank H. Hess and S. Xu for sharing their FIB–SEM data. U.M., L.F., T.Z. and S.W.N. are supported by the Waitt Foundation, core grant application no. NCI CCSG (CA014195). U.M. is a Chan-Zuckerberg Initiative Imaging Scientist and supported by NSF NeuroNex Award No. 2014862, and National Institutes of Health (NIH) grant no. R21 DC018237. C.R.S. is supported by NIH F32 GM137580. J.H. and F.M. are supported by the Wicklow AI in Medicine Research Initiative. K.H. is supported by NSF grant nos. 1707356 and NSF NeuroNex Award no. 2014862 and NIH/NIMH grant no. 2R56MH095980-06. Research in the laboratory of G.P. is supported by the Parkinson's Foundation (PF-JFA-1888) and NIH/NIGMS grant no. R35GM128823. S.B.Y. is funded by NIH grant no. T32GM007240. Y.K. was supported by Japan Society for the Promotion of Science KAKENHI grant nos. 17H06311 and 19H03336, and by AMED grant no. JP20dm0207084. We acknowledge the Texas Advanced Computing Center at The University of Texas at Austin for providing GPU resources that have contributed to the research results reported within this paper. We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the NVIDIA Quadro M5000 and NVIDIA Titan V used for this research.

Author contributions

L.F. designed the research, performed or participated in software development, all experiments and analyses, and wrote the paper. F.M. and J.H. designed the research and, together with K.K., A.A.L., Z.L. and A.S., contributed to software development. J.M., K.H., S.W.N. and Y.K. collected EM data. C.R.S., T.Z. and M.W. collected MitoTracker data. S.Y. and G.P. collected neuronal mitochondria data. S.W.N. and L.K. performed vesicle segmentation analysis. C.R.S., S.Y. and G.P. performed neuronal mitochrondrial mobiltiy analysis. S.W.N., S.Y., G.P., A.A.L., Z.L. and A.S. contributed to data visualization. U.M., K.H. and Z.Z. contributed with resources. U.M. and K.H. secured funding. U.M. conceived and designed the research, participated in software development, all experiments and analyses, oversaw the project and wrote the paper.

Competing interests

U.M. and L.F. have filed a patent application covering some aspects of this work (International Patent WO2020041517A9: 'Systems and methods for enhanced imaging and analysis', inventors U.M. and L.F., published on 1 October 2020). The rest of the authors declare no competing interests.

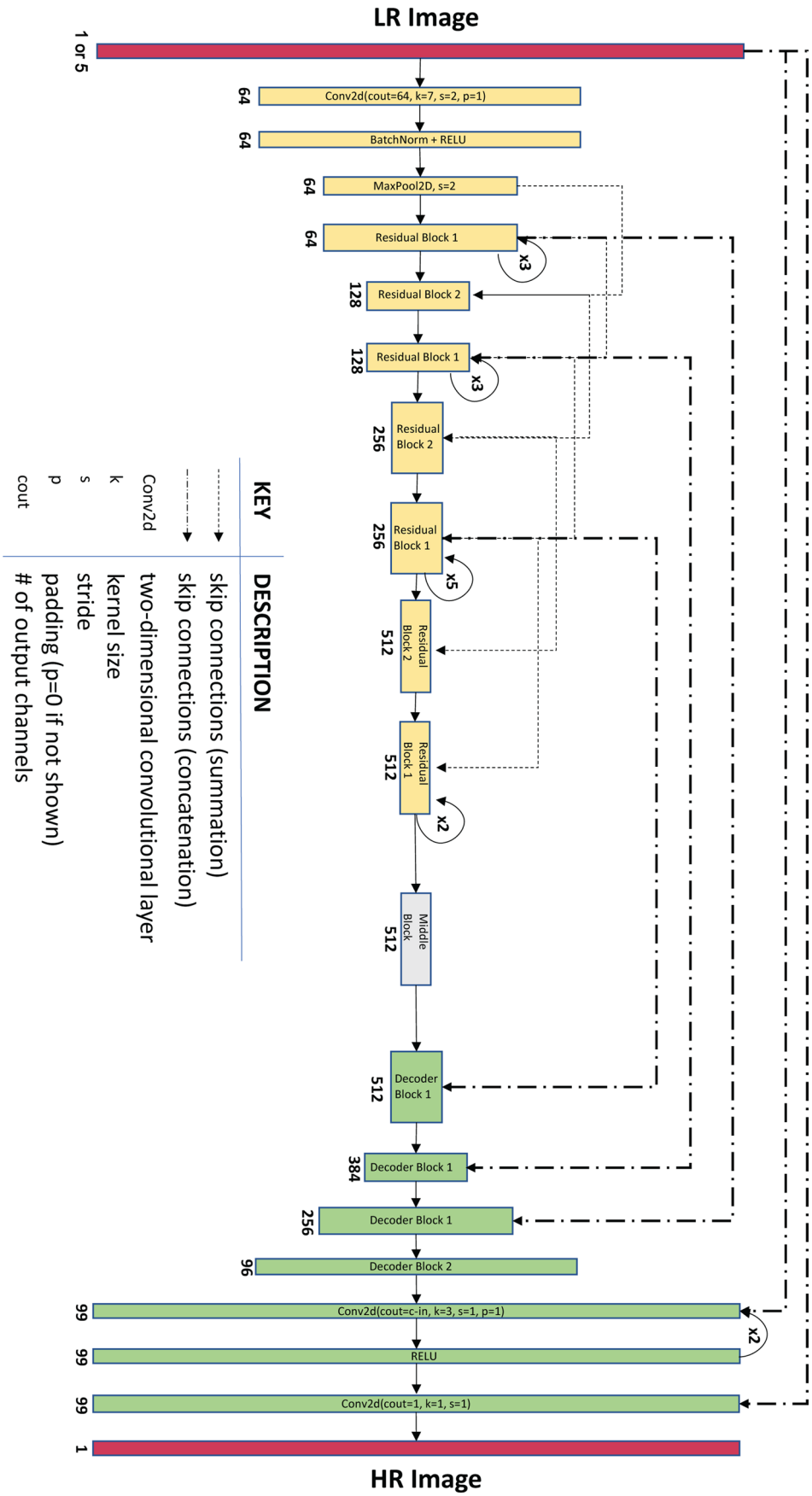
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41592-021-01080-z.

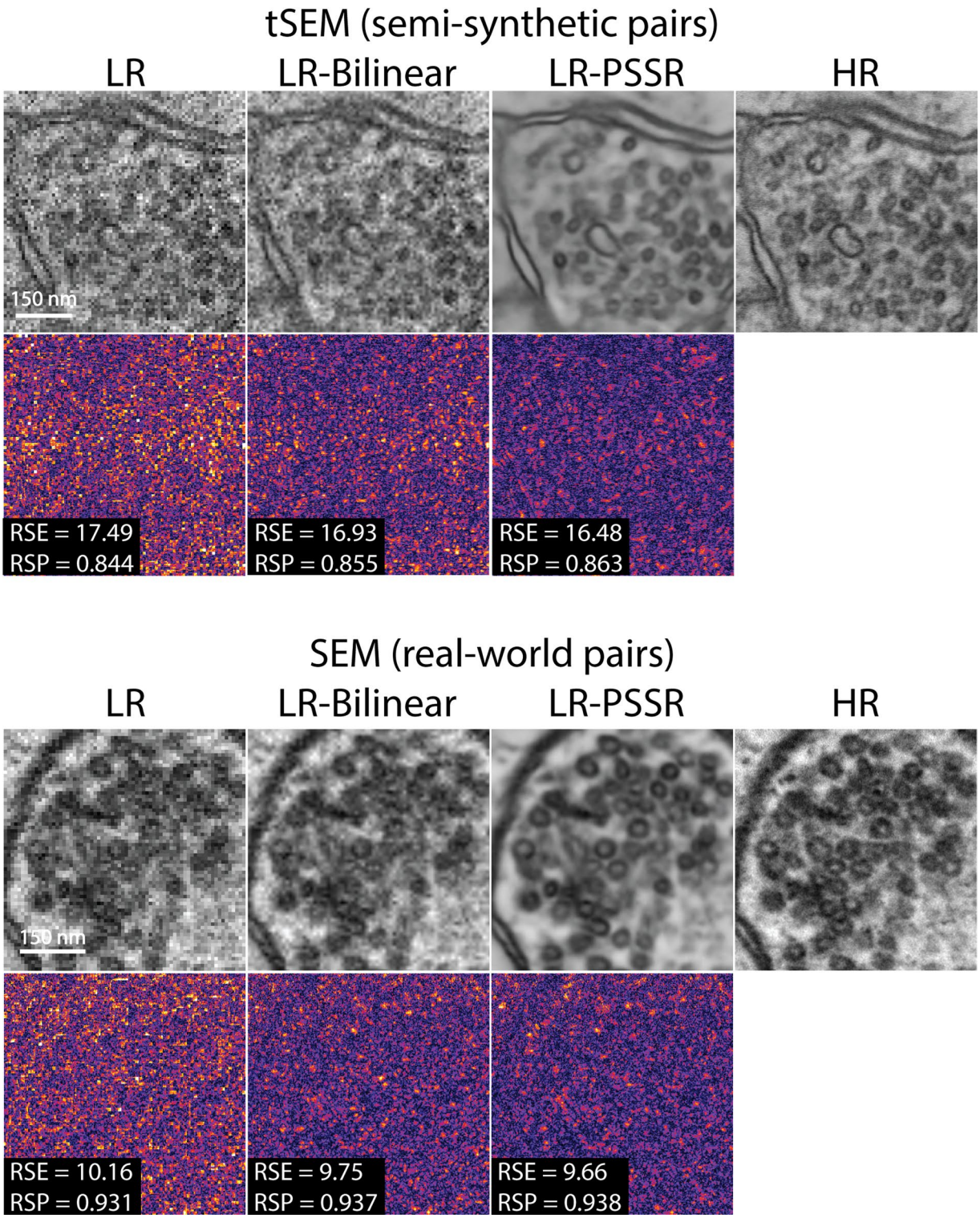
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-021-01080-z.

Correspondence and requests for materials should be addressed to U.M.

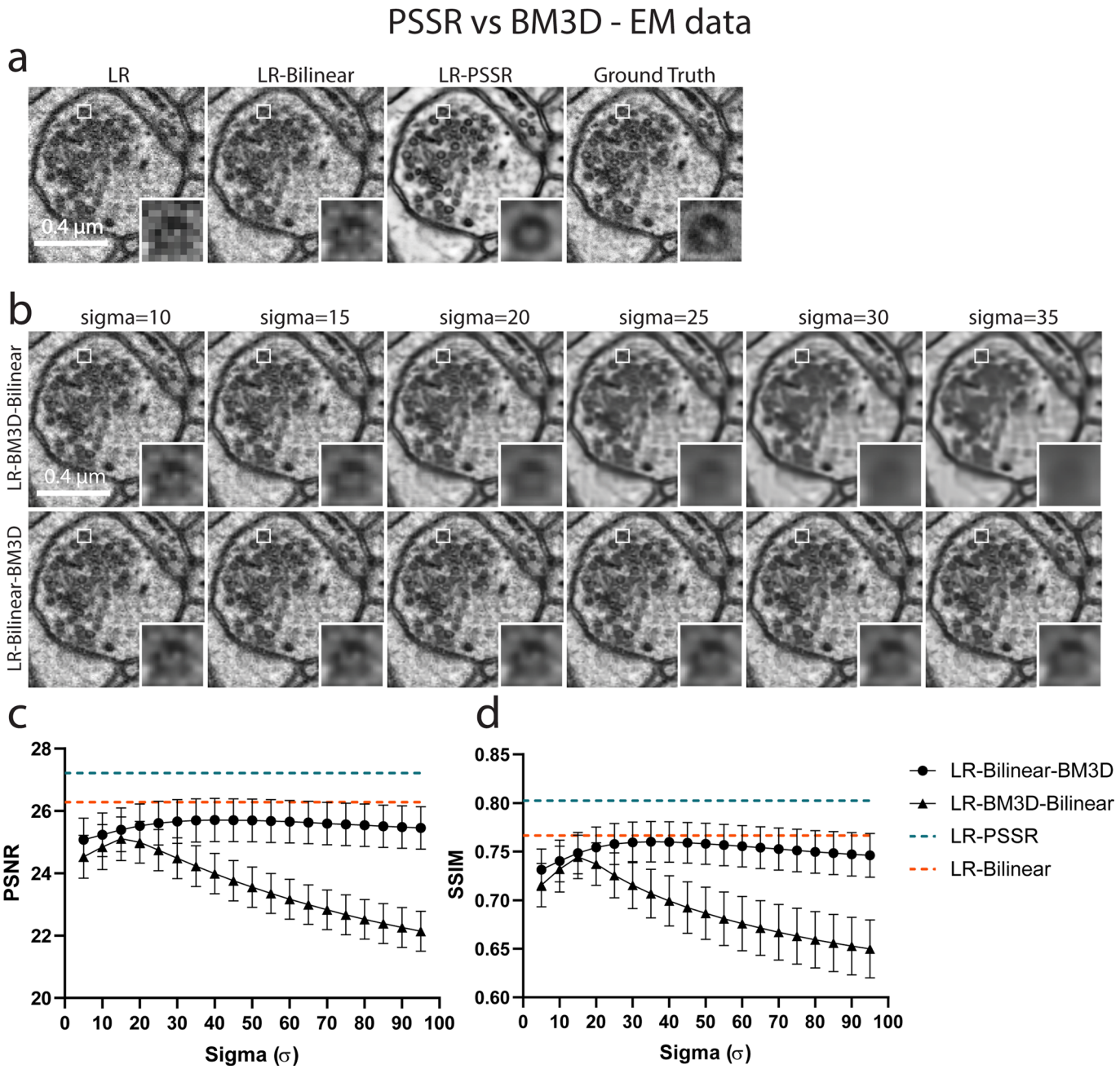
Peer review information *Nature Methods* thanks Jie Tian and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.



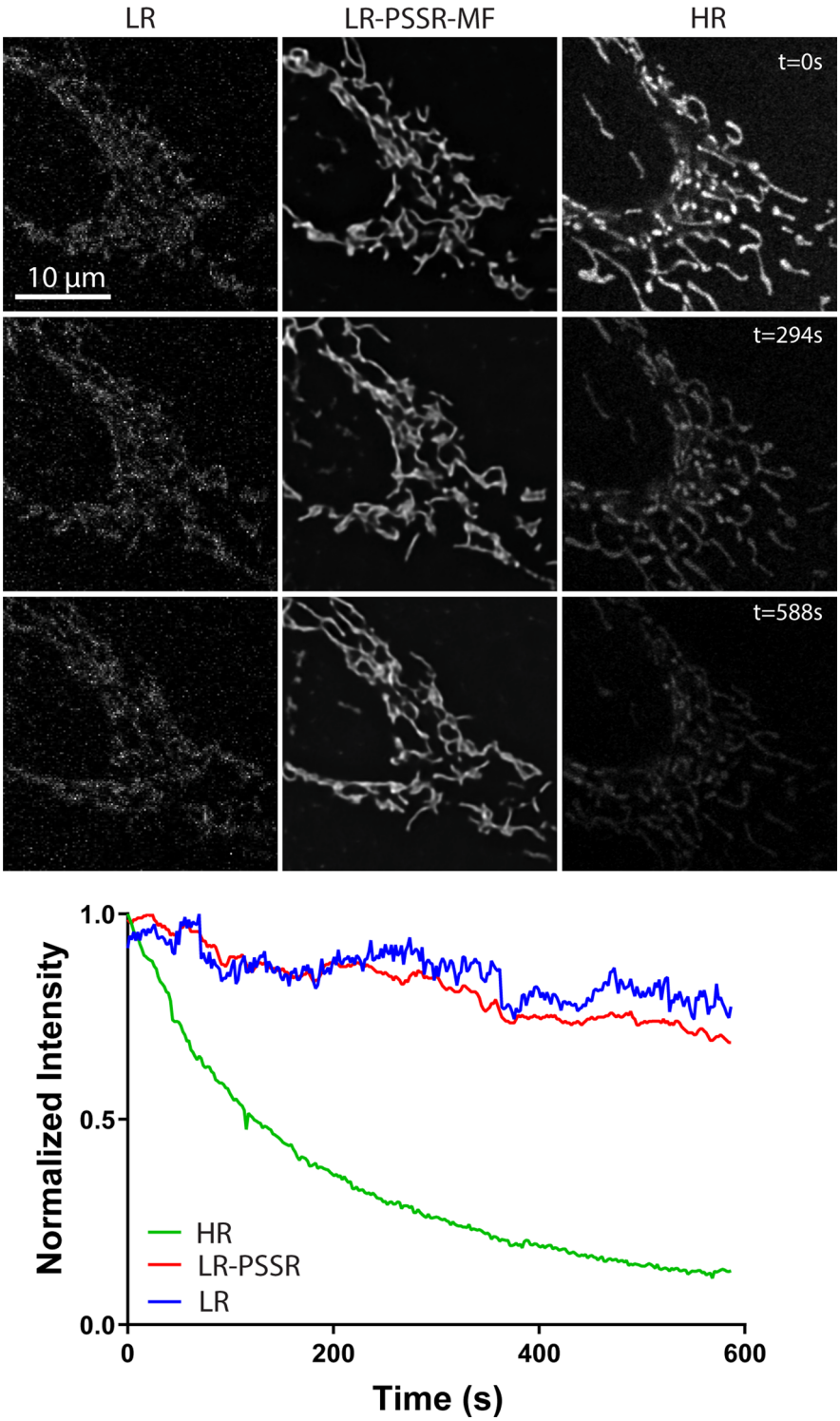
Extended Data Fig. 1 | PSSR Neural Network architecture. Shown is the ResNet-34 based U-Net architecture. Single-frame PSSR (PSSR-SF) and multi-frame PSSR (PSSR-MF) have 1 or 5 input channels, separately.



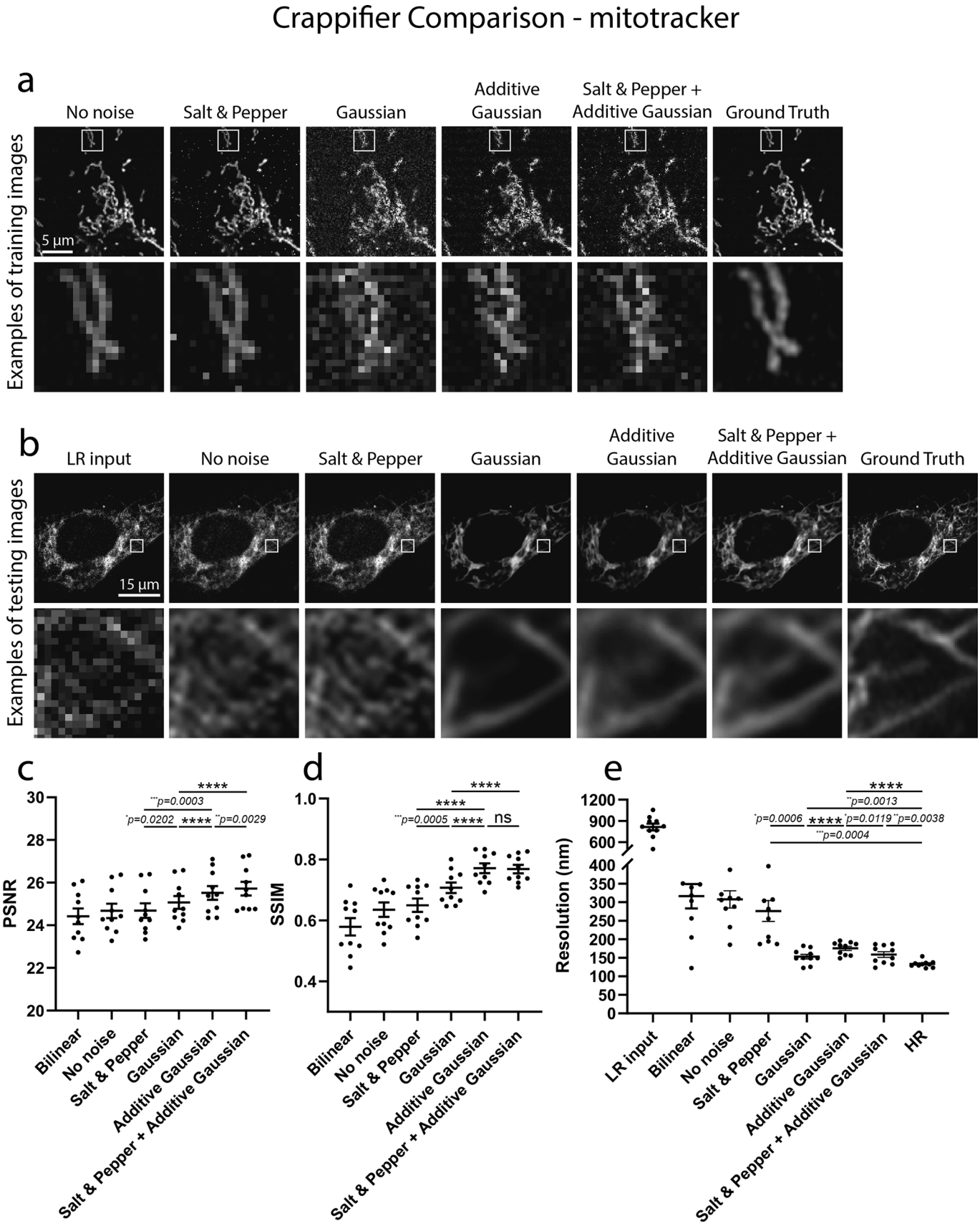
Extended Data Fig. 2 | NanoJ-SQUIRREL error-maps of EM data. NanoJ-SQUIRREL was used to calculate the resolution scaled error (RSE) and resolution scaled Pearson's coefficient (RSP) for both semi-synthetic and real-world acquired low (LR), bilinear interpolated (LR-Bilinear), and PSSR (LR-PSSR) images versus ground truth high-resolution (HR) images. For these representative images from Fig. 2, the RSE and RSP images are shown along with the difference images for each output.



Extended Data Fig. 3 | Comparison of PSSR vs. BM3D on EM data. PSSR restoration was compared to the Block-matching and 3D filtering (BM3D) denoising algorithm. BM3D was applied to low-resolution real-world SEM images before (LR-BM3D-Bilinear) and after (LR-Bilinear-BM3D) bilinear upsampling. A wide range of Sigma ($\sigma \in (0, 95]$, with step size of 5), the key parameter that defines the assumed zero-mean white Gaussian noise in BM3D method, was thoroughly explored. Images of the same region from the LR input, bilinear upsampled, PSSR restored, and Ground truth is displayed in (**a**). Results of LR-BM3D-Bilinear (**b**, top row) and LR-Bilinear-BM3D (b, bottom row) with sigma ranging from [10, 15, …, 35] are shown. PSNR and SSIM results of LR-BM3D-Bilinear and LR-Bilinear-BM3D across the explored range of sigma are plotted in (**c**) and (**d**). Metrics for bilinear-upsampled and PSSR-restored images of the same testing set are shown as dashed lines in orange (LR-Bilinear: PSNR = 26.28 ± 0.085; SSIM = 0.767 ± 0.0031) and blue (LR-PSSR: PSNR = 27.21 ± 0.084; SSIM = 0.802 ± 0.0026). n = 12 independent images for all conditions. Values are shown as mean ± SEM.

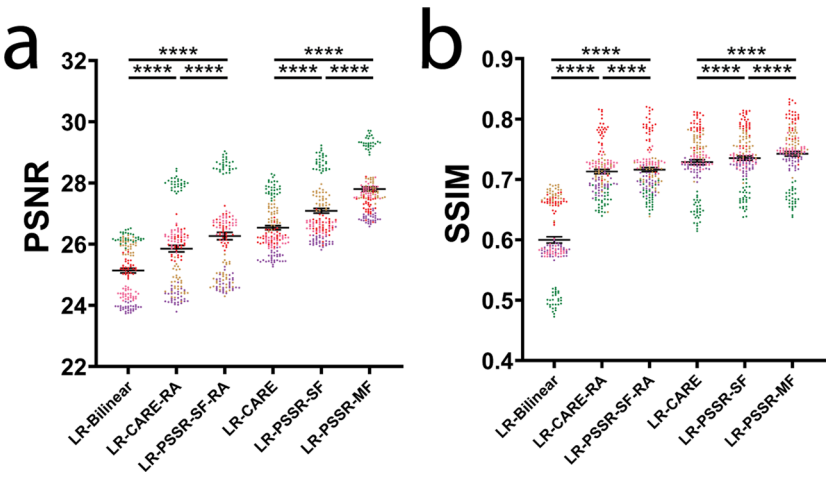


Extended Data Fig. 4 | Undersampling substantially reduces photobleaching. U2OS cells stained with mitotracker were imaged every 2 seconds with the same laser power (2.5µW) and pixel dwell time (~1µs), but with 16x lower resolution (196 x 196 nm xy pixel size) than full resolution Airyscan acquisitions (~49 x 49 nm xy pixel size). Mean intensity plots show the relative rates of fluorescence intensity loss over time (that is photobleaching) for LR, LR-PSSR, and HR images.

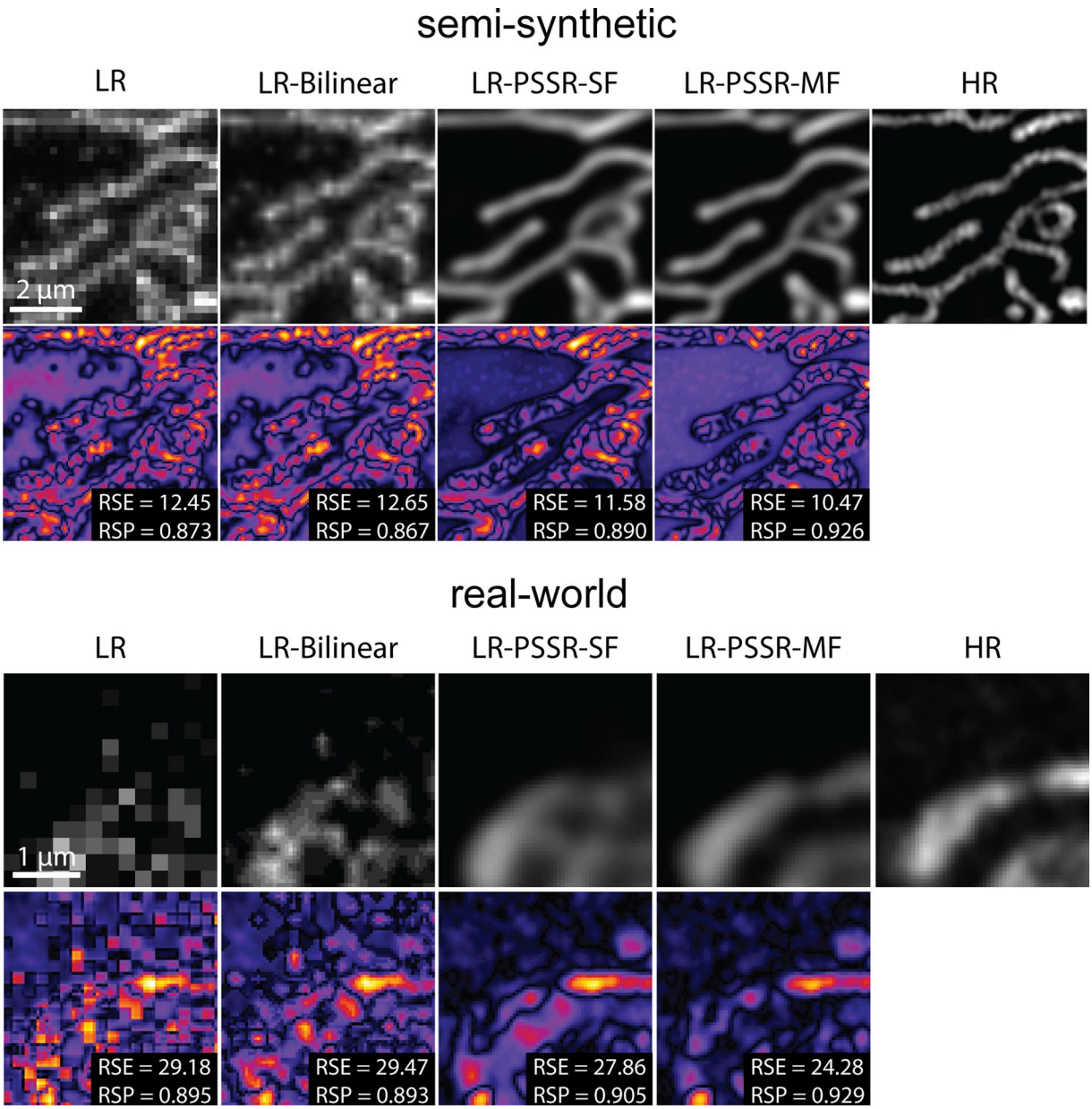


Extended Data Fig. 5 | See next page for caption.

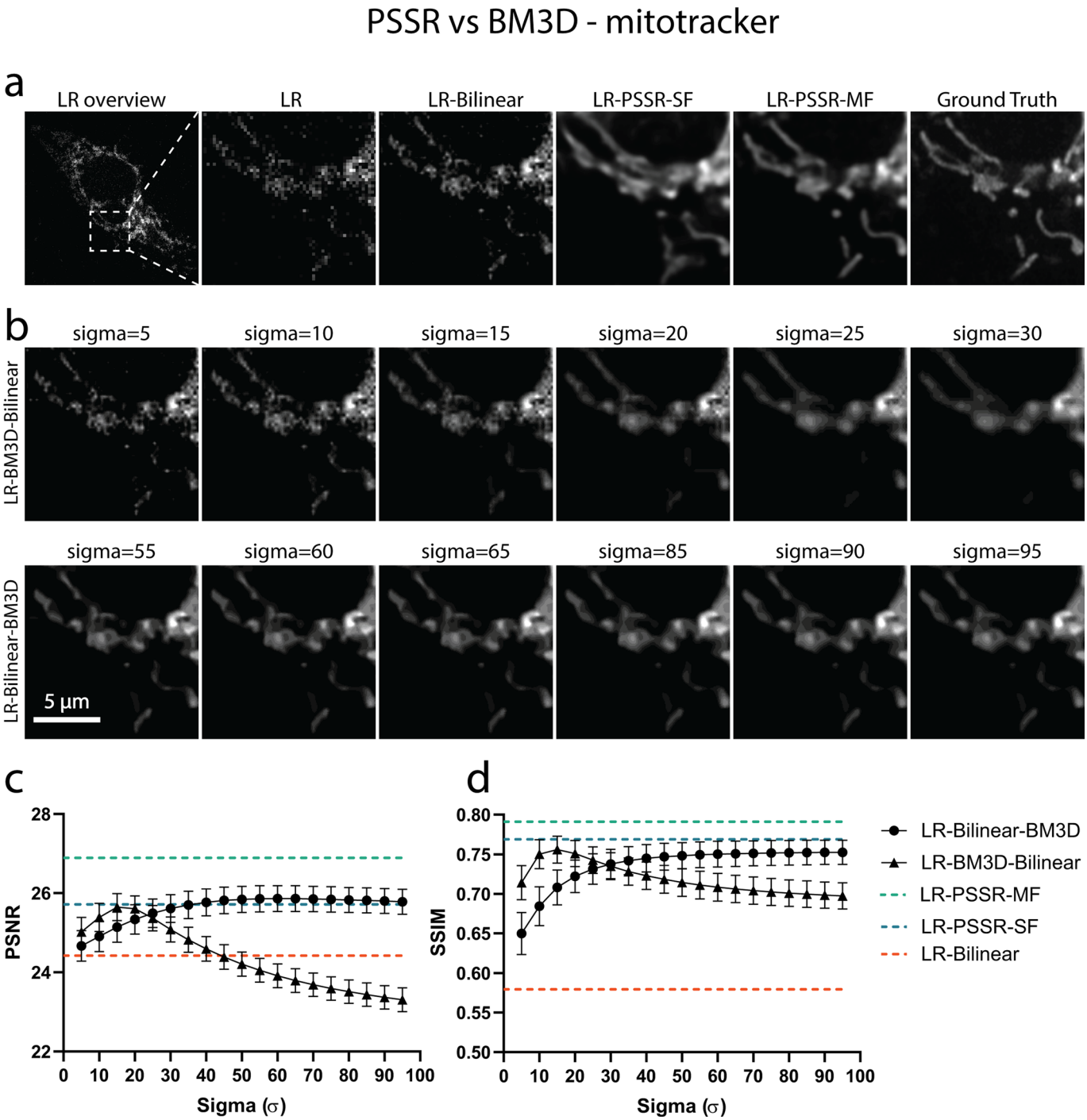
Extended Data Fig. 5 | Evaluation of crappifiers with different noise injection on mitotracker data. Examples of crappified training images, visualized results and metrics (PSNR, SSIM and FRC resolution) of PSSR-SF models that were trained on high- and low-resolution pairs semi-synthetically generated by crappifiers with different noise injection were presented. **a**, Shown is an example of crappified training images generated by different crappifiers, including 'No noise' (no added noise, downsampled pixel size only), Salt & Pepper, Gaussian, Additive Gaussian, and a mixture of Salt & Pepper plus Additive Gaussian noise. High-resolution version of the same region is also included. **b**, Visualized restoration performance of PSSR models that used different crappifiers (No noise, Salt & Pepper, Gaussian, Additive Gaussian, and a mixture of Salt & Pepper plus Additive Gaussian noise). LR input and Ground Truth of the example testing ROI are also shown. PSNR (**c**), SSIM (**d**) and FRC (**e**) quantification show the PSSR model that used 'Salt & Pepper + Additive Gaussian' crappifier yielded the best overall performance ($n = 10$ independent time-lapses of fixed samples with $n = 10$ timepoints for all conditions). All values are shown as mean ± SEM. P values are specified in the figure for $0.0001 < p < 0.05$. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, ns = not significant; Two-sided paired *t*-test.



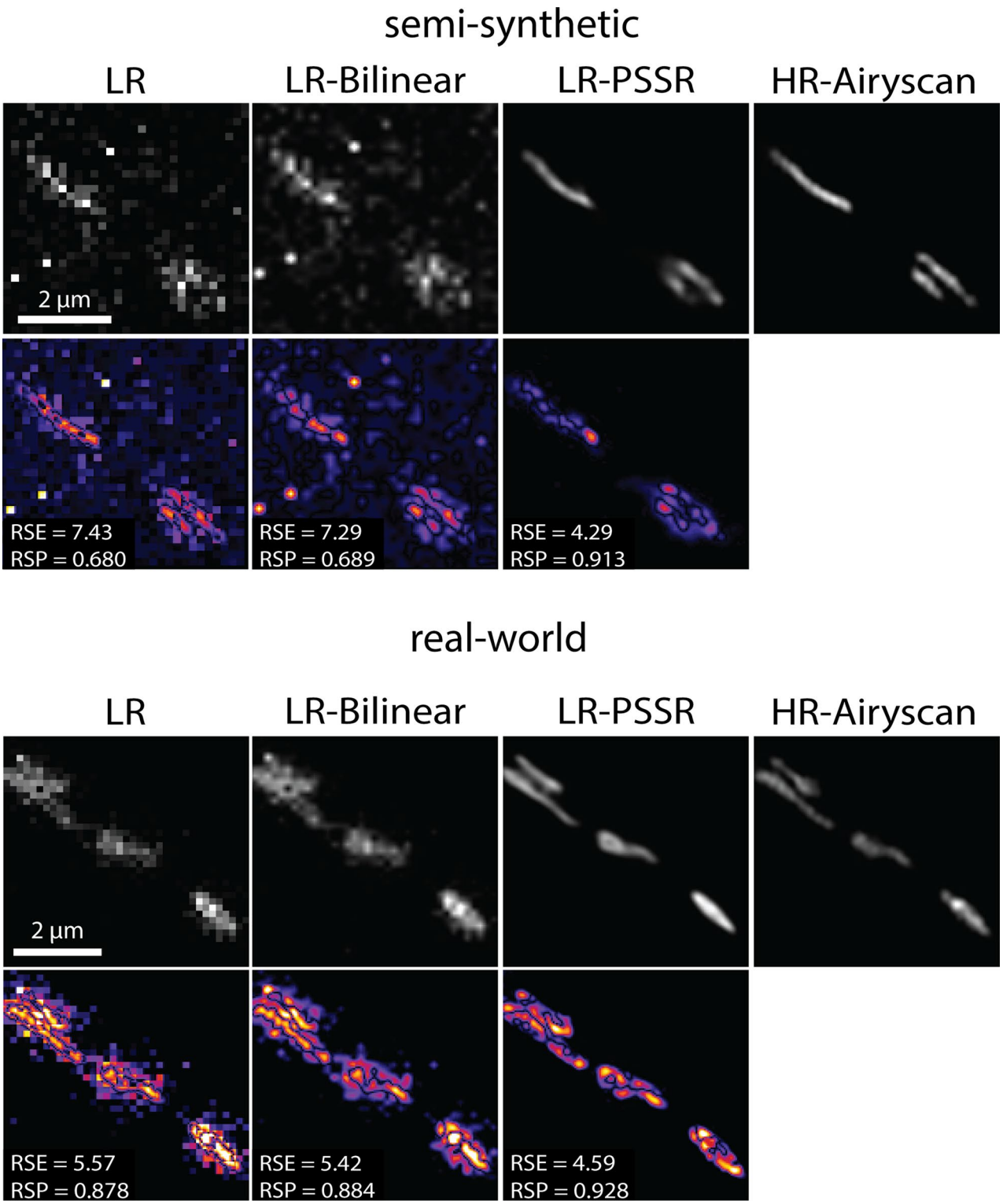
Extended Data Fig. 6 | Quantitative comparison of CARE and PSSR-SF with PSSR-MF and Rolling Average (RA) methods for timelapse data. PSNR (**a**) and SSIM (**b**) quantification show a decrease in accuracy when applying RA to LR-CARE and LR-PSSR-SF, while multi-frame PSSR provides superior performance compared to LR-PSSR-SF and CARE before and after RA processing. Data points were color-coded based on different cells. See Fig. 4c for visualized comparisons, and Supplementary Video 6 for a video comparison of the entire timelapse for CARE, LR-PSSR-SF, LR-PSSR-SF-RA, and LR-PSSR-MF. N = 5 independent timelapses with n = 30 timepoints each, achieving similar results. All values are shown as mean ± SEM. ****$p < 0.0001$; Two-sided paired *t*-test.



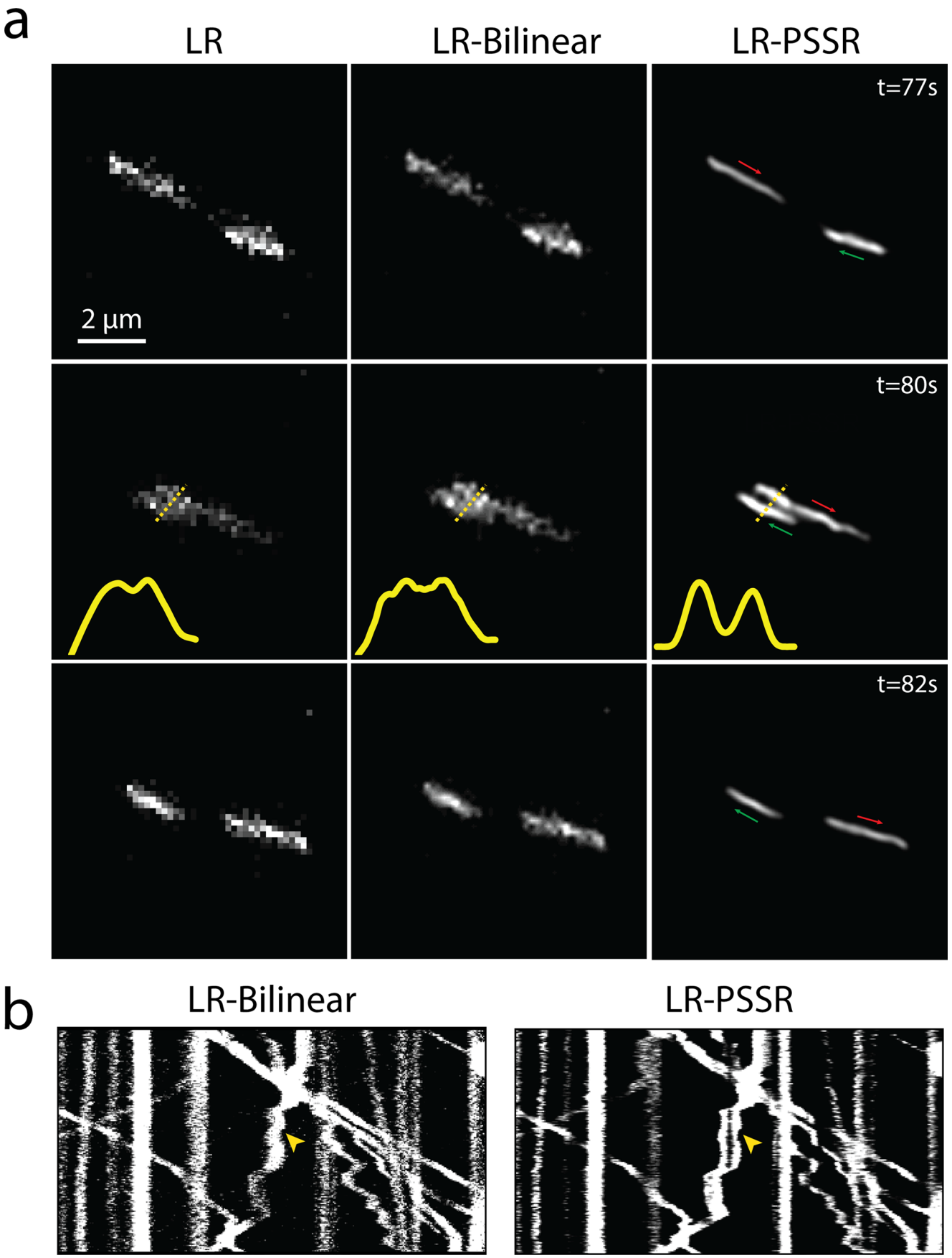
Extended Data Fig. 7 | NanoJ-SQUIRREL error-maps of MitoTracker data. NanoJ-SQUIRREL was used to calculate the resolution scaled error (RSE) and resolution scaled Pearson's coefficient (RSP) for both semi-synthetic and real-world acquired low (LR), bilinear interpolated (LR-Bilinear), and PSSR (LR-PSSR) images versus ground truth high-resolution (HR) images. For these representative images from Fig. 4, the RSE and RSP images are shown along with the difference images for each output.



Extended Data Fig. 8 | Compare PSSR with BM3D denoising method on mitotracker data. PSSR restored images was compared to results of applying BM3D denoising algorithm to low-resolution real-world mitotracker images before (LR-BM3D-Bilinear) and after (LR-Bilinear-BM3D) bilinear upsampling. A wide range of Sigma ($\sigma \in (0, 95]$, with step size of 5) was thoroughly explored. Examples of the same region from the LR input, bilinear upsampled, PSSR-SF restored, PSSR-MF restored, and Ground truth are displayed (**a**, top row). Images from the top 6 results (evaluated by both PSNR and SSIM values) of LR-BM3D-Bilinear (**a**, middle row) and LR-Bilinear-BM3D (**a**, bottom row) are shown. PSNR and SSIM results of LR-BM3D-Bilinear and LR-Bilinear-BM3D across the explored range of sigma are plotted in (**b**) and (**c**). Metrics resulted from bilinearly upsampled, PSSR-SF restored and PSSR-MF restored images of the same testing set are shown as dash lines in orange (LR-Bilinear: PSNR = 24.42 ± 0.367; SSIM = 0.579 ± 0. 0287), blue (LR-PSSR-SF: PSNR = 25.72 ± 0.323; SSIM = 0.769 ± 0.0139) and green (LR-PSSR-MF: PSNR = 26.89 ± 0.322; SSIM = 0.791 ± 0.0133). As it shows, in this fluorescence mitotracker example, BM3D performs better than bilinear upsampling with carefully defined noise distribution, whereas its general performance given both PSNR and SSIM is overall worse than single-frame PSSR (LR-PSSR-SF). Excitably, our multi-frame PSSR (LR-PSSR-MF) yields the best performance. n = 10 independent timelapses of fixed samples with n = 6-10 timepoints each for all conditions. Values are shown as mean ± SEM.



Extended Data Fig. 9 | NanoJ-SQUIRREL error-maps of neuronal mitochondria data. NanoJ-SQUIRREL was used to calculate the resolution scaled error (RSE) and resolution scaled Pearson's coefficient (RSP) for both semi-synthetic and real-world acquired low (LR), bilinear interpolated (LR-Bilinear), and PSSR (LR-PSSR) images versus ground truth high-resolution (HR) images. For these representative images from Fig. 5, the RSE and RSP images are shown along with the difference images for each output.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | PSSR facilitates detection of mitochondrial motility and dynamics. Rat hippocampal neurons expressing mito-dsRed were undersampled with a confocal detector using 170 nm pixel resolution (LR) to facilitate faster frame rates, then restored with PSSR (LR-PSSR). **a**, before and after timepoints of the event shown in Fig. 5 wherein two adjacent mitochondria pass one another but cannot be resolved in the original low-resolution (LR) or bilinear interpolated (LR-Bilinear) image but are clearly resolved in the LR-PSSR image. **b**, kymographs of a LR vs LR-PSSR timelapse that facilitates the detection of a mitochondrial fission event (yellow arrow).

nature research

Corresponding author(s): Uri Manor

Last updated by author(s): Jan 20, 2021

Reporting Summary

Nature Research 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 Research 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	1) tSEM high-resolution training images were imaged using a STEM detector on a ZEISS Supra 40 scanning electron microscope with a 28kV accelerating voltage and an extractor current of 102uA (gun aperture 30um). Images were acquired with a 2nm pixel size and a field size of 24576 x 24576 pixels with Zeiss ATLAS. 2) For our EM testing data, paired LR and HR images of the adult mouse hippocampal dentate gyrus middle molecular layer neuropil were acquired from ultrathin sections (80nm) collected on silicon chips and imaged in a ZEISS Sigma VP FE-SEM. 3) Serial block face scanning electron microscope testing images were acquired with a Gatan 3View system installed on the ZEISS Sigma VP FE-SEM. 4) High resolution Airyscan images were acquired using 2x Nyquist pixel size of 42 - 59 nm/pixel (depending on the wavelength) in SR mode (i.e. a virtual pinhole size of 0.2 AU), then processed using ZEISS Zen software (v2.3, blue edition) with auto-filter settings.
Data analysis	Final models were generated using fast.ai v1.0.55 library (https://github.com/fastai/fastai) and Pytorch 1.1.0. PSSR source code and documentation are available for download on GitHub (https://github.com/BPHO-Salk/PSSR) and are free for use under the BSD 3-Clause License. ImageJ (v1.52n) plugin SIFT was used for EM segmentation data alignment. ImageJ (v1.52n) plugin was used for kymograph analysis.

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

The entirety of our training and testing data sets and data sources are available at Texas Data Repository (https://doi.org/10.18738/T8/YLCK5A).

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

Life sciences study design

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

Sample size	Sample sizes were chosen to ensure sufficient statistical power while maintaining reasonable workload and costs for human segmentation tasks. Given our p-values ranged from 0.0005<p<0.05, we were satisfied that our sample sizes were above and beyond what was needed to reliably measure differences between conditions.
Data exclusions	One image was excluded from the neuronal mitochondria FRC analysis, due to the presence of an extreme outlier value in the LR condition that caused the average LR resolution to appear much lower than it is (even though the values for LR-Bilinear, LR-PSSR, and HR were all within normal range) (Figure 5b, bottom). A ROUT outlier test was performed to confirm that the excluded data point was a bona fide outlier. (Motulsky & Brown, BMC Bioinformatics, 2006)
Replication	We used multiple images from the same conditions and trained several independent models to ensure reproducibility. All attempts at replication were successful. The number of independent experiments performed are indicated in the figure legends, as excerpted below. Specifically, 1) Training experiments described in Figure 1a-b contains 40 independent image pairs, achieving similar results. 2) Testing data for PSNR and SSIM quantification described in Figure 1c-d contains 8 independent images, achieving similar results. 3) Testing data for FRC resolution quantification in Figure 1e contains 16 independent images, achieving similar results. 4) Semi-synthetic EM testing data described in Figure 2b contains 66 independent images, achieving similar results. 5) Real-world EM testing data described in Figure 2c contains 42 independent images, achieving similar results. 6) EM vesicle segmentation experiments described in Figure 3e were conducted with 10 independent bouton regions in each condition (LR-Bilinear, LR-PSSR and HR), achieving similar results. 7) Flickering analysis described in Figure 4b were conducted with 6 independent timelapses with n=80-120 timepoints each, achieving similar results. 8) Semi-synthetic mitotracker testing data described in Figure 4f (left) contains 8 independent timelapses with n=80-120 timepoints each, achieving similar results. 9) Real-world mitotracker testing data described in Figure 4f (right) contains 10 independent timelapses with n=10 timepoints each, achieving similar results. 10) Fission event counting analysis was conducted with 8 independent timelapses with n=80-120 timepoints each, achieving similar results. 11) Semi-synthetic neuronal mitochondria testing data described in Figure 5b (top) contains 7 independent timelapses with n=100 timepoints each, achieving similar results. 12) Real-world neuronal mitochondria testing data described in Figure 5b (bottom) contains 6 independent timelapses with n=12 timepoints each, achieving similar results. 13) Mitochondrial mobility analysis described in Figure 5f-i were conducted with n=76-216 mitochondria from 4 axons and 3 independent transfections. 14) BM3D EM quantification experiment described in Extended Data Figure 3 was conducted with 12 independent images for all conditions, achieving similar results. 15) Testing data for PSNR, SSIM and FRC quantification described in Extended Data Figure 5b-e contains 10 independent images for all conditions, achieving similar results. 16) PSNR and SSIM quantification described in Extended Data Figure 6 was conducted with 5 independent timelapses with n=30 timepoints each, achieving similar results. 17) BM3D mitotracker quantification experiment described in Extended Data Figure 8 was conducted with 10 independent timelapses of fixed samples with n=6-10 timepoints each for all conditions, achieving similar results.
Randomization	We randomized LR-Bilinear, LR-PSSR, and HR electron microscopy images, then distributed them to two blinded human experts for manual segmentation of pre-synaptic vesicles. Likewise, images in different conditions (LR-Bilinear, LR-PSSR and HR) were randomized and were passed to two blinded experts for mitochondria fission event counting. No samples/organisms were allocated into separate experimental groups because we were never comparing more than one biological condition.

Blinding	Investigators were blinded to group allocation during image collection and data analyses (including manual segmentation of presynaptic vesicles from our EM data and mitochondrial fission events from our live fluorescence timelapse imaging).

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	U2OS cells were purchased from ATCC.
Authentication	Cell lines were authenticated by ATCC but not by us.
Mycoplasma contamination	We regularly test for mycoplasma and none was detected.
Commonly misidentified lines (See ICLAC register)	None.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Primary hippocampal neurons used for our neuronal mitochondria imaging were collected from E18 embryonic Sprague Dawley rats. For EM imaging of rat hippocampal tissue a 7-month male Long Evans rat and a 8-week old male Wistar rat were used. All other images were generated from tissues and images generated for previously published studies.
Wild animals	The study did not involve wild animals
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All animal work was approved by the Institutional Animal Care and Use Committee (IACUC) of the Salk Institute for Biological Studies.

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