

RESOLFT time lapse imaging empowered by deep learning

Guillaume Minet

KTH Royal Institute of Technology

Anirban Ray

Fondazione Human Technopole

Francesca Pennacchietti

KTH Royal Institute of Technology <https://orcid.org/0000-0003-1769-972X>

Giovanna Coceano

KTH Royal Institute of Technology <https://orcid.org/0000-0001-6579-0857>

Florian Jug

Fondazione Human Technopole <https://orcid.org/0000-0002-8499-5812>

Ilaria Testa

`ilaria.testa@scilifelab.se`

KTH Royal Institute of Technology <https://orcid.org/0000-0003-4005-4997>

Article

Keywords:

Posted Date: December 3rd, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-8059028/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on October 11th, 2026. See the published version at <https://doi.org/10.1038/s41467-026-78580-z>.

RESOLFT time lapse imaging empowered by deep learning

Guillaume Minet¹, Anirban Ray^{2,3}, Francesca Pennacchietti¹, Giovanna Coceano¹, Florian Jug² and Ilaria Testa¹

¹*Department of Applied Physics and Science for Life Laboratory, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden*

²*Fondazione Human Technopole, 20157 Milan, Italy*

³*Technische Universität Dresden, 01069 Dresden, Germany*

* Corresponding author: Ilaria Testa

Email: ilaria.testa@scilifelab.se

Abstract

Capturing the dynamic interplay of proteins and organelles within cells requires microscopy methods with high spatial and temporal resolution. RESOLFT nanoscopy demonstrated great potential for live cell recording by using patterned light and reversibly switching fluorescent probes. Despite the latest improvement in both switching probes and optical schemes, the length and speed of the time-lapse is still limited by photobleaching and sampling. Here, we present two new approaches based on deep learning to restore low-SNR or under-sampled images, which pushes the experimental boundaries of RESOLFT nanoscopy achieving up to 5 times longer imaging with 10 times lower dose of light per frame, or a 4-fold increase in imaging speed. The increased speed and length of time lapses enable to follow vesicles budding in and out from intact mitochondria, as well the actin meshwork formation and dissociations opening up new possibility for dynamic data acquisition at the sub-organelle level.

Main

Fluorescence nanoscopy has revolutionized our view of the subcellular organization by specifically resolving organelles and protein complexes down to the nanoscale¹. Yet, monitoring sub-100 nm dynamics in live-cells remain challenging due to the increased photobleaching and associated photo-toxicity of nanoscopy techniques, as well as the usual compromise in temporal resolution that comes with increased spatial resolution.

All sub-100 nm nanoscopy techniques rely on ON-OFF switching of fluorophores and for live-cell imaging, reversibly switchable fluorescent probes (rsFPs)²⁻⁵ offer the distinct advantage of requiring relatively low doses of light ($\text{W}\cdot\text{km}^2/\text{cm}^2$) to initiate such transitions. Reversible Saturable Optical Fluorescent Transition (RESOLFT)⁶⁻⁸ takes advantage of rsFPs to record whole-cell timelapses with sub-100 nm resolution⁹⁻¹¹. However, the rsFPs have a limited cycling capability, often referred to as photoswitching fatigue, that ultimately limits the length of observation. Additionally, the imaging speed is limited both by the switching kinetics of the

rsFPs and the increased number of raw images needed to reconstruct the final super-resolved image. Here, we apply deep-learning-based methods to circumvent each one of those limitations: denoising to minimize photobleaching, and up-sampling of sub-sampled acquisitions to increase the recording speed.

We focus our work on RESOLFT, which has in its favor a simpler image formation, since it is coordinate-targeted based. This allows a robust image reconstruction, less prone to reconstruction artefacts in very noisy conditions, and offers a straightforward framework for the up-sampling training. More specifically, we used the MoNaLISA implementation⁹, which represents the current state-of-the-art with parallelized RESOLFT imaging, combining multi-foci patterns for both ON-switching and read-out along with superimposed periodic line patterns for OFF-switching (Fig. 1a), achieving live-cell recordings with 45-60 nm spatial resolution at ~ 0.5 Hz speed in a $\sim 40 \times 40 \mu\text{m}^2$ FOV.

Results

Prolonged imaging

To reduce photobleaching, we aimed at reducing the ON-switching intensity, which modulates the effective concentration of observed fluorescent molecules and therefore the intensity of the recorded signal¹². In a typical RESOLFT experiment, 80 – 100 % of the rsFPs are ON-switched, thus maximizing the Signal-to-Noise Ratio (SNR). Here, we propose to ON-switch only 10 – 20 % of the rsFPs, thus acquiring low-SNR images, and to restore high-SNR images afterwards with post-processing denoising. This not only minimizes photobleaching, but also mitigate phototoxicity that arise at the short irradiation wavelength used for ON-switching, typically 405 nm¹³.

Denoising has given rise to numerous deep-learning-based methods, including both supervised approaches, that rely on paired clean and noisy data, and unsupervised (or self-supervised) methods that learn to denoise without explicit clean targets. Although unsupervised methods are especially attractive for live cell nanoscopy because paired datasets are notoriously difficult to obtain, supervised methods generally offer superior performance. We thus aimed to find a denoising method that leverages supervised learning whenever high-quality ground truth is available, while still maintaining robust performance in unsupervised settings when it is not. The ideal strategy should also provide effective noise reduction without compromising the spatial resolution.

To find the best method, we systematically assessed the spatial resolution of predicted and experimentally generated GT images in two ways: by measuring the full width at half-maximum (FWHM) of fluorescently labeled vimentin filaments and calculating the image's Fourier Ring Correlation (FRC). The images of the very same sample region were acquired with low ($\sim 30 \text{ W/cm}^2$) and high ($\sim 300 \text{ W/cm}^2$) ON-switching energy (Fig. 1b). This paired dataset was used to train the supervised deep learning networks and evaluate the performances of both supervised and unsupervised methods (Fig. 1c).

We found that unsupervised (or self-supervised) methods such as Noise2Void (N2V)¹⁴ or Self-Inspired Noise2Noise (SN2N)¹⁵ do not maintain the spatial resolution: the average FWHM is 120 nm for SN2N, and 93 nm for N2V, while the GT's average FWHM is 58 nm (Fig. 1d). The same can be concluded from the FRC curves, where the curves of SN2N and N2V are clearly shifted to the lower frequencies compared to the GT's (Fig. 1d). Supervised methods like Content-Aware Image Restoration (CARE)¹⁶ or the more recent UNet-RCAN¹⁷ perform better than the unsupervised but still tend to smooth cellular structures and lose the finest features, resulting in higher average FWHM than the GT (73 nm for CARE, 70 nm for UNet-RCAN) and FRC curves dropping at lower frequencies. Interestingly, the differences between CARE and UNet-RCAN are small, contrary to what was reported before¹⁷.

Therefore, we propose here to use Hierarchical DivNoising (HDN)¹⁸, an unsupervised denoising technique which leverages a Ladder Variational Autoencoder (LVAE) architecture combined with an experimentally obtained noise model, and its supervised version, HDN^{sup}. We found that contrary to all other methods, HDN and HDN^{sup} maintain the spatial resolution with average FWHMs of 56 nm and 53 nm respectively. This is also confirmed by FRC curves: HDN closely matches the GT, and although HDN^{sup} is dropping slightly below the GT, it is the most shifted towards the higher frequencies compared to other supervised methods. This strong performance, both in supervised and unsupervised settings, highlights the adaptability of the HDN framework to varying data constraints: HDN^{sup} offers superior restoration, especially on inputs with the lowest SNR (see Supplementary Fig. S1), while HDN is a robust alternative when GT is not available.

We then applied our HDN^{sup} network that we trained on the fixed vimentin dataset, to living cells. Since the images were recorded with ~10 times less energy of UV-Vis light, photobleaching was substantially reduced, resulting in up to ~5 times longer time lapse imaging compared to conventional RESOLFT done with higher ON-switching light (Fig. 2a, b). Of note, these experiments were performed with endogenous levels of expression of vimentin molecules. For brighter structures, such as mitochondria labelled with rsEGFP2 targeting the outer membrane, we used the unsupervised version to extend timelapses over more than 100 frames (Supplementary Fig. S4, Supplementary video 1). Finally, we used HDN^{sup} to image complex and interlaced structures such as the actin meshwork, which is often too dim to be imaged for more than a few frames in conventional RESOLFT. We first acquired a high-quality training dataset on fixed cells expressing LifeAct-rsEGFP2. After validating the capability of HDN^{sup} to accurately restore the meshwork (Fig. 2c), we applied it to live cells and were able to observe the dynamics of the meshwork for at least 50 frames (Fig. 2d, e, Supplementary video 2).

Fast-MoNaLISA

Then, we aimed at improving the temporal resolution of MoNaLISA, which is key when imaging live-cells, since it determines the type of dynamics which is possible to monitor. Temporal resolution in all scanning microscopes is inherently limited by the number of scanning steps needed to record the desired image of the sample, which inversely scales with step size for a given region of interest (Fig. 3a). The optimal step size p is typically chosen

according to the Nyquist-Shannon theorem to avoid aliasing or redundant information, and it is twice the frequency limit of the system, which is roughly half of the expected resolution, thus $p \sim 20 - 30$ nm in RESOLFT imaging.

Current computational methods aiming at improving temporal resolution propose to predict new timepoints in between existing recording¹⁹, or to recover dynamics happening during a multi-shot imaging²⁰. However, they do not aim at reducing the recording time. On the hardware side, smart acquisition schemes have been successfully applied to point-scanning STED²¹ or RESOLFT²², but they rely on the fast response of single-point detector and are hardly applicable to parallelized, camera-based systems.

Here, we introduce a new method, Fast-MoNaLISA, to speed up the acquisition time by skipping scanning steps and then restoring subsampled images with an up-sampling network. This has the significant benefit of improving temporal resolution at the same illumination cost. Although a similar approach was applied to diffraction-limited confocal microscopy²³, it was never shown to work for sub-100 nm nanoscopy images. One specificity of our method is that we reduce the sampling while keeping the same OFF-switching dose, therefore still imprinting the same resolution in the sample.

Since each pixel of a reconstructed MoNaLISA image can be mapped to one specific scanning position, the training of the up-sampling network is straightforward (Fig. 3b, Supplementary Fig. S7). Starting from a Nyquist-sampled image, our GT, we apply a down-sampling step by removing a subset of pixels, thus creating a subsampled image, which is the input into the network. Importantly, this down-sampling step does not include any blurring, as is usual in common up-sampling network training setup. However, this means that the input's pixels will have the exact same values as some of the GT's pixels, including any noise present in the GT. Therefore, to prevent the network from learning the exact position of the original pixels, which would lead to a patterned noise (Supplementary Fig. S8), the pixel selection step is randomized. This makes it impossible for the network to reproduce the GT's noise, and leads to a built-in denoising, essentially a variation of a "Noise2Noise effect".

By tuning the proportion of removed pixels, we can choose the sampling ratio, that we denote S . As S gets smaller, the gain in imaging speed increases, but the restoration task gets harder, less information being provided to the network. We studied here three sampling ratios of $S = 75\%$, $S = 50\%$, and $S = 25\%$, respectively corresponding of speed increase of 1.3-fold, 2-fold, and 4-fold. For the network architecture, we chose the previously mentioned UNet-RCAN architecture, but keeping the last up-sampling layer of the original RCAN²⁴. RCAN was initially designed for up-sampling tasks, and we found that keeping the first U-Net network helps the network to converge.

We first tested our approach on the same structure used previously, fixed cells expressing vimentin-rsEGFP2, with an ideal SNR, corresponding to $> 80\%$ ON-switching level. As expected, the prediction accuracy on the finest structures improves when the sampling ratio S increases, especially from $S = 25\%$ to $S = 50\%$. This can be seen from the example line profile (Fig. 3c, insets), the relative FRC between the prediction and the GT (Fig. 3d), as well as peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) (Fig. 3e).

However, we found that higher sampling ratios are not always beneficial in the case of lower SNR (Supplementary Fig. S12). We hypothesize that this is due to a lower denoising effect for higher sampling ratios, which becomes more important when the SNR decreases. Importantly, the gain from $S = 50\%$ to $S = 75\%$ resulted to be minor or none for all SNR levels, suggesting that sampling half of the information is enough to restore subsampled images with the highest accuracy. For lower sampling ($S = 25\%$), the network still allows to recover most of the information, but there is more risk of error predictions, especially for low SNR inputs.

Since the goal of Fast-MoNaLISA is to image the rapid dynamics of trafficking molecules and organelles within cells, we hypothesized that imaging data containing the temporal evolution of the imaged cellular structure, or “temporal context”, could improve the accuracy of the predictions, especially for the lower sampling $S = 25\%$. To test this idea, we implemented a variation of our previous training, that we call multi-frames training, where we use a sub-sampled timelapse as an input instead of a single snapshot (Fig. 4a, Supplementary Fig. S9). The network is still trained to predict one single timepoint, the middle timepoint T_0 of the timelapse, but now learning additional information provided by the new temporal context.

We tested this multi-frames approach on mitochondria dynamics for different time lapse length, and saw significant improvement with $N=5$ frames, compared to the previous single frame training (Supplementary Fig. S9). Visually, we observed a global increase in image sharpness with the multi-frame predictions, especially when imaging mitochondrial-derived vesicles (MDVs), or mitochondria-vesicle contacts. This was confirmed by PSNR, SSIM and FRC curves (Supplementary Fig. S9). Additionally, the denoising effect is more accurate: small, dim vesicles are sometimes completely removed with the single-frame prediction, considered as noise, which is not the case for the multi-frames (Supplementary Fig. S9).

To demonstrate the utility of our approach, we used the Fast-MoNaLISA with $S = 25\%$ and a multi-frames prediction, to observe fast mitochondria dynamics. First, with 2D single plane imaging, we could image a full MoNaLISA FOV of $\sim 35 \times 35 \mu\text{m}^2$ at a speed of 2 Hz, 4 times faster than regular MoNaLISA. With mitochondria moving fast, reaching higher speed is key to avoid motion artefacts (Supplementary video 3). This allowed us to observe the temporary formation of small mitochondrial membrane protrusions elongating and retracting very rapidly, clearly happening too fast to be seen at lower frame time (Fig. 4b, Supplementary Fig. S10, Supplementary video 4). We characterized the length and dynamic of all elongating and retracting mitochondria movements for 3 different imaging speed: 2 Hz (fast-MoNaLISA), 1 Hz (regular MoNaLISA with smaller FOV for faster dwell time), and 0.5 Hz (regular MoNaLISA, full FOV). The 2 Hz imaging clearly allows us to detect, faster (1.5-4 s) and smaller (80-200 nm) protrusions that are completely missed in the other imaging speeds (Fig. 4c).

Mitochondria are spatially distributed throughout the whole 3D cell volume. Acquisition of images in different axial position as in 3D stacks are thus often needed to comprehensively monitor and resolve the whole network of organelles. Using the Fast-MoNaLISA, we could speed up the whole 3D mitochondrial network imaging compared to conventional recordings, thus enabling volumetric rate acquisition of up to 0.5 Hz, or even 1 Hz with a smaller FOV

(Fig. 4d). At this recording speed, we were able to observe a large variety of mitochondria dynamics including the formation of MDVs budding out and to follow their fate over time. MDVs have been shown to deliver specific cargoes to different subcellular locations, such as the endo-lysosomal system²⁵. More recently, it has been shown that MDVs carry functional ATP synthase and therefore were proposed as a potential way to deliver functional protein complexes to malfunctioning mitochondria²⁶. The ability of MDVs to fuse with existing mitochondria has not been demonstrated yet and is still under debate. Fast-MoNaLISA can help investigate the fate of MDVs and the proposed ability to exchange material within the mitochondrial network. We often detect vesicles contacting multiple mitochondria prior to fusions (Fig. 4e, h, Supplementary video 5), or a vesicle fusing in and budding-out a few seconds later (Fig. 4f). Many other examples are shown in Supplementary Fig. S11.

Vesicles dynamics need volumetric imaging to properly follow vesicle from the field of view and to differentiate between a vesicle and a 2D section of a mitochondrion's tip. However, without the Fast-MoNaLISA, the volume rate is too slow to follow the vesicles movements. We estimated that Fast-MoNaLISA enables the detection of at least 3 times more mitochondria-MDVs interactions with our Fast-MoNaLISA, compared to the regular one (Fig. 2g, Methods).

Discussion

In summary, we presented a new hybrid computational-experimental approach that is capable of providing comprehensive time-lapse imaging in 2D and 3D at high spatial (x-y ~ 60 nm) and temporal (~2 Hz) resolution of whole cells. Our methods add new imaging ability to a diverse set of structures from cytoskeleton to mitochondrial vesicles. By decreasing either the ON switching UV-Vis illumination doses or the number of scanning steps we monitored subcellular dynamics in extended temporal windows and volumetrically. Longer and faster time-lapse imaging added new opportunities for RESOLFT live cell nanoscopy opening up its application to dimer structures which could only be visualised as snapshot or fast-moving molecules which needs to be monitor in 3D for longer time.

Overall, the use of image computation in RESOLFT imaging is key to enable building up dynamic dataset of whole cells which can be used for both cell biological studies of selected proteins or to build dynamic model of cells by integrating high spatial information to the evolving temporal component.

We also believe the methods presented are applicable and can be beneficial for other imaging modalities, and we thus made the code publicly available with training examples (see code Availability below). HDN can be used to denoise any microscopy dataset and shows especially good performances to restore nanoscopy images. It also has the advantage of being usable both in a supervised and unsupervised manner in the same overall training structure. The subsampled image restoration can be beneficial to speed up any scanning-based imaging technique such as STED or spinning-disk confocal, and is especially attractive for camera-based detection, where the dwell-time is mainly limited by camera read-out times. Although its training method is supervised, it is important to note that the subsampled inputs are generated from regular Nyquist-sampled images, meaning a paired dataset acquisition is not necessary. Moreover, the

training set-up is applicable to any other networks and could potentially lead to even better results in the future, with the constant development of new deep-learning architectures.

Methods

Networks architectures

We aimed at keeping most of the code in the same code structure using PyTorch²⁷, to easily compare the performances of the different methods. This was possible for HDN, CARE and UNet-RCAN denoising task, as well as the upsampling task. This was not the case for SN2N and N2V, for which we used directly their original implementations (respectively <https://github.com/WeisongZhao/SN2N> and <https://github.com/juglab/n2v>).

HDN was implemented following the exact architecture from the original publication (<https://github.com/juglab/HDN>), which uses a Ladder Variational Auto Encoder (LVAE). We used 5 latent layers, with 6 residual blocks per layer, and a latent space dimension of 32 for each layer for all datasets.

CARE was implemented following <https://github.com/CSBDeep/CSBDeep> but adapted for PyTorch. We used a U-Net with 4 downsampling layers, and 64 initial convolutional filters.

For the UNet-RCAN denoising, we started from the original PyTorch RCAN implementation (<https://github.com/yulunzhang/RCAN>), and modified it according to the UNet-RCAN code (https://github.com/vebrahimi1990/UNet_RCAN_Denoising). We used a U-Net with 3 downsampling layers and 64 initial convolutional filters. For the RCAN we used 64 initial convolutional filters, 8 Residual Groups, and 12 Residual Channel Attention Blocks (RCAB). The usual skip connections of the U-Net were replaced by Channel Attention Blocks (CAB), as described in the UNet-RCAN implementation.

For the upsampling task, we used the same U-Net and UNet-RCAN network architectures (without the CAB skip connections for the U-Net), but with an additional upsampling block at the end, consisting of a transposed convolution, an activation layer, and a final convolutional block.

HDN training and inference

The training of HDN necessitates a probabilistic noise model $p(\mathbf{x}|\mathbf{s})$, which describes the conditional probability of observing a noisy value $\mathbf{x}$ for a given, clean signal $\mathbf{s}$ (see Noise Model generation below). From this, the network is trained by optimizing the loss, which is a weighed sum of two loss functions:

$$L = L_{rec} + \beta \cdot L_{KL}$$

L_{rec} is the reconstruction loss, which evaluates how likely is the generated image, considering the noise model. For unsupervised training, it is calculated from the log-likelihood between the prediction and the input. For supervised training, it can be adapted by calculating the log-likelihood between the prediction and the ground-truth. L_{KL} is the Kullback–Leibler (KL) divergence, which enforces the latent space distribution to be close to a Normal distribution.

The weight β was adjusted empirically to get the best spatial resolution (See Supplementary Fig. S2). For unsupervised training, it was empirically set to between $\beta=0.5$ and $\beta=0.4$, depending on the dataset. For supervised training, the weight is much lower, typically $\beta=0.001$. The full parameters for all HDN trainings can be found in Supplementary Table 1.

For inference, we always generate multiple samples and average them to get the final prediction result. For unsupervised training, we always used 50 samples. For supervised training, the lower weight on the L_{KL} means the samples diversity will significantly be reduced. We thus reduced the number of samples to 10 for HDN^{sup} predictions.

Other Networks Training

For CARE training, we used the MSE loss, as used in the original implementation. For all other supervised trainings, we used the loss function used in the original UNet-RCAN denoising, the weighted summation of Charbonnier loss and Edge loss ($L = L_{char} + \alpha L_{edge}$), with $\alpha = 0.05$. This loss function has been shown to prevent the smoothing effect that happens when training with the MSE loss¹⁷. We observed benefits from using this function both with denoising and upsampling tasks. Other training parameters, such as learning rates, batch size, epochs, etc. can all be found in the Supplementary Table 1.

Noise Model Generation

The noise model $p(x|s)$ describes the conditional probability of observing a noisy value x for a given clean signal value s . This can be built experimentally from a paired dataset of clean-noisy images. Briefly, this is done by constructing a 2D histogram (e.g. 256x256 bins) with the clean signal values on the y-axis, and the noisy observations on the x-axis. Each row j , once normalized, thus gives the probability $p(x|s_j)$. This histogram is then parametrized with a Gaussian Mixture Model (GMM).

Although this step requires a paired dataset, it is important to note that it is only necessary to acquire it once for every imaging system and can then be used for any other HDN training, keeping the method essentially unsupervised.

Paired Datasets for Denoising

To obtain a paired dataset with clean and noisy images, we fixed the cells (see Cell Culture below). Note that the fixation limits the number of cycles of the rsFPs, meaning photobleaching will happen faster than in live cells. However, it is still possible to acquire high-quality RESOLFT images.

For the vimentin dataset, the noisy images were first acquired at ~10-20 % ON-switching (~30W/cm²). Then, 5 images were acquired with ~80-90 % ON-switching (~300W/cm²). During this pseudo-timelapse, the signal will decrease rapidly. We used that signal decrease to evaluate HDN and HDN^{sup} results at different SNR levels (Supplementary Fig. S1), and to evaluate the upsampling predictions for the different SNR levels (Supplementary Fig. S12). To create the best ground-truth possible, the first 3 images of the pseudo-timelapse were averaged.

For the actin meshwork dataset, we used the same procedure, but the first image was acquired at ~50% ON-switching. The next 5 were acquired at ~90% ON-switching and averaged to create a high-quality ground-truth.

Training Datasets Preparation

For each dataset, we first split the dataset in training, validation, and testing splits (See Supplementary Table 1 for datasets sizes of each training). We then extracted 64x64 or 128x128 patches and removed patches containing less information. This was done with a procedure previously described¹⁷, which consists of normalizing the L2 norm of each patch to the maximum L2 norm of the whole training dataset and discarding the patches with a normalized L2 smaller than 0.1.

For the upsampling task, we then generate the subsampled patches. For the single-frame static training (Fig. 3b, Supplementary Fig. S7), we generate the 4 possible subsampled patches by simply selecting 1 pixel every 2 in each direction, with the 4 possible starting points: (x=0,y=0), (x=1,y=0), (x=0,y=1), (x=1,y=1). For $S = 25\%$ we choose randomly 1 out of the 4 patches. For $S = 50\%$ we choose randomly 2 of the 4 patches and shuffled their order randomly. For $S = 75\%$, we choose randomly 3 of the 4 patches and shuffled their order randomly.

For the multi-frames temporal training (Fig. 4a, Supplementary Fig. S9), each acquired timelapses is first split into smaller timelapses of length N , before the patch extraction. Then, for each timepoint of the small timelapse-patch, we generate the subsampled patch as described above for $S = 25\%$.

Quantitative evaluation of the prediction

For resolution assessment, we traced $N=30$ line-profiles on vimentin filaments on FIJI, with a thickness of 3 pixels. We fitted them with a Lorentzian function with a custom MATLAB script. For the FRC comparative measurements, we used the single image FRC (*IFRC*) method, which we validated to give similar results than two images FRC (*2FRC*) on our data. For PSNR and SSIM metrics, we used the metrics module of the *scikit-image* library in Python.

MoNaLISA imaging

All images were acquired on one of the two custom-built MoNaLISA setups we have. Both setups follow the original implementation, with a few minor differences. The first setup has an OFF-period pattern $P_{OFF} = 238$ nm and multi-foci pattern $P_{MF} = 714$ nm. The second setup has an OFF-period pattern $P_{OFF} = 312$ nm and a multi-foci pattern period $P_{MF} = 625$ nm. Images were acquired on both setups interchangeably, without a systematic preference for one over the other. We used ImSwitch²⁸ (<https://github.com/ImSwitch>) for hardware control.

SNR quantification

The SNR quantification was done from line profile measurements, by taking the ratio of the peak value of the gaussian over the standard deviation of the background (Supplementary Fig. S5, S6).

Image Analysis of Mitochondria dynamics

For the protrusion length and duration analysis (Fig. 4c), we manually measured in FIJI the maximum protrusion length, and counted over how many frames it lasted, for all back-and-forth movements we could observe and detect by eyes. This was done on 5 cells for each considered speed.

For the MDV-mitochondria interactions number of events (Fig. 4g), we counted each time we can follow a vesicle and follow its fate with relative certainty. The number of events for “Regular” imaging was estimated by keeping only 1 volume for every 4 of the “Fast” imaging. This is only an estimation, because it does not consider the additional scanning artefacts of the slower scanning, which makes the interpretation of the mitochondria-MDV contacts even harder and is thus probably overestimating the number of events for the regular imaging.

Cell Culture

Wild-type U2OS (ATCC HTB-96) cells and U2OS with endogenously tagged vimentin-rsEGFP2 (ref. Monalisa) were cultured in DMEM (Thermo Fisher Scientific; 41966029) supplemented with 10% (vol/vol) FBS (Thermo Fisher Scientific; 10270106) and 1% penicillin-streptomycin (Sigma Aldrich; P4333) and maintained at 37 °C and 5% CO₂ in a humidified incubator. For transfection with LifeAct-rsEGFP2 and rsEGFP2-Omp25 plasmids (Ref. Monalisa), 2 × 10⁵ cells per well were seeded on coverslips in a six-well plate. After 1 day, cells were transfected using Lipofectamine LTX Reagent with PLUS reagent (Thermo Fisher Scientific; 15338100) according to the manufacturer’s instructions. 24–36 h after transfection, cells were washed in PBS solution, placed with phenol-red-free DMEM or Leibovitz’s L-15 medium (Thermo Fisher Scientific; 21083027) in a chamber, and imaged.

For fix imaging experiments, cells were washed once with PBS and then fixed in 4% paraformaldehyde (PFA) in phosphate buffered saline (PBS) (pH7.4) at room temperature for 20 minutes.

Data availability

Training datasets, trained models and data presented are available in the following Zenodo repository: <https://doi.org/10.5281/zenodo.17132698>.

Code availability

The code is available at the GitHub repository:
https://github.com/GuillaumeMinet/deepL_resolft.

Corresponding Author

Ilaria Testa – SciLifeLab and KTH Royal Institute of Technology, 17165, Solna, Sweden;
Email: ilaria.testa@scilifelab.se

Author Contributions

I.T. conceived and supervised the project. F.J. supervised the project and developed the neural networks. G.M. collected the training data, implemented the code, conducted the experiments, and performed the data analysis. A.R. provided guidance on coding, the design of the training setup, and fine-tuning of the training parameters. F.P. contributed guidance across all aspects of the project. G.C. prepared the biological samples and provided biological expertise. I.T. and G.M. wrote the manuscript with input from all authors.

Acknowledgment

We thank the DL@MBL: Deep Learning for Microscopy Image Analysis Course organized at the Marine Biological Laboratory for helping in the start of this project. This work was supported by EU ERC InsPiRe (101002490) and CZI Chan Zuckerberg Initiative, Dynamic Imaging, iNano 2023 (2303-2508) to IT.

References

1. Sahl, S.J., Hell, S.W. & Jakobs, S. Fluorescence nanoscopy in cell biology. *Nat Rev Mol Cell Biol* **18**, 685-701 (2017).
2. Grotjohann, T. et al. rsEGFP2 enables fast RESOLFT nanoscopy of living cells. *Elife* **1**, e00248 (2012).
3. Brakemann, T. et al. A reversibly photoswitchable GFP-like protein with fluorescence excitation decoupled from switching. *Nat Biotechnol* **29**, 942-947 (2011).
4. Habuchi, S. et al. Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa. *Proc Natl Acad Sci U S A* **102**, 9511-9516 (2005).
5. Pennacchietti, F. et al. Fast reversibly photoswitching red fluorescent proteins for live-cell RESOLFT nanoscopy. *Nat Methods* **15**, 601-604 (2018).
6. Hell, S.W., Jakobs, S. & Kastrup, L. Imaging and writing at the nanoscale with focused visible light through saturable optical transitions. *Applied Physics A* **77**, 859-860 (2003).
7. Hofmann, M., Eggeling, C., Jakobs, S. & Hell, S.W. Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proc Natl Acad Sci U S A* **102**, 17565-17569 (2005).
8. Grotjohann, T. et al. Diffraction-unlimited all-optical imaging and writing with a photochromic GFP. *Nature* **478**, 204-208 (2011).
9. Masullo, L.A. et al. Enhanced photon collection enables four dimensional fluorescence nanoscopy of living systems. *Nat Commun* **9**, 3281 (2018).
10. Boden, A. et al. Volumetric live cell imaging with three-dimensional parallelized RESOLFT microscopy. *Nat Biotechnol* **39**, 609-618 (2021).
11. Casas Moreno, X. et al. Multi-foci parallelised RESOLFT nanoscopy in an extended field-of-view. *J Microsc* **291**, 16-29 (2023).
12. Marín-Aguilera, G. et al. All-Optical Strategies to Minimize Photo-Bleaching in Reversibly Switchable Fluorescent Proteins. *bioRxiv*, 2025.2001.2020.633912 (2025).
13. Waldchen, S., Lehmann, J., Klein, T., van de Linde, S. & Sauer, M. Light-induced cell damage in live-cell super-resolution microscopy. *Sci Rep* **5**, 15348 (2015).
14. Krull, A., Buchholz, T.O. & Jug, F. in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2124-2132 (2019).
15. Qu, L. et al. Self-inspired learning for denoising live-cell super-resolution microscopy. *Nat Methods* **21**, 1895-1908 (2024).

16. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat Methods* **15**, 1090-1097 (2018).
17. Ebrahimi, V. et al. Deep learning enables fast, gentle STED microscopy. *Commun Biol* **6**, 674 (2023).
18. Mangal, P., Mauricio, D., Peyman, M. & Florian, J. in International Conference on Learning Representations (2022).
19. Priessner, M. et al. Content-aware frame interpolation (CAFI): deep learning-based temporal super-resolution for fast bioimaging. *Nat Methods* **21**, 322-330 (2024).
20. Cao, R., Divekar, N.S., Nunez, J.K., Upadhyayula, S. & Waller, L. Neural space-time model for dynamic multi-shot imaging. *Nat Methods* **21**, 2336-2341 (2024).
21. Heine, J. et al. Adaptive-illumination STED nanoscopy. *Proc Natl Acad Sci U S A* **114**, 9797-9802 (2017).
22. Dreier, J. et al. Smart scanning for low-illumination and fast RESOLFT nanoscopy in vivo. *Nat Commun* **10**, 556 (2019).
23. Fang, L. et al. Deep learning-based point-scanning super-resolution imaging. *Nat Methods* **18**, 406-416 (2021).
24. Zhang, Y. et al. in Proceedings of the European Conference on Computer Vision (ECCV) (2018).
25. Soubannier, V., Rippstein, P., Kaufman, B.A., Shoubridge, E.A. & McBride, H.M. Reconstitution of mitochondria derived vesicle formation demonstrates selective enrichment of oxidized cargo. *PLoS One* **7**, e52830 (2012).
26. Hazan Ben-Menachem, R. et al. Mitochondrial-derived vesicles retain membrane potential and contain a functional ATP synthase. *EMBO Rep* **24**, e56114 (2023).
27. Paszke, A. et al. in Proceedings of the 33rd International Conference on Neural Information Processing Systems Article 721 (Curran Associates Inc., 2019).
28. Casas Moreno, X., Al-Kadhimi, S., Alvelid, J., Bodén, A. & Testa, I. ImSwitch: Generalizing microscope control in Python. *Journal of Open Source Software* **6**, 3394 (2021).

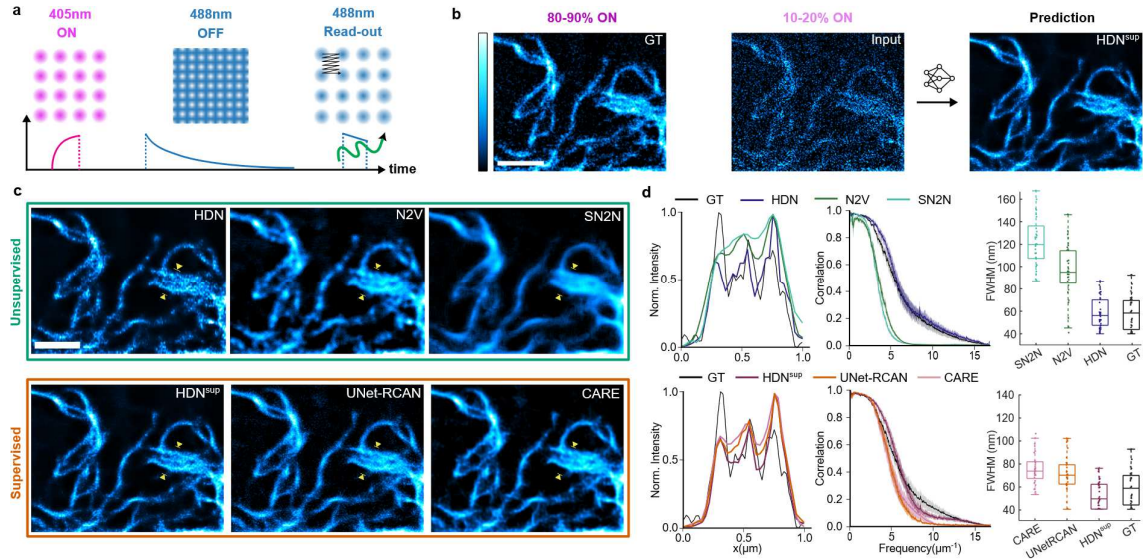


Figure 1. Denoising of low ON-switching MoNaLISA images.

a Sequence of MoNaLISA imaging. Black arrows shown between 488 nm read-out foci represent the scanning pattern used during imaging. **b** MoNaLISA imaging of fixed U2OS cells endogenously expressing vimentin-rsEGFP2 for: 80-90 % ON-switching (left), 10-20 % ON-switching (middle), and 10 – 20 % ON-switching + denoising with our proposed method. Scale bar, 1 μm . **c** Denoised predictions of 10 – 20 % ON-switching MoNaLISA for HDN, N2V and SN2N as unsupervised methods (top row) and HDN^{sup}, UNet-RCAN and CARE as supervised ones (bottom row). Scale bar, 1 μm . **d** Resolution analysis of denoising results for unsupervised (top row) and supervised methods (bottom row). Line profile example with normalized intensity (left column), single-image Fourier Ring Correlation (middle column), and FWHM fit of line profiles of 30 filaments.

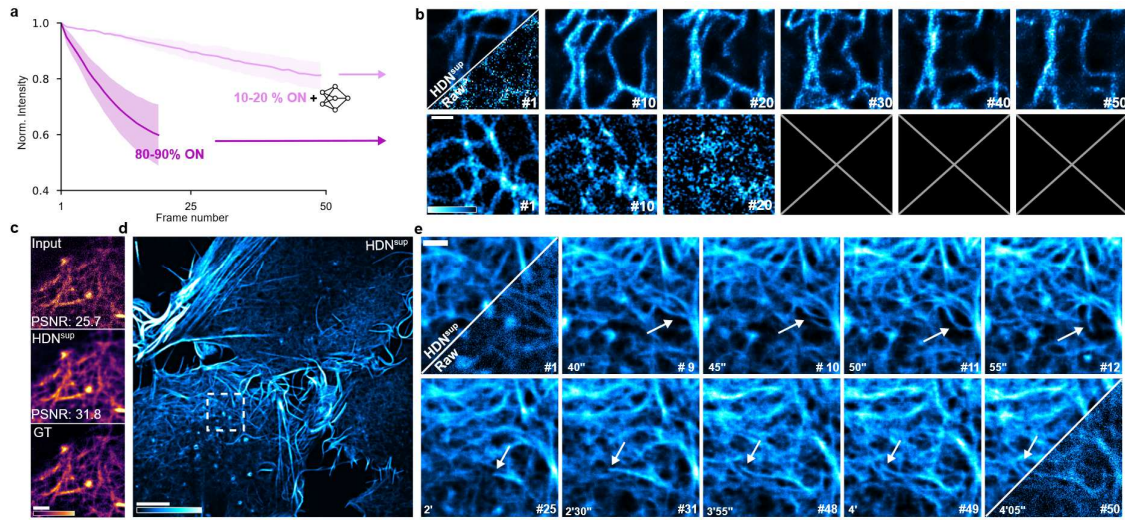


Figure 2. Prolonged MoNaLISA imaging.

a Photobleaching profile for 80 – 90% ON-switching (dark magenta) and 10 – 20% ON-switching (light magenta) imaging of U2OS cells endogenously expressing vimentin-rsEGFP2 (mean $\pm$ std, $N = 5$ cells). **b** Timelapse imaging with 10 – 20 % ON-switching (top row) and 80 – 90% ON-switching (bottom row). The 80 – 90% ON-switching timelapse was stopped after the 20th frame because of the photobleaching. Scale bar, 500 nm. **c** Representative raw input image, HDN^{sup} prediction, and corresponding GT image of the Actin meshwork in fixed U2OS cells expressing LifeAct-rsEGFP2. Input was acquired with ~ 50 % ON-switching. GT is an average of 5 frames at ~ 90 % ON-switching. Scale bar, 1 μm . **d** HDN^{sup} prediction of one full FOV. Scale bar, 5 μm . **e** Magnified region (see inset h) showing the actin meshwork evolving over 50 frames, acquired every 5 s. White arrows highlights meshwork movements. For the first and last frame, the lower right corner shows the raw image data. Scale bar, 1 μm .

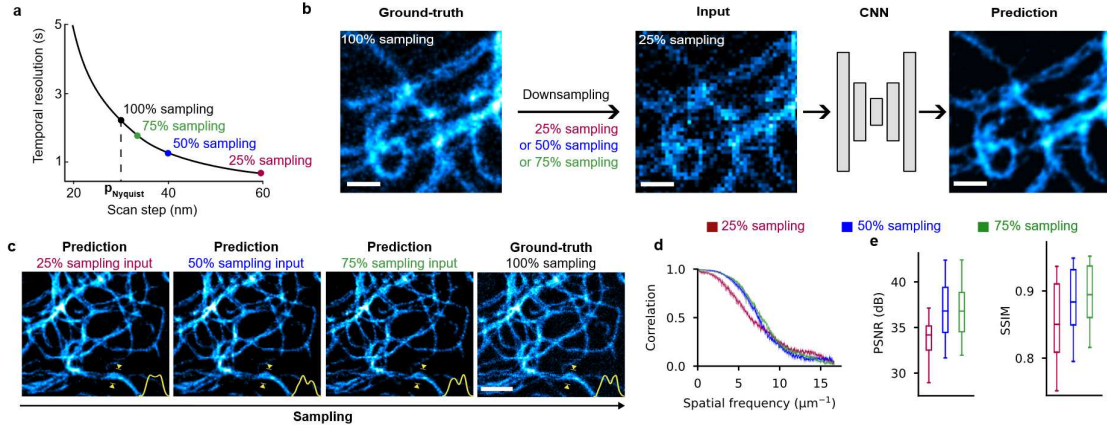


Figure 3. Fast-MoNaLISA – single frame training and results.

a Temporal resolution dependency over the scanning pixel size. Frame rate calculated for a multi-foci period of 625 nm and a dwell time of 5 ms. **b** Training setup for Fast-MoNaLISA. Starting from a Nyquist-sampled image (the GT), we select a subset of pixels to create a subsampled input of a given sampling ratio S (75%, 50% or 25%). **c** Fast-MoNaLISA results on U2OS cells endogenously expressing vimentin-rsEGFP2, across sampling ratios and SNR levels. Quantification of the SNR levels are indicated in the GT images. The line profiles across the white arrows in the top row images is reported in orange. Scale bar, 1 μm . **d** Relative FRC between predictions and GT (mean $\pm$ std, $N = 3$ cells). **e** PSNR and SSIM performances for all sampling ratios and SNR levels.

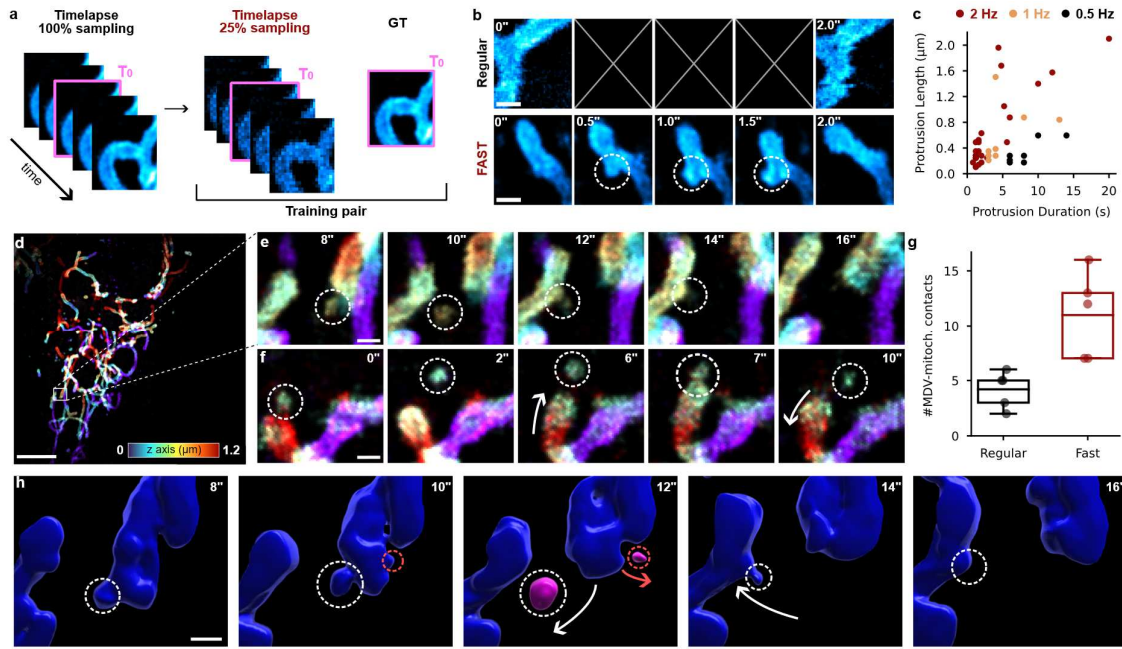


Figure 4. Multi-frames Fast-MoNaLISA applied to fast mitochondria imaging.

a Multi-frames training concept: a subsampled ($S = 25\%$) timelapse with 5 consecutive frames is used to restore the centre timepoint T_0 . The surrounding timepoints give temporal context and help the restoration task. **b** Regular vs Fast-MoNaLISA imaging examples of mitochondria in U2OS cells expressing rsEGFP2-Omp25. Regular MoNaLISA images are raw images acquired at 0.5 Hz (empty grey rectangles with the cross correspond to the timepoints not accessible at this imaging speed). Fast images are multi-frames predictions of $S = 25\%$ sampled images acquired at 2 Hz. White dotted circles highlight a fast protrusion happening during the time acquisition of the slow modality. Scale bar, 500 nm. **c** Characterization of the protrusion's lengths and duration for 3 imaging speeds: 2 Hz (red), 1 Hz (orange), 0.5 Hz (black). **d** Full FOV of fast volumetric recordings of mitochondria, Z axis color-coded. Scale bar, 10 μm . **e** Zoomed-in timelapse region of **d** showing a MDV budding-out and contacting another mitochondrion. **f** Zoomed-in region from a different cell showing MDV re-contacting same mitochondrion after several seconds. Scale bar, 500 nm. **g** Number of mitochondria-MDVs interactions seen per cell with Regular and Fast-MoNaLISA, $N = 5$ cells. **h** Volumetric modelling of region shown in **e**. White circles and arrows highlight the path of one MDV, while red circles and arrow follow a second vesicle. Top-down view, scale bar, 500 nm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [1prolongedmitochondriaimaging.mp4](#)
- [2actinmeshworkdynamics.mp4](#)
- [3mitochondriamotionartefacts.mp4](#)
- [4mitochondrialfastdynamics2d.mp4](#)
- [5volumetricmitochondriavesiclesbudding.mp4](#)
- [Supplementary.pdf](#)