Enabling Real-Time Fluctuation-Based Super Resolution Imaging

Miyase Tekpınar[1], Jelle Komen[1], Hana Valenta[3], Ran Huo[1], Klarinda de Zwaan[4], Peter Dedecker[3], Nergis Tomen[2], Kristin Grußmayer[1*]

[1*]Department of Bionanoscience and Kavli Institute of Nanoscience, Delft University of Technology, Delft, Netherlands.
[2]Department of Intelligent Systems, Delft University of Technology, Delft, Netherlands.
[3]Department of Chemistry, KU Leuven, Leuven, Belgium.
[4]Biomedical Engineering, Nanoscopy for Nanomedicine, TU Eindhoven, Eindhoven, Netherlands.

*Corresponding author(s). E-mail(s): k.s.grussmayer@tudelft.nl;
Contributing authors: m.tekpinar@tudelft.nl; j.j.m.komen@tudelft.nl; hana.valenta@kuleuven.be; r.huo@tudelft.nl; b.k.d.zwaan@tue.nl; peter.dedecker@kuleuven.be; n.tomen@tudelft.nl;

Abstract

Live-cell imaging captures dynamic cellular processes, but many structures remain beyond the diffraction limit. Fluctuation based super-resolution techniques reach beyond the diffraction limit by exploiting correlations in fluorescence blinking. However, existing methods require the acquisition of hundreds of frames and involve computationally intensive post-processing, which can take tens of seconds, thereby limiting the suitability for real-time sub-diffraction imaging of fast cellular events. To address this, we use a recurrent neural network model, which integrates sequential low-resolution frames to extract spatio-temporally correlated signals. By creating a synthetic dataset consisting of blinking emitters and using super-resolution optical fluctuation imaging (SOFI) to reconstruct our super-resolution targets, we developed a deep-learning based real-time super-resolution fluctuation imaging method (RESURF). Our method significantly improves temporal resolution by reducing the required number of frames down to only 8 frames and also doubles the spatial resolution. By testing on different datasets, we show that our method offers the flexibility of training with either synthetic or experimental dataset and can generalize to different structures. We also demonstrate a 400-fold reduction in computational latency compared to SOFI, by achieving inference time less than 30 ms, highlighting its promise both as an efficient high-throughput imaging and real-time solution for live-cell super-resolution imaging.

Keywords: Real-time super-resolution imaging, low latency, AI, convolutional gated recurrent unit (convGRU), fluctuation microscopy, high-throughput imaging

Live-cell fluorescence imaging captures dynamic cellular behaviors and aims to maximize both spatial and temporal resolution while minimizing sample damage [1], enabling advancements in fundamental cell biology. However, visualizing these cellular dynamics is constrained by the diffraction limit, which restricts the spatial resolution to around 200-300 nm. Super-resolution (SR) techniques can overcome this barrier and achieve higher spatial resolutions. Well-known techniques include Stimulated Emission Depletion (STED) microscopy [2], Structured Illumination Microscopy (SIM) [3], and Single Molecule Localization Microscopy (SMLM) [4, 5]. While STED and SIM require a complex microscopy setup, SMLM can be performed with a wide-field microscope. By using the stochastic and sparse blinking of fluorophores, which they switch between bright "on-" state and dark "off-" state, individual molecules are localized from diffraction-limited image sequences. Subsequently, their positions are used to generate a super-resolution image with a spatial resolution of a few tens of nanometers. However the reconstruction of a super-resolution image requires many thousands of frames, typically corresponding to data acquisition over minutes to hours. This elongated experimental duration leads to relatively low temporal resolution and moderate speeding-up of the imaging usually entails higher laser illumination. Thus, this prevents observation of fast processes and can lead to high phototoxicity [6, 7]. On the other hand, fluctuation-based super-resolution techniques, which also utilize blinking emitters, typically require only hundreds of frames to reconstruct an SR image as they can handle high densities of blinking fluorophores [8]. Widely used examples of fluctuation-based methods are super-resolution optical fluctuation imaging (SOFI) [8–10] and SRRF (Super-Resolution Radial Fluctuations) [11, 12].

SOFI leverages the correlation of blinking from a fluorophore over time and space and the independence of different fluorophores [9, 13]. Dertinger et al. showed that n^{th} order cumulants of a diffraction limited image sequence of blinking emitters result in a point spread function (PSF) raised to the power of n, which brings a $\sqrt{n}$-fold better resolution. This resolution enhancement can be further increased to n-fold in n^{th}-order SOFI by applying linearization and deconvolution or Fourier reweighting after the cumulant calculation [14]. Higher orders require better statistics, thus more frames and longer acquisition time than second-order SOFI. This can be challenging because of the limited signal-to-noise ratio (SNR) as a result of the bleaching of the sample. Therefore, there have been previous attempts to reduce the number of frames. One such approach is SACD (super-resolution imaging based on autocorrelation with two-step deconvolution), which applies background subtraction and deconvolution as a pre-processing step to conventional SOFI to reduce noise and improve signal quality prior to cumulant analysis [13]. Another common technique is SRRF (Super-Resolution Radial Fluctuations), which operates by computing gradient maps and applying a radiality transform that exploits the radial symmetry of the PSF. A final cumulant-based analysis of the image sequence produces a super-resolved, noise-suppressed image. The enhanced version, eSRRF, further assists users in selecting parameter settings that minimize the risk of artifacts [12].

Even-though the above-mentioned methods require shorter imaging time than SMLM, their temporal resolution are typically in the range of seconds to minutes. Under low laser illumination necessary for prolonged imaging and thus for low SNR conditions, they require longer measurements to reach high SNR image reconstruction. In addition, the SR image is mostly reconstructed offline, after the experiment and the computational latency of these methods is limited to the range of seconds even with GPU support. Therefore they are not suitable for long-duration, real-time, live cell imaging. In this study, we aim to offer a deep-learning model to reconstruct a super-resolution image in real-time from few, ultra low-SNR microscopy images by using deep learning.

Deep learning has been extensively explored to increase image quality and imaging speed [15]. Most methods employ the U-Net architecture [16], a type of convolutional neural network (CNN), because of its high performance in image segmentation, denoising, and deconvolution[17–21]. Despite the growing number of studies that apply deep learning in microscopy, few models focus on extracting temporal information. A notable study called deformable phase-space alignment

(DPA) time-lapse image SR (TISR) neural network, used a convolutional recurrent feature propagation to generate time-lapse SIM imaging. By using phase-space alignment in addition to frequency domain, they achieved the aggregate of information from sequential frames with an inference time of a few hundreds ms [22]. However, the method is still limited by the acquisition speed of SIM imaging which is in the range of hundreds of milliseconds to seconds per frame. As in most deep-learning-based SR techniques, DPA-TISR also uses an experimental dataset for training, preventing it from being widely used for low-SNR and dynamic imaging conditions. In addition to CNNs, recurrent neural networks (RNNs) are rarely used in microscopy. Nevertheless, some studies have shown the potential of RNNs to accelerate SMLM image reconstruction [23, 24]. An example of RNN, specifically, a long-short-term memory (LSTM) network, based SR technique is DBlink for video reconstruction from SMLM data [25]. By feeding localization maps into the LSTM, DBlink generates super-resolution videos at rates of up to 66 frames per second (FPS). However, this method remains constrained to SMLM sample conditions and relies on localization-based post-processing, restricting it use to off-line image analysis.

In this study, we propose to use the multi-image super-resolution gated recurrent unit (MISRGRU) architecture to achieve real-time fluctuation-based super-resolution (RESURF) [26]. Originally designed to predict single super-resolution images from multiple low-resolution images in remote sensing, MISRGRU utilizes a convolutional gated recurrent unit (convGRU), a powerful RNN architecture capable of leveraging both temporal and spatial information, making it compatible with any fluctuation-based super-resolution technique. MISRGRU is flexible in that it supports training and inference with a varying number of input images. By design, the output can be easily interpolated to any image size, unlike for the commonly used U-Net which requires pre-processing. This allows for straightforward training to produce n-fold larger output images, facilitating n-fold resolution improvement.

Here, we first create a foundation model by training MISRGRU with large synthetic data, where we have full access to ground truth structures for evaluation. We aim to utilize the foundation models for transfer learning of new models and also as a baseline to evaluate the performance of future models. Our synthethic dataset features microtubule-like structures, dynamic out-of-focus background and a realistic sCMOS camera noise model. In this study, we focus on reconstructing second-order SOFI images as the target output to limit computational demands and maximize imaging speed. We tested different input frame counts for the network to determine the model performance and the temporal resolution. Compared to a hundred or more frames typically required for second-order SOFI, our model achieved a two-fold spatial resolution improvement for simulated microtubules using significantly fewer frames and reduced computational latency. We also showed the generalization of the model to different patterns using both simulated and experimental data.

Furthermore, we demonstrate that we can optimize the foundation models via transfer learning for a range of live-cell and fixed-cell data labeled with different blinking fluorescent proteins and organic dyes. Using realistic simulations with the parameters close to experimental data, the network out-performed SOFI and eSRRF in our evaluation metrics, in different structures labeled by fluorescence proteins. We also showed that the performance of the models could easily be reproduced by using an experimental image sequence to retrain instead of simulations. This allows users flexibility in training strategies that can be conveniently adapted to facilitate dynamic and high-throughput experiments.

Results

Enabling real-time super-resolution

SOFI typically requires several hundreds of frames (100-500) depending on fluorophore blinking characteristics and the signal-to-noise ratio (SNR) to reach its theoretical two-fold resolution enhancement and to produce high-SNR second-order SOFI reconstructions [27]. However, this frame count and the associated computational time limit SOFI's ability to achieve high temporal resolution for real-time super-resolution sample screening. To address this limitation, we leverage deep learning to reduce the number of required input frames

and enable real-time analysis, while maintaining image quality comparable to high-SNR SOFI reconstructions.

The deep-learning network architecture is based on the MISRGRU model, which comprises an encoder to extract feature representations of low-resolution frames, followed by a convolutional gated recurrent network (convGRU) based fusion layer to process the features sequentially to capture the correlations between input images. Global average pooling (GAP) combines them into a single representation, and a decoder reconstructs the high-resolution image by upsampling the fused feature maps. The input of the network consists of a time series of diffraction-limited frames, the target output being second-order SOFI reconstructions of the structures, as illustrated in Figure 1. To validate the network, we created a realistic synthetic dataset by generating structural patterns resembling microtubules randomly decorated with fluorophores at a certain density. We simulated fluorophore blinking dynamics, including on- and off-states, and incorporated an sCMOS camera noise model along with a non-stationary background to closely mimic experimental imaging conditions (see Methods). The dataset includes a range of SNR levels (photon number/frame: 40,60,80,100) and includes varying off-times (600, 1200, 2000 ms), while maintaining a constant average on-time of 10 ms to mimic gentle, low SNR live-cell conditions labeled by fluorescence proteins. For generating target images, we used 100 diffraction-limited frames to compute second-order cross-cumulants, followed by linearization and Richardson-Lucy deconvolution for post-processing [14]. These SOFI images were then employed as targets during network training. Following an approach similar to Chen et al., we trained separate models using 8, 15, 20, and 25 input frames to assess the network's ability to reconstruct super-resolved images under different temporal constraints[21]. We also tested both mean-absolute error (L1) and Fourier loss (F) and showed that Fourier loss performs better in reaching higher RSP scores for slow-blinking low-SNR conditions (Supplementary Figure 1).

We determined the performance of the models with different input frames by applying a set of established evaluation metrics, as described in more details in the Performance evaluation part of the Methods section. We estimated the resolution of predicted super-resolution (SR) images using decorrelation analysis, a widely adopted technique in super-resolution microscopy for resolution quantification [28]. We measured the similarity between the predicted and ground-truth (GT) or wide-field (WF) images using resolution-scaled Pearson correlation coefficient(RSP) and resolution-scaled error (RSE), a standard method in SR imaging for quantitative comparison of image quality [29, 30]

In figure 1b, we show the resolution and RSP scores of the models with different number of inputs and compare it to the SOFI method. All models approximate the theoretical two-fold spatial resolution improvement of second-order SOFI. The RSP scores reveal that all models surpass SOFI in preserving structural integrity. While the Pearson correlation coefficient slightly decreases with fewer input frames, we find that the 8-frame model offers a favorable trade-off between temporal resolution and spatial accuracy. We also optimized the network to limit the computational latency by decreasing the size of the model by reducing the number of hidden channels. In figure 1c, we show that the RSP and resolution scores change only slightly for the reduced model size and the latency can be reduced down to 27 ms for an image size of 512x512 pixels. Despite the inherent complexity of sequential data processing in recurrent neural networks, the optimized MISRGRU model achieved a significantly lower end-to-end latency of 27 ms. This represents an approximate 400-fold acceleration in super-resolution reconstruction compared to the SOFI method. These findings highlight the strong potential of MISRGRU for real-time, live-cell super-resolution imaging.

Visual examples of reconstructions under varying SNR levels and blinking conditions are provided in Figure 2. As expected, under the slowest blinking condition (Figure 2a), Model-8f misses some structures, simply because of the limited number of emitter switching to on-state in this short period of time. As we simulated the images with a Gaussian PSF and with FWHM of 220 nm, we first calculated the resolution of noise-free wide-field images (260 nm) and assigned the half of this value (130 nm) for the two-fold resolution improvement goal. However, image decorrelation

analysis can be affected by the noise and background levels. Therefore, in Figure 2b, we see very high resolution values in wide-field images of slow blinking conditions compared to fast-blinking conditions with worse SNR and SBR. SOFI reconstructions using 100 frames fail to achieve a two-fold resolution improvement under low SNR and slow blinking conditions. On the other hand, Model-20f and Model-8f maintain the resolution improvement under different conditions. This showed that by being trained with a simulation dataset varied in SNR levels, the network learned to extract only the meaningful, correlated signal between the subsequent frames but not the uncorrelated noise. In Figure 2c, we tested the resolution-scaled Pearson correlation (RSP) scores of super-resolution reconstruction by taking both ground-truth and averaged wide-field frames as a reference in two separate analysis. The difference between the two analysis showed us that stationary background pixels in wide-field images degrade the RSP scores if the SR reconstruction performs a good background elimination. Wide-field images have very poor SNR levels under slow blinking low SNR conditions. As a result, the RSP estimations using wide-field images resulted in lower scores than the ones using ground-truth as a reference. In the latter case, we can see the high performance of both models under extreme noise conditions (Figure 2c) compared to SOFI, more aligned with the images shown in Figure2a. This offers a practical benefit for real-time super-resolution fluctuation (RESURF) live-cell imaging by tolerating lower laser intensities, thereby reducing phototoxicity and extending image acquisition durations.

In Supplementary Figure 1, we showed the performance of the network for the SNR conditions in and out of range of the training set. Models achieve generalization to different, higher and extremely low SNR levels which were not seen during training. Another model for higher SNR levels than the tested conditions can be easily retrained with a small dataset (see Transfer learning with small experimental dataset). In Supplementary Figure 5a, we evaluated the dependence of the model on the size of the training data set and showed that the model converges to the minimum training error already around the training set size of 2000. In addition, we also tested the model's performance for different combinations of training datasets (k-fold cross validation analysis) and showed that models have high consistency, showing very similar scores in RSP and resolution analysis (Supplementary Figure 5c). Furthermore, we tested the performance of the models for different image sizes and demonstrated that model performance does not degrade for increasing image sizes, even though it is only trained using images with a size of 128x128 pixels (Supplementary Figure 6). This offers a widely-applicable method for a large variety of typical imaging conditions using modern scientific cameras without additional effort for retraining of the model.

RESURF models generalize from simulations to different structures in live cells labelled by fluorescence proteins

Considering the dynamic features and poor SNR conditions of live-cell experiments, collecting a dataset for training and processing the data to create super-resolution target images can be very challenging. Therefore, we generated a new synthetic dataset closely mimicking our experiments with different photoswitchable fluorescence proteins (ffDronpa and SkylanS) as explained in 8. We used the specific, known parameters of the live-cell experiments to simulate their blinking kinetics and the full-width-half-maximum for the Gaussian-shaped point-spread-function obtained from our experimental data. A more detailed list of simulation parameters and network evaluation with simulations are provided in Supplementary Figure 2. As the SNR and SBR levels of the simulations were extremely low, 100 frames were not enough to create high SNR SOFI images as target for training. To mitigate the risk that the network could learn background pixels containing deconvolution artifacts caused by SOFI analysis, we applied additional filtering on SOFI images by using a method called wavelet-based background and noise subtraction (WBNS) [31].

In Figure 3 and 4, we show the performance of RESURF in fast-bleaching live-cell measurements by testing it with different intracellular structures and with different photoswitchable fluorescence proteins. As no ground-truth (GT) image is available in this setting, we employed resolution-scaled

Pearson correlation (RSP) to assess structural fidelity and resolution-scaled error (RSE) to show the difference between the reference image and resolution scaled version of the super-resolution image [29]. As previously mentioned for Figure 2, RSP analysis is sensitive to background pixels. Considering that the majority of the field-of-view in most microscopy experiments will belong to background pixels, we repeated the RSP and RSE analysis by using averaged wide-field (AVG) and post-processed wide-field (AVG-WBNS) as reference images. We compared the model results with SOFI images generated by 100 frames, SOFI images after post-processing with WBNS (SOFI-WBNS) and eSRRF images generated by 20 frames. Because of the higher background levels in this new dataset, we additionally incorporated L1 loss together with Fourier loss (Model-FL) during training and compared it to the model using only Fourier loss (Model-F) (see section Network and training scheme).

In Figure 3b, we evaluate the super-resolution (SR) reconstructions by different methods. In the top row, we show the image decorrelation analysis graph for each SR image. The high background level in SOFI reconstruction influences image decorrelation analysis by resulting in a worse resolution estimation than SOFI-WBNS. eSRRF can eliminate background better than SOFI, however it cannot reach the resolution level that SOFI-WBNS and the models achieve. Two very close microtubules that clearly demonstrate this difference are indicated by white arrows and also shown in Figure 3c. Following the SR images, resolution-scaled error (RSE) maps for two different reference images are given in the last two rows, together with the average RSE score and resolution-scaled Pearson correlation coefficient (RSP). The RSP and pixel errors from RSE analysis are higher in the background when the reference image has a higher background and the super-resolution reconstruction eliminates the background. This effect can be clearly seen by comparing the RSE maps of SOFI and SOFI-WBNS for different reference images. Therefore, using AVG-WBNS as a reference gives more representative results for our models. The complete movie of the cell in Figure 3a is shown in Supplementary video 1.

As can be seen in Figure 4a and c, our models trained only on microtubule structures generalize well to mitochondria and actin structures. This is confirmed by evaluation scores almost identical compared to microtubule structures. We also note that Model-F and Model-FL perform similarly where high frequency spatial components such as microtubule and mitochondria images are needed to be separated from low-frequency spatial components such as large cellular membrane or background signal. As Model-FL is trained with both Fourier and pixel-level loss function L1 while targeting background-subtracted SOFI images, it does not recognize the low-frequency spatial components(4c,d) and thus performs better in background elimination. Overall, both models achieved generalization to different structures and did not show any sign of memorizing the patterns.

For experiments with high-background conditions, eSRRF serves as a complementary benchmarking method. SOFI and eSRRF have distinct advantages and limitations. SOFI typically requires larger number of frames and additional post-processing to perform well under high background levels, whereas eSRRF is known to achieve more effective background suppression with fewer frames. However, eSRRF can generate artifacts by collapsing closely spaced structures into one another [32]. This effect is particularly evident in the mitochondria images shown in Figure 4.

To achieve better generalization and avoid hallucinations, we followed several strategies. First, in addition to different SNR levels, we simulated different background levels and random emitters in different axial positions. Second, we optimized the SOFI targets to represent the information in the diffraction limited input images. And lastly, we kept the emitter density levels moderate and the number of input frames low, thus there are more random single emitters on the focal plane to avoid the structural bias. In addition to successful models, we also showed the counter examples in Supplementary Figure 3. When we trained the model to target the SOFI results using high SNR and an order of magnitude more frames, the model showed more hallucinations in the background. Next, we trained the network by increasing the emitter density. The resulting model learned to connect random emitters to create filamentous structures, showing high structural bias.

In summary, we calibrated our training dataset to have a large variety of SNR and background levels, while avoiding complete structures in SOFI

targets and eliminating background via post-processing when necessary. As a result of our training strategy and the sequential processing ability in the convGRU layer, the models learn to extract the correlation between low-resolution images rather than recognizing similar patterns to the ones in the training dataset.

Applying RESURF models on public datasets with different fluorescent labels

We further investigated the generalization of models trained with synthetic data to different microscopes by using public datasets. As we did not vary the PSF width in our simulations, we selected datasets with a FWHM close to our synthetic images (between 2.5 and 3.2 pixels). Indeed, our models require retraining to perform well in very different optical configurations (see section Transfer learning with small experimental dataset). First, we tested a dataset which uses self-blinking dyes to label microtubule structures with our model trained for live-cell conditions [33]. This experiment was aimed at higher orders of SOFI analysis, using statistics from many frames with slower and more sparse blinking conditions than the datasets we tested in previous sections. As there are not enough emitters activated during 20-frames to reconstruct a complete structure, we averaged consecutive model predictions. By using either 100 or 8000 frames in total (by averaging 5 and 400 SR predictions), we compared them with the corresponding SOFI images. The conditions of the experiment were ideal for SOFI and both model and SOFI reconstructions showed comparable results (Fig. 5). Next, we used two datasets originally acquired for the deep-learning based super-resolution technique called SFSRM[34]. SFSRM used single frames to reach SMLM level resolutions. While the method can reach high temporal resolution, as stated in the study, this is an ill-posed problem and comes with strict limitations. Especially in low SNR levels, the method requires denoising prior to inference, making its application more complicated for live-cell conditions. The data set includes highly dynamic two intracellular structures labeled by fluorescent proteins, requiring high temporal information. Therefore we used 8-frames model and we did not apply any pre-processing before the model inference. We reconstructed a super-resolution movie of microtubule-vesicle interactions during vesicle transport and mitochondrial fission at endoplasmic reticulum contact site (Fig. 6). We showed that RESURF model can be used for wide-range of SNR, background and blinking conditions and is applicable to datasets from different microscopes with similar optics without retraining.

Transfer learning with small experimental dataset

The previous models are trained only with simulations and focused on low and ultra low SNR conditions with a narrow PSF FWHM (around 2.7 pixels). We re-trained the network to achieve reliable SR prediction for high SNR DNA-PAINT experiments with a significantly larger PSF (FWHM of 3.8 pixels). We used transfer learning for the new model and trained it with experimental microtubule data acquired from fixed-cell samples. Therefore, we were able to evaluate the model's performance for high-SNR applications and examine its consistency compared to models trained with simulations only. As we recently showed, DNA-PAINT labeling can be used to tailor fluorescent blinking to achieve both SMLM and SOFI imaging to achieve high-throughput multi-modal super-resolution imaging [35]. As a test case, we chose a fast-blinking condition and as a result of abundant exchangeable bright emitters inside the DNA-PAINT buffer, it allows long-term high-SNR data acquisition. We showed that RESURF speeds up both data-acquisition and computation time, making it suitable for real-time high-throughput imaging.

Using pre-trained weights from synthetic data allows to leverage learned features, speeding up the training process and improving generalization, which is especially useful when constrained by limited microscopy data [25, 36, 37]. We trained the model using a dataset of only 500 patches of 128x128 pixels. We created ground-truth SOFI images by using 500 frames and trained the models using 20 frames input. As shown in Supplementary Figure 4 and for live-cell experiments, limiting the number of frames to be used for target SOFI reconstruction is a crucial step to

prevent hallucinations and encourage generalization. Intuitively, using larger number of frames for SOFI could lead to better target images for training. However, it also brings a risk of hallucinations if the input images do not have the same emitters activated during 20 frames. This can encourage the model memorize the patterns instead of extracting temporal correlations. Therefore, although our dataset contains 10,000 frames of acquisition, we limit the target SOFI reconstructions to 500 frames to reduce the risk of hallucinations. The results are presented in Figure 7 and the model predictions are compared to both SOFI with 500 and 10,000 frames. Similar to the simulations, the models effectively approximates the filaments by doubling the resolution and shows high RSP score. This proves that RESURF is suitable for retraining with small datasets to learn different microscopy settings allowing easy adaptation and fast training for different experimental conditions.

Discussion

Live-cell imaging aims to achieve high spatial and temporal resolution while minimizing phototoxicity and sample damage. Although single-molecule localization microscopy (SMLM) is widely employed for both imaging and tracking at the molecular scale, it compromises temporal resolution and imposes stress on the sample due to the prolonged acquisition times. As a faster alternative, fluctuation-based imaging techniques enhance resolution using only hundreds of frames, offering a more favorable trade-off between resolution and imaging speed while better preserving sample health throughout the experiment.

To advance this approach, we adopted a deep-learning architecture with recurrent neural network called multi-image super-resolution gated recurrent unit (MISRGRU), a deep learning model designed to enhance both spatial and temporal resolution while remaining adaptable to a range of imaging conditions. By the inherent ability of recurrent networks to process sequential data, MISRGRU could serve as a real-time alternative to any fluctuation-based super resolution imaging method. We evaluated its performance using SOFI targets, as SOFI is a fluctuation-based technique with well-understood image formation. It offers advantages for live-cell imaging, including compatibility with denser fluorophore labeling, operating under lower SNR conditions, and reduced frame requirements compared to SMLM. Despite these strengths and GPU acceleration for cumulant calculations, instantaneous SOFI is still limited in temporal resolution due to its computational demands. In addition, Richardson-Lucy deconvolution is used as a post-processing. As an iterative method, it increases the computation time with the number of iterations. As a result, SOFI computations take tens of seconds to process a region of 512x512 pixels in our benchmarking system. To overcome this, we optimized and trained MISRGRU using synthetic and real datasets to reconstruct second-order SOFI images with approximately 400-fold acceleration in processing time using only 8 to 20 frames. We developed a real-time fluctuation-based super-resolution (RESURF) which allows forecasting super-resolution images of raw data during the ongoing experiment and reduces background artifacts while doubling the spatial resolution.

Our model leverages temporal information to effectively capture correlated fluorophore blinking events. Using both synthetic and experimental datasets, we observed that our method achieves a favorable balance between spatial resolution, structural fidelity, and temporal performance. Under extreme low signal-to-noise ratio (SNR) conditions, RESURF outperforms SOFI, suggesting its potential for live-cell imaging with reduced laser intensity, thereby minimizing photodamage and allowing for longer imaging sessions. In highly dynamic imaging scenarios, using a smaller number of input frames, in this case 8, offers a practical advantage, as increasing the frame count can lead to motion blur due to structural displacement over time. This enables the reconstruction of high-quality, high-frame-rate movies that preserve spatial clarity in fast-moving scenes. Its real-time reconstruction performance could be utilized as a valuable pre-processing tool for downstream applications such as segmentation, classification, or rare event detection. In addition, the MISRGRU architecture offers several practical benefits. First, independent of the training configuration, the model can be used with images of any size as input and does not require retraining. Second, the model can be trained to interpolate the input image to any size of output. This allows

using higher-order SOFI images as a target to further optimize RESURF for higher resolution improvement.

Moreover, the models were able to generalize to different intracellular structures, including mitochondria and actin, as well as vesicles and endoplasmic reticulum acquired from a different microscopy system, as tested using public datasets. By optimizing the simulation conditions to include more random emitters and target images to correspond to low-resolution frames, we prevented the structural bias and achieved a deep-learning model to extract temporal correlation between input frames.

To further improve the robustness of the model, especially under challenging imaging conditions such as those encountered in live-cell cytosolic environments, several strategies can be pursued. One promising direction is enhancing the quality of SOFI targets used during training through advanced post-processing techniques, such as improved linearization and through sophisticated pre-processing methods. These improvements are especially critical in ultra low-SNR regimes, where SOFI reconstructions tend to degrade in quality. Even-though we trained our models using a wide range of SNR conditions, we mostly focused on low and ultra-low SNR conditions. In addition, we did not normalize our datasets during training and used 16-bit inputs. By this, we aimed to keep the brightness differences between independent inputs consistent in the inference time. However, this also makes the model more sensitive to very high brightness levels of input images and can lead to saturated predictions. To increase the network's performance, it requires retraining for higher SNR conditions, which was showcased by using transfer learning with a small dataset of a high-SNR DNA-PAINT sample.

We first pre-trained the model only on synthetic datasets to create a foundation model. This offers a great opportunity for the live-cell experiment where collecting training data can be very challenging because of the poor SNR conditions and dynamic behavior of the sample. Next, we fine-tuned foundation models using a small number of data either from simulations or experiments. Leveraging pre-trained weights from synthetic data enabled the model to retain useful feature representations, thereby accelerating convergence and enhancing generalization. This is particularly valuable when only limited experimental microscopy data are available or determining the simulation parameters is not feasible. The model demonstrated comparable performance across real and synthetic data in terms of Pearson correlation, spatial resolution, and suppression of background artifacts.

We optimized our training to eliminate background artifacts via post-processing target SOFI images (synthetic dataset) or using longer image sequences to reconstruct the SOFI image (experimental dataset). For heterogeneous experimental conditions, it is more challenging to optimize the parameters for the WBNS method to balance between residual background artifacts and missing structures. In addition to alternative post-processing techniques, several strategies could be followed to increase the model's performance. First, multiple models could be trained for different background and SNR conditions instead of one global model and used different models for different regions of the images. This approach has recently been successfully implemented for SMLM [38]. An alternative way could be increasing the number of learnable parameters by increasing number of channels and using complementary targets, instead of one SOFI image, to balance between missing structures and background artifacts. This could allow the model to learn more complex tasks. However, optimizing a larger model comes with the cost of higher computational latency.

In summary, we present a recurrent neural network-based architecture tailored for real-time fluctuation super-resolution imaging (RESURF) within the microscopy domain. The proposed model is capable of reconstructing high-quality super-resolution images from as few as eight input frames, while minimizing background artifacts and achieving a two-fold spatial resolution enhancement near the theoretical limit. In addition to image reconstruction, the model can serve as a powerful pre-processing tool for downstream tasks such as image segmentation, classification, and rare event detection. By leveraging pre-trained weights from synthetic datasets and fine-tuning with either synthetic or fixed-cell microscopy data, the network is well-suited for live-cell imaging scenarios. Thanks to its low

latency, the model holds strong potential for real-time sample screening and for enhancing spatio-temporal resolution during live experiments. Furthermore, fast super-resolution streaming could help with high-throughput imaging and opportunities for smart microscopy.

Methods

Simulations

The MATLAB based SOFI simulation tool was adapted to more accurately represent noise and background effects, following the methodology described in [39]. To account for electronic noise, a gain and standard deviation map of the sCMOS camera used in our experimental setup was generated and incorporated into the simulation [40, 41]. In addition, a non-stationary background model was implemented in which background signals arise from out-of-focus blinking emitters. Inspired by the approach used in SACD [13], the distribution of emitters in the focal plane was geometrically shifted and transformed to generate a corresponding background distribution, which was subsequently convolved with a broader point spread function (PSF) to simulate out-of-focus blur. For the simulations prepared for live-samples, we added random emitters in focal plane and in some random axial planes outside focal plane.

The simulation begins by generating a binary mask of the target structure, which consists only filamentous patterns for training datasets of all the models presented in this study. We also created hollow mitochondria structures to test the the generalization of the model to different structures. These patterns are produced by modifying the pattern generation algorithms originally developed for the DBlink method and to study mitochondrial dysfunction [25, 42]. Emitters were randomly distributed across the simulation area at a density of 5000 fluorophores/μm^2 , and only those overlapping with the binary mask were retained. The spatial grid for emitter placement was set to 1280×1280 pixels with a resolution of 10 nm per pixel. Filament structures with thicknesses ranging from 10 nm to 40 nm were simulated.

The temporal blinking behavior of individual emitters was simulated using a continuous-time Markov process [39] , resulting in high-resolution video sequences. The emitters have average on-time of 10 ms, and averaged off-times of 600 ms, 1200 ms, or 2400 ms to model different blinking speeds. Signal-to-noise ratio (SNR) was further controlled by adjusting the Ion value, which represents the average number of photons emitted per frame by each fluorophore. Ion values of 100, 80, 60, and 40 were used to reflect different SNR conditions, mimicking the photophysics of fluorescent proteins under low illumination.

Both in-focus and out-of-focus emitter planes were convolved with Gaussian PSFs, having full-width at half-maximum (FWHM) values of 220 nm and 2000 nm, respectively. The maximum background contribution per pixel was calculated based on the mean brightness of fluorophores in the focal plane, using a fixed ratio between in-focus and out-of-focus signals. The resulting noise-free image stack was then downsampled by a factor of 10. Finally, Poisson-distributed photon noise and sCMOS-specific electronic noise were added to each frame to produce the final simulated dataset.

These parameters are changed for fluorescence protein simulations to represent higher background and low SNR conditions as shown in Supplementary Figure 1 and 2. The on-time and off-time are adjusted for the fluorescence proteins and specifically for ffDronpa [43]. As a result, the on-time ratio happens to be faster than the simulations for the foundation models which could decrease the frequency of the appearance of random single emitters. Therefore we lowered the emitter density to 1000 fluorophores/μm^2. As we used the foundation model to initialize the training, we only simulated 200 different microtubules patterns. Then, we varied the SNR and background levels by generating 10 different images for each pattern. Unlike previous simulations, we also added random emitters in the background, both in the focal plane and in different axial positions out of focus. In each 10-mers of simulation, we only added one high SNR condition and we maintained low SNR and high background conditions for the rest.

Network and training scheme

Multi-image super-resolution (MISR) methods have been extensively studied in satellite imaging [44]. MISRGRU was first introduced in 2020 to the

computer vision community to improve the quality of satellite imagery [26]. This model leverages multiple low-resolution images and an encoder-decoder framework to generate a super-resolved output. As illustrated in Figure 1, the architecture is composed of an encoder, a convolutional gated recurrent unit (convGRU) serving as a fusion layer, and a decoder. We inherited this architecture and modified to reduce size thus improve the latency by excluding the image registration layer and reducing the number of input channels from 64 to either 24 or 8. We also optimized the loss function by testing mean-absolute error as a spatial a spatial loss (L_1), Fourier loss ($L_{\mathcal{F}}$)[45]. During the training for the live-samples labeled by fluorescence proteins, we combined these two loss function to optimize structural integrity and resolution of predictions. We updated the weights for spatial and Fourier loss in very 20 epochs by using gradient normalization [46].

$$L_1 = \left|\hat{Y}_{u,v} - Y_{u,v}\right| \tag{1}$$

$$L_2 = \left|\hat{Y}_{u,v} - Y_{u,v}\right|^2 \tag{2}$$

$$L_{\mathcal{F}} = L_{\mathcal{F}_A} + L_{\mathcal{F}_\angle} \tag{3}$$

$$L_{\mathcal{F}_\mathcal{A}} = \frac{2}{UV} \sum_{u=0}^{U/2-1} \sum_{v=0}^{V-1} \left| |\hat{Y}|_{u,v} - |Y|_{u,v} \right| \tag{4a}$$

$$L_{\mathcal{F}_\angle} = \frac{2}{UV} \sum_{u=0}^{U/2-1} \sum_{v=0}^{V-1} \left| \angle\hat{Y}_{u,v} - \angle Y_{u,v} \right| \tag{4b}$$

Here, both SR images $\hat{Y}$ and the target image Y are transformed into the Fourier space by applying fast Fourier transform (FFT), where the absolute amplitude difference $L_{\mathcal{F}_\mathcal{A}}$ and absolute phase difference $L_{\mathcal{F}_\angle}$ are calculated. Due to symmetry in the Fourier space (Hertimitian symmetry), only half of the spectral components is considered.

We trained the foundation model with simulated datasets of 128×128 pixel images. The dataset was split into training (2,000 samples), evaluation, and test sets (480 samples each). Model training was performed using the Adam optimizer with a learning rate (η) of 0.001 and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. He Normal initialization was used for the model weights [47]. To mitigate overfitting, early stopping was applied with a maximum of 400 training epochs and a patience threshold of 10. A batch size of 10 was selected to balance GPU memory usage and computational efficiency. All training and testing were conducted on an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory.

To extend the model to experimental data, we used the pre-trained weights from the foundation model as initialization of the models we trained for live-cell and DNA-PAINT measurements. To further reduce inference latency, we applied TensorRT acceleration to all models trained within the PyTorch framework [48]. This optimization included techniques such as layer fusion and precision calibration, including the use of lower-precision formats like fixed-point floating point, to maximize computational efficiency.

Performance evaluation

We evaluate the performance of SOFI reconstructions and AI model generated predictions using three complementary methods. First, we estimate image resolution through image decorrelation analysis, a technique that operates on a single image and relies on partial phase correlation to determine spatial resolution [28]. Second, we employ the resolution-scaled Pearson correlation (RSP) metric to quantify the similarity between the super-resolution (SR) image and a reference image [49]. Prior to correlation computation, the SR image is adjusted to match the resolution of the reference image using an estimated resolution-scaling function (RSF), providing a more robust comparison than traditional metrics like structural similarity index (SSIM). For experimental data, the reference is the average projection of the image sequence (wide-field). As we noticed that the RSP score can under-represent the structural similarity in images with high-background, we also used a wavelet-based background subtraction method (WBNS) to create a background free alternative reference image [31]. We used 3 levels in wavelet decomposition to estimate the background and skipped the denoising part of the algorithm. In addition, for simulated test set, we also calculated the RSP scores between ground-truth and super-resolution predictions. Lastly, to

visualize the pixel-level difference between reference image and the super-resolution prediction, we used resolution-scaled error map (RSE). Similar to RSP, absolute difference between pixel values are calculated after matching the resolution between two images and averaged to calculate RSE score. We noticed that the boundary pixels in model predictions for actin data (in Figure 4d) can sometimes show hot pixels. This is likely due to the extremely high signal levels in actin experiments, which is not represented in our training set. Therefore, we cropped the five boundary pixels in actin images and used the smaller, cropped region during the model evaluation with RSP and resolution analysis.

Data analysis

SOFI

We reconstructed SOFI images using sofipackage based on cross-cumulant-based algorithm which is available from https://github.com/kgrussmayer/-sofipackage. Following n-th order cross-cumulant analysis, images reach to resolution improvement of $\sqrt{n}$-fold. Next, we apply linearization and Richardon-Lucy deconvolution as a post-processing [14, 27, 50]. To compensate the non-linearity in brightness, we either use linear-scaling or adaptive linearization. While the linear scaling takes the n-th root of the brightness in n-th order SOFI, adaptive linearization incorporates blinking properties to enable the effective use of the dynamic range. For Richardson-Lucy deconvolution (RLD), we used a gaussian kernel with the same FWHM we used for simulations and 10 iterations to reach two-fold resolution improvement. For foundation models, we use adaptive linearization and 2.2 pixels of FWHM for RLD. For high background simulations, imitating live-cell condition, we use linear-scaling and 2.7 pixels of FWHM for RLD and follow with a wavelet-based background and noise subtraction (WBNS) method to eliminate the background and ringing artifacts caused by deconvolution [31]. We used 3 levels decomposition for background estimation and 2 levels of decomposition for noise estimation.

For synthetic training datasets created for live-cell experiments, we followed WBNS method with a binary mask to eliminate the background pixels [51]. To create the binary mask, we applied Li-thresholding on the high-SNR versions in each 10-mers dataset.

eSRRF

We used the Fiji plugin of eSRRF to get guidance in finding the best parameters for radius (R) and sensitivity level (S), and operated the rest of the analysis using the Python tool, nanopyx [11, 30]. eSRRF has different types of temporal analysis. In all figures and evaluations, we used averaged eSRRF reconstructions with two factor interpolation.

Sample preparation and data acquisition

DNA-PAINT imaging

Fast DNA-PAINT data used in Figure 7 were acquired using the method previously published [35]. Briefly, COS7 cells were cultured in DMEM with standard supplements, seeded on coverslips, and fixed. Microtubules in cells were stained firstly with primary antibodies against alpha-tubulin, followed by nanobodies conjugated with 5×R1 DAN-PAINT docking strands (sequence 5'-3': TCCTCCTCCTCCTCCTCCT). Cells were imaged in buffer consisting of 20 nM R1 imager strands (sequence 5'-3': AGGAGGA – Atto655), 5% (v/v) ethylene carbonate and 500 mM NaCl in PBS. Custom-built TIRF microscope was used for imaging at 638 nm excitation ($2.5kW/cm^2$); 10,000 frames were recorded at 10 ms exposure using MicroManager. Full description of the nanobody production, labeling protocol and microscope specifications are available in the method section of [35].

Live-cell imaging

For live-cell imaging, HeLa cells were grown at $37°C$, in 5% CO2 atmosphere in DMEM (ThermoFisher) complemented with 10% fetal bovine serum (FBS) and GlutaMAX (ThermoFisher). For imaging experiments, cells were seeded in 35-mm glass-bottom dishes (MatTek) and after 24 h, they were transiently transfected using FuGENE HD Transfection Reagent (Promega) according to the manufacturer's protocol using following plasmids; MAP4-ffDronpa (created in

the Dedecker Lab), DAKAP-SkylanS (created in the Dedecker Lab) and SkylanS-betaActin (purchased from Addgene). 24 hours after the transfection, the cells were washed with PBS (ThermoFisher) and then imaged in PBS on a Nikon Eclipse Ti-2 Inverted Microscope equipped with a 1.4 NA oil immersion objective (100 CFI Apochromat Total Internal Reflection Fluorescence) and a ZT405/488/561/640rpcv2 dichroic mirror with a ZET405/488/561/640 nm emission filter (both Chroma Technology). The imaging was performed always in TIRF mode using 488 nm and 405 nm lasers (Oxxius L6Cc laser combiner). Each measurement consisted of the acquisition of stacks of 500 or 1000 images. MAP4-ffDronpa and DAKAP-SkylanS are recorded at 30 ms and SkylanS-betaActin is recorded at 20 ms exposure.

Acknowledgements

We thank the TU Delft Department of Bionanoscience for support and the TU Delft AI lab initiative for funding BIOlab. K.S.G was supported by an ERC grant (QScope, 101165129) funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Contributions

M.T, J.K and K.S.G conceived the idea. M.T and J.K designed the simulations and evaluation with input from K.S.G and N.T. J.K tested MISGRU network and optimized them to create foundation models. M.T performed further analysis to evaluate the performance and improved the simulations for live-cell experiments and generalization of the models. M.T and J.K trained the model for transfer learning with experimental data. M.T analyzed public datasets to test generalization. H.V designed and performed the live-cell experiments. B.K.Z and R.H. designed and performed the DNA-paint experiments. M.T wrote the manuscript. K.S.G, N.T and P.D reviewed the manuscript. N.T supervised the deep network implementation and evaluation. K.S.G supervised the microscopy experiments, data analysis and network evaluation.

References

[1] Pylvänäinen, J. W., Gómez-de Mariscal, E., Henriques, R. & Jacquemet, G. Live-cell imaging in the deep learning era. *Current Opinion in Cell Biology* **85**, 102271 (2023). URL https://www.sciencedirect.com/science/article/pii/S0955067423001205.

[2] Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters* **19**, 780 (1994). URL https://opg.optica.org/abstract.cfm?URI=ol-19-11-780.

[3] Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy: SHORT COMMUNICATION. *Journal of Microscopy* **198**, 82–87 (2000). URL https://onlinelibrary.wiley.com/doi/10.1046/j.1365-2818.2000.00710.x.

[4] Betzig, E. *et al.* Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* **313**, 1642–1645 (2006). URL https://www.science.org/doi/10.1126/science.1127344.

[5] Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793–796 (2006). URL https://www.nature.com/articles/nmeth929.

[6] Zheng, X. *et al.* Current challenges and solutions of super-resolution structured illumination microscopy. *APL Photonics* **6**, 020901 (2021). URL https://doi.org/10.1063/5.0038065.

[7] Jacquemet, G., Carisey, A. F., Hamidi, H., Henriques, R. & Leterrier, C. The cell biologist's guide to super-resolution microscopy. *Journal of Cell Science* **133**, jcs240713 (2020). URL https://doi.org/10.1242/jcs.240713.

[8] Pawlowska, M., Tenne, R., Ghosh, B., Makowski, A. & Lapkiewicz, R. Embracing the uncertainty: the evolution of sofi into a diverse family of fluctuation-based super-resolution microscopy methods. *Journal of Physics: Photonics* **4**, 012002 (2021).

[9] Dertinger, T., Colyer, R., Iyer, G., Weiss, S. & Enderlein, J. Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *Proceedings of the National Academy of Sciences* **106**, 22287–22292 (2009). URL https://pnas.org/doi/full/10.1073/pnas.0907866106.

[10] Basak, S. *et al.* Super-resolution optical fluctuation imaging. *Nature Photonics* 1–9 (2025).

[11] Gustafsson, N. *et al.* Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nature Communications* **7**, 12471 (2016). URL https://www.nature.com/articles/ncomms12471.

[12] Laine, R. F. *et al.* High-fidelity 3D live-cell nanoscopy through data-driven enhanced super-resolution radial fluctuation. *Nature Methods* 1949–1956 (2023).

[13] Zhao, W. *et al.* Enhanced detection of fluorescence fluctuations for high-throughput super-resolution imaging. *Nature Photonics* **17**, 806–813 (2023).

[14] Geissbuehler, S. *et al.* Mapping molecular statistics with balanced super-resolution optical fluctuation imaging (bSOFI). *Optical Nanoscopy* **1**, 4 (2012). URL http://optnano.springeropen.com/articles/10.1186/2192-2853-1-4.

[15] Cunha, I., Latron, E., Bauer, S., Sage, D. & Griffié, J. Machine learning in microscopy–insights, opportunities and challenges. *Journal of cell science* **137** (2024).

[16] Ronneberger, O., Fischer, P. & Brox, T. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F. (eds) *U-net: Convolutional networks for biomedical image segmentation.* (eds Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241 (Springer International Publishing, Cham, 2015).

[17] Jin, L. *et al.* Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nature Communications* **11**, 1934 (2020). URL https://www.nature.com/articles/s41467-020-15784-x.

[18] Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnology* **36**, 460–468 (2018). URL https://www.nature.com/articles/nbt.4106.

[19] Qu, L. *et al.* Self-inspired learning to denoise for live-cell super-resolution microscopy. URL http://biorxiv.org/lookup/doi/10.1101/2024.01.23.576521.

[20] Speiser, A. *et al.* Deep learning enables fast and dense single-molecule localization with high accuracy. *Nature Methods* **18**, 1082–1090 (2021). URL https://www.nature.com/articles/s41592-021-01236-x.

[21] Chen, J. *et al.* Deep-learning accelerated super-resolution radial fluctuations (SRRF) enables real-time live cell imaging. *Optics and Lasers in Engineering* **172**, 107840 (2024). URL https://www.sciencedirect.com/science/article/pii/S014381662300369X.

[22] Qiao, C. *et al.* A neural network for long-term super-resolution imaging of live cells with reliable confidence quantification. *Nature Biotechnology* 1–10 (2025).

[23] Nehme, E., Weiss, L. E., Michaeli, T. & Shechtman, Y. Deep-storm: super-resolution single-molecule microscopy by deep learning. *Optica* **5**, 458–464 (2018). URL https://opg.optica.org/optica/abstract.cfm?URI=optica-5-4-458.

[24] Li, J., Tong, G., Pan, Y. & Yu, Y. Spatial and temporal super-resolution for fluorescence microscopy by a recurrent neural network. *Opt. Express* **29**, 15747–15763 (2021). URL https://opg.optica.org/oe/abstract.cfm?URI=oe-29-10-15747.

[25] Saguy, A. *et al.* DBlink: dynamic localization microscopy in super spatiotemporal resolution via deep learning **20**, 1939–1948. URL https://www.nature.com/articles/s41592-023-01966-0.

[26] Rifat Arefin, M. *et al.* Multi-image super-resolution for remote sensing using deep recurrent networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 816–825 (2020).

[27] Deschout, H. *et al.* Complementarity of palm and sofi for super-resolution live-cell imaging of focal adhesions. *Nature communications* **7**, 13693 (2016).

[28] Descloux, A., Grußmayer, K. S. & Radenovic, A. Parameter-free image resolution estimation based on decorrelation analysis. *Nature Methods* **16**, 918–924 (2019). URL https://www.nature.com/articles/s41592-019-0515-7.

[29] Culley, S. *et al.* Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nature Methods* **15**, 263–266 (2018). URL https://www.nature.com/articles/nmeth.4605.

[30] Saraiva, B. M. *et al.* Efficiently accelerated bioimage analysis with nanopyx, a liquid engine-powered python framework. *Nature methods* **22**, 283–286 (2025).

[31] Hüpfel, M., Yu. Kobitski, A., Zhang, W. & Nienhaus, G. U. Wavelet-based background and noise subtraction for fluorescence microscopy images. *Biomedical Optics Express* **12**, 969–980 (2021).

[32] Alva, A. *et al.* Fluorescence fluctuation-based super-resolution microscopy: Basic concepts for an easy start. *Journal of Microscopy* **288**, 218–241 (2022).

[33] Grußmayer, K., Lukes, T., Lasser, T. & Radenovic, A. Self-blinking dyes unlock high-order and multiplane super-resolution optical fluctuation imaging. *ACS nano* **14**, 9156–9165 (2020).

[34] Chen, R. *et al.* Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging. *Nature Communications* **14**, 2854 (2023).

[35] de Zwaan, K. *et al.* High-throughput single-molecule microscopy with adaptable spatial resolution using exchangeable oligonucleotide labels. *ACS nano* (2024).

[36] Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016). http://www.deeplearningbook.org.

[37] Demircan-Tureyen, E., Akbulut, F. P. & Kamasak, M. E. Restoring fluorescence microscopy images by transfer learning from tailored data. *IEEE Access* **10**, 61016–61033 (2022).

[38] Saguy, A. *et al.* One-click reconstruction in single-molecule localization microscopy via experimental parameter-aware deep learning. *npj Imaging* **3**, 61 (2025).

[39] Girsault, A. *et al.* Sofi simulation tool: a software package for simulating and testing super-resolution optical fluctuation imaging. *PLoS One* **11**, e0161602 (2016).

[40] Zhang, Z., Wang, Y., Piestun, R. & Huang, Z.-L. Characterizing and correcting camera noise in back-illuminated scmos cameras. *Optics express* **29**, 6668–6690 (2021).

[41] Liu, S. *et al.* scmos noise-correction algorithm for microscopy images. *Nature methods* **14**, 760–761 (2017).

[42] Brenner, B. *et al.* Quantifying nanoscopic alterations associated with mitochondrial dysfunction using three-dimensional single-molecule localization microscopy. *Biomedical optics express* **15**, 1571–1584 (2024).

[43] Dedecker, P. *et al.* Fast and reversible photoswitching of the fluorescent protein dronpa as evidenced by fluorescence correlation spectroscopy. *Biophysical journal* **91**, L45–L47 (2006).

[44] Lepcha, D. C., Goyal, B., Dogra, A. & Goyal, V. Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Information Fusion* **91**, 230–260 (2023).

[45] Fuoli, D., Gool, L. V. & Timofte, R. Fourier space losses for efficient perceptual image super-resolution. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)* 2340–2349 (2021).

[46] Chen, Z., Badrinarayanan, V., Lee, C.-Y. & Rabinovich, A. *Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks*, 794–803 (PMLR, 2018).

[47] Glorot, X. & Bengio, Y. Teh, Y. W. & Titterington, M. (eds) *Understanding the difficulty of training deep feedforward neural networks.* (eds Teh, Y. W. & Titterington, M.) *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Vol. 9 of *Proceedings of Machine Learning Research*, 249–256 (PMLR, 2010). URL https://proceedings.mlr.press/v9/glorot10a.html.

[48] Pytorch. Tensor rt. https://github.com/pytorch/TensorRT (2024).

[49] Culley, S. *et al.* Author correction: Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nature Methods* **17**, 1167–1167 (2020).

[50] Dertinger, T., Colyer, R., Vogel, R., Enderlein, J. & Weiss, S. Achieving increased resolution and more pixels with Superresolution Optical Fluctuation Imaging (SOFI). *Optics Express* **18**, 18875 (2010). URL https://opg.optica.org/oe/abstract.cfm?uri=oe-18-18-18875.

[51] Li, C. & Lee, C. Minimum cross entropy thresholding **26**, 617–625. URL https://linkinghub.elsevier.com/retrieve/pii/003132039390115D.

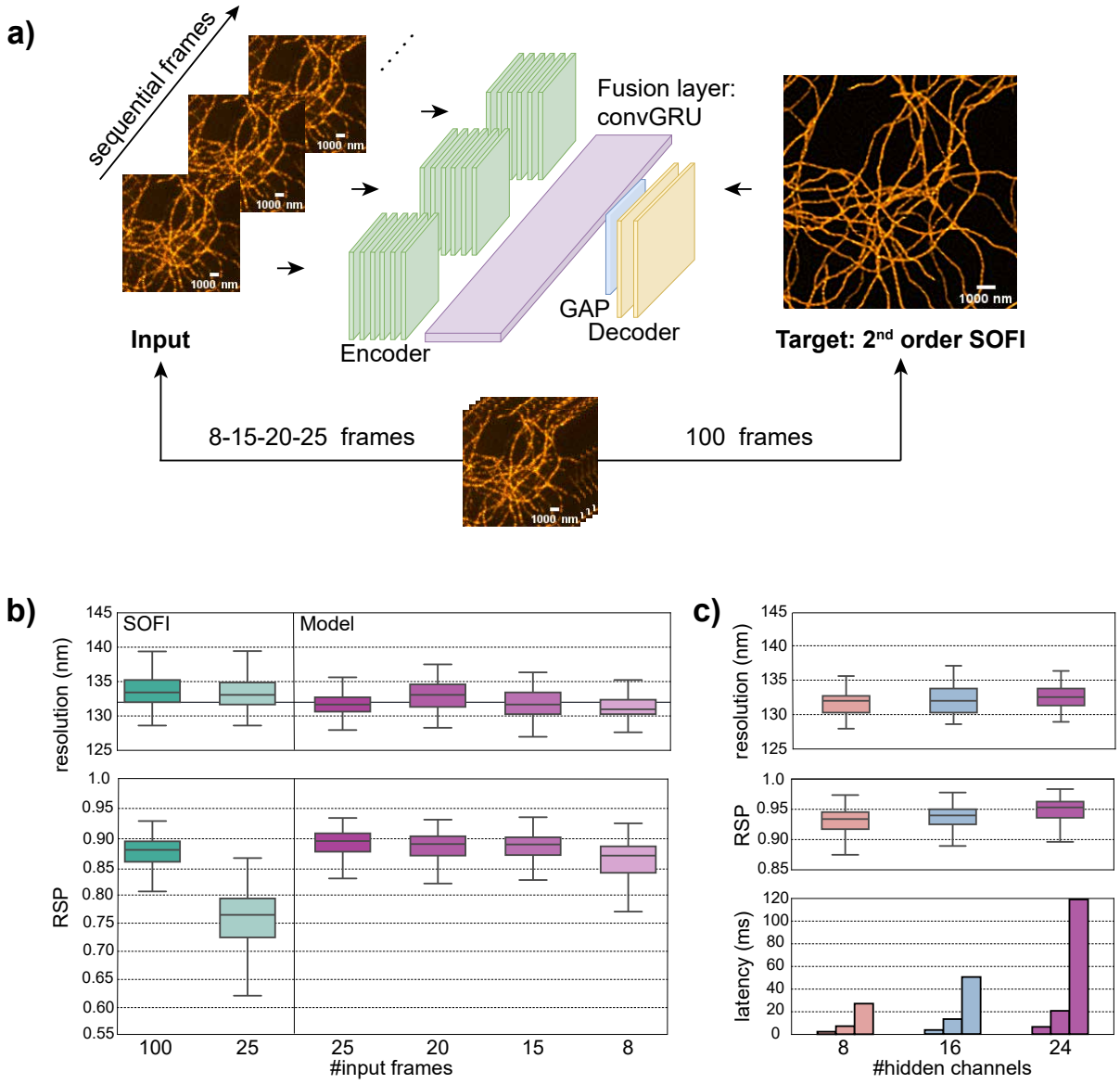
a)
sequential frames
1000 nm
Input
Encoder
Fusion layer:
convGRU
GAP
Decoder
Target: 2nd order SOFI
8-15-20-25 frames
100 frames
b)
SOFI
Model
resolution (nm)
RSP
#input frames
c)
resolution (nm)
RSP
latency (ms)
#hidden channels

Fig. 1: **The network architecture and optimization for RESURF**. **a.** The network architecture (MISRGRU) and training strategy. It is trained with simulated diffraction limited image sequences and corresponding second order SOFI reconstructions. Multiple models (foundation models) are trained by using different number of input frames (8,15,20,25) and, different number of hidden channels. **b.** Spatial resolution estimated via image decorrelation analysis and resolution-scaled Pearson correlation coefficient for SOFI (green) and the models (purple) using different numbers of input frames. Each boxplot depicts the median value (center line), 25th and 75th percentile (box edges), and the 1.5 IQR (interquartile range) (whiskers). The line at 132 nm marks the half-resolution of the widefield average image. **c.** Performance evaluation of models configured with different numbers of hidden channels: 8 (red), 16 (blue), and 24 (purple), all using 20 frames input. Latency is the average inference time over 500 repetitions: from left to right, each bar represents a different field of view (FOV) size: 128×128, 256×256, and 512×512 pixels. All models shown in this figure are trained by using F loss. The comparison of different loss functions can be found in Supplementary Figure 1. Scale bar: 1 µm.

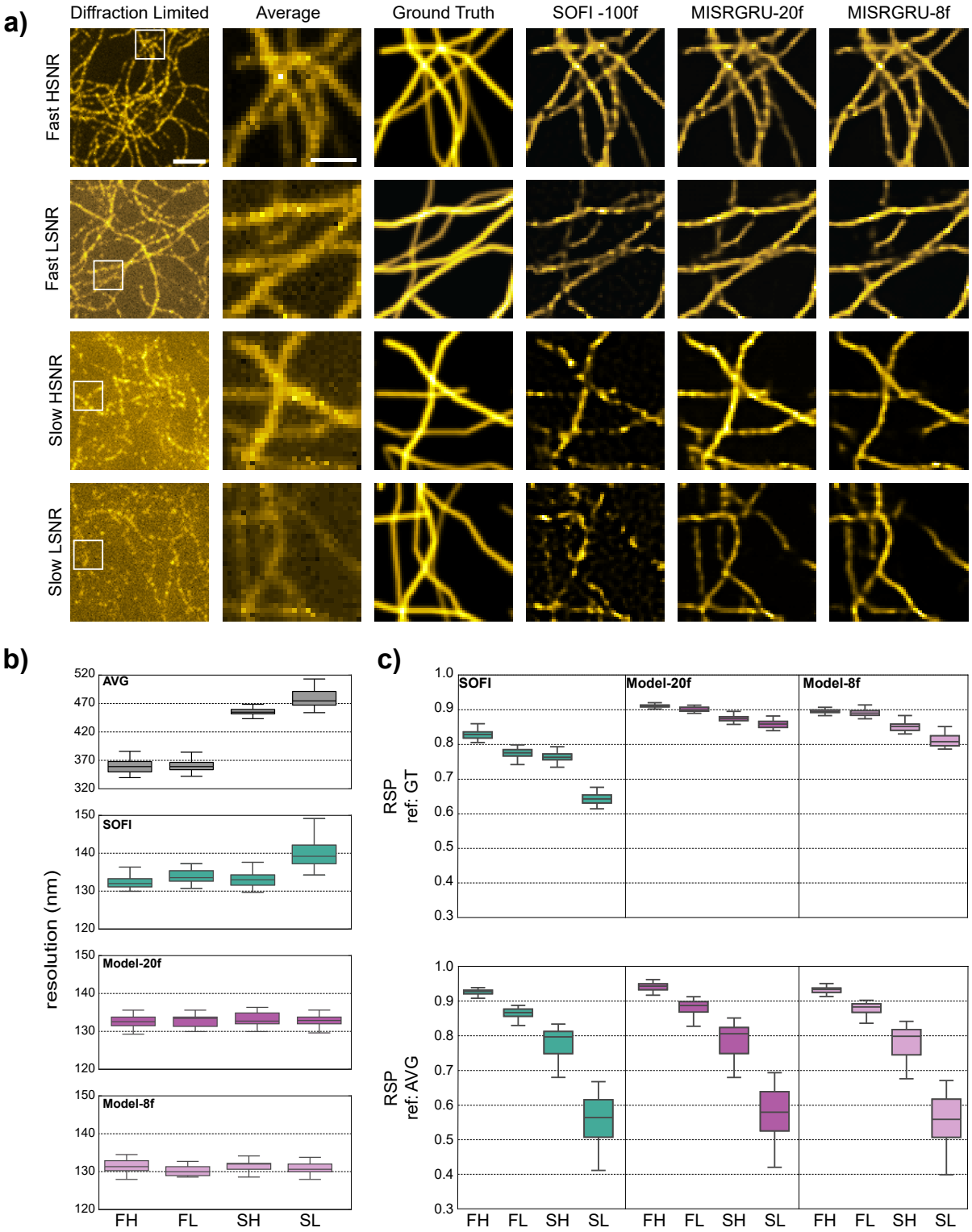
a)
Diffraction Limited
Average
Ground Truth
SOFI -100f
MISRGRU-20f
MISRGRU-8f
Fast HSNR
Fast LSNR
Slow HSNR
Slow LSNR
b)
resolution (nm)
AVG
SOFI
Model-20f
Model-8f
FH
FL
SH
SL
c)
RSP ref: GT
RSP ref: AVG
SOFI
Model-20f
Model-8f

Fig. 2: **Performance of foundation models. a.** Qualitative comparison of models and SOFI. Each row corresponds to a different SNR (photons/frame): 100(high-SNR) or 40 (low-SNR), and and blinking conditions (on time/ off time in ms): 10/600 (fast) or 10/2400 (slow). From left to right: a single diffraction-limited input frame (128x128 pixels or 12.8x12.8 μm), the average of cropped area (30x30 px), the ground truth image demonstrating two-fold resolution enhancement, second-order SOFI reconstructions using 100 frames, and Model-20f and Model-8f. **b, c.** Corresponded resolution estimations and RSP scores of averaged wide-field, SOFI and the models for different conditions. RSP analysis is repeated for two different reference image: ground-truth (GT), and averaged wide-field (AVG) FH: Fast high SNR, FL: Fast low SNR, SH: Slow high SNR, SL: Slow low SNR. Pixel size in AVG: 100nm, Scale bars: 3 µm(diffraction limited), 1 µm (average).

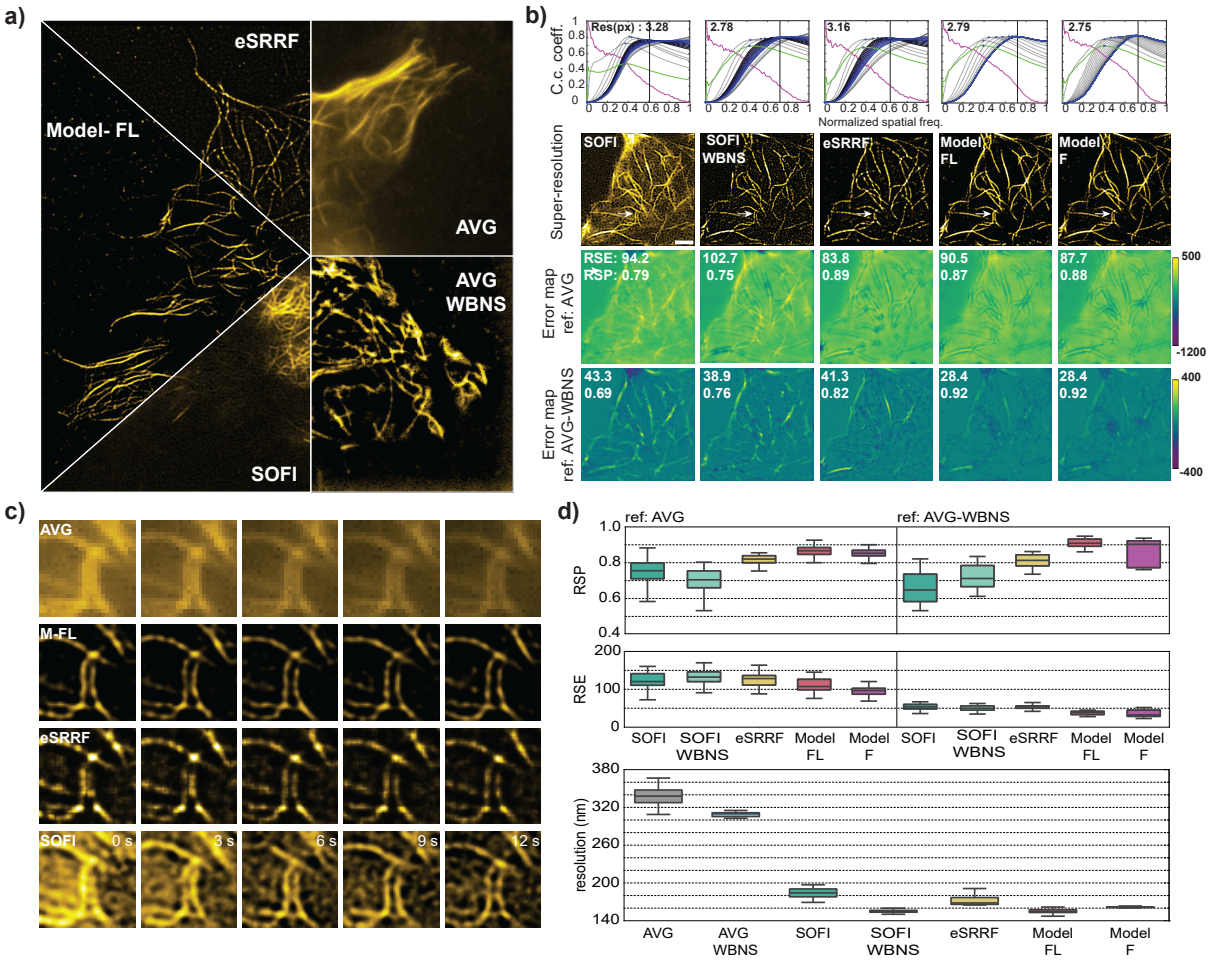

Fig. 3: **RESURF reconstructs SR images with high temporal resolution in a fast-bleaching live-sample a.** Qualitative comparison of SR reconstruction of microtubule structures using Model-FL, SOFI and eSRRF (HeLa cell, MAP4-ffDronpa). **b.** Influence of background on image decorrelation analysis and resolution-scaled error (RSE) for cropped regions of 128x128 pixels. First row shows decorrelation analysis for resolution estimation. Following SR reconstructions by different method RSE estimations for two reference images:(i) averaged projection of wide-field frames (AVG), (ii) background subtracted AVG images using wavelet-based noise and background subtraction (AVG-WBNS) **c.** SR reconstruction from a small dynamic region of 30x30 pixels (1.5x1.5 μm). **d.** Evaluation scores for resolution estimation, RSP and RSE by using AVG and AVG-WBNS as reference images. Each box plot is created by using 50 random crops of 128x128 from the large FOV in (a). Model-F: trained with Fourier loss, Model-FL: trained with combination of Fourier and L1 loss. Pixel size in AVG: 108 nm, Scale bars: 1.6 µm (b), 2 µm (c)

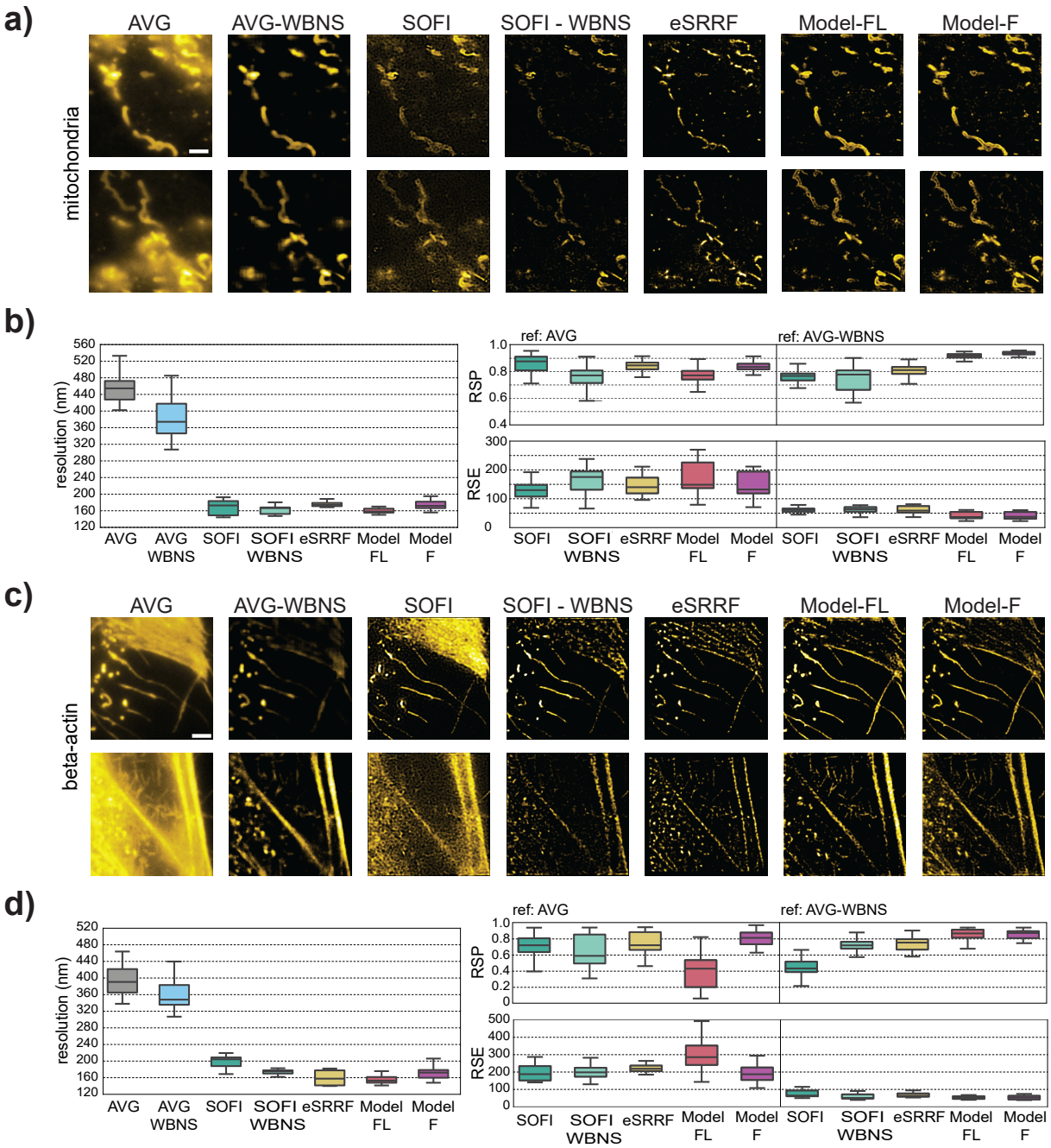

Fig. 4: **Generalization of RESURF to different intracellular structures. a,b.** Qualitative comparison of our models, SOFI and eSRRF for a region of 128x128 pixels for mitochondria (DAKAP-SkylanS) (a) and actin (SkylanS-betaActin) in HeLa cells (c). **(b,d.)** Correspondong evaluation scores for resolution estimation, RSP and RSE by using AVG and AVG-WBNS as reference images. Model-F: trained with Fourier loss, Model-FL: trained with combination of Fourier and L1 loss.Pixel size: 108 nm. Scale bars: 2 µm

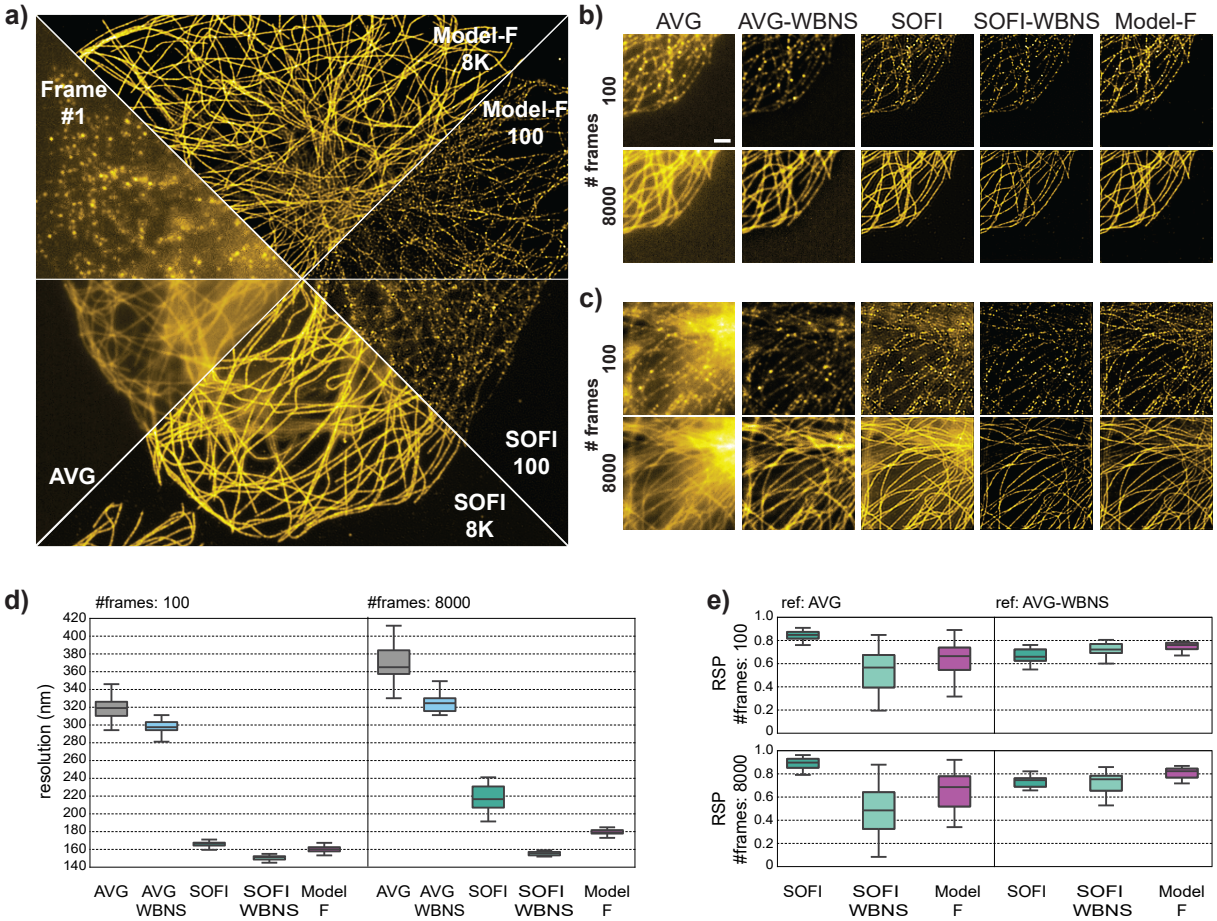

Fig. 5: **Generalization of RESURF to slow-blinking samples tested by a public SOFI dataset expressing microtubules and labelled by self-blinking dyes[33] . a,b,c.** COS-7 cell, immunostained for microtubules with Abberior FLIP 565. Qualitative comparison of SR reconstructions using 100 or 8000 frames by our models, SOFI and eSRRF. Small regions of 128x128 pixels are cropped from a high structural density (c) and a low structural density (b) area. **d,e.** Quantitative comparisons via resolution estimation and resolution-scaled Pearson correlation (RSP). SOFI-8000 is the average of SOFI reconstructions using 500 frames. Model-F for 100 and 8000 frames are calculated by averaging the model reconstructions used by 20-frames inputs. Scale bars: 2 µm

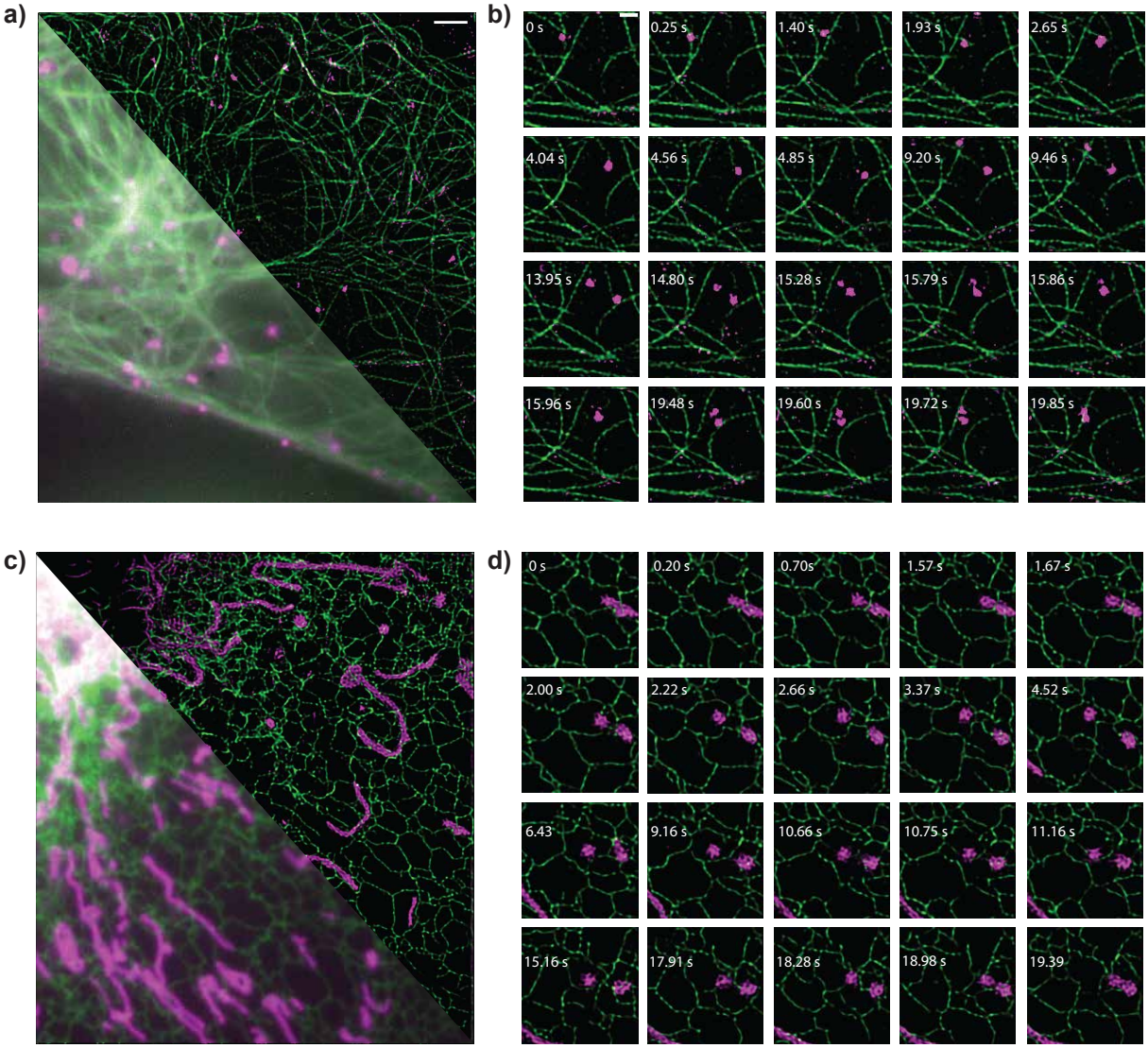

Fig. 6: **RESURF observation of dynamic biological processes in public dataset**[34] **. a.** Two-color image of microtubules (green) and vesicles (magenta) in Beas2B cells. **b.** Example of vesicle transportation dynamics. **c.** Two-color image of endoplasmic reticulum (ER-green) and mitochondria (magenta) in Beas2B cells. **d.** Example of mitochondrial fission at ER-mito contact site. Wide-field images are averaged projection of first 8-frames and RESURF results are reconstructed by rolling a 8-frames window and using Model-F. Time-labels belong to the first frame in each window. Scale bars: 5 μm (a), 1 μm (b)

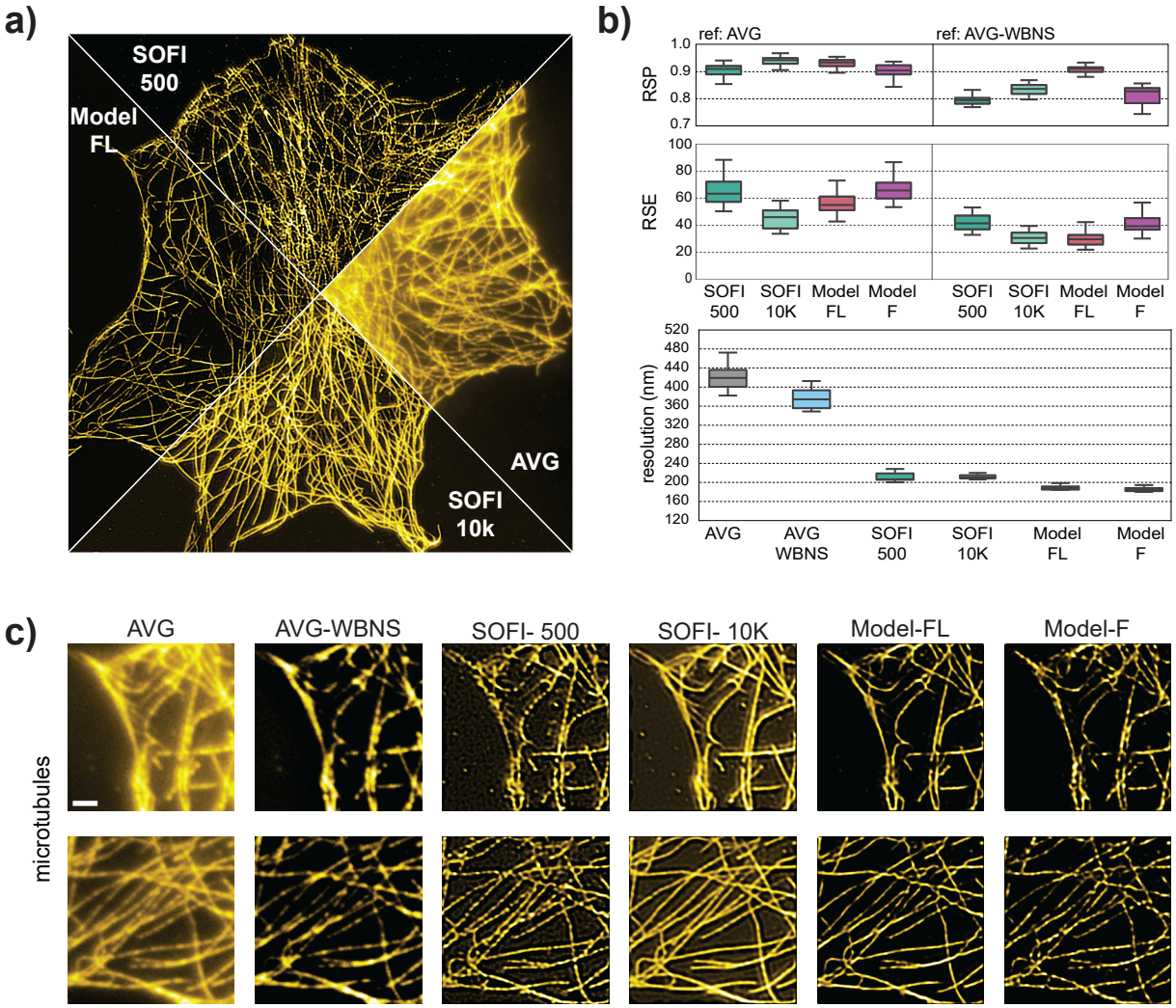

Fig. 7: **RESURF fine-tuned by small experimental dataset enables high-throughput SR imaging for fast-blinking DNA-PAINT samples. a.** Super-resolution reconstructions and averaged wide-field image of microtubules in a COS7 cell, labeled by DNA-PAINT **b.** Corresponded evaluation scores of resolution estimation and resolution-scaled Pearson correlation (RSP) for two different reference images (average (AVG)and post-processed average (AVG-WBNS) of diffraction limited frames). **c.** Qualitative comparison of SR reconstructions by our models, SOFI with 500 and 10,000 frames for a cropped region of 128x128 pixels in wide-field. Averaged wide-field and model predictions are reconstructed by using 20 frames. Scale bar: 2 µm