Deep learning augmented single molecule localization microscopy reconstruction : enhancing robustness and moving towards live cells

Jiachuan Bai

Sorbonne Université

Ecole doctorale

Imagerie et Modélisation, Institut Pasteur, Paris, France

Deep learning augmented single molecule localization microscopy reconstruction

Enhancing robustness and moving towards live cells

Par Jiachuan BAI

Thèse de doctorat de bio-informatique

Dirigée par Christophe ZIMMER

Présentée et soutenue publiquement le 07/11/2023

Devant un jury composé de :

Dr. Florian Jug	Rapporteur
Pr. Laure Blanc-Féraud	Rapporteuse
Pr. Isabelle Bloch	Présidente du jury
Dr. Charles Kervrann	Examinateur
Dr. Christophe Zimmer	Directeur de thèse

“学习到的知识是你一生最宝贵的财富”

献给爷爷

Table of contents

I. Introduction

I.1. Optical microscopy

In the world of biology discovery, seeing is believing.

For centuries, the human eye has served as the primary tool. The human naked eye can resolve around 50 µm under optimal conditions. For example, it can see a strand of hair at arm's distance (Frisk, 2012). Yet, in the world of biology, most structures and organisms are smaller than this limit. Examples include viruses, which range in size from 20 to 1 500 nm (Colson *et al.*, 2017), or the microtubules with a diameter of around 25 nm (Wade, 1997), or the nuclear pore complexes, which measure around 100-150 nm in diameter (Wente and Rout, 2010).

All of these structures are far below the resolution limit of the naked eye. **Figure 1** provides a visual representation of the relative scale of the biological structures and molecules, and it shows what our eyes can see as well as structures and molecules that exist on a scale much smaller than the width of a human hair.

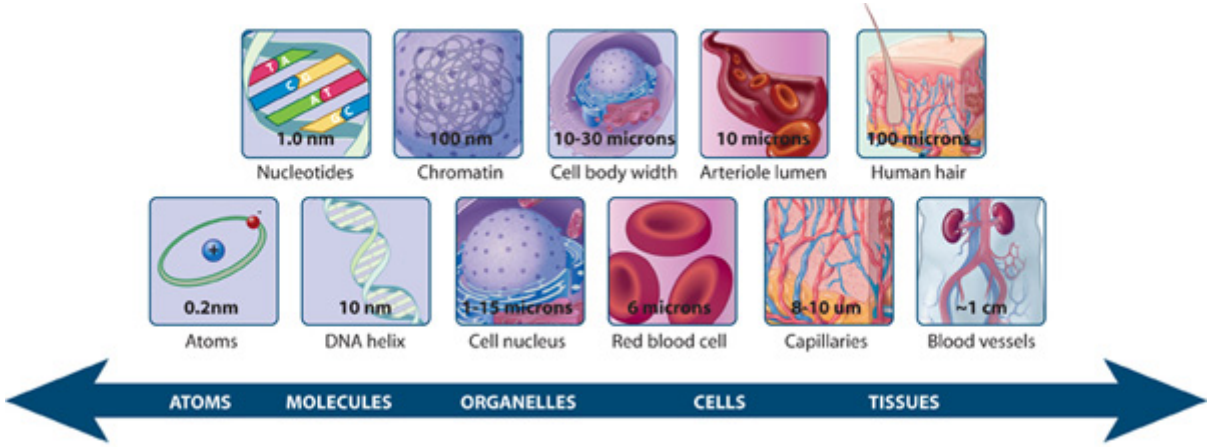


Figure 1. The scale of biological molecules and structures.
Image from (*What Is a Cell? | Learn Science at Scitable*, no date)

The limitations of our naked eyes lead us to rely on scientific tools. Between the 16th and 17th centuries, the first light microscopes were developed by Hans and Zacharias Janssen in around 1590 (Orchard and Nation, 2015). A microscope is equipped with a series of lenses capable of magnifying small biological structures. Thanks to this magnification, microscopy offers a resolution (page 9) around 200 nm, which is much better the naked eye, thus has become a central technique in cell biology.

I.1.1. Principle of optical microscopy

By converting emitted or transmitted light into a magnified image, optical microscopy reveals the "invisible" to the naked eyes. For instance**, Figure 2** shows the principle of a simple two-lens optical microscope using a ray diagram. The light source is collected by the objective lens, then transmitted through the sample to magnify. The transmitted light is transferred to the eyepiece lens, then the observer can see the image. Microscopy image can be captured by a camera installed in the optical system, which can be considered as two-dimensional matrix. Each element of this matrix is called pixel, that represents the smallest unit of an image, containing information such as intensity value and color.

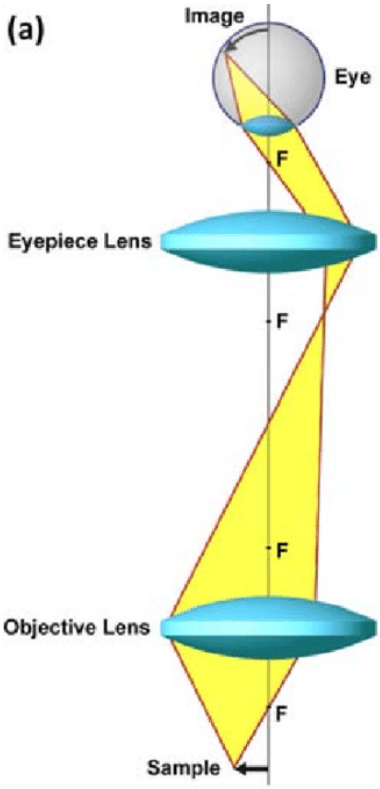


Figure 2. Ray diagram of the simplest two-lens microscope.
Image from (Lee *et al.*, 2011)

The numerical aperture of the objective lens, that is function of its size, can't be infinite, leading to imperfect restitution of images through an optical setup. The Point Spread Function (PSF) is a three-dimensional pattern coming from an infinitely small point light source through the optical system (will be detailed on page 9). Mathematically, microscopy images are formed by the convolution of the PSF with the sample. This convolution process, however, can lead to

the blurring and spreading of the original sample image, thus affecting the resolution of the microscopy image (Abbe, 1873).

The focal plane is at a specific distance from the objective lens where the sample is in sharp focus, the sample at this distance will be imaged with maximum clarity. The signals from regions of the sample below or above the focal plane overlap with the signals from the focal plane. As a result, the microscopy image has mixed signals: some are sharp, while others are blurred, which further impacts the resolution of the image.

I.1.2. Broad classifications of optical microscopy

Optical microscopy can be broadly categorized into two groups: bright-field and fluorescence microscopy:

In bright-field microscopy, the light source goes through the sample and is collected by the objective lens. Depending on the refractive index of the tissues, the light's direction and speed change. When the sample is illuminated, the regions of the sample with a higher refractive index than the medium will bend the light more and appear darker in the formed image. Conversely, regions with a lower refractive index compared to the medium allow more light to pass through, appearing brighter. This contrast allows us to distinguish the sample from the medium and visualize its different components.

Unlike bright field light microscopy that relies on the contrast of the specimen, in fluorescence microscopy (more details on page 11), we use fluorescent molecules also known as fluorescent dyes or fluorophores to label (or stain, tag) the specific biological structures or molecules of interest. The sample is illuminated with light of a specific wavelength and the fluorescent molecules absorb this excitation light and emit light at another wavelength (Lichtman and Conchello, 2005). Then the microscopy filters out the excitation light and collects the emitted fluorescence from the sample. This results in an image that only the labeled structures or molecules are visible.

Figure 3 illustrates the principle of fluorescence via the Jablonski diagram, where the fluorophore absorbs a photon, reaching an excited energy state and then emits a longer wavelength while returning to ground state. Each time the fluorophore returns from the excited

state to the ground state, there is a small chance it become permanently non-fluorescent. This irreversible process is called photobleaching, leading to decreased fluorescent signal over time.

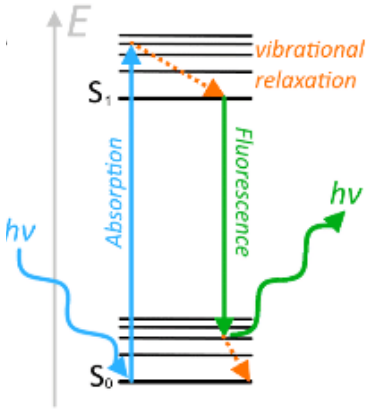


Figure 3. Jablonski diagram.
The fluorophore initially absorbs a photon (blue) at ground state (S_0), raising an electron to an excited energy state (S_1). Once in this excited state, the electron undergoes vibrational relaxation (orange). Due to the unstable nature of the excited state, the electron returns to the ground state typically within nanoseconds. The energy quantum of this difference is emitted via a photon (green) of a wavelength longer than the absorbed wavelength. Image from (Hochreiter, Pardo-Garcia and Schmid, 2015)

Figure 4 illustrates the difference between bright-field and fluorescent microscopy images. In the panel A, the bright-field image provides general contrast, but is not specific to cellular structures. In the panel B, the fluorescent image displays fluorophore-labeled nuclei and cell cytoplasm, offering a high-contrast visualization of specific cellular components.

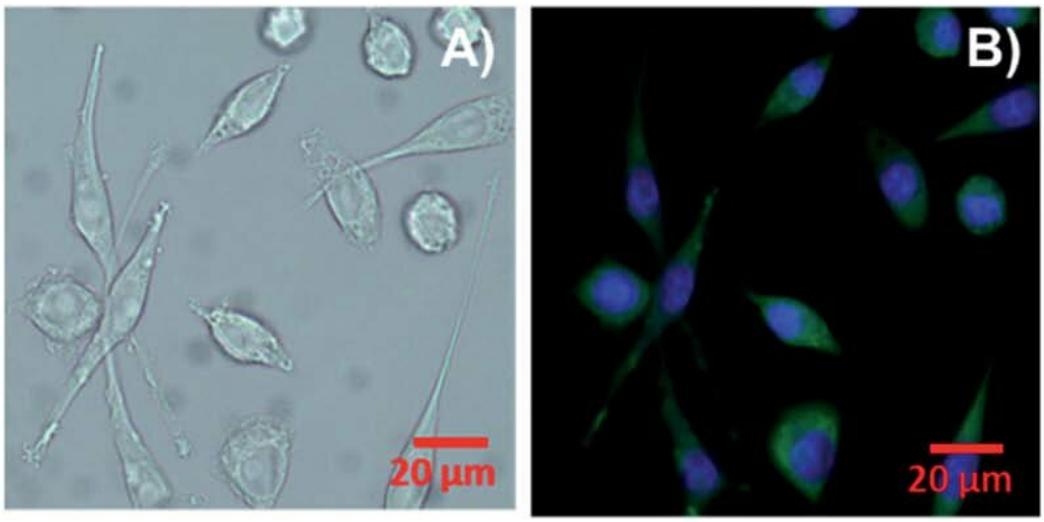


Figure 4. Images of mouse alveolar macrophage cells
This figure shows the bright-field (A) and fluorescence (B) microscopic images of mouse alveolar macrophage cells (RAW 264.7 Cells). Nuclei are blue and cell cytoplasm is green. Image from (Johnson *et al.*, 2021)

This thesis focus on super-resolution microscopy technique (page 13), which is an advanced form of fluorescence microscopy.

I.1.3. Point Spread Function (PSF)

When light passes through the microscope's aperture, it undergoes a phenomenon called diffraction, resulting in the spreading and bending of the light waves (Abbe, 1873). The Point Spread Function (PSF) is the three-dimensional image formed by an infinitely small point light source going through the optical system. Due to diffraction, the resulting image consists of a central bright region, known as the Airy disk, surrounded by rings (**Figure 5**) (Cole, Jinadasa and Brown, 2011). Overall, the PSF determines the resolution of the microscope.

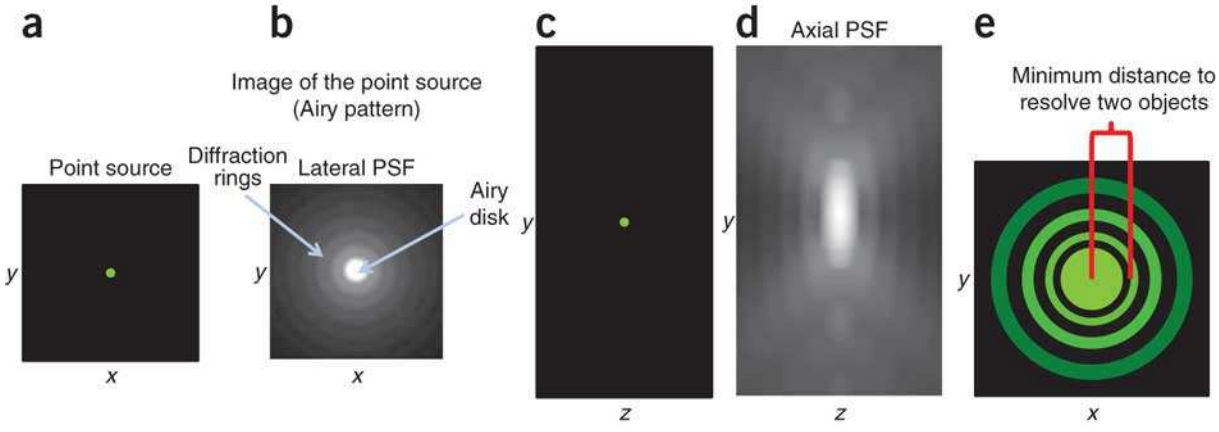


Figure 5. PSF schematics and theoretical functions.
Point source in the xy plane (a) and the yz plane (c). Simulated images for a theoretical PSF for the xy plane (b) and the yz plane (d). (e) The diffraction pattern of the minimum distance that the two objects can be apart. The red lines indicate the center of the Airy disk and the center of the first dark diffraction band within the Airy pattern. Image from (Cole, Jinadasa and Brown, 2011).

I.1.4. Resolution

The resolution in optical microscopy is defined by the minimal distance at which two objects can be distinguished through the optical system. Due to the diffraction, the image of a point source of light does not appear as a single point but as an Airy disk. When two objects are so close that their Airy disks are overlapped, these objects become indistinguishable. As a result, the resolution of conventional optical microscopy is typically limited by the diameter of the airy disk, which is approximately 200-300 nm.

The equation that describes the resolution limit of optical microscopy is given by the following formula:

$$d \approx \frac{\lambda}{2NA}$$

Where λ is the wavelength of the light used for microscopy, and NA is the numerical aperture of the objective. Although the diffraction-limited resolution of optical microscopy has long been considered as a fundamental constraint, the super-resolution microscopy techniques (page 13) have been developed to surpass this limit and provide higher resolution.

I.1.5. Image noise

The noise occurred in images acquired through optical systems is Poisson noise. Mathematically, the signal-to-noise ratio (SNR) is given by:

$$SNR = \sqrt{N}$$

where N represents the number of photons collected.

The more photons that are collected by the camera, the better the SNR becomes. A higher SNR means that the true signal can be more easily differentiated from the background noise. In practice, increasing the excitation laser power increases the photon collection, thus optimizes the image quality.

I.2. Fluorescence microscopy

As discussed in Section above, the fluorescence microscopy allows visualizing specific structures or molecules of interest within a sample, thanks to labeling methods that allows fluorophores bound to the target molecules, such as proteins or DNA.

I.2.1. Labeling

Since most molecules in the cells are not fluorescent, we need to label the molecules of interest with fluorescent molecules. There are several techniques for labeling, including:

- **Direct immunofluorescence** (Coons, Creech and Jones, 1941)

  Antibodies are specialized proteins produced by the immune system to recognize and bind to foreign substances (antigen) (Janeway, 2001). In fluorescent microscopy, antibodies can specifically recognize a target molecule. In direct immunofluorescence method, the antibodies are chemically conjugated with a fluorescent dye (**Figure 6.a**).

- **Indirect immunofluorescence** (Coons, Creech and Jones, 1941)

  We apply an unlabeled primary antibody to bind the target molecule in cells. A secondary antibody, which is conjugated with a fluorescent dye is then applied. These secondary antibodies are bound to the primary thus allowing visualization of the target molecule (**Figure 6.b**).

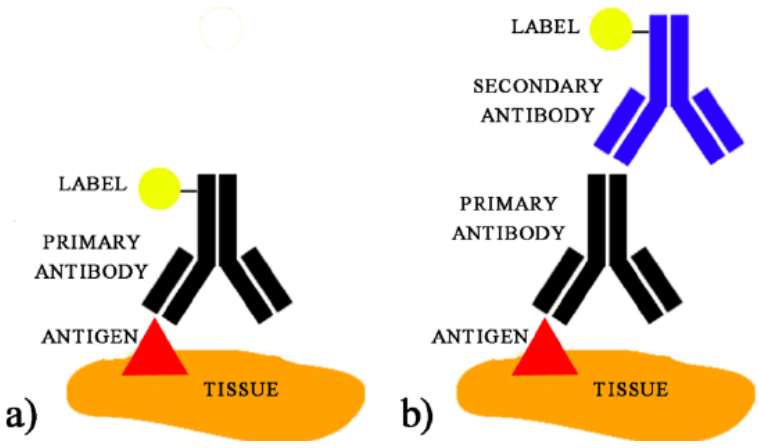


Figure 6. Schematic representation of direct immunofluorescence (a) and indirect immunofluorescence (b).
Image from (Vázquez-Gutiérrez and Langton, 2015)

- **Fluorescent protein fusion** (Chalfie *et al.*, 1994)

  The gene encoding the fluorescent protein is fused to the gene of the protein of interest. This fusion enables protein of interest to become fluorescent.

- **Click Chemistry** (Kolb, Finn and Sharpless, 2001)

  The molecule or structure of interest will be attached with a “clickable” chemical tag. The fluorescent dye with a complementary "clickable" tag undergoes a “click” reaction with this tag. This chemical process allows the attachment of the fluorophore with the targeted molecules.

I.2.2. Types of fluorophore

Fluorophores come in various forms:

- **Organic dyes**

These are small molecules that can be conjugated (attached) to antibodies, nucleic acids, or proteins such as Alexa Fluor dyes (Panchuk-Voloshina *et al.*, 1999) and Cyanine dyes (Mujumdar *et al.*, 1993).

- **Fluorescent proteins**

  These are proteins that are naturally fluorescent. The most famous example is Green Fluorescent Protein (GFP) (Chalfie *et al.*, 1994) and its variants such as mCherry and mOrange (Shaner *et al.*, 2004).

Organic fluorophores generally emit brighter fluorescence and has larger range of excitation and emission wavelengths than fluorescent proteins, and they are often smaller and more chemically stable.

I.3. Super-resolution microscopy

As mentioned above on page 9, conventional fluorescence microscopy is limited by the diffraction to 200-300 nm. Thus, it cannot resolve the structure or molecules at nanoscales.

The development of super-resolution microscopy as an advanced extension of fluorescence microscopy, has overcome the diffraction limit, providing a higher resolution. **Figure 7** presents the diffraction-limited and super-resolution microscopy images of a nuclear pore complex. Due to diffraction, the intricate details of the structure are not clearly resolved in the low-resolution image (**Figure 9.f**). Super-resolution microscopy, however, can provide a much higher level of detail, enabling the visualization of the finer features of the nuclear pore complex.

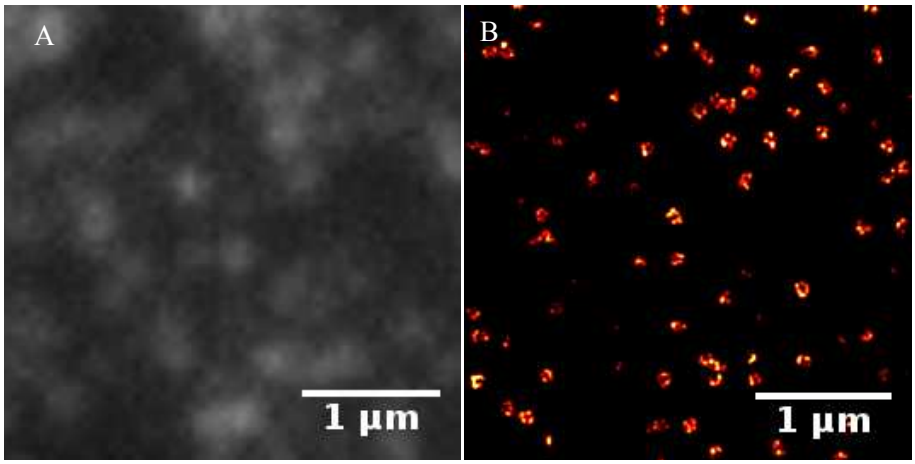


Figure 7. Images of nuclear pore complex.
(A) diffraction-limited image. (B) Super-resolution (SMLM) microscopy image. The SMLM technique will be detailed on page 16. Images acquired by M. Singh, Imaging and modelling unit, Institut Pasteur.

I.3.1. Different techniques of super-resolution microscopy

Three super-resolution techniques will be introduced in this section: Stimulated Emission Depletion (STED) (Hell and Wichmann, 1994), Structured illumination microscopy (SIM) (Gustafsson, 2000; Heintzmann, Jovin and Cremer, 2002) and Single-molecule localization microscopy (SMLM) (Lelek *et al.*, 2021). Each of these approaches uses a different method to achieve the super-resolution:

- **Stimulated Emission Depletion (STED) Microscopy:**

Stimulated emission is a process where an excited fluorophore is induced to a lower energy state by the presence of a photon. As illustrated in **Figure 8.A**, STED microscopy uses two high-intensity laser beams simultaneously: the first laser beam excites the fluorescent molecules in the sample, the second one is a donut-shaped depletion beam (STED beam) that de-activate the fluorescence emission at the periphery through stimulated emission. This allows narrowing the effective point spread function and thus improving resolution.

- **Structured Illumination Microscopy (SIM)**

  SIM uses typically light with a sinusoidal pattern to illuminate the sample. This patterned light interacts with the structures, creating a more complex interference image. Multiple images are recorded with different phases and orientations of the patterned excitation light. Then, by using computational methods, this super-resolution information is extracted from these interference pattern. **Figure 8.B** depicts the principle of the SIM technique.

- **Single-Molecule Localization Microscopy (SMLM)**

  SMLM relies on the stochastic switch-on-off event of fluorophore. Using this random blinking event, the localization of individual fluorophores can be estimated to construct super-resolution images (**Figure 8.C**). In this thesis, we applied this super-resolution technique to our research, and will explain the principle of SMLM in more details on page 16.

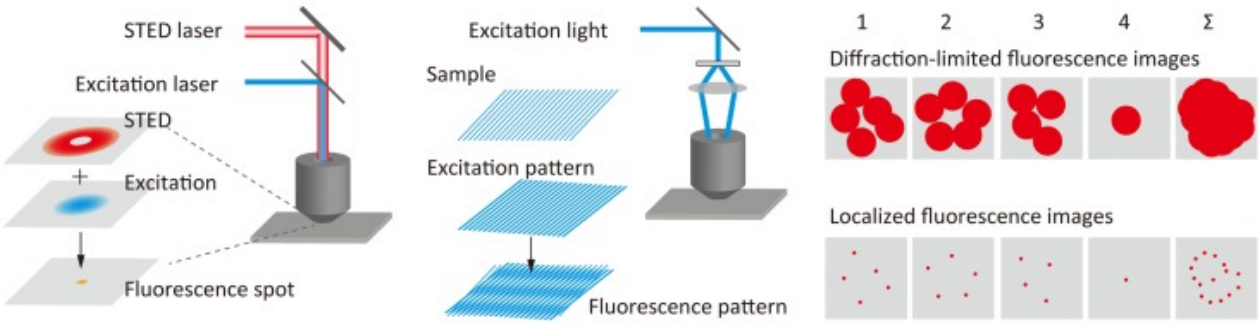


Figure 8. Principle of super-resolution microscopy techniques.
Left: STED microscopy. Middle: SIM. Right: Localization microscopy. Image from (Habuchi, 2014)

I.3.2. Comparison of STED, SIM and SMLM

STED microscopy can achieve resolution below 50 nm. Additionally, the image acquisition is relatively fast and the cost of post-processing computation is relatively low. However, the sample preparation is of moderate complexity. This is because STED relays on high-intensity lasers, limiting the choice of fluorophores to avoid the photobleaching. Additionally, the optical setups are complex because two laser beams should be perfectly aligned. SIM can provide resolution up to 100-150 nm. Its advantages are the easy sample preparation and the image acquisition is the fastest. SMLM stands out with the best resolution (20 nm or better). The trade-off is that it requires a longer acquisition time (often more than 30 minutes) and involves more complex sample preparation and expensive computations for localization estimation (Habuchi, 2014; Schermelleh *et al.*, 2019; Liu, Hoess and Ries, 2022).

Table 1 provides a comparison of these super-resolution microscopy techniques, highlighting their differences in terms of spatial and temporal resolution, as well as the complexity of sample preparation.

Table 1. Comparison of STED, SIM and SMLM.
In this table, red indicates the least favorable option, green represents the most favorable option, and yellow signifies a moderate option.

	STED	SIM	SMLM
Spatial resolution	Below 50 nm	100-150 nm	20 nm or better
Temporal resolution	Moderate	Fast	Slow
Sample preparation	Moderate	Easy	Complex
Computational post-processing cost	Low	Moderate	High
Optical set-up	Complex	Moderate	Easy

I.4. Single molecule localization microscopy (SMLM)

As mentioned on page 15, among different super-resolution microscopy techniques, Single Molecule Localization Microscopy (SMLM) offers the highest spatial resolution (e.g., 20 nm or better). It allows to resolve biological structures at or near the molecular scale, thus has become a powerful tool for studying cell biology.

I.4.1. Principle of SMLM

As illustrated in Figure 9, the SMLM relies on the principle that the spatial coordinates of individual fluorescent molecules can be accurately estimated if their PSFs do not overlap. To avoid this overlap, the emissions of fluorescent molecules are separated by time. This separation is typically achieved by the appearance and disappearance of PSFs. In practice, we need to sequentially acquire thousands of diffraction-limited image frames of the same field of view (FoV), ensuring that most fluorescent molecules are in the ON state in at least one frame of the acquisition sequence. These images are then computationally processed to detect all ON molecules and compute their coordinates by estimating the centroid position of each PSF. Finally, all the localizations are aggregated to construct a single, high-resolution image.

Very often we acquire a widefield image at the beginning of the acquisition, which is a conventional diffraction-limited fluorescence microscopy image (**Figure 9.f**)

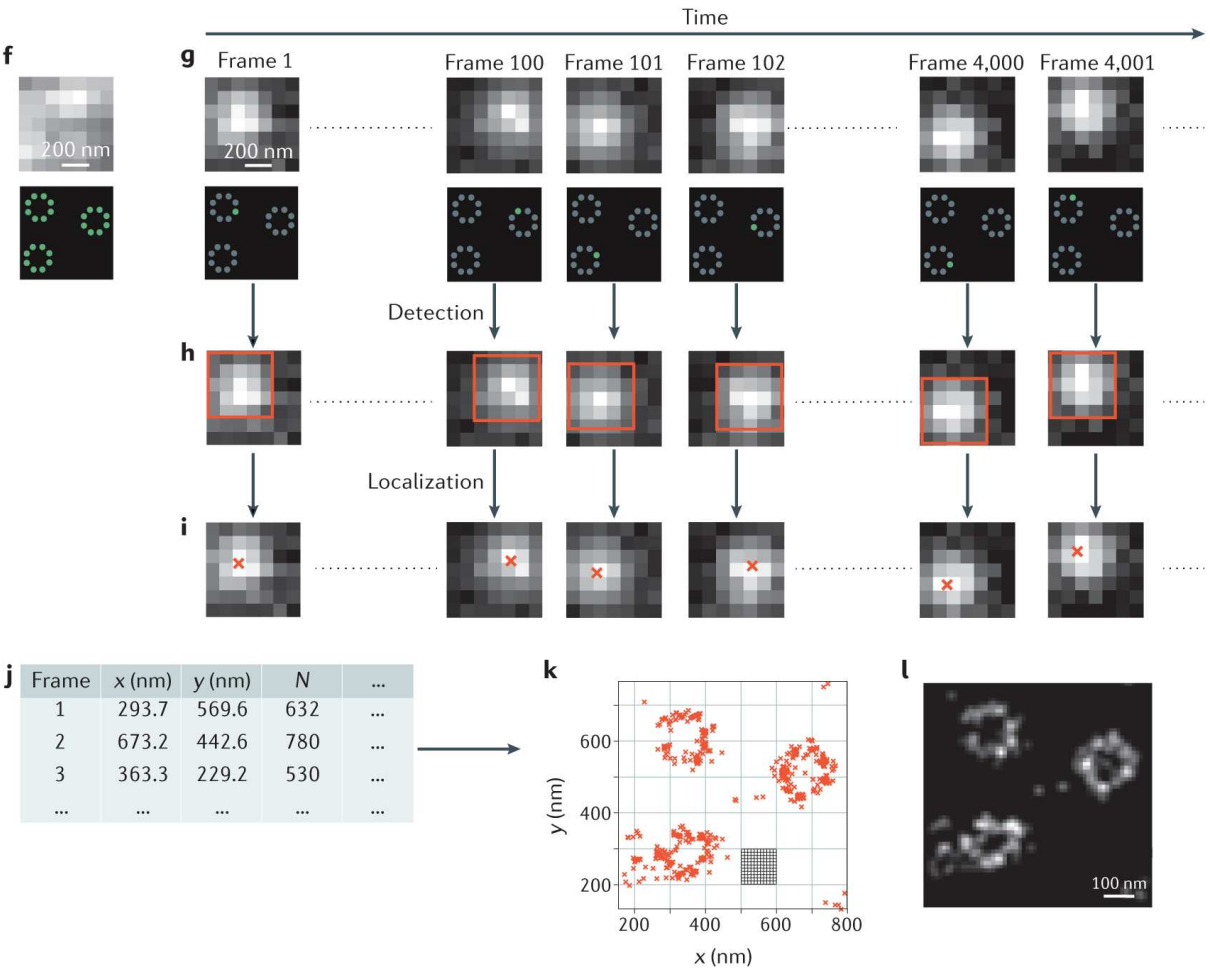


Figure 9. Principle of Single Molecule Localization Microscopy.
(f) diffraction-limited image of the nuclear pores complex, with all fluorescent molecules switched ON. (g) Sequential diffraction limited images of the same field of view (FoV) as in (f), where only a few molecules are ON at the same time, while others are OFF. (h, i) In each frame, single molecules are computationally detected (h) and localized (i). (j) Localization table, where each row represents a distinct localization event and columns indicate x, y coordinates and additional information such as frame number and N. (k, l) The molecules are aggregated into a picture. E.g. a scatter plot (k), or 2D histogram (l). When we put all this information together, we get a 'super-resolution' image that lets us see the ring-like structure of nuclear pores. Image from (Lelek *et al.*, 2021).

I.4.2. Fluorophores for SMLM

It exists different fluorophores to the appearance and disappearance of PSFs (Lelek *et al.*, 2021):

- **Photoswitchable fluorophores**

Photoswitchable fluorophores, such as Alexa Fluor 647, Cy5 or Dronpa (Andresen *et al.*, 2007), are molecules capable of undergoing transitions between “ON” and “OFF” states multiple times when exposed to specific wavelengths of light. This switching

between “ON” and “OFF” states occur in presence of an oxygen scavenger buffer solution (photoswitching buffer) that facilitates the photoswitching behavior of fluorescent molecules.

- **Photoactivatable fluorophores**

  Photoactivatable fluorophores, such as PA-GFP (Photoactivatable Green Fluorescent Protein), PAmKate and PAmCherry (Lukyanov *et al.*, 2005), are initially in the “OFF” state and can be switched to “ON “state either spontaneously or via UV light. However, once in “ON” state, they generally cannot be switched back to the “OFF” state until the photobleaching.

- **Photoconvertible fluorophores**

  Switched to the “ON” state, photoconvertible fluorophores can change their emission color to another when exposed to light of a specific wavelength. Some well-known photoconvertible are Eos and its derivatives, Dendra and mMaple, etc. (Lelek *et al.*, 2021).

- **Spontaneously blinking dyes**

  Spontaneously blinking dyes can be switched between “ON” and “OFF” states without requiring a photoswitching buffer. These dyes include the silicon rhodamine dye HMSiR, HEtetTFER and FRD, etc. (Lelek *et al.*, 2021).

- **Temporarily binding dyes**

  Unlike the dyes mentioned above, that bind irreversibly to the molecules of interest, temporarily binding dyes spontaneously bind and unbind with the target molecules. When dyes are bound, PSFs appear, and when dyes are unbound, the PSFs disappear. Any organic dye (discussed on page 12) that exhibits sufficient fluorescence, such as Cy3B, can be used as temporarily binding dyes (Lelek *et al.*, 2021).

I.4.3. Different techniques of SMLM

Depending on the dyes used, SMLM can be achieve using different techniques such as:

- **Photo Activated Localization Microscopy (PALM)** (Betzig *et al.*, 2006)

  PALM uses photoactivatable or photoswitchable fluorescent proteins. At the beginning of the image acquisition, most fluorophores are in an "OFF" state, and a subset is photoactivated with a specific wavelength of light to emit fluorescence. Because PALM uses encoded fluorescent proteins, it is ideal for live-cell imaging.

- **Stochastic Optical Reconstruction Microscopy (STORM)** and its variants (Rust, Bates and Zhuang, 2006; Heilemann *et al.*, 2008)

  STORM and its variants, such as direct Stochastic Optical Reconstruction Microscopy (dSTORM), use photoactivatable or photoswitchable organic dyes instead of fluorescent proteins through suitable photoswitching buffers. STORM and related techniques achieve higher resolution than PALM, because organic dyes are generally both smaller and brighter than fluorescent proteins.

- **Points Accumulation In Nanoscale Topography (PAINT)** (Sharonov and Hochstrasser, 2006)

  PAINT does not need photoswitching buffer, it capitalizes on the switching between free diffusion and immobilization binding of fluorophores to the target. DNA-PAINT (Schnitzbauer *et al.*, 2017) is the most well-known variant of PAINT (Lelek *et al.*, 2021), where the transient immobilization is achieved by hybridization of DNA strands. The temporarily binding dyes described earlier are applied in PAINT-based methods.

- **Minimum Emission Fluxes (MINFLUX)** (Balzarotti *et al.*, 2017)

  MINFLUX merges the principles of localization microscopy, like PALM or STORM, using the depletion lasers as stimulated emission depletion (STED) microscopy **(**page 13**)**. This technique uses a donut-shaped excitation beam combined with a three-point estimator to determine the position of emitter accurately, as illustrated in **Figure 10**.

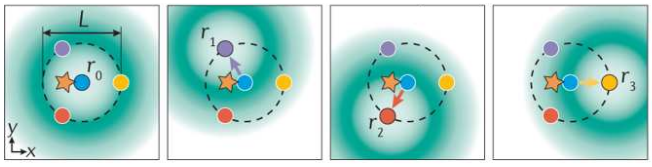


Figure 10. MINFLUX excitation concept.

A doughnut-shaped excitation beam (green) is moved sequentially to four probing positions r0, r1, r2 and r3 (colored circles; probing range L) near a single fluorophore (orange star). If the center of the doughnut perfectly aligns with the position of the fluorophore, no photons will be emitted. Image from (Sahl and Hell, 2019).

I.4.4. Limitation of SMLM

While SMLM has enabled researchers to achieve high resolutions imaging in microscopy, it still has its limitations. During SMLM acquisition, many thousands of diffraction-limited low resolution frames with exposure times of ~10-100ms are required to reconstruct a single high-resolution image. Consequently, although SMLM allows to visualize the biological structures at near nanoscale, image acquisition in SMLM is slow.

Furthermore, the quality of the reconstructed images may be influenced by various factors such as noise due to limited number of collected photons per fluorophore, the potential overlapping PSFs and sample drift that occurs during the image acquisition. Additionally, SMLM can be limited by phototoxicity. Phototoxicity refers to the cellular damage that occurs when a biological sample is exposed to excitation light during imaging.

I.4.5. Frustration pyramid for SMLM

The limitation mentioned above often leads to 'frustration pyramid'. The frustration pyramid describes the trade-offs between spatial resolution, temporal resolution, the size of Field of View (FoV), signal-to-noise ratio (SNR) and phototoxicity in the context of SMLM. Balancing these components is crucial for high-quality SMLM imaging, especially in live cells.

Figure 11 shows the frustration pyramid. For example, enhancing spatial resolution or SNR necessitates longer acquisition time, sacrificing the temporal resolution. And in order to visualize larger FoV, either spatial or temporal resolution needs to be reduced. As for recovering the fast-cellular dynamics, we need to enhance the temporal resolution of SMLM. However, increasing temporal resolution leads to lower spatial resolution and SNR. Additionally, the phototoxicity increases when using higher intensity excitation light or longer exposure time.

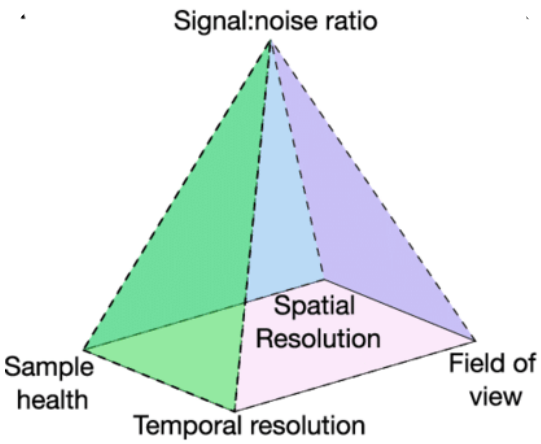


Figure 11. Pyramid of frustration.
Image from (Brameshuber *et al.*, 2022)

I.4.6. SMLM imaging in live cells

Building on the trade-offs outlined above, SMLM imaging in live cells is challenging. These challenges are mainly due to the rapid cellular motion and relatively low temporal resolution of SMLM. The limited temporal resolution of SMLM makes it hard to collect enough localizations to provide a super-resolution snapshot before the biological structure rearranges. Additionally, the need for high laser (introduced on page 10) is problematic for live cells, because it leads to photobleaching and phototoxicity, especially when using UV activation (Shroff *et al.*, 2008). As a result, SMLM imaging in live cells has mostly been limited to proofs of principle on relatively slow-moving structure (Shroff *et al.*, 2008; Jones *et al.*, 2011; Huang *et al.*, 2013; Holden *et al.*, 2014).

There are other computational approaches for live-cell imaging:

- **Super-Resolution Radial Fluctuations** (SRRF) method (Gustafsson *et al.*, 2016) relies on the radial symmetry in each pixel of the image sequences during the acquisition. The radial symmetry of a pixel is high if the intensity value of a pixel and its neighbors lie along a circle. SRRF analyses the fluctuations in radial symmetry of each pixel over the entire acquisition. It then uses this information to assign the likelihood that each pixel belongs to a structure of interest.
- **Super-resolution Optical Fluctuation Imaging** (SOFI) (Dertinger *et al.*, 2009) relies on temporal fluctuations in fluorescence intensity at each pixel of the acquired image sequence during the fluorophore switched between “ON” and “OFF” states. SOFI computes higher-order correlation of these fluctuations, meaning the correlation

between the intensity in one pixel and another pixel across multiple time points. The information from these higher-order correlations is then used to reconstruct a new image with a high resolution.

SRRF and SOFI allows the reconstruction of images in live cells. However, they generally provide lower spatial resolution compared to SMLM and the quality of the images depends on SNR of the original sequences.

In response to these challenges, deep learning (page 23) has become a promising approach for enhancing the temporal resolution of SMLM imaging (Möckl, Roy and Moerner, 2020), which will be discussed in the next section.

I.5. Deep learning

Deep learning is a subset of artificial intelligence (AI), and it has revolutionized various fields, such as computer vision, natural language processing, speech recognition and robotics. Deep learning focuses on the use of artificial neural networks to learn representations of data with multiple levels of abstraction (LeCun, Bengio and Hinton, 2015).

I.5.1. Structure of neural networks

Artificial neural networks, inspired by the structure and function of the human brain, are made up of connected neurons (McCulloch and Pitts, 1943). These neurons, when properly trained with a large amount of data, can learn to approximate complex non-linear representations within the data.

Neurons are the fundamental building block of artificial neural networks. A neuron consists of three main components: inputs, an activation function and an output. The Output (y) is the result of applying the activation function (f) to the weighted ($w_{1,2,\dots N}$) sum of N inputs ($x_{1,2,\dots N}$) (**Figure 12**):

$$y = f(z)$$

$$z = \sum_{j=1}^{N} w_j x_j + b$$

Where b is the bias. Various activation functions can be used in artificial neurons, such as the sigmoid ($y = S(z) = \frac{1}{1+e^{-z}}$), hyperbolic tangent (tanh: $y = \frac{e^{2z}-1}{e^{2z}+1}$), and rectified linear unit (ReLU $y = max(0, y)$) functions. The output can serve as input to other neurons in a network.

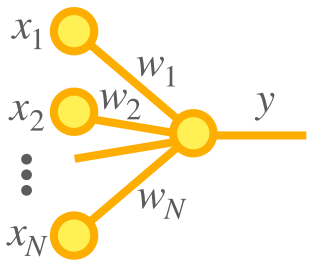


Figure 12. Operation of a single artificial neuron.
This figure illustrates the operation of a neuron (middle). The Output (y) is the result of applying the activation function to the weighted ($w_{1,2,...N}$) sum of N inputs ($x_{1,2,...N}$). Image from (Krenn *et al.*, 2023)

A neural network is structured into layers, with each layer containing a number of neurons. The neurons in one layer are connected to the neurons in the subsequent layer. Overall there are three types of layer: the input layer, hidden layers and the output layer.

Consider an example of image classification, where we aim to train a deep learning network to distinguish between image of cats and dogs:

- The input layer receives the input of neural network, which could be the pixel values of the image.
- The output layer produces the final output of the network, which in this case is two classes: cat (output = 0) and dog (output = 1).
- Hidden layers are the layers between the input and output layers. The neurons in these layers transfer features, such as edges and shapes, from the input layer to the output layer. The term "deep" in deep learning refers to the depth of neural networks, representing by the number of hidden layers

Three different types of layers are convolutional layer, pooling layers and fully connected layers:

Convolutional layers consist in sliding a small window (also known as filter or kernel) across the image. Each convolution operation yields a 2D feature map, also known as a feature map (also known as channel). Mathematically, the convolution is defined as a sum of the elementwise multiplication of the kernel values and the input values (**Figure 13**). The number of filters in a convolutional layer determines the number of feature maps in the output of the layer.

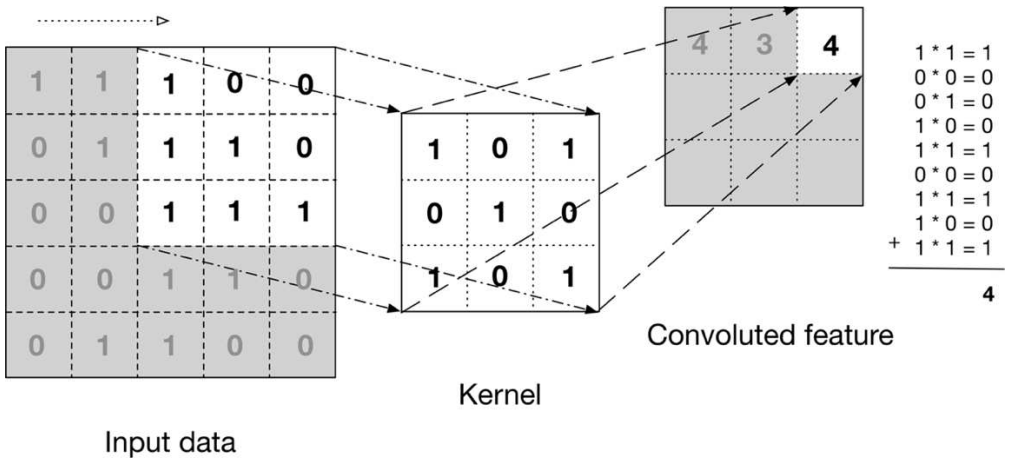


Figure 13. Convolution operation.
A 3X3 pixel kernel applied on input image, resulting in convoluted feature. Image from (Mandal, 2021)

- **Pooling layers** down-sample the spatial dimensions of the input, thereby reducing the computational cost and enhancing the network's translation invariance. This means that the network can extract features from an image irrespective of their positions. **Figure 14** illustrates the two different types of pooling layers: max pooling and average pooling.

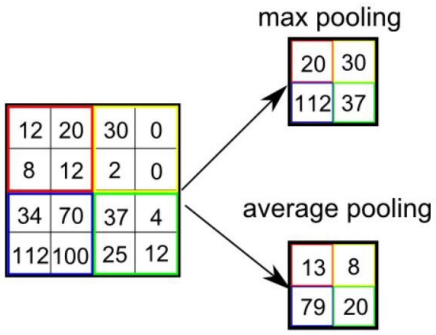


Figure 14. Pooling layer.
The max pooling (above) computes the maximum pixel value within a region of the image, defined by the size of the kernel. And the average pooling (below) calculates the mean pixel value within the region. Image from (Mandal, 2021).

- **Fully connected layers** connect each neuron in one layer to every neuron in another layer. They are generally utilized in the final stages of the network for classification tasks (Figure 15).

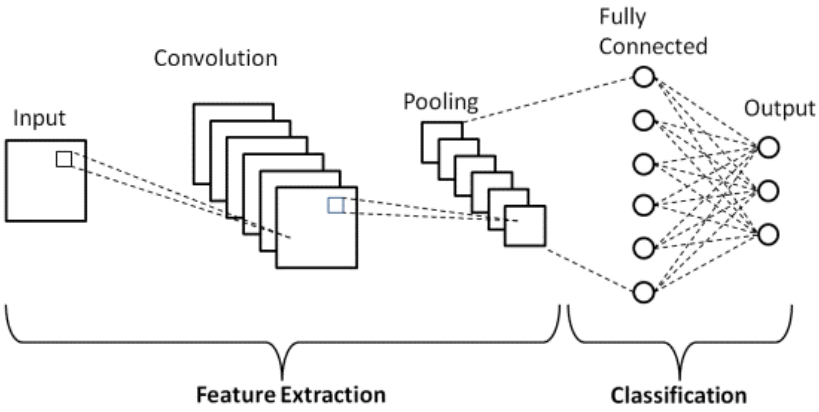


Figure 15. Illustration of a typical CNN for classification task.
The network is composed of convolution layers, pooling layers and fully connected layers. The output layer computes the classes of images. Image from (Phung and Rhee, 2019).

I.5.2. Parameters

Parameters of deep learning is learned through training phase so that the models make accurate predictions. These include weights and biases of neurons in neural networks (page 23), filters in Convolutional Neural Networks (page 25) and gate values in Recurrent Neural Networks (page 36).

I.5.3. Training

Deep learning models are computer programs that are trained to recognize patterns in data or make predictions (*Machine Learning Models: What They Are and How to Build Them*, 2023). In the training process of a deep learning model

In the training process of a deep learning model, the algorithm performs a sequence of computational steps to refine its parameters. This process begins with forward propagation, where data is fed into the input layer and passed through each layer of the network until an output is produced at the output layer.

Next, a loss function, also known as cost or objective function, quantifies the difference between the network's predictions and the ground truth, particularly in the context of supervised learning (Goodfellow, Bengio and Courville, 2016) (page 28).

After computing the loss function and its gradient according to the parameters of the neural network, the model undergoes gradient backpropagation to adjust it’s the parameters.

These gradients indicate the contribution of each neuron on the loss function. Based on these computed gradients, the parameters are then adjusted such that the loss function is minimized. The loss minimization leverages gradient-based optimization techniques, such as stochastic gradient descent, to iteratively update the model parameters (Rumelhart, Hinton and Williams, 1986).

Iteratively though forward propagation and backpropagation, the deep learning model progressively improves accuracy of the predictions, effectively learning from the training data.

I.5.4 Validation and test phases

During the validation phase, we evaluate the model's prediction on unseen validation data (page 27). When model overfits, its performance decrease. Monitoring the model's performance on the validation data allows to stop the training when the validation performance starts to degrade, thus preventing overfitting.

The test phase evaluates the performance of trained model on completely new data, after all tuning and training processes are complete. This involves in performing the trained model on real-world data.

I.5.5. Data partition

The data is the core of deep learning, however, simply having data is insufficient. Effective data partition ensures that the deep learning models generalize effectively to new, unseen data. This section introduces the data partition in deep learning:

- **Training data**

  Training data is used to train the deep learning model, during the training phase (page 26). The insufficient training data may lead to underfitting, meaning the model fails to learn the underling representation of the data (page 29).

- **Validation data**

  In general, we split the data into training and validation data, however, the validation data is not used during the training. The loss function computed from the validation data

can determinate weather the model overfits. The overfitting is when a model fails to predict unseen data (page 29).

- **Testing data**

  After training and validation, the testing data offers assessment of the model's performance, reflecting whether the model can be generalized to unseen data.

- **Cross-validation**

  Cross-validation is used to

  ensure that the model's performance doesn't depend on the data partition. In practice, the original training data is partitioned into multiple subsets. The model is trained and validated for multiple times, each time using a different subset as the validation data and the remaining subsets as the training data. The model performance is averaged over the cross-validation.

I.5.6. Different learning paradigms

Deep learning models are trained using either labeled, unlabeled, or mixed data. Based on the data, deep learning has four different learning paradigms: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning (Chollet, 2018):

- **Supervised learning**

  When the input of training data is paired with the correct output (also known as ground truth), we apply supervised learning. Model is trained on this labelled data to learn to predict on new, unseen labeled data. This paradigm is commonly applied in various deep learning applications, including image classification(Simonyan and Zisserman, 2015; Krizhevsky, Sutskever and Hinton, 2017), language translation (Sutskever, Vinyals and Le, 2014; Vaswani *et al.*, 2017), and optical character recognition(Kim, 2014).

- **Unsupervised learning:**

Model is trained on an unlabeled dataset, meaning that the ground truth is not provided during training. Because the output variables are unavailable, unsupervised learning focuses on extracting meaningful patterns from the input data. Unsupervised learning can be used for data visualization (Maaten and Hinton, 2008), data denoising by learning the properties of the noise (Krull, Buchholz and Jug, 2019; Sheth *et al.*, 2021), and clustering(Xie, Girshick and Farhadi, 2016).

- **Self-supervised learning**

  Self-supervised learning combines both labelled and unlabeled data for training, the labels are often automatically generated from the input data itself (X. Chen *et al.*, 2023). Similarly, using future input data as a target is also considered self-supervised learning, such as using past frames to predict the next frame in a video (Sermanet *et al.*, 2018).

- **Reinforcement learning**

  In reinforcement learning, an "agent" interacts with its environment to maximize rewards or minimize penalties based on the actions it takes. For example, it can be applied to play Atari games (Mnih *et al.*, 2013) or to compete in the game of Go at the highest level (Silver *et al.*, 2017).

I.5.7. Challenges of deep learning

Despite considerable progress in the field of deep learning, it also faces several challenges, such as:

- **Underfitting**

  Underfitting occurs when models fail to capture the patterns of the training data. The output quality of model is not in good agreement with the ground truth during the training and validation phase (Goodfellow, Bengio and Courville, 2016).

- **Overfitting**

Overfitting occurs when a model captures not only the patterns of the training but also its noise. The model performs well on training data, but not on validation or testing data (Goodfellow, Bengio and Courville, 2016).

- **Model mismatch**

The model mismatch is when deep learning models exhibit artifacts when testing them on data that diverges from the training distribution (Recht *et al.*, 2018).

I.6. Deep learning for computer vision

Given that this thesis focuses on the application of deep learning to SMLM images, this section is dedicated to introducing various deep learning models specifically designed for computer vision tasks.

Computer vision is one of the applications of AI, it allows interpreting and understanding visual information form of images or videos. The goal of computer vision is to replicate and even surpass the human visual system's capabilities. Computer vision has wide-ranging applications that we enumerate in the following section.

I.6.1. Different applications of computer vision

Computer vision tasks include:

- Image classification: assigning a label to an image based on its content (Boland and Murphy, 2001; He *et al.*, 2016; Krizhevsky, Sutskever and Hinton, 2017).
- Object detection: locating and identifying objects within an image (Redmon *et al.*, 2016; Ren *et al.*, 2016; Cheng *et al.*, 2018).
- Image segmentation: assigning a class label to each pixel on an image (Ronneberger, Fischer and Brox, 2015; Chen *et al.*, 2018).
- Image generation: generating new images that similar to a given dataset (Goodfellow *et al.*, 2014).
- Image translation: transforming an image from one domain to another (Ledig *et al.*, 2017; Karras, Laine and Aila, 2019)
- Image restoration: Image denoising (Zhang *et al.*, 2017; Krull, Buchholz and Jug, 2019) or image deconvolution (Xu *et al.*, 2014).

I.6.2. Convolutional Neuron Networks (CNNs)

Convolutional Neural Networks (CNNs) are one category of the deep learning models that are widely used for computer vision tasks. The convolution operation allows CNNs to automatically and adaptively learn spatial hierarchies of features, like edges, corners, textures,

or more complex structures depending on the depth of the layer in the network (Lecun *et al.*, 1998).

I.6.3. U-net

U-Net is a CNNs-based architecture that was initially designed for biomedical image segmentation. It is designed to process images in a way that learns both fine details and larger contextual information. The **Figure 16** illustrated the results of cell segmentation using U-net architecture.

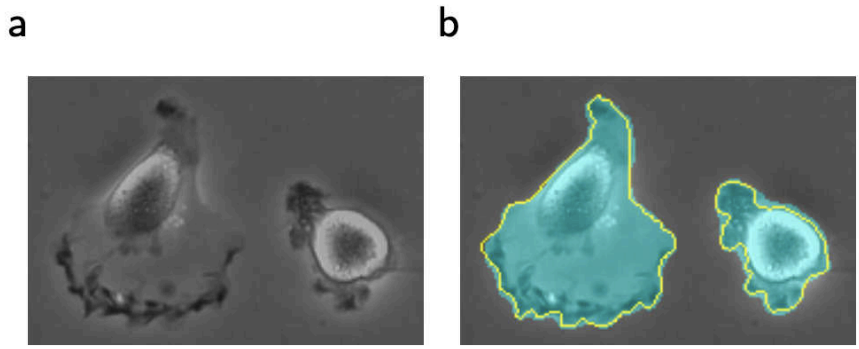


Figure 16. Result of U-net segmentation (Ronneberger, Fischer and Brox, 2015).
(a) input image. (b) Segmentation result (cyan mask) with manual ground truth (yellow border). Image from (Ronneberger, Fischer and Brox, 2015).

The architecture resembles the letter 'U' and comprises of two parts: the encoder path and the decoder path (**Figure 17**). The encoder path has a series of convolutional layers followed by a ReLU activation function and max-pooling layers. As we move deeper into the encoder path, the spatial dimensions of the feature maps decrease, and the number of channels increases. The decoder path mirrors the encoder path in reverse. It consists of a series of up-sampling operations followed by convolutional layers. Additionally, at each level of the encoder path, the feature maps from the same level of the decoder path are concatenated, named as skip connections. These skip connections allows the network to preserve fine details in the output (Ronneberger, Fischer and Brox, 2015).

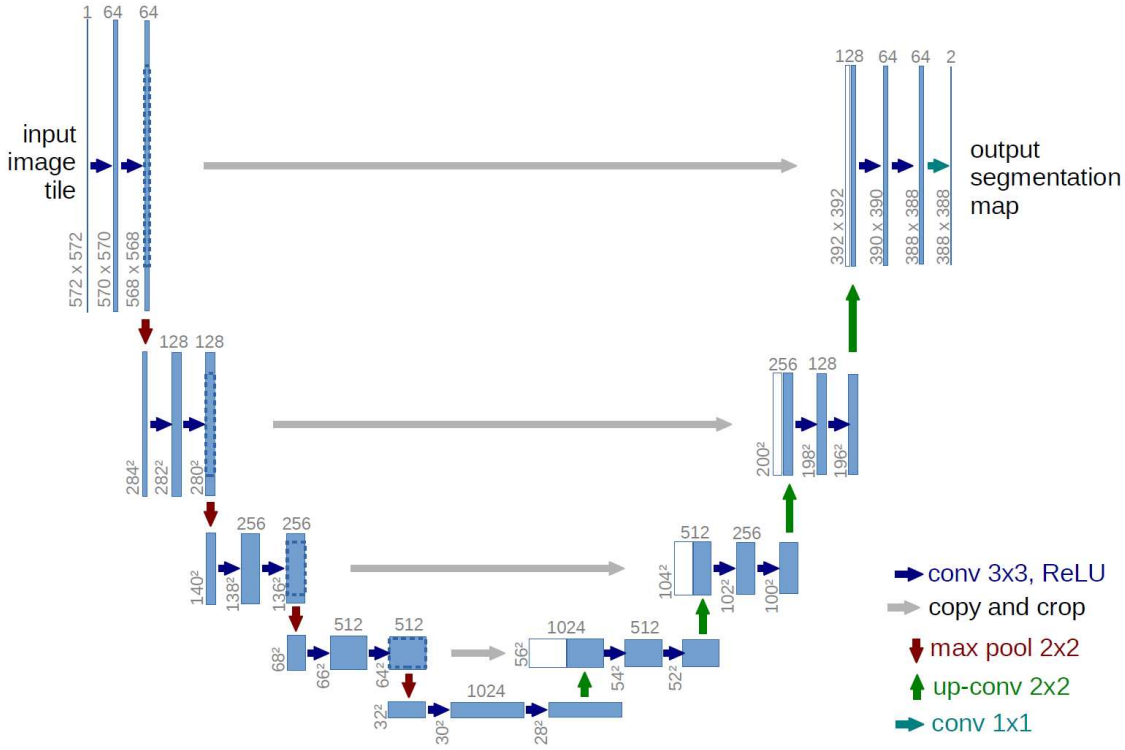


Figure 17. U-net architecture.
Blue boxes are the multi-channel feature map whose number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes are concatenated feature maps. The arrows denote the different operations. Image from (Ronneberger, Fischer and Brox, 2015).

I.6.4. Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) is an unsupervised deep learning model for generating data. GANs are made of two neural networks, a generator and a discriminator, that are trained against one another through an adversarial process. The generator is trained to generate data from random noise. It does this by gradually improving its ability to generate fake data (that look real to fool the discriminator) based on feedback from the discriminator. The discriminator is trained to distinguish between real data and fake data generated by the generator and compute a probability that the data is real. The discriminator is trained to correctly classify real and fake data, while the generator is trained to produce data that the discriminator cannot classify as real (**Figure 18**).

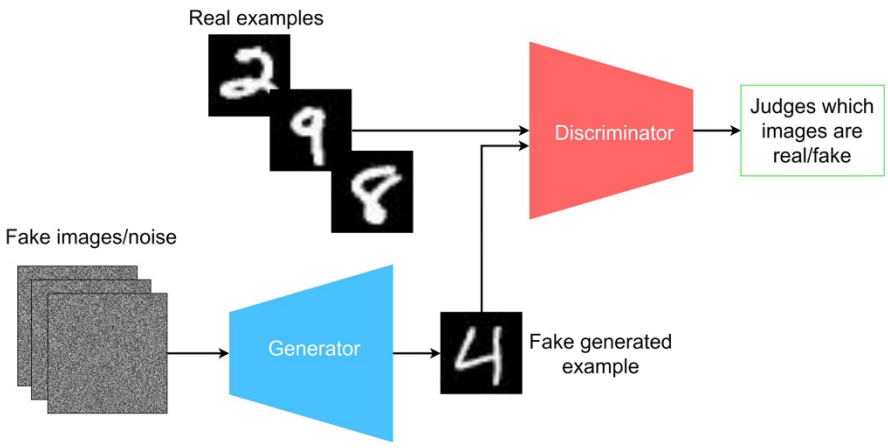


Figure 18. GAN model structure.
Image from (Caper, 2022)

Mathematically, the training process can be described by:

$$min_G max_D V(D,G) = \mathbb{E}_{x \sim p_{dada(x)}}[logD(x)] + \mathbb{E}_{z \sim p_{z(z)}}\left[\log\left(1 - D\left(G(z)\right)\right)\right]$$

Where D(x) is the discriminator's probability estimation of a real data x to be real, G(z) is the data generated by the generator given a noise vector z, D(G(z)) is the discriminator's probability estimation of a fake data to be real (Goodfellow *et al.*, 2014).

I.6.5. Pix2Pix

Pix2Pix is a Conditional Generative Adversarial Network (cGAN). Unlike GAN, it is a supervised learning model. In pix2pix, the generator is conditioned on the input image to generate a corresponding image close to the ground truth. The discriminator is conditioned on the same input image and trained to distinguish real images and fake images generated by the generator. Pix2Pix can perform image-to-image translation, where the input image serves as a guide for the image generation in a controlled manner (Isola *et al.*, 2018).

As shown in **Figure 19**, the model is trained to predict the photo from the hand-drawing edges. The generator and discriminator are trained against each other and both are conditioned on the input image, which is the edges.

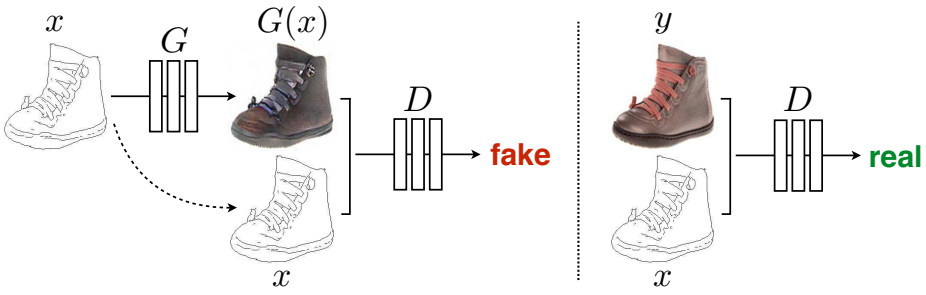


Figure 19. Conditional GAN.
Model is trained to predict realistic photo from the edges. Both generator and discriminator are conditioned with the input image (edges). Image from (Isola *et al.*, 2018).

I.6.6. Exploiting the temporal information

As mentioned on page 21, SMLM's temporal resolution limits its applications for cellular dynamic studies. In order to enhance temporal resolution, in this thesis, we will benefit from deep learning based methods. Several deep learning architectures have been specifically designed to handle temporal information for computer vision tasks:

3D Convolutional Neural Networks (3D CNNs):

3D CNN is an extension of traditional 2D CNN (discussed on page 31). By adding the temporal information as the third dimension, the model can learn spatiotemporal features (Tran *et al.*, 2015). Like 2D CNN, the translation invariance of 3D CNN allows it to capture the patterns in both spatial and temporal dimensions.

Recurrent Neural Networks (RNNs):

RNNs use sequential data. In an RNN, the output from a previous step is fed into the network as input for the next step, allowing the network to maintain the information from the past. The simplest form of RNNS**,** called vanilla RNN, is composed **of** input layers, recurrent (or repeating) layers and the output layers, as illustrated in **Figure 20**. However, for vanilla RNN, it is challenging to handle long sequences due to the vanishing and exploding gradient (Rumelhart, Hinton and Williams, 1986). For example, when the sequential length is large (e.g. h_0 *and* h_t), the capacity of handling the information diminishes.

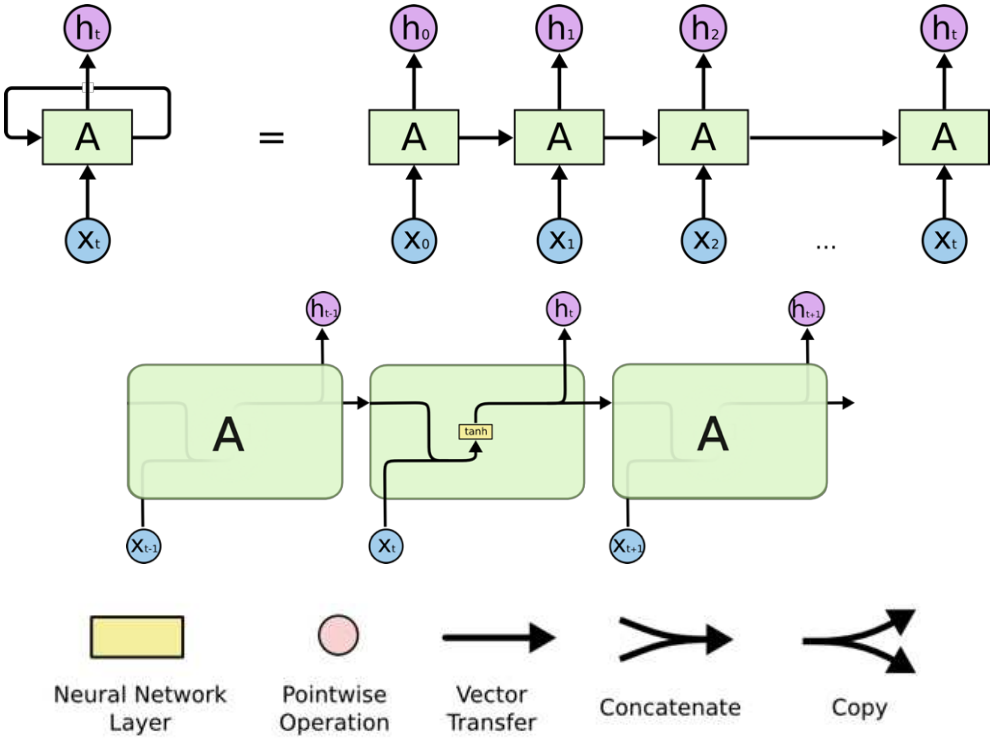


Figure 20. Illustration of a vanilla RNN.
The upper panel shows the vanilla RNN structure, where X_t is the input layer, h_t is the output layer. The block A is the recurring module that allows to pass the sequential data from one step to the next step. The lower panel shows in details how this repeating module works. The output from the previous step is copied and concatenated with the input, then this concatenated information is fed into the network as the input for the next step. Images from (*Understanding LSTM Networks -- colah's blog*, no date).

Long Short-Term Memory (LSTM)

LSTMs are a variant of RNNs. Compared to the vanilla RNN, which has one layer for each recurrent layer, the LSTM has four as depicted in **Figure 21**.

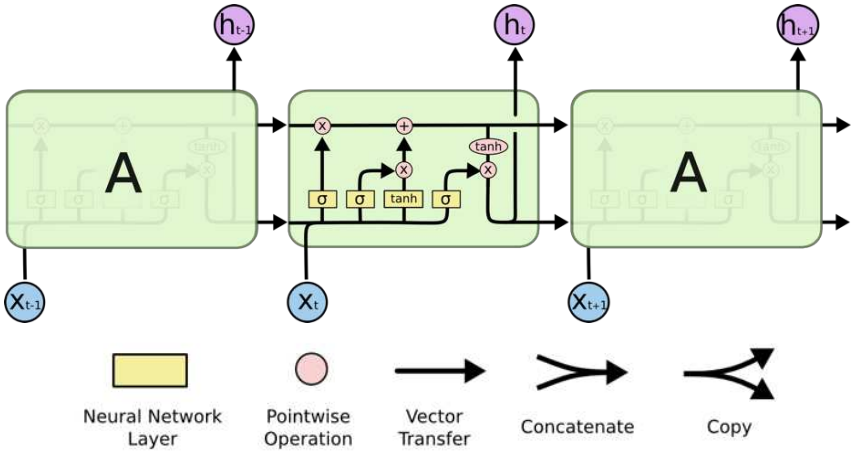


Figure 21. The recurrent module of LSTM networks.
Image from (*Understanding LSTM Networks -- colah's blog*, no date).

As shown in **Figure 21**, the horizontal bar on the top of the diagram is cell state, it allows the information passing through the network during the training. While the gates (yellow blocks) update or remove the information from the cell state.

Frist, the output from the previous step is concatenated with the current input and serve as input for the recurrent module. This new input passes through a sigmoid layer called the "forget gate". The output of this layer is ranged between 0 (completely forget this information from cell state) and 1 (completely remember this information).

Then, another sigmoid layer called the "input gate" computes which values of the input will be updated to the cell state. Then a tanh layer scales the input value between -1 and 1 and this will be multiplied by the output of the input gate. The output of this step will be added on the cell state.

Finally, we apply a sigmoid layer which decides what information of the cell state will serve as output. Then, we apply the tanh layer on cell state and multiply it by the output of the sigmoid gate (called output gate).

Thus, LSTMs allow to forget or update the information through the training process, they can better handle the data with long sequence (Hochreiter and Schmidhuber, 1997).

Vision transformer for videos:

The Transformer (Vaswani *et al.*, 2017) was initially introduced for natural language processing. Unlike RNNs relying on recurrent layers to process sequential data, Transformers utilize attention mechanisms to handle dependencies between input and output.

Before feeding the input into the Transformers, we pre-process the input data into a set of elements (tokens), this process is called tokenization. Then these tokens are mapped to fixed-size continuous vectors using word embedding. In order to provide the information about the order and position of tokens in the sequence, positional encoding is added to the embedding vectors.

Taking an example of the processing of "Hello, world.". This sequence could first be tokenized to tokens ["Hello", "world"], and mapped to continuous vectors: for example, "Hello": [0.1, 0.3, 0.2] and "world": [0.5, 0.1, 0.7], and the position encoding for these tokens

could be [0.0, 0.0, 0.1] and [0.0, 0.0, 0.2] respectively. Thus, the input will be [0.1, 0.3, 0.2] + [0.0, 0.0, 0.1] = [0.1, 0.3, 0.3] for "Hello" and [0.5, 0.1, 0.7] + [0.0, 0.0, 0.2] = [0.5, 0.1, 0.9] for "world".

The fundamental concept of the Transformers is the attention mechanism, which is composed of key (K), value (V) and query (Q). Taking an example of searching for a website on Google. When we search (query) for a website, the Google search engine will map the query against the information (keys) associated with websites. Then the algorithm will list the websites (values) according to the relevance. In practice, the K, Q, and V are three copies of the input data. The attention mechanism denoted as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where d_k is the dimension of Q and K. Attention is computed by weighted values (V). The weights are calculated by the dot products of the query (Q) with all keys (K), scaled by $\sqrt{d_k}$, and apply a softmax function. As illustrated in left panel of **Figure 22**, the Q, K and V are computed by multiplying with weight matrices W_Q, W_K and W_V respectively. The output of such attention mechanism is called one "head". Instead of single attention, the original paper used "multi-head attention" mechanism by concatenating h different attention outputs. This multiple attention is achieved by operating different weight matrices. The output of the multi-head attention is then projected back into the original dimensions (**Figure 22** right panel).

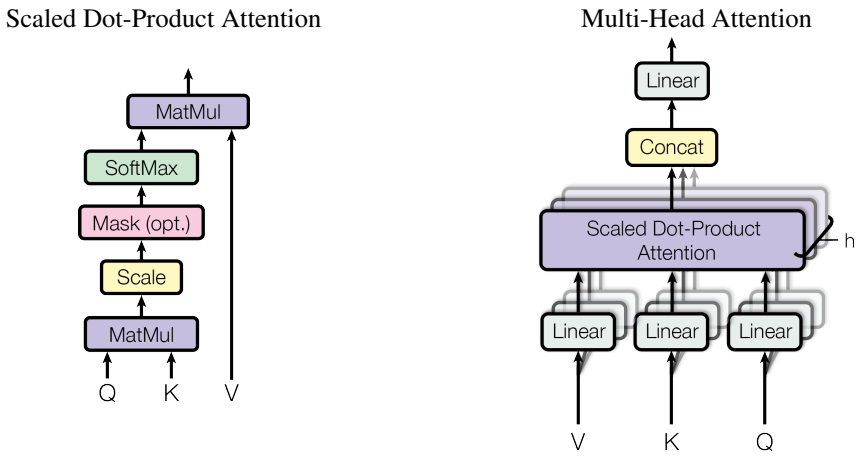


Figure 22. Scaled Dot-Product Attention (left). Multi-Head Attention (right).
Image from (Vaswani *et al.*, 2017).

The **Figure 23** represents the architecture of the Transformer, which is overall an encoder-decoder structure. The encoder has N layers, and each of them has multi-head attention layer followed by layer normalization and feed-forward neural network (composed by point-wise, fully connected layers) followed by layer normalization. Similar to the encoder, the decoder has also N identical layers. Each layer of the decoder inserts a multi-head attention mechanism taking the output of the encoder as input. During the training, the parameters are updated over iterations. The parameters consist of the linear projects in the multi-head attention, the parameters of the feed-forward networks and layer normalization.

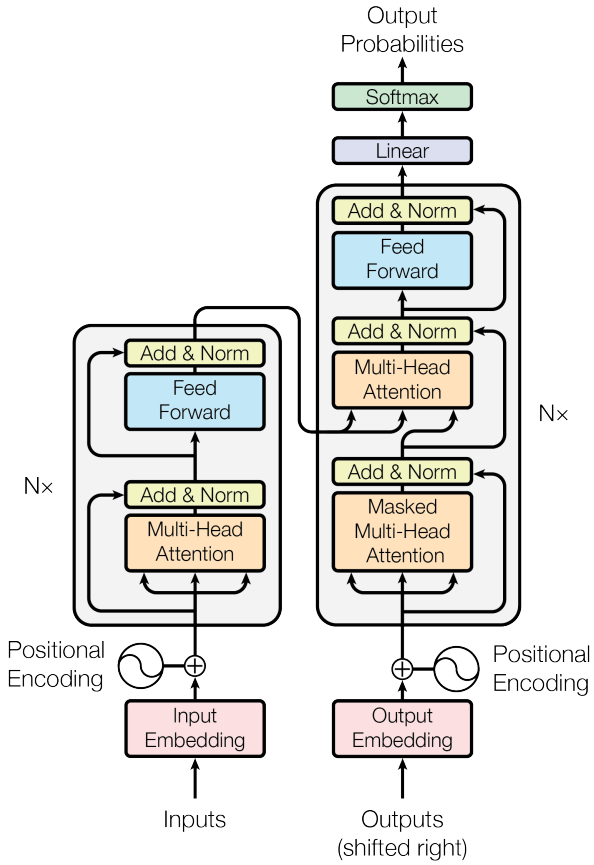


Figure 23. The architecture of Transformer.
Image from (Vaswani *et al.*, 2017).

Transformer models has been adapted for video analysis tasks. The videos are embedded into a sequence of token as the input of models. And the self-attention mechanisms allow to capture dependencies in the temporal dimension (Arnab *et al.*, 2021; Liu, Ning, *et al.*, 2021).

I.7. Deep learning-assisted Microscopy

Several deep learning algorithms have been employed to microscopy techniques:

I.7.1. Deep learning for SMLM localization

Deep-STORM

Deep-STORM is a model developed to reconstruct super-resolved images from low-resolution diffraction-limited images.

Deep-STROM is a fully convolutional encoder-decoder network. The encoder is composed of three 3×3 convolutional layers followed with 2×2 max-pooling layer. The decoder is composed of three deconvolution layers, each layer has 2 × 2 upsampling followed with 3×3 convolutional layer (**Figure 24**) (Nehme *et al.*, 2018).

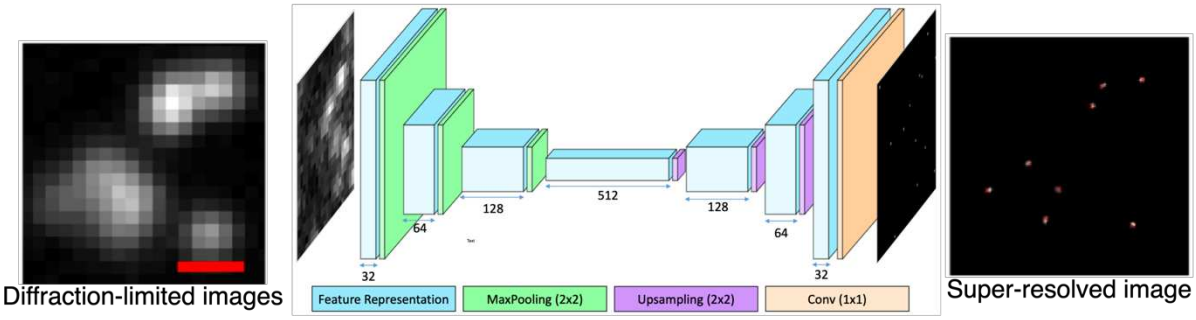


Figure 24. Deep-STORM convolutional architecture.
Deep-STORM is trained to generate super-resolved images from diffraction-limited images of blinking emitters during the SMLM acquisition. Image adapted from (Nehme *et al.*, 2018).

Deep-STORM is faster than traditional localization methods, even when emitters are overlapped, the method exhibits good performance. Deep-STORM is robustness to changes in emitter density and SNR of the input data. However, it does not output a list of localization coordinates like traditional localization methods, this may limit its applications.

DeepLoco

DeepLoco is trained to compute the 2D and 3D localizations from single frame of low-resolution image.

The DeepLoco model is composed of three parts: 1) A fully convolutional network that alternates between convolutions and spatial down-sampling three times. 2) A two-layer fully-connected ResNet that uses skip connections to bypass one or more layers during the forward and backward passes (He *et al.*, 2016). 3) Two linear layers that output K sources, where K is larger than the number of real sources. One linear layer outputs the weights, while another linear layer outputs the predicted spatial locations (**Figure 25**) (Boyd *et al.*, 2018).

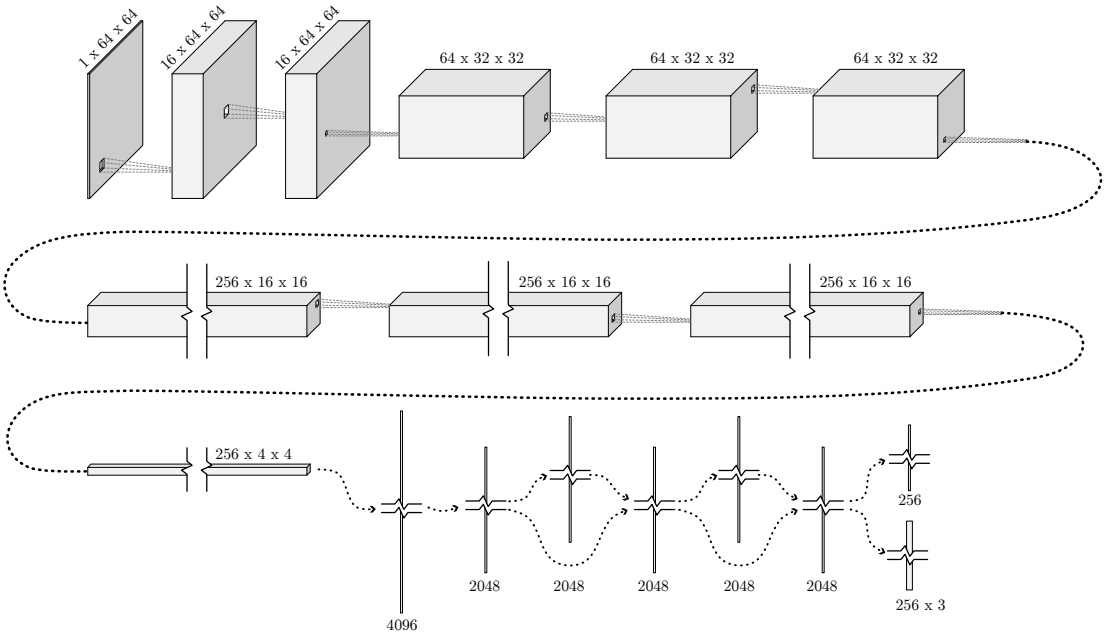


Figure 25. DeepLoco architecture.
The first two convolutions use 5 x 5 kernels and the other convolutions use 3 x 3 kernels. Image from(Boyd *et al.*, 2018).

Authors compared DeepLoco to the baseline methods which are the winners of 2D and 3D localization challenges. DeepLoco is orders of magnitude faster than existing approaches, while achieving comparable accuracy. However, compared to Deep-storm, the DeepLoco architecture is more complex, thus the computational cost is higher.

DECODE

DECODE is trained to localize single emitters at high density in three dimensions with high accuracy. The output represents for each pixel the probability of detecting an emitter and the emitter's spatial coordinates, brightness, the uncertainty of those predictions, and an optional background map (Speiser *et al.*, 2021). DECODE consists of two stacked U-nets, three

consecutive SMLM frames fed into first frame analysis U-net, and the outputs are concatenated and fed into the second temporal context U-net (**Figure 26**).

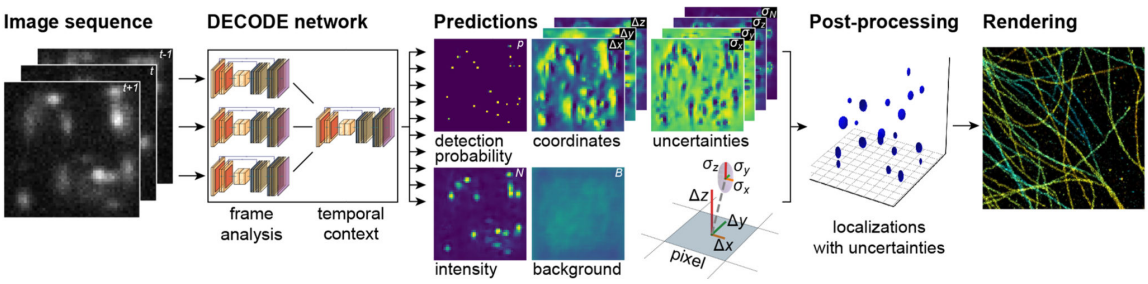


Figure 26. DECODE architecture.
The DECODE (left) is trained to localize single emitters from multiple frames of SMLM acquisition (right). Image from (Speiser *et al.*, 2021).

DECODE performs simultaneous detection and localization of emitters, outperforming other algorithms in a publicly available benchmark. It enhances both localization and detection accuracy and outputs emitter coordinates and their uncertainties. The uncertainty estimation is unique feature compared to DeepLoco and Deep-storm for super-resolution image reconstruction and for SMLM data analysis. DECODE's performance on images of emitters of dense data can substantially boost localization density or reduce imaging times.

I.7.2. Deep learning for microscopy image reconstruction enhancement

U-net based microscopy image enhancement

CARE (Content-Aware Image Restoration)

CARE (Weigert *et al.*, 2018) is a residual version of U-net architecture, aiming to restore image in fluorescence microscopy. CARE offers enhancements of image restoration for different microscopy imaging applications. Additionally, CARE computes the uncertainty estimation that provides insights into the reliability of the restored images. However, CARE cannot be applied to all image restoration problems. For instance, the intensity-based quantifications using care may might produce inaccurate or misleading results due to the nonlinear nature of the networks (Weigert *et al.*, 2018).

U-net-based method for SIM

Traditional SIM (page 13) acquires either 9 or 15 images to generate super-resolution. Another U-net-based deep learning method was developed by (Jin *et al.*, 2020) to reconstruct SIM images with fewer images. The u-Nets achieved results comparable to conventional SIM with fewer images. Their methods were demonstrated on live-cell super-resolution imaging with reduced photobleaching.

ANNA-PALM

ANNA-PALM stands for Artificial Neural Network Accelerated PhotoActivated Localization Microscopy. It aims to reconstruct super-resolved SMLM images from faster acquisition than conventional SMLM reconstruction. ANNA-PALM is inspired by Pix2Pix (as discussed on page 34). As illustrated in Figure 27, the training data consists of input: sparse SMLM images (generated from fewer low-resolution frames), and ground truth: dense SMLM images (high quality SMLM images with long acquisition times). The widefield diffraction-limited images, usually taken at the beginning of the SMLM image acquisition, can be served as input. The wide-filed image provides complementary information, allows better reconstruction quality. This has enabled, for example, to reduce the acquisition time of microtubules by a factor of up to 100, allowing to generate high-quality super-resolved images in few seconds of acquisition time only (Ouyang *et al.*, 2018).

However, the original ANNA-PALM method and its demonstration faced several limitations. First, ANNA-PALM had only been tested on images of limited diversity from a single laboratory. Second, the method is prone to exhibiting artifacts when applied to images obtained using different experimental conditions or protocols than the training data. Third, ANNA-PALM had only been demonstrated on fixed-cells and initial results on live cell images were unsatisfying. Because original ANNA-PALM is a 2D U-net-based cGAN, it cannot use the temporal information of the SMLM acquisition.

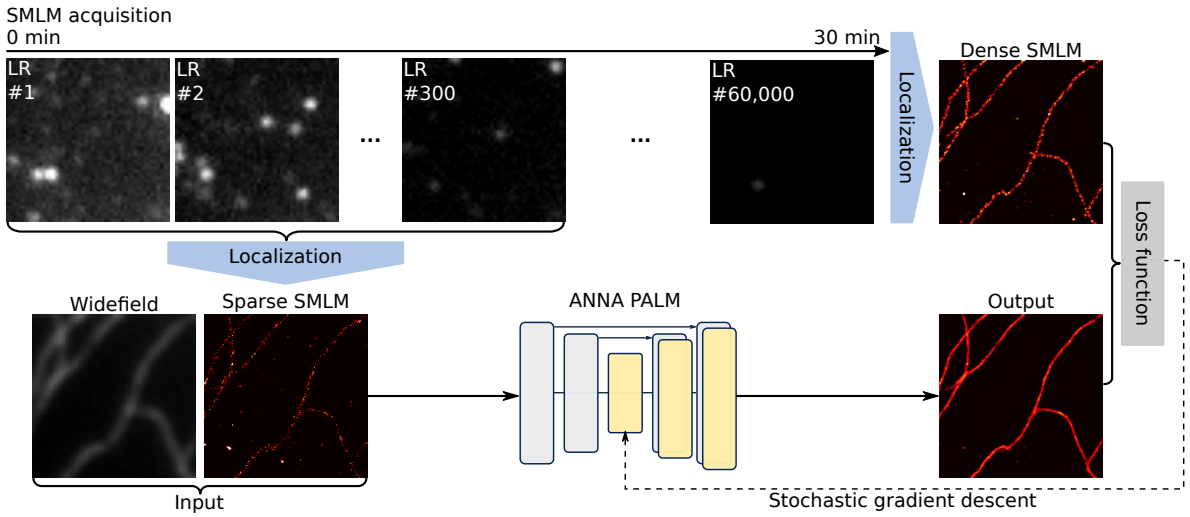


Figure 27. Simplified schematic of the ANNA-PALM training strategy.
A standard SMLM image is generated from a large number of low resolution image frames (LR, frames 1 to 60,000). This SMLM image defines the ground truth for training ANNA-PALM. Localizations computed from a smaller number of low resolution frames (here, 300), along with the corresponding widefield image (if available), are fed as input to ANNA-PALM. The loss function measures the consistency of the output image with the ground truth and the consistency of the generated output image with the widefield image (if available). Images adapted from (Ouyang *et al.*, 2018).

Deep CNN

(Kumar Gaire *et al.*, 2020) developed a deep CNN with ResNet framework, allows reconstructing high-density multicolor super-resolution images from low-density, contaminated multicolor spectroscopic SMLM (sSMLM) images with much fewer frames. sSMLM extracts the spatial locations and the corresponding spectral information of single-molecule blinking events (Zhang *et al.*, 2015). Using the deep CNN, the study was able to reconstruct two-color and three-color imaging

DBlink

DBlink is a model to reconstruct super-resolved video from low-resolution frames of SMLM data acquisition (Saguy *et al.*, 2023). As shown in **Figure 28**, DBlink is a hybrid deep learning architecture that combines CNNs and bi-directional LSTM networks. The bi-directional LSTM networks consist of two passes: the forward pass (green arrows) and the backward pass (red arrows), in which the information passes through frame by frame. The output frames of those passes are fed into a CNN as two input channels, and the CNN outputs the super-resolved video.

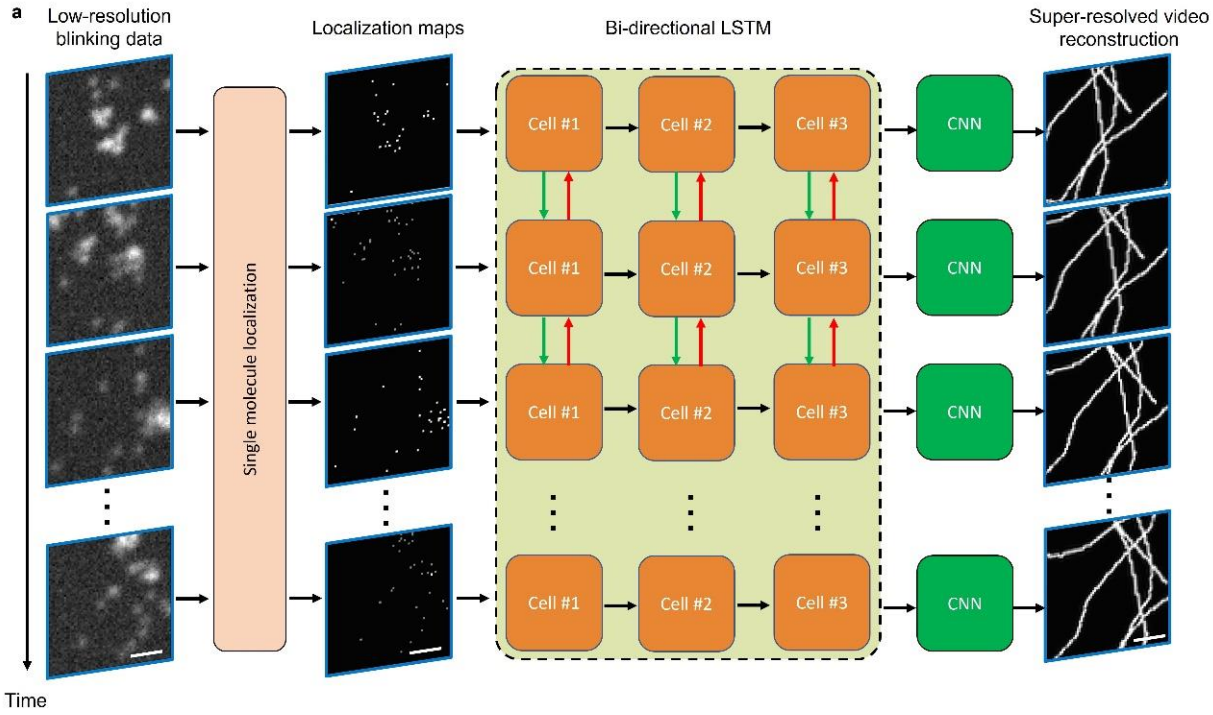


Figure 28. Architecture of DBlink.
The super-resolved localization maps are computed by Deep-STORM from low-resolution frames containing stochastic blinking events. The localization maps serve as input to a CNN-LSTM network. The output is super-resolution video reconstruction of the structure. Scale bar, 2.5 µm. Image from (Saguy *et al.*, 2023).

DBlink enables the reconstruction of super-resolved movies from the localization maps of single molecules in an SMLM experiment, thereby allowing the recovery of nanoscale structures in live cells. As highlighted on page 36, the strength of LSTM is its capability to process sequential data. Nevertheless, capturing spatial features within a sequence is more challenging for LSTMs compared to CNNs, as discussed on page 31. To address this limitation, the DBlink model integrates CNNs. However, while CNNs extracts spatial information, they might compromise or dilute the temporal information processed by the LSTMs. An additional consideration is that since DBlink's training data generated from simulations, the quality of the model's reconstructions is dependent on how closely the simulated training data aligns with real-world experimental data.

I.8. Objectives of my thesis

As introduced on page 43, ANNA-PALM allows to reconstruct high-quality super-resolved images from much fewer low-resolution frames. However, the original ANNA-PALM method faced several limitations. First, ANNA-PALM had only been tested on 7 images from our laboratory. Second, the method exhibits artifacts when applied to images obtained using different experimental conditions or protocols than the training data. Third, ANNA-PALM had only been demonstrated on fixed cells.

The objectives of my Ph.D. thesis are to address these limitations by: 1) improving the robustness of ANNA-PALM reconstructions when applied to data obtained from distinct laboratories, and 2) extending ANNA-PALM to reconstruct super-resolved time-lapse image sequences for dynamic biological structures in live cells. The results of my research will be detailed in the next section.

II. Results

II.1. ShareLoc platform

II.1.1. Background

As previously mentioned on page 16, the super-resolved SMLM image is generated from molecule coordinates using many thousands of raw low-resolution images. Consequently, the size of SMLM data often amounts to several gigabytes. This makes the SMLM data sharing challenging, either for raw data or molecule coordinates. While the SMLM technique is widely used for biological research (Lelek *et al.*, 2021), SMLM datasets are not usually publicly accessible. Numerous repositories, such as Figshare (http://figshare.com/http://figshare.com), Zenodo (http://zenodo.org), or IDR (Williams *et al.*, 2017), have been developed for sharing biological data. However, they are not specifically for gathering, visualizing and exploiting SMLM data. To address this, Wei Ouyang developed ShareLoc, an online platform (shareloc.xyz) dedicated to the collection and reuse of SMLM datasets acquired by the microscopy community.

II.1.2. Summary of methods and results

The FAIR principle stands for Findable, Accessible, Interoperable, and Reusable (Wilkinson *et al.*, 2016), emphasizes the efficient and effective reuse of the data. Data sharing based on this principle can enhance collaborative endeavors both within the community and across disciplinary boundaries.

ShareLoc platform ensures the FAIR principle in several ways:

Zenodo repository

ShareLoc is backed by Zenodo a data repository developed under the European OpenAIRE program and operated by CERN (the European Organization for Nuclear Research). With a Zenodo login, users can upload and store SMLM data (localizations and/or raw images) via the ShareLoc platform. When users upload data to the platform, a DOI (Digital Object Identifier) is automatically generated.

SMLM data uploading

During the uploading, the platform the platform features an auto-completion tagging system and allows for metadata assignment to SMLM data. Users also have the option to cite

relevant papers, further promoting transparency and context for data analysis. A ShareLoc administrator will review and approval the uploaded data, which will be displayed on the ShareLoc website. Once approved, the data becomes publically available for download, visualization, export, and reanalysis.

*.smlm format

Depending on the algorithms used for molecule localization, the format for SMLM localization data varies. To standardize the format and reduce the file size for faster transmission and loading times, Wei Ouyang introduced a lossless compressed file format with the *.smlm extension (**Supplementary Note 1** on page 61). This format can store arbitrary metadata and can is adaptable for 3D or multicolor SMLM data. Users have the option to convert their data to the *.smlm format during both upload and download.

Data annotation

Data annotation establishes a link between the image content and biological information, such as the target protein, cell line, and experimental conditions (e.g., drug treatment). This could include technical parameters like the fixation protocol, resolution, and laser excitation, ensuring the reproducibility of SMLM data. Therefore, annotating the data is important for data sharing. To facilitate rapid data annotation, ShareLoc platform leverages an auto completion tagging system (**Supplementary Note 2** on page 64).

EMBL-EBI ontology lookup service

ShareLoc's tagging system is built upon the EMBL-EBI ontology lookup service (https://www.ebi.ac.uk/ols/index) (Jupp *et al.*, no date). The ontology for the SMLM data annotation represents vocabularies within the SMLM domain, and the hierarchy relationships between them. For instance, as illustrated in **Figure 29**, the nucleoporin Nup133 is part of the nuclear pore. Additionally, as depicted in **Figure 30**, the nuclear pore is part of nucleus which is a cellular component.

Gene Product	Symbol	Qualifier	GO Term
UniProtKB:A0A7K7RBK5	Nup133	enables	GO:0017056 (F) structural constituent of nuclear pore
UniProtKB:A0A7K7RBK5	Nup133	part_of	GO:0005643 (C) nuclear pore

Figure 29. Illustration of ontology of nucleoporin "Nup133"
This figure illustrates the ontology of nucleoporin "Nup133". On the second column, we can see the Nup133 is part of nuclear pore. Image from (*QuickGO::Annotation List*, no date).

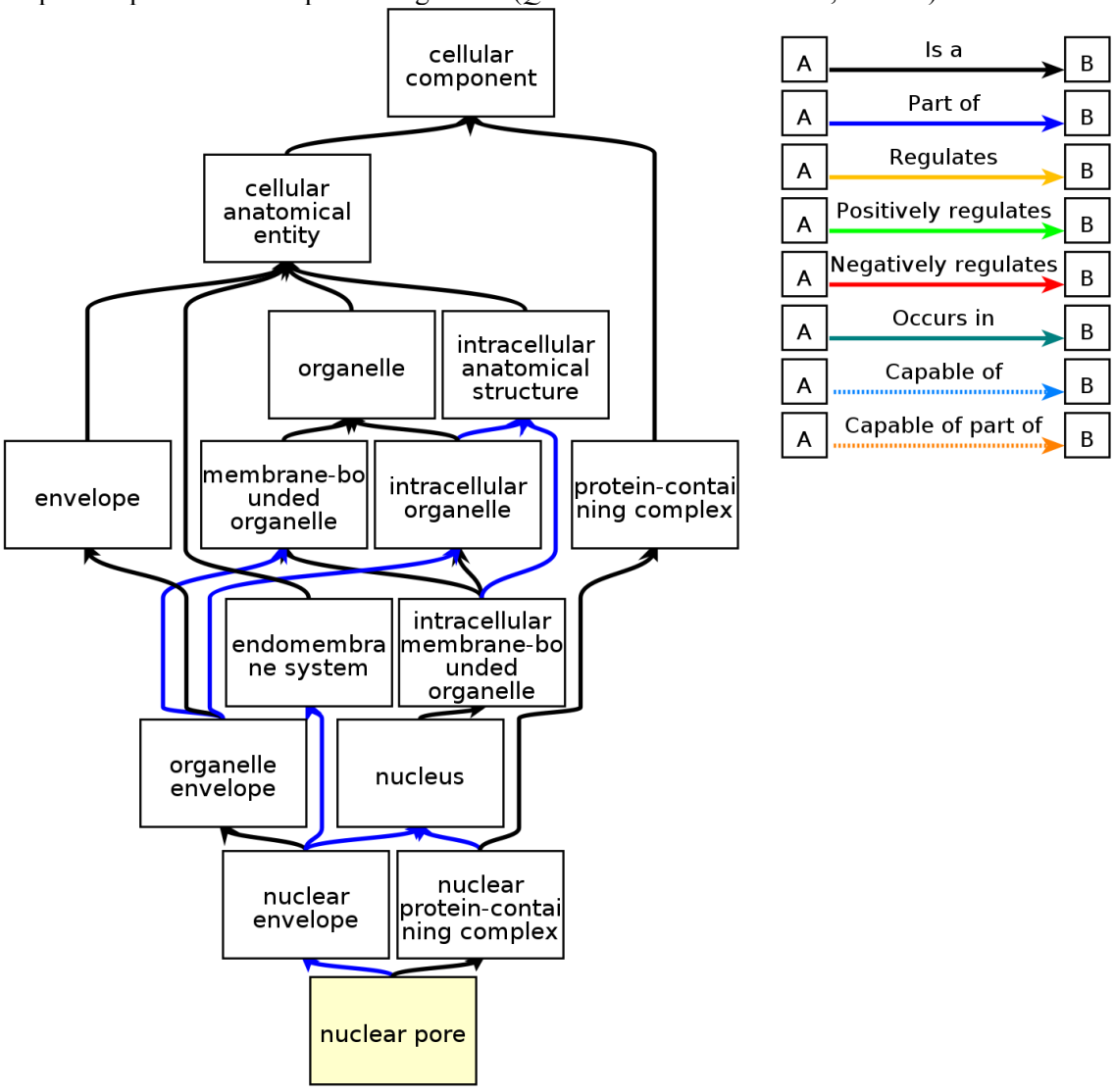


Figure 30. Chart of hierarchy relationship for "nuclear pore"
This figure illustrates the hierarchy of "nuclear pore" with other vocabularies. The bottom box is nuclear pore, the flashes represent different relationships between "nuclear pore" and other structures. Image from (*QuickGO::Term GO:0005643*, no date).

In addition to existing ontologies related to biological structures (such as cell lines U2OS and U373, biological structures including microtubules, mitochondria and nuclear-pore) and the SMLM experimental conditions (like fixation solutions such as glyoxal and methanol), I developed a specialized ShareLoc ontology. The ShareLoc ontology serves as a supplementary resource, as illustrated in **Figure 31**. This ontology is automatically generated and can be updated through continuous integration (CI) on GitHub. The CI process permits the frequent integration of code that generates the ontology, thus allowing ShareLoc users to dynamically add tags as needed (**Supplementary Note 2** on page 64).

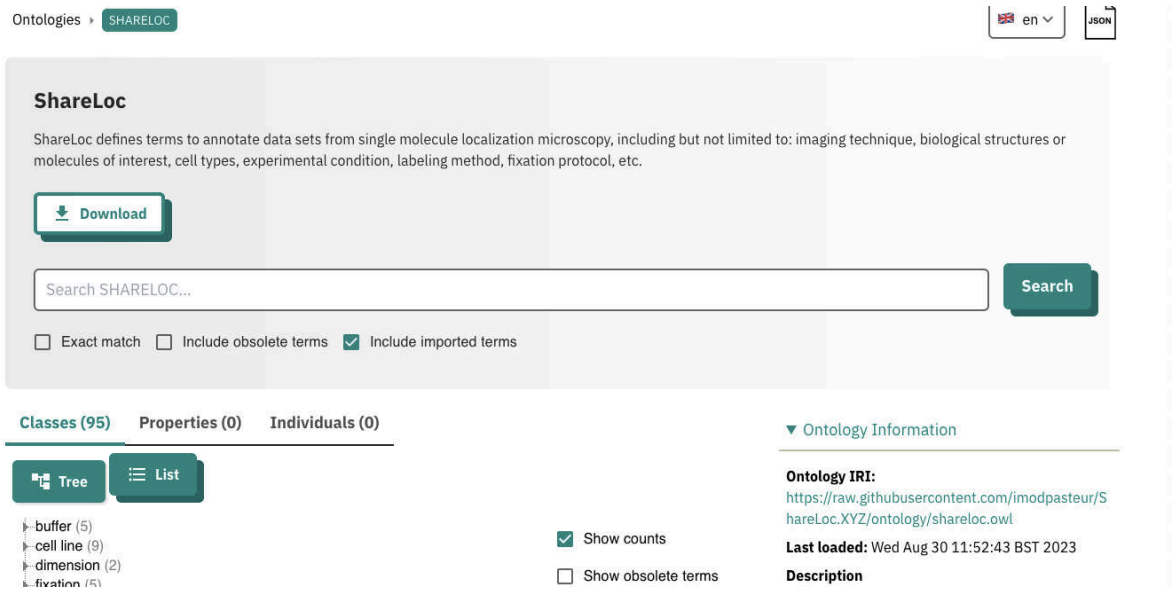


Figure 31. Screenshot of ShareLoc ontology on EMBL-EBI ontology lookup service website. Image taken from: (*ShareLoc*, no date).

SMLM image visualization

For quick and fluid visualization of SMLM localizations, ShareLoc features a browser-based viewer based on 'potree'. Potree is a method Potree is a technique capable of managing billions of localizations in a 3D point-cloud representation, offering features like zooming, panning, and rotation (Schutz, Krosl and Wimmer, 2019). This viewer requires no installations or updates and is compatible with all operating systems, including mobile devices (**Figure 1c** on page 58).

ShareLoc allows users to share SMLM data through a link, via e-mail or Twitter. The platform currently contains a large collection of SMLM data, comprising more than 1.5 billion localizations from more than 11 million raw images across 270 fields of view. The collection

includes images of microtubules, actin filaments, nuclear pores, clathrin-coated pits and mitochondria (**Supplementary Fig. 2** on page 67). We believe that ShareLoc can helps to accelerate the development of new analytical methods for SMLM and the reduce the redundant SMLM data acquired by the community for new biological information, ensuring reproducibility in SMLM imaging research and broadening its applications in life sciences.

The published paper detailing these results will be provided below and perspectives of ShareLoc platform will be discussed on page 137.

II.1.3. Contribution

I contributed by testing and validating the functionalities of the ShareLoc platform, including uploading and downloading the SMLM localization data, drafting associated metadata, developing ShareLoc ontology and reporting bugs. Further, I wrote relevant documentation and made tutorial videos (available at: https://www.youtube.com/watch?v=tFaEXWmv01g&list=PLYwa6cS54RkkyS17FoaZcrX-awRkqp6q5).

Article 1 (published): Ouyang, W.*, **Bai, J.***, Singh, M.K., Leterrier, C., Barthelemy, P., Barnett, S.F.H., Klein, T., Sauer, M., Kanchanawong, P., Bourg, N., Cohen, M.M., Lelandais, B., Zimmer, C., 2022. ShareLoc — an open platform for sharing localization microscopy data. Nat Methods 19, 1331–1333. https://doi.org/10.1038/s41592-022-01659-0 (* equal contribution)

https://doi.org/10.1038/s41592-022-01659-0

ShareLoc – an open platform for sharing localization microscopy data

Single-molecule localization microscopy (SMLM) has matured into one of the most widely used super-resolution imaging methods and has been used to address a broad spectrum of biological research questions[1]. This success has inspired the community to develop numerous computational techniques to extract localizations from raw images or turn them into biologically meaningful quantities[1–3]. The development of further analytical methods could greatly benefit from easy access to SMLM data generated worldwide. This is especially true for machine learning approaches and notably deep learning, whose performance hinges strongly on the amount of training data. However, despite the vast number of SMLM studies[1], the overwhelming majority of SMLM data remains inaccessible to the community. Repositories such as Figshare or Zenodo, or the added-value repository IDR[4,5], are generic in purpose and not optimally suited for gathering, visualizing and exploiting SMLM data in a manner consistent with the FAIR principles (findability, accessibility, interoperability and reusability)[6]. An important impediment for sharing SMLM data is that each super-resolution image is built by computing molecular coordinates from many thousands of raw, low-resolution images (Fig. 1a), which often total many gigabytes in size. Even localization files, albeit much smaller than the raw data, are usually too large to be sent by e-mail. Furthermore, they come in various formats, complicating reanalysis, while metadata are often unstructured or simply absent. Another restriction of generic data repositories is the lack of specific tools for visualizing SMLM data. Owing to these limitations, only a minuscule fraction of the globally acquired SMLM data is easily and publicly available, preventing their reanalysis and slowing the development of more powerful analysis methods. Here, we present ShareLoc (https://shareloc.xyz), an open platform designed to facilitate the sharing, visualization, annotation, and community-based reutilization of SMLM data (see online quick tutorial movie[7]).

ShareLoc consists of (i) a storage service backed by Zenodo, a widely used open-access repository operated by CERN, and (ii) an extendable system of web plugins built upon ImJoy[8], a state-of-the-art platform for developing and deploying interactive data science tools. With a Zenodo login, users can easily upload and store SMLM data (localizations and/or raw images up to 50 gigabytes) through the ShareLoc platform, automatically generating a digital object identifier (DOI) (Fig. 1b). Upon review and approval by a ShareLoc administrator, the new data will be shown on the ShareLoc website and available for download, visualization, export, and reanalysis[7]. Each dataset can be linked to other datasets or analysis tools, and by default SMLM data are linked to a dedicated viewer plugin (see below).

To facilitate rapid transmission of SMLM data, we developed a losslessly compressed binary file format (extension *.smlm) that substantially reduces localization file sizes and loading times, is portable, compact and flexible (for example, it can store arbitrary metadata), and is easily adjusted for 3D or multicolor data (Supplementary Note 1 and Supplementary Fig. 1a, b). Although ShareLoc uses the *.smlm format internally, the platform also supports the import of SMLM data in various standard formats and export as CSV files[7], and has features for batch downloading and conversion of multiple datasets[9] (see also Supplementary Software 1). Furthermore, we implemented a rapid WebGL-powered viewer plugin in ImJoy[8], based on 'potree', a method designed for quick and fluid visualization of very large 3D point clouds that can handle billions of localizations and includes zooming, panning and rotation features[7,10]. This viewer is entirely browser-based, requires no installations or updates and operates across all operating systems, including mobile devices[7] (Fig. 1c). Moreover, it allows quasi-instantaneous visualization without requiring download of the entire dataset, akin to Google Maps, thereby facilitating access for users with low Internet bandwidth[7].

Data on ShareLoc can be easily shared as a simple link, for example, by e-mail or Twitter[7] (Fig. 1c). ShareLoc already contains a public repository of SMLM data totaling more than 1.5 billion localizations from more than 11 million raw images in 270 fields of view, including images of microtubules, actin filaments, nuclear pores, clathrin-coated pits and mitochondria, which are easily browsed from a 'gallery' view (Fig. 1c, Supplementary Fig. 2). We implemented a simple and flexible tagging system that allows users to annotate images using more than 7 million terms from pre-existing biomedical ontologies accessible through the EMBL-EBI ontology lookup service, but also to create new tags, where needed, in a dedicated new 'SHARELOC' ontology. A search bar allows users to find data using specific tags such as 'microtubules', 'dna-paint' or '3d'[7] (Supplementary Note 2).

The ShareLoc platform helps to address all four principles of FAIRness, which are currently lacking for SMLM[6]. Findability is enabled by ShareLoc's rich and flexible annotation and searching features. Accessibility is facilitated by the compact *.smlm format and the simple and fluid, low-bandwidth-compatible viewer. Interoperability is guaranteed by the open-source format and tools to import, export and convert SMLM data in widely used file formats. Reusability is ensured by the unified file format, structured metadata and links with ontologies. In another paper (Bai et al., unpublished data), we will illustrate the benefit of data sharing on ShareLoc by showing how reanalysis of data from multiple labs can improve the robustness of super-resolution image reconstruction by deep learning. More generally, we believe that ShareLoc is poised to accelerate the development of new analytical methods for SMLM and the mining of SMLM data acquired by the community for new biological information, and will help to promote reproducible research in single-molecule imaging and its numerous applications in the life sciences.

Data availability

All data used in this study are available on the ShareLoc platform (https://shareloc.xyz/#/?type=dataset&tags=) and on Zenodo (https://zenodo.org/communities/

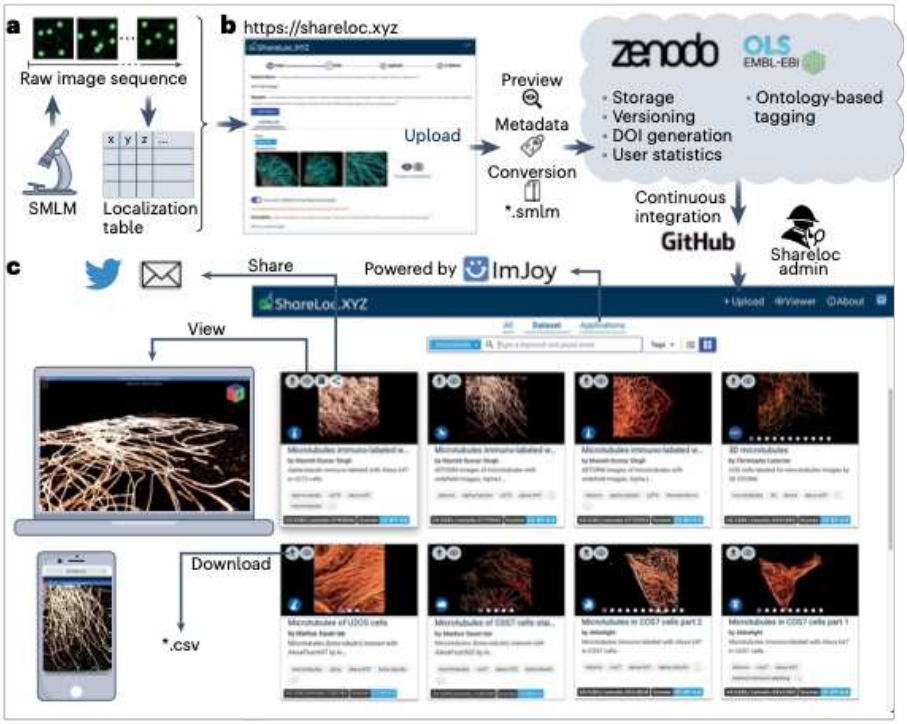


Fig. 1 | ShareLoc facilitates sharing, viewing, tagging and searching of SMLM data. a, In SMLM, a sequence of raw, low-resolution (diffraction-limited) images of single, photoswitching, fluorescent molecules is acquired and the molecular localizations are extracted computationally[1], typically resulting in a localization table. **b**, The online ShareLoc platform (https://shareloc.xyz) allows cloud storage of SMLM data on Zenodo with digital object identifiers (DOI) and enables easy annotation of data with tags searchable in the EMBL-EBI ontology lookup service (OLS), as well as conversion to the compact *.smlm format. Continuous integration on GitHub automatically notifies ShareLoc administrators of uploaded data and upon review (and, if needed, curation), the data are made available on the ShareLoc platform. **c**, ShareLoc provides gallery views of uploaded SMLM data (searchable by tags) and enables quick data sharing by e-mail or social media, interactive download as CSV files and quasi-instantaneous interactive 3D visualization through a dedicated ImJoy plugin that runs in a web browser, on computers or mobile devices, even in low-bandwidth settings. Datasets can be connected to other applications powered by ImJoy, allowing additional visualization or analysis capabilities.

shareloc/?page=1&size=20), and each dataset has a unique DOI.

Code availability

The source code underlying the ShareLoc platform is available on Zenodo at https://doi.org/10.5281/zenodo.5045318 and on GitHub at https://github.com/imodpasteur/ShareLoc.XYZ/. Specifications and implementations of the *.smlm file format are available on Zenodo at https://zenodo.org/record/6576521 and on GitHub at https://github.com/imodpasteur/smlm-file-format/releases/tag/v0.2.0. Tools for batch download and to parse *.smlm are available on GitHub at https://github.com/imodpasteur/shareloc-utils#shareloc-utilities. Scripts used in Supplementary Fig. 1 to compare file sizes and downloading times are available at https://github.com/imodpasteur/smlm-file-format/tree/baiddd-patch-1/data/file_format_benchmark.

Wei Ouyang[1,2,12] ✉, **Jiachuan Bai**[1,3,12], **Manish Kumar Singh**[1,4], **Christophe Leterrier**[5], **Paul Barthelemy**[6], **Samuel F. H. Barnett**[7,8,11], **Teresa Klein**[9], **Markus Sauer**[9], **Pakorn Kanchanawong**[7,8], **Nicolas Bourg**[6], **Mickael M. Cohen**[4], **Benoît Lelandais**[1,10] and **Christophe Zimmer**[1] ✉

[1]Imaging and Modeling Unit, Institut Pasteur, Université Paris Cité, Paris, France. [2]Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH – Royal Institute of Technology, Stockholm, Sweden. [3]Sorbonne Université, Collège Doctoral, Paris, France. [4]Laboratoire de Biologie Moléculaire et Cellulaire des Eucaryotes, Sorbonne Université, CNRS, UMR8226, Institut de Biologie Physico-Chimique, Paris, France. [5]Aix Marseille Université, CNRS, INP UMR7051, NeuroCyto, Marseille, France. [6]Abbelight, Cachan, France. [7]Mechanobiology Institute, National University of Singapore, Singapore, Singapore. [8]Department of Biomedical Engineering, National University of Singapore, Singapore, Singapore. [9]Department of Biotechnology and Biophysics, Biocenter, University of Würzburg, Würzburg, Germany. [10]Bioinformatics and Biostatistics Hub, Institut Pasteur, Université Paris Cité, Paris, France. [11]Present address: Max Planck Institute for Medical Research, Heidelberg, Germany. [12]These authors contributed equally: Wei Ouyang, Jiachuan Bai.
✉e-mail: wei.ouyang@scilifelab.se; czimmer@pasteur.fr

Published online: 21 October 2022

References

1. Lelek, M. et al. *Nat. Rev. Methods Primers* **1**, 1–27 (2021).
2. Nicovich, P. R., Owen, D. M. & Gaus, K. *Nat. Protoc.* **12**, 453–460 (2017).
3. Zhang, Z., Nishimura, Y. & Kanchanawong, P. *Mol. Biol. Cell* **28**, 333–345 (2017).
4. Williams, E. et al. *Nat. Methods* **14**, 775–781 (2017).
5. Sarkans, U. et al. *Nucleic Acids Res.* **46**, D1266–D1270 (2018). D1.
6. Wilkinson, M. D. et al. *Sci. Data* **3**, 160018 (2016).
7. Bai, J. *Quick Tutorial of ShareLoc* https://www.youtube.com/watch?v=tTOkqzr74wg
8. Ouyang, W., Mueller, F., Hjelmare, M., Lundberg, E. & Zimmer, C. *Nat. Methods* **16**, 1199–1200 (2019).
9. Ouyang, W. *ShareLoc Utilities* https://github.com/imodpasteur/shareloc-utils#shareloc-utilities (2022).
10. Schütz, M., Krösl, K. & Wimmer, M. Real-time continuous level of detail rendering of point clouds. In *2019 IEEE Conf. Virtual Reality and 3D User Interfaces (VR)* 103–110 (2019).

Acknowledgements

We thank H. Harmse (EMBL-EBI) for help in linking ShareLoc to the EBI ontology lookup service; M. Schuetz and F. Woitzel for development and sharing of the potree and Fairy Dust viewers, respectively; M. Lelek for advice on metadata; J. Swedlow, J. Moore and S. Besson for stimulating several improvements of the platform. C.L. acknowledges funding from CNRS ATIP AO2016. S.F.H.B. and P.K. acknowledge funding support from Mechanobiology Institute seed funding, Singapore Ministry of Education (MOE2019-T2-2-014) and National Research Foundation (QEP-P7). M.K.S., M.M.C. and C.Z. acknowledge funding by Agence Nationale de la Recherche (grant ANR 17 CE13 0026 02) and M.M.C. by Labex DYNAMO (ANR-11-LABX-0011-DYNAMO) and MITOFUSION (ANR-19-CE11-0018). J.B., B.L. and C.Z. acknowledge funding by Institut Pasteur, by the Région Ile-de-France in the framework of DIM ELICIT, and by the Inception program (Investissement d'Avenir grant ANR-16-CONV-0005).

Author contributions

W.O. designed and implemented the ShareLoc platform. J.B. tested the platform, curated data, and wrote documentation. M.K.S., C.L., P.B., S.F.H.B., T.K., M.S., P.K., N.B., M.M.C. and B.L. provided SMLM data. C.Z. coordinated the project and wrote the manuscript with input and edits from W.O., J.B. and B.L. All authors read and approved the manuscript.

Competing interests

N.B. is shareholder of Abbelight. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-022-01659-0.

Peer review information *Nature Methods* thanks Jason Swedlow and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Supplement for

ShareLoc – an open platform for sharing localization microscopy data

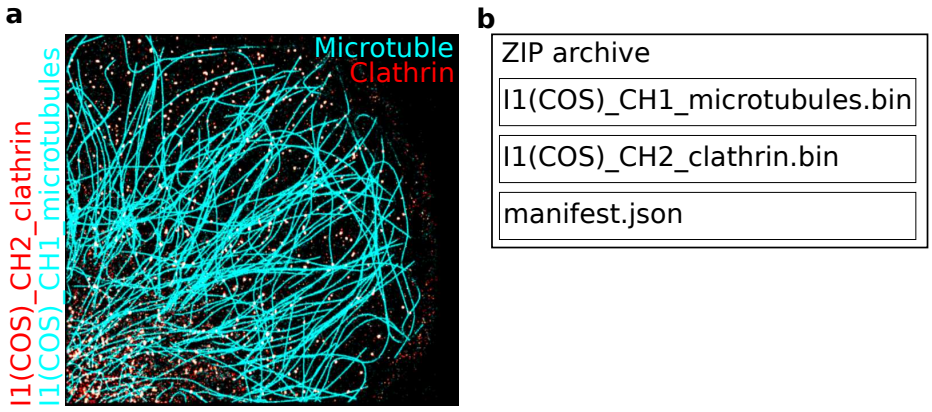
Wei Ouyang[#], Jiachuan Bai[#], et al.

Supplementary Note 1: The *.smlm data format

This Note details our motivation for developing the *.smlm data format and its design principles.

Storing and transmitting the data underlying typical SMLM experiments is challenging, because of the enormous file size. A typical experiment requires a sequence of ~10^4-10^5 individual raw (diffraction limited) images for each super-resolution image and data volume can reach ~100 GB per experiment or more[1]. With fast, kHz-rate cameras, a SMLM system can output Terabytes per hour. The localization data computed from the raw images are much smaller, but still typically occupy several GB, because by default, localizations are generally stored as comma separated value (CSV) files, which trade off file size to optimize human readability. As a result, typical SMLM data files are too large for simple sharing methods such as e-mail.

To address this, we designed a specific data format called 'SMLM' (*.smlm), which instead optimizes for size rather than human readability, and can reduce file size by 6-fold or more compared to the CSV format. For example, a 7.16 GB *.csv localization table can be replaced by a 1.34 GB *.smlm file without any loss of information. To achieve this size reduction, we adopted two strategies. First, we use binary encoding rather than text encoding, implying that all numbers are represented by data types such as integer, float or double. Second, we compress the binary file using generic lossless compression algorithms that come with the standard zip file format[2]. In order to compensate the lack of human readability of binary formats, we add a complementary text file (hereafter called 'manifest file') to describe the structure of the localization table and store metainformation. See an example layout of a *.smlm file below.



An example SMLM image (**a**) and the structure of its corresponding *.smlm file (**b**). **a**) Dual color 3D SMLM image of clathrin (red) and microtubules (cyan), from the Leterrier lab. **b**) Layout of the corresponding *.smlm file. The *.smlm file is a zip archive that contains two binary localization files (one for each color channel), and a manifest file with the fixed name 'Manifest.json'. Note that the *.smlm format can also be used to store raw images, for example widefield images corresponding to the SMLM image.

The manifest file is implemented in the widely supported JavaScript Object Notation (*.json) format and serves two purposes. First, it stores information that is needed for

a reader program to properly read the binary data, such as the data type (e.g. float or double), the number of rows, or the header. Second, it provides a human understandable summary of the localization table that does not require loading or unzipping the whole file.


```
{
    "format_version": "0.2",
    "formats": {
        "smlm-table(binary)-0": {
            "name": "smlm-table(binary)-0",
            "type": "table",
            "mode": "binary",
            "extension": ".bin",
            "columns": 6,
            "headers": ["frame","x","y","z","uncertainty_xy","uncertainty_z_nm_"],
            "dtype": ["float32","float32","float32","float32","float32","float32"],
            "shape": [1,1,1,1,1,1],
            "units": [],
            "description": "the default smlm format for localization tables"
        },
        "smlm-table(binary)-1": {
            "name": "smlm-table(binary)-1",
            "type": "table",
            "mode": "binary",
            "extension": ".bin",
            "columns": 6,
            "headers": ["frame","x","y","z","uncertainty_xy","uncertainty_z_nm_"],
            "dtype": ["float32","float32","float32","float32","float32","float32"],
            "shape": [1,1,1,1,1,1],
            "units": [],
            "description": "the default smlm format for localization tables"
        }
    },
    "files": [
        {
            "name": "I1(COS)_CH2_clathrin.bin",
            "format": "smlm-table(binary)-0",
            "channel": "default",
            "rows": 382733,
            "offset": {
                "x": 0,
                "y": 0
            },
            "metadata": {},
            "exposure": -1,
            "type": "table",
            "min": {
                "frame": 1,
                "x": 1560.800048828125,
                "y": 1531.4000244140625,
                "z": -404.3999938964844,
                "uncertainty_xy": 0.30000001192092896,
                "uncertainty_z_nm_": 0.6999999880790710
            },
            "max": {
                "frame": 39999,
                "x": 40180.1015625,
                "y": 39407.1015625,
                "z": 403.5,
                "uncertainty_xy": 14.5,
                "uncertainty_z_nm_": 29
            },
            "avg": {
                "frame": 19057.150465206876,
                "x": 15180.081666312117,
                "y": 18308.978733591597,
                "z": -164.58818115778257,
                "uncertainty_xy": 3.9080617543854737,
                "uncertainty_z_nm_": 7.816527450217723
            },
            "hash": "e360f81d1f950555ba3aeff2981392dd",
            "tags": []
        },
        {
            "name": "I1(COS)_CH1_microtuble.bin",
            "format": "smlm-table(binary)-1",
            "channel": "default",
            "rows": 769747,
            "offset": {…
            },
            "metadata": {},
            "exposure": -1,
            "type": "table",
            "min": {…
            },
            "max": {…
            },
            "avg": {…
            },
            "hash": "6f3a97652596e81dcc63ae6a333630bf",
            "tags": []
        }
    ],
    "sample": "",
    "labeling": "",
    "license": "CC BY 4.0",
    "citeAs": "",
    "hash": "90c9563de5d6272ee535bd6b4220872a"
}
```

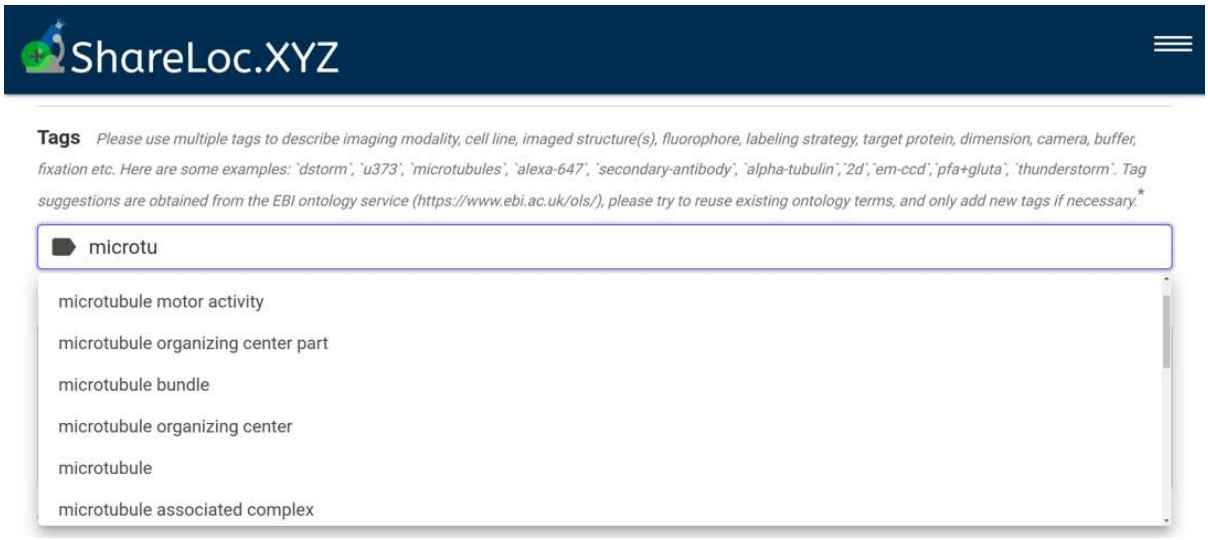

Example manifest file for the dual color SMLM image above (only excerpts are shown). The file contains several fields, including 'formats' and 'files'. The 'formats' field stores a set of formats describing the header (here: “frame”, “x”, “y”, etc.), the data type (in this case, float32 values), the shape, etc. of each column. The 'files' field specifies which format is used to read the corresponding localization file, and also includes meta information for each binary file, such as the number of rows (here: 382,733 and 769,747, corresponding to the number of localizations for clathrin and microtubules, respectively), the number of raw image frames (39,999 for clathrin), the minimum, maximum and average values of computed x,y,z coordinates, etc.

An important feature of the *.smlm format is its flexibility. Because the structure of the binary files is specified in the manifest file, it can be easily adapted to different types of applications. For example, in addition to the required columns specifying *x* and *y* coordinates, one can easily add a third column to specify the axial (*z*) coordinate in a 3D SMLM image, as in the example above[3]. This only requires changing the “headers” in the ‘formats’ field. It is similarly straightforward to add more columns, e.g. to specify additional parameters for each localization, such as the intensity, localization uncertainty, etc. It is even possible to add additional columns to store the raw pixel values of image patches centered on each localization. Another illustration of this flexibility are multicolor images. In multicolor SMLM, localizations from each color are typically stored in separate localization files. These can be stored in the same *.smlm file, by specifying their names in the ‘files’ field of the manifest file and indicating the

same format in the ‘format’ field, as also shown in the above example. Similarly, a *.smlm file can store both a localization table and the corresponding widefield image (in this case, the ‘formats’ field specifies both the table format and the image format). Another related example is tiled imaging, where images from many adjacent (or partly overlapping) fields of view are assembled into a much larger effective field of view. The many localization tables corresponding to each individual field of view can be saved together in a single *.smlm file, by specifying the names of the individual localization files in the ‘files’ field. In this case, each localization file can be given an ‘offset’ field to store the spatial offset (e.g. the coordinates of the upper left corner) of each field of view.

Supplementary Note 2: ShareLoc data tagging system and ontology

For optimal reuse, shared SMLM data should be annotated with metadata containing all information required to reproduce the data, including the experimental conditions, image acquisition and analysis protocols[4]. To facilitate such annotations, ShareLoc provides a simple and flexible, ontology-based tagging system. Users can choose tags as they wish, but in order to unify annotations and improve data searchability and reusability, ShareLoc encourages the use of terms from pre-existing ontologies. For example, terms such as “microtubules”, “nocodazole”, “formaldehyde” or “Alexa Fluor 647” are defined in pre-existing ontologies and we encourage users to use such pre-existing terms for data tagging wherever relevant. In practice, ShareLoc facilitates this using an auto-complete feature that automatically proposes ontology terms available through the EMBL-EBI ontology lookup service (EBI OLS; https://www.ebi.ac.uk/ols/index) as illustrated in the screenshot below.



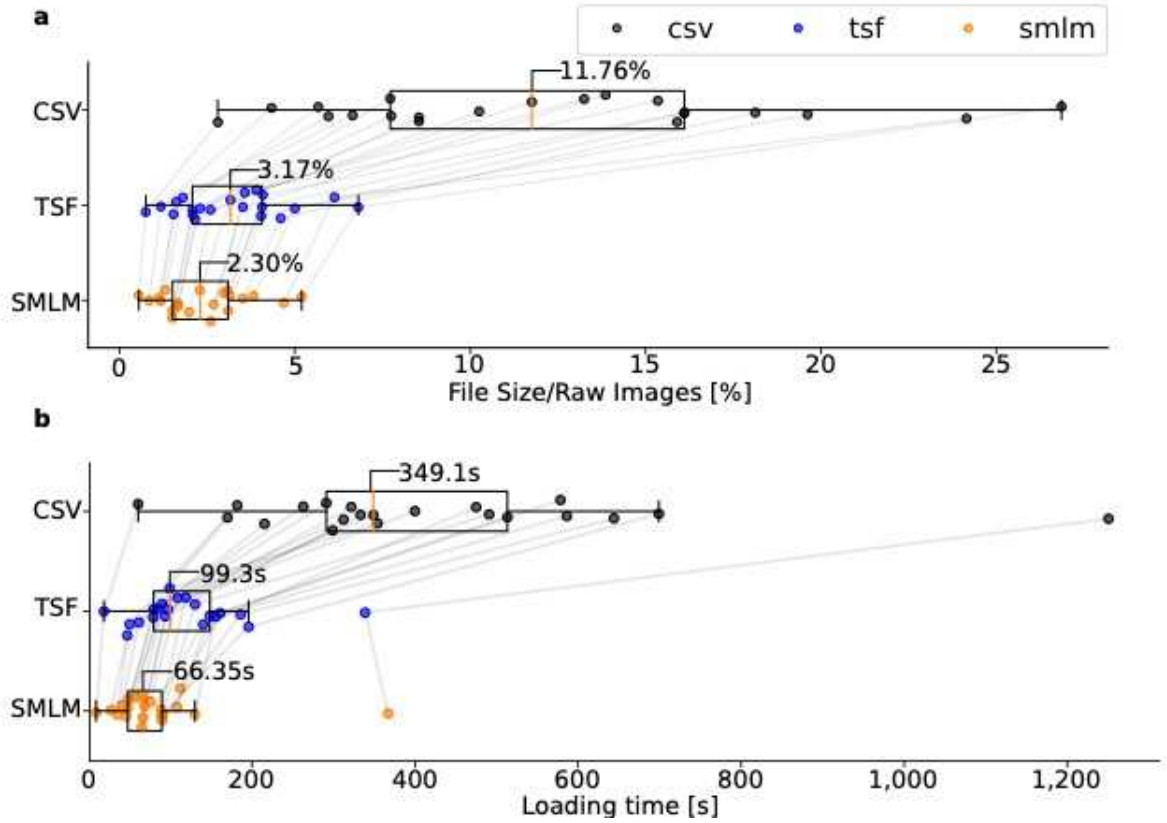
Typing “microtu” in the “Tags” box during data annotation yields a list of suggested ontology terms matching this character string.

To address the need for SMLM-relevant terms that were absent from pre-existing ontologies, e.g. “dSTORM” or “DNA-PAINT”, we also created a new dedicated ontology called SHARELOC, available at: https://www.ebi.ac.uk/ols/ontologies/shareloc. In this ontology, tags are grouped in a number of categories, such as imaging modality, cell line, fluorophore, etc (see Table below). Technically, the SHARELOC ontology is defined as a shareloc.tsv file available on GitHub at: https://github.com/imodpasteur/ShareLoc.XYZ/blob/main/src/shareloc.tsv. Users can add new terms to this ontology by creating a pull request on Github, which causes the shareloc.tsv file to be updated and automatically converted into a shareloc.owl file, available on Github at: https://github.com/imodpasteur/ShareLoc.XYZ/tree/ontology/shareloc.owl, thereby

allowing the SHARELOC ontology to be imported into OLS. This continuous integration is performed using ROBOT commands available on GitHub at: https://github.com/imodpasteur/ShareLoc.XYZ/tree/main/utils/convert-owl.sh.
On the ShareLoc home page, next to the search bar, tags can also be selected to filter the gallery view and search for specific data sets. For example, selecting the tags "microtubule", "alexa-647" and "nih3t3", will display only the SMLM data of microtubules imaged with Alexa-647 in NIH3T3 cells.

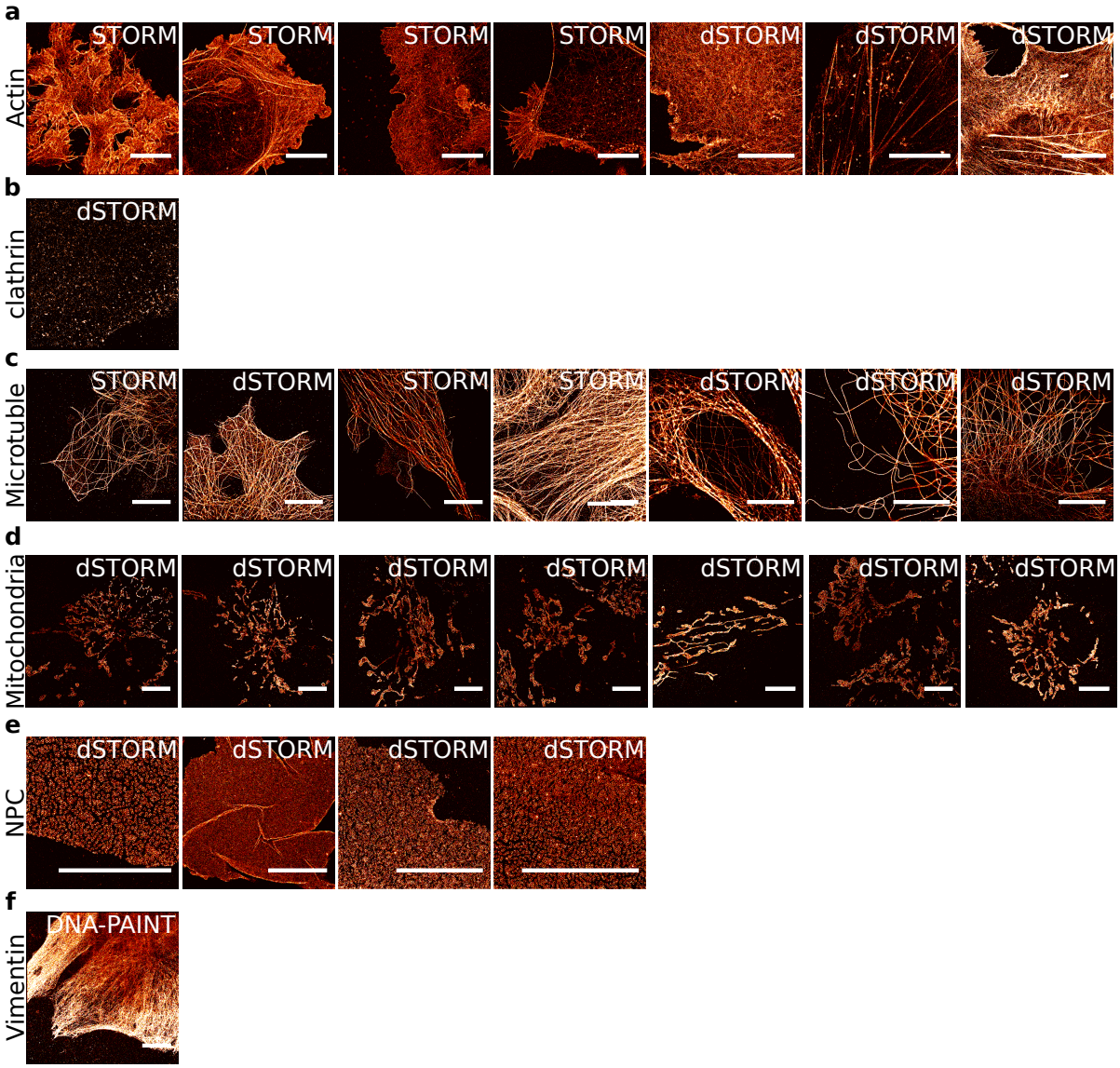
Category	**Tag**
Modality	palm, storm, dstorm, dna-paint, paint
Cell line	u2os, u373, cos7, ref52, msck, hela, nih3t3, yeast, e-coli
Fluorophore	alexa-647, alexa-555, alexa-532, cy5, cy3b, cf680, cf660, cf568, hmsir, atto-488, mos2, mapple, dendra2, pa-gfp, pa-mcherry, dronpa, nile-red, jf-646, jf-568, pa-jf568, pa-jf646
Labeling Strategy	transient-transfection, stable-knock-in, endogenous-labeling-crispr, direct-immuno-labeling, indirect-immuno-labeling, halo-tag, snap-tag, tetracysteine-tag, phalloidin
Structure	actin, microtubules, mitochondria, nuclear-pore, clathrin, vimentin, plasma membrane, intermediate filament, dna
Target molecule	nup133, nup96, tom22, alpha-tubulin, beta-tubulin, actin, vimentin, wga, escrt, ftsz, h2, h3, h1, polye
Dimension	2d, 3d
Camera	em-ccd, scmos
Reconstruction software	thunderstorm, smap, smoiphot, zola-3d, picasso, quickpalm
Buffer	glox, catalase, mea, abbelight-safe-reagent, idylle-everspark
Fixation	4% pfa, 1% pfa, pfa+gluta, glyoxal, methanol
Other	All other customized tags

Examples of tags available in the SHARELOC ontology and organized by categories.



Supplementary Figure 1: Localization data size and loading times for different file formats

a,b) File sizes (**a**) and loading times (**b**) of n=21 different localization data sets, for three file formats: CSV (*.csv), TSF (*.tsf) and SMLM (*.smlm). The file sizes are expressed as percentage of the size of the corresponding raw image sequences. Each dot corresponds to a distinct data set. Dots for the same data sets in different formats are connected by grey lines. Boxplots show the first and third quartiles as box edges, the median as orange bars (median value as indicated), and whiskers show the full data range (except outliers).



Supplementary Figure 2: ShareLoc contains diverse SMLM data sets

Examples of SMLM data available on the ShareLoc website. The data include images of actin filaments (**a**), clathrin coated pits (**b**), microtubules (**c**), mitochondria (**d**), nuclear pore complexes (NPC) (**e**), and vimentin (**f**). The contrast of some images has been adjusted for better visibility.

Supplementary References

1. Ouyang, W. & Zimmer, C. The imaging tsunami: Computational opportunities and challenges. *Curr. Opin. Syst. Biol.* **4**, 105–113 (2017).
2. Harnik, D., Khaitzin, E., Sotnikov, D. & Taharlev, S. A Fast Implementation of Deflate. in *2014 Data Compression Conference* 223–232 (2014). doi:10.1109/DCC.2014.66.
3. Aristov, A., Lelandais, B., Rensen, E. & Zimmer, C. ZOLA-3D allows flexible 3D localization microscopy over an adjustable axial range. *Nat. Commun.* **9**, 2409 (2018).
4. Lelek, M. *et al.* Single-molecule localization microscopy. *Nat. Rev. Methods Primer* **1**, 1–27 (2021).

II.2. Improving the robustness of ANNA-PALM

II.2.1. Background

As mentioned on page 29, a central question for deep learning methods is the robustness of model when input testing data diverges from the distribution on which it was trained. This is particularly relevant when considering variations of technical changes in SMLM data acquisition from different laboratories. For example, using different dyes with different emission wavelength might lead to different image resolution. The Full Width at Half Maximum (FWHM) of SMLM images measures the width at which the intensity profile is half of its peak value, which can be an indicator of the resolution. As depicted in **Figure 32**, the FWHM distribution for SMLM images varies across different laboratories. Moreover, biological perturbation, such as drug treatments, can also influence the distribution of super-resolved SMLM images. For instance, Nocodazole drug treatment causes the depolymerization of microtubules, leading to changes in the microtubule network (**Supplementary Figure 6** on page 110).

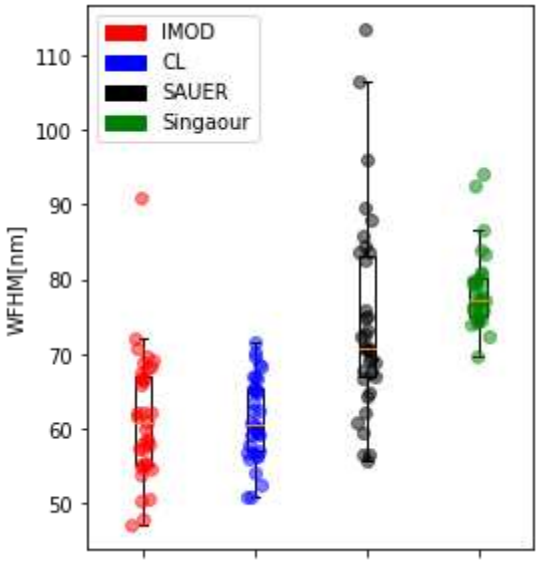


Figure 32. WFHM of SMLM images from different laboratories.
This figure displays the distribution of the WFHM values computed from dense SMLM images across different laboratories. Each color corresponds to a distinct laboratory. The WFHM values (y-axis) were computed using the intensity profiles drawn along lines perpendicular to the microtubules, with a total of 35 intensity profiles. Further details are available at https://docs.google.com/spreadsheets/d/1uohyqnrblcOshnzndooFcPhu3qkRtUScCt9h-BjlZXQ/edit?usp=sharing.

As illustrated in **Figure 33**, when test the ANNA-PALM model on an image obtained in a laboratory that was not represented in the training data, the reconstructed super-resolution

images by ANNA-PALM exhibit artifacts, e.g. at the intersections of microtubules. Therefore, it is crucial to provide a robust ANNA-PALM model when the test data is out-of-distribution.

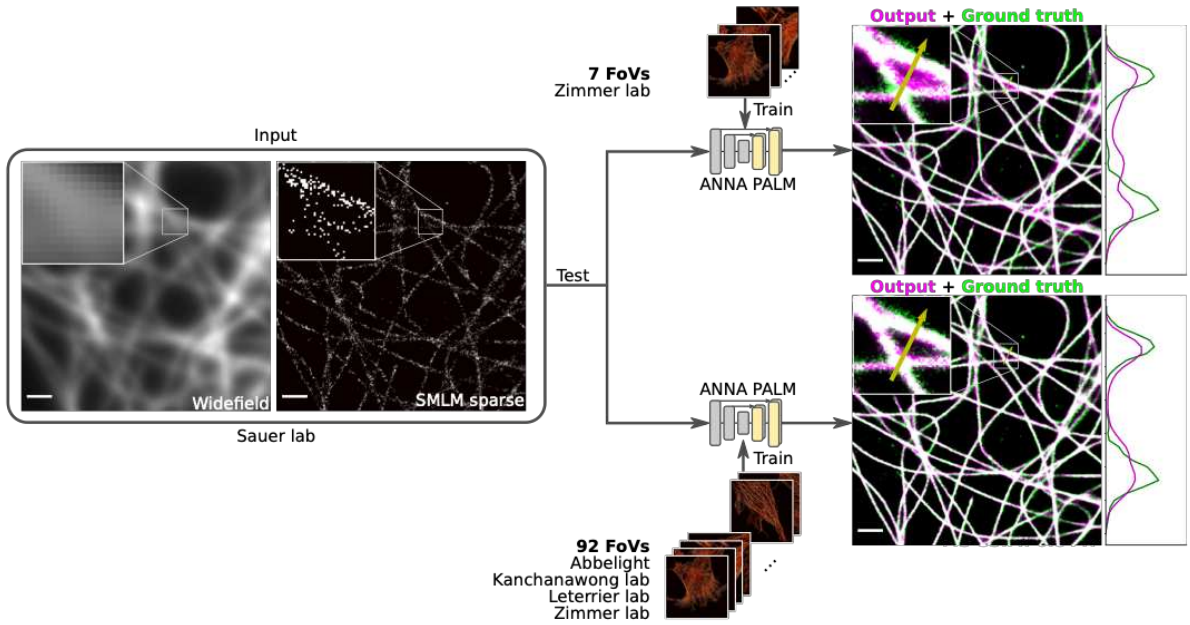


Figure 33. Training on shared SMLM data makes reconstruction of super-resolution images by deep learning more robust.
This figure illustrates a version of ANNA-PALM trained on 92 fields of view (FoVs) from four different labs (bottom) to the original ANNAPALM, which was trained on 7 FoVs from only one lab (top). On the left are ANNA-PALM's inputs: a sparse SMLM image of microtubules from 300 low-resolution frames and its corresponding widefield image, from a fifth, independent lab. On the right, the super-resolution images reconstructed by ANNAPALM (in pink) overlaid with the ground truth dense SMLM image from 40,000 low-resolution frames (in green). Intensity profiles along the indicated yellow arrow are displayed on the right. Scale bar: 1μm.

II.2.2. Summary of methods and results

An obvious approach to improve robustness of deep learning models is to retrain them using a larger and more varied training set (Hestness *et al.*, 2017). Since ShareLoc platform contains a large collection of SMLM data of microtubules (**Supplementary Figure 2** on page 103), I leveraged this data and re-trained ANNA-PALM (Ouyang *et al.*, 2018).

The the performance of the deep learning model depends on the quality of the reconstructed images. However, objectively evaluating the reconstruction quality is challenging (Maier-Hein *et al.*, 2023). I investigated different metrics for the evaluation:

- **Visual inspection**

By overlaying the deep learning model's output on the ground truth, we can visually assess the reconstruction quality. As shown in the left panels of **Figure 33**, this merged representation is an image in RGB (Red Green Blue) format. The ground truth is in the green channel and the models' reconstruction is in blue and red channels. The mixture of the red and blue channel appears magenta. Consequently, the ground truth appears green and the output appears magenta, and the overlapped areas of those two will be shown in white, representing the agreement between reconstructions and ground truth. The green areas represent false negatives and the pink ones display false positives.

- **Intensity profile**

When we draw a line at the same location on both the ground truth and the model's output, the intensity value of pixel along the line can be represented by the intensity profile line. Such representation is illustrated in **Figure 33**, where the y-axis is intensity values of the pixels, while the x-axis indicates the position along the line.

- **MS-SSIM** (Multi-Scale Structural Similarity)

MS-SSIM (Multi-Scale Structural Similarity) is an advanced version of the SSIM (Structural Similarity Index). SSIM allows to measure the similarity of two images at one scale while the MS-SSIM asses the similarity at various scales(Wang, Simoncelli and Bovik, 2003; Zhou Wang and Bovik, 2009). SSIM can provide more accurate quality assessment compared to traditional metrics such Mean Squared Error (MSE), as illustrated in **Figure 34**.

SSIM can be calculated by:

$$SSIM(x,y) = l(x,y) \cdot c(x,y) \cdot s(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_x\sigma_y + c_2)(\sigma_{xy} + c_3)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)(\sigma_x\sigma_y + c_3)}$$

where x and y are two images to compare, μ is mean of pixel value of x and y, σ^2 is the variance of pixel value of x and y and σ_{xy} is the covariance of x and y. This metric allows to compare the luminance (l), contrast (c) and structure (s)

between two images across multiple scales. It ranges between 0 and 1, where 1 represents the perfect agreement between x and y.

Figure 34. Comparisons of MSE (Mean Square Error) to SSIM and its variation.
This figure illustrates the comparison of SSIM and its variation CW-SSIM (Complex Wavelet-SSIM) compared to the metric MSE. Panel (a) shows the original image and panels (b)-(l) show the images with different distortions. For example, as illustrated in Figure 34, panels (c) and (f) have the same MSE, however, visually (f) is more similar to original image in panel (a). Image from (Zhou Wang and Bovik, 2009).

I first investigated whether increasing the quantity of SMLM training data for ANNA-PALM increases the quality of super-resolved images reconstructed by ANNA-PALM. In order to focus only on the effect of data quantity, I restricted the analysis to data from identical conditions acquired by Manish Singh (the same cell lines, culture conditions and imaging protocols). From the 81 available SMLM images, I randomly selected 21 for testing, while up to 60 were used for training (**Supplementary Figure 3** on page 104).

I compared ANNA-PALM models trained on a single SMLM image randomly chosen from these 60 images to the model trained on all 60 images. All models' reconstructions are in

much better agreement with the ground truth than the sparse SMLM input image. Visually, the quality of reconstructed images is improved when trained on 60 images. Quantitatively, the reconstruction quality of model trained on 60 images improved significantly (measured by MS-SSIM) (**Figure 1** on page 98**, Supplementary Figure 4** on page 105).

Then I assessed the robustness of ANNA-PALM to imaging protocols changes, I tested its performance on data from a laboratory not represented in the training data. I re-trained ANNA-PALM on up to 92 microtubule images from 4 laboratories, compared to the original model, which was trained on 7 images from our laboratory only. As shown in Table 2, I tested the re-trained models on the data from fifth laboratory and applied the rotations for models M1 to M4 (more details see **Table 1** on page 94).

Table 2. Retraining paradigm.

Datasets	Baseline M0	Model M1	Model M2	Model M3	Model M4
Abbelight (19)		Test	Train	Train	Train
Christophe Leterrier's lab (44)		Train	Test		
Christophe Zimmer's lab (7)	Train		Train		
Markus Sauer's lab (8)				Test	
Pakorn Kanchanawong's lab (22)				Train	Test

Overall, re-training ANNA-PALM on the larger and more diverse dataset shared on ShareLoc improves reconstruction quality, when testing images are acquired from the fifth labs. For three out of four rotations, the retrained models (M1-M4) performed significantly better than the baseline model M0. For model M3, the median MS-SSIM increased slightly but was not significant, likely owing to the small number of test images (n=8) (**Figure 2** on page 99**, Supplementary Figure 5** on page 106).

Finally, I investigated the robustness of ANNA-PALM models trained on a large number of technically diverse SMLM data (Table 2) to experimental perturbation of microtubule structures. The U2OS cell was incubated with 1μM of Nocodazole for 20 minutes. This treatment stimulated the. I applied the baseline model (M0) and models (M1-M4) above to sparse SMLM images of Nocodazole-treated cells, while only untreated cells were used as training data.

The baseline model (M0) already performed well, as it recovered virtually all microtubule filaments with few false negatives. However, it also hallucinated some filaments in regions of background noise. The reconstruction quality of chemically perturbed microtubule

images improves when ANNA-PALM was trained on a larger and technically diverse dataset (**Figure 3** on page 100**, Supplementary Figure 7** on page 111).

A manuscript detailing these results has been submitted and will be provided below and perspectives of ShareLoc platform will be discussed on page 138.

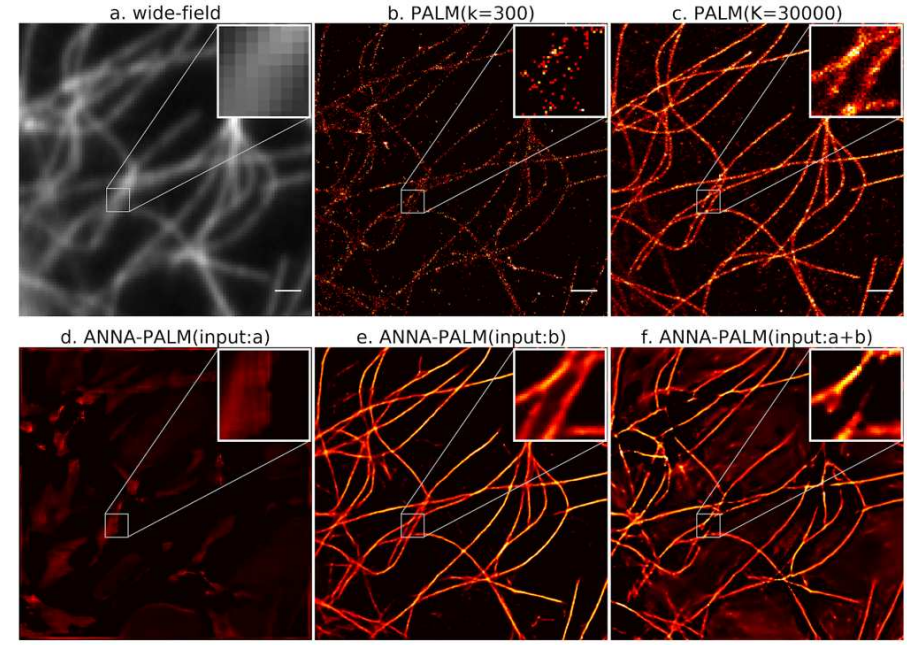
II.2.3. Problematics during the retraining

As previously discussed on page 43, both the sparse SMLM images and the widefield images are used as input for ANNA-PALM. Although the widefield images are optional, they can serve as complementary information to help reconstruct super-resolved images of higher quality (Ouyang *et al.*, 2018).

However, when retraining and testing ANNA-PALM using the SMLM data available on the ShareLoc platform, different dyes were used for different experiment, leading to different resolutions for widefield images. The original image registration method of ANNA-PALM (Reddy and Chatterji, 1996), implemented for widefield images of consistent resolution, is not suitable anymore. When widefield image resolutions varies, misalignment occurred between the widefield and dense SMLM images. This misalignment led to noticeable artifacts in the super-resolved SMLM images reconstructed by ANNA-PALM.

To address this, I developed an adapted image registration method for different resolutions. As illustrated in **Figure 35**, the registration between widefield and dense SMLM images is essential for high quality ANNA-PALM reconstruction. After the adapted image registration, the reconstruction quality improves.

Before correct image registration



After correct image registration

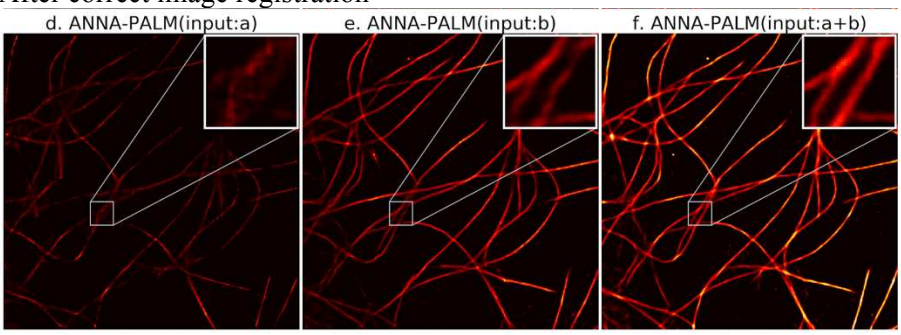


Figure 35. Registration of widefield and sparse SMLM data is essential for high quality reconstructions.

This figure shows the comparison of ANNA-PALM reconstruction quality before (middle) and after (bottom) correct image registration. (a) and (b) are the input of ANNA-PALM, and (c) is the ground truth. (a) Widefield image. (b) Sparse SMLM image. (c) Dense SMLM image. (d)-(f) are the ANNA-PALM reconstructions corresponding to different input data: (d) is the ANNA-PALM reconstruction taking only the widefield image (a) as input); (e) is the ANNA-PALM reconstruction taking only the sparse SMLM image (b) as input and (f) is the ANNA PALM reconstruction taking both the sparse image and the widefield image (a+b) as inputs. Note that the reconstruction in (f) contains multiple artifacts, due to a misalignment between widefield image and sparse image.

Article 2 (submitted): **Bai, J.***, Wei, O.*, Singh, M.K., Lelandais, B., Zimmer, C., n.d. Sharing localization microscopy data makes deep learning-based image reconstruction more robust. (* equal contribution)

Sharing localization microscopy data makes deep learning-based image reconstruction more robust

Jiachuan Bai[1]*, Wei Ouyang[2]*, Manish Kumar Singh[1], Benoît Lelandais[1,3], Christophe Zimmer[1]

[1]Institut Pasteur, Université Paris Cité, CNRS UMR 3691, Imaging and Modeling Unit, F-75015 Paris, France

[2]Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, Sweden

[3]Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France.

*equal contribution

Correspondence: czimmer@pasteur.fr

Keywords

Single molecule localization microscopy, deep learning, data sharing, open microscopy.

Abstract

In single molecule localization microscopy (SMLM), sequences of low-resolution image frames are computationally transformed into images with a resolution improved by ~10-100 times. Because tens of thousands of low-resolution frames are typically required to assemble a single super-resolution image, the throughput and speed of SMLM remain extremely low. We previously demonstrated that a computational strategy based on deep learning (ANNA-PALM) can reconstruct high-quality super-resolution images of cellular structures such as microtubules from much fewer low-resolution frames, thereby increasing imaging throughput by up to ~100-fold. However, the quality of deep learning-based reconstructions is generally strongly dependent on the quantity of training images and their consistency with the test images. If training and testing data are inconsistent, e.g. when a model is applied to images acquired in another laboratory using different imaging protocols, or under different biological conditions, deep learning methods are prone to generating reconstruction artifacts. Robustness to such changes between training and testing data can in principle be improved by training models on larger and more diverse data sets, but an empirical demonstration for SMLM data is lacking. Here, we leverage Shareloc (shareloc.xyz), a recently developed online platform for SMLM data sharing, to retrain ANNA-PALM models on roughly ten times more SMLM images of microtubules than before, originating from five distinct labs. We demonstrate empirically that this retraining improves ANNA-PALM reconstruction robustness compared to the original ANNA-PALM model when applied to microtubule images from a distinct lab or to images of chemically disrupted microtubule networks. Thus, SMLM data sharing enhances the quality and robustness of super-resolution image reconstruction by deep learning.

Introduction

Single molecule localization microscopy (SMLM)[1] methods (including PALM[2], (d)STORM[3,4], DNA-PAINT[5,6] and MINFLUX[7]) improve the spatial resolution of light microscopy down to a few tens of nanometers or less. Thanks to this considerable resolution improvement and its relative ease of implementation, SMLM is widely employed in life science research to image biological structures at near molecular resolution[1]. SMLM relies on the sequential acquisition of widefield, low resolution image frames, in which only a small random subset of fluorescent molecules emit light simultaneously, and hence appear as spatially isolated diffraction-limited spots. The spatial isolation of fluorescent spots from each other allows the coordinates of individual molecules to be computed with high precision and accuracy. The aggregated coordinates are then computationally assembled into a single super-resolution image. In order to generate a high-quality super-resolution image, the imaged structures must be sampled by a large number of single molecule localizations, which typically requires imaging ~10^4-10^5 low resolution frames[8]. As a result, SMLM is extremely slow, severely limiting imaging throughput or applications to studying cellular dynamics in live cells.

The success of deep learning based artificial intelligence (AI) for processing imaging data has inspired the development of deep learning methods to improve various aspects of super-resolution image reconstruction[9–17]. We previously developed ANNA-PALM[9], a deep learning method that reconstructs (predicts) high-quality super-resolution images from a considerably reduced number of low-resolution frames (**Figure S1**). Using ANNA-PALM, we demonstrated that super-resolution images of microtubules can be reconstructed from ~10^2 times fewer single molecule frames than usually required, e.g. from ~300 frames instead of ~30,000 frames, without trading off spatial resolution. This is only possible because strong prior information about the images is injected in the reconstruction process to compensate for the lacking single molecule localizations. ANNA-PALM learns this prior information by training a neural network on high quality super-resolution images obtained using standard SMLM imaging procedures, i.e. typically from ~10^4-10^5 low resolution frames.

As for any machine learning algorithm, a central question is whether the trained model is robust to technical changes in data acquisition and whether it can generalize to data obtained under different experimental conditions. In our initial demonstration[9], we trained ANNA-PALM on only seven SMLM images of microtubules obtained from a single lab (ours), using dSTORM. To explore the model's robustness, we assessed its performance at reconstructing images of microtubules acquired using DNA-PAINT, another SMLM technique, using different primary antibodies, different fluorescent dyes and different cameras, and obtained excellent reconstruction performance[9]. Nevertheless it remains to be determined to what extent ANNA-

PALM, when trained on images from one laboratory, is transferrable to data acquired in other laboratories. Transferability of trained deep learning models across laboratories is important, as it removes the need to retrain them and thus considerably facilitates their uptake. However, transferability is not guaranteed a priori because deep learning methods are notoriously sensitive to apparently insignificant changes in imaging parameters.

A related issue is robustness to biological perturbations. In our first study[9], we tested this robustness by treating cells with Taxol, which stabilizes microtubules and to Nocodazole, which depolymerizes them. In both cases, ANNA-PALM reconstructions were in good agreement with the ground-truth. However, for Nocodazole-treated cells the reconstructions exhibited minor hallucinations (short dim filaments without clear counterparts in the ground truth image), suggesting that this experiment probed the limits to robustness. Such limitations call for ways to make ANNA-PALM more robust to changes in biological structures.

Thus, an important goal is to further improve and test the robustness of ANNA-PALM-based super-resolution image reconstruction to both technical and biological changes. An obvious strategy to improve a deep model's robustness is to train it on a larger and more diverse data set. However, this strategy was hitherto challenged by the scarcity of publicly accessible SMLM data sets. Recently, we developed ShareLoc, an open online platform (shareloc.xyz) designed to facilitate the sharing and annotation of SMLM data[18]. Taking advantage of this platform, we retrained ANNA-PALM models on much larger and more diverse data sets and empirically evaluated the reconstruction quality and robustness to changes of laboratories or to a biological perturbation. We find that the retrained models exhibit higher reconstruction quality, with fewer artifacts, especially when applied to SMLM data from labs outside of the training data or to previously unseen biological conditions. Our report thus provides more robust ANNA-PALM models for SMLM images of microtubules and shows how open data sharing and reuse improves AI-based super-resolution image reconstruction.

Results

Assembling a large and diverse set of microtubule SMLM images from ShareLoc

From the ShareLoc platform, we downloaded 181 distinct SMLM images of microtubules, corresponding to 181 fields of view, and acquired in five different laboratories. In total, these data included ~2×10^9 single molecule localizations from ~8×10^6 low resolution image frames. These SMLM data were obtained for different cell lines, with different imaging protocols, that included labeling with different antibodies and/or dyes, different sample preparation and fixation techniques, cameras, objectives, laser excitation, and localization software, etc.

(**Table 1**, **Table S1** and **Figure S2**). All these differences in imaging and acquisition parameter can potentially affect neural network learning and lead to different model outputs for the same test data. Hereafter, we trained ANNA-PALM models on different subsets of these data, and assessed their reconstruction performance on held-out test data. We first used test data from the same lab as the training data, but thereafter assessed model robustness by testing on data acquired in different labs or in different biological conditions.

ANNA-PALM architecture and training strategy

ANNA-PALM is a deep learning algorithm inspired by the pix2pix network[19], a U-net based conditional generative adversarial network (cGAN). ANNA-PALM is trained to generate a high-quality super-resolution image from rapidly acquired single molecule localization data (**Figure S1**). More specifically, the training data consist in SMLM data of a specific biological structure (e.g., microtubules) computed from large numbers (e.g. 10^4-10^5) of consecutive low-resolution frames using standard localization algorithms. The SMLM images obtained from the entire sequence of frames are hereafter referred to as 'dense' SMLM images and serve either as target output for the neural network or as ground-truth for testing. By contrast, the input training data are images generated from a much smaller number of consecutive low-resolution frames (e.g. 200-500) from the beginning of the same sequence. Because these images have the same size as dense super-resolution images but are based on much fewer localizations, we refer to these images as 'sparse' SMLM images. ANNA-PALM takes a sparse SMLM image as input and the U-net is trained to generate an output image that is as close as possible to the corresponding dense SMLM image, as measured by a loss function that combines the L1 norm and a multi-scale structural similarity index (MS-SSIM)[20]. Because ANNA-PALM is a cGAN, the U-net generator network is trained in an adversarial way against a discriminator network, which aims to determine whether its input image is real or artificial, conditioned on the sparse SMLM image. If a widefield image is available for the same field of view, it is also included in the input data, in addition to the sparse SMLM image, and both images are fed to both the U-net and the discriminator network. Furthermore, an additional term in the loss function measures the consistency between the generated super-resolution image and the widefield (low-resolution) image (if available) using a third neural network. All three neural networks (the U-net generator, the discriminator, and the network for widefield images) are trained simultaneously[3]. Our training strategy uses different types of on-the-fly data augmentation techniques. We employ generic methods such as rotations, translations and elastic deformations, but also methods specific to SMLM data, such as creating many different sparse SMLM images for the same dense SMLM image by random subsampling of low-

resolution image frames, and adding random false localizations in the input data to mimic detection noise[9].

ANNA-PALM reconstruction quality increases with quantity of training images

We first asked if increasing the quantity of SMLM training data increases the quality of super-resolution images reconstructed by ANNA-PALM (**Figure 1, Figure S4**). In order to focus on the effect of data quantity alone, independently of the variability of technical imaging conditions or experimental conditions, we restricted this analysis to microtubule data acquired on the same cell types, in the same culture conditions and using the same imaging protocols (same microscope, same labeling, same fixation, same localization software) obtained within a single lab (ours). Out of 81 available SMLM images, we randomly chose 21 for testing and used up to 60 of the remaining images for training (**Figure S3**).

We first considered how ANNA-PALM performed when trained on a single SMLM image, randomly chosen among the 60. Because the performance of such a model depends strongly on which training image was chosen, we considered 15 distinct models, each trained on one of the 60 images. We applied these 15 models to each of the 21 test images, using a sparse SMLM image (obtained from 300 low resolution frames) and the corresponding widefield image as inputs (**Figure 1A**). We compared the output images to the corresponding dense SMLM images (obtained from 15,000 to 60,000 frames, with an average of 58,568 frames; **Table S1**), which we considered as ground truth images (**Figure 1B,C**). Remarkably, ANNA-PALM models trained on a single SMLM image already showed rather good reconstruction performance, since the vast majority of the microtubules present in the ground-truth images were correctly recovered by the ANNA-PALM predictions (**Figure 1B,C** and **Fig S4B**). The reconstructed images represented much better characterizations of microtubule filaments than the sparse SMLM input data. More quantitatively, the MS-SSIM of ANNA-PALM reconstructions (which lies between 0 and 1 and equals 1.0 for images in perfect agreement with the ground truth) ranged from 0.55 to 0.95, with a median value of 0.86 (and interquartile range 0.06), whereas the unprocessed sparse SMLM image had much lower MS-SSIM, ranging from 0.01 to 0.43 only, with a median of 0.1 ($p<10^{-7}$) (**Figure 1E** and **Figure S4C**). The high quality of models trained on just one image may be come as a surprise, since deep learning methods, when trained from scratch, typically require large amounts of training data to achieve satisfying performance. We attribute this high performance to the relative simplicity of microtubule structures, combined with the large size of each super-resolution image (from 858 by 858 pixels to 2600 by 2600 pixels, corresponding to a field of view of 17.15 μm by 17.15 μm or 52 μm by 52 μm, respectively) and the resulting high internal redundancy,

combined with our aggressive data augmentation. Nevertheless, the reconstructed images also tended to display artifacts, including hallucinated filaments due to noise in the sparse SMLM images and blurry regions where filaments are close to each other, especially at microtubule intersections (**Figure 1CD** and **Fig S4B**).

We then compared these reconstructions to those obtained with a model trained on all 60 training images. Examples of reconstructed images are shown in **Figure 1C** and in **Figure S4B**. Visual inspection indicates that the quality of reconstructed images is markedly improved in comparison with the models trained on a single image. For example, artifacts such as hallucinated filaments mentioned above or blurred microtubule extremities were mostly absent when using the model trained on 60 images (**Figure 1C,D**). Quantitative evaluation on the 21 test images confirmed that reconstruction quality improved significantly, from a median MS-SSIM of 0.64 to a median of 0.86 ($p<10^{-7}$) (**Figure 1E**). Thus, our data empirically confirm that training on a larger SMLM data set of technically similar images significantly increases the quality of ANNA-PALM reconstructions.

Training ANNA-PALM on ShareLoc data improves robustness to imaging protocols

To assess the robustness of ANNA-PALM to changes in imaging protocols, we then tested its performance on data from a laboratory not represented in the training data (**Figure 2**). We first considered n=21 sparse SMLM images of microtubules from lab K as testing data (obtained from 300 low resolution frames). We then used ShareLoc data to retrain ANNA-PALM on 78 SMLM images of microtubules from four labs (A, L, Z, S), leading to a new model hereafter called M4. (Note that more images would be available for training from lab Z, see above, but we limited the number of images to avoid excessive overrepresentation of data from a single lab). As a baseline model, we also considered our original ANNA-PALM model (hereafter called M0), which was trained on seven SMLM images of microtubules from our lab (lab Z) only (**Figure 2B**). **Figure 2C** shows an example SMLM image from lab K reconstructed with ANNA-PALM models M0 and M4 in comparison with the dense ground truth SMLM (obtained from 300 low resolution frames). Both models generated images in very good agreement with the ground truth and represented a very strong improvement over the sparse SMLM images, as measured by MS-SSIM (**Figure 2A,C,F**). However, model M0 also occasionally generated some low-intensity hallucinations (**Figure 2C,D**). These hallucinations mostly disappeared in the reconstructions of model M4. The improvement of the new model was quantitively confirmed for the n=17 images from lab K, as model M4 led to a very significant improvement of the MS-SSIM over M0 from a median of 0.63 to a median of 0.85 ($p=10^{-3}$, **Figure 2F**). Visual inspection of other examples confirmed that the new model corrected many imperfections of

the baseline model (**Figure S5**). Next, we rotated the split between training and testing data among labs, using as test sets data from labs A, L, S and K, in turn, and training four models M1-M4 on the data from the four remaining labs, making use of all available images (between n=56 and n=92 images; **Figure 2E**). All models (M0-M4) reconstructed images in much better agreement with the ground truth than the sparse SMLM input image (**Figure 2F** and **Figure S5B**). For three out of four rotations, the new models (M1-M4) performed significantly better than the baseline model M0, with the median MS-SSIM increasing from 0.59 to 0.82 for M1 ($p<10^{-4}$, n=19) and from 0.56 to 0.85 for M2 ($p<10^{-10}$, n=44), (see above for M4) (**Figure 2D**). For model M3, the median MS-SSIM increased slightly from 0.76 to 0.83, but this was not significant, likely owing to the small number of test images (p=0.34, n=8).

Overall, these results indicate that training ANNA-PALM on the larger and more diverse data set provided by ShareLoc improves reconstruction quality of images acquired outside of the labs providing the training data. Thus, leveraging the quantity and diversity of SMLM images available online makes ANNA-PALM models significantly more robust to technical variations in image acquisition.

Training ANNA-PALM on ShareLoc data improves robustness to biological perturbations

Finally, we asked whether the ANNA-PALM models trained on a large number of technically diverse SMLM data are also more robust to experimental alterations of microtubule structures. This is not guaranteed *a priori*, since the training data contain variations in imaging parameters, but not variations in biological conditions. We therefore examined the performance of the four models above at reconstructing images of microtubules subject to a strong experimental perturbation. Specifically, we incubated U2OS cells with 1 μM of Nocodazole, a microtubule depolymerizing drug, for 20 minutes. This treatment stimulated the depolymerization of microtubules, which led to strong changes in the morphology of filaments (**Figure S6**). We applied both the baseline model M0 and models M1-M4 above to sparse SMLM images (with corresponding widefield images) of Nocodazole-treated cells (**Figure 3A-C**). We stress that only untreated cells were used to define the training data, hence the Nocodazole-induced perturbations have not previously been seen by the neural networks and represent a novel event for all models. Despite this, the baseline model (M0) already performed well, as it recovered virtually all microtubule filaments with few false negatives, as shown in **Figure 3C** (left) and **Figure S7**). However, it also hallucinated some filaments in regions of background noise (**Figure 3C,D,** left, **Figure S7**). This hallucinatory tendency is unsurprising, given that the model was trained exclusively on images of untreated cells, which

typically display much denser microtubule networks than Nocodazole-treated cells (**Figure S6**). Remarkably, however, most of these artifacts disappeared when using ANNA-PALM model M4 above, which was trained on 78 images from labs A, L, Z and S, and the overall reconstruction quality improved both visually and according to MS-SSIM (**Figure 3C,D,** right). Analysis of n=19 images of Nocodazole-treated cells confirmed a strong and very significant improvement of the reconstructed images from a median MS-SSIM of 0.73 to 0.83 using model M4 compared to the baseline model M0 ($p<10^{-3}$, n=19) (**Figure 3F**). The same conclusion held true with models M1-M3, which were trained on subsets of 81, 52 and 92 images, respectively, from four of the five labs, again exclusively containing untreated cells (**Figure 3E, Figure S7**). In all cases, ANNA-PALM reconstructions were much better than with the baseline model, with the median MS-SSIM improving from 0.73 to between 0.88 and 0.90 (M1: $p<10^{-5}$; M2: $p<10^{-4}$; M3: $p<10^{-5}$; M4: $p<10^{-3}$; n=19) (**Figure 3F**).

Thus, our results indicate that when ANNA-PALM is trained on a larger and technically more diverse set of microtubule images, the model also provides high quality reconstructions of chemically perturbed microtubule networks, despite never having been exposed to such an experimental condition. More generally, it suggests that training on diverse data from the community through ShareLoc allows image reconstruction models to generalize to novel and biologically meaningful alterations of the imaged structures.

Discussion

In this report, we took advantage of the ShareLoc platform to retrain ANNA-PALM models on a large and diverse set of SMLM images acquired by multiple independent laboratories. We assessed the performance of these models at generating high quality super-resolution images of microtubules from sparse, rapidly acquired single molecule data, in comparison with the baseline model[9] , which was trained on only seven images from a single lab. We first showed that training ANNA-PALM on a larger number of images improves reconstruction quality when applied to test images from the same lab (**Figure 1** and **Figure S4**). While this result is not surprising qualitatively, our data establish that the improvement is notable and quantitatively significant for an increase of training data size by roughly one order of magnitude. We then showed that training ANNA-PALM on more images from multiple labs improves the ability of ANNA-PALM to reconstruct images from another lab without retraining, i.e. improves robustness to changes in technical imaging parameters (**Figure 2, Figure S5**). This ability of ANNA-PALM to reconstruct images from another laboratory without retraining is significant, since it broadens the scope of users to teams without the computational skills or computing resources (e.g. GPUs) required for model retraining.

Finally, we showed that training ANNA-PALM on ShareLoc data also strongly improves image reconstruction for chemically perturbed microtubules, even when such perturbations were entirely absent from the training data (**Figure 3, Figure S7**). Indeed, the reconstructed images displayed a strong disruption of the microtubule network in response to the Nocodazole treatment, thus revealing a phenotype that was not previously seen by the model. This is particularly meaningful, because it addresses the concern that deep learning-based image reconstruction may only recover phenotypes already present in the training data and may be biased against the discovery of new phenotypes[21]. The improved robustness of ANNA-PALM to altered phenotypes increases its applicability to high-throughput phenotypic screening approaches, for example in drug discovery[22]. Thus, training on diverse data such as those provided by ShareLoc, increases the potential of ANNA-PALM to uncover biologically meaningful changes in molecular structures. In order to facilitate the reuse of these models, we make them available on the ShareLoc platform for download.

As a main limitation of our study, we acknowledge that thus far our claims are empirically supported only by analyses of microtubule images. Our focus on microtubules was motivated mostly by the availability of more than 100 microtubule SMLM images on ShareLoc. However, as we previously showed, the methodology underlying ANNA-PALM is generic and applicable to multiple other structures, such as nuclear pores or mitochondria[9]. Therefore, we expect that similar improvements in reconstruction quality will be possible for various biological structures once more data are available on ShareLoc to train adequate reconstruction models. We hope that this demonstration of how SMLM data accessibility leads to more powerful deep learning models will further stimulate the community to enrich the ShareLoc platform with SMLM data of a wide range of molecular structures. With the anticipated upscaling of the quantity and variety of data on ShareLoc, we foresee that it will be possible to train much more powerful and robust ANNA-PALM models capable of analyzing a larger variety of biological structures with even larger gains of acquisition time. This in turn will help to overcome the current contradictions between temporal and spatial resolution and facilitate high-throughput and live cell super-resolution imaging.

Acknowledgements

We thank C. Thépenier for useful comments on the manuscript. J.B. and C.Z. acknowledge funding by Région Ile-de-France in the framework of DIM ELICIT, by the Inception program (Investissement d'Avenir grant ANR-16-CONV-0005), by the company Abbelight and by Institut Pasteur. M.S. and C.Z. also acknowledge funding by Agence Nationale de la Recherche (project ANR 21 CE45-003-02 'Qdynamics').

Declaration of interests

321 W. Ouyang and C. Zimmer are holders of a patent on ANNA-PALM.

References

1. Lelek, M., Gyparaki, M.T., Beliu, G., Schueder, F., Griffié, J., Manley, S., Jungmann, R., Sauer, M., Lakadamyali, M., and Zimmer, C. (2021). Single-molecule localization microscopy. Nat Rev Methods Primers *1*, 39. 10.1038/s43586-021-00038-x.
2. Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science *313*, 1642–1645. 10.1126/science.1127344.
3. Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods *3*, 793–796. 10.1038/nmeth929.
4. Heilemann, M., van de Linde, S., Schüttpelz, M., Kasper, R., Seefeldt, B., Mukherjee, A., Tinnefeld, P., and Sauer, M. (2008). Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. Angew. Chemie. Int. Ed *47*, 6172–6176.
5. Jungmann, R., Steinhauer, C., Scheible, M., Kuzyk, A., Tinnefeld, P., and Simmel, F.C. (2010). Single-Molecule Kinetics and Super-Resolution Microscopy by Fluorescence Imaging of Transient Binding on DNA Origami. Nano Lett. *10*, 4756–4761. 10.1021/nl103427w.
6. Reinhardt, S.C.M., Masullo, L.A., Baudrexel, I., Steen, P.R., Kowalewski, R., Eklund, A.S., Strauss, S., Unterauer, E.M., Schlichthaerle, T., Strauss, M.T., et al. (2023). Ångström-resolution fluorescence microscopy. Nature *617*, 711–716. 10.1038/s41586-023-05925-9.
7. Balzarotti, F., Eilers, Y., Gwosch, K.C., Gynnå, A.H., Westphal, V., Stefani, F.D., Elf, J., and Hell, S.W. (2017). Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. Science *355*, 606–612. 10.1126/science.aak9913.
8. Legant, W.R., Shao, L., Grimm, J.B., Brown, T.A., Milkie, D.E., Avants, B.B., Lavis, L.D., and Betzig, E. (2016). High-density three-dimensional localization microscopy across large volumes. Nature Methods *13*, 359–365. 10.1038/nmeth.3797.

References

1. Lelek, M., Gyparaki, M.T., Beliu, G., Schueder, F., Griffié, J., Manley, S., Jungmann, R., Sauer, M., Lakadamyali, M., and Zimmer, C. (2021). Single-molecule localization microscopy. Nat Rev Methods Primers *1*, 39. 10.1038/s43586-021-00038-x.
2. Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science *313*, 1642–1645. 10.1126/science.1127344.
3. Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods *3*, 793–796. 10.1038/nmeth929.
4. Heilemann, M., van de Linde, S., Schüttpelz, M., Kasper, R., Seefeldt, B., Mukherjee, A., Tinnefeld, P., and Sauer, M. (2008). Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. Angew. Chemie. Int. Ed *47*, 6172–6176.
5. Jungmann, R., Steinhauer, C., Scheible, M., Kuzyk, A., Tinnefeld, P., and Simmel, F.C. (2010). Single-Molecule Kinetics and Super-Resolution Microscopy by Fluorescence Imaging of Transient Binding on DNA Origami. Nano Lett. *10*, 4756–4761. 10.1021/nl103427w.
6. Reinhardt, S.C.M., Masullo, L.A., Baudrexel, I., Steen, P.R., Kowalewski, R., Eklund, A.S., Strauss, S., Unterauer, E.M., Schlichthaerle, T., Strauss, M.T., et al. (2023). Ångström-resolution fluorescence microscopy. Nature *617*, 711–716. 10.1038/s41586-023-05925-9.
7. Balzarotti, F., Eilers, Y., Gwosch, K.C., Gynnå, A.H., Westphal, V., Stefani, F.D., Elf, J., and Hell, S.W. (2017). Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. Science *355*, 606–612. 10.1126/science.aak9913.
8. Legant, W.R., Shao, L., Grimm, J.B., Brown, T.A., Milkie, D.E., Avants, B.B., Lavis, L.D., and Betzig, E. (2016). High-density three-dimensional localization microscopy across large volumes. Nature Methods *13*, 359–365. 10.1038/nmeth.3797.

9. Ouyang, W., Aristov, A., Lelek, M., Hao, X., and Zimmer, C. (2018). Deep learning massively accelerates super-resolution localization microscopy. Nat. Biotechnol. *36*, 460–468. 10.1038/nbt.4106.

10. Nehme, E., Weiss, L.E., Michaeli, T., and Shechtman, Y. (2018). Deep-STORM: super-resolution single-molecule microscopy by deep learning. Optica *5*, 458. 10.1364/OPTICA.5.000458.

11. Boyd, N., Jonas, E., Babcock, H., and Recht, B. (2018). DeepLoco: Fast 3D Localization Microscopy Using Neural Networks (Biophysics) 10.1101/267096.

12. Kumar Gaire, S., Zhang, Y., Li, H., Yu, R., Zhang, H.F., and Ying, L. (2020). Accelerating multicolor spectroscopic single-molecule localization microscopy using deep learning. Biomed. Opt. Express *11*, 2705. 10.1364/BOE.391806.

13. Speiser, A., Müller, L.-R., Hoess, P., Matti, U., Obara, C.J., Legant, W.R., Kreshuk, A., Macke, J.H., Ries, J., and Turaga, S.C. (2021). Deep learning enables fast and dense single-molecule localization with high accuracy. Nat Methods *18*, 1082–1090. 10.1038/s41592-021-01236-x.

14. Narayanasamy, K.K., Rahm, J.V., Tourani, S., and Heilemann, M. (2022). Fast DNA-PAINT imaging using a deep neural network. Nat Commun *13*, 5047. 10.1038/s41467-022-32626-0.

15. Qiao, C., Li, D., Guo, Y., Liu, C., Jiang, T., Dai, Q., and Li, D. (2021). Evaluation and development of deep neural networks for image super-resolution in optical microscopy. Nat Methods *18*, 194–202. 10.1038/s41592-020-01048-5.

16. Saguy, A., Alalouf, O., Opatovski, N., Jang, S., Heilemann, M., and Shechtman, Y. (2023). DBlink: dynamic localization microscopy in super spatiotemporal resolution via deep learning. Nat Methods, 1–10. 10.1038/s41592-023-01966-0.

17. Chen, R., Tang, X., Zhao, Y., Shen, Z., Zhang, M., Shen, Y., Li, T., Chung, C.H.Y., Zhang, L., Wang, J., et al. (2023). Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging. Nat Commun *14*, 2854. 10.1038/s41467-023-38452-2.

18. Ouyang, W., Bai, J., Singh, M.K., Leterrier, C., Barthelemy, P., Barnett, S.F.H., Klein, T., Sauer, M., Kanchanawong, P., Bourg, N., et al. (2022). ShareLoc — an open platform for sharing localization microscopy data. Nat Methods *19*, 1331–1333. 10.1038/s41592-022-01659-0.

19. Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A.A. (2018). Image-to-Image Translation with Conditional Adversarial Networks.

20. Zhao, H., Gallo, O., Frosio, I., and Kautz, J. (2017). Loss Functions for Image Restoration with Neural Networks. IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING *3*, 47–57. 10.1109/TCI.2016.2644865.

21. Belthangady, C., and Royer, L.A. (2019). Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. Nat Methods *16*, 1215–1225. 10.1038/s41592-019-0458-z.

22. Krentzel, D., Shorte, S.L., and Zimmer, C. (2023). Deep learning in image-based phenotypic drug discovery. Trends in Cell Biology. 10.1016/j.tcb.2022.11.011.

TABLES

Lab label	A	K	L	S	Z
Number of SMLM images with or without widefield image	19 with widefield	22 (17 with widefield, 5 without widefield)	44 (44 with widefield)	8 (8 with widefield)	7+81 (81 with widefield were used for Fig. 1,S4; 6 with widefield + 1 without widefield were used for Fig. 2,3,S5,S7)
Cell lines	COS7	MDCK, REF52, NIH3T3	COS7	U2OS, COS7	U-373 MG
Primary antibodies	Mouse anti-alpha-tubulin	Rabbit anti-alpha-tubulin		Mouse anti-beta-tubulin or rabbit anti-alpha-tubulin	Rat anti-alpha-tubulin
Fluorophores	Alexa-647, CF680	Alexa 647	Alexa-647	Alexa-647, Alexa-532	Alexa-647
Localization software	ThunderSTORM	ThunderSTORM		rapidSTORM	ThunderSTORM
PI or CEO	Jean-Baptiste Marie	Pakorn Kanchanawong	Christophe Letterier	Markus Sauer	Christophe Zimmer
Institution	Abbelight, Cachan, France	University of Singapore, Singapore	Aix Marseille Université, CNRS, INP UMR7051, Marseille, France	University of Würzburg, Würzburg, Germany	Institut Pasteur, Université Paris Cité, Paris, France

Table 1: Overview of SMLM data sets used to train or test ANNA-PALM

The number of low-resolution frames underlying each SMLM image ranges from 4,947 to 100,000 with a mean of 41,932 and a median of 40,000. For more details, see the spreadsheet Table S1.

FIGURE LEGENDS

Figure 1: Training ANNA-PALM on larger data sets improves super-resolution image reconstruction

(A) Input: a widefield image and a sparse SMLM image of immunolabeled microtubules not previously seen by the ANNA-PALM model were used as test data. The sparse SMLM image was obtained from the first 300 frames of an SMLM experiment and contains 21,703 localizations. The images show a 10.24 μm by 10.24 μm region of interest, and the inset 1 μm by 1 μm.

(**B**) Two ANNA-PALM models, M and M', were trained on SMLM images of microtubules from the same lab (lab Z). Model M was trained on a single SMLM image (i.e. a single field of view, FoV) from lab Z. Model M' was trained on 60 distinct SMLM images from lab Z.

(**C**) ANNA-PALM output: the output images generated by the two models are shown in pink, superposed to the ground-truth, shown in green. The ground-truth is the dense SMLM image obtained from 60,000 frames and contains 1,358,693 localizations. The agreement between output and ground truth images is measured by the multi-scale structural similarity index (MS-SSIM). For more examples, see **Figure S4B**.

(**D**) Normalized intensity profiles of the two ANNA-PALM images and the ground-truth image along the yellow arrows shown in the insets of (**C**). Note that M incorrectly predicted a third peak of lower intensity between the two main peaks in region (i) (asterisk) and incorrectly predicted a valley of above-background intensity in between the two main peaks in region (ii) (asterisk), whereas M' correctly recovered two peaks separated by background in both cases.

(**E**) Boxplots compare the MS-SSIM of ANNA-PALM reconstructions for models trained on a single image (M) or on 60 images (M'), or the unprocessed sparse SMLM input images, relative to the ground-truth, for n=21 test images from lab Z. To robustly assess ANNA-PALM performance when trained on a single image, we trained 15 models on 15 distinct, randomly chosen single images (not overlapping the test image set). Each dot corresponds to a single test image and indicates the MS-SSIM averaged over these 15 models. Horizontal orange lines show medians. Lower and upper box edges correspond to the 25 and 75% percentiles and whiskers show the full data range except for outliers. Grey lines connect values for identical test images. The filled black dots and connected by thicker lines correspond to the image shown in (**C**). Indicated p-values are from a signed-rank Wilcoxon test. Scale bars in (**A**), (**C**): 0.5 μm.

Figure 2: Training ANNA-PALM on ShareLoc data improves reconstruction robustness to imaging protocols

(**A**) Input: a widefield image and a sparse SMLM image of immunolabeled microtubules acquired by lab K. The sparse SMLM image was obtained from the first 300 frames of an SMLM experiment and contains 8,850 localizations. The images show a 10.24 μm by 10.24 μm region of interest, and the inset 1 μm by 1 μm.

(**B**) Two ANNA-PALM models were trained: model M0 was trained on the same seven images from lab Z as previously[9]. Model M4 was trained on 78 images from labs A, L, Z, and S. The training data included no images from lab K.

(**C**) ANNA-PALM output: the output images generated by the two models are shown in pink, superposed to the ground-truth shown in green. The ground-truth is the dense SMLM image obtained from 20,000 frames and contains 578,850 localizations. For more examples, see **Figure S5**.

(**D**) Normalized intensity profiles of the two ANNA-PALM images and the ground-truth image along the yellow arrow shown in the insets of (**C**). Note that the two main intensity peaks were well recovered by both models in both regions; but with model M0, a third peak of lower intensity, corresponding to hallucinated filaments in (**C**), was predicted between the two main peaks (asterisks), whereas no such artifact was generated by model M4.

(**E**) Overview of data sets used for training and testing the different models. Model M0 (trained on seven images from lab Z) was tested on images from each of the four other labs (A, L, S, and K), in turn. Models M1-M4 were trained on between 56 and 92 images pooled from four labs in different combinations, and tested on images from the remaining fifth lab, as indicated.

(**F**) Boxplots compare the MS-SSIM of ANNA-PALM reconstructions using model M0 or models M1-M4 or the sparse SMLM data relative to the ground truth, on the same test data. Each dot corresponds to a single test image. The number of test images n is indicated. Medians are shown as horizontal orange lines. Lower and upper box edges correspond to the 25 and 75% percentiles and whiskers show the full data range except for outliers. Grey lines link data points corresponding to the same image. Indicated p-values are from Wilcoxon signed-rank tests. Note that the median reconstruction quality improved for all four models M1-M4 compared to M0. The increase is significant for all models except M3. Scale bars in (**A**), (**C**): 0.5 μm.

Figure 3: Training ANNA-PALM on ShareLoc data improves reconstruction robustness to experimental perturbations

(**A**) Input: a widefield image and a sparse SMLM image of immunolabeled microtubules in cells exposed to a Nocodazole treatment. The sparse SMLM image was obtained from the first 300 frames of the sequence and contains 5,652 localizations. The images show a 10.24 μm by 10.24 μm region of interest, and the inset 1 μm by 1 μm.

(**B**) Two ANNA-PALM models were trained: model M0 (left) was trained on the same seven images from lab Z as previously[9]. Model M4 (right) was trained on 78 images from labs A, L, Z and S (as in **Figure 2B**). The training data only included untreated cells.

(**C**) ANNA-PALM output: the output images generated by the two models are shown in pink, superposed to the ground-truth in green. The ground-truth is the dense SMLM image obtained from 60,000 frames and contains 759,390 localizations.

(**D**) Normalized intensity profiles of the two ANNA-PALM images and the ground-truth image along the yellow arrows shown in the insets of (**C**). Note that both models correctly recovered the main intensity peak, but M0 incorrectly predicted secondary bumps (asterisks), corresponding to hallucinated structures in (**C**), which were absent from reconstructions by M4.

(**E**) Overview of data sets used for training. Models are identical to those used in **Figure 2**, but applied to Nocodazole-treated cells from lab Z, as indicated.

(**F**) Boxplots compare the MS-SSIM of ANNA-PALM reconstructions relative to the ground truth using model M0 or models M1-M4, on n=19 images of Nocodazole-treated cells. Each dot corresponds to a single test image. Medians are shown as horizontal orange lines. Lower and upper box edges correspond to the 25 and 75% percentiles and whiskers show the full data range except for outliers. Grey lines link data points corresponding to the same image. Indicated p-values are from Wilcoxon signed-rank tests. Note that the median reconstruction quality improves for all four models M1-M4 compared to M0 and that this increase is significant in all four cases. Scale bars in (**A**), (**C**): 0.5 μm.

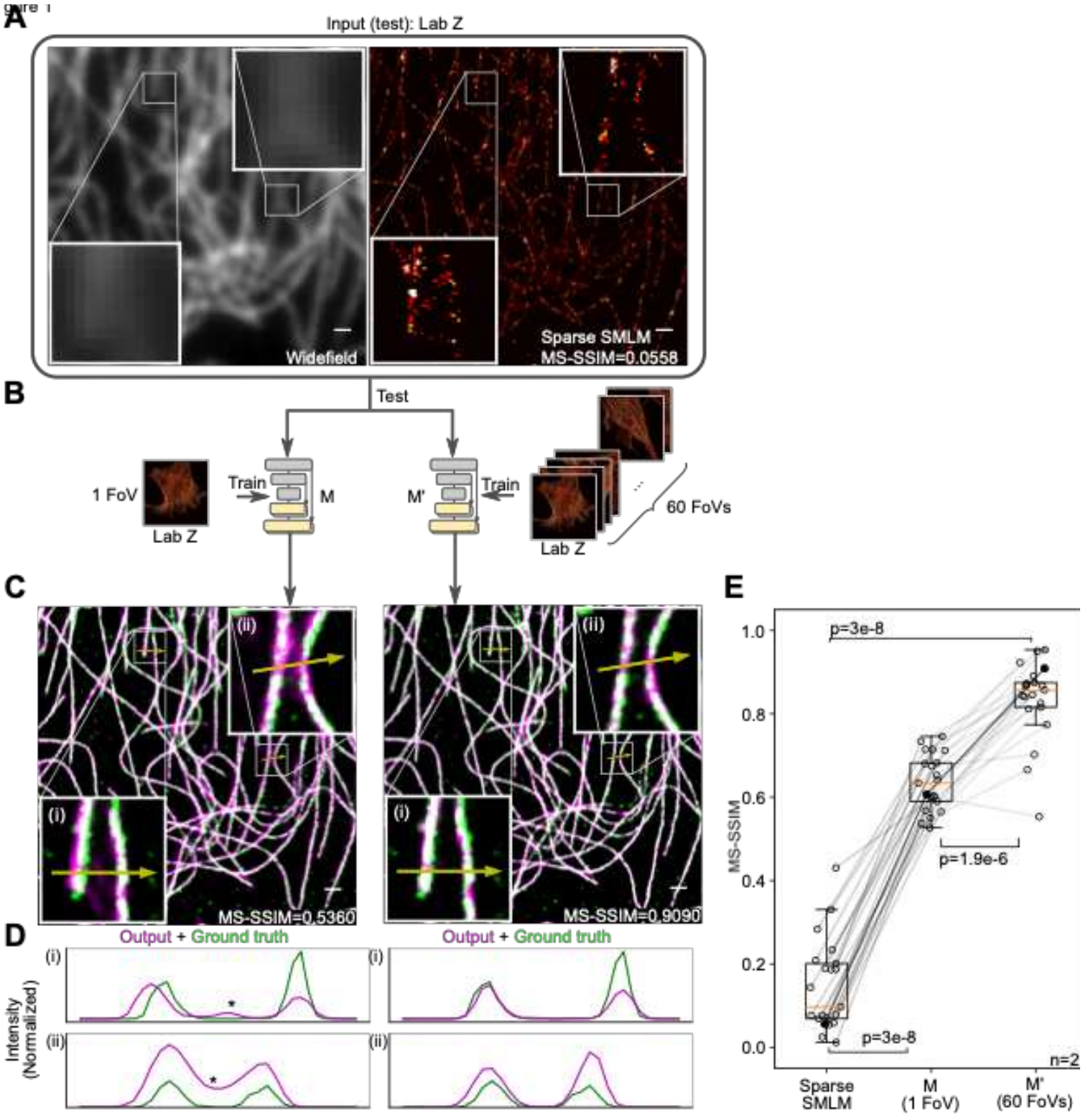
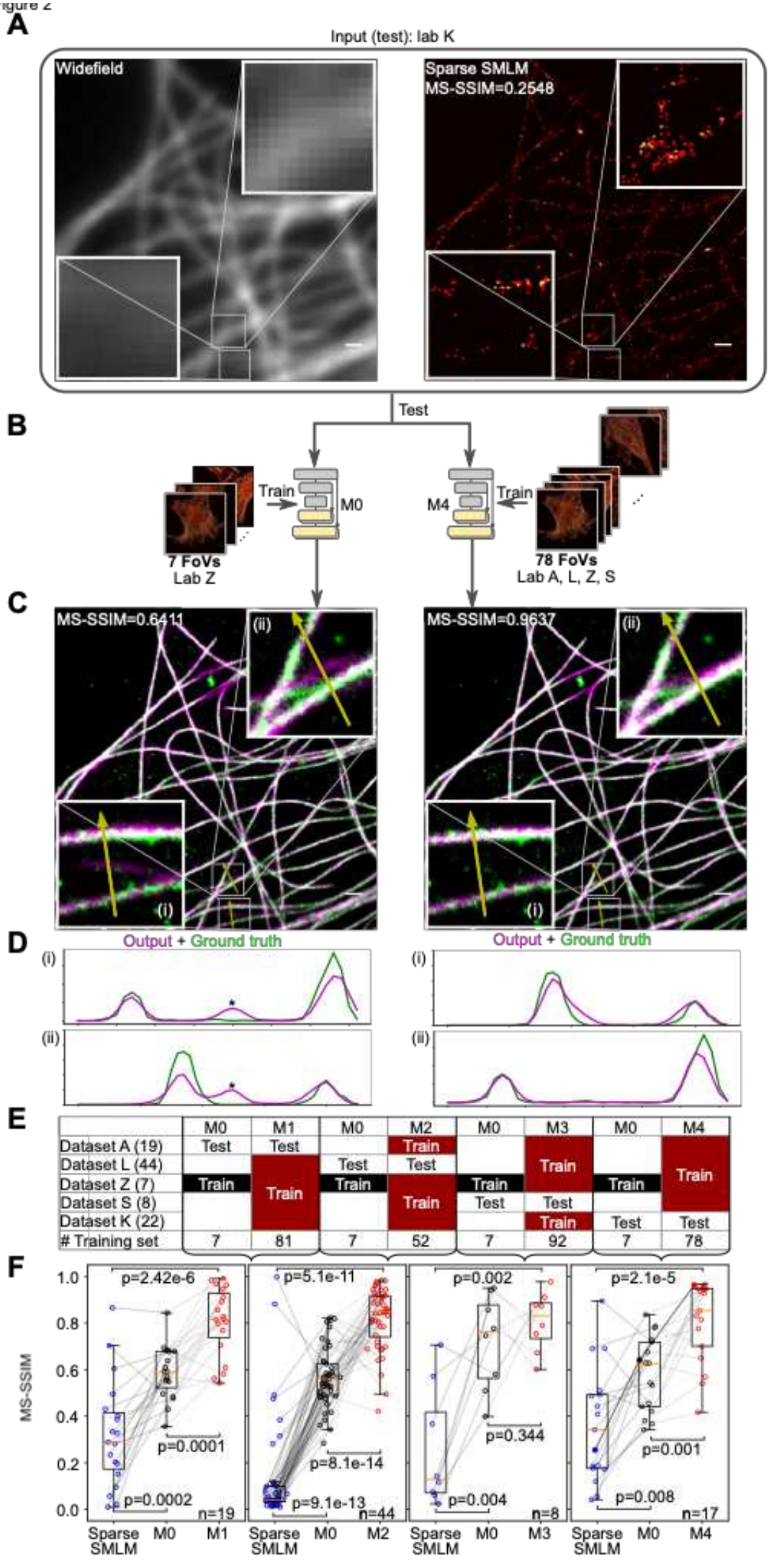
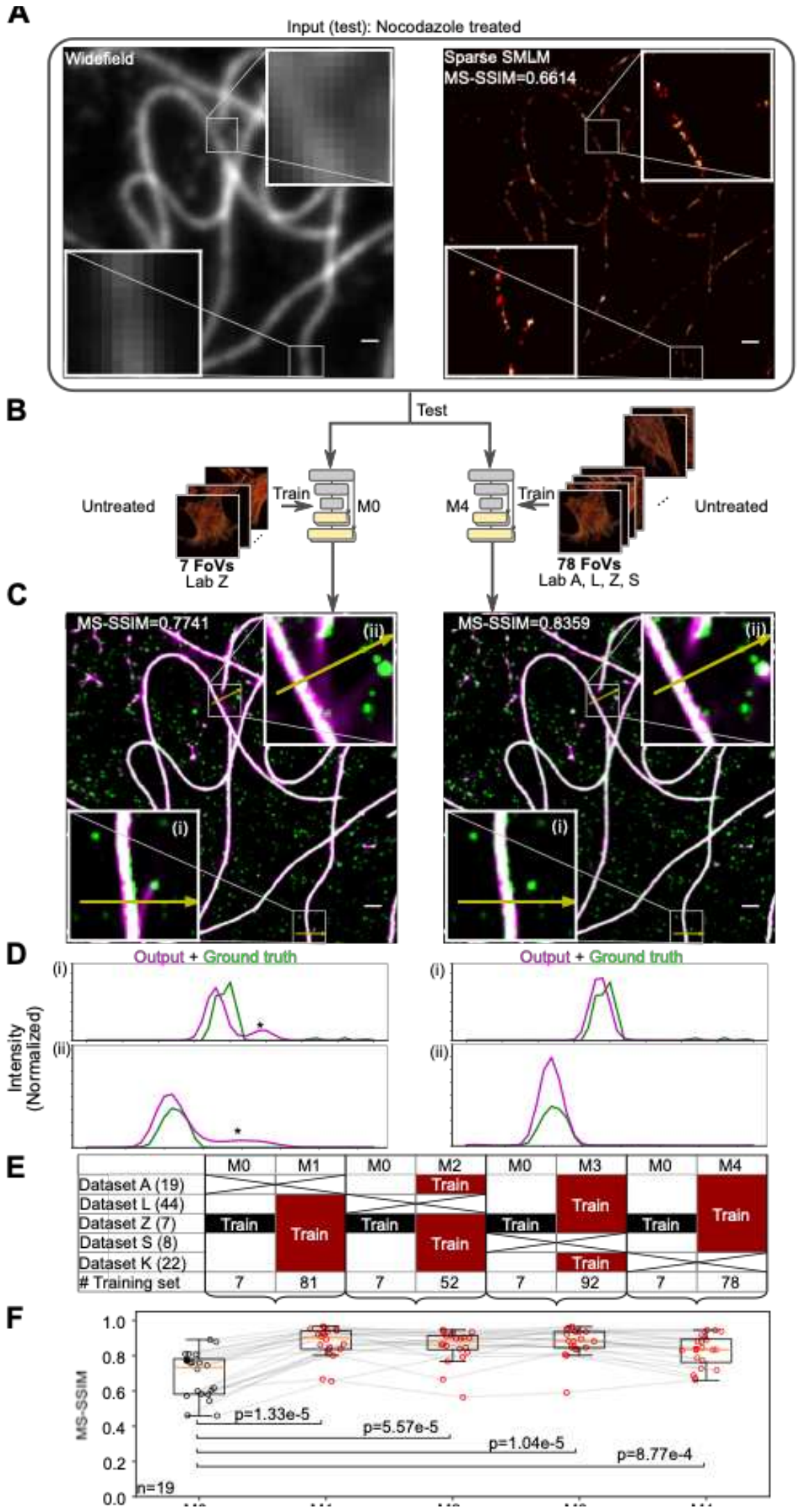

A
Input (test): Lab Z
Widefield
Sparse SMLM
MS-SSIM=0.0558
B
Test
1 FoV
Lab Z
Train
M
M'
Train
60 FoVs
Lab Z
C
(i)
(ii)
MS-SSIM=0.5360
MS-SSIM=0.9090
D
Output + Ground truth
Output + Ground truth
Intensity (Normalized)
E
MS-SSIM
p=3e-8
p=1.9e-6
p=3e-8
n=2
Sparse SMLM
M (1 FoV)
M' (60 FoVs)

Figure 2



	M0	M1	M0	M2	M0	M3	M0	M4
Dataset A (19)	Test	Test		Train		Train		Train
Dataset L (44)		Train	Test	Test				
Dataset Z (7)	Train		Train		Train		Train	
Dataset S (8)				Train	Test	Test		
Dataset K (22)						Train	Test	Test
# Training set	7	81	7	52	7	92	7	78


A
Input (test): Nocodazole treated
Widefield
Sparse SMLM
MS-SSIM=0.6614
B
Test
Untreated
Train
M0
M4
Train
Untreated
7 FoVs
Lab Z
78 FoVs
Lab A, L, Z, S
C
MS-SSIM=0.7741
MS-SSIM=0.8359
(i)
(ii)
D
Output + Ground truth
Output + Ground truth
Intensity (Normalized)
E
M0 M1 M0 M2 M0 M3 M0 M4
Dataset A (19)
Dataset L (44)
Dataset Z (7)
Dataset S (8)
Dataset K (22)
Training set
7 81 7 52 7 92 7 78
F
MS-SSIM
p=1.33e-5
p=5.57e-5
p=1.04e-5
p=8.77e-4
n=19

Supplemental information

Sharing localization microscopy data makes deep learning-based image reconstruction more robust

Jiachuan Bai[1]*, Wei Ouyang[2]*, Manish Kumar Singh[1], Benoît Lelandais[1], Christophe Zimmer[1]

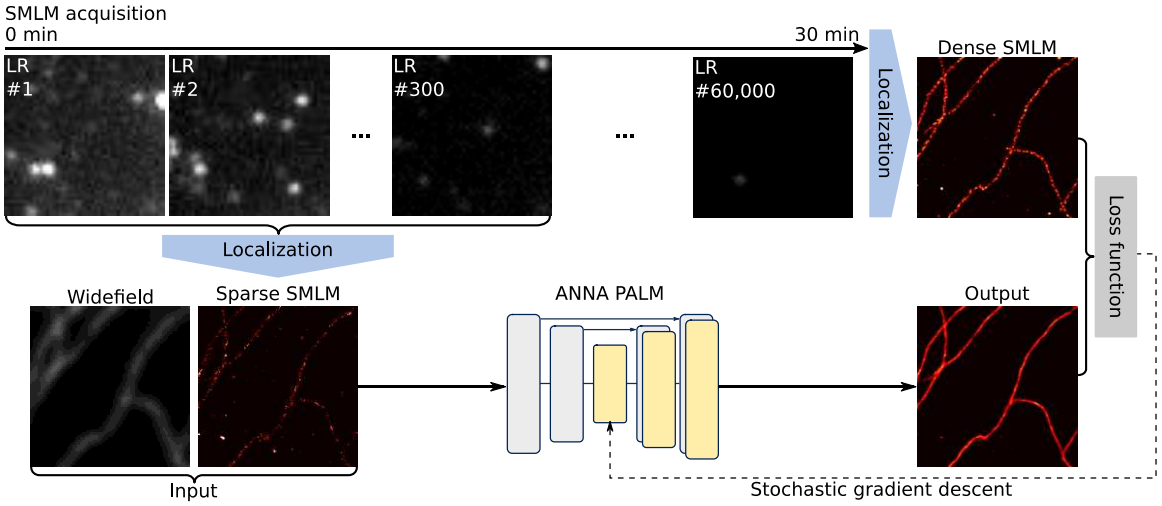


Figure S1: ANNA-PALM training strategy

This Figure is a simplified schematic of the strategy used to train ANNA-PALM. A standard ("dense") SMLM image is generated based on the localizations of single molecules computed from a large number of low resolution image frames (LR, frames 1 to 60,000). This dense SMLM image defines the ground truth for training ANNA-PALM. Localizations computed from a smaller number of low resolution frames (here, 300) are used to create a "sparse SMLM" image. Along with the corresponding widefield image (if available), the sparse SMLM image is fed as input to a U-net. The U-net is trained using stochastic gradient descent to output an approximation of the ground truth image, as defined by a loss function that measures the consistency of the output image with the ground truth. The U-net is trained in an adversarial manner against a conditional discriminator (not shown here) and the loss function also contains a term that measures the consistency of the generated output image with the widefield image (if available). For a more detailed description of ANNA-PALM, we refer to Figure 1 of Ouyang et al. Nat Biotech 2018.

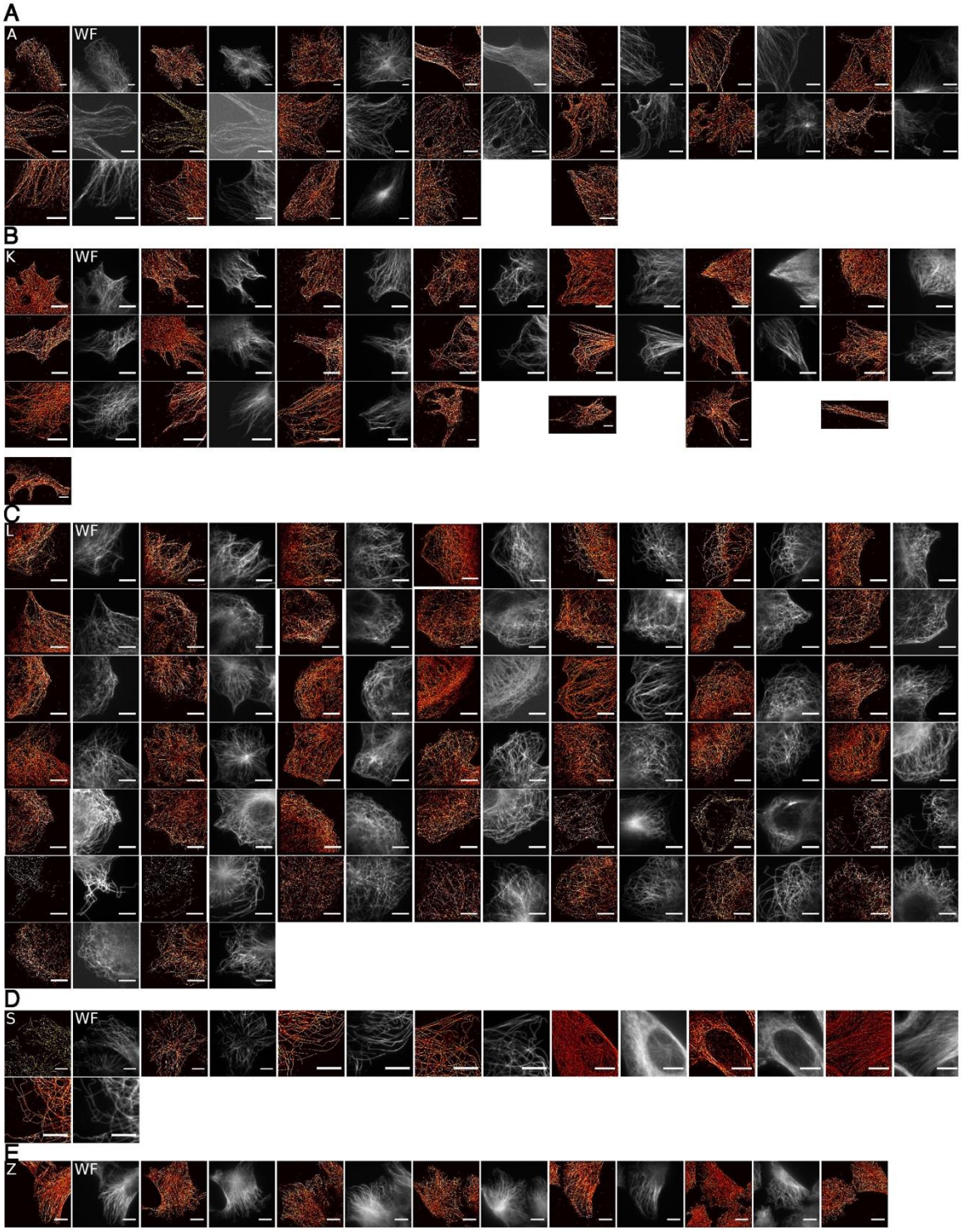


Figure S2: SMLM images of microtubules from multiple labs

This gallery shows SMLM images and corresponding widefield images (WF) of microtubules obtained from laboratories A (**A**), K (**B**), L (**C**), S (**D**) and Z (**E**). See Table 1 and Table S1. Scale bars are 10 μm.

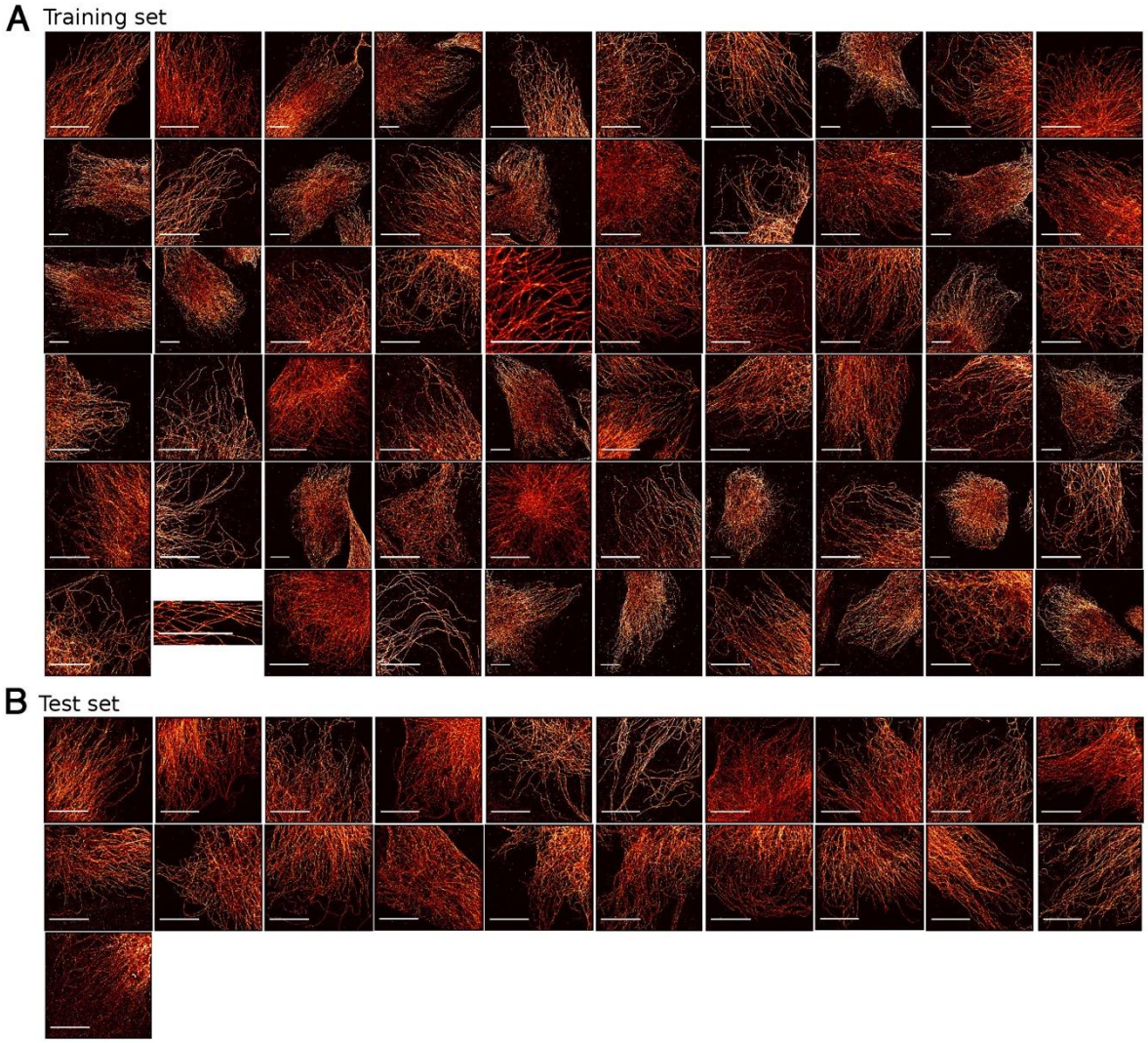


Figure S3: SMLM images of microtubules from lab Z

This gallery view shows SMLM images of microtubules acquired in lab Z for this study, distinguishing images used for training (**A**) and testing (**B**). Scale bars are 10 μm.

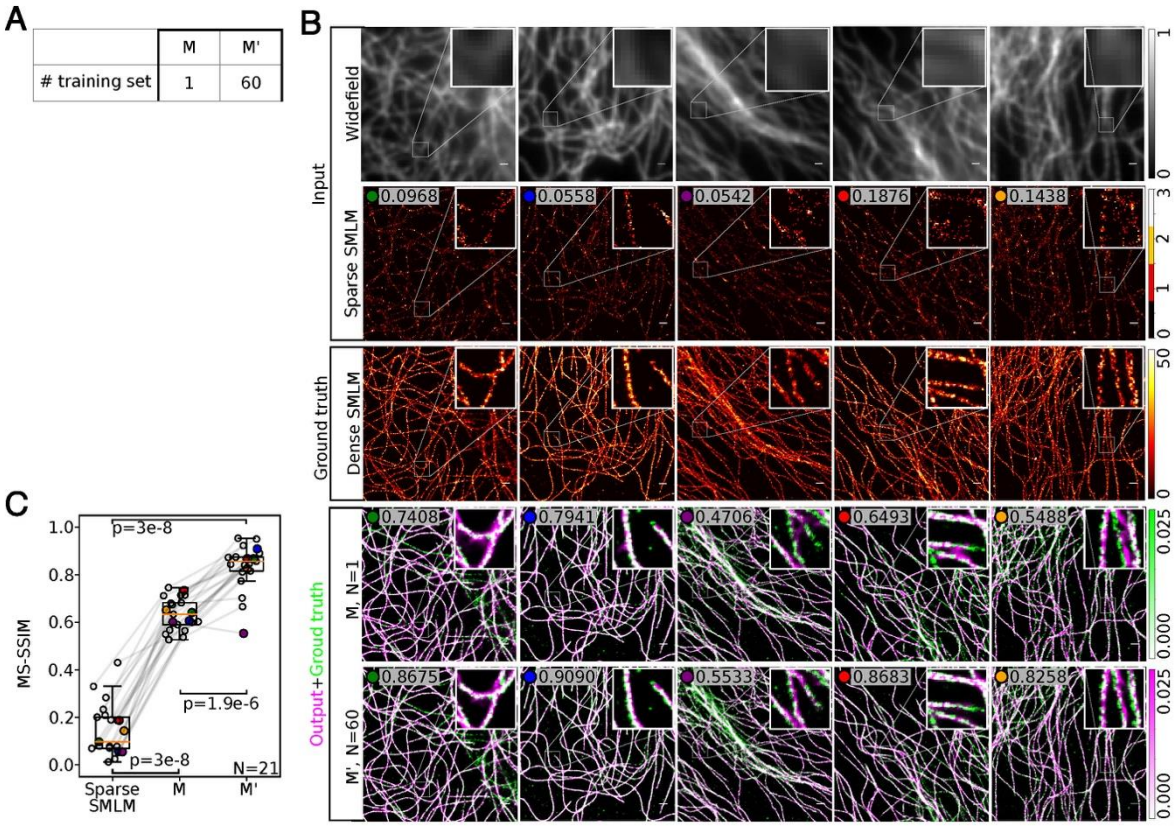


Figure S4: Training ANNAPALM on larger data sets improves reconstruction quality

(**A**) This Figure compares ANNAPALM reconstructions by models trained on only one (M) vs. 60 images (M').

(**B**) Examples of paired widefield and sparse SMLM images used as inputs to ANNA-PALM models M or M', along with the corresponding ground truth images (dense SMLM image obtained using all available frames). In the two bottom rows, the ANNA-PALM output image is shown in pink superposed to the ground-truth image in green. The MS-SSIM of the ANNA-PALM image relative to the ground truth image is indicated on the top left.

(**C**) Boxplots compare the MS-SSIM of ANNA-PALM reconstructions for models trained on a single image (M) or on 60 images (M'), or the unprocessed sparse input images (sparse SMLM) relative to the ground truth. This panel is identical to **Figure 1E,** except for the colored dots, which correspond to the images labeled with a dot of the same color in (**B**).

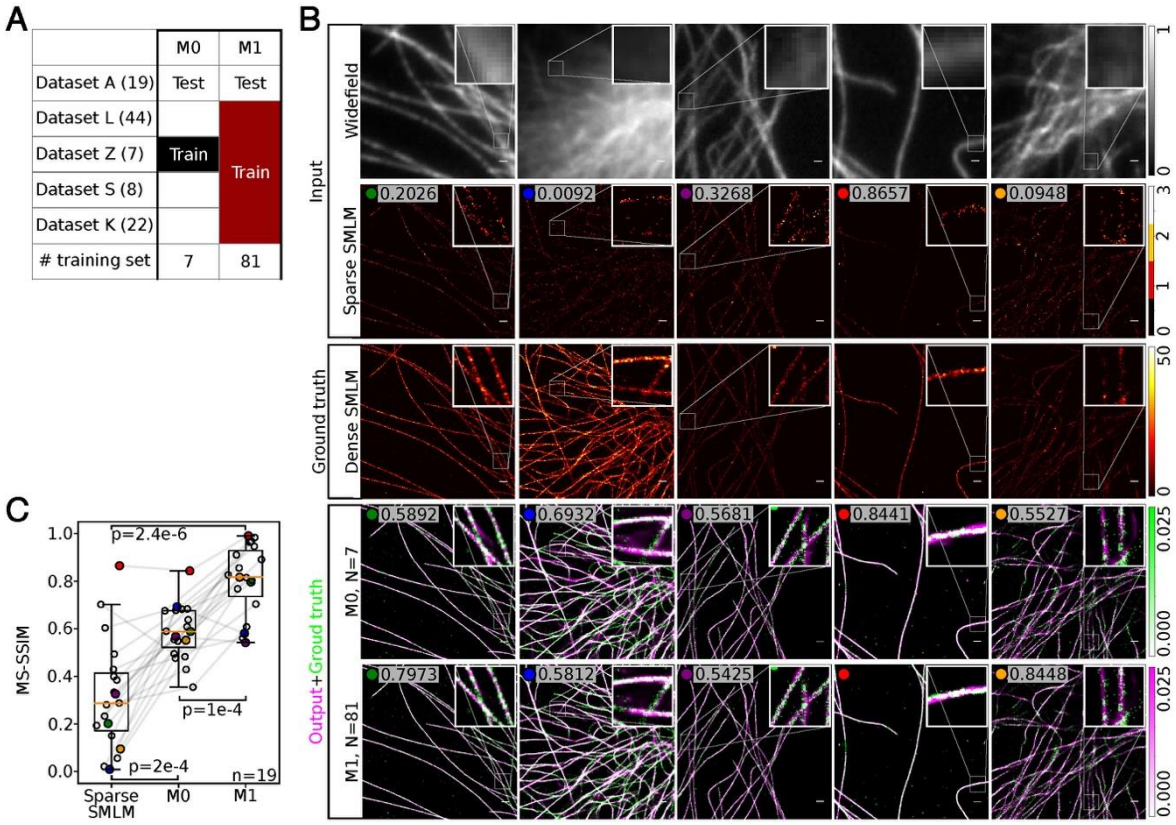


Figure S5: Training ANNA-PALM on ShareLoc data increases robustness to imaging protocols

(**A**) Overview of data sets used for training or testing models M0 and M1.

(**B**) Examples of paired widefield and sparse SMLM images used as inputs to ANNA-PALM models M0 or M1, along with the corresponding ground truth images (dense SMLM image, obtained using all available frames). In the two bottom rows, the ANNA-PALM output image is shown in pink superposed to the ground-truth image in green. The MS-SSIM of the ANNA-PALM image relative to the ground truth image is indicated on the top left.

(**C**) Boxplots compare the MS-SSIM of ANNA-PALM reconstructions using models M0 or M1 or the sparse SMLM input data, relative to the ground truth, as in **Figure 2F**. Each dot corresponds to a distinct image and colored dots correspond to the images labeled with a dot of the same color in (**B**).

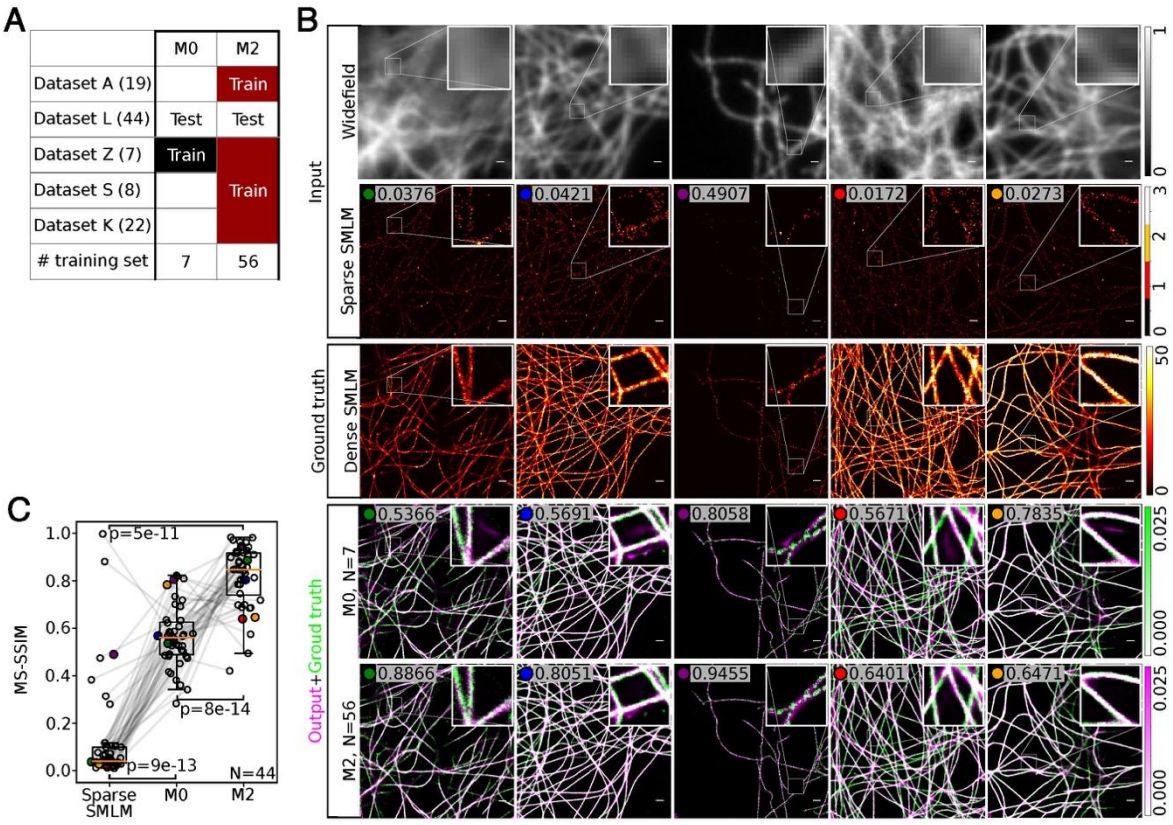


Figure S5 (continued): Same as above but for models M0 and M2.

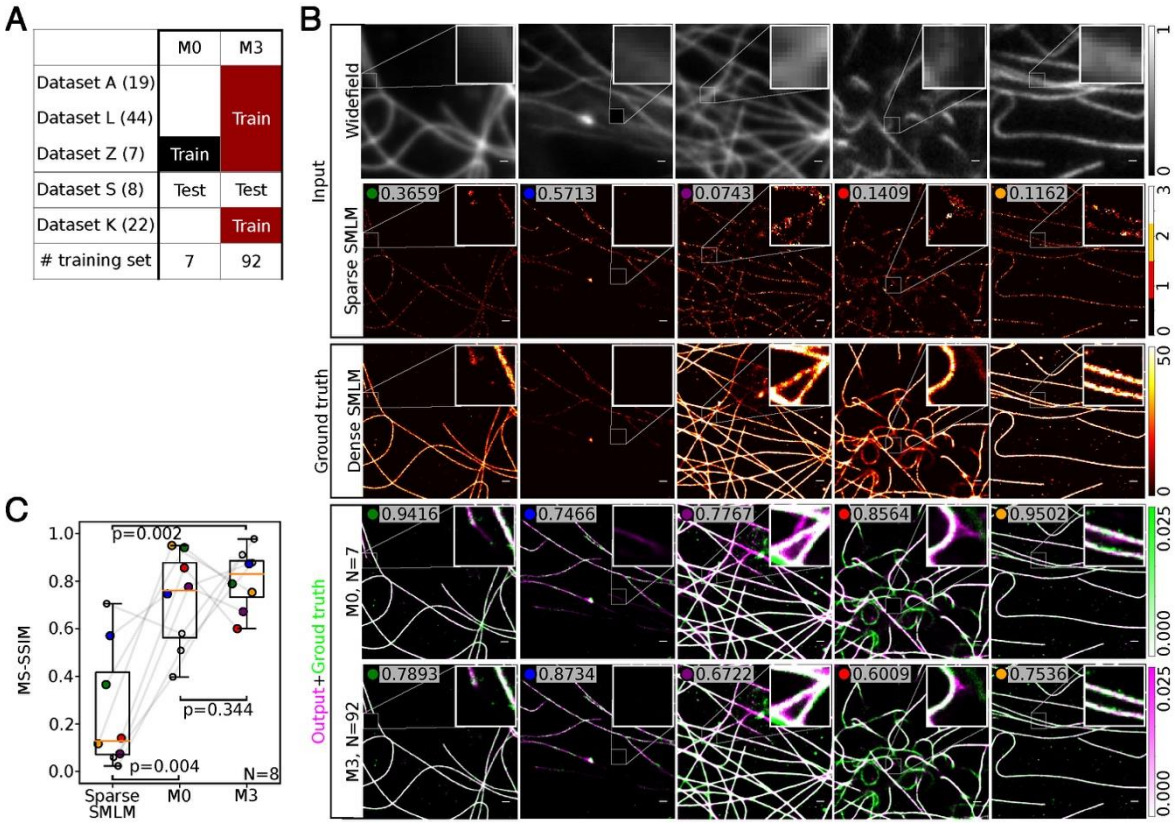


Figure S5 (continued): Same as above but for models M0 and M3.

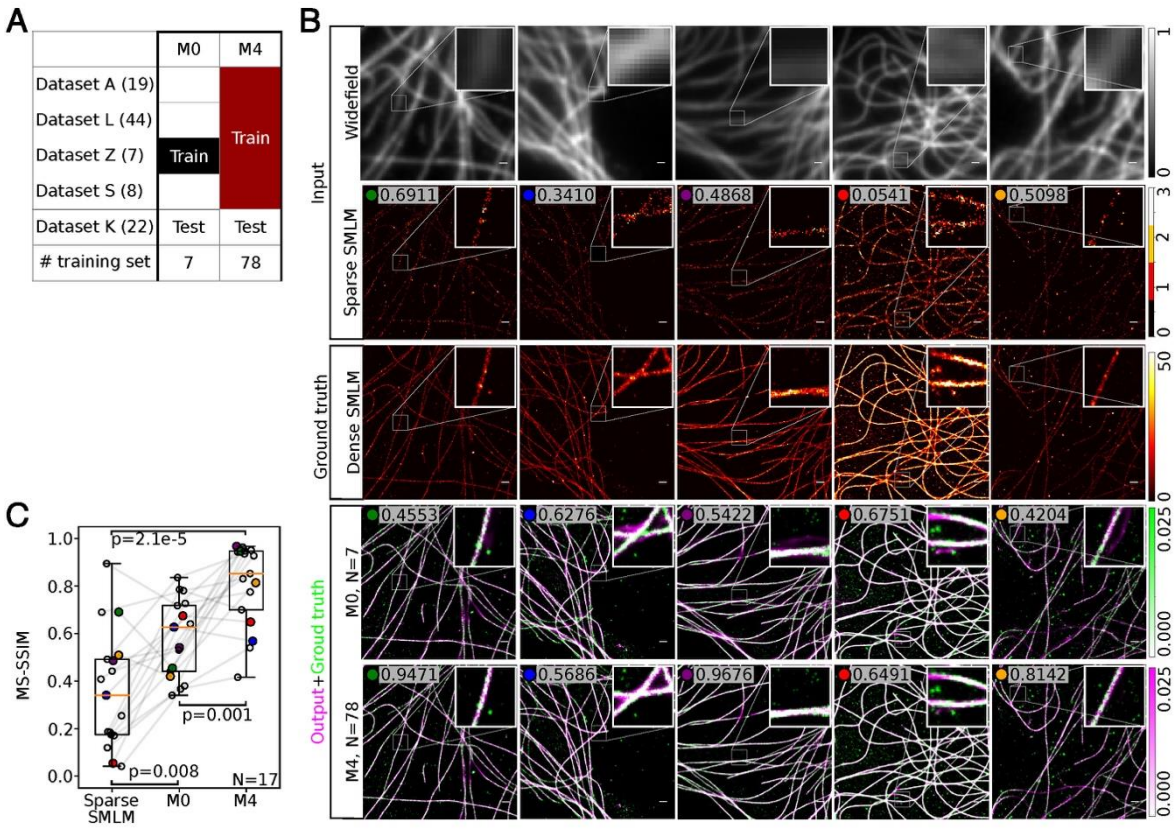


Figure S5 (continued) : Same as above but for models M0 and M4.

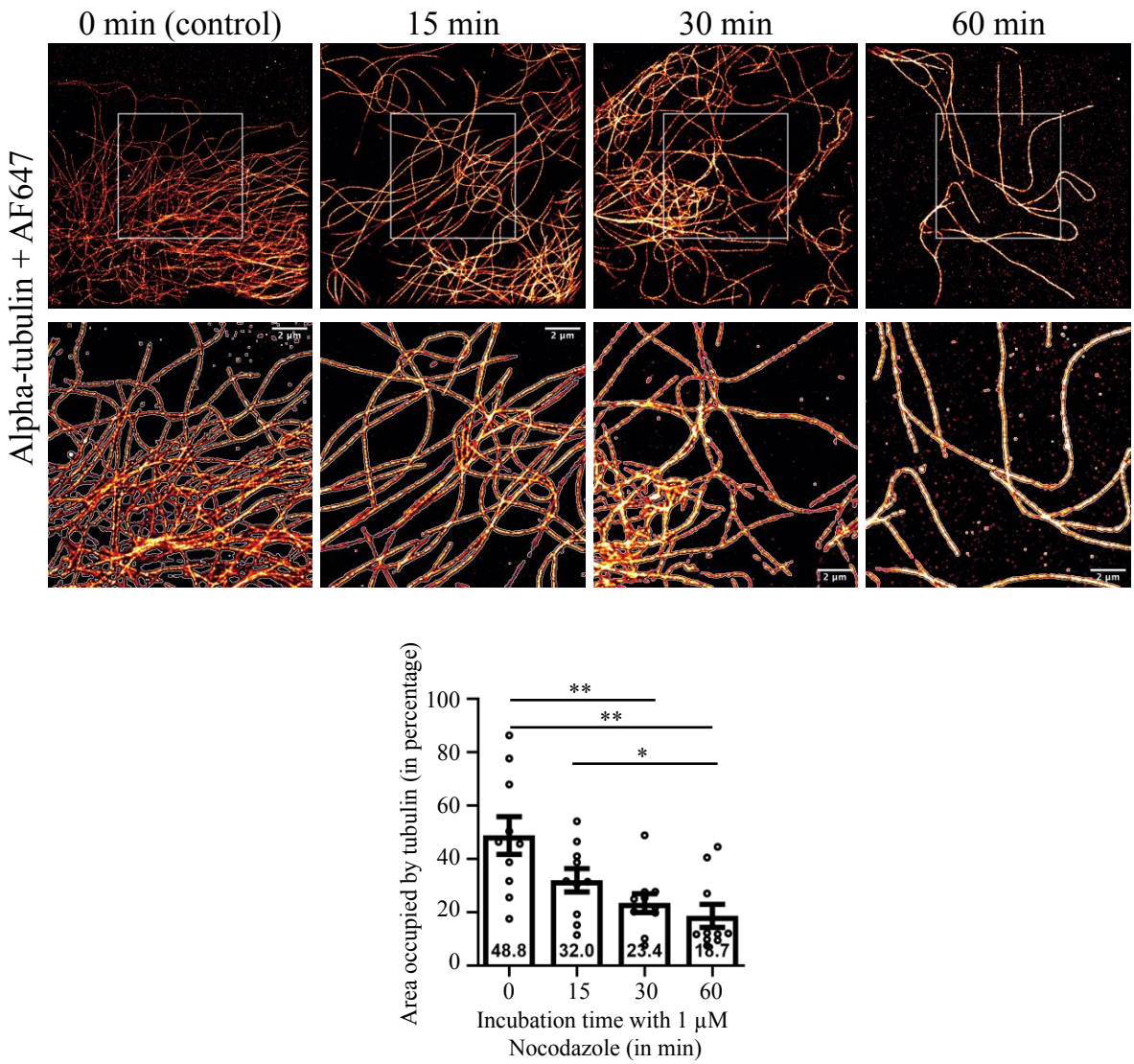


Figure S6: Nocodazole disrupts microtubule networks

Top row shows SMLM images of microtubules exposed to 1 µM of Nocodazole for durations of 15, 30 or 60 minutes and an untreated control. Bottom plot shows the fraction of image area occupied by microtubules for different Nocodazole treatment durations. This area was measured by applying the same Gaussian filter and threshold to all images and considering a region of interest in the same location. Each dot corresponds to a distinct image (n=10 images for each treatment duration). Bars show mean +/- standard error of the mean. Statistical tests are Mann-Whitney t-test with * indicating $p<0.05$, and ** indicating $p<0.01$.

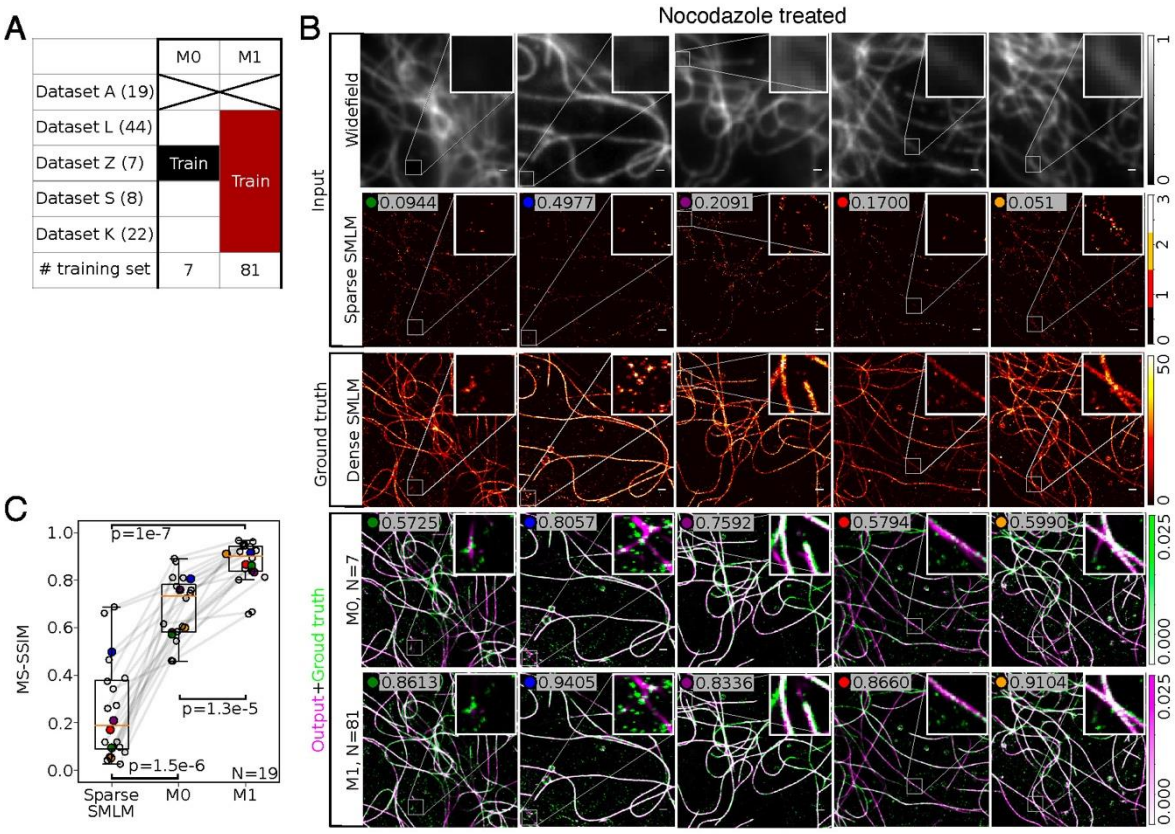


Figure S7: Training ANNA-PALM on ShareLoc data improves reconstruction robustness to experimental perturbations

(**A**) Overview of data sets used to train models M0 and M1.

(**B**) Examples of paired widefield and sparse SMLM images used as inputs to ANNA-PALM models M0 or M1, along with the ground truth images (dense SMLM image, obtained using all available frames). In the two bottom rows, the ANNA-PALM output image is shown in pink superposed to the ground-truth image in green. The MS-SSIM of the ANNA-PALM image relative to the ground truth image is indicated on the top left.

(**C**) Boxplots compare the MS-SSIM of ANNA-PALM reconstructions using models M0 or M1 or the sparse SMLM input data, relative to the ground truth, as in **Figure 2F**. Each dot corresponds to a distinct image and colored dots correspond to the images labeled with a dot of the same color in (**B**).

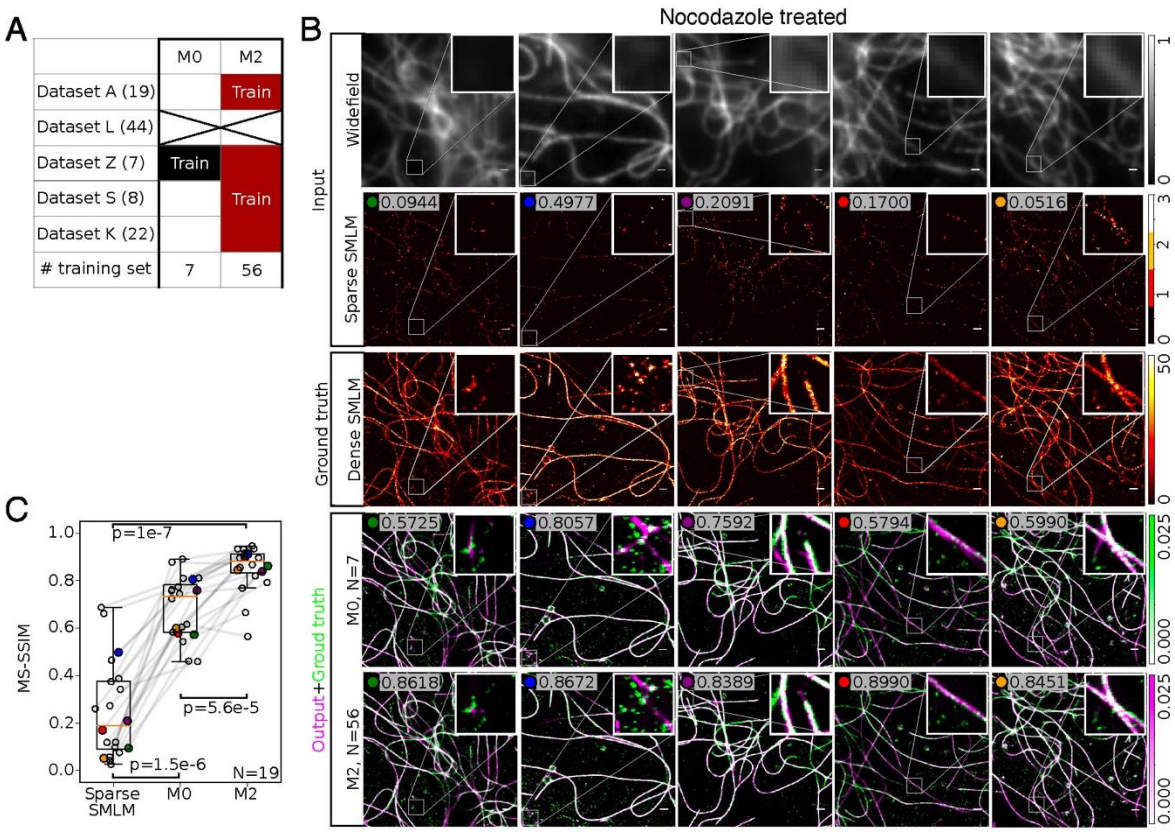


Figure S7 (continued): Same as above but for models M0 and M2.

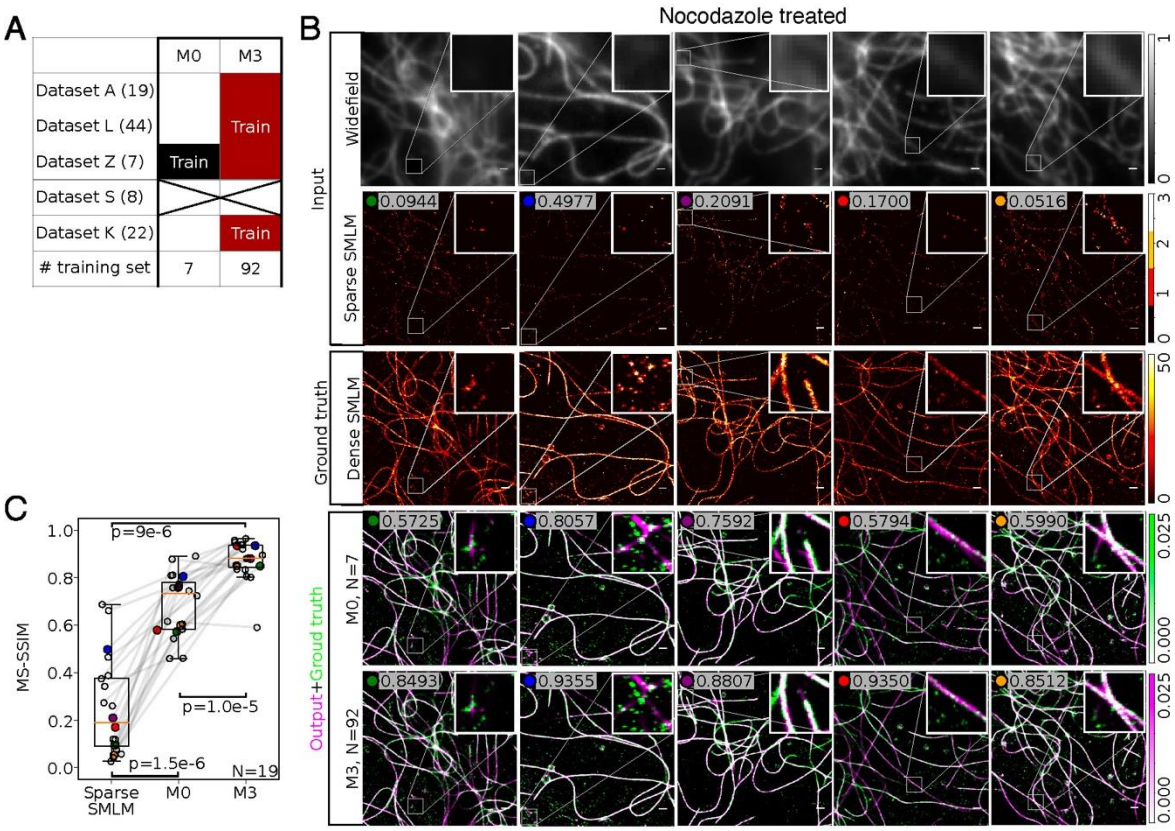


Figure S7 (continued): Same as above but for models M0 and M3.

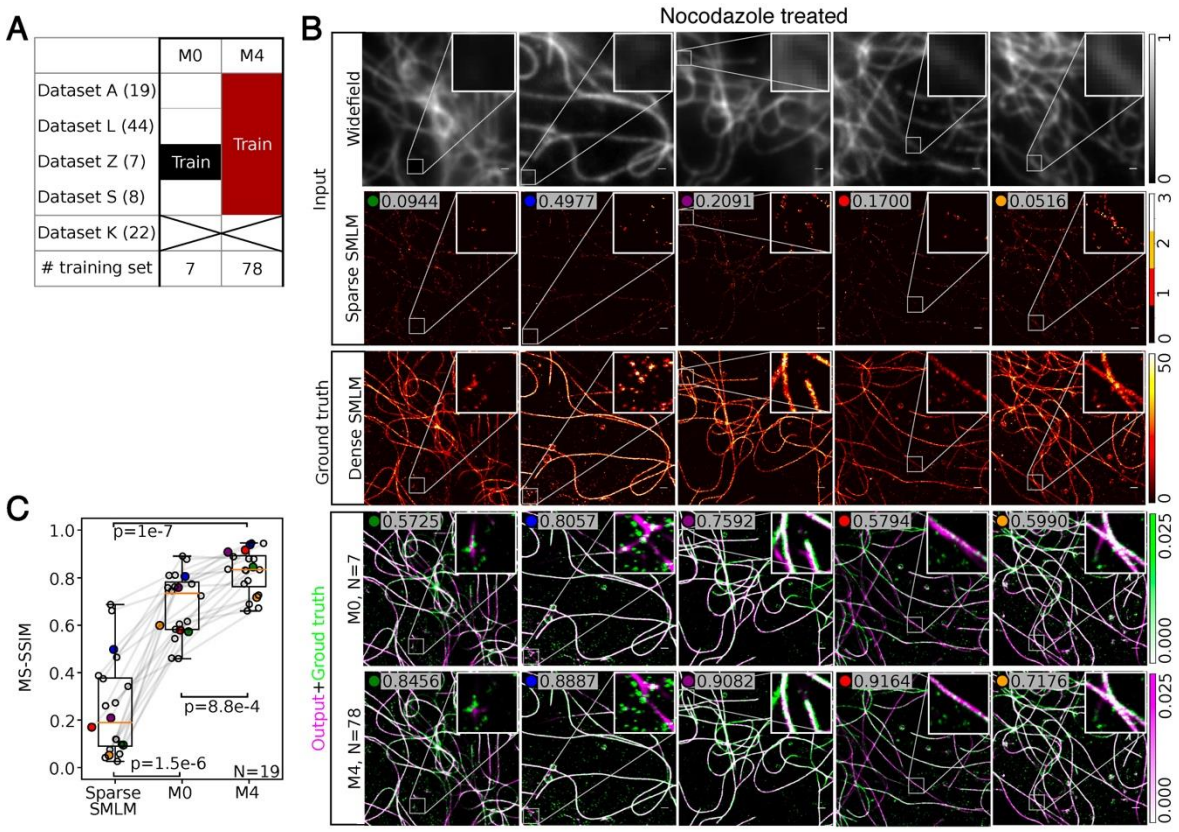


Figure S7 (continued): Same as above but for models M0 and M4.

II.3. Extending ANNA-PALM to reconstruct super-resolved structural dynamics

II.3.1. Background

As mentioned on page 21, SMLM acquisition relies on many low-resolution images, mostly limiting applications to fixed cells or structures undergoing slow dynamics. To reconstruct super-resolution movies of moving structures (e.g. microtubules) in live cells is very challenging.

As illustrated in **Figure 36**, to ensure high temporal resolution and avoid motion blur, each time point of the reconstructed super-resolution movie is defined from only a small number of consecutive low-resolution frames. This leads to a strong under-sampling of the structure by single molecule localization events, leading to what we refer to as a sparse SMLM movie. Each time point of reconstructed super-resolution movie is generated by ANNA-PALM using sparse SMLM movie as input.

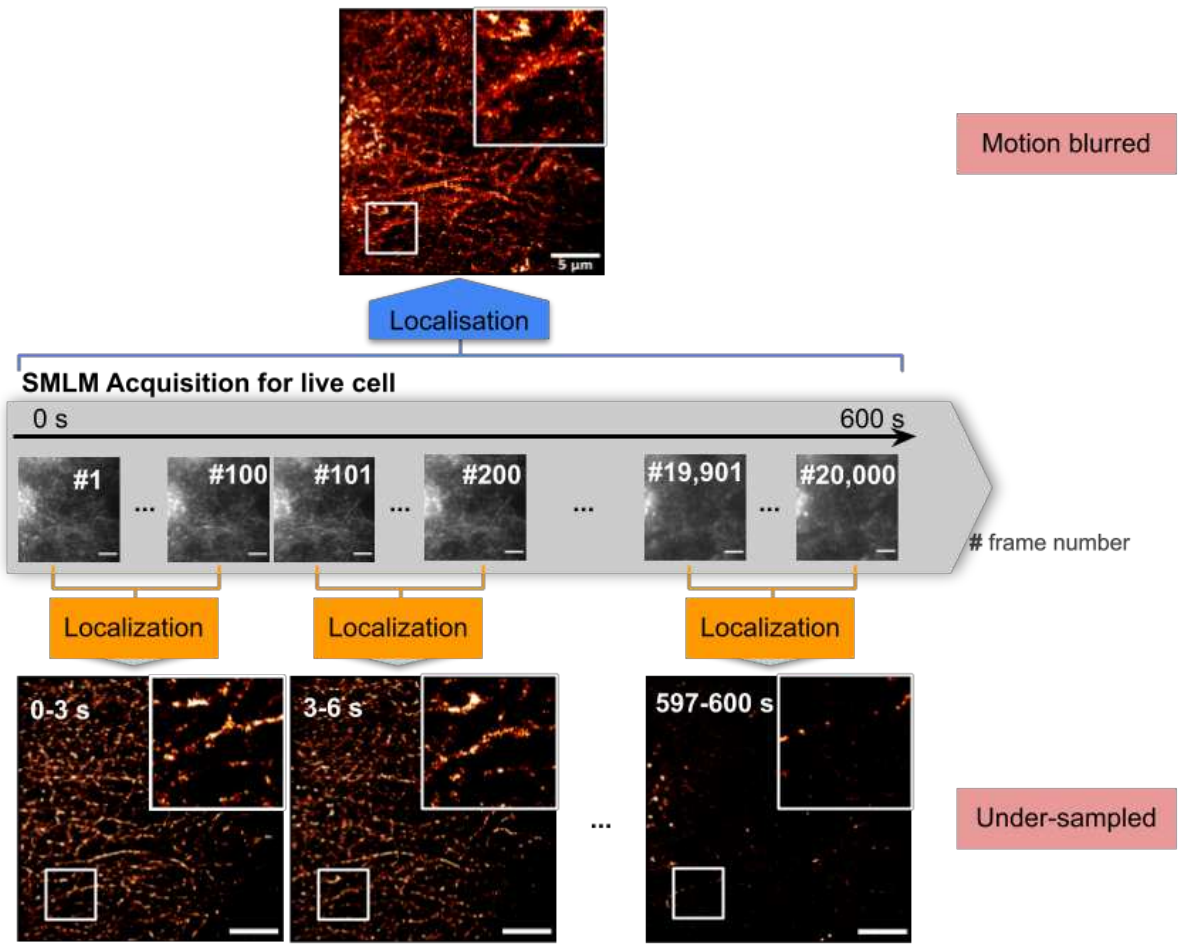


Figure 36. Trade-offs of SMLM imaging in live cells.

This figure illustrates the challenges of live cell SMLM imaging of microtubules. The blue path shows an image generated from the entire sequence of 20,000 low-resolution frames (gray panel). This image is blurred by the dynamics of the structures. By contrast, the sparse SMLM movie generated from the localizations extracted in a smaller number of consecutive low-resolution frames (orange path) is under-sampled and is generally not a super-resolution movie. Scale bar: 1μm. Images acquired by Manish Singh.

As previously discussed, ANNA-PALM is capable of reconstructing high quality super-resolution images from strongly under-sampled localization data. Initially I attempted to apply the original ANNA-PALM method to sequentially reconstruct each time point of the super-resolution movie based on the individual time point from the sparse, under-sampled movie. For this purpose, I analyzed live-cell SMLM data of microtubules stained with mEos3.2 in U2OS cell lines, as illustrated in **Figure 37**.

I first trained ANNA-PALM on sparse/dense SMLM images of fixed samples of these mEos3.2-labeled U2OS cells. For testing, I stacked consecutive sparse SMLM images (obtained from non-overlapping sets of 100 consecutive frames) together to form a movie (**Figure 37** middle panel). This stacked movie was then fed as an input to ANNA-PALM, where each sparse SMLM image was processed independently. A predicted super-resolution movie was then obtained from the consecutive output images.

However, the reconstructed movie was jittery, as evident from Video (available at: https://drive.google.com/file/d/1Xs8GIMJmKFJaXRX_yJnOB62rYzh4uAHo/view?usp=sharing). This is because the temporal sequence of image acquisition was not taken into account in the reconstruction process. Recognizing this limitation, I revised the ANNA-PALM architecture to harness and utilize this vital temporal information.

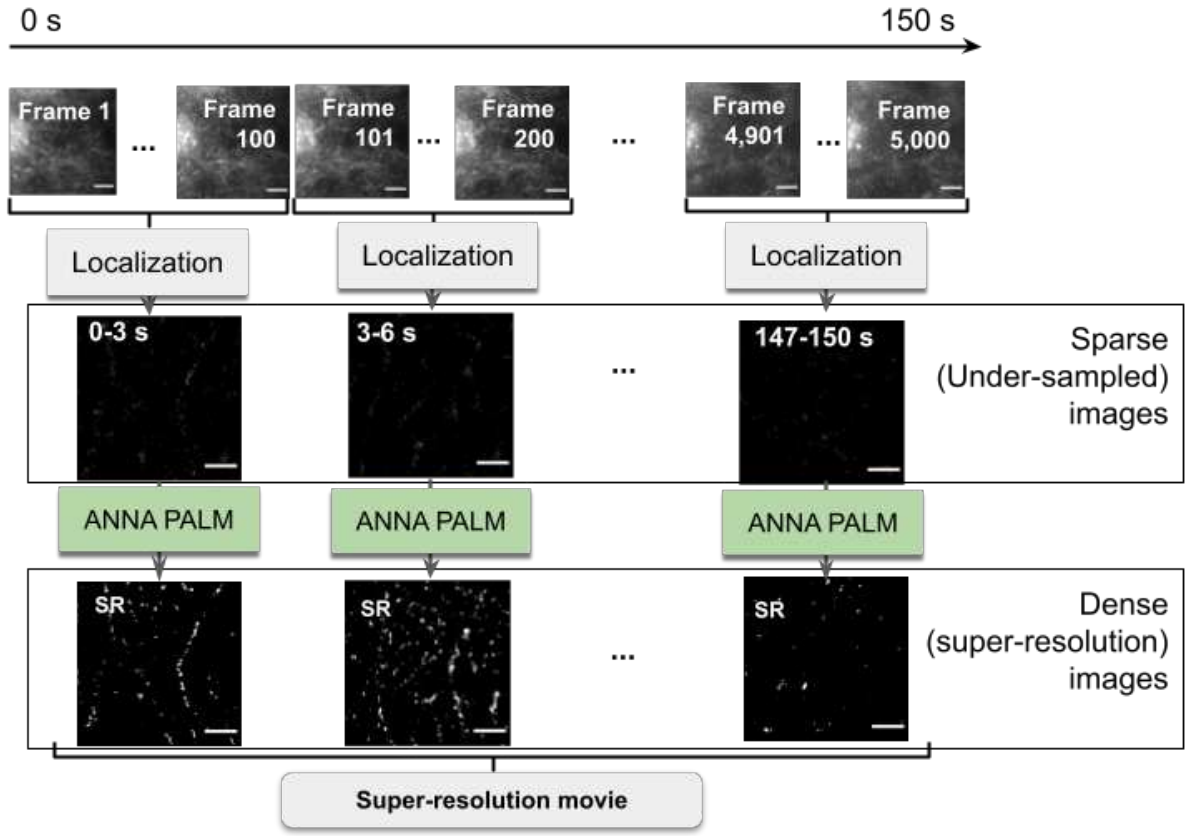


Figure 37. Using original ANNA-PALM to reconstruct live cell SMLM data yields suboptimal reconstructions.
The top panel shows a low-resolution single molecule sequence of live U2OS cells with microtubules labeled with mEos3.2. The entire acquisition consists of 5,000 frames with an exposure time of 30 ms, amounting to a total of 150 seconds of acquisition. In the middle panel, a sequence of sparse SMLM images is generated, where each sparse SMLM image is obtained from 100 consecutive low resolution frames, hence corresponding to a time interval of 3s. These sparse SMLM images are separately fed as input to ANNA-PALM (green), generating a super-resolved movie (bottom panel).

II.3.2. Exploiting temporal information using 3D CNNs

As discussed on page 31, 2D CNNs can capture the spatial features and are invariant to translation. 3D CNNs can be used to extract temporal information in movies by considering time as the third dimension (Tran *et al.*, 2015; Sakkos *et al.*, 2018). The 3D CNN can capture changes in pixel intensities across time.

Model structure

Based on the ANNA-PALM (2D) architecture (as discussed on page **43**), I implemented a model with an end-to-end 3D cGAN. This model is depicted in **Figure 38** and referred to as

"ANNA-PALM 3D". The generator is based on the U-net architecture, and the discriminator is a deep 3D convolution network.

The training dataset comprises movies with 128 time points, each consisting of a 128x128-pixel image. The kernel for each convolution layer has a size of 4x4x4 pixel, and it slides by 2 pixels each time (stride = 2). which effectively reduces the size of the subsequent feature maps by half for each deeper layer. To avoid overfitting, dropout is applied on three early deconvolution layers. With dropout, random subsets of neurons are dropped out at each iteration (Srivastava *et al.*, 2014). 3D ANNA-PALM architecture has 178,332,418 trainable parameters, a significant increase compared to the 44,599,042 parameters found in its 2D CNNs with using a 128x128-pixel image as input.

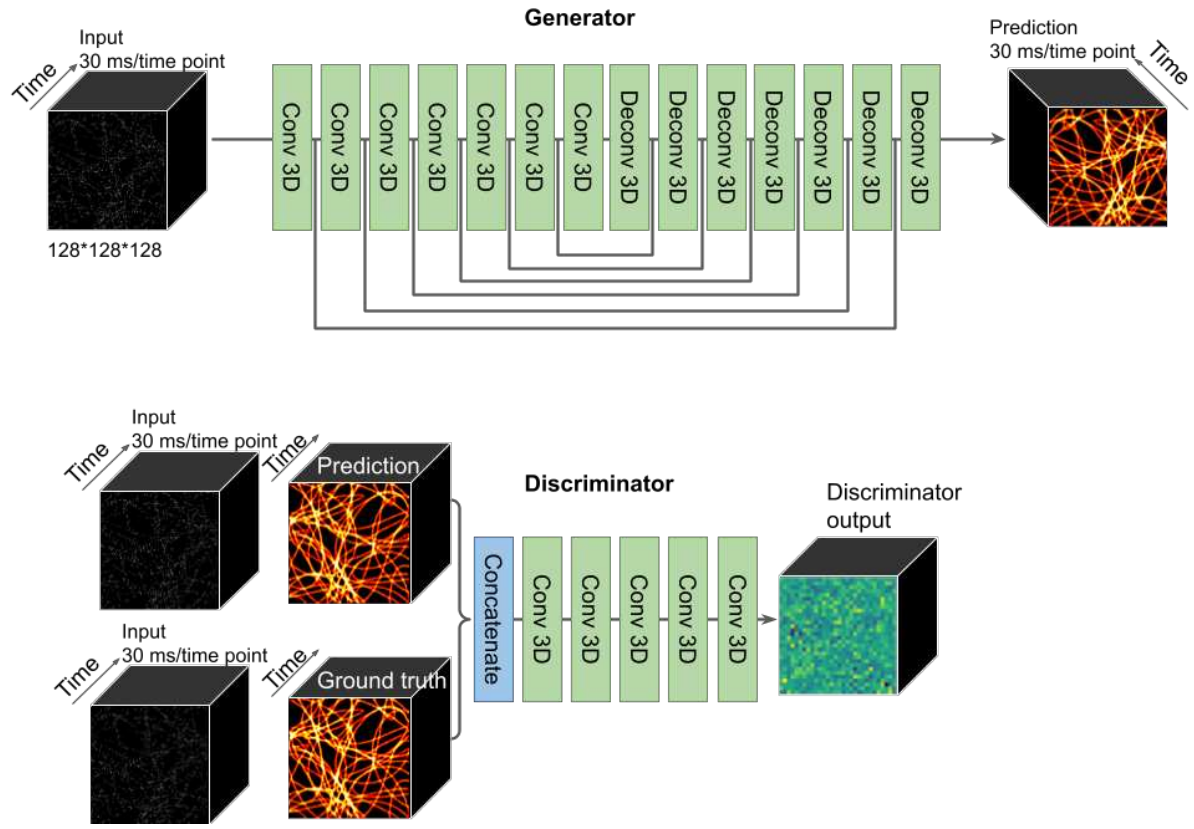


Figure 38. The architecture of ANNA-PALM 3D.
This figure displays the architecture of ANNA-PALM 3D. The generator (top) is a U-net-based 3D CNNs. The discriminator (bottom) is a 3D CNNs network. The kernel size for each convolution layer is 4x4x4 pixels with stride = 2.

The input of the generator is the sparse SMLM movie and its output is a reconstructed super-resolved movie. The input data is concatenated with output and ground truth dense SMLM movie and fed into the discriminator. The super-resolved movie reconstructed by this 3D ANNA-PALM model can be considered as the 3-dimensional interpretation from the sparse

SMLM movie where the first two dimensions are each time point of the SMLM movie and the third dimension is time (**Figure 39**).

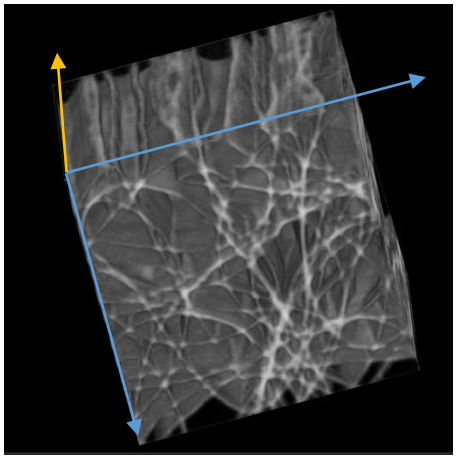


Figure 39. Illustration of SMLM movie considered as 3D representation.
This figure shows different angle of 3D representation of SMLM movie simulation through Napari, a python-based image viewer (Ahlers *et al.*, 2023). The yellow flesh represents temporal dimension, and the blue fleshes represent spatial dimensions.

Loss function

The loss function of the 3D ANNA-PALM is expressed as follows:

$$L = \beta\left(\alpha\left(1 - MS\_SSIM(< Y >, < \hat{Y} >)\right) + (1 - a)G * \left|< Y > - < \hat{Y} >\right|\right)$$

$$+\gamma(L_D + L_G)$$

where α, β and γ are hyperparameters that represent the weights of the different loss terms. Inspired by validated cGAN networks, such as 2D ANNA-PALM (Ouyang *et al.*, 2018) and pix2pix (Isola *et al.*, 2018), I initiated set the values of these hyperparameters as α = 0.84, β = 50 and γ = 1 during the training. Here, G denotes a Gaussian smoothing window, Y stands for the ground truth image and $\hat{Y}$ represents the prediction of 3D ANNA-PALM model. The term L_D and L_G refer to the generator and discriminator losses respectively, which are mentioned denoted as:

$$L_D = \mathbb{E}[logD(Y)] + \mathbb{E}\left[\log\left(1 - D\left(G(\hat{Y})\right)\right)\right]$$
$$L_G = \mathbb{E}\left[\log\left(D\left(G(\hat{Y})\right)\right)\right]$$

Where $D(Y)$is the discriminator's probability estimation of ground truth Y to be real, $G(\hat{Y})$ is the model's prediction and $D\left(G(\hat{Y})\right)$ is the discriminator's probability estimation of predicted data to be false for discriminator, real for generator (Goodfellow *et al.*, 2014). The training process involves the min-max optimization on page 33.

Training and testing strategies

Developing a training strategy for 3D ANNA-PALM is challenging, because unlike for fixed cells, a clear ground truth, i.e. a dense SMLM movie, is lacking. The absence of ground truth also makes it difficult to assess reconstruction quality. To address these challenges, I proposed a method to generate super-resolution SMLM movies for training and validation in a way that can capture complex structural dynamics.

The strategies consist in:

- Training phase:

  The model is trained on paired sparse and dense SMLM movies generated from either synthetic or experimental images in fixed cells. These movies undergo deformations over time to simulate structural dynamics by deforming and superposing fixed images (will be explained on page 123).

  To avoid possible confusions:

  - Dynamic data generated from synthetic static images will be called "**semi-simulated SMLM movie**"
  - Dynamic data generated from experimental static images will be called "**hybrid-simulation-experimental SMLM movie**".

- **Testing phase:**

  The trained model is tested on SMLM movies generated from synthetic or experimental images in live cells. These movies exhibit the structural dynamics that are more realistic/complex than simulated dynamic and hybrid-simulation-experimental SMLM movies.

To avoid possible confusions:

- Dynamic data generated from synthetic dynamic data will be called "**simulated dynamic SMLM movie**
- Movies generated from experimental live cell images will be called "**experimental SMLM movie**".

Consequently, I evaluated the 3D ANNA-PALM model's performance on the semi-simulated/experimental SMLM movies. The goal was to discern whether the model could accurately recover structural dynamics that is not present in the training dataset. Then I compared the performance of the 3D ANNA-PALM model to the original 2D ANNA-PALM model **Figure 37** for 2D model reconstruction.

II.3.3. Training data generation

Strategy to generate ground truth super-resolution movies

As mentioned above, the ground truth for both semi-simulated and hybrid-simulation-experimental SMLM movies are generated based on static real images, which are either static simulated images, or real images of fixed cells. As illustrated in **Figure 40**, I deformed, shifted and superposed two super-resolution images of microtubules in fixed cells. By overlaying these two images I created a complex dynamic of structures, in which the intersection of two filaments changes over time and filaments can pass each other.

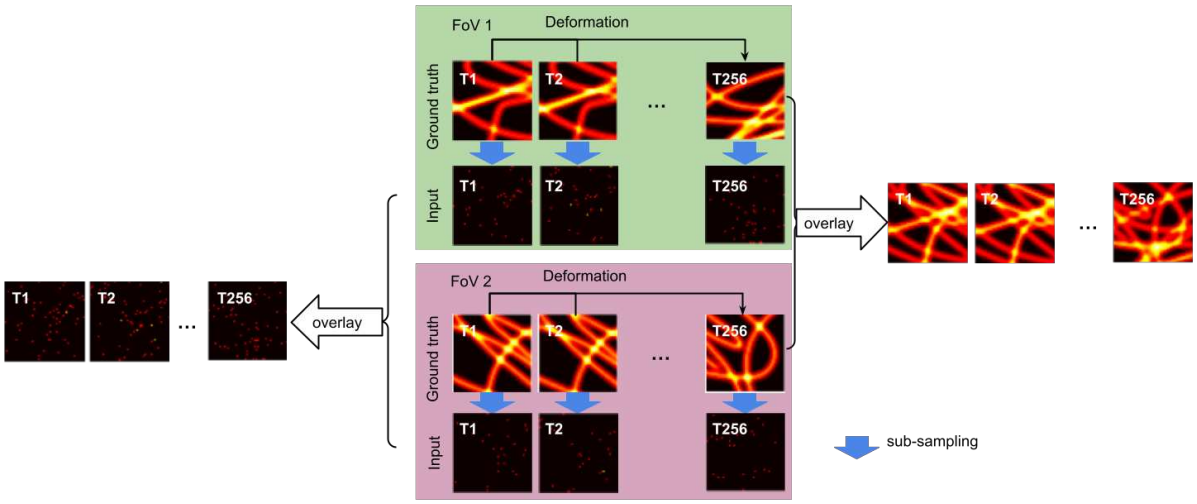


Figure 40. Illustration of training data simulation.

The ground truth SMLM movie (right) is generated by overlaying two fields of view (FoV), to which nonlinear deformations are applied over time. The input data (left) is generated by sub-sampling the ground truth images of each FoV. Each FoV measures 2.56µm *2.56µm.

The deformation applied on fixed cell images is based on a fourth-order polynomial function of the x and y coordinates of each pixel. As illustrated in **Figure 41**, for a pixel located at (x, y), the deformation is given by:

$$d_{x,y}(x,y) = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + c_5y + c_6y^2 + c_7y^3 + c_8y^4$$

where c_i are coefficients randomly generated following a uniform probability distribution, their cumulative absolute value remaining below a threshold α. For simulated dynamic movies of microtubules, I set $\alpha = 2$, because much smaller α results in a deformation resembling a simple translation and a much larger α causes the two FoVs to only overlap at the beginning of the movie.

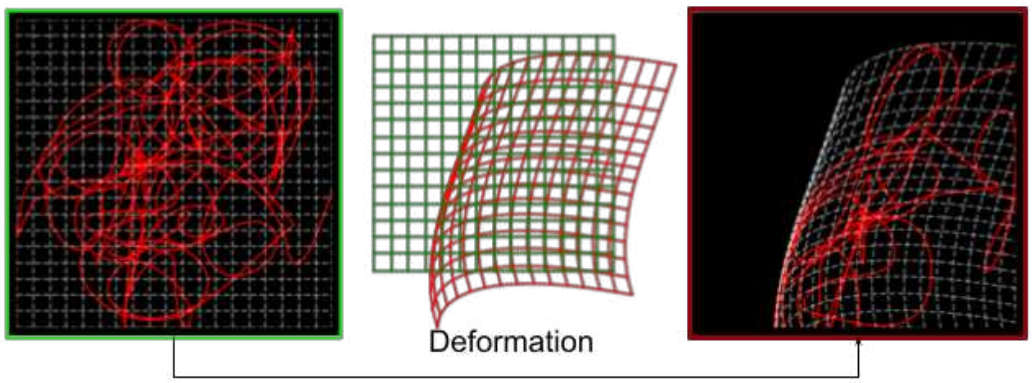


Figure 41. Illustration of 4th order deformation.
This figure illustrates a synthetic static microtubule image before (green) and after (red) 4th order polynomial deformation. The FoV measures 2.56µm *2.56µm.

For semi-simulated SMLM movies, the synthetic SMLM images of microtubules in fixed cells are generated using the same approach as in the original ANNA-PALM paper. As illustrated in Figure 42, the dense SMLM image is generated from a molecular dynamics simulation of polymers. Then a sparse synthetic SMLM image is generated by applying Poisson sampling to mimic the stochastic localizations of SMLM acquisitions (Ouyang *et al.*, 2018).

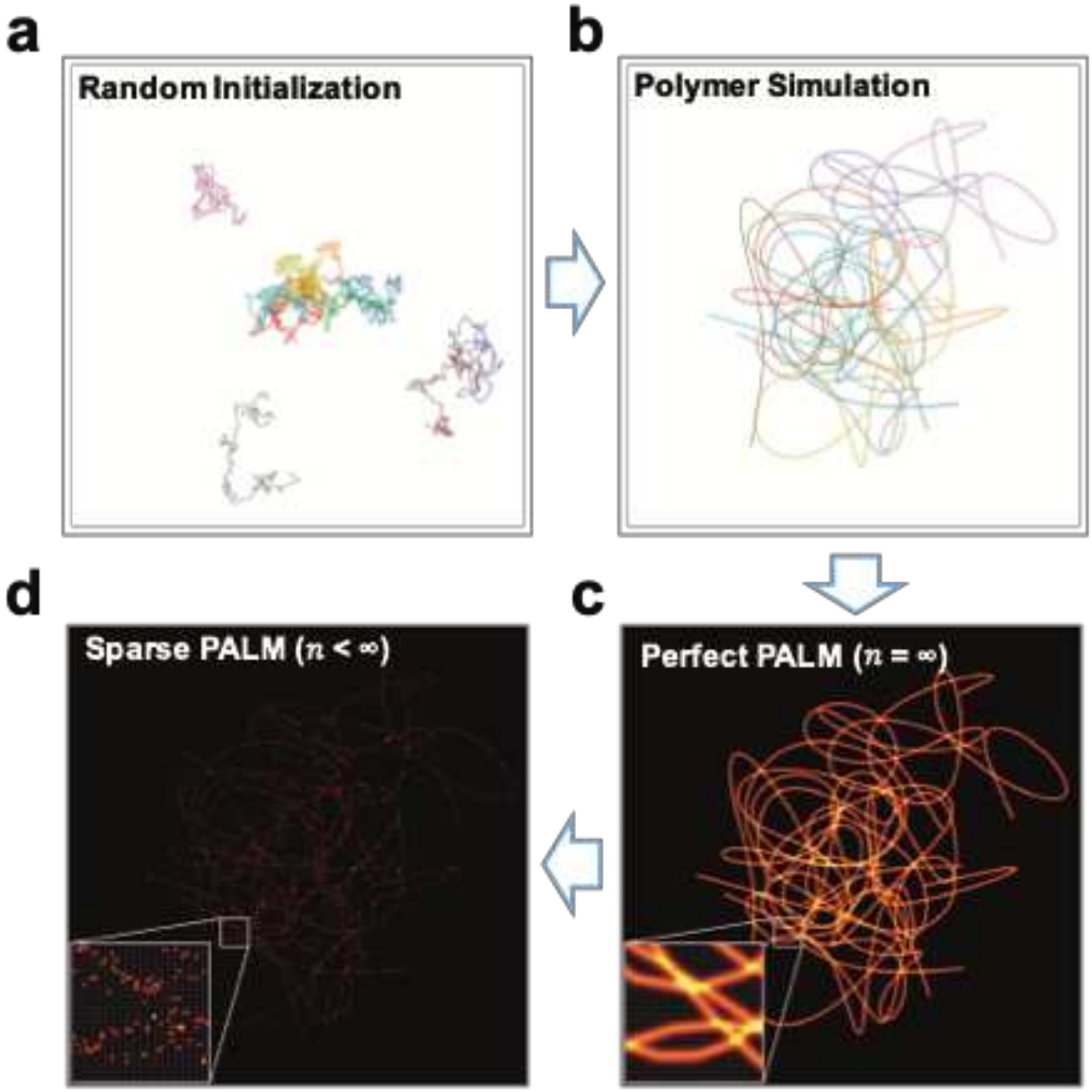


Figure 42. Synthetic dense and sparse SMLM images.
Synthetic SMLM images of microtubules are generated from simulated polymer chains. (a). Polymer chains are initialized randomly (each color represents a different polymer chain). (b). The stochastic temporal evolution of these chains leads to polymers adopting random semiflexible configurations. (c). The dense SMLM image is generated by convolving the zero-thickness curves in (b) with a Gaussian kernel of standard deviation 1.5 pixel. (d) Poisson sampling is applied to this probability density to generate a simulated sparse SMLM image. Images from (Ouyang *et al.*, 2018).

In total, I generated 274 semi-simulated SMLM movies as training data, examples of which are shown in **Figure 43**.

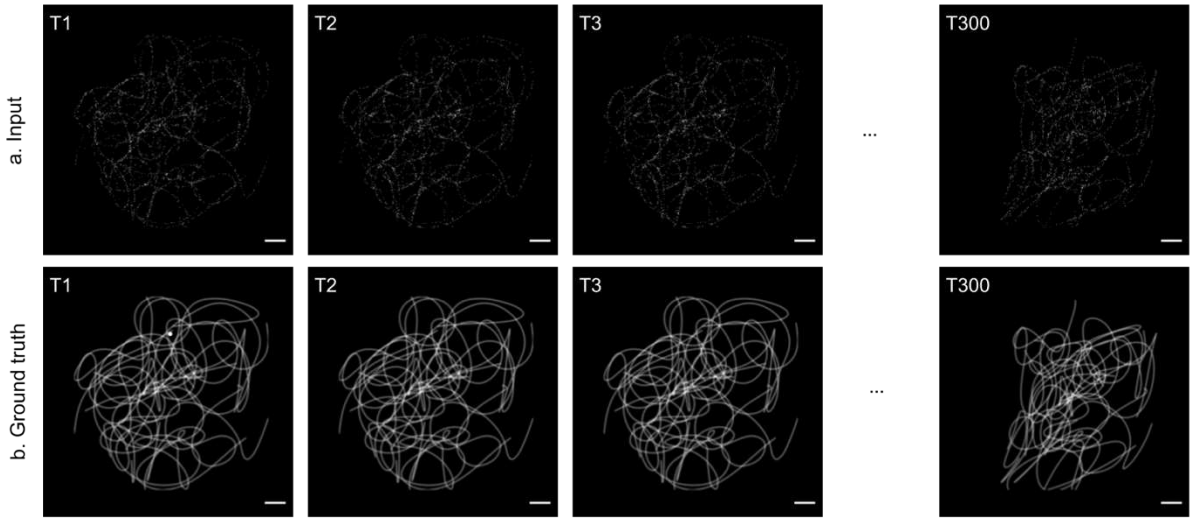


Figure 43. Illustration of training data for synthetic microtubules.
This figure shows a full dynamics simulation**:** a semi-simulated SMLM movie. The panel (a) displays the simulated sparse SMLM movie, and the panel (b) presents the simulated dense SMLM movie. The movie contains 300 time points.

Subsequently, I applied the same deformation and superposition approaches to the experimental data in fixed cells to generate hybrid-simulation-experimental SMLM movies. By leveraging the ShareLoc platform, I was able to generate a total of 79 SMLM movies. An example of training data is illustrated **Figure 44**.

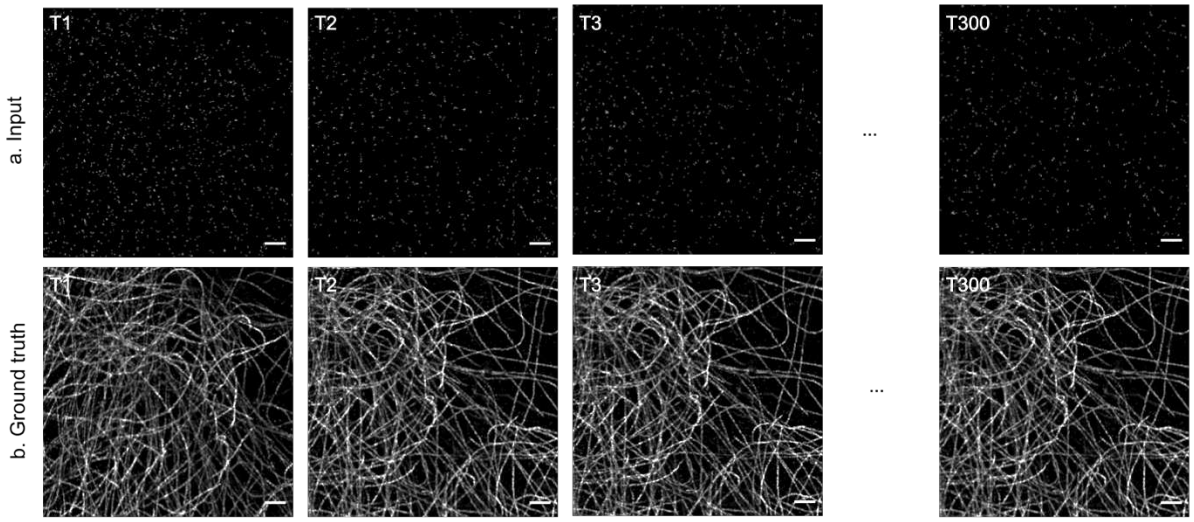


Figure 44. Illustration of training data for experimental data.
This figure shows a full dynamics simulation**:** a hybrid-simulation-experimental SMLM movie. The panel (a) displays the simulated sparse SMLM movie, and the panel (b) presents the simulated dense SMLM movie. The movie contains 300 time points. The experimental image is microtubules stained with “Alexa 647” in fixed cells.

Furthermore, the implementation of 3D ANNA-PALM involves on-the-fly data augmentation for the training data, meaning that data augmentation is performed during the training process. Specifically, this augmentation includes image rotation and random cropping at each training iteration.

Testing data generation

The testing data generation consists in 1) simulated dynamic generated from the polymer simulation with stochastic temporal evolution (Figure 42**.c**), and 2) experimental SMLM movies generated from real live-cell SMLM acquisition.

For the simulated dynamic SMLM movies, the structural dynamics are more complex than in the semi-simulated movies. This is because in the molecular dynamics simulation, each filament has its own stochastic movement. Different parts of the same polymer can move in different directions, therefore there is relatively little correlations between movements, whereas in the movies generated by deformations and superposition of two static images there are more correlations. In the molecular dynamics, the motions are more individual, whereas in the superposition movies the movements are more collective.

Regarding the experimental SMLM movies, each time point is generated using X frames of SMLM acquisition in live cells. In our current experimental setup, the value of X can be 30, 100, or 500 (**Figure 37** top and middle rows).

II.3.4. Results for synthetic microtubules

I first trained and evaluated 3D ANNA-PALM model on semi-simulated SMLM movies of microtubules and then tested the model on the simulated dynamic SMLM movies, benchmarking the testing results against the original 2D ANNA-PALM model. For visual inspection of reconstruction quality, I generated kymographs. By plotting pixel intensities along a defined line across multiple frames, kymograph illustrates the structural changes over time.

Figure 45 presents the results of the comparisons using kymographs. Through this tool, we can observe the capabilities of the 3D ANNA-PALM model in comparison to the original 2D ANNA-PALM model. The continuous lines on the kymographs, represent the dynamic of microtubules. When there is a clear, continuous line in the kymograph, it indicates a successful

recovery of the microtubule dynamics, because the simulated motions are smooth and did not involve jumps.

As shown in

Figure 45.a, the input data is sparse and does not reveal the dynamics of microtubules. This makes it challenging to understand the structural dynamics of the microtubules in live cells.

As shown in **Figure 45.c, d**, the 3D ANNA-PALM model consistently recovers the microtubule structure at each time point, while the 2D model reconstructions exhibit artifacts. This is because 3D ANNA-PALM extracts the spatial-temporal features. The kymographs confirm the enhancement of performance by incorporating temporal information and are relevant to the results of my first experiment that mentioned on page 119.

As shown in **Figure 45.c**, the super-resolution movie reconstructed by 2D ANNA-PALM model is jittery. Not only the reconstructed microtubules appear distorted, not consistent their characteristic shape, but there are also artifacts present in the reconstructions.

As shown in **Figure 45.d**, the 3D ANNA-PALM model outputs a more refined super-resolution movie. At each time point, the motions in these reconstructions are considerably smoother, and the shapes of the microtubules are generally well reconstructed

The kymographs further confirm the value of incorporating temporal information, which aligns with the observations made in my first experiment detailed on page 119.

However, while the 3D model offers significant improvements over the 2D reconstructions, it is not flawless. The model could be refined in the future, and this will be detailed on page 142.

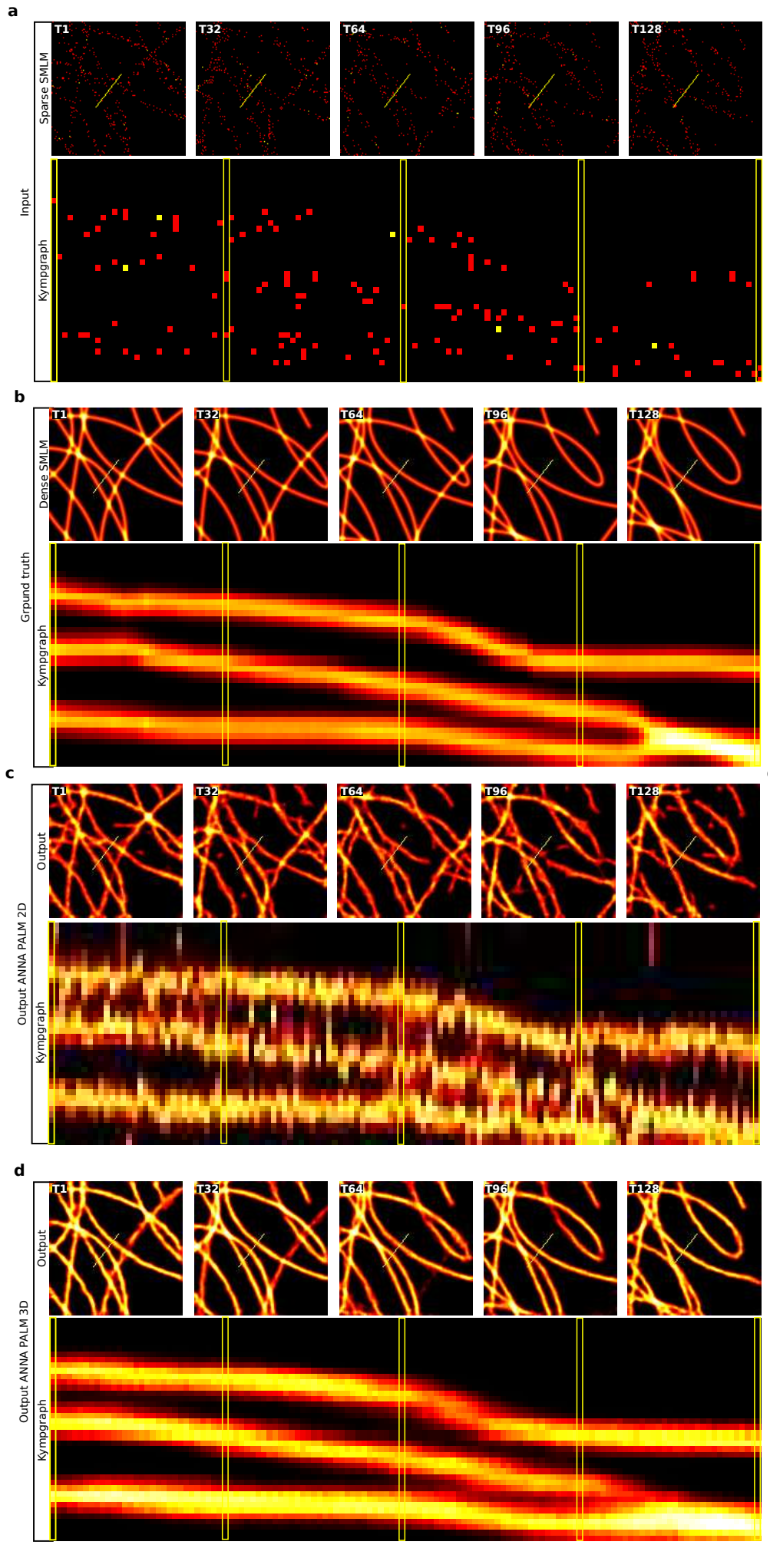

a
T1
T32
T64
T96
T128
Sparse SMLM
Input
Kympgraph
b
Dense SMLM
Grpund truth
c
Output
Output ANNA PALM 2D
d
Output ANNA PALM 3D

Figure 45. Super-resolution movie reconstruction using 2D vs 3D ANNA-PALM
Each panel of this figure shows 5 time points of the data, and the kymograph along the yellow line. (a) The input of the models. (b) The ground truth. (c) The reconstruction of the original ANNA-PALM. (d) the reconstruction of 3D ANNA-PALM. Each FoV measures 2.56µm *2.56µm.

II.3.5. Quantitative analyses

We expect to reconstruction quality as function of localization rate and the velocity could be a valuable insight for the super-resolved movies reconstructed by 3D ANNA-PALM. However, in practice, we need information about the velocity and localization rate of experimental SMLM movie in live cells.

Estimation of the localization from experimental data

I computed the localization rate by counting localization number in each time point of the SMLM experimental movie. Nevertheless, determining the velocity is challenging due to the absence of a super-resolution movie. The average localization rate ranges between 110Hz to 380Hz per FoV of 2.56µm*2;56µm.

Estimation of the velocity from experimental data

I came up with a method to approximate the velocity. We acquire one widefield image after every X low-resolution SMLM frames (**Figure 46**). For instance, if X equals 100 frames and the exposure time for each low-resolution frame is 30ms, we can obtain one wide-field image every 3 seconds. We can then use this widefield movie to make an approximate estimation of the velocity. I manually measured the velocity of microtubules in the experimental data. In the widefield movies, I plotted the intensity profile along a line perpendicular to the fast-moving microtubules in each image. I then measured the displacement of the peak of the profile intensity over time. Given the time interval between two images, the velocity could be calculated. The maximum measured velocity ranged from approximately 60 to 90 nm/s.

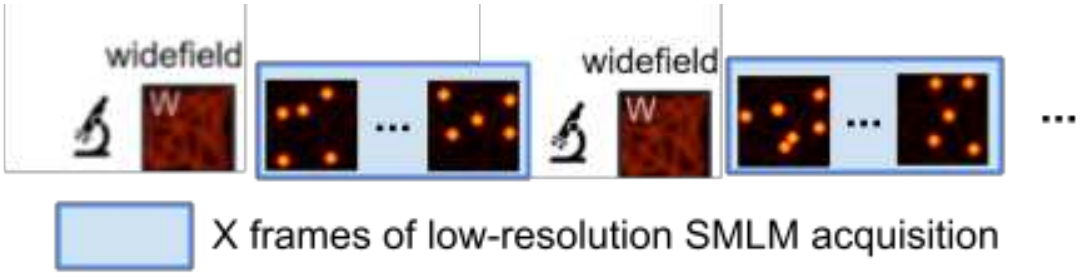


Figure 46. Widefield movie acquisition.

This figure illustrates the SMLM imaging in live cells where acquisition of widefield images alternates with acquisition of a number X of single molecule images.

Based on this information, I varied the average velocity of synthetic microtubules in the testing data from 0 to 220 nm/s and the average rate of single molecule localization from 30 to 1 600 Hz per FoV of 2.56µm*2;56µm. I then quantitatively assessed the reconstruction quality as function of those two factors, using the MS-SSIM metric.

As shown in **Figure 47.b**, the reconstruction quality depends on both the localisation rate and the velocity. Over all, the reconstructed super-resolved SMLM movies are in much better agreement with the ground truth than the sparse SMLM movies. When the localization rate increases, each time point of the sparse SMLM movie is better sampled, leading to improved reconstruction quality. Conversely, when the velocity increases, the temporal information across consecutive frames in the input movie becomes more ambiguous, resulting in decreased reconstruction quality. According to MS-SSIM, for realistic parameter ranges 3D ANNA-PALM 3D produces reconstructions can be consider as good quality.

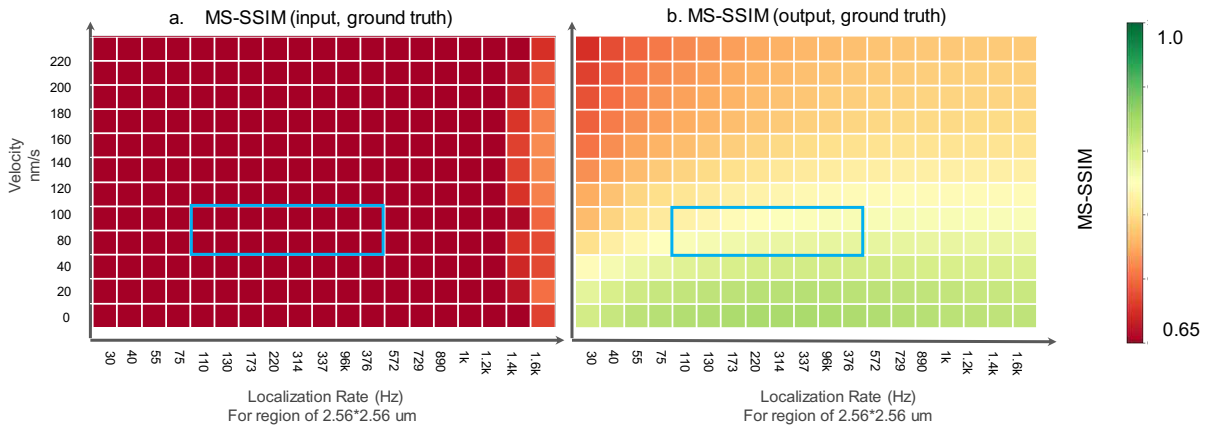


Figure 47. Quantifying ANNAPALM 3D reconstruction quality as function of velocity and localization rate.

In this figure, the x-axis represents the localization rate (in Hz per 2.56µm*2;56µm), while the y-axis represents the velocity of the moving structure (in nm/s). The color scheme of the graph indicates MS-SSIM values, where green represents good reconstruction quality, and red indicates poor reconstruction

quality. Panel a shows the MS-SSIM of sparse movies and ground truth movies, panel b refers to the MS-SSIM between super-resolution movies reconstructed by 3D ANNA-PALM and the ground truth. The velocity and localization rate of experimental data in live cells are in the rectangular region on the figure.

II.3.6. Preliminary results for experimental data of microtubules

This section presents the preliminary results of experimental SMLM movies of microtubules reconstructed by 3D ANNA-PALM model. As mentioned above, the training data consisted in hybrid-simulation-experimental sparse and dense SMLM movies (see **Figure 44**) and the testing data consisted in experimental SMLM movies. Each time point of the hybrid-simulation-experimental and experimental movies is generated from 50 frames of low-resolution frames from SMLM acquisition, corresponding to 1.5 seconds.

However, while testing the trained 3D ANNA-PALM model on the experimental data, the results proved sub-optimal (see **Figure 48**). The yellow arrows in the figure indicate the hallucination of the reconstruction.

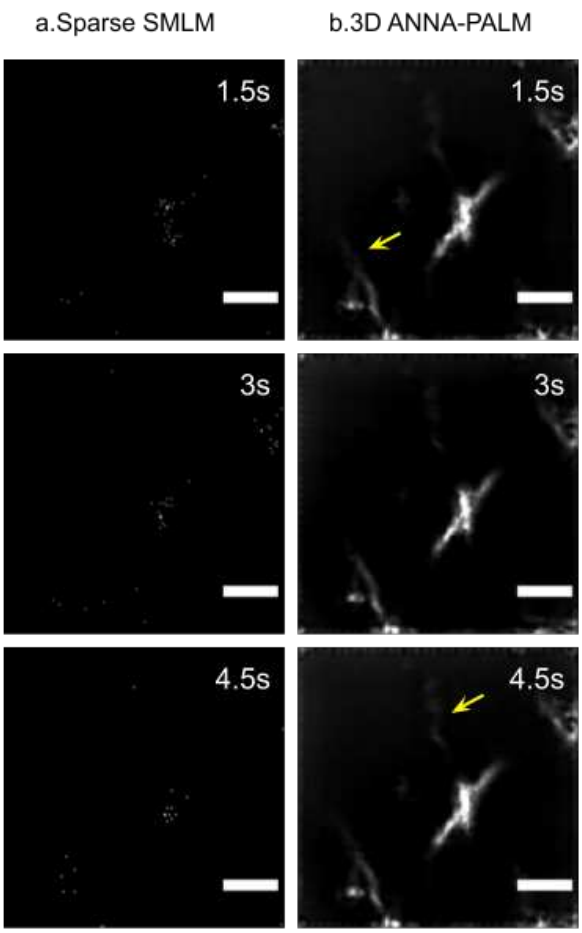


Figure 48. Preliminary results of live cell SMLM reconstruction using 3D ANNA-PALM
(a) and (c) show time points of input sparse SMLM movie, while (b) and (d) display the 3D ANNA-PALM reconstruction. Each time point corresponds to 1.5s. Scale bar: 0.5 μm.

Considering the quantitative results on simulations shown above (page 130), this poor reconstruction quality is most likely due to the non-consistency of velocity and localization rate between training and testing data: the model has been optimized for the conditions it was trained on, thus, the trained model might struggle to accurately reconstruct movies. More work is needed to diagnose the reasons and improve the performance of 3D ANNA-PALM on experimental data. We will discuss this in more detail on page 142.

III. Discussion

My Ph.D. research focuses on deep learning augmented SMLM reconstruction, aiming to improve the robustness of ANNA-PALM and extending it to reconstruct super-resolution movies in live cells. In this section I will discuss the results, their limitation as well as some perspectives.

III.1. ShareLoc platform

As discussed above on page 51, ShareLoc platform makes SMLM data publicly available, ensuring the FAIR principles (Findability, Accessibility, Interoperability and Reproducibility) in several ways: built upon Zenodo, ShareLoc automatically generates a DOI (Digital Object Identifier) when users upload data, ensuring the accessibility and findability. Additionally, the platform facilitates the assignment of metadata when uploading SMLM data, ensuring the interoperability and reproducibility.

ShareLoc promotes collaboration within the biological research community, since researchers can visualize, reuse, and reproduce SMLM data, aided by annotated protocols. Furthermore, the platform could enhance the developments of analytical tools like deep learning-assisted methods for SMLM data, as introduced on page 40.

The ontology-based tagging system for metadata on ShareLoc allows to easily annotate the SMLM data (page 52). Based on this feature, ShareLoc could help search for structural homologs. By leveraging the hierarchal relationships among tags, we can potentially connect the related SMLM data. This could facilitate the research of relevant data.

Another perspective is to extend the platform beyond the storage and visualization features. We could integrate plugins to provide computational tools, for example, to retrain models such as ANNA-PALM or localization algorithms. Researchers could benchmark, analyze, and derive insights all within the platform, without downloading large SMLM datasets.

There are serval other generic open-access data repositories, such as FigShare (http://figshare.com/) Dryad (White *et al.*, 2008), and IDR (Williams *et al.*, 2017) etc. A potential avenue for future development could involve collaboration between these platforms pooling their resources and minimizing data redundancies. This collaboration could minimize redundancy of data stored on different servers.

III.2. Enhancing the robustness of ANNA-PALM

III.2.1. Summary

In deep learning methods, large and diverse training data enable the model to generalize the underlying distribution of the data (Goodfellow, Bengio and Courville, 2016). As outlined on page 71, depending on the different experimental conditions and the acquisition protocols, the SMLM data distribution, such as resolution or morphology of biological structures can be changed.

Data sharing allows to increase the diversity and the quantity of the training data. In this thesis, I have demonstrated that re-training ANNA-PALM on larger and more diverse data significantly improves model robustness (page 72). Especially important is our observation that when the test data is obtained under different experiment conditions, the retrained model outperforms the model trained on homogenous data.

III.2.2. Adding evaluation metrics

As mentioned on page 74, I used MS-SSIM to assess the quality of image predicted by ANNA-PALM. While MS-SSIM remains our primary metric, alternative evaluations (Maier-Hein *et al.*, 2023) such as colocalization analysis (Dunn, Kamocka and McDonald, 2011) and IoU (Intersection Over Union) could be applied to provide a more complete assessment of reconstruction quality.

Colocalization analysis allows to measure the spatial correlation between two channels of a dual-color image, determining if there is a significant, non-random overlap between two channels. In practice, colocalization can be used to assess the consistency of the ANNAPALM reconstruction with the ground truth image.

The output of the model has different properties in terms of brightness, contrast and intensity value of pixels, etc. Consequently, to compare the IoU of the image generated from

model to the ground truth, it is necessary to segment images into binary form. As illustrated in **Figure 49**, the IoU is calculated by dividing the overlapping region of the ground truth and the output by the area of their union.

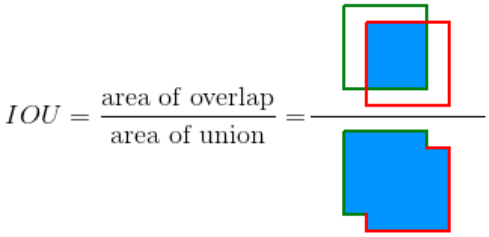


Figure 49. Illustration of index intersection over union (IoU).
Image from (Padilla, Netto and da Silva, 2020).

The IoU index is dependent on the segmentation algorithm. I investigated different approaches to segment the images: Otsu's thresholding method and adaptive Gaussian thresholding methods (Otsu, 1979; Sankur, 2004; *OpenCV: Image Thresholding*, no date), as depicted in **Figure 50**.

Otsu's method determines the binarization threshold by minimizing the intra-class variance of the two groups of pixel intensities. In the adapted Gaussian thresholding method, the threshold is a cross-correlation with a Gaussian window of the neighborhood of a pixel minus constant C. In the future, we could leverage IOU to assess the reconstruction quality of ANNA-PALM models.

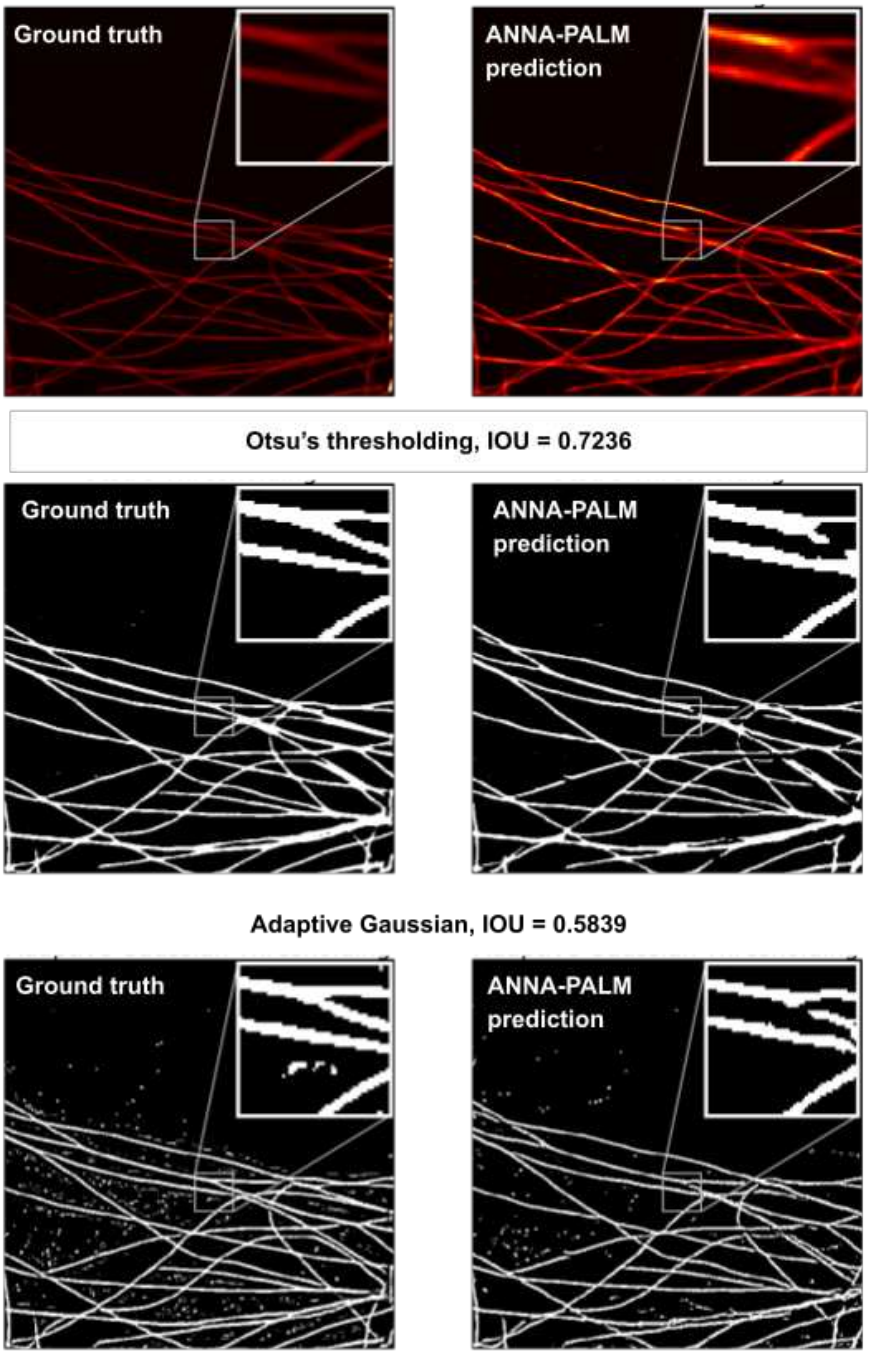


Figure 50. Different segmentation algorithms yield different IOU scores.
The images on the left side represent the ground truth, and on the right side are ANNA-PALM predictions. The upper panels display the original images; the middle panels display segmentations obtained by Otsu's thresholding method, and the lower panels show segmentations obtained by adaptive Gaussian methods.

III.2.3. Transfer learning

In order to save computational cost during the training phase, future work could also leverage transfer learning (Pan and Yang, 2010) or online training (Castro *et al.*, 2018):

Transfer learning involves fine-tuning a pre-trained model for a similar yet distinct task. This is based on the observation that early layers in deep neural networks capture generic data representations, while the later layers are more specific to the task (Yosinski *et al.*, 2014). In practice, the early layers of the neural network are "frozen", meaning the parameters are held fixed after training on a first data set. The later layers are updated during training on a second data set. Fine-tuning the later layers of a pre-trained model enhances the performance for a particular task. The pertaining allows the model to reach better performance faster on the second data set than if it was trained from scratch.

For example, if one wants to train ANNA-PALM to reconstruct dense SMLM images of microtubules with a new drug treatment, instead of training the model from scratch, we could use a model pre-trained on data shared on ShareLoc (as discussed on page 137) and fine-tune the later layers to enhance the performance with much fewer training iterations.

SMLM data collection expends when users upload new data on ShareLoc. We could pre-train a more generic model that can handle many different types of structures and conditions on this dataset. Ideally this pre-trained model evolves continuously while the training data becoming larger. This could be achieved by leveraging online training.

With online training, as fresh data emerges on platforms like ShareLoc, we do not need to retrain the model from scratch. The online learning allows models to learn new task without storing the old data (Castro *et al.*, 2018). When new data comes in, the model's parameters are updated based on the new data alone. However, the model may forget features learned from the previous data since it is always updating based on the newest data.

In the future, I intend to leverage transfer and online learning, with a focus on determining the best training conditions that facilitate the integration of new data while retaining important features learned from prior datasets.

III.3. Extending ANNA-PALM for live cell imaging

I have extended the ANNA-PALM model by using a 3D CNN-based architecture. This extension aims to reconstruct high-quality super-resolution SMLM movies in live cells, thereby capturing structural dynamics. The model consists of a conditional GAN (cGAN), where both the generator and discriminator are conditioned on the input data.

I simulated training data by overlaying two deformed dense SMLM images from either synthetic or experimental images in fixed cells. For synthetic microtubule images, results show that adding temporal information as a third dimension yielded much better reconstruction quality compared to the original 2D ANNA-PALM model.

Based on the synthetic microtubule data, I then quantitatively investigated the reconstruction quality as function of velocity of the moving structures and the localization rate. The reconstruction quality was evaluated using the MS-SSIM metric. The results indicated that the reconstruction quality increases when the structures move slower and when the localization rate is high. In order to determine the realistic range of these parameters for the experimental data, I computed the localization rate from experimental live-cell SMLM data and manually estimated the velocity from the widefield images alternating with the low resolution single molecule images.

Building on these results, I then applied the 3D ANNA-PALM model to real-world SMLM data of microtubules in live cells. I applied the same training strategy, where the model was trained on dynamic simulations obtained from experimental SMLM images of microtubules in fixed cells. However, the reconstruction quality did not meet expectations despite its potential demonstrated for synthetic microtubules data. To address this and to further refine the model, future work could focus on the following areas:

Once we have validated the 3D ANNA-PALM on SMLM data of microtubules in live cells, we could further enhance the model for more reliable and more general biological applications by:

- **Estimating the reconstruction uncertainty** (page 147)
- **Training and testing on different biological structures** (page 148)

III.3.1. Refining the dynamic simulations

As outlined on page 123, the dynamic simulation of semi-simulated SMLM movies is based on the deformed, overlaid fixed-cell images. However, when I applied similar deformations to generate hybrid-simulation-experimental SMLM movies, the dynamics seems to be insufficient. Thus, the simulation settings, such as velocity, localization rate and the deformation, could be further investigated, in order to generate more realistic dynamics.

Furthermore, a more sophisticated approach of generating dynamic ground truth super-resolution movies may help the model to output more accurate reconstructions. For this purpose, we propose to leverage an adversarial deep learning based approach. As depicted in **Figure 51**, we could use a cGAN to generate an SMLM movie using a single experimental dense SMLM image in fixed cells as input. The generator would use a recurrent U-net structure. The output of the network, a super-resolved movie, is then transformed into a low-resolution movie. This can be achieved either by a simple Gaussian blur or a neural network predicting the low-resolution image from the super-resolution image (Ouyang *et al.*, 2018). Then the discriminator distinguishes if the low-resolution movie is similar to the real-world widefield movie. The reconstructed super-resolution movies could serve as the ground truth for training data and should exhibit complex dynamics consistent with the dynamics observed in widefield movies.

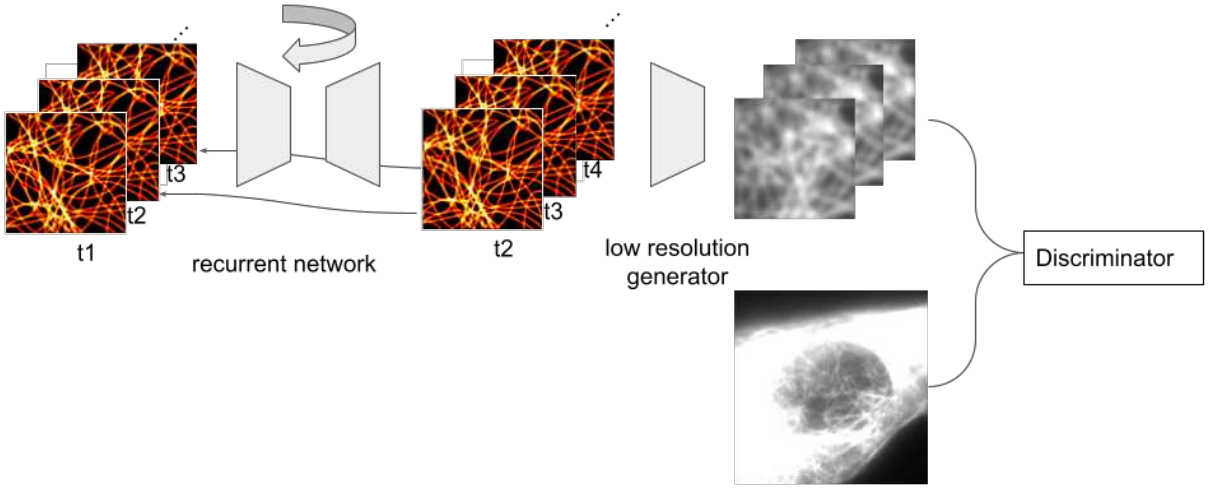


Figure 51. A proposed GAN strategy to simulate more realistic dynamics.
This Figure shows a potential neural network to generate the SMLM movies from experimental dense SMLM image (t1) in fixed cells. The generator consists in a recurrent U-net and a deep neural network that outputs low-resolution movie (top). The discriminator distinguishes if the low-resolution movie is similar to the real-world widefield movie (bottom).

III.3.2. Incorporating widefield images

Utilizing widefield images can provide complementary information that can enhance the quality of model's output (Ouyang *et al.*, 2018). By incorporating widefield images, the model may achieve a better performance for reconstructing biological structures and capturing their dynamics. As a next step, we aim to not only use the widefield movies acquired during the SMLM data acquisition for training but also as an indicator to evaluate the quality of reconstructions (R. Chen *et al.*, 2023).

III.3.3. Optimizing the use of temporal information

The temporal dimension is valuable for reconstructing the structural dynamics in live cells. Recent advancements in computer vision, particularly vision transformers (introduced on page 39), present a promising avenue to explore. The self-attention mechanism in transformers, which processes all embedded patches of movie simultaneously, might allow for better exploitation of temporal information (Arnab *et al.*, 2021; Liang *et al.*, 2021).

I tried to innovate on the 3D ANNA-PALM model by replacing its U-net based generator with a video vision transformer consisting of six transformer encoder blocks. Inspired by (Arnab *et al.*, 2021), the input data is embedded as shown in **Figure 52**.

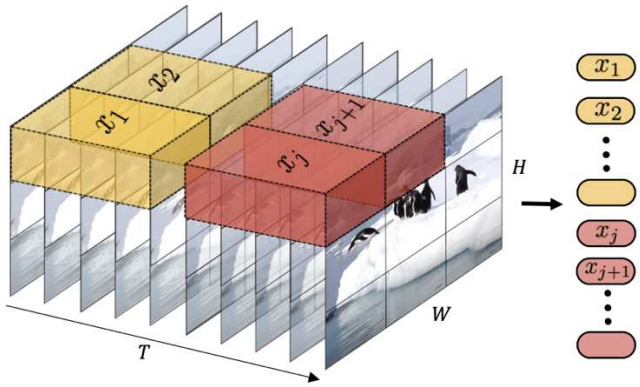


Figure 52. Illustration of input data embedding.
This figure shows the input data embedding for the video vision-transformer. The operation is not only for the spatial dimensions, but also for the temporal dimensions. Image from (Arnab *et al.*, 2021)

Preliminary results, as presented in **Figure 53**, show this simple transformer's output, the structure is overall recovered, however the output exhibits some hallucinations on the background.

Considering CNNs' ability to handle spatial, a possible perspective for future research is an exploration of the introduction of CNNs into vision transformers (Wu *et al.*, 2021). This integration may potentially enhance the performance of vision transformers, hence improve the reconstruction quality.

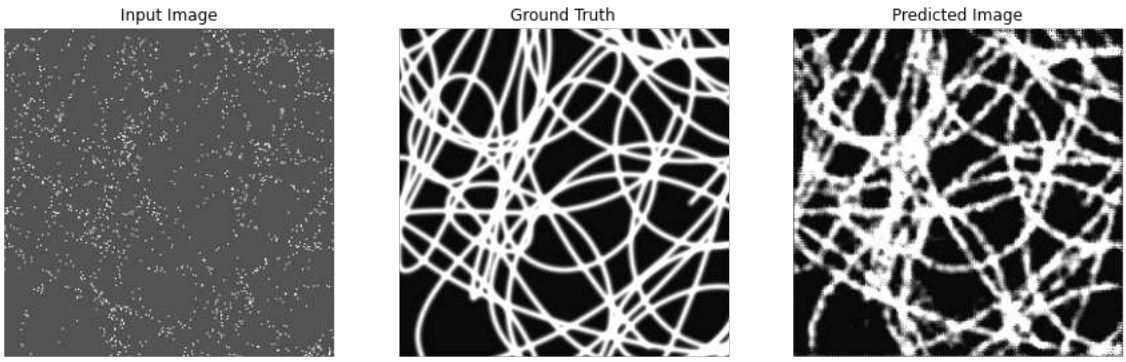


Figure 53. Preliminary result for video vision transformer architecture.
This figure displays the first frame of the movie for the input data (left panel), the ground truth (the middle panel) and the model's prediction (right panel).

III.3.4. Validating the 3D ANNA-PALM on live-cell experimental data

As mentioned before, unlike for fixed cell imaging, the ground truth for live cell SMLM imaging is absent. Hence validation must be done by other methods. To validate the 3D ANNA-PALM on live-cell experimental data imaging, we could potentially use specific imaging technique such as dual-color imaging in live cells. For example, we could perform dual color imaging of alpha and beta tubulins labeled with distinct dyes. These proteins form dimers that assemble into microtubules, hence the SMLM images of alpha-tubulin or beta-tubulin are essentially identical (**Figure 54**). For validation, we could separately feed SMLM sparse movies of alpha and beta tubulins into our 3D ANNA-PALM model. We could then assess the model by evaluating if the super-resolved movies reconstructed by the model are colocalized. This could allow us to gain additional insights into 3D ANNA-PALM's performance.

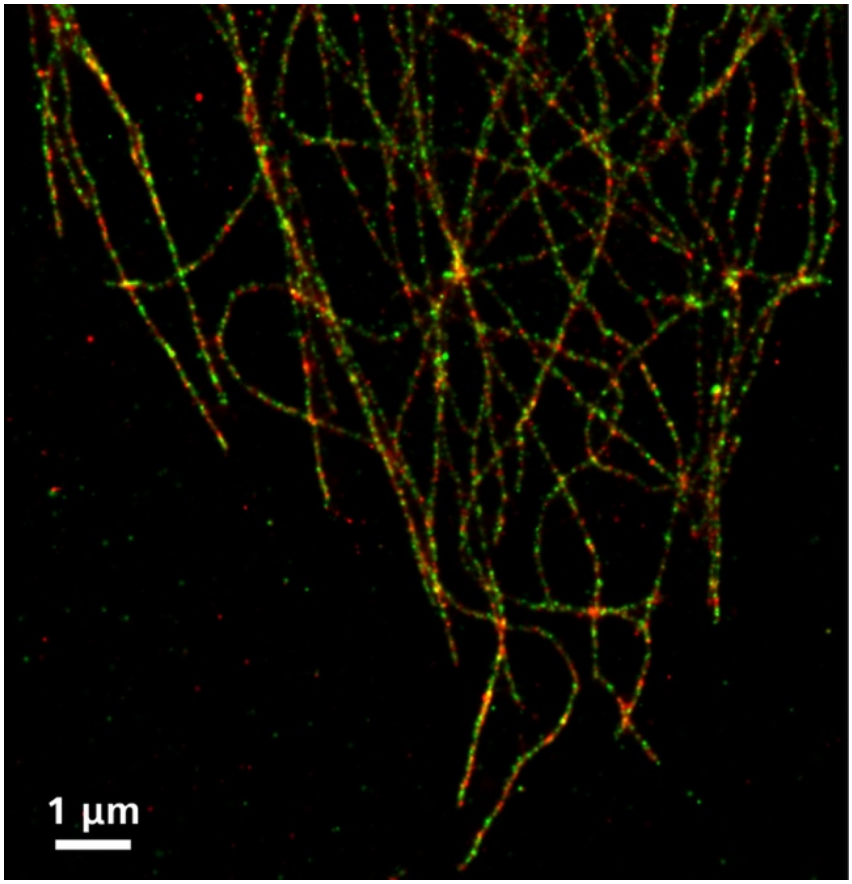


Figure 54. Dual-color imaging of alpha- and beta-tubulin as a tool for ANNAPALM validation. This Figure shows dual-color SMLM imaging of microtubules in fixed cells, with the two channels acquired simultaneously. Alpha tubulin was labelled with Alexa 555 (red) and beta tubulin with Alexa

488 (green). Scale bar: 1µm. Image from (*ZEISS Elyra 7 with Lattice SIM^2 Super-Resolution Microscope*, no date).

III.3.5. Estimating the uncertainty

As discussed on page 72, the model mismatch is a central challenge for deep learning, thus computing uncertainty estimation of model's output can provide an idea about how much we can trust the model's predictions. Uncertainty predictions are of particular importance in the medical domain (Ovadia *et al.*, 2019; Laves, Tölle and Ortmaier, 2020).

Uncertainties can be estimated using Bayesian Neural Networks (BNNs) (Jospin *et al.*, 2022). In BNNs, the weights are computed as probability distributions. BNNs rely on Bayes' theorem, denoted as:

$$p(\omega|D) \propto p(D|\omega)p(\omega)$$

where $p(\omega|D)$ is the posterior probability, representing the probability of the weights ω given the data D, $p(\omega)$ is the prior of the weights and $p(D|\omega)$ is the likelihood representing the probability of data given the weights.

In practice, it is difficult to compute the posterior probability. Thus, I applied dropout as a Bayesian approximation (Gal and Ghahramani, 2016). With dropout, random subsets of neurons are dropped out at each iteration. This is usually done during training to avoid overfitting (Srivastava *et al.*, 2014). However, for the Bayesian approximation, dropout is applied during testing. The model's output is then computed as the mean of N forward passes, while the uncertainty is computed as the variance of these N passes.

Figure 55 displays preliminary results from performing N forward passes while using dropout during the test stage (N=20). A larger variance indicates higher uncertainty: for example, reconstructions at intersections exhibit higher uncertainty. In the future, and upon successful validation, we could leverage the uncertainty estimation to better assess whether we can trust the model's prediction.

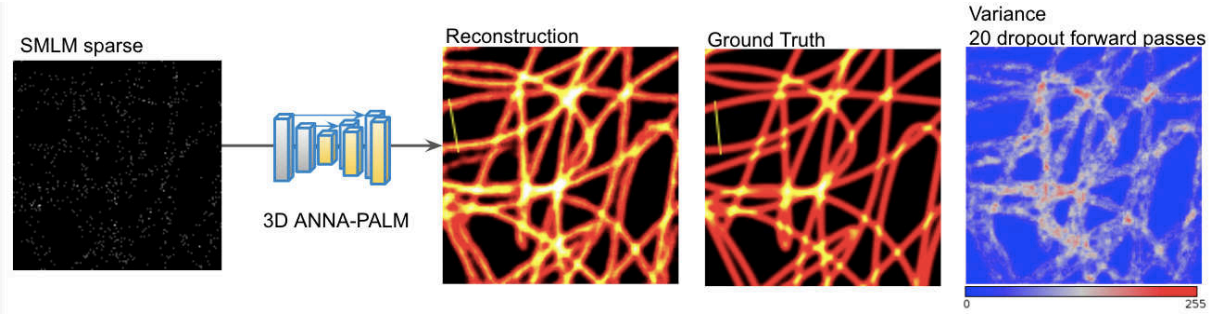


Figure 55. Preliminary result for uncertainty estimation.
The first panel is the input for 3D ANNA-PALM. The second panel is the reconstruction of the model, computed as the mean of 20 forward passes. The third panel represents the ground truth, and the fourth panel illustrates the variance of 20 passes.

III.3.6. Moving onto different biological structures

Our current focus was on the demonstrations of 3D ANNA-PALM on microtubules, mainly because they are widely used to test super-resolution methods, and because they are relatively simple and highly redundant structures. However, the methodology is not restricted to microtubules or filamentary structures. Thus, future work will aim to generalize the model to handle different biological structures such as mitochondria or the nuclear pore complex. The established training strategy could be adapted to SMLM data of various biological structures.

III.3.7. Conclusion

Once the 3D ANNA-PALM model is validated for experimental SMLM data in live cells, we believe that our approach may pave the way for super-resolution imaging of dynamic structures in live cells.

Additionally, benchmarking 3D ANNA-PALM model against methods developed by other research groups, such as DBlink (Saguy *et al.*, 2023) will offer insights into the relative performance of these approaches.

IV. References

Abbe, E. (1873) ‘Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung’, *Archiv für Mikroskopische Anatomie*, 9(1), pp. 413–468. Available at: https://doi.org/10.1007/BF02956173.

Ahlers, J. *et al.* (2023) ‘napari: a multi-dimensional image viewer for Python’. Zenodo. Available at: https://doi.org/10.5281/ZENODO.3555620.

Andresen, M. *et al.* (2007) ‘Structural basis for reversible photoswitching in Dronpa’, *Proceedings of the National Academy of Sciences*, 104(32), pp. 13005–13009. Available at: https://doi.org/10.1073/pnas.0700629104.

Arnab, A. *et al.* (2021) ‘ViViT: A Video Vision Transformer’. arXiv. Available at: http://arxiv.org/abs/2103.15691 (Accessed: 10 April 2023).

Balzarotti, F. *et al.* (2017) ‘Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes’, *Science*, 355(6325), pp. 606–612. Available at: https://doi.org/10.1126/science.aak9913.

Betzig, E. *et al.* (2006) ‘Imaging Intracellular Fluorescent Proteins at Nanometer Resolution’, *Science*, 313(5793), pp. 1642–1645. Available at: https://doi.org/10.1126/science.1127344.

Boland, M.V. and Murphy, R.F. (2001) ‘A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells’, *Bioinformatics*, 17(12), pp. 1213–1223. Available at: https://doi.org/10.1093/bioinformatics/17.12.1213.

Boyd, N. *et al.* (2018) *DeepLoco: Fast 3D Localization Microscopy Using Neural Networks*. preprint. Biophysics. Available at: https://doi.org/10.1101/267096.

Brameshuber, M. *et al.* (2022) ‘Understanding immune signaling using advanced imaging techniques’, *Biochemical Society Transactions*, 50(2), pp. 853–866. Available at: https://doi.org/10.1042/BST20210479.

Caper, H. (2022) ‘Generative adversarial networks explained’, *IBM developer*. Available at: https://developer.ibm.com/articles/generative-adversarial-networks-explained/.

Castro, F.M. *et al.* (2018) ‘End-to-End Incremental Learning’. arXiv. Available at: http://arxiv.org/abs/1807.09536 (Accessed: 7 September 2023).

Chalfie, M. *et al.* (1994) ‘Green Fluorescent Protein as a Marker for Gene Expression’, *Science*, 263(5148), pp. 802–805. Available at: https://doi.org/10.1126/science.8303295.

Chen, L.-C. *et al.* (2018) ‘DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), pp. 834–848. Available at: https://doi.org/10.1109/TPAMI.2017.2699184.

Chen, R. *et al.* (2023) ‘Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging’, *Nature Communications*, 14(1), p. 2854. Available at: https://doi.org/10.1038/s41467-023-38452-2.

Chen, X. *et al.* (2023) ‘Context Autoencoder for Self-Supervised Representation Learning’. arXiv. Available at: http://arxiv.org/abs/2202.03026 (Accessed: 30 August 2023).

Cheng, H.-J. *et al.* (2018) ‘Using Deep Learning to Identify Cell and Particle in Live-Cell Time-lapse Images’, in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Madrid, Spain: IEEE, pp. 1327–1331. Available at: https://doi.org/10.1109/BIBM.2018.8621567.

Chollet, F. (2018) *Deep learning with Python*. Shelter Island, New York: Manning Publications Co.

Cole, R.W., Jinadasa, T. and Brown, C.M. (2011) 'Measuring and interpreting point spread functions to determine confocal microscope resolution and ensure quality control', *Nature Protocols*, 6(12), pp. 1929–1941. Available at: https://doi.org/10.1038/nprot.2011.407.

Colson, P. *et al.* (2017) 'Mimivirus: leading the way in the discovery of giant viruses of amoebae', *Nature Reviews Microbiology*, 15(4), pp. 243–254. Available at: https://doi.org/10.1038/nrmicro.2016.197.

Coons, A.H., Creech, H.J. and Jones, R.N. (1941) 'Immunological Properties of an Antibody Containing a Fluorescent Group.', *Experimental Biology and Medicine*, 47(2), pp. 200–202. Available at: https://doi.org/10.3181/00379727-47-13084P.

Dertinger, T. *et al.* (2009) 'Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI)', *Proceedings of the National Academy of Sciences*, 106(52), pp. 22287–22292. Available at: https://doi.org/10.1073/pnas.0907866106.

Dunn, K.W., Kamocka, M.M. and McDonald, J.H. (2011) 'A practical guide to evaluating colocalization in biological microscopy', *American Journal of Physiology-Cell Physiology*, 300(4), pp. C723–C742. Available at: https://doi.org/10.1152/ajpcell.00462.2010.

Frisk, M.L. (2012) 'Visible to the Naked Eye', *Science Translational Medicine*, 4(159). Available at: https://doi.org/10.1126/scitranslmed.3005261.

Gal, Y. and Ghahramani, Z. (2016) 'Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning', *arXiv:1506.02142 [cs, stat]* [Preprint]. Available at: http://arxiv.org/abs/1506.02142 (Accessed: 4 April 2022).

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep learning*. Cambridge, Massachusetts: The MIT Press (Adaptive computation and machine learning).

Goodfellow, I.J. *et al.* (2014) 'Generative Adversarial Networks'. arXiv. Available at: http://arxiv.org/abs/1406.2661 (Accessed: 16 August 2023).

Gustafsson, M.G.L. (2000) 'Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. SHORT COMMUNICATION', *Journal of Microscopy*, 198(2), pp. 82–87. Available at: https://doi.org/10.1046/j.1365-2818.2000.00710.x.

Gustafsson, N. *et al.* (2016) 'Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations', *Nature Communications*, 7(1), p. 12471. Available at: https://doi.org/10.1038/ncomms12471.

Habuchi, S. (2014) 'Super-Resolution Molecular and Functional Imaging of Nanoscale Architectures in Life and Materials Science', *Frontiers in Bioengineering and Biotechnology*, 2. Available at: https://doi.org/10.3389/fbioe.2014.00020.

He, K. *et al.* (2016) 'Deep Residual Learning for Image Recognition', in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, pp. 770–778. Available at: https://doi.org/10.1109/CVPR.2016.90.

Heilemann, M. *et al.* (2008) 'Subdiffraction-Resolution Fluorescence Imaging with Conventional Fluorescent Probes', *Angewandte Chemie International Edition*, 47(33), pp. 6172–6176. Available at: https://doi.org/10.1002/anie.200802376.

Heintzmann, R., Jovin, T.M. and Cremer, C. (2002) 'Saturated patterned excitation microscopy—a concept for optical resolution improvement', *Journal of the Optical Society of America A*, 19(8), p. 1599. Available at: https://doi.org/10.1364/JOSAA.19.001599.

Hell, S.W. and Wichmann, J. (1994) 'Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy', *Optics Letters*, 19(11), p. 780. Available at: https://doi.org/10.1364/OL.19.000780.

Hestness, J. *et al.* (2017) 'Deep Learning Scaling is Predictable, Empirically', *arXiv:1712.00409 [cs, stat]* [Preprint]. Available at: http://arxiv.org/abs/1712.00409 (Accessed: 15 January 2022).

Hochreiter, B., Pardo-Garcia, A. and Schmid, J. (2015) 'Fluorescent Proteins as Genetically Encoded FRET Biosensors in Life Sciences', *Sensors*, 15(10), pp. 26281–26314. Available at: https://doi.org/10.3390/s151026281.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long Short-Term Memory', *Neural Computation*, 9(8), pp. 1735–1780. Available at: https://doi.org/10.1162/neco.1997.9.8.1735.

Holden, S.J. *et al.* (2014) 'High throughput 3D super-resolution microscopy reveals *Caulobacter crescentus* in vivo Z-ring organization', *Proceedings of the National Academy of Sciences*, 111(12), pp. 4566–4571. Available at: https://doi.org/10.1073/pnas.1313368111.

Huang, F. *et al.* (2013) 'Video-rate nanoscopy using sCMOS camera–specific single-molecule localization algorithms', *Nature Methods*, 10(7), pp. 653–658. Available at: https://doi.org/10.1038/nmeth.2488.

Isola, P. *et al.* (2018) 'Image-to-Image Translation with Conditional Adversarial Networks', *arXiv:1611.07004 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1611.07004 (Accessed: 14 February 2022).

Janeway, C. (ed.) (2001) *Immunobiology: the immune system in health and disease ; [animated CD-ROM inside]*. 5. ed. New York, NY: Garland Publ. [u.a.].

Jin, L. *et al.* (2020) 'Deep learning enables structured illumination microscopy with low light levels and enhanced speed', *Nature Communications*, 11(1), p. 1934. Available at: https://doi.org/10.1038/s41467-020-15784-x.

Johnson, L.M. *et al.* (2021) 'Fabrication of polyethylene terephthalate (PET) nanoparticles with fluorescent tracers for studies in mammalian cells', *Nanoscale Advances*, 3(2), pp. 339–346. Available at: https://doi.org/10.1039/D0NA00888E.

Jones, S.A. *et al.* (2011) 'Fast, three-dimensional super-resolution imaging of live cells', *Nature Methods*, 8(6), pp. 499–505. Available at: https://doi.org/10.1038/nmeth.1605.

Jospin, L.V. *et al.* (2022) 'Hands-on Bayesian Neural Networks -- a Tutorial for Deep Learning Users', *IEEE Computational Intelligence Magazine*, 17(2), pp. 29–48. Available at: https://doi.org/10.1109/MCI.2022.3155327.

Jupp, S. *et al.* (no date) 'A New Ontology Lookup Service at EMBL-EBI'.

Karras, T., Laine, S. and Aila, T. (2019) 'A Style-Based Generator Architecture for Generative Adversarial Networks'. arXiv. Available at: http://arxiv.org/abs/1812.04948 (Accessed: 16 August 2023).

Kim, Y. (2014) 'Convolutional Neural Networks for Sentence Classification'. arXiv. Available at: http://arxiv.org/abs/1408.5882 (Accessed: 30 August 2023).

Kolb, H.C., Finn, M.G. and Sharpless, K.B. (2001) 'Click Chemistry: Diverse Chemical Function from a Few Good Reactions', *Angewandte Chemie International Edition*, 40(11), pp. 2004–2021. Available at: https://doi.org/10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5.

Krenn, M. *et al.* (2023) 'Artificial intelligence and machine learning for quantum technologies', *Physical Review A*, 107(1), p. 010101. Available at: https://doi.org/10.1103/PhysRevA.107.010101.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) 'ImageNet classification with deep convolutional neural networks', *Communications of the ACM*, 60(6), pp. 84–90. Available at: https://doi.org/10.1145/3065386.

Krull, A., Buchholz, T.-O. and Jug, F. (2019) 'Noise2Void - Learning Denoising from Single Noisy Images', *arXiv:1811.10980 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1811.10980 (Accessed: 15 October 2020).

Kumar Gaire, S. *et al.* (2020) 'Accelerating multicolor spectroscopic single-molecule localization microscopy using deep learning', *Biomedical Optics Express*, 11(5), p. 2705. Available at: https://doi.org/10.1364/BOE.391806.

Laves, M.-H., Tölle, M. and Ortmaier, T. (2020) 'Uncertainty Estimation in Medical Image Denoising with Bayesian Deep Image Prior', *arXiv:2008.08837 [cs, eess]* [Preprint]. Available at: http://arxiv.org/abs/2008.08837 (Accessed: 15 October 2020).

Lecun, Y. *et al.* (1998) 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE*, 86(11), pp. 2278–2324. Available at: https://doi.org/10.1109/5.726791.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', *Nature*, 521(7553), pp. 436–444. Available at: https://doi.org/10.1038/nature14539.

Ledig, C. *et al.* (2017) 'Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network', in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, pp. 105–114. Available at: https://doi.org/10.1109/CVPR.2017.19.

Lee, T. *et al.* (2011) 'Optical Microscopy of Soft Matter Systems'. arXiv. Available at: http://arxiv.org/abs/1108.3287 (Accessed: 21 August 2023).

Lelek, M. *et al.* (2021) 'Single-molecule localization microscopy', *Nature Reviews Methods Primers*, 1(1), p. 39. Available at: https://doi.org/10.1038/s43586-021-00038-x.

Liang, J. *et al.* (2021) 'SwinIR: Image Restoration Using Swin Transformer'. arXiv. Available at: http://arxiv.org/abs/2108.10257 (Accessed: 7 April 2023).

Lichtman, J.W. and Conchello, J.-A. (2005) 'Fluorescence microscopy', *Nature Methods*, 2(12), pp. 910–919. Available at: https://doi.org/10.1038/nmeth817.

Liu, S., Hoess, P. and Ries, J. (2022) 'Super-Resolution Microscopy for Structural Cell Biology', *Annual Review of Biophysics*, 51(1), pp. 301–326. Available at: https://doi.org/10.1146/annurev-biophys-102521-112912.

Liu, Z. *et al.* (2021) 'Video Swin Transformer'. arXiv. Available at: http://arxiv.org/abs/2106.13230 (Accessed: 12 July 2022).

Lukyanov, K.A. *et al.* (2005) 'Photoactivatable fluorescent proteins', *Nature Reviews Molecular Cell Biology*, 6(11), pp. 885–890. Available at: https://doi.org/10.1038/nrm1741.

Maaten, L. van der and Hinton, G. (2008) 'Visualizing Data using t-SNE', *Journal of Machine Learning Research*, 9(86), pp. 2579–2605.

Machine Learning Models: What They Are and How to Build Them (2023) *Coursera*. Available at: https://www.coursera.org/articles/machine-learning-models (Accessed: 24 September 2023).

Maier-Hein, L. *et al.* (2023) 'Metrics reloaded: Recommendations for image analysis validation'. arXiv. Available at: http://arxiv.org/abs/2206.01653 (Accessed: 5 September 2023).

Mandal, M. (2021) 'Introduction to Convolutional Neural Networks (CNN)', *Analytics Vidhya*, 1 May. Available at: https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/ (Accessed: 2 September 2023).

McCulloch, W.S. and Pitts, W. (1943) 'A logical calculus of the ideas immanent in nervous activity', *The Bulletin of Mathematical Biophysics*, 5(4), pp. 115–133. Available at: https://doi.org/10.1007/BF02478259.

Mnih, V. *et al.* (2013) 'Playing Atari with Deep Reinforcement Learning'. arXiv. Available at: http://arxiv.org/abs/1312.5602 (Accessed: 30 August 2023).

Möckl, L., Roy, A.R. and Moerner, W.E. (2020) 'Deep learning in single-molecule microscopy: fundamentals, caveats, and recent developments [Invited]', *Biomedical Optics Express*, 11(3), p. 1633. Available at: https://doi.org/10.1364/BOE.386361.

Mujumdar, R.B. *et al.* (1993) 'Cyanine dye labeling reagents: Sulfoindocyanine succinimidyl esters', *Bioconjugate Chemistry*, 4(2), pp. 105–111. Available at: https://doi.org/10.1021/bc00020a001.

Nehme, E. *et al.* (2018) 'Deep-STORM: super-resolution single-molecule microscopy by deep learning', *Optica*, 5(4), p. 458. Available at: https://doi.org/10.1364/OPTICA.5.000458.

OpenCV: Image Thresholding (no date). Available at: https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html (Accessed: 22 September 2023).

Orchard, G. and Nation, B. (eds) (2015) *Cell structure and function*. Oxford, United Kingdom: Oxford University Press (Fundamentals of biomedical science).

Otsu, N. (1979) 'A Threshold Selection Method from Gray-Level Histograms', *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), pp. 62–66. Available at: https://doi.org/10.1109/TSMC.1979.4310076.

Ouyang, W. *et al.* (2018) 'Deep learning massively accelerates super-resolution localization microscopy', *Nature Biotechnology*, 36(5), pp. 460–468. Available at: https://doi.org/10.1038/nbt.4106.

Ovadia, Y. *et al.* (2019) 'Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift', *arXiv:1906.02530 [cs, stat]* [Preprint]. Available at: http://arxiv.org/abs/1906.02530 (Accessed: 8 January 2020).

Padilla, R., Netto, S.L. and da Silva, E.A.B. (2020) 'A Survey on Performance Metrics for Object-Detection Algorithms', in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Niteroi, Brazil: IEEE.

Pan, S.J. and Yang, Q. (2010) 'A Survey on Transfer Learning', *IEEE Transactions on Knowledge and Data Engineering*, 22(10), pp. 1345–1359. Available at: https://doi.org/10.1109/TKDE.2009.191.

Panchuk-Voloshina, N. *et al.* (1999) 'Alexa Dyes, a Series of New Fluorescent Dyes that Yield Exceptionally Bright, Photostable Conjugates', *Journal of Histochemistry & Cytochemistry*, 47(9), pp. 1179–1188. Available at: https://doi.org/10.1177/002215549904700910.

Phung and Rhee (2019) 'A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets', *Applied Sciences*, 9(21), p. 4500. Available at: https://doi.org/10.3390/app9214500.

QuickGO::Annotation List (no date). Available at: https://www.ebi.ac.uk/QuickGO/annotations?geneProductId=A0A7K7RBK5 (Accessed: 20 September 2023).

QuickGO::Term GO:0005643 (no date). Available at: https://www.ebi.ac.uk/QuickGO/term/GO:0005643 (Accessed: 20 September 2023).

Recht, B. *et al.* (2018) 'Do CIFAR-10 Classifiers Generalize to CIFAR-10?' Available at: https://doi.org/10.48550/ARXIV.1806.00451.

Reddy, B.S. and Chatterji, B.N. (1996) 'An FFT-based technique for translation, rotation, and scale-invariant image registration', *IEEE Transactions on Image Processing*, 5(8), pp. 1266–1271. Available at: https://doi.org/10.1109/83.506761.

Redmon, J. *et al.* (2016) 'You Only Look Once: Unified, Real-Time Object Detection', in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, pp. 779–788. Available at: https://doi.org/10.1109/CVPR.2016.91.

Ren, S. *et al.* (2016) 'Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks'. arXiv. Available at: http://arxiv.org/abs/1506.01497 (Accessed: 16 August 2023).

Ronneberger, O., Fischer, P. and Brox, T. (2015) 'U-Net: Convolutional Networks for Biomedical Image Segmentation', in N. Navab et al. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 234–241. Available at: https://doi.org/10.1007/978-3-319-24574-4_28.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) 'Learning representations by back-propagating errors', *Nature*, 323(6088), pp. 533–536. Available at: https://doi.org/10.1038/323533a0.

Rust, M.J., Bates, M. and Zhuang, X. (2006) 'Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)', *Nature Methods*, 3(10), pp. 793–796. Available at: https://doi.org/10.1038/nmeth929.

Saguy, A. *et al.* (2023) 'DBlink: dynamic localization microscopy in super spatiotemporal resolution via deep learning', *Nature Methods* [Preprint]. Available at: https://doi.org/10.1038/s41592-023-01966-0.

Sahl, S.J. and Hell, S.W. (2019) 'High-Resolution 3D Light Microscopy with STED and RESOLFT', in J.F. Bille (ed.) *High Resolution Imaging in Microscopy and Ophthalmology*. Cham: Springer International Publishing, pp. 3–32. Available at: https://doi.org/10.1007/978-3-030-16638-0_1.

Sakkos, D. *et al.* (2018) 'End-to-end video background subtraction with 3d convolutional neural networks', *Multimedia Tools and Applications*, 77(17), pp. 23023–23041. Available at: https://doi.org/10.1007/s11042-017-5460-9.

Sankur, B. (2004) 'Survey over image thresholding techniques and quantitative performance evaluation', *Journal of Electronic Imaging*, 13(1), p. 146. Available at: https://doi.org/10.1117/1.1631315.

Schermelleh, L. *et al.* (2019) 'Super-resolution microscopy demystified', *Nature Cell Biology*, 21(1), pp. 72–84. Available at: https://doi.org/10.1038/s41556-018-0251-8.

Schnitzbauer, J. *et al.* (2017) 'Super-resolution microscopy with DNA-PAINT', *Nature Protocols*, 12(6), pp. 1198–1228. Available at: https://doi.org/10.1038/nprot.2017.024.

Schutz, M., Krosl, K. and Wimmer, M. (2019) 'Real-Time Continuous Level of Detail Rendering of Point Clouds', in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Osaka, Japan: IEEE, pp. 103–110. Available at: https://doi.org/10.1109/VR.2019.8798284.

Sermanet, P. *et al.* (2018) 'Time-Contrastive Networks: Self-Supervised Learning from Video', in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD: IEEE, pp. 1134–1141. Available at: https://doi.org/10.1109/ICRA.2018.8462891.

Shaner, N.C. *et al.* (2004) 'Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein', *Nature Biotechnology*, 22(12), pp. 1567–1572. Available at: https://doi.org/10.1038/nbt1037.

ShareLoc (no date). Available at: https://www.ebi.ac.uk/ols4/ontologies/shareloc (Accessed: 3 September 2023).

Sharonov, A. and Hochstrasser, R.M. (2006) 'Wide-field subdiffraction imaging by accumulated binding of diffusing probes', *Proceedings of the National Academy of Sciences*, 103(50), pp. 18911–18916. Available at: https://doi.org/10.1073/pnas.0609643104.

Sheth, D.Y. *et al.* (2021) 'Unsupervised Deep Video Denoising', *arXiv:2011.15045 [cs, eess, stat]* [Preprint]. Available at: http://arxiv.org/abs/2011.15045 (Accessed: 31 January 2022).

Shroff, H. *et al.* (2008) 'Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics', *Nature Methods*, 5(5), pp. 417–423. Available at: https://doi.org/10.1038/nmeth.1202.

Silver, D. *et al.* (2017) 'Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm'. arXiv. Available at: http://arxiv.org/abs/1712.01815 (Accessed: 30 August 2023).

Simonyan, K. and Zisserman, A. (2015) 'Very Deep Convolutional Networks for Large-Scale Image Recognition', *arXiv:1409.1556 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1409.1556 (Accessed: 14 January 2021).

Speiser, A. *et al.* (2021) 'Deep learning enables fast and dense single-molecule localization with high accuracy', *Nature Methods*, 18(9), pp. 1082–1090. Available at: https://doi.org/10.1038/s41592-021-01236-x.

Srivastava, N. *et al.* (2014) 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting', *Journal of Machine Learning Research*, 15(56), pp. 1929–1958.

Sutskever, I., Vinyals, O. and Le, Q.V. (2014) 'Sequence to Sequence Learning with Neural Networks'. arXiv. Available at: http://arxiv.org/abs/1409.3215 (Accessed: 30 August 2023).

Tran, D. *et al.* (2015) 'Learning Spatiotemporal Features with 3D Convolutional Networks', in *2015 IEEE International Conference on Computer Vision (ICCV)*. *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile: IEEE, pp. 4489–4497. Available at: https://doi.org/10.1109/ICCV.2015.510.

Understanding LSTM Networks -- colah's blog (no date). Available at: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (Accessed: 2 September 2023).

Vaswani, A. *et al.* (2017) 'Attention Is All You Need', *arXiv:1706.03762 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1706.03762 (Accessed: 15 January 2022).

Vázquez-Gutiérrez, J.L. and Langton, M. (2015) 'Current potential and limitations of immunolabeling in cereal grain research', *Trends in Food Science & Technology*, 41(2), pp. 105–117. Available at: https://doi.org/10.1016/j.tifs.2014.10.002.

Wade, R. (1997) 'Microtubule structure and dynamics', *Current Opinion in Cell Biology*, 9(1), pp. 12–17. Available at: https://doi.org/10.1016/S0955-0674(97)80146-9.

Wang, Z., Simoncelli, E.P. and Bovik, A.C. (2003) 'Multiscale structural similarity for image quality assessment', in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. Conference Record of the 37th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA: IEEE, pp. 1398–1402. Available at: https://doi.org/10.1109/ACSSC.2003.1292216.

Weigert, M. *et al.* (2018) 'Content-aware image restoration: pushing the limits of fluorescence microscopy', *Nature Methods*, 15(12), pp. 1090–1097. Available at: https://doi.org/10.1038/s41592-018-0216-7.

Wente, S.R. and Rout, M.P. (2010) 'The Nuclear Pore Complex and Nuclear Transport', *Cold Spring Harbor Perspectives in Biology*, 2(10), pp. a000562–a000562. Available at: https://doi.org/10.1101/cshperspect.a000562.

What Is a Cell? | Learn Science at Scitable (no date). Available at: https://www.nature.com/scitable/topicpage/what-is-a-cell-14023083/ (Accessed: 9 August 2023).

White, H.C. *et al.* (2008) 'The Dryad Data Repository: A Singapore Framework Metadata Architecture in a DSpace Environment', *International Conference on Dublin Core and Metadata Applications*, pp. 157–162.

Wilkinson, M.D. *et al.* (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 3(1), p. 160018. Available at: https://doi.org/10.1038/sdata.2016.18.

Williams, E. *et al.* (2017) 'Image Data Resource: a bioimage data integration and publication platform', *Nature Methods*, 14(8), pp. 775–781. Available at: https://doi.org/10.1038/nmeth.4326.

Wu, H. *et al.* (2021) 'CvT: Introducing Convolutions to Vision Transformers'. arXiv. Available at: http://arxiv.org/abs/2103.15808 (Accessed: 11 September 2023).

Xie, J., Girshick, R. and Farhadi, A. (2016) 'Unsupervised Deep Embedding for Clustering Analysis'. arXiv. Available at: http://arxiv.org/abs/1511.06335 (Accessed: 30 August 2023).

Xu, L. *et al.* (2014) 'Deep Convolutional Neural Network for Image Deconvolution', in Z. Ghahramani et al. (eds) *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper_files/paper/2014/file/1c1d4df596d01da60385f0bb17a4a9e0-Paper.pdf.

Yosinski, J. *et al.* (2014) 'How transferable are features in deep neural networks?' arXiv. Available at: http://arxiv.org/abs/1411.1792 (Accessed: 7 September 2023).

ZEISS Elyra 7 with Lattice SIM^2 Super-Resolution Microscope (no date). Available at: https://www.zeiss.com/microscopy/en/products/light-microscopes/super-resolution-microscopes/elyra-7.html (Accessed: 23 September 2023).

Zhang, K. *et al.* (2017) 'Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising', *IEEE Transactions on Image Processing*, 26(7), pp. 3142–3155. Available at: https://doi.org/10.1109/TIP.2017.2662206.

Zhang, Z. *et al.* (2015) 'Ultrahigh-throughput single-molecule spectroscopy and spectrally resolved super-resolution microscopy', *Nature Methods*, 12(10), pp. 935–938. Available at: https://doi.org/10.1038/nmeth.3528.

Zhou Wang and Bovik, A.C. (2009) 'Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures', *IEEE Signal Processing Magazine*, 26(1), pp. 98–117. Available at: https://doi.org/10.1109/MSP.2008.930649.

V. Table of figures

VI. Table of tables

VII. Acknowledgement

First and foremost, I would like to express my deepest gratitude to my thesis supervisor, Christophe ZIMMER, for his unwavering support, patience, and guidance throughout this research journey, even during the most challenging moments. The countless hours we spent discussing and refining this work have not only contributed to its quality but have also been moments of profound learning for me.

A heartfelt acknowledgment goes to all the members of my lab. A special mention to Benoît LELANDAIS, whose constructive suggestions and feedback were indispensable in shaping this work. Additionally, my deepest thanks to Manish SINGH for his invaluable contribution in data acquisition.

My appreciation extends to Ouyang WEI for the enlightening discussions that added depth to this work, Abbelight, Christophe LETERRIER's lab, Markus SAUER's lab and Pakorn KANCHANAWONG's lab for their collaboration, assistance.

I wish to express my sincere appreciation to my TAC members, Jean-Luc STARCK, Sandrine ETIENNE-MANNEVILLE, and Antoine COULON for their constructive feedback and suggestions that immensely improved the quality of this research.

I would like to thank the administrative and technical staff of DSI department, especially Youssef GHORBAL and Konstantin PETROV, for their timely help and support throughout my academic journey.

To Glo, Jie, Shahinez, Steve and Yufei, thank you for standing by my side, through peaks and valleys of this journey. I also want to thank to IMOD mafia especially Hoa, Isabella and Solène, for all the expeditions inside and outside Paris.

Lastly, I cannot end this without expressing my heartfelt gratitude to my family, especially my parents, for their unconditional love and support, sacrifices, and belief in me. 我最亲爱的爸爸妈妈， 奶奶，柯和姑姑，谢谢你们的爱和支持， 我爱你们。

Abstract:

While microscopy has been a central technique for cell biology since centuries, it has long been limited by diffraction to a resolution of ~200-300 nm. As a consequence, many molecular structures, such as viruses, nuclear pores, or microtubules were left unresolved. Single-molecule localization microscopy (SMLM) offers a high spatial resolution (e.g., 20 nm or better), allowing to resolve biological structures at or near the molecular scale. However, SMLM acquisition necessitates acquiring many thousands of low-resolution frames, mostly limiting its applications to fixed cells or to structures undergoing slow dynamics. To overcome this limitation, Ouyang et al. (2018) developed a deep learning-based approach called ANNA-PALM that can reconstruct a super-resolution image from much fewer low-resolution frames. However, the original ANNA-PALM method faced several limitations. First, ANNA-PALM had only been trained and tested on images from our laboratory. Second, the method exhibits artifacts when applied to images obtained using different protocols or experimental conditions than the training data. Third, ANNA-PALM had only been demonstrated on fixed cells.

The objectives of my Ph.D. thesis are to address these limitations by 1) improving the robustness of ANNA-PALM reconstructions when applied to data obtained from distinct laboratories and 2) extending ANNA-PALM to reconstruct super-resolved time-lapse image sequences for dynamic biological structures in live cells.

1. Improving the robustness of ANNA PALM: an obvious approach to improve robustness is to retrain the model using a larger and more varied data set. However, SMLM datasets are not usually publicly accessible. To address this, our lab developed ShareLoc, an online platform (shareloc.xyz) that allows the gathering and reuse of SMLM datasets acquired by the microscopy community. I first validated the platform's functionalities, curated SMLM data, implemented a ShareLoc ontology, and wrote relevant documentation. Next, I took advantage of ShareLoc data to retrain ANNA-PALM on larger and more diverse images and quantitatively evaluated the image reconstruction quality compared to the original model. I demonstrated that the robustness and reconstruction quality of ANNA-PALM significantly improved, notably when applied to images of microtubules taken under biological perturbation conditions never seen by the model during training.

2. Extending ANNA-PALM to reconstruct super-resolution movies of moving structures in live cells: achieving high-quality super-resolution reconstructions of structural dynamics is challenging. To avoid motion blur, each frame of the reconstructed movie is defined from localizations in only a small number of consecutive low-resolution frames. This leads to a strong under-sampling of the structures by single molecule localization events and does not enable super-resolution. Although ANNA-PALM can reconstruct high-quality super-resolved images from under-sampled localization data, training ANNA-PALM for live cells is more difficult, because a clear ground truth is lacking. The absence of ground truth also makes it difficult to assess reconstruction quality. To address these challenges, I first developed a method to generate ground truth super-resolution movies from static SMLM images obtained from long acquisition sequences. I implemented and tested this strategy using both simulated and experimental SMLM images. Second, I extended the ANNA-PALM architecture to 3D data, where the third dimension is time, in order to incorporate temporal information. I used simulations of microtubule dynamics to quantitatively evaluate the reconstruction quality of this approach in comparison with the original 2D ANNA-PALM, and as function of structure velocity and localization rates. The results show that incorporating temporal information considerably improves reconstruction quality and suggest that high quality live cell super-resolution movies of microtubules can in principle be reconstructed with realistic SMLM imaging parameters. Once validated on experimental data, our method should help pave the way for super-resolution imaging of dynamic structures in live cells.

Keywords: [Deep learning, SMLM, Live cell imaging, Super-resolution microscopy]

[Reconstruction de microscopie de localisation à molécules uniques par apprentissage profond - Amélioration de la robustesse et progression vers l'étude de cellules vivantes]

Résumé :

La microscopie optique est une technique centrale en biologie cellulaire, mais la diffraction limite sa résolution à environ 200-300 nm. Par conséquent, de nombreuses structures moléculaires telles que les virus, les pores nucléaires ou les microtubules, ne peuvent pas être résolues. La microscopie de localisation à molécule unique (SMLM) offre une résolution spatiale élevée (20 nm ou mieux), permettant de résoudre les structures biologiques à l'échelle moléculaire. Cependant, la SMLM nécessite l'acquisition de plusieurs milliers d'images à base résolution, ce qui limite ses applications aux cellules fixes ou aux structures à dynamique lente. Pour surmonter cette limitation, Ouyang et al. (2018) ont développé une approche d'apprentissage profond appelée ANNA-PALM, capable de reconstruire des images super-résolutives à partir d'un nombre beaucoup plus réduit d'images à basse résolution. Cependant, la méthode ANNA-PALM originale présente plusieurs contraintes. Tout d'abord, cette méthode présente des artefacts lorsqu'elle est appliquée à des images obtenues avec des protocoles ou dans des conditions expérimentales différents des données d'entraînement. Par ailleurs, ANNA-PALM n'a été démontrée que sur des cellules fixes.

Ma thèse vise à résoudre ces limitations : 1) en améliorant la robustesse d'ANNA-PALM pour des données issues de laboratoires ou de conditions expérimentales différents et 2) en l'étendant à la visualisation super-résolutive de structures biologiques dynamiques dans les cellules vivantes.

1. Amélioration de la robustesse d'ANNA PALM : notre laboratoire a développé ShareLoc, une plateforme en ligne (shareloc.xyz) qui permet la collecte, la visualisation et la réutilisation des données SMLM acquises par la communauté de microscopie. Dans un premier temps, j'ai validé les fonctionnalités de la plateforme, effectué la curation des données SMLM, implémenté l'ontologie ShareLoc et rédigé la documentation. Ensuite, j'ai ré-entraîné ANNA-PALM sur des images en plus grande quantité et plus variées partagées sur ShareLoc. J'ai évalué quantitativement la qualité des images reconstruites par rapport au modèle original. J'ai démontré une amélioration significative de la robustesse et de la qualité des reconstructions par ANNA-PALM, en particulier pour des images de microtubules prises dans des conditions biologiques différentes de celles des images d'entraînement.

2. Extension d'ANNA-PALM à la reconstruction de films super-résolutifs de la dynamique structurelle en cellules vivantes : pour éviter le floutage dû à la dynamique des structures, chaque image super-résolutive du film reconstruit est basée sur un faible nombre d'images consécutives à basse résolution, ce qui conduit à un sous-échantillonnage important de la structure, qui restreint la résolution. Bien qu'ANNA-PALM puisse reconstruire des images super-résolutives à partir de images sous-échantillonnées, pour les cellules vivantes l'entraînement du modèle et l'évaluation de la qualité des reconstructions sont plus difficiles en raison de l'absence de vérité terrain. Pour relever ces défis, j'ai développé une méthode pour créer des vérités terrain super-résolutives dynamiques à partir d'images SMLM statiques. J'ai appliqué cette stratégie à des données simulées ainsi qu'à des données expérimentales de microtubules. Ensuite, j'ai étendu l'architecture d'ANNA-PALM à des données 3D dont la troisième dimension est le temps, afin d'exploiter l'information temporelle. J'ai utilisé des simulations de microtubules en mouvement pour évaluer quantitativement la qualité des reconstructions par cette approche, en comparaison avec la méthode ANNA-PALM 2D originale, et en fonction de la vitesse des structures et des taux de localisation. Les résultats montrent que l'incorporation d'informations temporelles améliore considérablement la qualité des reconstructions et suggèrent que des films super-résolutifs de grande qualité de microtubules dans des cellules vivantes peuvent en principe être reconstruits avec des paramètres d'imagerie SMLM réalistes. Une fois validée sur des données expérimentales, notre méthode devrait ouvrir la voie à l'imagerie super-résolutive de structures dynamiques en cellules vivantes.

Mots clés : [Apprentissage profond, SMLM, Imagerie en cellules vivantes, Microscopie haute résolutive]